Three essays on the transformative role of technology in financial markets

Khaladdin Rzayev

A thesis submitted to

University of Edinburgh Business School

University of Edinburgh

For the degree of Doctor of Philosophy

October 2019

Supervisor:

Gbenga Ibikunle

To Narmin and Elvin, of course.

Papers adapted from thesis

2017 **Rzayev, K.**, and Ibikunle, G. (2017). Order aggressiveness and flash crashes. Conference Paper and Presentation at Market Microstructure Seminar, 19th – 24th June 2017, Université Della Svizerra Italiana. **Peer-reviewed**.

2017 **Rzayev, K.**, and Ibikunle, G. (2017). Order aggressiveness and flash crashes. Conference Paper and Presentation at 2nd European Capital Markets Workshop, 9th July 2018, Cass Business School. **Peer-reviewed**.

2018 **Rzayev, K.**, and Ibikunle, G. (2018). A state-space modelling of the information content of trading volume. **Journal of Financial Markets**, *Forthcoming*.

2018 **Rzayev, K.**, and Ibikunle, G. (2018). A state-space modelling of the information content of trading volume. Conference Paper and Presentation at the American Finance Association 2019 Annual Meeting, 4th – 6th January 2019, Atlanta.

2018 **Rzayev, K.**, and Ibikunle, G. (2018). A state-space modelling of the information content of trading volume. Conference Paper and Presentation at the European Financial Management Association 2018 Annual Meeting, 27th – 30th June 2018, Milan.

2018 **Rzayev, K.**, and Ibikunle, G. (2018). A state-space modelling of the information content of trading volume. Conference Paper and Presentation at the 8th Behavioural Finance and Capital Markets Conference, 24th – 26th September 2018, La Trobe University

2019 **Rzayev, K.**, and Ibikunle, G. (2019). Need for Speed? International transmission latency, liquidity and volatility. Conference Paper and Presentation at 3rd European Capital Markets Workshop, 5th July 2019, DCU Business School**. Peer-reviewed**.

2019 **Rzayev, K.**, and Ibikunle, G. (2019). Need for Speed? International transmission latency, liquidity and volatility. Conference Paper and Presentation at the American Finance Association 2020 Annual Meeting, 3rd – 5th January 2020, San Diego. **Peer-reviewed**.

2019 **Rzayev, K.**, and Ibikunle, G. (2019). Need for Speed? International transmission latency, liquidity and volatility. Conference Paper and Presentation at the Paris Financial Management Conference, 16th – 18th December 2019, Paris. **Peer-reviewed**.

2019 **Rzayev, K.**, and Ibikunle, G. (2019). Need for Speed? International transmission latency, liquidity and volatility. Conference Paper and Presentation at the 27th Conference on the Theories and Practices of Securities and Financial Markets, 6th – 7th December 2019, Kaohsiung. 27th SFM 2nd Best Paper Award. **Peer-reviewed**.

Acknowledgement

I would like to express my sincere gratitude and special appreciation to my PhD supervisor, Dr Gbenga Ibikunle for the continuous support and contribution. It has been an honour to be his PhD student. His advice on both research as well as on my career/job hunting have been priceless. It is safe to say that without his guidance and constant feedback this PhD would not have been achievable. I would also like to thank my examiners, Professor Ian Marsh and Dr Angelica Gonzalez for their insightful comments and inspiration.

My special words of thanks should go to Professor Jo Danbolt, Dr Angelica Gonzalez, Professor Christine Cooper and Dr Aristotelis Boukouras for all the assistance they have provided me during my job search. My thanks also go out to the support I received from our IT and PhD support staff, especially Ms Angela Chitzanidi and Ms Susan Keatinge.

I am also very grateful to all my friends and colleagues in Accounting and Finance research group at the University of Edinburgh for all the fun we have had during my PhD journey. In particular, my heartfelt thanks to Dr Yuxin Sun, Dr Tom Steffen, Dr Ben Sila, Suhee Kim, Bagus Mudiantoro, Ziwei Xu, Saleh Ahmadov, Sabina Savadova, Giorgi Amirkhanashvili, Cavid Nadir, Farid Ibrahimov and Gunay Huseynalizadeh.

Lastly, and most importantly, I would like to thank my family for their encouragement and love. For my parents, brother and sister who incented me to strive towards my goal. Most of all for my loving and supportive wife Narmin whose support during my PhD is so appreciated. And to my wonderful son, Elvin for providing unending inspiration and being such a good little baby.

Contents

List of Abbreviations

ACD	**Autoregressive Conditional Duration**
AR	**Access Rule**
BVC	**Bulk Volume Classification**
CBOE	**Cboe Stock Exchange**
CCP	**Central Counterparty**
CS	**Component Share**
DiD	**Difference-in-Difference Framework**
DMMs	**Designated Market Makers**
DSs	**Designed Sponsors**
ESMA	**European Securities Market Authority**
ETFs	**Exchange Traded Funds**
HFMs	**High Frequency Market Makers**
HFTs	**High Frequency Traders**
HFT	**High Frequency Trading**
ILS	**Information Leadership Share**
IS	**Information Share**
ITS	**Inter-market Trading Systems**
TL	**Information Transmission Latency**
LSE	**London Stock Exchange**
MDH	**Mixture of Distribution Hypothesis**
MDR	**Market Data Rule**
MiFID	**Markets in Financial Instrument Derivatives**
NASDAQ	**Nasdaq Stock Market**
NBBO	**National Best Bid or Offer**
NYSE	**New York Stock Exchange**
OPR	**Order Protection Rule**
PS	**Partition Specific Gateway**
Reg NMS	**Regulation National Market System**
RIC	**Reuters Identification Code**
RMM	**Regulated Market Maker**
SEC	**U.S. Securities and Exchange Commission**
SLPs	**Supplemental Liquidity Providers**
SPR	**Sub-Penny Rule**
SROs	**Self-Regulatory Organizations**
TRTH	**Thomson Reuters Tick History**
VECM	**Vector Error Correction Model**
VPIN	**Volume-Synchronized Probability of Informed Trading**
XSE	**Xetra Stock Exchange**

List of Figures

List of Tables

Abstract

Financial markets are vital for capital allocation and as a consequence, for the wider economy. They perform two primary functions: liquidity and price discovery. Liquidity refers to the ability to trade large quantities of an instrument quickly, and with relatively little price impact. Therefore, it offers investors the flexibility to make investment decisions. Price discovery encompasses the price formation process in financial markets and is, therefore, critical for efficient capital allocation. Both these functions are linked to the functioning of the wider economy. Over the last decade, financial markets have been transformed with the help of technology and are now a completely different proposition. Specifically, technological advancements, such as high frequency trading (HFT), have altered the structure of financial markets, the strategies of traders, and the liquidity and price discovery processes. These changes and developments have ignited a heated debate among academics and regulators. While some researchers claim that HFTs increase the market efficiency by improving the liquidity and price discovery (see as an example, Brogaard et al., 2014b), others argue that they create adverse selection risks for slow traders and contribute to market instability by exacerbating illiquidity shocks, such as flash crashes (see as an example, Kirilenko et al., 2017). Motivated by these contrasting views, this thesis investigates these issues, and is therefore situated at the intersection of financial markets, technology and regulations. It specifically examines the topical issues around the transformative role of technology in financial markets by adopting novel and unique approaches. In the first study, I present a novel framework illustrating the links between order aggressiveness and flash crashes. My framework involves a trading sequence beginning with significant increases in aggressive sell orders relative to aggressive buy orders until instruments' prices fall to their lowest levels. Thereafter, a rise in aggressive buy orders propels the prices back to their pre-crash levels. Using a sample of S&P 500 stocks trading during the May 6 2010 flash crash, I show that the framework is correctly

specified and provides a basis for linking flash crashes to aggressive strategies, which are found to be more profitable during flash crashes. The second study is a methodological contribution to the financial econometrics literature, in which I propose a state space modelling approach for decomposing a high frequency trading volume into liquidity- and information-driven components. Using a set of high frequency S&P 500 stocks data, I show that the model is empirically relevant, and that informed trading is linked to a reduction in volatility, illiquidity and toxicity/adverse selection. Furthermore, I observe that my estimated informed trading component of volume is a statistically significant predictor of one-second stock returns; however, it is not a significant predictor of one-minute stock returns. I show that this disparity can be explained through the HFT activity, which eliminates pricing inefficiencies at high frequencies. The third study exploits the impact of the international transmission latency on liquidity and volatility by constructing a measure of the transmission latency between exchanges in Frankfurt and London and exploiting speed-inducing technological upgrades. I find that a decrease in the transmission latency increases the liquidity and volatility. In line with the existing theoretical models, I show that the amplification of liquidity and volatility is associated with the variations in adverse selection risk and aggressive trading. I then investigate the net economic effect of high latency, which lead to the finding that the liquidity deterioration effect of high latency dominates its volatility reducing effect. This implies that the liquidity enhancing benefit of increased trading speed in financial markets outweighs its volatility inducing effect.

Keywords: trading volume, permanent component, transitory component, market quality, time series models, state space modelling, transmission latency, microwave connection, high-frequency trading, liquidity, volatility, order aggressiveness, algorithmic trading, asymmetric information, extreme price movement, high-frequency data, logistic regression.

1. Introduction

1.1 A summary of the thesis

Technological advancements have significantly altered the nature of trading in financial markets; the structure of financial markets, the strategies of traders, and the mechanisms of liquidity and price discovery are completely different today. One of the most important implications of the evolution of technology in financial markets is HFT, which has grown tremendously over the past decade and now drives at least half of all the trading in major financial markets (see Brogaard et al., 2014b). In response to the changes in the structure of financial markets, and in order to ensure a stronger level of competition and a more efficient pricing system, new trading rules and regimes, such as the Regulation National Market System (Reg NMS) and the Markets in Financial Instrument Directives (MiFID), have been implemented. These new trading rules have further altered critical market processes and mechanisms, resulting in an even more complex market structure. Therefore, it is vital to examine the effects of these developments in financial markets. This thesis directly addresses the questions arising from these developments, and consequently fills a yawning gap in the existing literature.

The evidence regarding the impact of technological advancements on financial markets has hitherto been inconsistent. While some studies show the positive impact of HFTs on liquidity and price discovery (see as examples Brogaard et al., 2014b; Hendershott et al., 2011; Hoffmann, 2014), others suggest that HFTs can increase the adverse selection (and hence, deteriorate liquidity) and contribute to flash crashes (see as examples, Biais et al., 2015; Foucault et al., 2016; Foucault et al., 2017; Hendershott and Moulton, 2011; Kirilenko et al., 2017). Motivated by these contrasting predictions/findings, I conduct three studies examining the questions linked to the recent major developments and challenges in financial markets. Specifically, I investigate the role of aggressive traders in flash crashes in HFT-driven markets, the effects of informed/liquidity trading on financial markets at high frequency, and the relationship between speed/latency and the market quality. I directly contribute to the literature through the investigations described below. The contributions and research questions of the thesis are briefly discussed in the following paragraphs.

Chapter 2 contains an investigation of the contribution of aggressive orders to flash crashes by proposing a novel framework illustrating the links between order aggressiveness and flash crashes. The framework is a hypothetical three-stage trading strategy, which generates a similar price evolution process to flash crashes and provides more profit for aggressive traders under some conditions. The framework motivates three important predictions. First, excessive sell order aggressiveness creates a downward pulling effect on stock prices prior to and during the first half of flash crashes. Then, in the second stage of flash crashes, the balance of order aggressiveness shifts to the buy side which generates an increasing pressure on stock prices. Second, the framework predicts that order aggressiveness prior to flash crashes is linked to it. Third, aggressive orders are more profitable during extreme price movements; thus, traders tend to behave more aggressively during these periods. Thereafter, I test these predictions using ultra-high frequency trading data for the components of the S&P 500 stock index impacted by the May 6 flash crash, i.e. the biggest flash crash in the financial market history. The empirical results are consistent with the framework's predictions. First, I documented that, as predicted by the framework, during the first half of the flash crash stock prices going down as a result of excessive aggressive order imbalance favouring the sell side. Then, the aggressive order imbalance shifts to the buy side, which pushes the prices back to the pre-flash crash level. Second, the study shows that increased build-up order aggressiveness contributed to the May 6 2010 flash crash. Third, the findings suggest that aggressive trading is significantly more profitable during flash crashes than normal trading periods; the difference is both economically and statistically significant. Specifically, an informed/aggressive trader could earn up to a cumulative return in excess of 1,482 basis points (bps), based on my analysis of a sample of flash crash-affected stocks. This chapter differs from previous studies (see as examples, Easley et al., 2011; Jacob Leal et al., 2016; Kirilenko et al., 2017) in at least two aspects. First, my framework makes no assumptions regarding the liquidity constraints in the market, and, therefore, is more consistent with the reality of the HFT-driven financial markets. Second, while some studies show the contribution of aggressive traders to the flash crash (see as an example Mcinish et al., 2014), to the best of my knowledge, the theoretical framework and empirical analysis I present in this study are the first to explain the economic motivation for aggressive trading. Explicitly, the results suggest that higher profitability of aggressive orders may be a driver of the traders' aggressiveness during flash crashes.

In Chapter 3, I propose the state space modelling approach for decomposing a high frequency trading volume into liquidity- and information-driven components. More explicitly, I demonstrate trading volume as a sum of two unobserved series: a non-stationary permanent series and a stationary transitory series. I argue that uninformed/liquidity traders and informed traders can be respectively modelled by using estimated permanent and transitory components from the state space approach. I then test this argument, i.e. the empirical relevance of the proposed state space approach, by developing a set of univariate analysis and multivariate regression models. Using a set of high frequency S&P 500 stocks data, I find that the model is empirically valid and that informed trading reduces volatility, illiquidity and toxicity/adverse selection. The validity of the empirical relevance of the model is further confirmed by the price efficiency test prescribed in Chordia et al. (2002; 2008). Specifically, I demonstrate that one-second stock returns can be predicted from the estimated transitory component of the volume, which implies that the transitory component of the state space approach can indeed provide a signal about informed trading. However, I observe that both Chordia et al.'s (2008) and my informed trading proxies are not able to predict one-minute stock returns. By using the NASDAQ-provided HFT data, I show that HFTs eliminate pricing inefficiencies at minute intervals.

This chapter makes the following contributions to the existing literature. First, the proposed state space modelling approach is fundamentally different from the existing decomposition methods and has significant economic worth. Second, I examine the role of informed trading activity in market quality by using a relevantly new proxy – market toxicity – which is not well-documented in the literature. Finally, I present new evidence on the speed of price adjustment in HFT-driven markets.

The study reported in Chapter 4 involves estimating the information transmission latency (TL) between a home exchange in Frankfurt and a satellite exchange in London; subsequently, its effect on the liquidity and volatility of cross-listed stocks in the satellite market is examined. This study provides important empirical findings. First, I find that the TL between Frankfurt and London is 3–5ms. Second, and more importantly, my findings suggest that while higher transmission speed improves the market quality by increasing liquidity, it nevertheless raises volatility, and thus, harms the market quality. By adopting a quasi-experimental setting, I show that the positive relationship between the transmission speed and liquidity/volatility is causal. Furthermore, I show that channels proposed by various theoretical models, i.e. *adverse selection avoidance* and *aggressiveness*, can explain the main findings – the positive relationship between speed and liquidity/volatility. Finally, I report that the net economic impact of high speed is positive for the economy.

My contributions to the existing literature are as follows. First, the study is the first to empirically estimate the TL between the two biggest European financial centres, Frankfurt and London. The highly fragmented characteristic of the European trading venues makes it particularly important for this region. Second, I provide causal evidence on the direct impact of speed on volatility, which is unclear in the literature. Third, the latency metric I constructed, TL, is more relevant for fragmented financial markets, and it implies that it has further economic insight and is more consistent with the reality of trading in modern financial markets.

Finally, the study investigates the net economic effect of high speed, which is unclear in the HFT literature and thus, has many important implications for the wider economy.

Chapters 2, 3 and 4 individually focus on a single issue, as stand-alone studies. I use the US and European stocks as sample data for empirical tests across the three chapters. The next part of this section discusses the literature related to my studies.

1.2 Financial markets: microstructure and technology

Financial market microstructure is the branch of financial economics, dedicated to the study of pricing dynamics of financial securities and the mechanisms used for trading those securities. According to Madhavan (2000, p.205) *"market microstructure studies the process by which investors' latent demands are ultimately translated into prices and volumes"*. Whereas O'Hara (1995) views market microstructure as involving an investigation of the way in which price discovery is affected by trading mechanisms. Furthermore, O'Hara (2003) argues that market microstructure mainly analyses two important functions of financial markets: liquidity and price discovery. Financial markets impact the wider economy through these two channels, and thus, well-functioning financial markets are essential for the stability of the financial system and long-term economic growth. Hence, examining the microstructure of financial markets is vital and has many important implications.

O'Hara (2015) suggests that technological advancements have altered the entire system of financial markets; therefore, investigating the implications of these changes for market microstructure is important. The author further argues that researchers must change their directions to reflect the fundamental differences in the new market microstructure, i.e. the high frequency market microstructure.

While the intersection of financial markets and technology has several implications, in this section, I provide the literature review for two of them, namely HFT and market fragmentation, as they are directly linked to the studies investigated in this thesis; the most important result of the intersection of financial markets and technology is HFT. The evidence of the impact of HFT on financial markets has so far been inconsistent. While some studies find that HFTs improve the liquidity and price discovery process (see as examples, Brogaard et al., 2014b; Hendershott et al., 2011), others suggest that they increase the adverse selection risks for slow market makers and exacerbate price crashes (Biais et al., 2015; SEC, 2010).

For example, Jovanovic and Menkveld (2016) propose a model where HFTs interact with investors in a limit order book; the authors assume that HFTs are equipped with hard information about the common values of assets. The study shows that HFTs use their speed advantage to avoid being adversely selected and thereby, increase liquidity and reduce the adverse selection in financial markets. Furthermore, their findings suggest that well-designed double auctions maximise the welfare impact of HFTs. Similar to Jovanovic and Menkveld (2016), Hoffmann (2014) also shows that HFTs can reduce adverse selection by avoiding it and increasing the trade. However, the study finds that if the proportion of HFTs is not given exogenously, they can cause social welfare loss. The predictions of the abovementioned studies are empirically confirmed by Hendershott et al. (2011). Hendershott et al. (2011) conducts one of the first studies that examined the role of Algorithmic Trading (AT) in the market quality. By using the New York Stock Exchange's automated quote dissemination in 2003 as an exogenous shock, the study finds that AT improves the liquidity and reduces (increase) the noise (efficient) price discovery. Similar to Hendershott et al. (2011), Brogaard et al. (2014b) also find that HFTs increase (reduce) the efficient (noise) price discovery. The study conducts an analysis using the NASDAQ-provided HFT data, and by employing the state space approach, described in Menkveld et al. (2007), to decompose the price discovery into efficient and noise parts.

The negative impact of HFT on financial markets is modelled by Biais et al. (2015), Foucault et al. (2016), and Foucault et al. (2017). Biais et al. (2015) develop a simple model where financial institutions have heterogeneous private valuations and private information.

While heterogeneous private valuations create additional gains from trades, private information causes adverse selection for slow traders. More importantly, the authors show that as a result of the negative externality of fast trading, i.e. adverse selection impact, investments in fast trading exceed their utility and hence, fast trading reduces social welfare. The study proposes various approaches that allow all traders, slow and fast, to obtain social gains. Foucault et al. (2016) build on Kyle (1985) and allow one speculator (slow or fast) and one competitive dealer in the model. The study shows that when the speculator is fast, the dealer has a higher adverse selection risk and thus, the market is less liquid. Furthermore, in the model, the speed of price discovery is independent on the types of speculators (fast or slow), since the fast speculator's trades have less (more) information content for long-run (short-run) price changes, and these two effects offset each other. Similar to Biais et al. (2015) and Foucault et al. (2016), Foucault et al. (2017) also show that HFT can raise adverse selection in financial markets. The study argues that there are two possible channels for arbitrage opportunities: (1) demand and supply shocks (price pressures) and (2) asynchronous adjustments in asset prices. The study models the latter channel and demonstrates that while HFTs improve price efficiency in financial markets, they increase the adverse selection for dealers. The prediction of the mentioned theories, i.e. positive relation between HFT and the adverse selection cost, is empirically confirmed by Hendershott and Moulton (2011). Hendershott and Moulton (2011) find that although automation trading systems make the prices more efficient, these systems increase adverse selection and thereby, raise the cost of immediacy.

Kirilenko et al. (2017) empirically show that HFTs behave differently than traditional market makers during the flash crash. More specifically, the study reveals that during the May 6 flash crash, the trading strategies of HFTs are based more on quote sniping and latency arbitrage rather than traditional market making strategies. This implies that HFTs may adversely select/stale quotes of slower market makers during extreme price movements; therefore, while HFTs did not trigger the flash crash, they contributed towards it (see also SEC, 2010). By contrast, Brogaard et al. (2014a) show that although the profits of HFTs are high during extreme price movements, there is little evidence of HFTs causing extreme price movements. Furthermore, the study states that while HFTs provide liquidity when a single stock experiences extreme price movements, they demand liquidity in multi stocks case, i.e. when there are simultaneous extreme price movements in multiple stocks. In addition to these studies, Easley et al. (2011), Kyle and Obizhaeva (2016), Madhavan (2012), and Menkveld and Yueshen (2017) show other technology-linked factors that are susceptible to contribute to flash crashes. Further literature about the impact of HFTs on financial markets and wider economy can be found in Menkveld (2016).

Another noteworthy outcome of the intersection of financial markets and technology is market fragmentation. Market fragmentation implies that stock trading is spread across multiple financial markets. Although the impact of this phenomenon on market quality has been extensively investigated in the market microstructure field, the studies do not have a consensus. On the one hand, O'Hara and Ye (2011) show that market fragmentation improve the market quality by lowering the transaction costs and increasing the execution speeds (see also Battalio, 2012). On the other, Bennett and Wei (2006) find that when stocks switch from a more fragmented market (NASDAQ) to a less fragmented one (NYSE), they experience the better market quality and more efficient price discovery.

While market fragmentation is not directly studied in this thesis, one of the most important and HFT-related implications of it – the transmission latency between financial markets – is investigated. Shkilko and Sokolov (2016) examine liquidity when weather-related episodes disrupt the microwave transmission between the lead and lag markets; they find that adverse selection and trading costs decline and the liquidity improves during these periods, implying that a higher transmission speed between trading venues may harm the market quality. Menkveld and Zoican (2017) and Baron et al. (2018) also investigate the role of speed differentials in financial markets, however, by focussing more on a consolidated market structure rather than a fragmented one. More explicitly, Menkveld and Zoican (2017) theoretically model the HFT arms race. The study extends Budish et al.'s (2015) model by adding the impact of exchange speed to it. Their model explains the impact of exchange speed on the market quality by using two contrasting channels. In the first channel, high frequency market makers (HFMs) obtain speed advantages as a result of the improvements in exchange latency; by doing so, they avoid being adversely selected and improve liquidity. In the second case, high frequency speculators (HFSs) obtain speed advantages as a result of improvements in the exchange latency; thereby, they adversely select the HFMs' orders, leading them to widen the bid-ask spread to compensate for the increased adverse selection risk. The widening of the bid-ask spread implies a deterioration in liquidity. The authors further analyse the net effect of the exchange speed improvements and show that it depends on the news-to-liquidity ratio of the asset. Specifically, Menkveld and Zoican (2017) present that the former (latter) channel is strong when the news-to-liquidity ratio is high (low). Baron et al. (2018) study the relationship between speed differentials and the trading revenue in a single market and find that the fastest firms (HFTs) can generally earn the largest trading revenues.

1.3 Background

In this section, I provide detailed information on the structures of the markets investigated in the thesis. I further discuss a few of the key aspects and regulations that should be understood before reading the studies presented in this thesis.

1.3.1 U.S. Markets: The New York Stock Exchange (NYSE) and the Nasdaq Stock Market (NASDAQ)

As of January 2007, the NYSE is a hybrid market, and in terms of liquidity, it is considered the best exchange in the US.[1] The NYSE's market model mainly consists of three players: (1) the designated market makers (DMMs), (2) the floor brokers, and (3) the supplemental liquidity providers (SLPs). The DMMs improve price discovery and liquidity by maintaining well-coordinated markets. Furthermore, the DMMs minimise the order latency by matching the incoming orders from traders. Previously, the NYSE used the specialists instead of the DMMs. In order to eliminate the issue of front running, the NYSE replaced specialists by the DMMs. Explicitly, while the specialists could see all incoming orders, the DMMs do not have an advanced view and they thus behave like a regular market participant. As of 2017, there are five DMMs in the NYSE.

The floor brokers execute transactions, i.e. buying and selling stock, on behalf of their firm's clients. Their firm should be the member firm of the NYSE. The SLPs are established by the NYSE to improve liquidity and they are electronic, high-volume members of the marketplace. One of the most important obligations of the SLPs is to keep a bid price at the National Best Bid or Offer (NBBO) level, at least 10% of the trading day. Six order types are commonly used in the NYSE. They are immediate or cancel, displayed limit, displayed limit reserve, non-displayed limit, auction, and other. Almost half of all orders are immediate or cancel and displayed limit orders.

Similar to the NYSE, the NASDAQ is also considered a hybrid market. The NASDAQ is considered the largest trading venue by volume in the US.[2] While both of them are hybrid markets, in contrast to auction market system of the NYSE (at market open and close), the NASDAQ is a purely dealer's market. More explicitly, in the NASDAQ market, security

---

[1] https://www.nyse.com/markets/nyse
[2] http://www.nasdaqtrader.com/Trader.aspx?id=TradingUSEquities

transactions are executed through dealers. However, at market open and close periods, market participants can buy and sell securities from each other in the NYSE. Another important difference is in terms of their market models. As already noted, the NYSE relies on the DMMs and the SLPs for providing liquidity and efficient trading/price discovery processes. However, the NASDAQ is a purely electronic market and therefore, has different “traffic control” mechanism, i.e. the electronic market makers. In contrast to the NYSE, in the NASDAQ exchange, each stock has more than one market maker. The NASDAQ’s market maker buys and sells securities at the NASDAQ’s prices and can execute transactions for his/her own account as well as his/her client’s account. The NASDAQ uses the price-time priority rule to execute transactions and has various types of orders, such as intermarket sweep orders, post-only orders, supplemental orders, and more.

1.3.2 European Markets: Deutsche Börse Xetra (Xetra) and Cboe European Equities (CBOE)

Xetra is a main German stock exchange where 90% of security trading at all German exchanges are executed. Furthermore, 30% of all exchange traded funds (ETFs) in Europe are transacted at this trading venue. Xetra uses “continuous trading with auctions” for securities; this model is a combination of continuous trading and auctions. Through the continuous trading mechanism, liquid securities are executed immediately at the current market price. In addition to this continuous trading system, three auctions are organised during the day. By adopting auctions, Xetra aims to determine the price level for the exchange which can be used by institutional investors for valuing their trading positions. In addition to this system, Xetra offers the designated sponsors (DSs). The primary aim of the DSs is to provide liquidity for less liquid securities. Xetra also has the regulated market makers (RMMs) who provide liquidity to the market and have similar obligations to the DSs. The major difference between the RMMs and the DSs is their requirements. More specifically, while the RMMs should include a presence of at least 50% during continuous trading, this number is 90% in equities and 80% in ETFs for the DSs. The basic types of orders used by market participants in Xetra are market order, limit order, stop market order, stop limit order, and trailing stop order.

CBOE is the largest European trading venue in terms of the value traded, and was formed through the merging of BATS Europe and Chi-X Europe in 2011. CBOE offers an excess to 18 major European financial markets, and more than 6,000 securities are traded through this venue.

CBOE continuously accepts orders from 08:00 to 16:30 and further, has a periodic auctions book which aims to publish the prices and order quantity, prior to the order execution. CBOE uses the post trade model and offers the central counterparty (CCP) to provide liquidity to the venue. This market participant is used by trading participants to clear their trades. More specifically, the CCPs can be a buyer (seller) to each seller (buyer). This type of a trading model allows participants to trade and clear their trade with minimum cost and counterparty risk.

CBOE operates two lit (BXE-lit and CXE-lit) and two dark order books (BXE-dark and CXE-dark). Between these four order books, CBOE holds nearly 25% of the daily equities trading in European markets. The lit order books have four categories of orders: (1) visible orders, (2) interbook orders, (3) hidden orders and (4) peg orders. Each order category has different order types. For example, visible orders have visible limit order, post only order, and reserve order types. There are two order categories in the dark order books: minimum acceptable quantity and midpoint peg orders.

1.3.3 Regulations: The Regulation National Market Systems (Reg NMS) and the Markets in Financial Instrument Derivatives (MiFID)

As previously noted, the new trading rules and regimes have been proposed in response to the changes in the structure of financial markets. In this section, I discuss two important regimes, the Regulation National Market System (Reg NMS) and the Markets in Financial Instrument Directives (MiFID), which have been respectively implemented in the US and European financial markets.

According to SEC (2005), Reg NMS has four key provisions: (1) the order protection rule (OPR), (2) the access rule (AR), (3) the sub-penny rule (SPR), and (4) the market data rules (MDR). The OPR has been designed to protect limit orders against trade-throughs – a trade-through is an order with a suboptimal price. More specifically, these orders are executed at non-optimal prices, even if there is a better price on the same or other exchanges. It is important to note that the OPR protects only the displayed and automatically accessible orders. The SEC offers the OPR to strengthen price discovery and market efficiency by protecting the limit orders. The AR has been proposed to set better standards for ruling access to quotations in securities. This rule is very vital for the OPR, as market participants and trading centres need to have effective access to prices to protect them against a trade-through. The AR rule provides more efficient access to quotations using various ways. For instance, the rule allows participants to use private linkages in addition to a collective linkage facility. More specifically, before the AR rule, traders could use only a collective linkage facility, such as inter-market trading systems (ITS), to gain access to the quotations of exchange-listed stocks. These collective linkages are generally offered by trading venues. However, following the success of private linkages in electronic markets (for example, NASDAQ), the SEC decided to use the same connection systems for traditional exchange-listed stocks. In contrast to ITS, private linkages are offered by various connection providers and aim to obtain better access to the quotations displayed by exchanges.

The SPR was proposed to prevent market participants from "stepping ahead" of the displayed limit orders through trivial amounts. More specifically, the rule implies that exchanges cannot accept quotes that are prices in increments of less than \$0.01, if the quote price is more than \$1.00. The main aim of the rule is to improve liquidity by protecting liquidity providers from losing their priority owing to trivial amount changes in order prices. The MDR is designed to fairly allocate the market data revenues of self-regulatory organisations (SROs)[3] and, by doing so, promote the wide availability of market data. Before the MDR, the market data revenue was allocated according to the number of trades reported by the SROs. However, the MDR allocates more revenue to SROs who provide greater contributions to the best displayed quotes. Furthermore, in addition to the required best quotes and trades data distributed by exchanges through the different plans, the MDR allows exchanges to distribute their own data independently. This further promotes the wide availability of market data.

While the Reg NMS's OPR mandates the best execution by enforcing trade-through provisions, the European equivalent of the all-encompassing financial regulation, the MiFID, does not enforce trade-throughs. However, the European regulation/MiFID requires that investment firms ensure the best execution on behalf of their firms. This somewhat differs from the US regulation/Reg NMS, requiring that platforms forward orders to other platforms offering the best execution.

The MiFID has been designed to promote transparency, competition and integration in European financial markets. Two levels of this regulation have been implemented: MiFID and MiFID II. The MiFID has been effective from November 2007. The first important implication of the MiFID is pre- and post-trade transparency. Pre-trade transparency allows market participants to continuously monitor transactions and quotations; thereby, this transparency provides traders with the opportunity to immediately obtain new information about the fundamental value of assets. It is important to note that pre-trade transparency has been applied to lit order books only. Post-trade transparency requires that market participants report their post-trade information within three minutes of the relevant trade.

[3]SROs are multilateral trading platforms such as NYSE.

Another important implication of the MiFID has been proposed for market fragmentation. The MiFID has been designed to promote competition, and this competition leads to more fragmented financial markets in Europe. As pointed out, this implication raises an important difference between the MiFID and the Reg NMS. Specifically, while the Reg NMS requires orders to be executed at the best bid and ask prices – even if these best prices are available at different markets – the MiFID does not have this requirement. Rather, the MiFID requires investment firms to take all the necessary steps for the best execution, considering the costs, speed and likelihood of execution. Therefore, under the MiFID, orders might sometimes be executed at non-optimal prices.

The MiFID II has been proposed to address the shortcomings of the MiFID, considering the consequences of the 2008 financial crisis and was implemented in January 2018. One of the important implications of the MiFID II covers the HFT/AT. According to the rule, HFTs should disclose their algorithms and test them in specific environments prior to trading in financial markets. Furthermore, all HFTs that trade in European markets should register as investment firms. To prevent market participants from leaving HFTs during stressed periods, MiFID II has set clear "exit" conditions. The MiFID II has also introduced cancellation fees for HFTs to mitigate the adverse impact of some HFT strategies, such as quote stuffing. In addition to the HFT-related regulations, the MiFID II sets two important limits for dark trading. First, in any trading venue, the dark trading volume for each stock should not be more than 4% of the total trading volume for that particular stock in that particular venue. Second, the dark volume of any stock in all trading venues cannot exceed 8% of the total trading volume of that particular stock in all venues.

2. Order aggressiveness and flash crashes

2.1 Introduction

Flash crashes are characterised by high price volatility, a significant negative return in instruments' prices and are defined by a sharp price reversal (see Aldridge, 2010; Easley et al., 2011). The most notable flash crash in recent history occurred on May 6, 2010 (see Kirilenko et al., 2017). On this day, market indices such as the S&P 500, the Dow Jones Industrial Average, the Russell 2000, and the Nasdaq 100, fell significantly before rebounding within an extremely short period.

In the aftermath of the May 6 flash crash there has been a widespread concern that trading strategies commonly deployed by the fastest traders in financial markets – the so-called high frequency traders (HFTs) – induce or worsen price crashes.[4] Kirilenko et al. (2017) argue that, although there may be no evidence of HFTs causing the May 6 flash crash, they nevertheless exacerbated it by demanding immediacy. The immediacy demanded at a heightened pace in a liquidity-constrained environment appeared to have led to an unbearably high level of order flow toxicity, thereby worsening the price crash.[5] The aggressiveness of HFTs in demanding liquidity could therefore be argued to be a major contributing factor to the extent of the price crash recorded on May 6, 2010. However, to date, there has been no study directly linking order aggressiveness[6] to flash crashes, with no constraints placed on market agents. This chapter addresses this gap in the literature.

[4] About five months after the flash crash, on September 30 2010, the Commodity Futures Trading Commission (CFTC) and the Securities and Exchange Commission (SEC) released a study identifying an automated program executing the sale of 75,000 E-mini S&P 500 futures contracts as the main trigger for the flash crash (see SEC, 2010).

[5] Easley et al. (2011) highlight the key role played by order flow toxicity in the occurrence of the flash crash; they also propose a measure of order flow toxicity, which they call the Volume-Synchronized Probability of Informed Trading (VPIN).

[6] I define aggressive orders in line with the classification approach of Biais et al. (1995); specifically, aggressive orders are defined with respect to their sizes and tendency to cross the spread.

This chapter differs from previous studies (see as examples, Easley et al., 2011; Jacob Leal et al., 2016; Kirilenko et al., 2017) in at least two respects. Firstly, the links I draw between order aggressiveness and flash crashes make no assumptions regarding liquidity constraints in the market.[7] Secondly, although there are a few studies examining the role of trader aggressiveness in flash crashes (see as an example Mcinish et al., 2014), to my knowledge, the theoretical framework and empirical analysis I present in this study is the first to explain the economic motivation for aggressive trading. Specifically, I show that higher profitability of aggressive orders during flash crashes may be a driver of traders' aggressive trading behaviour during periods of extreme price movements.

My approach involves extending the approach of Menkveld (2013), developed to decompose the trading profit in a normal market environment into its spread and positioning components. Menkveld (2013) illustrates the decomposition of traders' profits by presenting two extreme cases – aggressive and passive market making trading strategies. The framework shows that traders adopting aggressive trading strategies incur losses during normal trading days and, therefore, the majority of traders – about 80% – tend to deploy passive market making trading strategies. The losses reported for aggressive traders on normal trading days is due to incoming market orders adversely selecting aggressive orders in the market (see also Glosten and Milgrom, 1985). My extension of this two-stage approach shows how aggressive trading strategies affect the price discovery process in financial markets.

The framework involves a trading sequence beginning with significant increases in aggressive sell orders relative to aggressive buy orders until instruments' prices fall to their lowest levels. Thereafter, a rise in aggressive buy orders propels prices back to their pre-crash levels. Using the predictions of the framework, I highlight the role of order aggressiveness in

[7] Jacob Leal et al. (2016) also develop an agent-based model of a limit-order book to show the impact of HFT on financial markets; their HFTs are assumed to deploy only predatory high frequency trading strategies (aggressive trading strategies). They conclude that aggressive HFTs are culpable in flash crashes. Consistent with Jacob Leal et al. (2016), Mcinish et al. (2014) show that the aggressive behaviour of Intermarket Sweep Orders contributed to the May 6, 2010 flash crash.

extreme price movements, such as flash crashes, and argue that order aggressiveness can contribute to flash crashes.[8] In this case, the framework shows that even in a liquid trading environment where there are no significant liquidity constraints, order aggressiveness can contribute to an environment of severe illiquidity such that prices become extremely volatile, as evident during the May 6, 2010 event.

Furthermore, the framework shows that profits in aggressive trading strategies are positive and large during extreme price movements such as flash crashes, and therefore the fraction and the number of aggressive orders should be higher during these periods when compared with normal trading periods. I decompose the profits of aggressive traders into their spread and positioning components and similar to Menkveld (2013), I show that traders are confronted with a position profit and, inevitably, a spread loss when they trade aggressively. However, unlike during normal trading periods, when markets are volatile, the position profit eclipses the spread loss, thus making aggressive trading ultimately profitable during periods of high price volatility. Since my framework involving a three-stage aggressive trading strategy, which results in a price collapse and a subsequent sharp price reversal, mimics the form of a flash crash, I argue that aggressive trading strategies can contribute to flash crashes (see also Mcinish et al., 2014). I test the foregoing arguments and framework predictions using ultra-high frequency trading data for the components of the S&P 500 stock index affected by the May 6 flash crash. The empirical results obtained are completely in line with the predictions of my framework.

Firstly, I find that a significant imbalance in order aggressiveness favouring sell orders ensues in the run-up to and during the flash crash. I document a significant increase in the number of aggressive sell orders relative to aggressive buy orders in the run-up to and during the flash crash until instruments' prices plummeted to their troughs. The increase in aggressive

[8] This argument is also motivated by the results of Griffiths et al. (2000) and Wuyts (2011), who show that aggressive orders have price impacts larger than those of other trades.

sell orders with no corresponding rise in aggressive buy orders precipitated the crash in instruments' prices. This finding is very important, since the total number of aggressive orders could be high; however, a significant price crash will only occur if aggressive sell orders significantly outstrip aggressive buy orders. This result is consistent with the official reporting following the flash crash (see SEC, 2010).

Secondly, I link the evolution of order aggressiveness to the flash crash within an econometric framework, showing that increased order aggressiveness is related to the May 6 2010 flash crash; hence, a build-up of aggressive orders ahead of the flash crash appears to be a contributory factor to the onset of the flash crash.

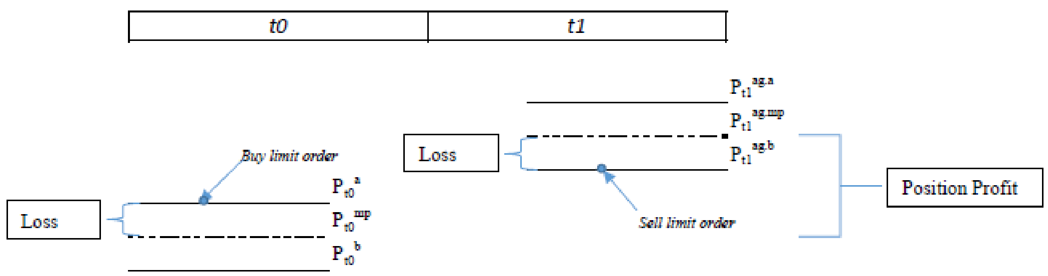
Thirdly, I show that aggressive trading is significantly more profitable during periods of high price volatility such as flash crashes, than during normal trading periods. The increase in profitability is economically meaningful. I find that an informed trader could earn up to a cumulative return in excess of 1,482 basis points (bps) based on my analysis of a sample of flash crash-affected stocks, this is significantly higher than possible during the non-flash crash periods. Consistent with this finding, the fraction of aggressive buy and sell orders during the May 6, 2010 flash crash is higher than the fraction of these kinds of orders during other periods under investigation. The actual number of aggressive sell and buy limit orders during the flash crash is also remarkably higher than during the surrounding periods (before and after the flash crash). The results are robust to alternative estimation approaches and model specifications, including estimation frequencies. Overall, the empirical results show that the framework is correctly specified and the arguments I present valid in the case of the flash crash I examine. Thus, my theoretical framework not only predicts the aggressive behaviour of HFTs during the flash crash, more importantly, it explains the economic intuition behind this aggressive behaviour. This is a further distinguishing element of this current contribution to the existing literature.

2.2 The approach

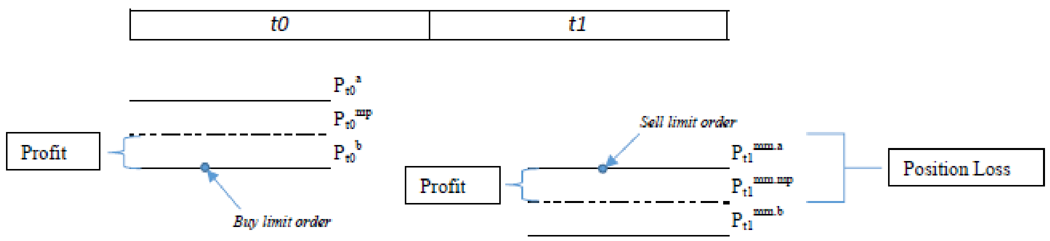
2.2.1 Motivation

Griffiths et al. (2000) and Wuyts (2011) find that aggressive orders generate larger price impacts. Given this finding, there is a case to be made for aggressive orders being culpable in inducing extreme price movements, such as flash crashes. However, this argument raises an interesting question about why aggressive orders do not always cause flash crashes, given that they are likely to be submitted repeatedly on any given day in financial markets. In order to examine this question and demonstrate the potential relationship between order aggressiveness and flash crashes, I extend the approach of Menkveld (2013). Following Sofianos (1995), Menkveld (2013) decomposes the profit of traders into two components: the spread component and the positioning component. Menkveld's (2013) framework focuses on two extreme cases involving aggressive trading on the one hand and passive market making on the other, by using a two-stage approach:

Aggressive trading strategy



Passive market making strategy



where P_{t0}^{a} is the ask price at time t_0, P_{t0}^{b} is the bid price at time t_0, P_{t0}^{mp} is the mid-price at time t_0, $P_{t1}^{ag.a}$, $P_{t1}^{ag.b}$ and $P_{t1}^{ag.mp}$ are the ask price, the bid-price and the mid-price at time t_1 under aggressive trading strategy, respectively, and $P_{t1}^{mm.a}$, $P_{t1}^{mm.b}$ and $P_{t1}^{mm.mp}$ are the ask price, the bid-price and the mid-price at time t_1 under passive (market-making) trading strategy, respectively.

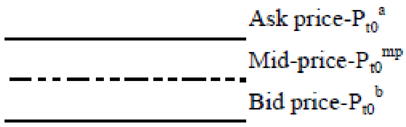
In the first extreme case, i.e. aggressive trading strategy, a trader consumes liquidity in order to pursue a fundamental value change, and then quickly follows this with a sell order. By submitting a buy limit order at the ask price and a sell limit order at the bid price, the trader will make a spread loss at t_0 and t_1, but will make a position profit at the end of the trading session. The trader will adopt this trading strategy if she expects a large position profit at the end of the trading session – this is necessary to compensate for the spread losses incurred from the first and second trading stages. However, adverse selection is a potential risk here, as the position profit could be negative if the trader's orders are adversely selected by an informed market order (see Glosten and Milgrom, 1985). Consistent with this argument, Menkveld (2013) finds that position profit is negative in the Dutch stock market during normal trading periods – periods of no or very low price volatility. In the second extreme case, i.e. the passive market making strategy, a trader acting as a market maker makes a profit from the spread in the first and second trading session, and a loss from her position at the end of trading.

In this chapter, I alter the strategies above and further extend the framework to decompose the profit of traders. Specifically, I employ a three-stage approach and alter the order of submitted orders to show the relationship between order aggressiveness and flash crashes; what this means is that while Menkveld's (2013) framework begins with a buy order, my approach begins with a sell limit order.

2.2.2 My three-stage approach

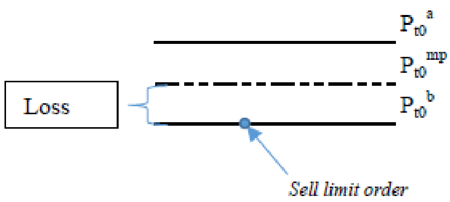
Trading at t_0

Traders submit sell limit orders at t_0 by following one of two trading strategies (passive and aggressive), while the subsisting bid and ask prices, with mid-price P_{t0}^{mp}, are set before traders come to the market:



$$P_{t0}^{mp} = \frac{P_{t0}^{a} + P_{t0}^{b}}{2} \tag{2.1}$$

I assume that a trader will submit a sell limit order at the prevailing best bid price if she wants to adopt an aggressive trading strategy, or a trader will submit a sell limit order at ask price is she wants to adopt a passive market making strategy. I focus on one of these extreme cases, an aggressive trading strategy, as I aim to illustrate the relationship between order aggressiveness and flash crashes. By submitting a sell limit order at the bid price, a trader will make a loss at t_0. The trading sequence is illustrated below:



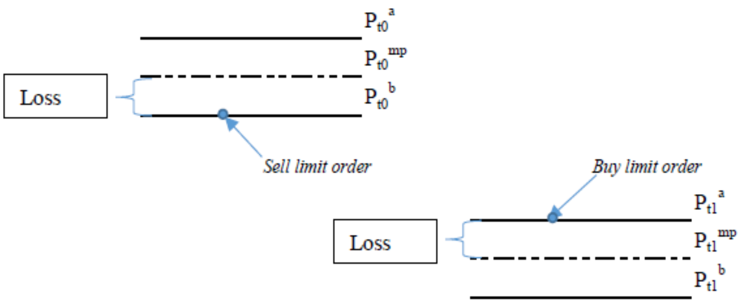
The loss of my hypothetical aggressive trader is therefore given as:

$$\pi_{t0}^{ag} = P_{t0}^{b} - P_{t0}^{mp} \tag{2.2}$$

Trading at t_1

Inevitably, different types of trading strategies in t_0 will have different impacts on ask and bid prices. This implies that bid and ask prices at t_1 will be different under either of the two extreme (passive and aggressive) strategies/cases. By submitting an aggressive sell limit order at t_0, the trader consumes liquidity, which in turn induces a price change. An aggressive trading strategy will therefore have a downward pulling effect on bid and ask prices, leading to bid and ask prices going down at t_1. However, if an aggressive order is adversely selected by an incoming informed market order, the price will go up at t_1 and the aggressive trader will incur a significant position loss. I therefore concentrate on the case where an aggressive order is not adversely selected (see Glosten and Milgrom, 1985). This assumption is in line with the recent literature, arguing that HFTs can predict adverse selection and therefore are able to avoid falling prey to it (see Hirschey, 2017; Hoffmann, 2014; Jovanovic and Menkveld, 2016). This assumption is critical for my framework to mimic the price evolution during a flash crash, i.e. price falling significantly from the level at t_0 to t_1.

During the second trading stage, the aggressive trader submits an aggressive buy limit order at the ask price:



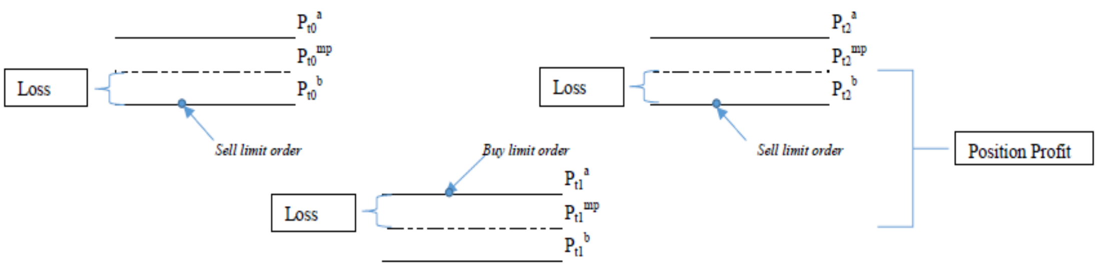
The submission of a buy order at the ask price will again lead to the trader incurring losses at t_1. The payout at this stage will be:

$$\pi_{t1}^{ag} = P_{t1}^{mp} - P_{t1}^{a} \tag{2.3}$$

Trading at t_2

As earlier stated, the deployed trading strategies will have varying impacts on ask and bid prices. An aggressive trading strategy at t_1 will generate an increasing pressure on bid and ask prices, thus bid and ask prices will appreciate subsequently at t_2 and reach initial position (t_0). If an aggressive order is adversely selected by an incoming informed market order, the price will go down at t_2 and the hypothetical trader will again incur significant position loss. Therefore, I again assume that an aggressive order is not adversely selected, to mimic the price evolution during a flash crash, i.e. price rebounds from t_1 to t_2 and attains the pre-flash crash level.

I further assume that the asset price at time t_2 will be equal to the asset price at time t_0. This is necessary for the sequence of events/price evolution to be consistent with a flash crash; i.e. a sudden/sharp fall in the price of an asset and a full rebound in price shortly afterwards:



By submitting a sell limit order at t_2's bid price, the aggressive trader makes a profit from her position and incurs losses from the bid-ask spread. Thus, her position profit and spread loss are as follows:

$$\text{Position Profit} \qquad \pi_{t2}^{ag.p} = P_{t2}^{mp} - P_{t1}^{mp} \tag{2.4}$$

$$\text{Spread Loss} \qquad \pi_{t2}^{ag.ba} = P_{t2}^{b} - P_{t2}^{mp} \tag{2.5}$$

$$\text{Total Profit} \qquad \pi_{t2}^{ag} = P_{t2}^{b} - P_{t1}^{mp} \tag{2.6}$$

To sum up these trading strategies thus far, I can examine the profitability of an aggressive trading strategy. By combining the above equations, I generate the following equations for an aggressive trading strategy, assuming that the bid and ask prices at time t_2 equal the bid and ask prices at time t_0 :

$$\begin{aligned}\pi^{ag} &= (P_{t0}^{b} - P_{t0}^{mp}) + (P_{t1}^{mp} - P_{t1}^{a}) + (P_{t0}^{b} - P_{t1}^{mp}) = \\ &= (P_{t0}^{b} - P_{t0}^{mp}) + (P_{t0}^{b} - P_{t1}^{a}) = \\ &= 2P_{t0}^{b} - P_{t0}^{mp} - P_{t1}^{a}\end{aligned} \tag{2.7}$$

Typically, a trader should pay the clearing fee and the aggressive exchange fee (usually imposed by exchanges on traders consuming liquidity) when she adopts an aggressive trading strategy. For simplicity, I assume that these fees are zero. As seen from Equation 2.7, the position profit of an aggressive trading strategy is high if there is a sharp reduction in the asset price at t_1 (P_{t1}^{a}). The interesting point is that this type of sharp reduction is consistent with the extreme price movements documented in the case of flash crashes. Therefore, I argue that although Menkveld (2013) shows that position profit is negative during normal trading days, it might be large and positive during extreme price movements. This implies that this kind of extreme price movement could be profitable for some traders. This argument is consistent with Brogaard et al. (2014a), who show that although HFTs do not cause extreme price movements such as flash crashes, these types of price movements is more profitable for HFTs. The argument raises an interesting question about why traders fail to always adopt an aggressive trading strategy and therefore obtain large and positive profitable positions or, more specifically, are there some other conditions that ensure that traders become aggressive? I argue that there should be other conditions, which are not necessarily directly linked with the traders themselves, which may lead to traders choosing an aggressive trading strategy. The important point to note is that the price decrease in t_1 should be very sharp in order to compensate for the losses from the spread. As already stated, that there will be a position loss if the order submitted by an aggressive trader is adversely selected by an incoming informed market order (see also Glosten and Milgrom, 1985). Therefore, traders must be sure that they do not face adverse selection risk when attempting an aggressive trading strategy. Indeed, this argument explains Brogaard et al.'s (2014) view regarding the profitability of extreme price movements for HFTs.

The ability of HFTs to make hay of volatile trading conditions as described above is not far-fetched. Hirschey (2017) argues that HFTs can anticipate buying and selling pressure, which could help them avoid being adversely selected when deploying aggressive trading strategies (see also Hoffmann, 2014; Jovanovic and Menkveld, 2016). Indeed, Hirschey (2017) finds that HFTs' aggressive sales and purchases consistently lead those of other investors. This implies that the framework I illustrate above is more likely to be successfully deployed when it is implemented at a high frequency.

2.2.3 Order aggressiveness and flash crashes

Thus far, I have demonstrated price evolution under an aggressive trading strategy. The sequence of aggressive trading strategy I describe is useful for understanding the contribution of order aggressiveness to flash crashes. Although, the sequence of orders is not based on the May 6 2010 flash crash, the aggressive trading strategy shares three notable characteristics with the May 6, 2010 flash crash. Firstly, the price movement under this strategy exactly mimics the price movements in the US financial markets during the flash crash, i.e. asset prices collapse and rebound very rapidly within a very short period of time. Secondly, the SEC (2010) finds that a large amount of seller-initiated E-mini contracts executed by algorithmic traders triggered the flash crash. My approach also begins with a sell limit order. Thirdly, consistent with recent empirical findings, my framework also predicts the aggressiveness of HFTs in demanding liquidity during flash crashes (see as an example Kirilenko et al., 2017). Inspired by these three commonalities, I argue that an aggressive trading strategy can contribute to flash crashes under certain conditions, mainly when there is excessive aggressiveness prior to flash crashes and aggressive traders can avoid adverse selection risk. In order to test my arguments, using relevant data, I examine the aggressiveness of the order flow during, and prior to, the May 6, 2010 flash crash. If, indeed, the predictions of the framework are consistent with the flash crash, then, firstly, there should be an excessive sell order aggressiveness in financial markets, which will create a downward pulling effect on prices. Thereafter, the excessive aggressiveness should shift to the buy side and as a result, prices will rise. Secondly, the fraction and number of aggressive buy and sell orders during the May 6, 2010 flash crash should be higher than the fraction of aggressive buy and sell orders during the surrounding periods. This is simply because, as I have shown, aggressive orders are more profitable during these periods. It is very important to note that I do not argue that the three-stage aggressive trading strategy I illustrate in this chapter is the reason for the May 6 flash crash. Rather, I argue that order aggressiveness prior and during the May 6, 2010 flash crash contributes to flash crashes (see SEC, 2010).

2.3 Data

2.3.1 Sample selection

In order to empirically test my hypotheses, as developed above, I focus on the biggest and most reported flash crash in the recent financial markets history, the May 6, 2010 flash crash experienced in the U.S. markets. The flash crash was one of the most turbulent periods in U.S. financial markets history and has been considered to be the most harmful flash crash to date, during which the biggest intraday point decline in the history of the Dow Jones Industrial Average was recorded. Instruments such as options, exchange-traded funds, and individual stocks, also suffered from the May 6, 2010 flash crash.[9] I focus on the May 6, 2010 flash crash because it provides an ideal ground for testing the relationship between a flash crash and pre-crash aggressiveness.

The data employed consists of ultra-high frequency tick-by-tick data for a selection of 53 S&P 500 stocks sourced from the Thomson Reuters Tick History (TRTH) database.

[9] According to SEC (2010), the May 6, 2010 flash crash lasted for approximately 36 minutes and could be viewed as consisting of two halves: (1) prices collapse and reach their lowest levels from 2:32 PM to 2:45 PM, (2) prices rebound and reach their pre-crash levels from 2:46 PM to 3:08 PM.

Appendix 4.B contains a detailed list of all stocks included in the sample. I obtain data for all messages recorded for May 6, 2010, but focus mainly on the period between 1:30 PM and 4 PM, since the flash crash started around 2:32 PM and lasted for about 36 minutes (see SEC, 2010). In the data, each message is recorded with a time stamp to the nearest 1/1000th of a second (millisecond). The following variables are included in the dataset: Reuters Identification Code (RIC), date, timestamp, price, volume, bid price, ask price, bid volume, and ask volume.

Although the S&P 500 index consists of 500 large companies listed on the NYSE and NASDAQ, I select only the 53 stocks deemed to have been severely affected by the flash crash. I employ these stocks, because only stocks affected by a flash crash are appropriate for testing the predictions of a framework depicting a flash crash. In addition, I select S&P 500 stocks because SEC (2010) also examines the impact of the flash crash on individual stocks by using a sample selected from this index. SEC (2010) shows that a large trader executing a sell program for 75,000 E-mini S&P 500 index futures contracts triggered the flash crash of May 6, 2010. As the performance of this index future is directly linked with the S&P 500 stocks, it is reasonable to select the components of S&P 500 for my analysis.

Once the raw data is obtained, I determine the prevailing best bid and best ask quotes for each transaction by using the order flow as downloaded. I then follow Chordia et al. (2001) and Ibikunle (2015) in applying a standard set of exclusion criteria to the data, thus deleting all inexplicable observations which might arise due to errors in data entry.

2.3.2 Sample Description

In order to better observe the dynamics of stocks during the flash crash, I classify the sample into three periods: before the flash crash (from 1:30 PM to 2:32 PM), the flash crash period (from 2:32:01 PM to 3:08 PM), and after the flash crash (from 3.08:01 PM to 4:00 PM).

Table 2. 1 Transactions' summary statistics and statistical tests

Panels A and B respectively present trading summary statistics and statistical tests of differences between the period of the flash crash and surrounding periods for 53 S&P 500 stocks affected by the May 6, 2010 flash crash. The statistical tests conducted are two-sample t-tests and pairwise Wilcoxon-Mann-Whitney U tests. The sample period covers 1:30 PM to 4 PM May 6, 2010. The time series on May 6, 2010 is divided into three: before the flash crash (from 1:30 PM to 2:32 PM), the flash crash period (from 2:32 PM to 3:08 PM), and after the flash crash (from 3.08 PM to 4 PM).

Panel A. Summary statistics

		Total transactions (000s)	Average per minute transactions (000s)
Number of Transactions	1:30 PM – 2:32 PM	186.6	3.0
	2:32 PM – 3:08 PM	329.9	8.9
	3:08 PM – 4 PM	405.8	7.8
	All	922.3	19.7

		Total trading volume (000s)	Average per minute trading volume (000s)
Trading Volume	1:30 PM – 2:32 PM	62878.8	1014.2
	2:32 PM – 3:08 PM	98185.5	2653.7
	3:08 PM – 4 PM	119209.9	2292.5
	All	280274.2	5960.4

		Total dollar trading volume ($'000,000)	Average per minute dollar trading volume ($'000,000)
Dollar Trading Volume	1:30 PM – 2:32 PM	2541.6	41.0
	2:32 PM – 3:08 PM	4332.2	117.1
	3:08 PM – 4 PM	5239.7	100.8
	All	12113.5	258.9

Panel B. Statistical tests

Trading volume	
Method	p-value
Two-Sample T tests	
Pooled	<0.0001

Satterthwaite	<0.0001
Wilcoxon-Mann-Whitney U tests	<0.0001

Dollar trading volume	
Method	**p-value**
Two-Sample T tests	
Pooled	<0.0001
Satterthwaite	<0.0001
Wilcoxon-Mann-Whitney U tests	<0.0001

Table 2. 2 Order quoting summary statistics

Table presents order quoting summary statistics for 53 S&P 500 stocks affected by the May 6, 2010 flash crash. The sample period covers 1:30 PM to 4 PM May 6, 2010. The time series on May 6, 2010 is divided into three: before the flash crash (from 1:30 PM to 2:32 PM), the flash crash period (from 2:32 PM to 3:08 PM), and after the flash crash (from 3.08 PM to 4 PM).

		Total number of shares at the bid side (000,000s)	Average shares/minute at the bid side (000,000s)
Number of shares in orders submitted at the bid side	1:30 PM – 2:32 PM	168.4	2.7
	2:32 PM – 3:08 PM	93.3	2.5
	3:08 PM – 4 PM	130.0	2.5
	All	391.7	7.7

		Total number of shares/minute at the ask side (000,000s)	Average shares/minute at the ask side (000,000s)
Number of Shares in orders submitted at the ask side	1:30 PM – 2:32 PM	168.3	2.7
	2:32 PM – 3:08 PM	85.7	2.3
	3:08 PM – 4 PM	123.9	2.4

All	377.8	7.4

Panel A of Table 2.1 presents the summary statistics of trading activities of the selected stocks. I observe a marked increase in average per minute transactions and trading volume during the flash crash, followed by a fall after the flash crash. This volatility is consistent with the modelled effects of the flash crash as presented in the framework. Prior to the flash crash, the average per minute trading volume is about 1 million. This increases by 161% during the flash crash and afterward falls by approximately 14%. Furthermore, the average per minute number of transactions and dollar trading volume during the flash crash are about three times higher than before the flash crash. After attaining the highest levels, average transaction and dollar trading volumes per minute fall by about 13%. I compute statistical tests to show the differences in trading volume and dollar trading volume between the period of the flash crash and surrounding periods. In Table 2.1's Panel B, I present the p-values of different statistical approaches, testing for the null that there is no difference between the trading activity during the flash crash and non-flash crash periods. For robustness, I construct two-sample t-tests and pairwise Wilcoxon-Mann-Whitney U tests. Both methods show that the difference between these two periods is statistically significant. Given that, in the market microstructure literature, changes in trade sizes are thought to reflect the changing composition of the traders/participants in a market, one may assume that the fraction of traders that submit aggressive orders increases during flash crash.

Table 2.2 presents the order submission summary statistics for my sample of stocks. Although average per minute trading volume increases sharply during the flash crash, the average volume of shares submitted in bid and ask orders over the same frequency decline during the flash crash. Firstly, this is consistent with what I would expect in t_1, following liquidity consumption in t_0. Secondly, when the ratio of shares in orders to trading volumes is calculated, I find that the ratio is 5.3 before the flash crash, indicating that approximately one in five submitted shares in the orders submitted is executed prior to the flash crash. The ratio quickly falls to 1.8 during the flash crash and increases 2.1 afterwards. Thus, the rate of order execution quickens during the flash crash as the search for liquidity intensifies. The estimate of 1.8 share in order to trade ratio shows that more than half of shares in orders submitted during the flash crash are executed. This result further supports my argument that traders become more aggressive during the flash crash or, at the very least, the proportion of aggressive traders in the market increases during the flash crash.

2.4 Empirical analyses, results and discussions

My aim in this section is to formally test hypotheses arising from my three central framework arguments. The first argument suggests that excessive aggressiveness in trading is culpable in the inducement of flash crashes; this implies a significantly increased volume of aggressive sell and buy orders in the period leading up to and during the flash crash. More specifically, my framework predicts that, firstly, there should be an excessive sell aggressiveness in the first half of the flash crash and this aggressiveness will create a downward pressure on prices. Then, the buy side should subsequently become more aggressive, which will inevitably create an upward pressure on prices. Secondly, my framework predicts that aggressive orders contribute to the severity of flash crashes if there is excessive aggressiveness in the market in the build-up to extreme price movements. Thirdly, the framework suggests that aggressive orders are more profitable during extreme price movements such as flash crashes. The implication here is that the fraction and number of aggressive orders in the lead up to and during flash crashes should be higher than the fraction and number of aggressive orders during other trading periods surrounding flash crashes.

2.4.1 The evolution of order aggressiveness

In order to proceed with the test of the arguments/hypotheses above, I need to identify an appropriate indicator or proxy for aggressive orders. This is required to be able to compute interval-based fractions and volume of aggressive orders in the market. For consistency with the existing literature, I employ an established approach as developed by Biais et al. (1995) to categorise limit orders according to their aggressiveness for my empirical analysis. The acceptance of this classification scheme in the market microstructure literature is underscored by its relatively wide use (see as examples Degryse et al., 2005; Griffiths et al., 2000; Hagströmer et al., 2014). The Biais et al. (1995) order classification algorithm involves dividing buy and sell orders into six groups by their level of aggressiveness; Category 1 orders are the most aggressive orders, while Category 6 orders are the least aggressive. A Category 1 buy order has a bid price higher than the best ask price and a quantity larger than the quantity available at the best ask price at its time of submission. These kinds of buy orders would normally walk across the order book. A Category 2 buy order has a bid price equal to the best ask price but has a target quantity exceeding the prevailing depth at the best ask price. Category 3 buy orders also have bid prices equal to the best ask prices, however their target quantities do not exceed the prevailing depth at the best ask price. The bid price of Category 4 buy orders is higher than the best bid price but less than the best ask price. The quantity of this order is not necessary for categorisation purposes. Categories 5 and 6 buy orders are the least aggressive. Like the Category 4 buy order, there are no quantity requirements for categorising Category 5 buy orders, however the bid prices of these orders are equal to the best bid prices. All buy orders not otherwise categorised above are classified as Category 6 orders; specifically, the prices of these orders are less than the best bid prices. Based on their classification, Category 4, 5 and 6 orders are not usually immediately executed, and are therefore considered passive.

The categorisation for the sell orders mirror those of the buy orders. The ask prices of the Category 1 sell orders are less than the best prevailing bid price and their sizes exceed the depths at the current best bid prices. The ask prices of the Category 2 and 3 sell orders equal to the best bid price. Furthermore, the target quantities of Category 2 orders are higher than the quantities available at the best bid prices, whereas the quantity of Category 3 sell orders are not. Consistent with the categorisation of buy orders, the prices of Category 4 sell orders lie within the best bid-ask spread, i.e. less than prevailing best ask prices. The prices of Category 5 sell orders equal the best ask price, while the remaining orders are classified as Category 6 sell orders. The prices of this latter group of sell orders are higher than the prevailing best ask prices.

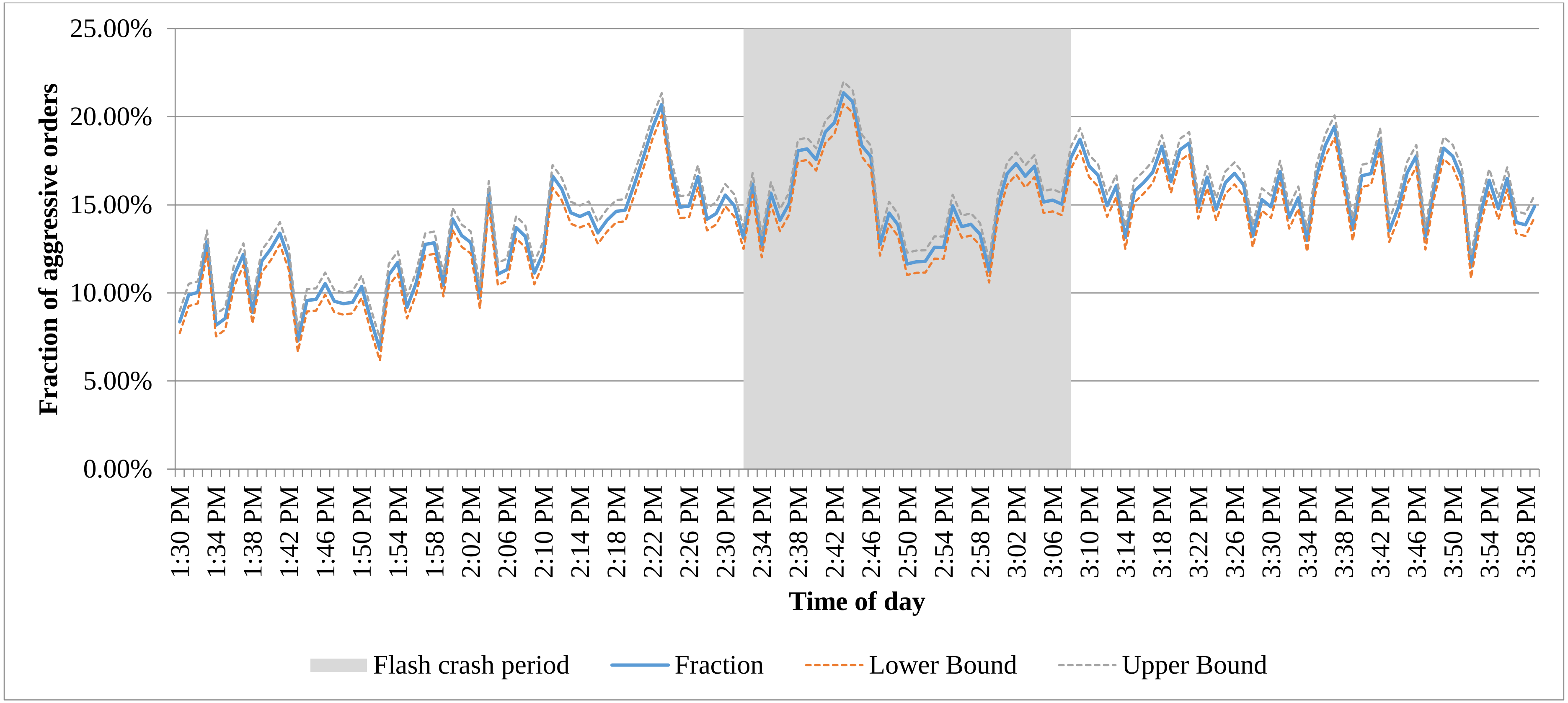
Degryse et al. (2005) show that the most aggressive order types (Categories 1 and 2) execute immediately and cause price movements. Although Category 3 orders are less aggressive than the first two classes of orders, they still usually result in prompt transactions, therefore these three types of orders (Categories 1, 2 and 3) can be considered as aggressive orders (see Degryse et al., 2005; Foucault, 1999). Thus, I focus on the first three types of orders. Specifically, I compute the sum of fractions of the aggressive order categories for the May 6, 2010 flash crash, as well as for the normal periods surrounding the flash crash. I then compare the volumes within a statistical framework to determine whether the fraction of aggressive orders during the flash crash is higher than the fraction of the same types of orders during normal periods.

Figure 2.1 presents the evolution of order aggressiveness during the day of the flash crash. I use 1-minute time intervals to construct both panels of the panels in the figure. In Panel A, I employ the standard errors of the cross-sectional means to construct 99% confidence bands for the order aggressiveness estimates in Panel A, to show the upper and lower bounds of the fraction of aggressive orders during the flash crash.

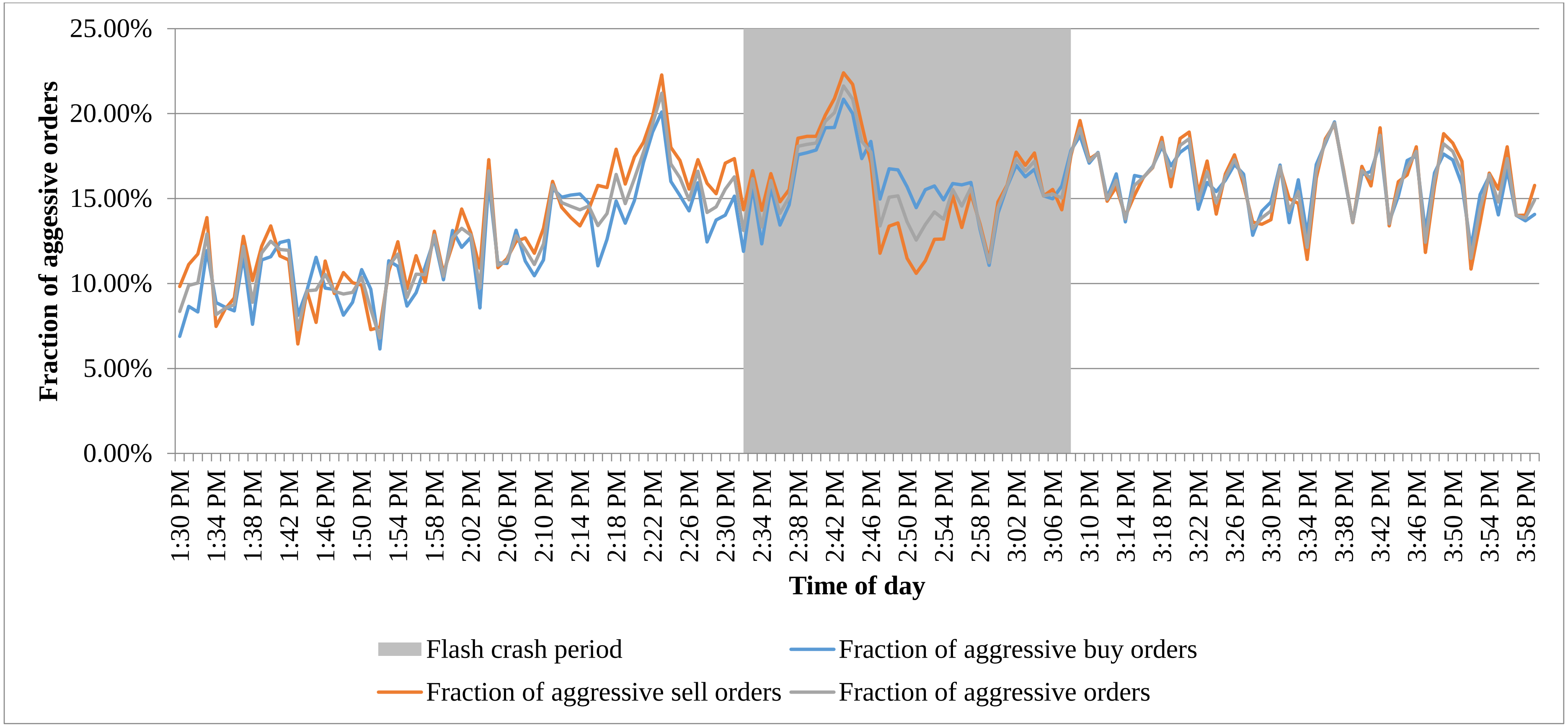
Figure 2. 1 Intraday evolution of the fraction of aggressive orders

Panels A and B depict the minute-by-minute evolution of the fraction of aggressive orders for 53 S&P 500 stocks affected by the May 6 2010 flash crash; Panel B presents the fraction of aggressive orders when disaggregated into buys and sells, as well as the fraction of all aggressive orders, while Panel presents only the fraction of all aggressive orders. 99% confidence bands are constructed for Panel A using the means of the minute-by-minute fractions of aggressive orders across the stocks in the sample. The sample period covers 1:30 PM to 4 PM May 6, 2010. The shaded area indicates the flash crash period.

Panel A. Fraction of total aggressive orders



Panel B. Fraction of total, buy and sell aggressive orders

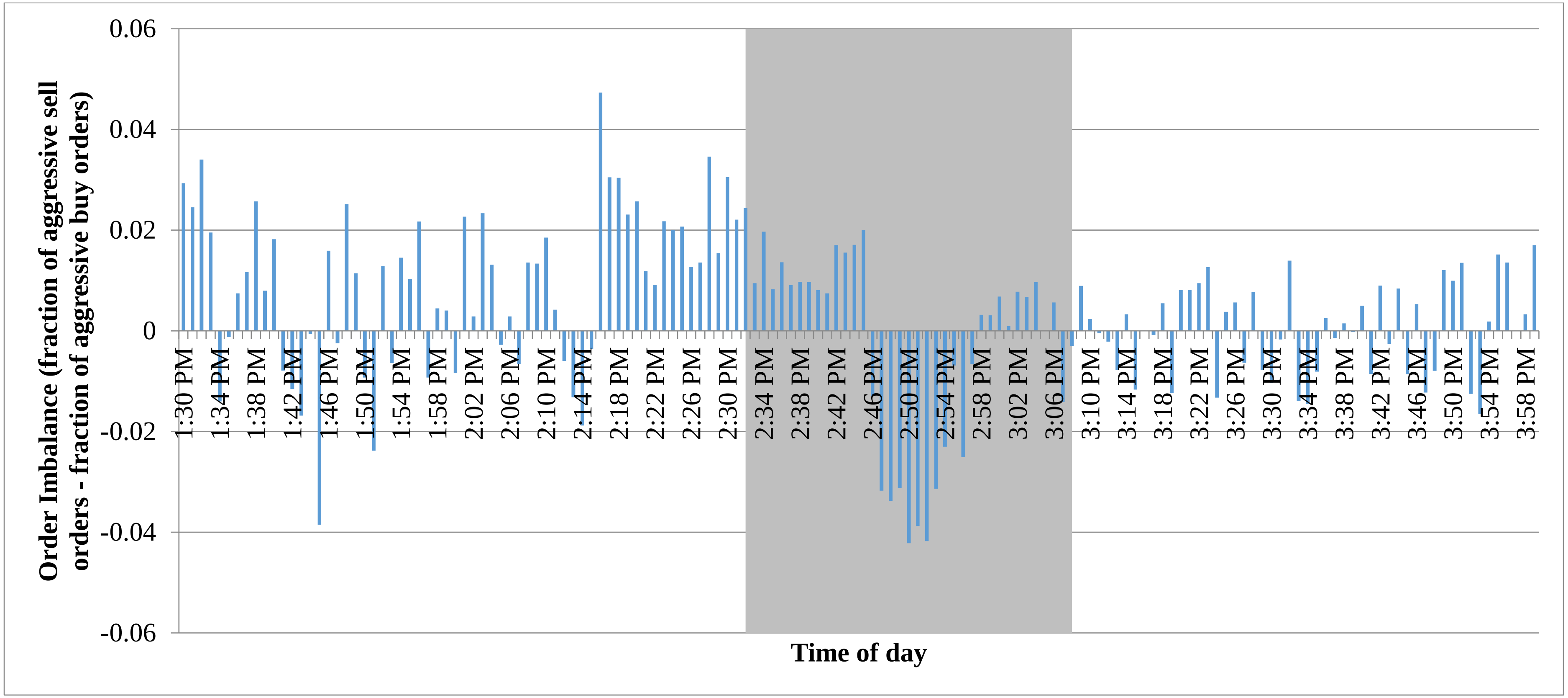


As evident in Panel A, the fraction of aggressive orders almost tripled during the flash crash from about 8% at 1:30 PM to 21.36% at 2:43 PM. The proportion of aggressive orders during the flash crash is, on average, higher than the surrounding time intervals. This finding suggests that aggressive trading activity is more prominent during the flash crash than in the surrounding periods. This result is consistent with the view that since aggressive orders might be more profitable during periods of extreme price movements, traders tend to show more aggressive behaviour during such periods. Furthermore, in Figure 2.1, I observe that the first of the two peaks of aggressive trading occurs just prior to the onset of the flash crash at about 2:23 PM, when the fraction of aggressive orders attains about 20.71% of the total order volume. This appears to underscore my intuition regarding the contribution of pre-flash crash order aggressiveness to the flash crash. I discuss the results of my formal test of this assertion in the next section.

Panel B makes the important distinction between buy and sell aggressive orders. Consistent with the framework's predictions, the sell side is more aggressive from 2:17 PM to 2:45 PM and then the buy side becomes more aggressive until 2:58 PM. This is not unexpected since SEC (2010) show that prices reached their lowest levels at 2:45 PM and the start to increase thereafter. This shows that the predictions of my framework are consistent with the empirical evidence and the arguments I make are valid in the case of the flash crash I examine. A clearer view of the balance between sell and buy aggressive orders is presented in Figure 2.2.

Figure 2. 2 Intraday evolution of aggressive order imbalance I

The figure presents the minute-by-minute evolution of aggressive order imbalance (difference between the fractions of aggressive sell and buy orders) for 53 S&P 500 stocks affected by the May 6 2010 flash crash. The sample period covers 1:30 PM to 4 PM May 6, 2010. The shaded are indicates the flash crash period. The shaded area indicates the flash crash period.



Consistent with the results of Figure 2.1, Figure 2.2 shows that, as predicted by my framework, there is a significant increase of aggressive sell orders until the stocks' price attained their lowest levels during the flash crash (at 2:45 PM) and thereafter the number of aggressive sell orders are outstripped by the number of aggressive buy orders until the prices reverted back to their pre-crash levels. Furthermore, Figure 2.2 shows that I observe a peak in aggressive order imbalance (the difference between aggressive sell and buy orders) at 2:17 PM; this implies that as predicted by my framework, aggressive orders' build-up ahead of the flash crash is a contributory factor to flash crashes.

However, it is important to note that, based on my predictions, a high fraction of aggressive orders during some specific days alone is not enough to influence extreme price movements such as a flash crash; flash crashes are more likely induced by a large amount of aggressive orders. Therefore, I also need to examine the number of aggressive orders during the flash crash day in order to adequately investigate the prediction made in my framework.

Figure 2. 3 Intraday evolution of aggressive orders

The figure presents the minute-by-minute evolution of the numbers of total, sell and buy aggressive orders for 53 S&P 500 stocks affected by the May 6, 2010 flash crash. The sample period covers 1:30 PM to 4 PM May 6 2010. The shaded area indicates the flash crash period. The shaded area indicates the flash crash period.

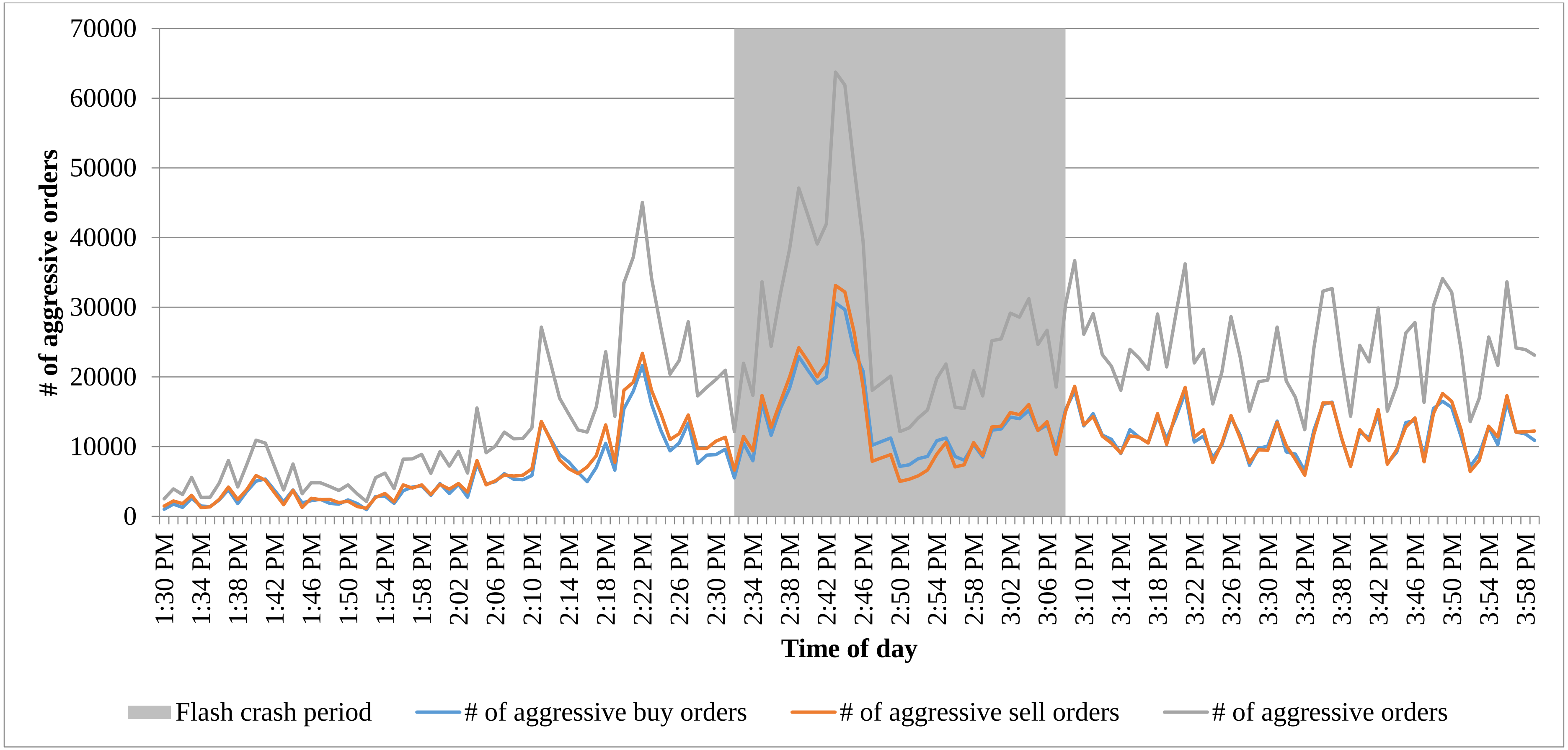
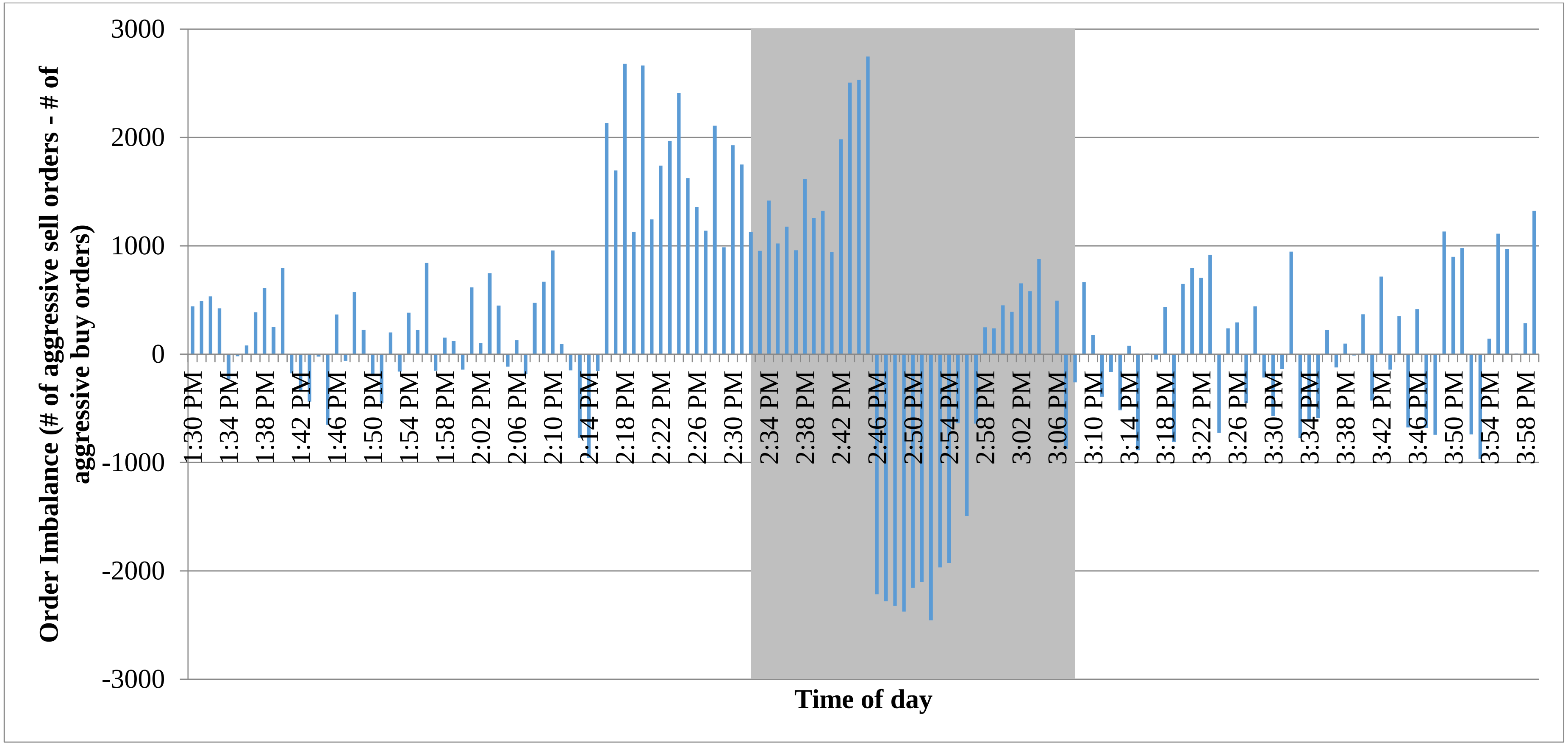


Figure 2.3 presents the evolution of the number of aggressive orders on May 6, 2010. As evident in the figure, there is a noteworthy rise in the number of aggressive orders as I approach the epicentre of the crash. The number of aggressive orders increases by about 6 times from the number at 2:00 PM (10,586/minute) to 62,760/minute at 2:43 PM, then falls precipitously to about 24,000/minute thereafter. Consistent with the data on the fraction of aggressive orders, I also observe a peak in the number of aggressive orders prior to the onset of the flash crash, at 2:22 PM (45,050/minute). This implies that, consistent with the predictions of my framework, excessive aggressiveness is likely to occur prior to flash crashes. Furthermore, as evident in Figure 2.2, I observe an excessive level of sell order aggressiveness from 2:17 PM to 2:45 PM and an excessive buy order aggressiveness thereafter. A review of the balance between aggressive sell and buy orders is useful in clarifying the changing of order dominance between the two order types. Thus, I compute aggressive order imbalance by the numbers of orders.

Figure 2. 4 Intraday evolution of aggressive order imbalance II

The figure presents the minute-by-minute evolution of aggressive order imbalance (difference between the number of aggressive sell and buy orders) for 53 S&P 500 stocks affected by the May 6 2010 flash crash. The sample period covers 1:30 PM to 4 PM May 6, 2010. The shaded area indicates the flash crash period. The shaded area indicates the flash crash period.



Similar to the picture painted in Figure 2.2, Figure 2.4 shows that the predictions of my framework are completely in line with the evolution of the number of buy and sell orders during a real flash crash. I observe a surge in sell order aggressiveness prior to and during the first half of the flash crash until the price levels of instruments reached their minimum levels. Thereafter, the number of aggressive buy orders start to increase relative to the number of aggressive sell orders until the prices regain their pre-crash levels. The implications of the findings presented in Figure 2.2 and Figure 2.4 are significant, since the total number of aggressive orders could be high for a number of reasons; however, a flash crash is unlikely to ensue if there are no significant differences in the fractions and numbers of aggressive buy and sell orders.

Thus far, the univariate empirical results presented have been generally consistent with the predictions of my framework concerning the relationship between order aggressiveness and flash crashes. Firstly, there is a significantly increased level of sell order aggressiveness prior to and during the first half of the flash crash and then, buy order aggressiveness gradually outstrips sell order aggressiveness. Secondly, there is excessive order aggressiveness prior to the flash crash. Thirdly, the number and the fraction of aggressive orders attain their highest levels during the flash crash and is in line with my argument that these types of orders might be more profitable during extreme price movements. Although the initial results suggest that my hypothesis on the predictive power of aggressive orders for flash crashes has merit, it is imperative that these results are formally tested within a multivariate framework.

2.4.2 Multivariate Analysis

Next, I formally investigate the relationship between aggressive orders and flash crashes within a multivariate framework. Specifically, I estimate the following regression model with stock-specific variables:

$$
\begin{aligned}
FC_{it} &= \alpha + \beta_{NAO} NAO_{it} + \beta_{\ln V} \ln V_{it} + \beta_{VPIN} VPIN_{it} + \beta_{VLT} VLT_{it} \\
&+ \beta_{OIB} OIB_{it} + \beta_{BAS} BAS_{it} + \beta_{MF} MF_{it} + \varepsilon_{it}
\end{aligned} \tag{2.8}
$$

where FC_{it} is a binary dependent variable and time, t, equals one-second.[10] I employ two cases of the Model (4.8). Firstly, I use the standard logit model; in this step, my aim is to test whether the build-up of aggressive orders ahead of the flash crash is linked to its onset. In the logit model, FC_{it} equals one for the pre-flash crash period (2:17 PM to 2:32 PM). Secondly, I employ the multinomial logit mode, which allows me to concurrently examine the relationship between both the pre-flash crash and flash crash periods on the one hand and contemporaneous order aggressiveness on the other. Thus, in the multinomial estimation of Model (8), FC_{it} equals one for the pre-flash period (2:17 PM – 2:32 PM), two for the flash crash period (2:32:01 PM – 3:08 PM) and zero otherwise. *NAO* is the number of aggressive orders obtained by using the order classification scheme described above. I estimate the above regression for sell (NASO) and buy (NABO) aggressive orders separately in order to capture the marginal impact of each type of order. Estimating the depth of the impact of each order type is important since according to the literature and my framework, aggressive sell orders should play a more important role in flash crashes (see SEC, 2010). As already noted, the first three categories of orders are earmarked as aggressive orders. This is the most important variable in my study, and according to my arguments, I expect to see a positive relationship between the number of aggressive orders and the pre-flash crash (FC_{it}=1) period (see also Griffiths et al., 2000; Mcinish et al., 2014; Wuyts, 2011).

Apart from the key variable, I employ some control variables in order to strengthen the consistency of my results. *lnV* is the natural logarithm of the number of shares traded for one/five second interval. This proxy is used to control for the effect of trading volume. The

[10] For robustness, I also employ five-second interval analysis and obtain qualitatively similar results. For parsimony, the results of the five-second estimation results are not presented; however, they are available on request.

VPIN metric is introduced as a real-time indicator of order flow toxicity. *VPIN* is a modified version of the Easley et al. (1996) and Easley et al. (1997) probability of an informed trade (PIN) metric and is proposed by Easley et al. (2011) as a measure of the probability of an informed trade in a high frequency environment. Easley et al. (2011) and Easley et al. (2012) highlight the role of order flow toxicity in the May 6, 2010 flash crash.[11] Easley et al. (2011; 2012) argue that *VPIN* can be used to predict flash crashes. By contrast, Andersen and Bondarenko (2014) show that *VPIN* is a poor predictor for flash crashes after controlling for volume. Therefore, including *VPIN* as a control variable in Model (8) offers another opportunity to examine the flash crash predictability potentials of *VPIN*. In addition to *VPIN*, *OIB* is also employed to control for the order flow toxicity. Note that multicollinearity is not an issue here, since the correlation coefficient between *VPIN* and *OIB* is very low, at 0.054 (see Table 2.3). SEC (2010), Kirilenko et al. (2017), and Easley et al. (2011), show that a large order imbalance was one of the contributing factors to the May 6, 2010 the flash crash, hence the inclusion of order imbalance as an explanatory variable is completely in line with the literature. *OIB* is calculated as the absolute value of the difference between the number of buy and sell trades, divided by the total number of trades (see Chordia et al., 2008). In order to obtain *OIB*, trades must first be classified into buys and sells. Generally, three types of trade classification schemes are used to classify trades; these are the tick rule, the Lee and Ready (1991) algorithm, and Easley et al. (2011; 2012) bulk volume classification (BVC) method. In this study, I employ the Lee and Ready (1991) algorithm for order classification.[12] Chakrabarty et al. (2015), in their comparative analysis of the aforementioned trade classification methods,

[11] Computing VPIN requires determining the number of buckets to be employed for volume classification and a buy/sell trade classification method. I use 200 buckets for volume classification, because Wu et al. (2013), who examine 16,000 various parameter combinations for evaluating the effectiveness of VPIN, concludes that 200 buckets yield optimal results. Buy and sell volumes are computed using the BVC approach proposed by Easley et al. (2011).

[12] For robustness, I also compute OIB using the other two methods and employ them in Model (8), the inferences drawn from those estimations are unchanged irrespective of which OIB computation approach I use.

conclude that the Lee and Ready (1991) algorithm method is a more accurate trade classification method than competing methods.

VLT is the one/five-second standard deviation of mid-price returns; this variable is introduced to control for trading volatility.[13] Prior contributions report extreme price volatility during the May 6, 2010 flash crash day (see as examples Easley et al., 2011; 2012; Kirilenko et al., 2017; SEC, 2010). Furthermore, an increase in the volatility of an instrument's price will increase its market risk, leading to a larger price impact as well as extreme price movements. *BAS* is the one/five second spread between the best ask and best bid prices, and is a proxy for liquidity. *BAS* tends to be narrow when liquidity is high; hence, under liquidity constraints, i.e. when *BAS* is wide, I therefore expect a larger price impact (see Borkovec et al., 2010). *MF* corresponds to market fragmentation. Madhavan (2012) and Golub et al. (2012) show that market fragmentation is one of the factors that contribute to flash crashes, and Menkveld and Yueshen (2017) underscore and further explain the results of Madhavan (2012). In this study, the inverse of the Herfindahl-Hirschman Index is used for capturing how fragmented each stock is across various venues for each corresponding interval.[14]

Table 2. 3 Correlation matrix of explanatory variables

The table presents the correlation matrix for the explanatory variables employed in the flash crash models. *NAO* is the number of aggressive orders, *NASO* is the number of aggressive sell orders, *NABO* is the number of aggressive buy orders, *VPIN* is the Volume-Synchronized Probability of Informed Trading, *VLT* is the standard deviation of the mid-price returns, *OIB* is the order imbalance, *BAS* is a bid-ask spread, *MF* represents market fragmentation, and *lnV* is the natural logarithm of the number of shares. The sample includes 53 S&P 500 stocks affected by the May 6, 2010 flash crash. The sample period covers 1:30 PM to 4 PM May 6, 2010.

	NAO	NASO	NABO	VPIN	VLT	OIB	BAS	MF	lnV
NAO	1								
NASO	0.91	1							
NABO	0.90	0.92	1						

[13] I employ mid-price returns in order to reduce bid-ask bounce (see Avramov et al., 2006).
[14] The index is defined as: $HHI_t = \sum_{k=1}^{K} \left(s_t^k\right)$, where s_t^k is volume share of venue *k* on day *t*. The value of the index ranges from 0 to 1; higher value implies less fragmentation.

VPIN	0.014	0.013	0.014	1					
VLT	0.093	0.094	0.093	0.146	1				
OIB	0.385	0.383	0.384	0.054	0.139	1			
BAS	-0.013	-0.012	-0.014	0.10	0.27	0.03	1		
MF	0.273	0.274	0.273	0.12	0.20	0.13	0.05	1	
lnV	0.394	0.395	0.394	0.08	0.30	0.27	0.05	0.58	1

Table 2.3 presents the correlation matrix of the explanatory variables; the low correlation coefficient estimates suggest that multicollinearity is not an issue with the regression model.

The results for both the logit and multinomial logit models' estimations are presented in Table 2.4 and Table 2.5 respectively.

Table 2. 4 Standard logit model for one second frequency

The predictive power of the number of aggressive orders on flash crashes is estimated using the following model:

$$FC_{it} = \alpha + \beta_{NAO} NAO_{it} + \beta_{\ln V} \ln V_{it} + \beta_{VPIN} VPIN_{it} + \beta_{VLT} VLT_{it}$$
$$+ \beta_{OIB} OIB_{it} + \beta_{BAS} BAS_{it} + \beta_{MF} MF_{it} + \varepsilon_{it}$$

The table reports logit regressions' coefficient estimates using one second frequencies. Results for standard logit model estimations are presented for the number of aggressive orders, aggressive sell orders and aggressive buy orders in the second, third and fourth columns respectively. FC_{it} equals zero from 1:30 PM to 2:17 PM, and from 2:32 PM to 4:00 PM, while it takes the value of one from 2:17 PM to 2:32 PM. *NAO*, *NASO* and *NABO* are the number of aggressive orders, number of aggressive sell orders and number of aggressive buy orders, respectively, *lnV* is the natural logarithm of the number of shares, *VPIN* is the Volume-Synchronized Probability of Informed Trading, *VLT* is the standard deviation of the mid-price returns, *OIB* is the order imbalance, *BAS* is the prevailing bid-ask spread and *MF* represents market fragmentation. Standard errors are presented in parentheses. The sample includes 53 S&P 500 stocks affected by the May 6, 2010 flash crash. The sample period covers 1:30 PM to 4 PM May 6, 2010. *** and ** correspond to statistical significance at the 0.01 and 0.05 levels, respectively.

Variables	**NAO**	**NASO**	**NABO**
	$4.98 \times 10^{-3***}$ (2.18×10^{-4})	$1.66 \times 10^{-2***}$ (7.27×10^{-4})	$7.12 \times 10^{-3***}$ (3.12×10^{-4})
lnV	$-1.23 \times 10^{-2***}$ (2.37×10^{-3})	$-1.23 \times 10^{-2***}$ (2.37×10^{-3})	$-1.23 \times 10^{-2***}$ (2.37×10^{-3})
VPIN	-1.2531^{***} (2.43×10^{-2})	-1.2531^{***} (2.43×10^{-2})	-1.2530^{***} (2.43×10^{-2})
VLT	-1460.6^{***} (45)	-1460.6^{***} (46.002)	-1460.5^{***} (45.998)
OIB	$-3 \times 10^{-5***}$ (5.1×10^{-6})	$-3.11 \times 10^{-5***}$ (5.1×10^{-6})	$-3 \times 10^{-5***}$ (5.1×10^{-6})
BAS	-1.34^{***} (6.48×10^{-2})	-1.3432^{***} (6.48×10^{-2})	-1.3431^{***} (6.48×10^{-2})
MF	1.70×10^{-3}	1.68×10^{-3}	1.7×10^{-3}

	(5.26×10^{-3})	(5.26×10^{-3})	(5.26×10^{-3})
Mc Fadden's R^2	0.025	0.0292	0.0251

The results presented in Table 2.4 show that, as predicted by my framework, aggressive orders are positively linked with the pre-flash crash period. The result holds for a combination of buy and sell aggressive orders as well as for each type of aggressive orders separately. The positive and statistically significant coefficients suggest that order aggressiveness in the lead up to the flash crash is linked to the onset of the crash. An essential point to note is that the relationship between aggressive orders and the pre-flash crash period is statistically significant even after controlling for volume, liquidity, order flow toxicity and volatility. This finding is important given recent findings by Andersen and Bondarenko (2014), showing that a *popular* metric for order flow toxicity, the *VPIN* metric, developed by Easley et al. (2011; 2012), is a poor predictor for flash crashes once trading activity is controlled for. The practical implication of this result is that traders seeking to avoid the adverse effects of a flash crash must act quickly to do so. However, their actions could be inevitably endogenous, leading to a self-fulfilling prophecy, as their actions could exacerbate what might already be proving to be a challenging and increasingly illiquid trading environment. As already noted, according to the existing literature and the predictions of my approach, I expect that sell orders to play a more important role in the flash crash (SEC, 2010) and therefore, estimation separate regressions for aggressive sell and buy orders may provide more insightful results. This expectation is confirmed by the magnitude of the coefficient estimates and explanatory power for both the buy and sell aggressive orders estimations. Firstly, the coefficient estimate for aggressive sell orders is 2.3 times higher than the coefficient for the number of aggressive buy orders. Secondly, according to the *McFadden's* R^2, the model with the sell order has a higher explanatory power.

The estimated coefficients for all the other explanatory variables, except *MF* (market fragmentation), are also significantly correlated with the pre-flash crash period; however, the aggressive orders variables (*NAO*, *NASO* and *NABO*) are the only positive and statistically significant variables. As already noted, my model allows us to test the flash crash predictability potential of *VPIN* after controlling for trading activity, liquidity and volatility. My findings show that *VPIN* is negatively correlated with the pre-flash period; increases in the value of the *VPIN* metric does not provide a signal about extreme volatility. This is in some ways an unsurprising result, since Andersen and Bondarenko (2014) also show that *VPIN* is negatively correlated with future short-term volatility after controlling for trading activity. The explanatory power of the standard logit model reported for the *NAO*, *NASO* and *NABO* regressions using *McFadden's* R^2, are 2.5%, 2.9% and 2.51% respectively. This is also unsurprising because of the following two reasons. Firstly, I employ one-second frequency for the estimations.[15] Secondly, although *McFadden's* R^2 is a similar measure of the goodness of fit to the classic R^2, the value of *McFadden's* R^2 tend to be remarkably lower than the value of R^2 (see David and Peter, 1979).

[15] *McFadden's* R^2 rises to about 4.5% when I estimate the regression at five-second frequencies; the results are not presented for parsimony, but are available on request.

Table 2. 5 Multinomial logit model for one second frequency

The predictive power of the number of aggressive orders is estimated using the following model:

$$
\begin{aligned}
FC_{it} &= \alpha + \beta_{NAO} NAO_{it} + \beta_{\ln V} \ln V_{it} + \beta_{VPIN} VPIN_{it} + \beta_{VLT} VLT_{it} \\
&+ \beta_{OIB} OIB_{it} + \beta_{BAS} BAS_{it} + \beta_{MF} MF_{it} + \varepsilon_{it}
\end{aligned}
$$

The table reports multinomial logit regressions' coefficient estimates using one second frequencies; Results for multinomial logit model estimations for the number of aggressive orders, the number of aggressive sell orders and the number of aggressive buy orders are presented in the second, third and fourth columns respectively. FC_{it} equals zero from 1:30 PM to 2:17 PM, and from 3:08 PM to 4:00 PM, while it takes the value of one from 2:17 PM to 2:32 PM (pre-flash crash period) and takes the value of two from 2:32 PM to 3:08 PM (the flash crash period). *NAO*, *NASO* and *NABO* are the number of aggressive orders, the number of aggressive sell orders and the number of aggressive buy orders, respectively, *lnV* is the natural logarithm of the number of shares, *VPIN* is the Volume-Synchronized Probability of Informed Trading, *VLT* is the standard deviation of the mid-price returns, *OIB* is the order imbalance, *BAS* is the prevailing bid-ask spread and *MF* represents market fragmentation. Standard errors are presented in parentheses. The sample includes 53 S&P 500 stocks affected by the May 6, 2010 flash crash. The sample period covers 1:30 PM to 4 PM May 6, 2010. *** and ** correspond to statistical significance at the 0.01 and 0.05 levels, respectively.

	NAO		NASO		NABO	
Variables	FC = 1	FC = 2	FC = 1	FC = 2	FC = 1	FC = 2
	6.30×10^{-3}***	3.81×10^{-3}***	2.09×10^{-2}***	1.27×10^{-2}***	9.0×10^{-3}***	5.45×10^{-3}***
	(2.33×10^{-4})	(1.79×10^{-4})	(7.74×10^{-3})	(5.95×10^{-4})	(3.32×10^{-3})	(2.55×10^{-4})
lnV	-2.75×10^{-2}***	4.71×10^{-2}***	-2.75×10^{-2}***	4.71×10^{-2}***	-2.74×10^{-2}***	4.71×10^{-2}***
	(2.43×10^{-3})	(1.77×10^{-3})	(2.43×10^{-3})	(1.77×10^{-3})	(2.43×10^{-3})	(1.77×10^{-3})
VPIN	-4.25×10^{-1}***	2.88***	-4.25×10^{-1}***	2.88***	-4.25×10^{-1}***	2.88***
	(2.53×10^{-2})	(1.72×10^{-2})	(2.53×10^{-2})	(1.72×10^{-2})	(2.53×10^{-2})	(1.72×10^{-2})
VLT	-1045.2 ***	1106.9 ***	-1045.5 ***	1106.8 ***	-1045.0 ***	1106.9 ***
	(47.81)	(19.25)	(47.81)	(19.24)	(47.81)	(19.24)
OIB	-2.00×10^{-5}***	2.4×10^{-5}***	-2.10×10^{-5}***	2.4×10^{-5}***	-2.10×10^{-5}***	2.4×10^{-5}***
	(5.27×10^{-6})	(2.74×10^{-6})	(5.27×10^{-6})	(2.74×10^{-6})	(5.27×10^{-6})	(2.74×10^{-6})
BAS	-5.2×10^{-1}***	2.05***	-5.2×10^{-1}***	2.05***	-5.2×10^{-1}***	2.05***
	(6.94×10^{-2})	(3.08×10^{-2})	(6.94×10^{-2})	(3.08×10^{-2})	(6.94×10^{-2})	(3.08×10^{-2})
MF	4.18×10^{-2}***	1.71×10^{-1}***	4.18×10^{-2}***	1.71×10^{-1}***	4.18×10^{-2}***	1.71×10^{-1}***
	(5.36×10^{-3})	(4.24×10^{-3})	(5.36×10^{-3})	(4.24×10^{-3})	(5.36×10^{-3})	(4.24×10^{-3})
Mc Fadden's R^2	0.066		0.0702		0.0665	

Table 2.5 presents the results for the multinominal logit model estimation. I employ this model to test the consistency of the standard logit model and in order to examine the relationship between contemporaneous order aggressiveness on the one hand and the pre-flash crash and the flash crash period on the other. This approach expectedly leads to a higher model explanatory power for the multinominal logit model estimation (*McFadden's* R^2 of 6.9% and 6.6% for the number of aggressive sell and buy orders, respectively) when compared with the standard logit model estimation reported in Table 2.4. Firstly, the findings in Table 2.5 are generally consistent with the results I present in Table 2.4; all the aggressive orders variables are positively and significantly related with the pre-flash crash period, which suggests a link between the number of aggressive orders and the onset of the flash crash. Furthermore, consistent with the findings from Table 2.4, the number of aggressive sell orders play a more important role in the flash crash. The only difference in the results is that while market fragmentation (*MF*) is not statistically significant in the standard logit model, it is significantly and positively correlated with the pre-flash period in the multinominal logit model. This implies that prior market fragmentation is related to flash crashes (see also Madhavan, 2012; Menkveld and Yueshen, 2017). The second set of results in Table 2.5, based on the flash crash period itself, are also interesting. The results show that the *NAO*, *NASO* and *NABO* are positively and significantly correlated with the flash crash period even after controlling for volume, liquidity and volatility. The positive and statistically significant estimates of the aggressive orders variables appear to confirm that increases in aggressive orders make flash crashes more likely to ensue. Specifically, the results suggest that the probability of flash crashes at time *t* rises as the number of aggressive orders increases at the same time. The evidence is in line with my approach that order aggressiveness plays an important role in flash crashes.

The regression results above, documenting the relationship between order aggressiveness and flash crashes, are consistent with the previous literature since they show that aggressive orders have a larger price impact than non-aggressive orders and that aggressive trading behaviour contributes to flash crashes (see as examples Griffiths et al., 2000; Mcinish et al., 2014; Wuyts, 2011).

The estimated coefficient estimates for all the other explanatory variables in Table 2.5 are also consistent with the existing literature on flash crashes. For example, the market toxicity metric, *VPIN*, has a statistically significant and positive relationship with the flash crash period. Taken together with the metric's documented relationship with the pre-flash crash period, the implication here is that while *VPIN*, may be a poor predictor of flash crashes when trading activity is controlled for (see also Andersen and Bondarenko, 2014), it nevertheless is positively correlated with flash crashes themselves. This suggests that market toxicity has a direct relationship with the flash crash; this evidence is in line with findings of Easley et al. (2011; 2012) that market toxicity plays an important role in the flash crash. Volatility exhibits a statistically significant and positive relationship with the flash crash. The positive coefficient is consistent with the stream of the market microstructure literature that states that an increase in the volatility of stock prices causes a larger price impact, since extreme price movements and flash crashes are characterized by extreme price volatility (see as examples Easley et al., 2011; Kirilenko et al., 2017; SEC, 2010). One plausible explanation of this positive relationship is that an increase in the volatility of stock prices increases the market risk, which in turn leads to larger spreads and extreme price movements.

The literature identifies order imbalance as one of the instigators of the May 6, 2010 flash crash (see as examples Easley et al., 2011; Kirilenko et al., 2017; SEC, 2010). Furthermore, Sun and Ibikunle (2016) find that order imbalance has information content and there is a significant and positive relationship between order imbalance and price impact in a high frequency trading environment. Thus, the positive relationship between *OIB* and the flash crash reported in Table 2.5 is unsurprising and is in line with the literature. The bid-ask spread, *BAS*, is also positively and statistically significantly related with the May 6, 2010 flash crash. This result is again unsurprising because existing literature finds that orders have a larger price impact when the bid-ask spread is wide (see Aitken and Frino, 1996) and, as already enumerated, liquidity constraints contribute to extreme price movements in the market. Furthermore, Borkovec et al. (2010), SEC (2010), and Menkveld and Yueshen (2017) find that the spread during the May 6, 2010 flash crash was uncharacteristically wide. Market fragmentation, *MF*, exhibits a statistically significant and positive relationship with the flash crash as well; this result can be justified that market fragmentation is important in explaining the anatomy of the flash crash. This result underscores the results of Madhavan (2012), Golub et al. (2012), and Menkveld and Yueshen (2017) that show that the flash crash is linked directly to market structure. When liquidity is fragmented across several venues, immediate access to counterparties becomes slightly more challenging given that orders may now need to be routed through several other channels in order for them to be filled.

I caution that evidence presented in Table 2.5 should be interpreted carefully. I do not claim to have found a causality between order aggressiveness and the flash crash. In addition, I do not claim that order aggressiveness was the main factor leading to the crash. However, my analysis shows that similar to other suggested factors, like order imbalance, order flow toxicity, market fragmentation, order aggressiveness has additional explanatory power for the flash crash.

2.4.3 Directional returns during the flash crash

I now turn my attention to the third mainline argument derived from my framework, which is that aggressive orders are more profitable during flash crashes. Earlier, I observe an increase in the volume of aggressive orders during the flash crash, I interpret this to be in response to their profitability during such periods. However, I also note that such increases may relate to the unwinding of untenable positions that arise as a result of extreme swings in instruments' valuations during a flash crash. In order to examine the veracity of my argument regarding the profitability of aggressive orders, I follow the approach proposed by Ederington and Lee (1995) to compute hypothetical returns attributable to an informed trader active during the flash crash and its surrounding periods (see also Caminschi and Heaney, 2014; Frino et al., 2017).

I estimate simple returns for each stock and sign the returns using a directional parameter ($DIR_{t,s}$), based on the assumption that the informed trader holds private information regarding the trajectory of the stocks' prices she trades. I define the directional return for each one-minute interval as

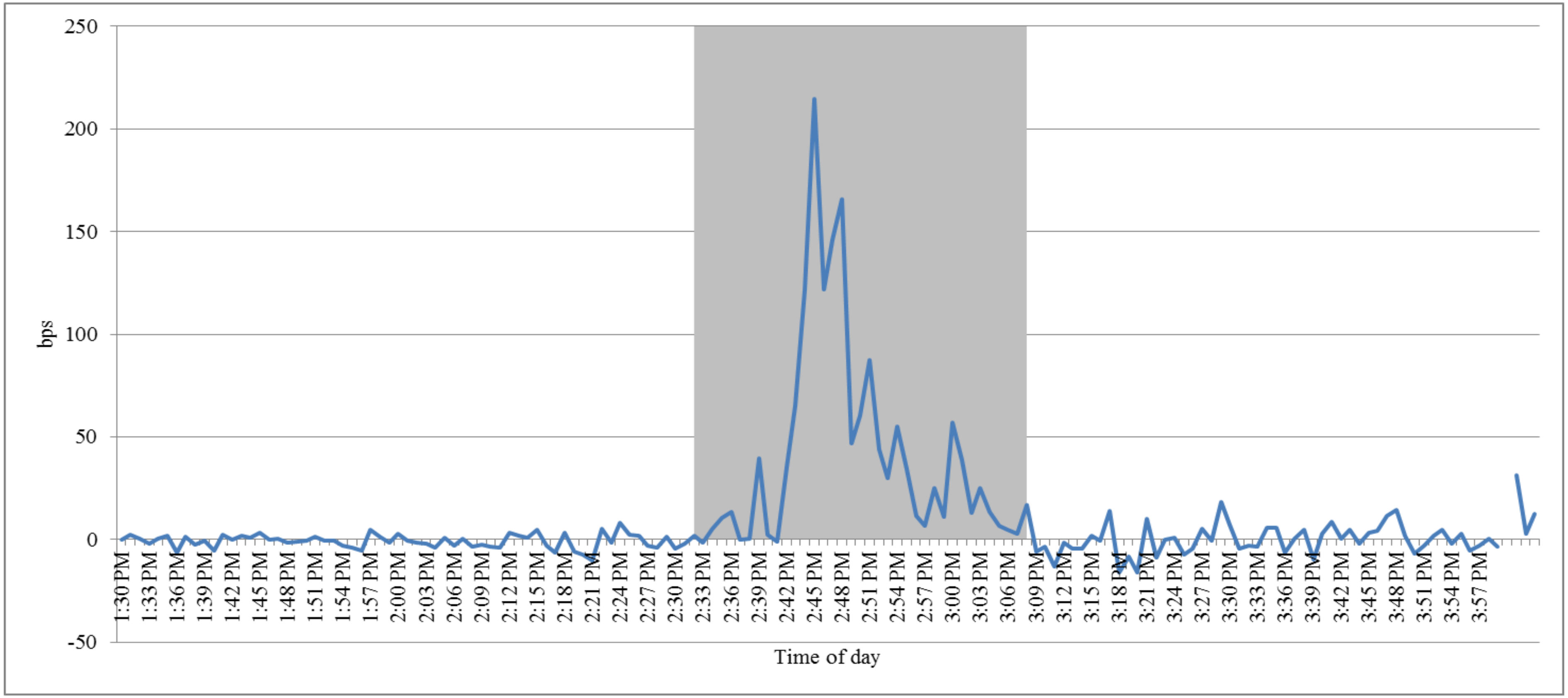
$$DR_{t,s} = R_{t,s} * DIR_{t,s} \tag{2.9}$$

where, $R_{t,s}$ represents simple return for stock s and time *t*. In order to define the directional parameter ($DIR_{t,s}$), firstly I compute the returns of each stock for the flash crash period (from 14:32 PM to 15:08 PM) ($R_{fc,s}$). The direction factor, $DIR_{t,s} = 1$ if $R_{fc,s} > 0$, $DIR_{t,s} = -1$ if $R_{fc,s} < 0$, and $DIR_{t,s} = 0$ if $R_{fc,s} = 0$. $DIR_{t,s} = 1$ (-1) indicates that the trader takes a long (short) position at time *t* for stock, s. I compute the average directional return, ADR_t, as the average of adjusted returns for all stocks for each one-minute interval. The cumulative average directional return, $CADR_t$ from 1:30 PM to 4:00 PM is estimated using the average directional returns.

Figure 2. 5 Directional returns

Panels A and B are minute-by-minute plots of average direction-adjusted returns and cumulative average direction-adjusted returns measures (in basis points) respectively for 53 S&P 500 stocks affected by the May 6 2010 flash crash. The sample period covers 1:30 PM to 4 PM May 6, 2010. The shaded area indicates the flash crash period.

Panel A. Average direction-adjusted returns



Panel B. Cumulative Average Adjusted Return

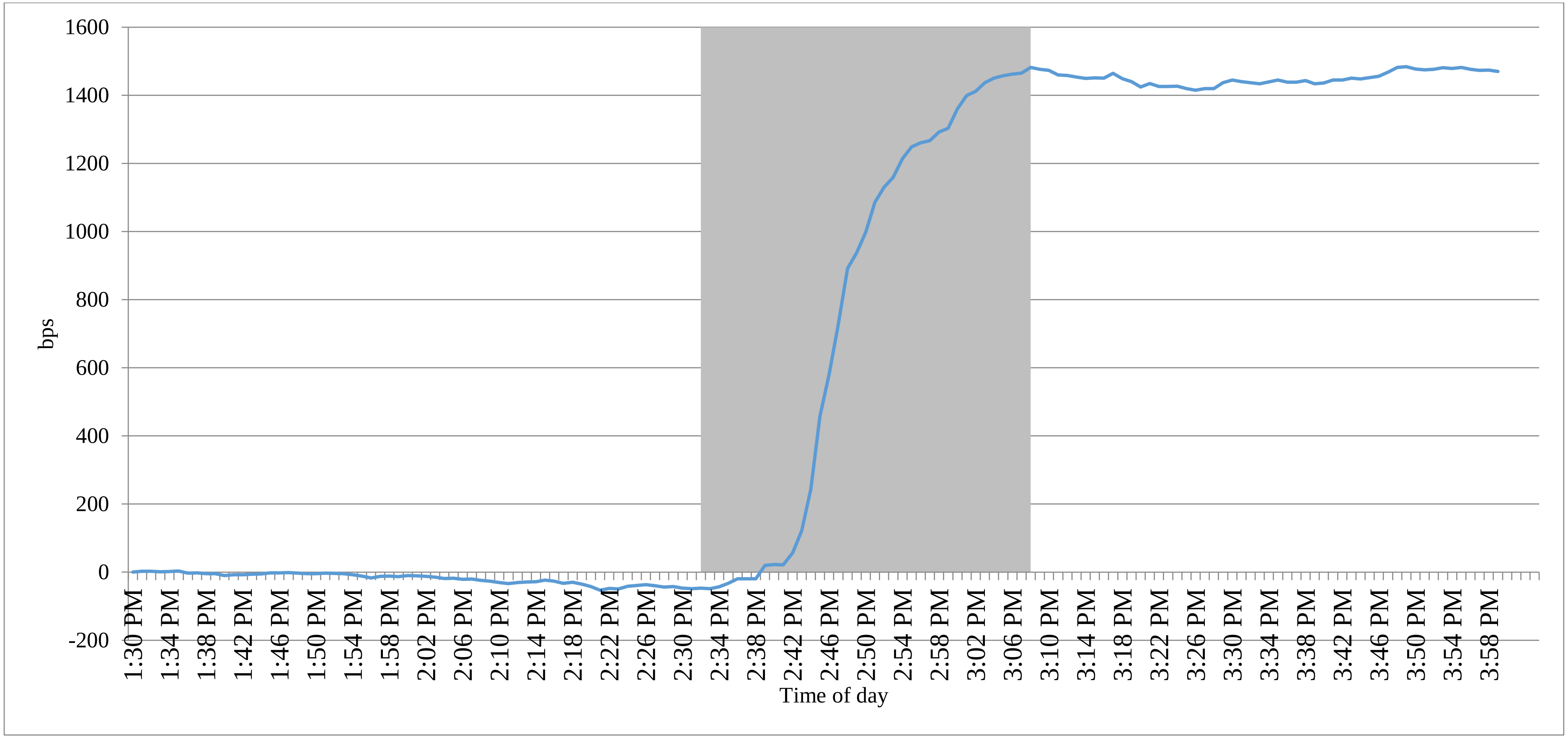


Table 2. 6 Average direction-adjusted returns

Table presents the average adjusted returns (AAR) in 10-minute batches for S&P 500 stocks. All return measures are reported in bps (1 bps = 0.01%). The t-value is the statistic of a one-sample t-test testing the null of the mean being equal to zero. The sample includes 53 S&P 500 stocks affected by the May 6, 2010 flash crash. The sample period covers 1:30 PM to 4 PM May 6, 2010. *** and ** correspond to statistical significance at the 0.01 and 0.05 levels, respectively.

From (Time)	**To (Time)**	**AAR**	**Sign**	**t-value**
1:30 PM	1:40 PM	-0.90		-0.13
1:41 PM	1:50 PM	0.56		0.08
1:51 PM	2:00 PM	-0.58		-0.08
2:01 PM	2:10 PM	-1.97		-0.30
2:11 PM	2:20 PM	-1.25		-0.19
2:21 PM	2:30 PM	-0.45		-0.06
2:31 PM	2:40 PM	6.95		1.06
2:41 PM	2:50 PM	97.59	***	14.91
2:51 PM	3:00 PM	36.19	***	5.53
3:01 PM	3:10 PM	11.27	*	1.85
3:11 PM	3:20 PM	-4.85		-0.75
3:21 PM	3:30 PM	2.01		0.31
3:31 PM	3:40 PM	-0.80		-0.12
3:41 PM	3:50 PM	4.07		0.62
3:51 PM	4:00 PM	-1.36		-0.21

Figure 2.5 reports the hypothetical returns attainable through aggressive (directional) trading in 53 selected S&P 500 stocks around the May 6, 2010 flash crash. Panel A shows the simple returns adjusted for direction of price movement over the flash crash period averaged across all 53 stocks, while Panel B shows the cumulative average direction-adjusted returns for the same stocks. As presented in Panel A, there are positive and significant directional returns during the flash crash. Remarkably, as predicted by my framework, the positive directional return is gained during the second half of the flash crash and only ends at the end of the flash crash at about 3:08 PM. The cumulative directional returns in Panel B shows the clear and continuous trend in adjusted returns during the flash crash period. This and the stabilisation of the cumulative returns following the conclusion of the flash crash support my arguments about the profitability of aggressive orders during periods of extreme price movements like flash crashes. The overall cumulative returns accruable to an informed trader during the flash crash is in excess of 1,482 basis points.

Table 2.6 reports the average direction-adjusted returns in 10-minute batches. Consistent with the insights from Figure 2.5, there is a positive and statistically significant adjusted returns, which commences in the second half of the flash crash and continues until the end of the flash crash. All estimated directional returns outside of the flash crash period are not statistically significant.

Overall, the directional returns analysis yields consistent results with the predictions of my framework, implying that aggressive orders are significantly more profitable during extreme price movements like flash crashes.

2.5 Conclusion

In this chapter, I develop a new framework for understanding the role of aggressive orders in flash crashes by extending the approach of Menkveld (2013). I then use ultra-high frequency data from 53 S&P 500 stocks affected by the May 6, 2010 flash crash to test the arguments motivated by the framework. The selection of the May 6, 2010 flash crash for my investigation is motivated by its recognition as the most significant flash crash in recent financial markets history. My main framework predictions/arguments are as follows. Firstly, there should be a significant increase in sell order aggressiveness prior to and during the first half of flash crashes, i.e. until instruments' price levels hit their lowest values and then the balance of order aggressiveness should shift to the buy side in the second half of the flash crash, i.e., until the prices re-attain their pre-crash levels. Secondly, my framework predicts that the build-up of order aggressiveness, which could be observed prior to extreme price volatility events, is inextricably linked to flash crashes. Thirdly, aggressive orders are more profitable during extreme price movements and thus traders tend to submit orders that are more aggressive during those periods.

In the formal test of the relationship between the number of aggressive orders and the pre-flash crash period, the empirical results are consistent with the predictions of my framework. Firstly, I find a significant increase in sell order aggressiveness prior to and during the first half of the May 6 2010 flash crash, thereafter the balance of order aggressiveness swings to the buy side, with traders submitting more aggressive buy orders relative to aggressive sell orders. The sell side is more aggressive until prices plummet to their lowest levels and then, the buy side becomes more aggressive in the run-up to prices regaining their pre-crash levels. Secondly, I find that the number of aggressive orders in the run up to the flash crash is positively and significantly related to the pre-flash crash period; thus, the build-up of order aggressiveness may contribute to the onset of flash crashes. Thirdly, the fraction and the number of aggressive orders during the flash crash are higher than the fraction and the number of orders during the surrounding periods due to the significantly larger (than other periods) profits accruable to informed investors during the flash crash. I estimate that for the stocks in my sample, an informed investor during the flash crash could achieve a return on his portfolio in excess of 1,482 bps, a return far larger than accruable during surrounding periods. This finding supports my argument that aggressive orders are more profitable markets are volatile and hence, traders tend to submit orders that are more aggressive during such periods.

While my findings show the contribution of aggressive orders to flash crashes, it is essential to note two points. First point is related to potential bias in the study. Explicitly, I investigate the role of aggressive orders in extreme price movements by focusing on stocks that impacted by the flash crash. While this method is consistent with the literature (see Easley et al., 2011), using only the flash crash affected stocks may lead to sample selection bias. Second, my findings should not be misconstrued as an endorsement of policies aimed at limiting aggressive orders or aggressive trading behaviours in financial markets. While I acknowledge that aggressive traders can induce extreme price movements, aggressive trading in itself could be a symptom of deeper underlying structural issues, which are not the focus of this study.

3. A state space modelling of the information content of trading volume

3.1 Introduction

Trading in financial markets is driven either by information or by the search for liquidity (see Admati and Pfleiderer, 1988). Liquidity traders do not trade on the basis of any specific information; their trading strategies are therefore not directly related to future payoffs. The trading strategies of informed traders, on the other hand, are based on private information and are directly related to future payoffs. The activities of these two fundamental types of traders have been extensively analysed in seminal papers in the larger financial markets literature, and more so in market microstructure papers. For example, Kyle (1985) predicts that the volatility of asset prices partially reflects inside information (informed trading) and is independent of liquidity-driven trading effects, while Glosten and Milgrom (1985) predict that the breadth of the bid-ask spread is primarily driven by informed trading, which incorporates adverse selection costs into the spread.[16]

More recently however, Kaniel and Liu (2006) have extended Glosten and Milgrom's (1985) model to show that informed traders with long-lived information are more likely to use limit orders than market orders. Therefore, informed traders' trading strategies, depending on the longevity of their information sets, may be negatively related with adverse selection. Using a comprehensive sample of trades from Schedule 13D filings by activist investors, Collin-Dufresne and Fos (2015) show that, consistent with Kaniel and Liu (2006), informed traders with long-lived information typically use limit orders, which leads to a negative correlation between adverse selection and informed trading (see also Collin-Dufresne and Fos, 2016).

[16] Consistent with Glosten and Milgrom (1985), Easley and O'Hara (1987) also suggest that stock illiquidity should increase in the presence of informed traders, as information asymmetry increases adverse selection, which widens the spread.

This chapter builds on the above predictions and findings by developing a general state space-based methodology for decomposing trading volume into unobservable liquidity-driven and information-driven components. According to Hendershott and Menkveld (2014), state space modelling is a natural tool for modelling an observed variable as the sum of two unobserved variables. While the application of state space modelling for decomposing price, owing to its efficiency, is very common in the finance literature (see as examples, Brogaard et al., 2014b; Hendershott and Menkveld, 2014; Menkveld et al., 2007), the approach has thus far not been directly applied to trading volume.[17] This is surprising given the preponderance of the literature on the strength of the relationship between price and trading volume (see as examples, Clark, 1973; Cornell, 1981; Epps and Epps, 1976; Harris, 1986; 1987; Karpoff, 1987).

The heavily evidenced relationship shown in the literature is linked to the joint dependence of price and volume on an underlying or set of underlying variable(s); this is the 'mixture of distribution hypothesis' (MDH) (see Clark, 1973; Harris, 1986). Harris (1986) argues that the underlying variable is the rate of flow of information. Hence, as new information arrives, traders act on it by revising their positions and consequently increase trading volume. Harris (1987), using data from NYSE, provides an empirical basis for the MDH. This implies that the theoretical basis for the application of state space modelling to price (i.e. that price reflects both information and non-information components) holds for volume.[18] However, it is important to note that while the information component of price is its permanent component, the information component of volume is transitory. This is simply because although new information implies a new permanent level of price it will only affect trading volume temporarily, since once prices reflect this information, informed traders will no longer hold an

[17] McCarthy and Najand (1993) apply state space modelling to the analysis of price and volume dependence in currency futures.

[18] A second explanation for the existence of the price-volume relationship is based on the sequential information models proposed by Copeland (1976), Jennings et al. (1981) and Smirlock and Starks (1984). The models suggest that volume improves forecasts of price variability and vice versa.

informational advantage and will therefore cease their trading based on the exploited information (see also Fama, 1970; Chordia et al., 2002; Suominen, 2001).

As discussed by Hendershott and Menkveld (2014), the state space approach holds significant economic value over other methods that could be appropriated for variable decomposition, such as autoregressive models (see as an example, Hasbrouck, 1991). Firstly, the estimation of the model using maximum likelihood is asymptotically unbiased and efficient. Secondly, maximum efficiency in dealing with missing values is achieved due to the use of the Kalman filter, which accounts for level changes across periods with missing observations, employed in the maximum likelihood estimation. This is a critical argument in the use of state space modelling in decomposing asset prices and trading volume in a high frequency trading environment such as the one I examine, since standard estimation approaches do not deal with missing observations. For example, estimating a vector autoregression implies truncation of the lag structure. Although standard approaches to decomposing trading volume may work well in a low-frequency environment, information in today’s markets travel at such ultra-high speeds that those standard approaches could potentially discard any additional information that could be obtained from high frequency data. Thirdly, following estimation the Kalman smoother, which is essentially a backward recursion after a forward recursion with the Kalman filter, facilitates a decomposition of any realised change in the series such that the estimated permanent or transitory component at any interval is estimated using all past, present, and future observations in the series. Thus, the purpose of filtering is to ensure that estimates are updated with the introduction of every additional observation (see also Durbin and Koopman, 2012).

In line with the expectation that asset price (and by extension, volume) is driven by informed trading and can therefore be decomposed into permanent and transitory components (see Brogaard et al., 2014b; Menkveld et al., 2007), I demonstrate that (observable) trading volume is a sum of two unobserved series. The first is a nonstationary series (the permanent component), and the second is a stationary series (the transitory component). I argue that the unobserved permanent component of trading volume is mainly driven by liquidity traders, whereas the unobserved transitory component is primarily driven by informed traders. The permanent component in the state space model is a nonstationary series and follows a random walk. Consistent with the literature (see as an example, Kyle, 1985), liquidity/uninformed traders trade randomly (i.e. the general reference to noise trading in the market microstructure literature), and thus I model the trading volume of liquidity traders as a random walk. Consequently, the non-random walk component of trading volume is modelled as trading volume due to informed trading activity.

In a test of the validity of the proposed state space-based volume decomposition approach, I use the estimated permanent and transitory components of trading volume to examine the impact of liquidity and informed trading activities on market quality metrics, such as volatility, liquidity, and toxicity. This part of my analysis serves as a joint test of the empirical relevance of the state space model and the impact of informed and liquidity trading on market quality. The relevance of my state space approach is underscored when my empirical findings are in line with the model predictions in the existing relevant theoretical market microstructure literature. I thereafter examine the predictive power of the estimated information-driven/transitory component of trading volume on short-horizon returns. This analysis furthers my aim of demonstrating the relevance of the state space approach to decomposing trading volume into informed and liquidity components. It is also a direct test of the efficiency of the price discovery process (see Chordia et al., 2005; 2008). Similar to the order imbalance metrics employed in Chordia et al. (2008), the transitory component, which also signals private information, is expected to be a predictor of short-horizon returns.

All the results obtained are generally consistent with my expectations. Based on my state space-estimated information and liquidity-driven components of trading volume, I find that after controlling for aggregate trading volume, stock price volatility and liquidity/toxicity are not driven by liquidity trading activity; however, it is impacted by informed trading activity. I also find that informed trading activity reduces price volatility and market toxicity and enhances liquidity. The results are robust to alternative estimation frequencies, approaches and proxies for volatility and liquidity. This finding is in line with the theoretical model developed by Collin-Dufresne and Fos (2016),[19] which predicts that the price volatility-informed trading relationship is influenced by two effects. On the one hand, informed trading reveals information, and this decreases uncertainty in financial markets, which reduces price volatility. On the other hand, aggressive trading behaviour on the part of informed traders could increase volatility. Thus, the net impact of informed trading on stock price volatility depends on which effect dominates. Under normal trading conditions, the former effect would naturally dominate. The results are also consistent with the empirical findings of Avramov et al. (2006) and Collin-Dufresne and Fos (2015), who find that price volatility and adverse selection are negatively correlated with informed trading. The negative relationships of informed trading with order flow toxicity and illiquidity are linked to informed traders' use of limit orders rather than (aggressive) market orders.

Furthermore, I find that the transitory component, as estimated using my state space approach, is a significant predictor of one-second stock returns. This implies that although financial markets are efficient in the long-term, there are short-term inefficiencies in markets because investors need time to absorb new information (see Chordia et al., 2008). However, I find that the horizon for short-term stock returns predictability has decreased substantially since the five-minute window reported by Chordia et al. (2008). The predictability of short-horizon returns now only holds on a per second basis, and no longer at the minutes-long threshold reported in earlier studies. I show that high frequency trading is the driver of this sharp reduction in the length of short-term return predictability.

[19] The rational expectation model developed by Wang (1993), via a different mechanism, also predicts a negative relationship between informed trading and stock price volatility.

Several streams of the literature relate to this study. One delineates traders into liquidity- and information-motivated traders (see as an example, Avramov et al., 2006), and another examines the role of the different types of traders on price volatility and liquidity/toxicity (see as examples, Daigler and Wiley, 1999; Van Ness et al., 2016). This chapter differs from both of these streams of the literature in at least three respects. Firstly, the approach of decomposing trading volume using state space modelling is fundamentally different to those employed in existing studies and holds noteworthy economic value/significance over other decomposition methods. Secondly, I examine the role of informed trading activity in the evolution of specific market quality metrics, including for a new market quality metric, market toxicity. Finally, and critically, I present new evidence on the speed of price adjustment in the presence of HFT-driven informed order flow.

3.2 Trading volume and the state space model

3.2.1 The application of state space modelling to trading volume

State space models are a natural tool for modelling an observed variable as the sum of two unobserved variables. The asymptotic unbiasedness and efficiency of their estimation, i.e. maximum likelihood via the Kalman filter (see Brogaard et al., 2014b; Hendershott and Menkveld, 2014), make them best suited to analysing high frequency time series.

In my setting, the state space model decomposes trading volume into two parts: the permanent component of trading volume, which is driven by liquidity trading, and the transitory component of trading volume, which is driven by information-motivated trading. Thus, liquidity-motivated trading is expected to constitute the permanent part of trading volume, while informed order flow is expected to make up the transitory part. In other words, uninformed/liquidity order flow is necessary for trading, while informed order flow is not as critical. These expectations are consistent with the predictions of the models of Glosten and Milgrom (1985) and Suominen (2001).

Firstly, the Glosten and Milgrom (1985) model predicts a partial market breakdown if there is an excessive level of informed traders in the market relative to liquidity traders. This is simply because when there is a dearth of liquidity traders in the market, market makers will aim to protect themselves against being adversely selected by widening the spread. Wider spreads make order execution more difficult and trading less likely. As suggested by Glosten and Milgrom (1985), this prediction is congruous with the well-known lemons problems proposed by Akerlof (1970). It simply implies that trading relies on the *permanent* presence of liquidity traders in the market. The permanent character of liquidity order flow is underscored by the well-known ‘no trade’ theorems. While trading may not be informationally efficient in the absence of informed trades, they can still occur because of the dispersion of beliefs inherent in uninformed order flow. This is not the case when liquidity-seeking order flow is unavailable in the market. Specifically, high levels of informed orders relative to liquidity orders implies that orders will cluster on one side of the order book, leading to no trade scenarios (see Brunnermeier, 2001), since there is no dispersion of belief in informed order flow. This is why Morris (1994) argues that no trade problems can be solved by adding liquidity traders to the market. Therefore, the permanent component of trading volume, as modelled using state space modelling, can be characterised as the liquidity component of trading volume. In addition, generally, the theoretical literature models liquidity traders as random traders (see as an example, Kyle, 1985). In line with this, in the state space representation, the permanent component is modelled as a (nonstationary) random walk.

Secondly, Suominen (2001) shows that after trading reveals the private information held by informed traders, liquidity traders will inevitably revise their pricing and thus become more cautious. This may result in a reduction in informed trading in the market. Furthermore, according to the Efficient Market Hypothesis (EMH), any new information is simultaneously absorbed by traders, and hence can only cause transitory (short-term) changes in trading volume (see Fama, 1970). Similarly, Chordia et al. (2002) argues that private information impacts liquidity temporarily in financial markets. Thus any changes in the information-driven component of trading volume, while having a durable impact on price (see Menkveld et al., 2007), should only affect trading volume temporarily. Consistent with this, in the state space representation, the stationary and transitory component of trading volume as modelled using state space modelling is adopted as a proxy for informed trading activity.

The above arguments provide a firm basis for my modelling approach. Additionally, it is useful to draw comparisons between my state space modelling approach and a related methodological stream of the financial economics literature. When investigating trading behaviour in financial markets, modelling may focus on the duration between transactions as a means of capturing trading intentions, such that the time stamp may be used as an explanatory variable in the mean function of durations. In addition, a cubic spline may be used to smooth out huge variations in the duration effects. Such a model is often regarded as a state space counterpart of the autoregressive conditional duration (ACD) model of Engle and Russell (1998) (see also Durbin and Koopman, 2012).[20] The ACD is suitable for analysing trading data with transactions at irregular intervals, and the model is extensively used in the market microstructure literature to test hypotheses about duration and transaction clustering. In my state space representation, the permanent characteristics of the nonstationary series imply constant duration, whereas the transitory structure of the stationary series requires non-constant duration between transactions. Since the permanent and transitory components of trading volume are motivated by liquidity and information trades respectively, there should be constant (non-constant) duration in liquidity (informed) trading activity. For example, as transactions duration decreases, I would expect an increase in the speed of price adjustment to new information (see Dufour and Engle, 2000). Specifically, if indeed my state space representation is empirically relevant, then I would expect that non-constant duration or duration clustering is

[20] Pacurar (2008) provides a review of the duration modelling literature.

driven by informed trading. The empirical findings in the literature (see as examples, Dufour and Engle, 2000; Engle, 2000; Russell and Engle, 2005; Zhang et al., 2001) are in line with this expectation, and therefore provide an additional set of arguments that further underscore the empirical relevance of my state space approach. However, ultimately, the ACD is an autoregressive model and consequently is less efficient for decomposing an observed variable into unobserved components than the state space modelling approach using maximum likelihood estimation via the Kalman filter (see Brogaard et al., 2014b; Durbin and Koopman, 2012; Hendershott and Menkveld, 2014).

3.2.2 The state space equation

I model trading volume as the sum of a non-stationary permanent (liquidity-driven) component and a stationary transitory (information-driven) component.[21] In its simplest form, the structure of the state space model for trading volume, a multiple of *I* stock prices, *T* intraday periods, and *D* intervals can be expressed as:

$$v_{i,t,\tau} = m_{i,t,\tau} + s_{i,t,\tau} \tag{3.1}$$

and

$$m_{i,t,\tau} = m_{i,t,\tau-1} + u_{i,t,\tau} \tag{3.2}$$

where

$$v_{i,t,\tau} = ln\left(TVolume_{i,t,\tau}\right), \tag{3.3}$$

for $i = 1,\dots,I$ and $\tau = 1,\dots,T$ and $t = 1,\dots,D$; both τ and t index event and calendar times respectively (see Menkveld, 2013). $TVolume_{i,t,\tau}$ is the volume traded in stock *i* at interval *t*

[21] In addition to modelling the natural logarithm of trading volume as an observable variable in the state space representation, for robustness, I also employ level trading volume, percentage changes in trading volume and first difference of trading volume. My inferences are unchanged irrespective of the approach I employ; indeed all the estimates obtained are qualitatively similar.

and period τ, $m_{i,t,\tau}$ is a non-stationary permanent component of the volume traded in stock *i* at interval *t* and period τ, $s_{i,t,\tau}$ is a stationary transitory component of the volume traded in stock *i* at interval *t* and period τ, and $u_{i,t,\tau}$ is an idiosyncratic disturbance error in stock *i* at interval *t* and period τ. $s_{i,t,\tau}$ and $u_{i,t,\tau}$ are assumed to be mutually uncorrelated and normally distributed. The structure of the model shows that only changes in $u_{i,t,\tau}$ affect trading volume permanently; $s_{i,t,\tau}$ is temporary because its effects are ephemeral. By using maximum likelihood (likelihood is constructed using the Kalman filter),[22] I can easily estimate $\sigma_{i,t}^{2u}$ and $\sigma_{i,t}^{2s}$, where *t* equals to one of one second, minute or hour. Specifically, I first partition my sample into one second, minute and hour intervals, then estimate $\sigma_{i,t}^{2u}$ and $\sigma_{i,t}^{2s}$ for these intervals by using trading volume at different periods (τ) during the intervals. This implies that, as in Menkveld et al. (2007), my permanent and transitory components ($\sigma_{i,t}^{2u}$ and $\sigma_{i,t}^{2s}$), as estimated using the state space model, are time variant (see Table 4 in Menkveld et al., 2007: 220). I impose the time variant structure, because I subsequently use the estimated components in multivariate predictive regressions. Brogaard et al. (2014b) also compute time variant permanent and transitory components of an observable variable (price).

According to the structure of my state space model, the permanent component of trading volume is due to the activity of the fraction of the market populated by liquidity traders, while the other fraction of the market populated by informed traders reflects the transitory component of trading volume. It implies that my estimated coefficients ($\sigma_{i,t}^{2u}$ and $\sigma_{i,t}^{2s}$), modelled as variances of permanent and transitory trading volume respectively, can be used as

---

[22] The Kalman filter evaluates the conditional mean and variances of the state vector $\boldsymbol{m_t}$ given past observations $V_{t-1} = \{\boldsymbol{v_1}, .., \boldsymbol{v_{t-1}}\}$: $\boldsymbol{a_{t|t-1}} = \mathrm{E}(\boldsymbol{m_t}|V_{t-1})$, $\boldsymbol{P_{t|t-1}} = \mathrm{var}(\boldsymbol{m_t}|V_{t-1})$, $\quad t = 1, .., N$.

In order to initialize the Kalman filter, I also have $\boldsymbol{a_{1|0}} = \boldsymbol{a}$ and $\boldsymbol{P_{1|0}} = \boldsymbol{P}$, where $\boldsymbol{m_1} \sim N(\boldsymbol{a}, \boldsymbol{P})$. This initialization works only if $\boldsymbol{m_t}$ is a stationary process. However, as in my case, often $\boldsymbol{m_t}$ is not a stationary process. Hence, “diffuse initialization” is done and estimated by numerically maximizing the log-likelihood. This is evaluated by the Kalman filter due to prediction error decompositions. It can be shown that when the model is correctly specified the standardized prediction errors are normally and independently distributed with a unit variance (see Durbin and Koopman, 2012 for further details).

proxies for the two fractions of trading volume, i.e. $\sigma_{i,t}^{2u}$ is a proxy for liquidity-motivated traders and $\sigma_{i,t}^{2s}$ is a proxy for information-motivated traders. Since informed trading occurs only occasionally relative to uninformed trading, which is more regular, I would expect $\sigma_{i,t}^{2s}$ to be higher than $\sigma_{i,t}^{2u}$.

Although a one-second interval is a suitable frequency to investigate high-frequency trading activity, it is a very short interval for trade-based measures such as trading volume; hence, I employ one-minute and one-hour interval analysis for robustness. Furthermore, any interval that has fewer than three transactions is excluded from the sample.

The value of my volume decomposition approach is inextricably linked to the relevance of the estimated transitory and permanent components as proxies for informed and uninformed trading respectively. Therefore, in order to test their empirical relevance, I employ a series of predictive multivariate regressions, which are discussed in the next section. Specifically, I test whether the estimated components of trading volume's impact on market quality proxies are consistent with the predicted and established patterns in the literature. The hypotheses related to these tests are developed in Section 3.2.3.

3.2.3 The empirical relevance of state space decomposition of trading volume: theory and hypotheses

This section develops three hypotheses for testing the relevance of my state space modelling approach.

3.2.3.1 Hypothesis I: state space model-estimated components of trading volume and volatility

Kyle (1985) presents a theoretical model for deriving equilibrium security prices when traders' information sets are asymmetric. The model predicts a constant volatility in a continuous auction system, reflecting information being incorporated into prices at a constant rate. Price volatility in part depends on the informed trader's information as incorporated into prices, and is *"unaffected by the level of noise trading"* (see Kyle, 1985: 1319).[23] Degryse et al. (2013) extend Kyle's (1985) model by adding a large liquidity trader to the framework. They show that when a market maker perceives order flow as uninformed, she does not revise prices, such that the liquidity trader benefits from a lower price impact. This prediction also suggests an insignificant level of uninformed trading-price volatility relationship. Crucially, this relationship relies on a risk neutrality assumption.

Hellwig (1980) takes a more apt approach by assuming that price reflects information derived from the auctioning activity of risk averse agents. This assumption yields a prediction of a positive relationship between liquidity trading and volatility (see also Collin-Dufresne and Fos, 2016; Daigler and Wiley, 1999). Considered together with the well-documented positive relationship between aggregate trading volume and stock price volatility (see as examples, Karpoff, 1987; Lamoureux and Lastrapes, 1990; Lee and Rui, 2002; Park, 2010), the implication of the above prediction is that, in a framework controlling for aggregate trading volume, the positive relationship between volatility and liquidity trading activity dissipates. This is because, as argued by Collin-Dufresne and Fos (2016) and Daigler and Wiley (1999), the positive relationship between trading volume and volatility is driven by liquidity trading. Furthermore, Hellwig (1980) shows that informed trading activity decreases volatility in financial markets (see also Avramov et al., 2006; Wang, 1993), implying a negative relationship between volatility and informed trading activity.

I would therefore expect that the negative relationship between informed trading and volatility will endure in a framework controlling for trading volume. Conversely, there should be no expectation of a statistically significant relationship between liquidity trading and

[23] Kalotychou and Staikouras (2009), reviewing several market microstructure models, argue that, consistent with Kyle's (1985) model, only informed traders contribute to volatility in the long-run.

volatility once volume is controlled for, since liquidity trading is the main driver of the trading volume-volatility relationship. I exploit these predicted relationships in a test of the validity of my state space modelling approach. Specifically, I test the following hypothesis:

Hypothesis I. The state space model-estimated transitory component of trading volume reduces volatility

3.2.3.2 Hypothesis II: state space model-estimated components of trading volume, liquidity and market toxicity

In the market microstructure literature, the bid-ask spread holds economic significance for the market maker (see as an example, Branch and Freed, 1977). Huang and Stoll (1997) show that the bid-ask spread incorporates three costs: the order processing cost, inventory holding cost, and the adverse selection cost. Huang and Stoll (1997) and Bollen et al. (2004) argue that order processing and inventory holding costs respectively are not related to the type of traders active in the market, since a market maker incurs those costs irrespective of who they trade with. However, the adverse selection cost is trader type-dependent. Glosten and Milgrom (1985) and Easley and O'Hara (1987) predict that the adverse selection cost is due to market makers facing adverse selection risk when they trade with informed traders. This means that the bid-ask spread is driven by informed trading activity. Order flow is considered toxic when market makers are adversely selected by informed traders in a high frequency environment (see Easley et al., 2011). Hence, market toxicity is seen as the high frequency equivalent of adverse selection risk. I would therefore expect market toxicity to rise in line with increases in the adverse selection cost and the widening of the bid ask spread. The widening of the bid-ask spread implies a reduction in liquidity.

While an increase in informed trading activity could lead to increased adverse selection risk for the market maker and induce a widening of the spread, this effect is often eclipsed by an overall increase in trading volume due to aggregate (uninformed and informed) trading activity. This is because informed trading mainly occurs in tandem with uninformed trading. According to Admati and Pfleiderer (1988), increases in uninformed/liquidity trading volume go hand in hand with induced informed trading volume, such that liquidity-seeking trading activity provides an opportunity for informed traders to camouflage their trades. This implies that informed traders would normally trade only when their trades could be disguised, and uninformed trading activity offers the opportunity for disguising informed trades. This is logical since if informed orders are identified ahead of execution, they would no longer be beneficial for informed traders and therefore could no longer be considered informed.

Kyle (1985) also states that an increase in noise trading induces a higher level of informed trading (see also Ibikunle, 2018). Ibikunle (2018) specifically provides empirical evidence that informed traders increase their trading activity in the presence of higher trading volumes, which is shown to be dominated by uninformed trading activity. Increased trading activity has the effect of enhancing liquidity and therefore inducing a narrowing of the bid-ask spread (see Barclay and Hendershott, 2003; Biais et al., 1999 for further empirical evidence). Hence, I would expect a positive relationship between market liquidity and informed trading activity. This expectation is consistent with Kyle (1981; 1984; 1985; 1989) showing that informed trading activity is positively related to market liquidity (see also Collin-Dufresne and Fos, 2015). Improvements in liquidity implies a narrowing of the bid-ask spread and by extension a reduction in market toxicity.

Furthermore, according to Collin-Dufresne and Fos (2016), informed traders with long-lived information mainly use limit orders. This helps them avoid detection and leads to a negative correlation between adverse selection and informed trading. Consequently, I test the following hypothesis:

Hypothesis II. The state space model-estimated transitory component of trading volume enhances liquidity and reduces market toxicity.

3.2.3.3 Hypothesis III: state space model-estimated transitory component of trading volume and short-horizon returns

According to Fama (1970), financial markets are largely informationally efficient over a daily horizon. Chordia et al. (2002; 2008) however argue that there are inefficiencies in markets at shorter horizons because traders need time to act on new information. Motivated by this, Chordia et al. (2002; 2008) examine the predictability of short-term returns from lagged order imbalance and find that, indeed, markets are inefficient over short periods. Chordia et al. (2002; 2008) use order imbalance in their own regressions investigating the predictability of short-horizon returns for two reasons. Firstly, order imbalance signals private information, which should result in a permanent price impact (this is also alluded to by Kyle, 1985). Secondly, large order imbalances exacerbate the inventory problem faced by the market maker, leading to quote revisions and changes in the bid-ask spread. Similarly, I argue that my transitory component of trading volume signals private information, and thus I expect the component to be a significant predictor of short-horizon stock returns and, by extension, an inverse predictor of market efficiency (see Chordia et al., 2008; Chung and Hrazdil, 2010).

The informative element of both Chordia et al.'s (2002; 2008) order imbalance measure and my own state space-based transitory component of trading volume[24] measure make them suitable predictors in the short-horizon return predictive regressions. Consequently, my third hypothesis is as follows:

Hypothesis III. The state space model-estimated transitory component of trading volume is a significant predictor of short-horizon returns.

[24] In addition, the idea that returns depend on trading volume (or its components) is consistent with the literature. The relationship between return and lagged trading volume is predicted by the sequential information arrival model developed by Copeland (1976) and Jennings et al. (1981). This model assumes that initially, new information is observed only by a trader, leading to her revising her beliefs and beginning to trade advantageously with the information. This informed trading activity generates a new equilibrium price, and therefore returns (price changes). Specifically, sequential information flow models argue that contemporaneous absolute stock returns can be predictable by lagged trading volume (see also Hiemstra and Jones, 1994).

3.3 Data and measures

3.3.1 Data

I use two sets of data in this study. The first consists of ultra-high frequency tick-by-tick data for the most active 100 S&P 500 stocks, as sourced from the Thomson Reuters Tick History (TRTH) database; trading activity is measured by dollar trading volume. It includes data for the trading days between October 2016 and September 2017. In the data, each message is recorded with a time stamp to the nearest millisecond. The following variables are included in the dataset: Reuters Identification Code (RIC), date, timestamp, price, volume, bid price, ask price, bid volume, and ask volume. I apply Lee and Ready's (1991) algorithm to classify trades as buyer- or seller-initiated.[25] The final dataset after cleaning[26] contains about 216.37 million trades, out of which 106.89 million (109.48 million) are buyer- (seller-) initiated. The total value of all trades captured in the analysis equals US$3.28 trillion.

The second dataset is used to execute additional out of sample tests of the validity of my state space modelling approach. It is a proprietary dataset obtained from NASDAQ, and contains transactions for 120 randomly selected NASDAQ and NYSE-listed stocks trading during all the trading days in 2009. The data is complementary to the first dataset I employ because it disaggregates transactions into those executed based on orders submitted by HFTs and non-HFTs. This is the same dataset described in detail by Brogaard et al. (2014b). The dataset contains the following information on each transaction included in the sample: date, time (in milliseconds), transaction size (shares), price, buy-sell indicator, and liquidity nature of the two sides to each trade (HH, HN, NH and NN). HH indicates a trade based on an HFT demanding liquidity and an HFT supplying the required liquidity. HN implies that an HFT demands liquidity and a non-HFT supplies liquidity, while NH is the opposite. NN refers to

[25] Chakrabarty et al. (2015) compare the different trades classification methods and conclude that Lee and Ready's (1991) is the most accurate method.

[26] I follow Chordia et al. (2001) and Ibikunle (2015) in applying a standard set of exclusion criteria to the data, with the aim of eliminating inexplicable values due to erroneous data entry.

trades where both counterparties are non-HFTs. I identify the sum of HH, HN and NH as HFT volume. Based on this classification, HFTs are counterparties in about 71.30% of all trades in the sample. The NASDAQ-provided dataset is only used in Section 3.5 of this chapter, where further justification for its use is outlined.

3.3.2 Measures and descriptive statistics

In order to conduct a joint test of the empirical relevance of my state space modelling approach and the impact of liquidity and informed trading on price volatility, liquidity, and market toxicity, I estimate a set of predictive regressions. Thus, apart from the state space-estimated permanent and transitory components of trading volume, my volatility, liquidity, and market toxicity measures are the main variables of interest. Below I elaborate on how these and other relevant variables are computed.

3.3.2.1 Volatility measures

Consistent with the literature, I use the absolute value of price changes, $\left|\Delta p_{i,t}\right|$, as the main proxy for stock price volatility (see Karpoff, 1987). $\Delta p_{i,t}$ is the difference in price change between the last transaction prices, *p*, for stock *i* at intervals *t* and *t-1*.

For robustness, I also proxy volatility using the standard deviation of stock returns $\sigma_{i,t}^{R}$ (see Barclay and Hendershott, 2003; 2008; Lamoureux and Lastrapes, 1990; Malceniece et al., 2018), where *R* is the midpoint-to-midpoint return with each midpoint computed using the best bid and ask quotes corresponding to each transaction in stock *i* during interval *t*; *R* is thus defined in event/transaction time. The standard deviation of these returns within each interval *t* is my volatility measure. This midpoint-based approach is used in order to reduce the incidence of bid-ask bounce, which transaction prices are susceptible to (see Avramov et al., 2006; Chordia et al., 2008). However, an alternative set of volatility estimates computed from transaction prices do not yield materially different results. Interval t corresponds to one of one second, minute or hour for both volatility proxies.

3.3.2.2 Liquidity measures

For robustness, I employ three spread measures as proxies for liquidity; the spread metrics are the effective spread ($Espread_{i,t}$), quoted spread ($Qspread_{i,t}$), and relative spread ($Rspread_{i,t}$). The $Rspread_{i,t}$ is obtained by dividing the difference between interval t's best ask and bid prices by the midpoint of both prices for stock i, while the $Qspread_{i,t}$ is simply the difference between interval t's best ask and bid prices for stock i. The $Espread_{i,t}$ is twice the absolute value of the difference between the last transaction price for stock i in interval t and the midpoint of the prevailing bid and ask prices when the transaction occurs for stock i. Interval t corresponds to one of one second, minute or hour for all liquidity proxies.

3.3.2.3 Market toxicity

I use the nominal order imbalance metric employed by Chordia et al. (2008) as a proxy for the level of order flow toxicity in the market; in this chapter, I call the measure $MT_{i,t}$. This is because existing order toxicity measures, such as the volume synchronised probability of informed trading (VPIN - see Easley et al., 2012), essentially capture the essence of order imbalance in the market and thus are highly correlated with $MT_{i,t}$. $MT_{i,t}$ is computed as the absolute value of the number of buyer-initiated trades minus the number of seller-initiated trades divided by the total number of trades for stock i during interval t, where t corresponds to one of one-minute or one-hour. I employ only minute and hour intervals because it is challenging to obtain enough trading volume for the lower volume stocks to compute unbiased order imbalance metrics within a one-second interval.

3.3.2.4 Volume measures

In my state space model, trading volume is the observable variable, which is then decomposed into unobservable proxies of liquidity trading activity $\left(\sigma_{i,t}^{2u}\right)$ and informed trading activity $\left(\sigma_{i,t}^{2s}\right)$ in stock *i* at time *t*. Thus, the proxies could be mechanically correlated with trading activity and volume. In order to ascertain that observed effects of the proxies are not due to aggregate trading volume, I need to include at least one proxy for trading volume/activity in my secondary models. This is particularly important in my framework since Andersen and Bondarenko (2014) show that the relationship between VPIN, also estimated from trading volume, and future short-term volatility is trivial after controlling for mechanic correlation between VPIN and trading volume. Controlling for trading volume/activity in my secondary models addresses the Andersen and Bondarenko (2014) criticism. I employ the natural logarithm of trading volume for stock *i* at time *t* ($TV_{i,t}$) as a proxy for trading volume. A second trading activity-related proxy is also included in my models: $BSI_{i,t}$. $BSI_{i,t}$ is the absolute value of the difference between buyer- and seller-initiated trades for stock *i* during interval *t*. According to Chordia et al. (2002), the metric adequately proxies trading activity because it strongly influences prices and liquidity.

Table 3. 1 Definitions of variables and descriptive statistics

The table defines the variables calculated for each stock-interval, *i*, *t*, and reports the descriptive statistics. All variables, except $MT_{i,t}$, are computed at a one-second frequency (*t* equals one-second). $MT_{i,t}$ is computed at a one-minute frequency. The sample contains the most active 100 S&P 500 stocks traded between October 1, 2016 through to September 30, 2017 on NYSE and NASDAQ.

Variable	Description	Mean	Median	Stand. Deviation
$Espread_{i,t}$	Effective spread for stock *i* at interval *t*. Computed as twice the absolute value of the difference between the last execution price and the midpoint of the prevailing bid and ask prices at interval t.	0.0090	0.0100	0.0462

$Rspread_{i,t}$	Relative spread for stock *i* at interval *t*. Computed as the difference between the best ask and bid prices divided by the midpoint of both prices during interval *t*.	0.0003	0.0002	0.0009
$Qspread_{i,t}$	Quoted spread for stock *i* at interval *t*. Computed as the difference between the best ask and bid prices during interval *t*.	0.0186	0.0100	0.0564
$BSI_{i,t}$	Absolute difference between buyer- and seller-initiated trades for stock *i* during interval *t*.	1584.05	424.00	35771
$\|\Delta p_{i,t}\|$	Absolute value of price change for stock *i* during interval *t*. Computed as the absolute value of the differences between last prices at intervals *t* and *t-1*.	0.0091	0.0090	0.0670
$R_{i,t}$	Midpoint-to-midpoint return for stock *i* during interval *t*. Computed as the difference between the midpoints corresponding to the last transactions at intervals *t* and *t-1* divided by the midpoint corresponding to the last transaction at interval *t-1*	$-0.412\text{x}10^{-6}$	0.00	0.0013
$\sigma_{i,t}^{R}$	Standard deviation of midpoint-to-midpoint returns for stock *i* during interval *t*; each midpoint during the interval *t* corresponds to a transaction occurring during the interval.	$0.92\text{x}10^{-4}$	$0.59\text{x}10^{-4}$	0.0009
$MT_{i,t}$	Market toxicity for stock *i* for interval *t*. Computed as the absolute value of the difference between the numbers of buy and sell trades divided by the sum of the numbers of buy and sell trades occurring during interval *t*.	0.5406	0.5037	0.3419

Table 3.1 presents the descriptive statistics for volatility, liquidity, market toxicity and volume metrics. Midpoint return estimates for stock *i* at time *t*, $R_{i,t}$s, are also presented. All measures except that of market toxicity ($MT_{i,t}$) are based on one-second computations; $MT_{i,t}$ is based on one-minute calculations. Consistent with recent evidence (see as an example, Malceniece et al., 2018), the spread measures are tight, with the average $Espread_{i,t}$, $Rspread_{i,t}$, and $Qspread_{i,t}$ corresponding to 0.009, 0.0004, and 0.018, respectively. Average midpoint returns

are weakly negative over my sample period. However, volatility is generally low irrespective of which proxy I focus on. The mean and median for $|\Delta p_{i,t}|$ are about 0.0092 and 0.009 respectively, while $\sigma_{i,t}^{R}$ is lower still at 0.00009.

3.4 Analysis of state space decomposition of trading volume

3.4.1 State space decomposition of trading volume: estimates

Table 3.2 presents the cross-sectional mean estimated values of the permanent (liquidity-driven) and transitory (information-driven) components of trading volume as decomposed using the state space model.

Table 3. 2 State Space Estimates

The table contains mean cross-sectional estimates of transitory (information-driven) and permanent (liquidity-driven) components of trading volume for the most active 100 S&P 500 stocks trading between October 1, 2016 and September 30, 2017. Stocks are divided into quartiles according to their level of trading activity; trading activity is based on trading volume. Quartile 1 contains the least active companies, while Quartile 4 contains the most active stocks. The estimates are based on the following state space model for decomposing trading volume into its transitory and permanent components:

$$v_{i,t,\tau} = m_{i,t,\tau} + s_{i,t,\tau} \,; m_{i,t,\tau} = m_{i,t,\tau-1} + u_{i,t,\tau}$$

where $v_{i,t,\tau} = ln(TVolume_{i,t,\tau})$, *i = 1,…,I (stocks)*, *t = 1,…,D (intervals)*, τ *= 1,…,T (periods)*, $TVolume_{i,t,\tau}$ corresponds to the trading volume of stock *i* at interval *t* and period τ, $m_{i,t,\tau}$ is a non-stationary permanent component of stock *i* at interval *t* and period τ, $s_{i,t,\tau}$ is a stationary transitory component for stock *i* at interval *t* and period τ and $u_{i,t,\tau}$ is an idiosyncratic disturbance error for stock *i* at interval *t* and period τ. $\sigma_{i,t}^{2s}$ and $\sigma_{i,t}^{2u}$ are the respective estimates of the transitory and permanent components of trading volume for stock *i* and interval *t*, estimated by maximum likelihood (constructed using the Kalman filter). Estimations are presented for one-second, one-minute, and one-hour frequencies (*t* equals one-second, one-minute and one-hour).

	Stock quartiles			
Variable	Least active	2	3	Most active
	One-second frequency (*t* equals one-second)			
$\sigma_{i,t}^{2s}$	1.02	1.24	1.37	1.51
$\sigma_{i,t}^{2u}$	0.46	0.49	0.53	0.78
	One-minute frequency (*t* equals one-minute)			
$\sigma_{i,t}^{2s}$	1.21	1.36	1.63	1.88
$\sigma_{i,t}^{2u}$	0.49	0.55	0.72	0.85
	One-hour frequency (*t* equals one-hour)			

$\sigma_{i,t}^{2_s}$	1.34	1.65	1.77	1.96
$\sigma_{i,t}^{2_u}$	0.51	0.59	0.76	0.97

The results are presented for the mean estimates based on one-second, one-minute, and one-hour estimations. For improved insight, I divide my sample into quartiles according to their level of trading activity; trading activity is measured by dollar trading volume. The stocks in Quartile 1 are the least active stocks, while Quartile 4 stocks are the most active. As expected, the mean $\sigma_{i,t}^{2_s}$ is consistently higher than the mean $\sigma_{i,t}^{2_u}$ across all quartiles, irrespective of the estimation frequency of the state space model. This is consistent with the structure of my state space modelling approach. Informed trades are modelled as transitory, occurring only when traders have an informational advantage in the market, while uninformed trades are a permanent fixture in markets. This implies a higher variance for informed trades, hence I would expect higher estimates for $\sigma_{i,t}^{2_s}$ relative to $\sigma_{i,t}^{2_u}$.

Informed traders are, strategically, more active when trading volume and liquidity trading are high, because higher trading volumes provide better "camouflage" for informed trades (see Admati and Pfleiderer, 1988). The estimates presented in Table 3.2 are consistent with this widely held view in the market microstructure literature. The mean variance of liquidity-motivated trades in Quartile 4 is higher than the mean variance of liquidity trades in all of the other quartiles, and is lowest in Quartile 1. This suggests that informed traders should be most active in Quartile 4 and least active in Quartile 1. The transitory component estimates in Table 3.2 are completely in line with this expectation. The mean transitory component in Quartile 4 are 1.51, 1.88 and 1.96 for the one-second, one-minute and one-hour estimations respectively. These estimates are 48%, 55.37% and 46.27% larger than the one-second, one-minute, and one-hour frequencies mean estimated values for Quartile 1 stocks at 1.02, 1.21, and 1.34 respectively.

Inferring from the Kyle (1985) and Glosten and Milgrom (1985) models, when uninformed traders are scarce in the market, the price discovery process becomes impaired or even breaks down. This is because the prospect of compensation for gathering information is reduced in markets where uninformed traders are few, and this leads to fewer than optimal potential informed traders being incentivised to acquire information. The absence of informed traders in the markets impairs the price discovery process, since their trades convey information to the market. Thus, both liquidity and informed traders are critical to the price discovery process. An approach that allows me to directly estimate the proportion of trading volume that can be attributed to both types of traders is therefore valuable in several contexts, not least in market reporting activities, investment management, and policy/regulations development. For example, firm managers' responses to the so-called *speeding ticket* (Price and Volume Query) often issued by some exchanges, such as the Australian Securities Exchange, focus mainly on explaining the evolution of trading volume.

3.4.2 State space decomposition of trading volume: analysis of empirical relevance

3.4.2.1 Hypothesis I: state space model-estimated components of trading volume and volatility

I estimate the multivariate predictive model presented in Equation (3.4) in order to examine the relationship between the state space estimated proxies of liquidity $\left(\sigma_{i,t}^{2u}\right)$ and informed $\left(\sigma_{i,t}^{2s}\right)$ trading activities on the one hand and volatility on the other. This is a direct test of *Hypothesis I* in Section 3.2.3.1.

$$\left|\Delta p_{i,t}\right| = \alpha + \beta_1 Espread_{i,t-1} + \beta_2 TV_{i,t-1} + \beta_3 BSI_{i,t-1} + \beta_4 \sigma_{i,t-1}^{2s} + \beta_5 \sigma_{i,t-1}^{2u} + \varepsilon_{i,t} \quad (3.4),$$

where all variables are as defined in Section 3.3.2. Equation (3.4) is estimated at one-second, one-minute, and one-hour intervals. $\sigma_{i,t-1}^{2s}$ and $\sigma_{i,t-1}^{2u}$, the proxies for informed and uninformed/liquidity trading respectively, are estimated from trading volume. Multicollinearity may therefore be of potential concern, since I include two proxies of trading activity ($TV_{i,t-1}$ and $BSI_{i,t-1}$) in the model. However, as shown in Table 3.3, this is not the case. Note that my state space representation models informed and uninformed trading volume as variances of transitory and permanent trading volume. I employ these variance measures as proxies of informed and uninformed trading volume in Equation (3.4), and subsequent models. Therefore, collinearity is not expected in the regression framework. Consistent with this view, the correlation coefficient estimates presented in Table 3.3 show that there are no multicollinearity issues in my empirical models.

Table 3. 3 Correlation matrix for variables

The table plots the correlation matrix of the variables employed in this study's models. One-second frequency (t equals one-second) is used to compute all variables. $Qspread_{i,t}$ is the quoted spread for stock i for interval t and is computed as the difference between the best ask and bid prices for interval t. $Rspread_{i,t}$ is the relative spread for stock i and interval t and is computed as the difference between the best ask and bid prices divided by the midpoint of both prices for interval t. $Espread_{i,t}$ is the effective spread for stock i for interval t and computed as twice the absolute value of the difference between the last execution price and the midpoint of the prevailing bid and ask prices for interval t, $TV_{i,t}$ is the natural logarithm of trading volume for stock i for interval t, while $BSI_{i,t}$ is the absolute difference between buyer- and seller-initiated traders for stock i for interval t. $\sigma_{i,t}^{2s}$ is the state space model-estimated transitory component of trading volume and is the proxy for informed trading in stock i during interval t, while $\sigma_{i,t}^{2u}$ is the state space model-estimated permanent component of trading volume and is the proxy for liquidity trading in stock i during interval t. $|\Delta p_{i,t}|$ is the absolute value of price change for stock i for interval t and is computed as the absolute value of the differences between the last prices at intervals t and t-1, while $\sigma_{i,t}^{R}$ is the standard deviation of midpoint-to-midpoint returns for stock i during interval t; each midpoint corresponds to a transaction. The sample contains the most active 100 S&P 500 stocks traded between October 1, 2016 and September 30, 2017 on NYSE and NASDAQ.

	$Qspread_{i,t}$	$Rspread_{i,t}$	$Espread_{i,t}$	$TV_{i,t}$	$BSI_{i,t}$	$\sigma_{i,t}^{2s}$	$\sigma_{i,t}^{2u}$	$\|\Delta p_{i,t}\|$	$\sigma_{i,t}^{R}$
$Qspread_{i,t}$	1								
$Rspread_{i,t}$	0.79909	1							
$Espread_{i,t}$	0.90722	0.72476	1						
$TV_{i,t}$	-0.04144	-0.06261	-0.01673	1					
$BSI_{i,t}$	0.00133	0.01265	0.00284	0.11090	1				
$\sigma_{i,t}^{2s}$	0.00000	0.00013	0.00007	0.00326	0.44342	1			
$\sigma_{i,t}^{2u}$	0.00000	-0.00001	0.00005	0.00021	-0.00001	-0.00000	1		
$\|\Delta p_{i,t}\|$	0.08621	0.05052	0.06789	0.01904	0.01101	0.00004	0.00008	1	
$\sigma_{i,t}^{R}$	0.12381	0.16384	0.11483	0.01700	0.01050	0.00004	0.00001	0.42911	1

As stated in Section 3.2.1, for robustness, I employ a second volatility proxy, i.e. the standard deviation of midpoint returns. Consistent with the literature, I include the proxy's lagged value as an additional explanatory variable (see as examples, Justiniano and Primiceri, 2008; Schwert, 1989) in Equation (3.5):

$$\sigma_{i,t}^{R} = \alpha + \beta_1 \sigma_{i,t-1}^{R} + \beta_2 Espread_{i,t-1} + \beta_3 TV_{i,t-1} + \beta_4 BSI_{i,t-1} + \beta_5 \sigma_{i,t-1}^{2s} + \beta_6 \sigma_{i,t-1}^{2u} + \varepsilon_{i,t} \quad (3.5),$$

where all variables are as previously defined. Equations (3.4) and (3.5) are estimated using time fixed effects with panel corrected standard errors. For robustness, I also include stock fixed effect and both stock and time fixed effects jointly. All of the estimation approaches yield qualitatively similar results.

Both of the volatility proxies I employ encapsulate all variation in stock prices; no distinction is made between permanent and temporary price changes. This approach is based on the extensive market microstructure literature stream investigating the impact of various market phenomena on market quality proxies (see as examples the recent works by Buti et al., 2011; Comerton-Forde and Putniņš, 2015; Malceniece et al., 2018). The purpose of this analysis is to test the empirical relevance of my state space modelling approach by verifying whether the estimated components of trading volume affect market quality variables as predicted in the literature (see Section 3.2.3), hence my adoption of the volatility measures developed in the existing literature.

The results obtained from the estimation of Equations (3.4) and (3.5) are presented in Table 3.4.

Table 3. 4 Predictive regressions of market volatility on lagged components of trading volume

The predictive power of one-second/minute/hour permanent and transitory components of trading volume is estimated using the following models:

$$|\Delta p_{i,t}| = \alpha + \beta_1 Espread_{i,t-1} + \beta_2 TV_{i,t-1} + \beta_3 BSI_{i,t-1} + \beta_4 \sigma_{i,t-1}^{2s} + \beta_5 \sigma_{i,t-1}^{2u} + \varepsilon_{i,t}$$

$$\sigma_{i,t}^{R} = \alpha + \beta_1 \sigma_{i,t-1}^{p} + \beta_2 Espread_{i,t-1} + \beta_3 TV_{i,t-1} + \beta_4 BSI_{i,t-1} + \beta_5 \sigma_{i,t-1}^{2s} + \beta_6 \sigma_{i,t-1}^{2u} + \varepsilon_{i,t}$$

where $|\Delta p_{i,t}|$ is the absolute value of price change for stock *i* and interval *t* and computed as the absolute value of the differences between last prices at intervals *t* and *t-1*, $Espread_{i,t-1}$ is the effective spread for stock *i* for interval *t-1* and computed as twice the absolute value of the difference between the last execution price and the midpoint of the prevailing bid and ask prices for interval *t-1*. $\sigma_{i,t-1}^{R}$ is the standard deviation of midpoint-to-midpoint returns for stock *i* during interval *t-1*; each midpoint corresponds to a transaction. $TV_{i,t-1}$ is the natural logarithm of trading volume for stock *i* and interval *t-1*, and $BSI_{i,t-1}$ is the absolute difference between buyer- and seller-initiated traders for stock *i* and interval *t-1*. $\sigma_{i,t-1}^{2s}$ and $\sigma_{i,t-1}^{2u}$ are state space model-estimated proxies (estimated using Kalman filter constructed maximum likelihood) for informed and uninformed trading activity respectively for stock *i* and interval *t-1*. The sample contains the most active 100 S&P 500 stocks traded between October 1, 2016 and September 30, 2017 on NYSE and NASDAQ. ***, ** and * correspond to statistical significance at the 0.01, 0.05 and 0.10 levels, respectively.

Panel A

Dependent Variable: $\|\Delta p_{i,t}\|$			
	One-second frequency	One-minute frequency	One-hour frequency
$Intercept$	0.643x10^{-2***} (625.22)	0.192x10^{-1***} (132.36)	0.116x10^{-1***} (21.50)
$Espread_{i,t-1}$	0.595x10^{-1***} (234.89)	0.463x10^{-1***} (77.39)	0.323x10^{-1***} (15.08)
$TV_{i,t-1}$	0.921x10^{-3***} (19.90)	0.402x10^{-2***} (5.74)	0.169x10^{-2***} (4.66)
$BSI_{i,t-1}$	0.571x10^{-6***} (166.23)	0.132x10^{-6***} (129.64)	0.727x10^{-6***} (12.20)
$\sigma_{i,t-1}^{2s}$	-0.325x10^{-4***} (-10.82)	-0.412x10^{-3***} (-7.33)	-0.326x10^{-2***} (-4.97)
$\sigma_{i,t-1}^{2u}$	0.864x10^{-5} (0.83)	-0.215x10^{-4} (-0.27)	-0.827x10^{-4} (-0.42)
Sample size *(n)*	29959938	8880028	204354
Fixed effects	Time	Time	Time
$\overline{R^2}$	0.87 %	2.49 %	5.58 %

Panel B

Dependent Variable: $\sigma_{i,t}^{R}$			
	One-second frequency	One-minute frequency	One-hour frequency
$Intercept$	0.724x10^{-4***} (413.26)	0.780x10^{-4***} (255.85)	0.848x10^{-4***} (14.40)
$\sigma_{i,t-1}^{R}$	0.333x10^{-1***} (64.52)	0.375x10^{-1***} (52.94)	0.616*** (88.49)
$Espread_{i,t-1}$	0.161x10^{-2***} (367.79)	0.155x10^{-2***} (53.28)	0.475x10^{-4} (1.53)
$TV_{i,t-1}$	0.753x10^{-5***} (16.46)	0.780x10^{-5***} (8.09)	0.107x10^{-4***} (5.18)
$BSI_{i,t-1}$	0.121x10^{-8***} (265.09)	0.255x10^{-8***} (140.46)	0.230x10^{-8***} (23.75)

$\sigma_{i,t-1}^{2s}$	-0.613x10^{-6***} (-17.75)	-0.729x10^{-5***} (-14.25)	-0.903x10^{-3***} (-12.37)
$\sigma_{i,t-1}^{2u}$	0.725x10^{-7} (0.08)	0.684x10^{-6} (0.02)	-0.578x10^{-3} (-0.01)
Sample size *(n)*	29959938	8880028	204354
Fixed effects	Time	Time	Time
$\overline{R^2}$	1.12%	3.24%	8.73%

The inferences drawn from the estimates in Table 3.4 are consistent across all frequency estimations. The coefficient estimates show that lagged $\sigma_{i,t}^{2s}$ is a significant predictor of the absolute value of price changes, $|\Delta p_{i,t}|$, and the standard deviation of stock returns, $\sigma_{i,t}^{R}$; the $\sigma_{i,t}^{2s}$ coefficients are statistically significant at the 0.01 level. The negative coefficient estimates indicate that increases in information-motivated trades reduce price volatility in financial markets. This result is consistent with the findings of Avramov et al. (2006), who find that stock price volatility is negatively correlated with informed traders (see also Hellwig, 1980; Wang, 1993). *Hypothesis I* is therefore upheld.

In contrast, $\sigma_{i,t}^{2u}$ is not a significant predictor of volatility once I control for volume. This is because the positive relationship between trading volume and volatility is driven by trading volume due to liquidity trading (see Collin-Dufresne and Fos, 2016; Daigler and Wiley, 1999).[27] The significant negative $\sigma_{i,t}^{2s}$ and the insignificant $\sigma_{i,t}^{2u}$ coefficient estimates imply a validation of my state space approach to decomposing trading volume into informed and liquidity-driven components.

I note that while the coefficient estimates are consistent for all estimation frequencies across both panels, the impact of $\sigma_{i,t}^{2s}$ is stronger for lower frequencies. For example, in Panel

[27] For robustness and in a test of the arguments presented by Collin-Dufresne and Fos (2016) and Daigler and Wiley (1999), i.e. that a positive volume-volatility relation is driven by liquidity trading, I exclude the trading volume proxy from a follow-up model. I find that once trading volume is not controlled for, the liquidity trading proxy becomes a positive and statistically significant predictor of volatility. For parsimony, I do not show this result, however it is available upon request.

A (B), the effect of $\sigma_{i,t}^{2s}$ on volatility proxies for the one-hour frequency estimation is 6.65 (124.28) and 98.80 (1,249) times larger than that of the one-minute and one-second frequency estimations respectively. These differences are due to more information being typically released over longer intervals. It is plausible that the market learns more about the developments relevant to an instrument over an hour than over a second or a minute, or at the very least, comes to terms more with new information over a longer time horizon. The estimated coefficients for all of the other explanatory variables are consistent with the existing literature.

The explanatory powers of the one-second regressions are low, with the $\overline{R^2}$ being only about 0.87% for $\left|\Delta p_{i,t}\right|$ in Panel A and 2.49% for $\sigma_{i,t}^{R}$ in Panel B. This is unsurprising and is because I estimate the models at a one-second frequency, with very little information being released during the very narrow window (see Chordia et al., 2008). Consequently, the $\overline{R^2}$ estimates are larger for the one-minute and one-hour frequencies, which are 3.24% and 8.73% respectively in Panel B.

3.4.2.2 Hypothesis II: state space model-estimated components of trading volume, liquidity and market toxicity

I next test *Hypothesis II* from Section 3.2.3.2. Specifically, I investigate the nature of the relationship between my state space model-estimated components of trading volume on the one hand, and liquidity and market toxicity on the other. For this purpose, I estimate the following multivariate predictive models:

$$Spread_{i,t} = \alpha + \beta_1\sigma_{i,t-1}^{p} + \beta_2 TV_{i,t-1} + \beta_3 BSI_{i,t-1} + \beta_4\sigma_{i,t-1}^{2s} + \beta_5\sigma_{i,t-1}^{2u} + \varepsilon_{i,t} \quad (3.6),$$

$$MT_{i,t} = \alpha + \beta_1\sigma_{i,t-1}^{p} + \beta_2 TV_{i,t-1} + \beta_3 BSI_{i,t-1} + \beta_4\sigma_{i,t-1}^{2s} + \beta_5\sigma_{i,t-1}^{2u} + \varepsilon_{i,t} \quad (3.7),$$

where $Spread_{i,t}$ corresponds to one of $Espread_{i,t}$, $Qspread_{i,t}$, and $Rspread_{i,t}$. All variables are as previously defined. Equation (3.6) is estimated at one-second, one-minute, and one-hour frequencies, while Equation (3.7) is estimated at one-minute and one-hour frequencies only. This is because trading activity during a one-second interval is minimal and not substantial enough to compute $MT_{i,t}$ within the interval in an unbiased manner.

Table 3. 5 Predictive regressions of market liquidity on lagged components of trading volume

The predictive power of the state space-estimated lagged permanent and transitory components of trading volume is estimated using the following model:

$$Spread_{i,t} = \alpha + \beta_1 \sigma^{R}_{i,t-1} + \beta_2 TV_{i,t-1} + \beta_3 BSI_{i,t-1} + \beta_4 \sigma^{2_s}_{i,t-1} + \beta_5 \sigma^{2_u}_{i,t-1} + \varepsilon_{i,t}$$

where $Spread_{i,t}$ corresponds to one of $Espread_{i,t}$, $Qspread_{i,t}$ and $Rspread_{i,t}$. $Espread_{i,t}$ is the effective spread for stock *i* at interval *t* and is computed as twice the absolute value of the difference between the last execution price and the midpoint of the prevailing bid and ask prices for interval *t*, $Rspread_{i,t}$ is the relative spread for stock *i* at interval *t* and is obtained by dividing the difference between the best ask and bid prices by the midpoint of both prices for interval *t*, $Qspread_{i,t}$ is the quoted spread for stock *i* for interval *t* and computed as the difference between the best ask and bid prices for interval *t*. $\sigma^{R}_{i,t-1}$ is the standard deviation of mid-price returns for stock *i* during interval *t-1* and calculated as the standard deviation of midpoint-to-midpoint returns during interval *t-1*; each midpoint corresponds to a transaction. $TV_{i,t-1}$ is the natural logarithm of trading volume for stock *i* during interval *t-1* and $BSI_{i,t-1}$ is the absolute difference between buyer- and seller-initiated traders for stock *i* during interval *t-1*. $\sigma^{2_s}_{i,t-1}$ and $\sigma^{2_u}_{i,t-1}$ are state space model-estimated proxies (estimated using Kalman filter constructed maximum likelihood) for informed and uninformed trading activity respectively for stock *i* and interval *t-1*. The sample contains the most active 100 S&P 500 stocks traded between October 1, 2016 and September 30, 2017 on NYSE and NASDAQ. ***, ** and * correspond to statistical significance at the 0.01, 0.05 and 0.10 levels, respectively.

Panel A

Dependent Variable: $RSpread_{i,t}$			
	One-second frequency	One-minute frequency	One-hour frequency
$Intercept$	0.443x10^{-3***} (232.56)	0.518x10^{-3***} (197.78)	0.594x10^{-3***} (48.25)
$\sigma^{R}_{i,t-1}$	0.574x10^{-1***} (243.51)	0.196x10^{-3***} (55.67)	0.321x10^{-1***} (75.11)
$TV_{i,t-1}$	0.498x10^{-5} (0.54)	0.978x10^{-4} (1.57)	-0.482x10^{-4} (-0.27)
$BSI_{i,t-1}$	0.135x10^{-8***} (290.72)	0.322x10^{-8***} (253.17)	0.868x10^{-8***} (65.88)
$\sigma^{2_s}_{i,t-1}$	-0.888x10^{-5***} (-21.43)	-0.964x10^{-4***} (-14.53)	-0.305x10^{-4***} (-12.64)
$\sigma^{2_u}_{i,t-1}$	-0.349x10^{-6} (-0.04)	-0.234x10^{-4} (-0.09)	-0.645x10^{-4} (-0.03)
Sample size *(n)*	29959938	8880028	204354
Fixed effects	Time	Time	Time
$\overline{R^2}$	1.20%	2.76%	19.49%

Panel B

Dependent Variable: $QSpread_{i,t}$			
	One-second frequency	One-minute frequency	One-hour frequency
$Intercept$	$0.177 \times 10^{-1***}$ (184.56)	$0.175 \times 10^{-1***}$ (95.13)	$0.148 \times 10^{-1***}$ (25.21)
$\sigma_{i,t-1}^{R}$	2.461^{***} (209.08)	2.27^{***} (102.41)	114^{***} (53.72)
$TV_{i,t-1}$	$-0.866 \times 10^{-3*}$ (-1.81)	-0.144×10^{-3} (-0.36)	-0.376×10^{-2} (-1.33)
$BSI_{i,t-1}$	$0.867 \times 10^{-7***}$ (199.53)	$0.954 \times 10^{-7***}$ (153.56)	$0.345 \times 10^{-7***}$ (34.17)
$\sigma_{i,t-1}^{2s}$	$-0.256 \times 10^{-4***}$ (-16.53)	$-0.352 \times 10^{-4***}$ (-9.51)	$-0.114 \times 10^{-3***}$ (-6.43)
$\sigma_{i,t-1}^{2u}$	0.157×10^{-6} (0.09)	0.370×10^{-6} (0.32)	-0.289×10^{-5} (-0.22)
Sample size *(n)*	29959938	8880028	204354
Fixed effects	Time	Time	Time
$\overline{R^2}$	1.05%	2.02%	18.36%

Panel C

Dependent Variable: $ESpread_{i,t}$			
	One-second frequency	One-minute frequency	One-hour frequency
$Intercept$	$0.792 \times 10^{-2***}$ (104.07)	$0.928 \times 10^{-2***}$ (72.07)	$0.916 \times 10^{-2***}$ (17.26)
$\sigma_{i,t-1}^{R}$	2.633^{***} (137.44)	15.66^{***} (74.68)	108.56^{***} (13.55)
$TV_{i,t-1}$	$-0.161 \times 10^{-3***}$ (-4.32)	-0.118×10^{-3} (-0.47)	-0.189×10^{-2} (-0.35)
$BSI_{i,t-1}$	$0.567 \times 10^{-7***}$ (125.39)	$0.646 \times 10^{-7***}$ (89.28)	$0.216 \times 10^{-6***}$ (27.35)
$\sigma_{i,t-1}^{2s}$	$-0.209 \times 10^{-4***}$ (-11.10)	$-0.321 \times 10^{-4***}$ (-12.92)	$-0.122 \times 10^{-3***}$ (-10.65)
$\sigma_{i,t-1}^{2u}$	-0.764×10^{-4} (-0.23)	-0.187×10^{-4} (-0.13)	-0.205×10^{-4} (-0.13)
Sample size *(n)*	29959938	8880028	204354
Fixed effects	Time	Time	Time
$\overline{R^2}$	1.77%	2.16%	16.43%

Panels A, B, and C of Table 3.5 show the results for Equation (3.6), where $Rspread_{i,t}$, $Qspread_{i,t}$ and $Espread_{i,t}$ correspond to $Spread_{i,t}$ respectively. The negative and statistically significant (p-value <0.01) $\sigma^{2s}_{i,t-1}$ coefficient estimates show that, consistent with *Hypothesis II* and the predictions of Kyle (1981; 1984; 1985; 1989), informed trading activity is positively linked to liquidity. By contrast, $\sigma^{2u}_{i,t-1}$'s coefficient estimates are not statistically significant, suggesting that $\sigma^{2u}_{i,t}$ is not a significant predictor of liquidity once I control for volume and order flow dynamics. This is in line with the estimates obtained in the estimation of Equation (3.5). The results are also consistent with the empirical findings of Collin-Dufresne and Fos (2015) and suggest that, as predicted by Kaniel and Liu (2006), informed traders use limit orders rather than market orders. The coefficients of all of the control variables are in line with the consistent literature. The consistency of the results with the literature emphasize the relevance of my state space modelling approach. Similar to the price volatility model, $\overline{R^2}$ values in Panels A, B, and C are generally small for the one-second and one-minute high frequency estimations, with estimates ranging from 1.05% to 2.76%. The low $\overline{R^2}$ values are due to the estimation frequencies. Hence, the one-hour frequency models have much higher levels of explanatory powers. In Panels A, B, and C, the $\overline{R^2}$ values are 19.49%, 18.36% and 16.43% respectively for the one-hour frequency estimations.

Table 3. 6 Predictive regressions of market toxicity on lagged components of trading volume

The predictive power of the state space-estimated lagged permanent and transitory components of trading volume is estimated using the following model:

$$MT_{i,t} = \alpha + \beta_1\sigma^{R}_{i,t-1} + \beta_2 TV_{i,t-1} + \beta_3 BSI_{i,t-1} + \beta_4\sigma^{2s}_{i,t-1} + \beta_5\sigma^{2u}_{i,t-1} + \varepsilon_{i,t}$$

where $MT_{i,t}$ is the proxy for market toxicity for stock *i* and interval *t* and is calculated as the absolute value of the difference between the numbers of buy and sell trades divided by the sum of the numbers of buy and sell trades occurring during interval *t*. $\sigma^{R}_{i,t-1}$ is the standard deviation of mid-price returns for stock *i* during interval *t-1* and calculated as the standard deviation of midpoint-to-midpoint returns during interval *t-1*; each midpoint corresponds to a transaction. $TV_{i,t-1}$ is the natural logarithm of trading volume for stock *i* during interval *t-1* and $BSI_{i,t-1}$ is the absolute difference between buyer- and seller-initiated traders for stock *i* during interval *t-1*. $\sigma^{2s}_{i,t-1}$and $\sigma^{2u}_{i,t-1}$ are state space model-estimated proxies (estimated using Kalman filter constructed maximum likelihood) for informed and uninformed trading activity respectively for stock *i* and interval *t-1*. The sample contains the most active 100 S&P 500 stocks traded between October 1, 2016 and September 30, 2017 on NYSE and NASDAQ. ***, ** and * correspond to statistical significance at the 0.01, 0.05 and 0.10 levels, respectively.

Dependent Variable: $MT_{i,t}$		
	One-minute frequency	One-hour frequency
$Intercept$	0.609***	0.564***
	(408.71)	(111.41)
$\sigma_{i,t-1}^{R}$	1.612***	1.690***
	(59.48)	(42.84)
$TV_{i,t-1}$	0.248x10^{-3}***	0.296x10^{-2}
	(8.54)	(1.28)
$BSI_{i,t-1}$	0.187x10^{-6}***	0.219x10^{-6}***
	(58.15)	(54.29)
$\sigma_{i,t-1}^{2s}$	-0.535x10^{-2}***	-0.152x10^{-2}***
	(-5.72)	(-24.11)
$\sigma_{i,t-1}^{2u}$	-0.242x10^{-3}	-0.745x10^{-3}
	(-0.62)	(-0.63)
Sample size *(n)*	8880028	204354
Fixed effects	Time	Time
$\overline{R^2}$	0.95%	3.99%

Table 3.6 presents the estimated coefficients for the model estimated at one-minute and one-hour frequencies. Consistent with the results in Tables 3.4 and 3.5, $\sigma_{i,t-1}^{2s}$ is negatively and statistically significantly related to $MT_{i,t}$ at the 0.01 level of statistical significance, however $\sigma_{i,t-1}^{2u}$ is not, once volume and liquidity are controlled for. The inverse relationship between the $MT_{i,t}$ and $\sigma_{i,t-1}^{2s}$ suggests that information-motivated trading volume reduces order flow toxicity in financial markets, even after controlling for the overall impact of trading volume and volatility. This is in line with the arguments that informed trading, which is dependent on uninformed trading activity, enhances liquidity (see Kyle, 1981; 1984; 1985; 1989). Another explanation for the ameliorating effect of informed trading on market toxicity is presented by Admati and Pfleiderer (1988), who show that when informed traders observe the same information signal (a very plausible scenario), they compete against each other to exploit the signal. This competition may lead to the market maker facing reduced adverse selection risk. When faced with reduced adverse selection risk, market makers will respond with tighter spreads, implying a reduction in toxic order flow.

Although all other control variables are significant in the one-minute frequency model estimation, the explanatory power of the regression is small, with the $\overline{R^2}$ being only about 0.95%, again owing to the high frequency of the model estimation. This view is underscored by the larger $\overline{R^2}$ value for the one-hour frequency estimation at 3.99%.

3.4.2.3 Hypothesis III: state space model-estimated components of trading volume and short-horizon returns

As outlined in Section 3.2.3.3, my third hypothesis suggests that $\sigma_{i,t}^{2_s}$ is a significant predictor of short-horizon stock returns. In a test of this hypothesis, I estimate the following regression model:

$$R_{i,t} = \alpha + \beta_1\sigma_{i,t-1}^{p} + \beta_2 Espread_{i,t-1} + \beta_3 TV_{i,t-1} + \beta_4 BSI_{i,t-1} + \beta_5\sigma_{i,t-1}^{2_s} + \varepsilon_{i,t} \tag{3.8}$$

where all of the variables are as previously defined. All variables are computed over a one-second frequency.

Table 3. 7 Predictive regressions of short horizon stock returns on lagged transitory component of trading volume

The predictive power of the state space-estimated lagged transitory component of trading volume is estimated using the following model:

$$R_{i,t} = \alpha + \beta_1\sigma_{i,t-1}^{R} + \beta_2 Espread_{i,t-1} + \beta_3 TV_{i,t-1} + \beta_4 BSI_{i,t-1} + \beta_5\sigma_{i,t-1}^{2_s} + \beta_6 MT_{i,t-1} + \varepsilon_{i,t}$$

where $R_{i,t}$ is the midpoint-to-midpoint return for stock *i* during interval *t* and is computed as the difference between the midpoints corresponding to the last transactions at intervals *t* and *t-1* divided by the midpoint corresponding to the last transaction at interval *t-1*. $\sigma_{i,t-1}^{R}$ is the standard deviation of mid-price returns for stock *i* during interval *t-1* and calculated as the standard deviation of midpoint-to-midpoint returns during interval *t-1*; each midpoint corresponds to a transaction. $Espread_{i,t-1}$ is the effective spread for stock *i* and interval *t-1* and computed as twice the absolute value of the difference between the last execution price and the midpoint of the prevailing bid and ask prices for interval *t-1* $TV_{i,t-1}$ is the natural logarithm of trading volume for stock *i* during interval *t-1* and $BSI_{i,t-1}$ is the absolute difference between buyer- and seller-initiated traders for stock *i* during interval *t-1*. $MT_{i,t-1}$ is a proxy for market toxicity for stock *i* and interval *t-1* and calculated as the absolute value of the difference between the numbers of buy and sell trades divided by the sum of the numbers of buy and sell trades for interval *t-1* $\sigma_{i,t-1}^{2_s}$ is a state space model-estimated proxy (estimated using Kalman filter constructed maximum likelihood) for informed trading activity for stock *i* and interval *t*-1. The sample contains the most active 100 S&P 500 stocks traded between October 1, 2016 and September 30, 2017 on NYSE and NASDAQ. ***, ** and * correspond to statistical significance at the 0.01, 0.05 and 0.10 levels, respectively.

Dependent Variable: $R_{i,t}$

	One-second frequency	One-minute frequency
$Intercept$	-0.551x10^{-5***} (-24.19)	-0.650x10^{-4***} (-6.68)
$\sigma^{R}_{i,t-1}$	0.513x10^{-3**} (2.20)	-0.103x10^{-4} (-1.04)
$ESpread_{i,t-1}$	0.388x10^{-3***} (75.37)	0.833x10^{-3***} (42.25)
$TV_{i,t-1}$	0.147x10^{-6***} (6.88)	0.771x10^{-5***} (7.82)
$BSI_{i,t-1}$	0.499x10^{-9***} (53.14)	0.100x10^{-8***} (46.25)
$\sigma^{2s}_{i,t-1}$	-0.165x10^{-6***} (-24.58)	-0.303x10^{-4} (-1.11)
$MT_{i,t-1}$		0.301x10^{-5} (0.82)
Sample size *(n)*	29959938	8880028
Fixed effects	Time	Time
$\overline{R^2}$	0.17%	0.65%

Table 3.7 presents the estimated coefficients for Equation (3.8). All of the coefficients, except β_1 (for $\sigma^{R}_{i,t-1}$), are statistically significant at the 0.01 level. This result is a validation of my third hypothesis and thus further emphasizes the empirical relevance of my state space modelling approach. The statistically significant relationship between $\sigma^{2s}_{i,t-1}$ and one-second $R_{i,t}$ implies that $\sigma^{2s}_{i,t}$, as obtained using the state space model approach, signals private information similar to the order imbalance metrics used by Chordia et al. (2008). The $\sigma^{2s}_{i,t-1}$ coefficient estimate is negative, suggesting that an increase in the level of informed trading eliminates/reduces return predictability/arbitrage (see Hellwig, 1980; Wang, 1993). The $\overline{R^2}$ is 0.17%. The low $\overline{R^2}$ is linked to the estimation frequency of the regression model, which is one second in this case.

An estimation of the model over a lower frequency, such as the one-minute interval, could also prove insightful. This is because the trading volume in my sample appears to be mainly driven by HFTs, given the sample period and the market I focus on (see Brogaard et al., 2014b). Thus, if HFTs are responsible for driving a substantial proportion of the informed trading volume, the predictability of stock return should be greatly diminished over a one-minute interval, since a one-minute interval cannot be considered a short-horizon in an HFT-driven market. I estimate the following regression at a one-minute frequency; the only difference to Equation (3.8) is the addition of $MT_{i,t}$, which can only be validly computed at a minimum frequency of about one minute:

$$R_{i,t} = \alpha + \beta_1 \sigma^{p}_{i,t-1} + \beta_2 Espread_{i,t-1} + \beta_3 TV_{i,t-1} + \beta_4 BSI_{i,t-1} + \beta_5 \sigma^{2_s}_{i,t-1} + \beta_6 MT_{i,t-1} + \varepsilon_{i,t} \quad (3.9)$$

In this model, I expect that the coefficients for the two information signal proxies, i.e. $MT_{i,t-1}$ and $\sigma^{2_s}_{i,t-1}$, will not be statistically significant at the one-minute interval because of the superfast trading systems of HFTs trading in S&P 500 stocks.

The final column of Table 3.7 presents the estimated coefficients for Equation (3.9). As predicted, $\sigma^{2_s}_{i,t-1}$'s coefficient is not statistically significant, owing to the lack of return predictability over a time period stretching into a minute. However, the $\overline{R^2}$ coefficient at 0.65% is larger than for the one-second frequency estimation in Equation (3.8). The lack of statistical significance for $\sigma^{2_s}_{i,t-1}$'s coefficient in the one-minute frequency regression model is due to the prevalence of HFT activity in the data I use, and the ability of HFTs to absorb and act on new information at a fast pace and thereby eliminate arbitrage opportunities. This leads to the elimination of return predictability at less than ultra-high frequencies. $MT_{i,t-1}$ is an information signal based on the order imbalance metric used by Chordia et al. (2008); however, in contrast to the results presented by Chordia et al. (2008), the metric is not statistically significant here. This shows that while one-second stock returns are predictable from lagged metrics that signal private information, one-minute stock returns are not predictable in financial markets dominated by HFTs.

A key finding here is that although $\sigma_{i,t}^{2_s}$ is a lag predictor of one-second stock returns, one-minute stock returns are not predictable using either $\sigma_{i,t}^{2_s}$ or the order imbalance metric $MT_{i,t}$ inspired by Chordia et al. (2008). Thus, the latter part of the findings is not consistent with the results presented by Chordia et al. (2008), who show that even five-minute stock returns can be predicted from past order imbalance. The inconsistency here is linked to the data period employed by both studies. While Chordia et al. (2008) employ a dataset covering the years 1993 to 2002, when HFTs were not the main drivers of trading in financial markets, the analysis in this section is based on a much more recent dataset from 2016 to 2017. In Section 3.5, I show that, based on 2009 data, 71% of NASDAQ and NYSE's trading volume is linked to HFT activity. It is therefore not surprising to find that in recent years, the speed of price adjustment through the incorporation of new information has become much higher.

3.5 High frequency trading and return predictability[28]

In Section 3.4.2.3, I argue that the lack of a statistically significant relationship between $\sigma_{i,t}^{2_s}$ and one-minute $R_{i,t}$ is due to HFTs driving a faster incorporation of information into prices. In this section, I substantiate this theory by addressing the role of HFTs in the elimination of return predictability. In comparison with non-HFTs, HFTs could be viewed as being informed, simply on the basis that they trade with either private or public information (e.g. the sudden arrest of a firm's CEO for fraudulent activities) at a faster pace than non-HFTs. This is referred to as latency arbitrage; it involves the exploitation of a trading time disparity between fast and slow traders, when that trade is executed solely because of a latency advantage. Ibikunle (2018) argues that this speed advantage is tantamount to an information advantage when traders trade at different speeds, since the end result remains the same – a set of traders exploit information

[28] I am grateful to an anonymous referee for suggesting this analysis.

(whether private or public) ahead of a different set of traders. Thus, exchanges with infrastructures that especially accommodate HFTs tend to display efficient prices ahead of others when instruments are traded simultaneously across those exchanges. This is the case with the analysis of price leadership in the London equity market conducted by Ibikunle (2018). Chaboud et al. (2014) and Brogaard et al. (2014b) also show that HFTs enhance informational efficiency by speeding up price discovery and eliminating arbitrage opportunities.

In order to capture the transitory nature of informed trading volumes linked to HFT activity, I design a test which reflects the extent of transitory informed trading in the market when arbitrageurs observe that instruments' prices have deviated from their underlying values. I note that, while HFTs could be considered informed in comparison with non-HFTs, not all HFTs employ arbitrage strategies. Menkveld (2013) and Hagströmer and Nordén (2013) show that the majority of HFTs (about 80%) typically apply market making strategies. Furthermore, in a market dominated by HFTs, the speed advantage will not consistently confer appreciable advantages over the also fast competition. Thus, my test is designed to capture the changes in HFT volumes attributable to informed HFT activity.

For the test, I use the transactions dataset for 120 NASDAQ and NYSE stocks obtained from NASDAQ. The data disaggregates transactions into HFT and non-HFT transactions for the year 2009. Employing the dataset, I re-estimate Equations (3.8) and (3.9) with one additional variable, $D_{HFT,i,t-1} * \sigma^{2_s}_{i,t-1}$:

$$R_{i,t} = \alpha + \beta_1 \sigma^{p}_{i,t-1} + \beta_2 Illiq_{i,t-1} + \beta_3 TV_{i,t-1} + \beta_4 BSI_{i,t-1} + \beta_5 \sigma^{2_s}_{i,t-1} + \beta_6 D_{HFT,i,t-1} * \sigma^{2_s}_{i,t-1} + \varepsilon_{i,t} \quad (3.10)$$

$$R_{i,t} = \alpha + \beta_1 \sigma^{p}_{i,t-1} + \beta_2 Illiq_{i,t-1} + \beta_3 TV_{i,t-1} + \beta_4 BSI_{i,t-1} + \beta_5 \sigma^{2_s}_{i,t-1} + \beta_6 MT_{i,t-1} + \beta_7 D_{HFT,i,t-1} * \sigma^{2_s}_{i,t-1} + \varepsilon_{i,t} \quad (3.11)$$

$D_{HFT,i,t-1} * \sigma_{i,t-1}^{2_s}$ is obtained by interacting a new variable, $D_{HFT,i,t-1}$, with the lag transitory component variable, $\sigma_{i,t-1}^{2_s}$. $D_{HFT,i,t-1}$is a dummy variable equalling one for stock *i* for interval *t-1* during periods of high HFT activity. In order to determine the intervals of high HFT activity, I compute the proportion of HFT trades to non-HFT trades using the designations (HFT/non-HFT) for the transactions in the NASDAQ data. A one-second or one-minute interval is designated as an interval of high HFT activity if the proportion of HFT trades for that interval is one standard deviation higher than the mean for the surrounding -60, +60 corresponding intervals. Intervals correspond to one-second or one-minute. No other interval is considered because the existing literature (see as an example, Chordia et al., 2008) shows that short horizon predictability is eliminated within a few minutes. The NASDAQ dataset, as pointed out by Brogaard et al. (2014b), does not identify all HFTs. Hence, for robustness, I employ an alternative measure of HFT activity in my analysis; this is the widely deployed proxy based on the ratio of messages to the number of transactions (see as examples, Boehmer et al., 2015; Malceniece et al., 2018). $Illiq_{i,t-1}$ is a proxy for one period lag illiquidity and corresponds to one of either the Amihud (2002) illiquidity ratio or $Espread_{i,t}$. As in Equations (3.8) and (3.9), Equations (3.10) and (3.11) are estimated at one-second and one-minute frequencies respectively.

If $D_{HFT,i,t-1} * \sigma_{i,t-1}^{2_s}$'s coefficient is negative and statistically significant, it implies that a transitory rise in HFT activity is informed and reduces return predictability. This conclusion will be especially strengthened if $\sigma_{i,t-1}^{2_s}$ is not statistically significant in Equations (3.10) and (3.11), since it would imply that the reduction in return predictability is primarily driven by transitory HFT volume. A result of this nature would be in line with one of the assumptions underlying my state space modelling approach, i.e. informed trading volume is transitory and only arises to exploit deviations in the price of an instrument from its fundamental value.

Table 3. 8 Predictive regressions of short horizon stock returns on lagged components of trading volume interacted with a dummy variable for high frequency trading

The predictive power of the state space-estimated lagged transitory component of trading volume (interacted with a dummy variable for high frequency trading activity) is estimated using the following model:

$$R_{i,t} = \alpha + \beta_1 \sigma_{i,t-1}^R + \beta_2 Illiq_{i,t-1} + \beta_3 TV_{i,t-1} + \beta_4 BSI_{i,t-1} + \beta_5 \sigma_{i,t-1}^{2s} + \beta_6 MT_{i,t-1} + \beta_7 D_{HFT,i,t-1} * \sigma_{i,t-1}^{2s} + \varepsilon_{i,t}$$

where $R_{i,t}$ is the midpoint-to-midpoint return for stock *i* during interval *t* and is computed as the difference between the midpoints corresponding to the last transactions at intervals *t* and *t-1* divided by the midpoint corresponding to the last transaction at interval *t-1*. $\sigma_{i,t-1}^R$ is the standard deviation of mid-price returns for stock *i* during interval *t-1* and calculated as the standard deviation of midpoint-to-midpoint returns during interval *t-1*; each midpoint corresponds to a transaction. $Illiq_{i,t-1}$ is a proxy for one period lag illiquidity and corresponds to one of the Amihud (2002) illiquidity ratio ($Amihud_{i,t-1}$) or $Espread_{i,t}$. $Amihud_{i,t-1}$ is computed as absolute return divided by trading volume for stock *i* during interval *t-1*. $Espread_{i,t-1}$ is the effective spread for stock *i* and interval *t-1* and computed as twice the absolute value of the difference between the last execution price and the midpoint of the prevailing bid and ask prices for interval *t-1* $TV_{i,t-1}$ is the natural logarithm of trading volume for stock *i* during interval *t-1* and $BSI_{i,t-1}$ is the absolute difference between buyer- and seller-initiated traders for stock *i* during interval *t-1*. $MT_{i,t-1}$ is a proxy for market toxicity for stock *i* and interval *t-1* and calculated as the absolute value of the difference between the numbers of buy and sell trades divided by the sum of the numbers of buy and sell trades for interval *t-1* $\sigma_{i,t-1}^{2s}$ is a state space model-estimated proxy (estimated using Kalman filter constructed maximum likelihood) for informed trading activity for stock *i* and interval *t*-1. $D_{HFT,i,t-1}$ is a dummy variable equalling one during periods of high HFT activity for stock *i* and interval *t-1*. A one-second or one-minute interval is designated as an interval of high HFT activity if HFT trades for that interval is one standard deviation higher than the mean for the surrounding -60, +60 corresponding intervals. The sample contains the most active 100 S&P 500 stocks traded between October 1, 2016 and September 30, 2017 on NYSE and NASDAQ. ***, ** and * correspond to statistical significance at the 0.01, 0.05 and 0.10 levels, respectively.

Panel A

	Dependent Variable: $R_{i,t}$	
	One-second frequency	One-minute frequency
$Intercept$	-0.301x10^{-3}***	0.445x10^{-2}**
	(-6.18)	(2.15)
$\sigma_{i,t-1}^R$	0.635***	0.061x10^{-1}
	(8.79)	(1.50)
$Amihud_{i,t-1}$	-2.998***	-1.554
	(-4.10)	(-0.37)
$TV_{i,t-1}$	0.193x10^{-4}***	0.299x10^{-3}**
	(4.05)	(2.49)
$BSI_{i,t-1}$	0.01x10^{-9}*	0.01x10^{-6}*
	(1.57)	(1.79)
$\sigma_{i,t-1}^{2s}$	-0.103x10^{-6}	0.493x10^{-4}
	(-1.44)	(1.33)
$D_{HFT,i,t-1} * \sigma_{it-1}^{2s}$	-0.371x10^{-4}***	-0.168x10^{-3}
	(-3.16)	(-1.28)
$MT_{i,t-1}$		-0.165x10^{-2}
		(-1.33)
Sample size *(n)*	8291971	2069787
Fixed effects	Time	Time
$\overline{R^2}$	0.45%	0.91%

Panel B

	Dependent Variable: $R_{i,t}$	
	One-second frequency	One-minute frequency
$Intercept$	$-0.211\text{x}10^{-5***}$ (-3.17)	$-0.163\text{x}10^{-4*}$ (-1.84)
$\sigma^R_{i,t-1}$	$0.283\text{x}10^{-2**}$ (2.60)	$-0.697\text{x}10^{-4}$ (-1.32)
$ESpread_{i,t-1}$	$0.288\text{x}10^{-3***}$ (17.74)	$0.486\text{x}10^{-3***}$ (3.15)
$TV_{i,t-1}$	$0.204\text{x}10^{-6**}$ (2.83)	$0.313\text{x}10^{-5***}$ (11.06)
$BSI_{i,t-1}$	$0.236\text{x}10^{-10}$ (0.85)	$0.482\text{x}10^{-8***}$ (13.47)
$\sigma^{2s}_{i,t-1}$	$-0.266\text{x}10^{-8}$ (-0.03)	$-0.270\text{x}10^{-7}$ (-1.14)
$D_{HFT,i,t-1} * \sigma^{2s}_{i,t-1}$	$-0.256\text{x}10^{-5***}$ (-3.49)	$-0.142\text{x}10^{-10}$ (-1.38)
$MT_{i,t-1}$		$0.137\text{x}10^{-5}$ (0.54)
Sample size *(n)*	29959938	8880028
Fixed effects	Time	Time
$\overline{R^2}$	0.21%	0.77%

I present the results based on the two approaches to computing D_{HFT} in Table 3.8; Panel A shows the results using the NASDAQ-defined HFT/non-HFT transactions, while Panel B shows the results using the ratio of messages to transactions HFT proxy. $Illiq_{i,t}$ corresponds to the Amihud (2002) illiquidity ratio and $Espread_{i,t}$ in Panels A and B respectively. Contrary to the results in Table 3.7, although it remains negative, $\sigma^{2s}_{i,t-1}$'s coefficients for the one-second frequency estimation in both panels are not statistically significant. However, when the transitory component variable is interacted with $D_{HFT,i,t-1}$, it becomes highly statistically significant, while retaining its negative sign. This implies that the reduction in the return predictably observed in the earlier analysis is driven by informed HFT activity. Consistent with the assumption underlying my state space modelling approach, the transitory component of trading volume, i.e. an increase in HFT volume above the mean, aids the speedy incorporation of information into instruments' prices and leads to the elimination of arbitrage opportunities. With this analysis, I ascertain that the transitory trading volume component relevant to eliminating return predictability in today's financial markets is the HFT kind.

3.6 Conclusion

In this chapter, I develop a state space model for decomposing trading volume into liquidity-driven (permanent) and information-driven (transitory) components. I argue that the permanent component of trading volume is driven by liquidity-seeking order flow, while the transitory component is driven by information-motivated order flow. In addition to providing a robust set of arguments grounded in the literature to support my theses, I further develop a set of multivariate regression models to formally test them. Firstly, I find that the transitory component of trading volume obtained from my state space model has a statistically significantly relationship with volatility, liquidity and market toxicity, even after controlling for volume. There is no such relationship observed for the permanent component once volume is controlled for. These results are consistent with an extensive stream of theoretical and empirical studies on the relationship of the informed and liquidity trading activity with volatility, liquidity and market toxicity. The consistency therefore implies that the permanent and transitory components, estimated using my state space modelling approach, can be viewed as encapsulating the liquidity- and information-motivated trades, respectively.

I also demonstrate that the transitory component is a significant predictor of short-horizon returns. This underscores the argument that the transitory component is a proxy for private information. However, in contrast to Chordia et al. (2008), I find that one-minute returns cannot be predicted using either the state space-estimated transitory component or the minute(s)-long order imbalance metrics employed by Chordia et al. (2008). This implies that in today's high frequency trading environment, arbitrage opportunities are eliminated at a much faster rate than in the early 2000s period examined by earlier studies. I show that this sharp decline in the window for return predictability is driven by HFT activity.

4. Need for Speed? International transmission latency, liquidity and volatility

"The rise of high-frequency traders has opened up a debate among investors, brokers and exchanges. Critics have long claimed that speed-driven traders unfairly hurt traditional investors… Supporters argue that faster traders are now a vital element of modern markets…"

Financial Times, 15th May 2019

4.1 Introduction

The speed of trading and, ultimately, of price adjustment, is an important factor in the price discovery process in financial markets. That factor, today, holds a significance that transcends market quality implications. It is the driving force behind a recent upsurge of latency arbitrage in modern financial markets, as markets become increasingly dominated by ultra-high-frequency algorithmic traders. However, speed may also be good for markets. The evidence of this has thus far been inconsistent. Some studies find that speed is good for liquidity and price discovery (see as examples Brogaard et al., 2014b; Hendershott et al., 2011; Hoffmann, 2014), while others suggest a positive relationship between speed and adverse selection cost (see as examples Biais et al., 2015; Foucault et al., 2016; Foucault et al., 2017; Hendershott and Moulton, 2011), implying a negative effect on market quality and liquidity in particular. Jovanovic and Menkveld (2016) show that better informed high-frequency traders (HFTs) can reduce welfare, and Kirilenko et al. (2017) argue that although HFTs did not trigger the flash crash, they nevertheless exacerbated it by demanding immediacy.

While the existing literature focuses on traders' execution speed in their examination of the role of speed on market quality, I focus on a new variable capturing the combination of microwave/fiber optic connection latency, traders' information execution time, and exchange latency. I call this variable of interest *Transmission Latency (TL)*. The distinction I make here is important since speed between different exchanges is not *only* dependent on the heterogeneous technological capacity of traders, but also depends on the connection latency between financial markets and exchange latencies of different financial markets. This implies that *TL* holds economic significance for market quality beyond what the factors linked to trader execution speed hold. Furthermore, modern financial markets are characterized by high fragmentation. This underscores how critically inter-venue speeds must be incorporated into any examination of market quality implications of speed. The economic insights this consideration could generate are likely substantial (see also Menkveld and Zoican, 2017). In addition, recent arguments by regulators and investors suggest that while higher information transmission speed offered by HFTs improves liquidity (and by extension, market quality), it nevertheless contributes to higher volatility and market risk, and hence impairs market quality.[29] Motivated by these contrasting arguments and the incomplete picture drawn by the existing literature, I investigate the effects of speed on the quality of financial markets by applying the measure of latency, *TL*.

The focus of my study is therefore closely related to the works of Shkilko and Sokolov (2016), Menkveld and Zoican (2017), and Baron et al. (2018). Shkilko and Sokolov (2016) examine liquidity when there are speed differentials among traders, and find that these differentials impair liquidity and volatility in financial markets. It is important to note that Shkilko and Sokolov's (2016) focus on the 2011-2012 period, during which microwave networks are only accessible to a small group of sophisticated trading firms. By contrast, I use more recent data allowing me to capture the latest changes in microwave technology, which has recently lost much of its exclusivity and is now available for a nominal fee. It implies that my study offers a clearer picture of the effects of speed on market quality in financial markets.

[29] https://www.reuters.com/article/us-highfrequency-microwave/lasers-microwave-deployed-in-high-speed-trading-arms-race-idUSBRE9400L920130501

Specifically, I directly estimate transmission latency between trading venues from transaction-level data and empirically examine the impact of estimated latency on liquidity and volatility.

Similar to my study, Baron et al. (2018) construct measures of latency from transaction-level data, and examine the performance and competition among HFTs. There are two important differences between my study and Baron et al.'s (2018). First, Baron et al. (2018) do not estimate transmission latency *between* financial markets, which is particularly important in today's highly fragmented markets. Specifically, Baron et al. (2018) estimate what they call *Decision Latency*, which is the difference between timestamps from a passive trade to a subsequent aggressive trade by the same firm, in the same security and at the same exchange. Secondly, and more importantly, their study analyzes the impact of latency on HFTs' trading performance, not liquidity and volatility, in financial markets. Menkveld and Zoican (2017) model the HFT arms race by adding the impact of *exchange speed* to Budish et al.'s (2015) model, and find that indeed, there is a nontrivial relationship between exchange speed and liquidity. It is important to note that in Menkveld and Zoican's (2017) model, exchange latency does not include the trader's execution latency, and thus is assumed constant for all traders. Their model identifies two channels through which exchange speed affects the bid-ask spread. Firstly, as a result of improvements in exchange latency, high-frequency market makers (HFMs) can quickly update their quotes and reduce their adverse selection risk, which implies that HFMs narrow the competitive spread, since speed allows them to face reduced adverse selection risk. Secondly, high-frequency speculators may still prevail in an arms race; therefore, HFMs would need to set a wider spread in order to compensate for higher adverse selection risk.

My study differs from Menkveld and Zoican (2017) in at least two aspects. Firstly, their study is a theoretical contribution. Secondly, while Menkveld and Zoican (2017) focus on the role of exchange latency in financial markets, my main variable of interest, *TL*, captures the *combined* effect of trader execution latency, exchange latency, and connection latency between exchanges.

My empirical approach involves first estimating the *TL* between the home exchange in Frankfurt (Xetra Stock Exchange – XSE) and a satellite exchange in London (Cboe Stock Exchange – CBOE), where XSE-listed stocks are cross-listed, and then examining its effect on liquidity and volatility of cross listed stocks in the satellite market. I thereafter investigate the channels, as informed by various theoretical models, through which my latency measure impacts market quality metrics.

My findings suggest that 49% (80%) of price-changing trades on CBOE occur within 3 (5) milliseconds (ms) of similar and proportional price-changing trade on XSE. This means that the existing microwave and fiber optic connections affect price responses on CBOE within 3-5ms of price changes on XSE. These estimates are consistent with the anecdotal evidence provided by industry practitioners active in both markets, since the latency (3-5ms) includes the traders' execution latencies, exchange latencies in CBOE and XSE, and connection latency between XSE and CBOE. For example, Perseus, one of the microwave connection providers between London and Frankfurt, states that a round trip latency via microwave and fiber optics between London and Frankfurt is 4.6ms and 8.4ms, respectively (see Footnote 29). The significance of these estimates is that analysis shows that higher *TL* leads to lower liquidity and volatility (i.e. speed enhances liquidity and increases volatility). The results are robust to alternative proxies for liquidity and volatility and more importantly, the magnitudes of these effects are economically meaningful. In order to address potential endogeneity concerns, I present causal evidence from a quasi-experimental setting, studying the impact of two technological upgrades by XSE on liquidity and volatility in CBOE. I compare the liquidity and volatility of stocks that are impacted by these updates with those that are not and show that, consistent with the previous results, increases in speed lead to enhanced liquidity and higher volatility.

The positive effect of speed on liquidity is linked to fast traders using their speed advantage to avoid adverse selection risk and thereby increasing liquidity. Another channel through which speed impacts market quality metrics is explained by the prediction of Roşu (2016), specifically that speed increases the aggressiveness of traders and this aggressiveness then leads to higher price volatility (see also Collin-Dufresne and Fos, 2016). Thus, it appears that while speed enhances market quality by enhancing liquidity, it impairs it by intensifying market volatility. This implies a trade-off between the benefits of speeds (liquidity improvements) and its unwanted effects (increased volatility). I therefore examine the net economic implication of latency on market quality, with liquidity and volatility as market quality characteristics. The analysis shows that while high latency can improve market quality by reducing volatility, its liquidity deterioration effect dominates its volatility reducing effect. This implies that the net effect of increasing (reducing) latency (speed) is an impairment of market quality.

My contributions to the existing literature are as follows. Firstly, my study is the first to empirically estimate *TL* between the two biggest European financial centers, Frankfurt and London, and by so doing corroborates the information provided on connection speed by the microwave and fiber optic connection providers (such as McKay Brothers). This exercise is particularly important in Europe, where financial markets have become increasingly fragmented across dominant national exchanges and a dominant London-based pan-European trading venue, CBOE. Secondly, I provide causal evidence on the direct impact of speed on market quality variables, such as volatility, which is unclear in the current literature. Thirdly, I complement the existing empirical literature that examines the relationship between trader and exchange speed on financial markets, by analyzing the combined role of traders' execution latency, exchange latency, and connection latency (microwave or fiber connections) between exchanges on liquidity and volatility of financial markets. The approach I take is more realistic and the measure I use is more relevant when measuring the impact of speed on market quality in a fragmented trading space – the reality of trading in modern financial markets. Finally, using a framework that controls for the undesirable (increased volatility) and desirable (enhanced liquidity) effects of speed, I show that the dominant effect of increased speed of trading is positive for market quality.

4.2 Institutional and technical backgrounds

4.2.1 Transmission latency between financial markets

In today's trading environment, information transmission speeds between trading venues play an important role in facilitating price discovery in an increasingly fragmented market place. A decade ago, the most common way to transmit information from Frankfurt to London was via a fiber optic cable; at this time fiber optics offered information transmission latencies of about 4.2ms.[30] Although fiber optic technology offers fast transmission, it is not the fastest. This is simply because with fiber optic technology, "information" (photons) travels through cables and it is difficult to place cables in a straight line between trading venues. For example, Shkilko and Sokolov (2016) show that until 2010 the fiber optic cabling between Chicago and New York exceeded the straight line distance between the two cities by about 200 miles. In contrast to fiber optic technology, with microwave technology, "information" (microwaves) travels through air. Hence, microwave networks offer information transmission speeds that are between 30 and 50% faster than with fiber optic technology. For example, microwaves shave about 1.9ms off the information transmission latency between Frankfurt and London when

[30] https://www.reuters.com/article/us-highfrequency-microwave/lasers-microwave-deployed-in-high-speed-trading-arms-race-idUSBRE9400L920130501

compared to fiber optics, a reduction from 4.2ms to 2.3ms.[31] It is therefore not surprising that the past decade has seen an emergence of the operation of microwave networks between major financial trading locations, such as London and Frankfurt.[32] Some of these networks are operated by specialist network providers (e.g., McKay Brothers), while others are operated directly by HFTs (e.g., Jump Trading).

31 https://www.quincy-data.com/product-page/#latencies

32 https://www.bloomberg.com/news/articles/2014-07-15/wall-street-grabs-nato-towers-in-traders-speed-of-light-quest

Figure 4. 1 A map of microwave networks connecting the British Isles to continental Europe

Microwave networks between the UK and continental Europe as mapped out by Laumonier (2016). The providers of the microwave networks are also indicated.

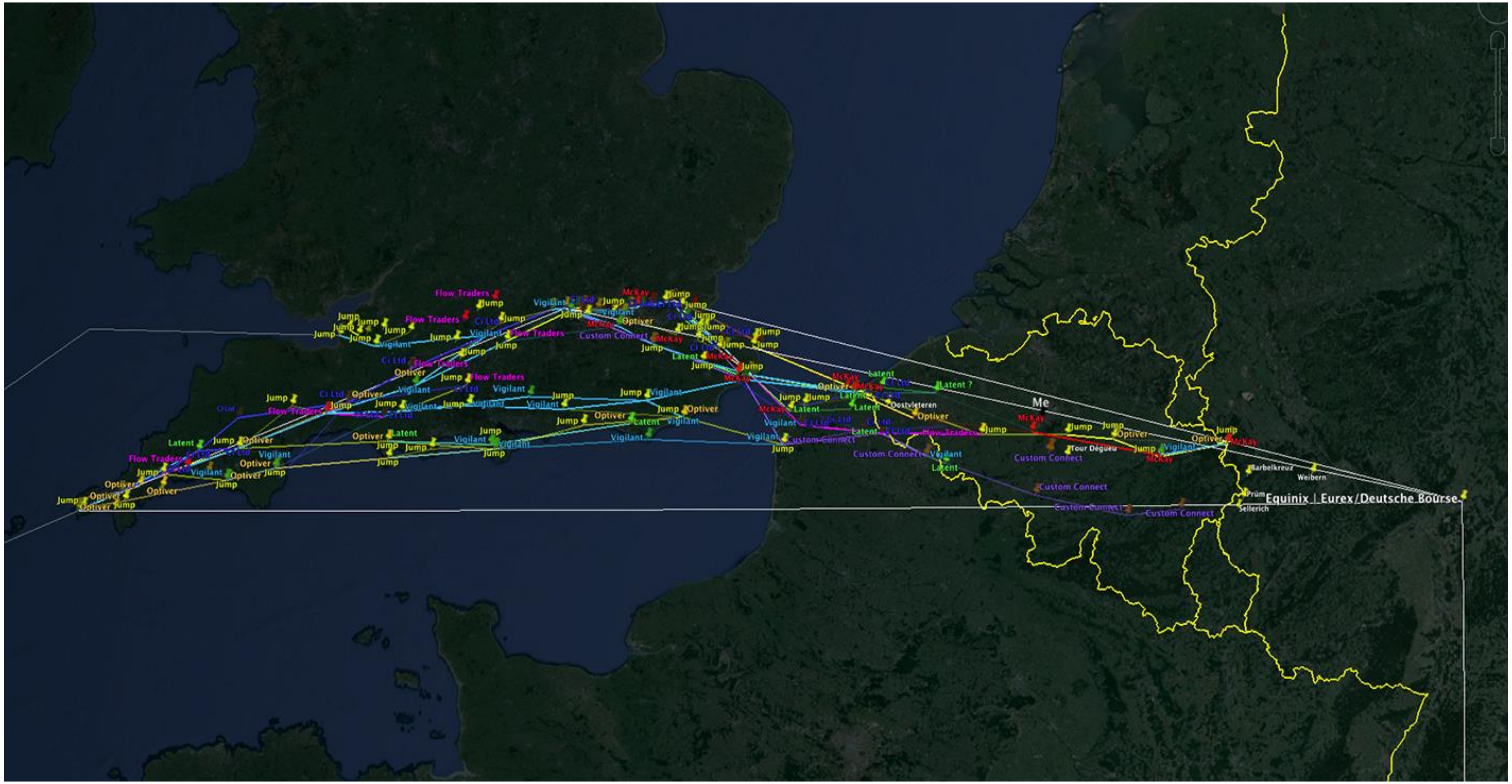


Figure 4.1 shows the microwave networks between the UK and Germany, and their respective providers (see Laumonier, 2016). Given the notable speed advantage of microwave networks, HFTs are ready to pay significant amounts of money to obtain several microseconds of speed advantage over their competitors.[33]

In this study, I estimate the information transmission latency between XSE and CBOE by using transaction-level data. My *TL* estimate is therefore composed of the following elements: (i) the connection latency between XSE and CBOE, (ii) the exchange latencies for XSE and CBOE, and (iii) the traders' execution latencies. Explicitly, the connection latency is the time it takes for information to travel via microwave/fiber optic connections between XSE and CBOE. The exchange latencies consist of the time it takes for the exchanges to process incoming and outgoing instructions. According to Menkveld and Zoican (2017), the exchange latency is the sum of gateway-processing latency and gateway-to-matching-engine latency. Gateway-processing latency equals the time spent inside the gateway application, and gateway-to-matching-engine latency is the time between an order's departure from the gateway and when the matcher begins processing the order. Finally, the transaction-level data from TRTH that I employ provides exact exchange timestamps for *executed* transactions. It thus also takes into account the time needed to execute transactions, which includes the traders' execution latencies, i.e. their signal processing and reaction times.

4.2.2 Technological upgrades on XSE

In order to address potential endogeneity concerns, I study the impact of two technological upgrades implemented by XSE on liquidity and volatility at CBOE. These technological upgrades are (1) the "New T7 Trading Technology" upgrade first offered on July

[33] https://www.businessinsider.com/locals-angry-at-flash-boy-traders-want-to-build-a-tower-taller-than-the-shard-2017-1?r=US&IR=T

3, 2017, and (2) the "Introduction of PS gateways" upgrade first offered on April 9, 2018.[34] The Deutsche Börse T7 Trading Technology system reduces order processing time significantly and should be captured by my *TL* measure. The PS (Partition Specific) gateways upgrade for all cash market instruments operates in parallel to the existing HF gateways. Usually, latency jitters on parallel inbound paths encourage multiplicity to reduce latency. However, this leads to greater system load and choking at busy times, and thus less predictable latencies may arise. The PS gateways upgrade introduces a single low-latency point of entry, which addresses this issue and consequently reduces exchange latency at XSE. This reduction should also be captured by *TL*. Since the two technological upgrades are introduced to reduce exchange latency at XSE, they could be employed as exogenous shocks in my quasi-natural experiment to examine the relationship between latency and market quality characteristics.

4.3 Data and latency estimation

My data source is the TRTH v2 (Datascope). The most important feature of the Datascope-sourced datasets that makes them highly suitable for my analysis is that they provide exact *exchange timestamps* in milliseconds for exchange-traded transactions and order flow. The main dataset employed in this study consists of ultra-high-frequency tick-by-tick data for the most active 100 German stocks that trade both on XSE in Frankfurt (home market) and on CBOE in London (satellite market). The dataset includes transaction-level data for trading days between March 2017 and August 2018. I select this period for two reasons. Firstly, Datascope does not provide exchange timestamps for European markets before June 2015. Secondly, as noted, to address potential endogeneity concerns, I employ a quasi-natural experiment approach using the two technological updates described above. The upgrade dates are July 3,

---

[34] The details of the upgrades can be found at https://www.xetra.com/dbcm-en/newsroom/press-releases/New-T7-trading-technology-goes-live-on-Xetra-144756 and https://www.xetra.com/resource/blob/228942/0bbe6323aa5436a88648d298d9b41512/data/143_17e.pdf

2017 and April 9, 2018. I then select a data coverage period spanning four months before and after the upgrades for the difference-in-difference (DiD) framework. The Datascope data contain standard transaction-level variables such as date, time (both TRTH and exchange timestamps), price, volume, bid price, ask price, bid volume, and ask volume.

From the raw data I determine the prevailing best bid and ask quotes for each transaction, enabling me to see the status of the order book at the time of each transaction. I divide the sample of 100 stocks into quartiles using their level of trading activity; trading activity is measured by euro trading volume.

4.3.1 Trading summary statistics

Table 4.1 reports trading activity statistics for XSE and CBOE.

Table 4. 1 Transactions' summary statistics and statistical tests

Panels A and B respectively present trading summary statistics for XSE and CBOE. Panel C reports the statistical tests of the trading summary differences between the XSE and CBOE. The statistical tests conducted are two-sample t-tests and pairwise Wilcoxon-Mann-Whitney tests. The sample consists of the 100 most active German stocks cross-listed on the XSE and CBOE. The sample period covers March 2017 to August 2018. Stocks are classified into quartiles using Euro trading volume.

Panel A

Trading activity: XSE				
	Average trading volume per stock (€'000,000)	Average trading volume per stock (000,000s)	Average transactions per stock (000s)	Average trade size per Stock (€'000)
Full sample	16,263.46	428.56	984.02	14.94
Least active	2,388.44	74.33	335.89	7.31
Quartile 2	4,717.94	145.04	557.78	10.92
Quartile 3	10,556.57	213.05	933.38	14.03
Most active	46,835.87	1,267.65	2,083.09	27.19

Panel B

Trading activity: CBOE				
Full sample	2,739.96	64.09	356.29	6.87
Least active	312.36	10.81	80.25	3.92
Quartile 2	667.55	18.67	165.23	5.72
Quartile 3	1,539.50	31.12	320.37	6.91
Most active	8,440.41	195.75	859.32	10.92

Panel C

		Trading activity (Full sample)		
XSE – CBOE	13,523.5***	364.47***	627.73***	8.07***
t-test p-value	< 0.001	< 0.001	< 0.001	< 0.001
W-M-W test p-value	< 0.001	< 0.001	< 0.001	< 0.001

Panels A and B of Table 4.1 present market activity statistics for XSE and CBOE respectively, and Panel C presents the difference in full-sample trading activity between the two stock exchanges along with p-values obtained using different statistical approaches (two-sample t-tests and Wilcoxon-Mann-Whitney tests). The p-values are reported for the null that there is no difference in trading activity between XSE and CBOE. Going by the number of transactions and nominal and euro-denominated trading volume, XSE appears to be more active than CBOE for the selected sample of stocks. This is expected since XSE is the home market for my selected sample of German stocks.

4.3.2 Price discovery

My latency (*TL*) estimation method assumes that information is transmitted from Frankfurt to London; an assumption supported by prior research (see as an example Grammig et al., 2005). Indeed, it is implausible to assume that the preponderance of firm-specific information about German companies originates from outside of Germany. The expectation that information for German stocks largely flows from Germany is also supported by the superior volume of transactions recorded for XSE compared to CBOE. Nevertheless, it is important to ascertain that XSE holds price leadership relative to CBOE for my sample of stocks, especially since the European markets have become increasingly fragmented over the past decade. This fragmentation has in some cases upended the natural expectation that superior trading activity confers higher levels of price discovery. For example, Ibikunle (2018) investigates price leadership for a sample of London Stock Exchange (LSE)-listed stocks cross-listed on CBOE, and finds that although LSE holds superior trading activity for the stocks, CBOE leads price discovery in those stocks for much of the trading day.

Table 4. 2 Price discovery analysis

This table presents the results for three different price discovery metrics estimating the share of price discovery for XSE and CBOE. *IS* is the information share metric as developed by Hasbrouck (1995), *CS* is the component share metric based on Gonzalo and Granger (1995), and *ILS* is the information leadership share as defined by Putniņš (2013). All estimates are computed based on price samples at the one-second frequency. The sample consists of the 100 most active German stocks cross-listed on XSE and CBOE. The sample period covers March 2017 to August 2018. Stocks are classified into quartiles using Euro trading volume.

	IS	CS	ILS
Full sample	0.69	0.64	0.61
Least active	0.63	0.60	0.56
Quartile 2	0.61	0.58	0.56
Quartile 3	0.68	0.64	0.58
Most active	0.76	0.71	0.61

Table 4.2 presents the results of the price leadership analysis between XSE and CBOE. For robustness, I employ three measures of price discovery computed using data price data sampled at the one-second frequency. The first and second measures are the information share metric (IS) developed by Hasbrouck (1995), and the component share metric (CS) developed by Gonzalo and Granger (1995).[35] These methods are based on the vector error correction model (VECM), and usually provide similar results if the VECM residuals are not correlated. However, as suggested by Yan and Zivot (2010), both metrics suffer from bias if noise levels differ across trading venues. Therefore, I employ the information leadership share metric (ILS) prescribed by Putniņš (2013), which corrects for the differential treatment of noise by the IS and CS measures and provides a cleaner measure of information leadership. The results are consistent with earlier studies, in that price discovery occurs mainly on XSE for German stocks; IS, CS and ILS estimates are 0.69, 0.64 and 0.61 respectively for the full sample of stocks. This

[35] I would like to acknowledge that the computation of the information follows the SAS codes that can be obtained from Joel Hasbrouck's website:
http://pages.stern.nyu.edu/~jhasbrou/EMM%20Book/SAS%20Programs%20and%20Data/Description.html

result implies that the majority of information is incorporated on XSE first. Therefore, my assumption regarding the information transmission direction appears valid. Table 4.2 further reports that the information share of XSE is typically highest for the most active stocks. This result is consistent with the empirical findings of Brogaard et al. (2014b), and suggests that HFTs are more active in the most active stocks.

4.3.3 Latency measurement

In general, latency can be considered as the delay between a signal and a response (see Baron et al., 2018). Following Laughlin et al. (2014), I define the signal as a price-changing trade in the home market, and the response as a near-coincident same direction price-changing trade in the satellite market. Laughlin et al. (2014) validly employ this method for futures-ETF pairs in the US financial markets, and I apply it to measure latency in the case of the 100 most active cross-listed German stocks between XSE and CBOE. According to the law of one price, the price of the cross-listed stocks should be the same regardless of location. Specifically, the difference between cross-listed security prices in different exchanges should simultaneously be eliminated in a no-arbitrage scenario and if markets are informationally efficient.[36]

The latency measurement approach involves first identifying the exact exchange timestamp for each price-changing trade on XSE. I then look for a near-coincident same direction price-changing trade on CBOE. In order to identify the near-coincident trade in CBOE I examine trades occurring within 10ms of each price-changing trade on XSE. I select the 10ms interval since the average information transmission latencies between Frankfurt and London are 2.3ms and 4.2ms for microwave and fiber optic connections, respectively.[37]

[36] One may argue that no-arbitrage limits and liquidity and trading cost can prevent market participants perfectly arbitraging price differences away. However, this argument cannot cause any serious concerns in my framework for two reasons. Firstly, I are using well-traded stocks in a major economy and secondly, on average, overwhelmingly, I would expect to see changes replicated across both platforms.

[37] https://www.reuters.com/article/us-highfrequency-microwave/lasers-microwave-deployed-in-high-speed-trading-arms-race-idUSBRE9400L920130501

Table 4. 3 Information transmission latency between XSE and CBOE

This table presents different statistics for the information transmission latency between XSE and CBOE. Panel A reports the number of responses on CBOE to price-changing trades on XSE for different time bins in milliseconds (ms) for the quartiles and full sample of stocks; stocks are classified into quartiles using Euro trading volume. Panel B presents the mean and standard deviation of the information transmission latency between XSE and CBOE for each quartile and the full sample of stocks. Panel C shows the average information transmission latencies for 21 trading days before and after a technological upgrade on July 3, 2017. The statistical tests conducted are two-sample t-tests and pairwise Wilcoxon-Mann-Whitney tests. The sample consists of the 100 most active German stocks cross-listed on XSE and CBOE. The sample period covers March 2017 to August 2018.

Panel A

Speed (ms)	Full sample		Least active		Quartile 2		Quartile 3		Most active	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
3	936,646	48.61	63,563	49.05	108,325	46.50	187,528	44.76	577,230	50.39
4	286,962	14.89	19,041	14.69	36,303	15.58	63,498	15.16	168,120	14.68
5	332,286	17.24	21,742	16.78	41,457	17.79	75,439	18.01	193,648	16.91
6	100,435	5.21	6,496	5.01	11,959	5.13	23,531	5.62	58,449	5.10
7	81,733	4.24	5,933	4.58	10,862	4.66	20,686	4.94	44,252	3.86
8	75,895	3.94	5,281	4.08	9,976	4.28	19,924	4.76	40,714	3.55
9	62,679	3.25	4,106	3.17	7,700	3.31	15,834	3.78	35,039	3.06
10	50,364	2.61	3,415	2.64	6,389	2.74	12,517	2.99	28,043	2.45

Panel B

Full sample		Quartile 1 (least active)		Quartile 2		Quartile 3		Quartile 4 (most active)	
Mean (ms)	St. Dev	Mean (ms)	St. Dev	Mean (ms)	St. Dev	Mean (ms)	St. Dev	Mean (ms)	St. Dev
4.39	1.86	4.39	1.87	4.45	1.88	4.55	1.94	4.32	1.83

Panel C

Period	Average latency for the full sample
Before upgrade	4.40
After upgrade	4.30

Difference	0.10***
t-test p value	< 0.001
W-M-W test p value	< 0.001

Panel A in Table 4.3 reports the number of responses on CBOE to the signals on XSE for various latencies. I exclude the responses that fall in the 2ms interval. This is because the 2ms interval is less than the theoretical limit of 2ms it should take light to travel in a vacuum between the two locations. The number of responses in this interval account for only 2% of all responses, hence the exclusion should not have any material impact on my analysis. Laughlin et al. (2014) argue that the responses at less than the speed-of-light can be considered as a proof of the predictive capacity of HFTs. I do not examine this argument since it is outside of the scope of this study.

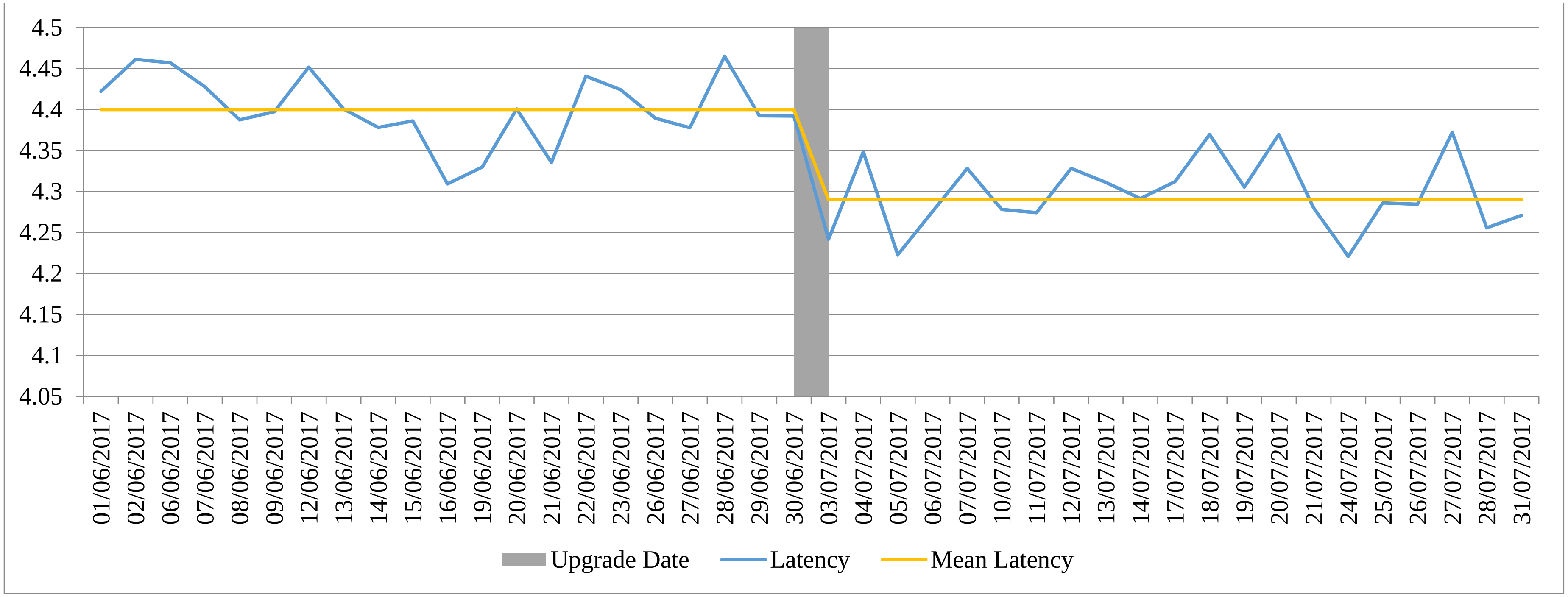
There are two important findings in Panel A. First, it shows that 48.61% (80.74%) of all responses (after excluding the [0 – 2ms] interval) fall within the 3ms (5ms) bin. These latencies are consistent with those provided by the microwave network and fiber optic connection providers, and corroborate the view that my latency measure indeed captures the transmission latency between the two trading venues. For example, McKay Brothers recently announced that their average microwave latency between the XSE (FR2) and CBOE (LD4) data centres is 2.3ms. Furthermore, it is generally acknowledged that the average latency via fiber optic connections is about 4.2ms (see Footnote 29). These announced latencies, 2.3ms and 4.2ms, are only transmission latencies between exchanges and do not take into account the exchange latencies and the traders' order execution latencies. Therefore, I expect the actual trading latencies to be closer to my estimated transmission latencies. Panel A's estimates suggest that traders are more likely to employ the faster microwave technology than fiber optic options for connecting Frankfurt and London. Secondly, on average, the most active stocks have quicker response times, with 50.39% (81.98%) of all responses falling in the 3ms (5ms) bin. This is unsurprising given that existing studies suggest that HFTs trade more in the most active stocks (see Brogaard et al., 2014b). Panel B in Table 4.3 presents the mean and standard deviation of latencies for the full sample and each quartile. The average latency for the full sample is 4.39ms and, consistent with Panel A in Table 4.3, the most active stocks have the lowest transaction latency.

The empirical relevance of my latency estimation is underscored by the literature (Baron et al., 2018; Laughlin et al., 2014), but I also directly test its precision by examining the latency evolution around the technology upgrade events. A downward adjustment of the latencies on the event dates would provide support to the accuracy of my estimation. Figure 4.2 illustrates the impact of the "New T7 Trading Technology" upgrade on my estimated latency variable, *TL*. The figure shows a sharp decrease in latency on the day of the upgrade, with the average latency falling by 0.105ms to 4.297ms – a reduction of 2.4%. In addition, Panel C in Table 4.3 tests the statistical significance of the difference between the latencies 21 trading days before and after the implementation of the upgrade. The estimates show that the average latency reduction is statistically significant.[38]

[38] Although not explicitly reported, the picture is comparable for the second technological upgrade. The "Introduction of PS gateways" leads to a significant latency reduction of 1.6%. The results are available on request.

Figure 4. 2 Information transmission latency over time

This figure plots the information transmission latency from June 2017 to July 2017. The period includes 21 trading days before and after a speed-inducing technological upgrade. The vertical bar indicates the technological upgrade, "New T7 Trading Technology", which took effect on July 3, 2017. The sample consists of the 100 most active German stocks cross-listed on XSE and CBOE.



The fact that my estimated latency variable decreases following the implemented upgrade provides suggestive evidence that my latency measure is empirically relevant and correctly captures the delay between a signal and a response.

4.4 Empirical findings and discussion

In this section, I examine the role of latency (speed) in fragmented financial markets by linking TL to liquidity and volatility.

4.4.1 Latency and Liquidity

Motivated by the contrasting theories on the impact of speed on market quality characteristics, I begin by testing whether TL is related to liquidity. I estimate the following regression model using fixed effects:

$$Spread_{i,t} = \alpha_i + \beta_t + \gamma latency_{i,t} + \sum_{k=1}^{5} \delta_k C_{k,i,t} + \varepsilon_{i,t} \tag{4.1}$$

where $Spread_{i,t}$ either corresponds to one of *quoted* *(*$Qspread_{i,t}$*)* or *effective* *(*$Espread_{i,t}$*)* *spread* for stock *i* and transaction *t*, α_i and β_t are stock and time fixed effects, $latency_{i,t}$ is the transmission latency between Frankfurt and London for stock *i* and transaction *t*, and $C_{k,i,t}$ is a set of *k* control variables which includes the standard deviation of stock returns *(*$Stddev_{i,t}$*)* for stock *i* and transaction *t* as a proxy for volatility, the inverse of price *(*$InvPri_{i,t}$*)* for stock *i* and transaction *t*, the natural logarithm of trading volume *(*$lnTV_{i,t}$*)* for stock *i* and transaction *t*, market depth *(*$Depth_{i,t}$*)* for stock *i* and transaction *t*, and momentum *(*$Momentum_{i,t}$*)* for stock *i* and transaction *t*. All of the variables are transactions-based (i.e. *t* represents trade time rather than clock time) because my measure of latency is transactions-based.

$Qspread_{i,t}$ is computed as the difference between ask and bid prices for stock *i* corresponding to transaction *t*, $Espread_{i,t}$ is measured as twice the absolute value of the difference between the transaction price and the prevailing bid-ask spread for stock *i* and transaction *t*, $Stddev_{i,t}$ is calculated as the standard deviation of returns for contemporaneous and previous transactions (transactions at time *t* and *t-1*) for stock *i*, $InvPri_{i,t}$ is the inverse of the transaction price for stock *i* and transaction *t*, $lnTV_{i,t}$ is the natural logarithm of trading volume for stock *i* and transaction *t*, $Depth_{i,t}$ is the sum of prevailing bid and ask sizes for stock *i* corresponding to transaction *t*, and $Momentum_{i,t}$ is the first lag of the stock return for stock *i* and transaction *t* (momentum for transaction *t* is the stock return corresponding to transaction *t-1*).

Table 4. 4 Summary statistics and correlation matrix for explanatory variables

This table reports the summary statistics and correlation matrix for the main explanatory variables. Panel A presents the mean and standard deviation of the main explanatory variables and Panel B shows the correlation matrix. All variables are computed for the CBOE. $Qspread$ is the quoted spread and computed as the difference between the best ask and bid prices, $Espread$ is the effective spread and computed as twice the absolute difference between the transaction price and the midpoint of the prevailing bid and ask prices, $AbsCha$ is the absolute value of price changes and computed as the absolute value of price differences between the contemporaneous and previous transactions, $Stddev$ is the standard deviation of stock returns for contemporaneous and previous transactions, $InvPri$ is the inverse price and computed as one divided by the transaction price, $lnTV$ is the natural logarithm of trading volume for each transaction, $Depth$ proxies the market depth and is computed as the sum of the prevailing ask and bid sizes for each transaction, and $Momentum$ is the first lag of stock returns for each transaction. The sample consists of the 100 most active German stocks cross-listed on XSE and CBOE. The sample period covers March 2017 to August 2018. Stocks are classified into quartiles using Euro trading volume.

Panel A

Variables	Full sample		Least active		Quartile 2		Quartile 3		Most active	
	Mean	St. Dev	Mean	St. Dev	Mean	St. Dev	Mean	St. Dev	Mean	St. Dev
$Qspread$ (bps)	454.24	1274	717.19	1445	709.86	2202	610.38	1216	289.61	544.66
$Espread$ (bps)	427.25	1190	670.24	1387	666.489	2063	559.01	997.11	275.43	515.22
$AbsCha$ (bps)	327.63	718.26	460.13	806.78	437.37	1145	371.46	629.75	255.59	444.52
$Stddev$ (bps)	13.35	275.99	20.90	140.18	15.90	315.42	30.99	348.32	8.88	271.96
$InvPri$ (bps)	302.16	340.52	363.80	557.58	217.24	134.89	423.11	319.73	307.01	329.44
$lnTV$	3.88	1.30	3.53	1.26	3.57	1.19	3.93	1.23	4.06	1.32
$Depth$	424.83	724.68	267.25	647.72	233.48	304.81	351.47	802.66	535.17	812.43
$Momentum$ (bps)	0.61	276.35	0.45	141.76	0.87	315.81	1.393	349.91	0.46	272.12

Panel B

	$Espread$	$Qspread$	$AbsCha$	$Stddev$	$InvPri$	$lnTV$	$Depth$	$Momentum$	$Latency$
$Espread$	1								
$Qspread$	0.96	1							
$AbsCha$	0.48	0.47	1						
$Stddev$	0.02	0.02	0.02	1					
$InvPri$	-0.16	-0.15	-0.20	0.00	1				
$lnTV$	-0.15	-0.14	-0.18	-0.00	0.47	1			

$Depth$	-0.10	-0.10	-0.12	-0.00	0.41	0.40	1		
$Momentum$	0.01	0.0	0.00	0.00	-0.00	-0.00	0.00	1	
$Latency$	0.02	0.02	0.00	0.00	0.00	-0.03	-0.01	0.00	1

Panels A and B in Table 4.4 report the mean and standard deviation estimates for all variables, and the correlation between the variables employed in the fixed effects model, respectively. As evident in Panel A, spread and volatility proxies are lower for the most active stocks. The narrower spreads on the most active stocks suggest that higher trading volume encourages traders to provide liquidity, i.e. HFTs are more active in the most active stocks (see also Brogaard et al., 2014b). Furthermore, the smaller absolute value of price changes and standard deviation of stock returns on the most active stocks are consistent with Kyle's (1985) model, in that informed traders participate more in the most active stocks, and this reduces price volatility (see Wang, 1993 for the relationship between informed trading and volatility). The low correlation coefficient estimates between the variables (except for the quoted and effective spreads, which is to be expected) suggest that I do not face multicollinearity issues in the regression models. It is important to note that all variables, except $latency_{i,t}$, are computed for CBOE. This is because, as discussed in Section 4.3.2, information is propagated from Frankfurt to London, hence the effects of latency can only be captured for the satellite market.

I estimate Equation (4.1) for the full sample of stocks and stock trading activity quartiles. I estimate the equation for stock quartiles because Jovanovic and Menkveld (2016) show that the relationship between exchange latency and financial markets may depend on the liquidity of stocks.

Table 4. 5 Latency and liquidity

This table reports the coefficient estimates from the following regression model:

$$Spread_{i,t} = \alpha_i + \beta_t + \gamma latency_{i,t} + \sum_{k=1}^{5} \delta_k C_{k,i,t} + \varepsilon_{i,t}$$

where $Spread_{i,t}$ corresponds to one of quoted ($Qspread_{i,t}$) or effective ($Espread_{i,t}$) spread for stock *i* and transaction *t*, α_i and β_t are stock and time fixed effects, $latency_{i,t}$ is the transmission latency between Frankfurt and London for stock *i* and transaction *t*. $C_{k,i,t}$ is a set of *k* control variables, which includes the standard deviation of stock returns ($Stddev_{i,t}$) for stock *i* and transaction *t* as a proxy for volatility, the inverse of price ($InvPri_{i,t}$) for stock *i* and transaction *t*, the natural logarithm of trading volume ($lnTV_{i,t}$) for stock *i* and transaction *t*, market depth ($Depth_{i,t}$) for stock *i* and transaction *t*, and momentum ($Momentum_{i,t}$) for stock *i* and transaction *t*. $Qspread_{i,t}$ is computed as the difference between ask and bid prices for stock *i* corresponding to transaction *t*, $Espread_{i,t}$ is measured as twice the absolute value of the difference between the transaction price and the prevailing bid-ask spread for stock *i* and transaction *t*, $Stddev_{i,t}$ is calculated as the standard deviation of returns for contemporaneous and previous transactions (transactions at time *t* and *t-1*) for stock *i*, $InvPri_{i,t}$ is the inverse of the transaction price for stock *i* and transaction *t*, $lnTV_{i,t}$ is the natural logarithm of trading volume for stock *i* at time *t*, $Depth_{i,t}$ is the sum of prevailing bid and ask sizes for stock *i* corresponding to transaction *t* and $Momentum_{i,t}$ is the first lag of the stock return for stock *i* at the time of transaction *t* (momentum for time *t* is the stock return at time *t-1*). The sample consists of the 100 most active German stocks that are cross-listed in XSE and CBOE. All variables, except latency, are computed for the CBOE. Stocks are classified into quartiles using Euro trading volume. The sample period covers March 2017 to August 2018. Standard errors are robust to heteroscedasticity and autocorrelation and t-statistics are reported in parentheses. *, ** and *** correspond to statistical significance at the 0.10, 0.05 and 0.01 levels respectively.

Panel A

	Dependent variable: $\boldsymbol{Qspread_{i,t}}$				
	Full sample	Least active	Quartile 2	Quartile 3	Most active
$latency_{i,t}$	0.988x10^{-3}***	0.112x10^{-3}***	0.111x10^{-3}***	0.166x10^{-3}***	0.656x10^{-3}***
	(25.49)	(6.67)	(7.52)	(12.83)	(26.87)
$Stddev_{i,t}$	0.280x10^{-1}***	0.144***	0.267***	0.381x10^{-1}***	0.139x10^{-1}***
	(9.90)	(6.39)	(12.10)	(4.22)	(8.50)
$InvPri_{i,t}$	0.280x10^{-3}	0.599	-0.475	-2.02	0.214
	(0.01)	(1.15)	(-0.78)	(-1.53)	(1.56)
$lnTV_{i,t}$	0.181x10^{-2}***	0.166x10^{-2}***	0.385x10^{-2}***	0.297x10^{-2}***	0.910x10^{-3}***
	(26.18)	(5.57)	(14.42)	(12.18)	(21.21)
$Depth_{i,t}$	0.162x10^{-5}***	0.743x10^{-5}***	0.340x10^{-5}***	0.137x10^{-4}***	0.397x10^{-6}***
	(10.84)	(12.37)	(4.19)	(12.15)	(5.01)

$Momentum_{i,t}$	0.233x10^{-1}***	0.372x10^{-1}*	0.118x10^{-1}	0.694x10^{-1}***	0.544x10^{-2}***
	(8.46)	(1.85)	(0.59)	(8.10)	(3.35)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes
$\overline{R^2}$	41.6%	24.8%	20.9%	48.5%	25.9%

Panel B

	Dependent variable: $\boldsymbol{Espread_{i,t}}$				
	Full sample	Least active	Quartile 2	Quartile 3	Most active
$latency_{i,t}$	0.671x10^{-3}***	0.632x10^{-3}***	0.605x10^{-3}***	0.105x10^{-2}***	0.525x10^{-3}***
	(18.43)	(4.72)	(4.22)	(8.55)	(22.69)
$Stddev_{i,t}$	0.248x10^{-1}***	0.142***	0.244***	0.369x10^{-1}***	0.109x10^{-1}***
	(9.35)	(7.92)	(11.40)	(4.33)	(7.05)
$InvPri_{i,t}$	-0.821x10^{-1}	-0.348x10^{-1}	-0.752x10^{-1}	-2.32*	0.173
	(-0.42)	(-0.08)	(-0.13)	(-1.87)	(1.33)
$lnTV_{i,t}$	0.841x10^{-3}***	0.552x10^{-3}**	0.201x10^{-2}***	0.101x10^{-2}***	0.497x10^{-3}***
	(12.91)	(2.33)	(7.80)	(4.35)	(12.22)
$Depth_{i,t}$	0.108x10^{-5}***	0.560x10^{-5}***	0.234x10^{-5}***	0.116x10^{-4}***	0.197x10^{-7}
	(7.70)	(11.77)	(2.99)	(10.94)	(0.26)
$Momentum_{i,t}$	0.229x10^{-1}***	0.169x10^{-1}	-0.241x10^{-1}	0.740x10^{-1}***	0.559x10^{-2}***
	(8.85)	(1.07)	(-1.25)	(9.14)	(3.63)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes
$\overline{R^2}$	40.9%	29.9%	19.7%	47.5%	25.5%

The results obtained from the estimation of Equation (4.1) are presented in Table 4.5. Standard errors are robust to heteroscedasticity and autocorrelation. The coefficient estimates show that there is a positive relationship between information transmission latency and both quoted and effective spreads. The results hold for all the stock quartiles as well as for the overall sample. This implies that the increases (decreases) in transmission latency (speed) are associated with deteriorations in liquidity. Specifically, the quoted and effective spreads widen by 10 and 7bps respectively for each one unit increase (decrease) in latency (speed). Both estimates are statistically significant at the 0.01 level. The magnitude of the association is also economically meaningful. For example, the results show that a 1ms decrease in latency is expected to reduce quoted (effective) spread by about 10/454 = 2.2% (7/427 = 1.6%). It simply implies that using microwave latency over fibre optic cables (the difference between these two transmission methods is about 1.9ms) for trading information transmission can potentially reduce quoted (effective) spread by 4.2% (3%). This is a substantial change in economic terms, especially, considering the staggering number of such trades that could be placed over the course of one day.

This result suggests that displayed liquidity (quoted spread) improves and trading cost (effective spread) decreases as a result of an acceleration in speed between the home and satellite markets. The results presented in Panels A and B of Table 4.5 are generally consistent, but there is a notable point of departure. While Panel A's estimates show that the effect of latency on spreads is larger in magnitude for the most active stocks compared to the least active stocks, Panel B's estimates show otherwise. Thus, Panel A's results suggest that the positive link between speed and liquidity improvements is mainly driven by the most active stocks, while Panel B's results suggest that the least active stocks are the main drivers of this relationship. This inconsistency may be linked to differences of intuition behind the computation of quoted and effective spreads. Quoted spread is considered the better estimate of trading cost if trades are executed at the quoted prices, while the effective spread is a better measure of trading cost when trades are executed inside the quoted spread (see Petersen and Fialkowski, 1994). Petersen and Fialkowski (1994) further show that the inaccuracy of the quoted spread when trades are executed inside the spread is notably stronger for the most active stocks. Thus, I caution that the evidence presented in Panel A, suggesting that the relationship between liquidity and speed is mainly driven by the most active stocks, should be interpreted carefully. $\overline{R^2}$s for the full sample for the quoted and effective spread regressions are 42% and 41% respectively, which is very high for estimations at transactions (sub-minute) frequency.

This result is consistent with the predictions of Hoffmann (2014) and Jovanovic and Menkveld (2016), and the results of the empirical studies of Hendershott et al. (2011) and Menkveld (2013). In general, the theoretical literature suggests two opposite impacts of latency on liquidity. On the one hand, high-frequency market makers may exploit higher speeds in updating their quotes faster and, hence, face a substantially reduced level of adverse selection risk (see as an example Hoffmann, 2014). On the other hand, speculative high-frequency traders can use higher speed to pick off limit orders of market makers, and thus, increase adverse selection risk (see as an example Biais et al., 2015). My results indicate that, as shown by Menkveld (2013), high-frequency traders generally tend to deploy market making strategies rather than speculative strategies. Being faster allows them to avoid being adversely selected and to manage their inventory more efficiently. This ability to reduce adverse selection risk implies a narrowing of the spread and an improvement in liquidity. In addition to the consistency with the liquidity literature stream, my results are also in line with the findings from the price discovery stream. Specifically, my results imply that high-frequency traders benefit from higher speed to eliminate price distortions quicker, and the improvement in efficient price discovery attracts more traders, thereby further increasing liquidity (see also Brogaard et al., 2014b).

4.4.2 Latency and volatility

Next, I estimate the following regression model using fixed effects in order to test the impact of latency on stock price volatility:

$$Volatility_{i,t} = \alpha_i + \beta_t + \gamma latency_{i,t} + \sum_{k=1}^{5} \delta_k C_{k,i,t} + \varepsilon_{i,t} \tag{4.2}$$

where $Volatility_{i,t}$ corresponds to either the absolute value of price changes ($AbsCha_{i,t}$) or the standard deviation of stock returns ($Stddev_{i,t}$) (see Karpoff, 1987). $AbsCha_{i,t}$ is computed as the absolute value of transaction price differences between transaction t and *t-1*. $C_{k,i,t}$ is a set of k control variables, which includes $Espread_{i,t}$, $InvPri_{i,t}$, $lnTV_{i,t}$, $Depth_{i,t}$, and $Momentum_{i,t}$. All of these variables are as previously defined. The only difference between Equations (4.2) and (4.1) is that instead of volatility, I use a liquidity proxy as one of the control variables rather than as a dependent variable.

Table 4. 6 Latency and volatility

This table reports the coefficient estimates from the following regression model:

$$Volatility_{i,t} = \alpha_i + \beta_t + \gamma latency_{i,t} + \sum_{k=1}^{5} \delta_k C_{k,i,t} + \varepsilon_{i,t}$$

where $Volatility_{i,t}$ corresponds to either absolute value of price change ($AbsCha_{i,t}$) or the standard deviation of stock returns ($Stddev_{i,t}$), α_i and β_t are stock and time fixed effects, $latency_{i,t}$ is the information transmission latency between Frankfurt and London and $C_{k,i,t}$ is a set of *k* control variables, which includes the effective spread ($Espread_{i,t}$) for stock *i* and transaction *t* as a proxy for liquidity, the inverse of price ($InvPri_{i,t}$) for stock *i* at time *t*, the natural logarithm of trading volume ($lnTV_{i,t}$) for stock *i* and transaction *t*, market depth ($Depth_{i,t}$) for stock *i* and transaction *t*, and momentum ($Momentum_{i,t}$) for stock *i* and transaction *t*. $AbsCha_{i,t}$ is computed as the absolute value of transaction price differences between the time of transaction *t* and transaction *t-1*, $Stddev_{i,t}$ is calculated as the standard deviation of returns for contemporaneous and previous transactions (transactions *t* and *t-1*) for stock *i*, $Espread_{i,t}$ is measured as twice the absolute value of the difference between the transaction price and the prevailing bid-ask spread for stock *i* and transaction *t*, $InvPri_{i,t}$ is the inverse of the price for stock *i* and transaction *t*, $lnTV_{i,t}$ is the natural logarithm of trading volume for stock *i* and transaction *t*, $Depth_{i,t}$ is the sum of prevailing bid and ask sizes for stock *i* corresponding to transaction *t*, and $Momentum_{i,t}$ is the first lag of the stock return for stock *i* and transaction *t* (momentum for transaction *t* is the stock return for transaction *t-1*).The sample consists of the 100 most active German stocks that are cross-listed in XSE and CBOE. All variables, except latency, are computed for the CBOE. Stocks are classified into quartiles using Euro trading volume. The sample period covers March 2017 to August 2018. Standard errors are robust to heteroscedasticity and autocorrelation and t-statistics are reported in parentheses. *, ** and *** correspond to statistical significance at the 0.10, 0.05 and 0.01 levels respectively.

Panel A

	Dependent variable: $\boldsymbol{AbsCha_{i,t}}$				
	Full sample	Least active	Quartile 2	Quartile 3	Most active
$latency_{i,t}$	- 0.699x10^{-4}***	- 0.297x10^{-3}***	- 0.274x10^{-3}***	0.142x10^{-4}	- 0.544x10^{-4}***
	(-3.20)	(-3.63)	(-3.49)	(0.21)	(-2.81)
$Espread_{i,t}$	0.129***	0.106***	0.117***	0.126***	0.173***
	(297.96)	(60.75)	(101.04)	(148.27)	(221.64)
$InvPri_{i,t}$	- 0.104	0.387	- 0.463	- 0.722	- 0.833x10^{-1}
	(-0.89)	(1.53)	(-1.43)	(-1.06)	(-0.77)
$lnTV_{i,t}$	0.522x10^{-3}***	0.826x10^{-3}***	0.568x10^{-3}***	0.902x10^{-3}***	0.344x10^{-3}***
	(13.39)	(5.72)	(4.01)	(7.16)	(10.12)
$Depth_{i,t}$	- 0.101x10^{-5}***	- 0.634x10^{-6}**	- 0.856x10^{-6}**	- 0.276x10^{-5}**	- 0.924x10^{-6}**
	(-12.03)	(-2.18)	(-1.99)	(-4.77)	(-14.71)

$Momentum_{i,t}$	- 0.342x10^{-2}**	- 0.158x10^{-1}*	- 0.241x10^{-1}**	- 0.258x10^{-2}	- 0.244x10^{-2*}
	(-2.21)	(-1.65)	(-2.29)	(-0.60)	(-1.89)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes
$\overline{R^2}$	41.8%	34.5%	28.6%	49.4%	30.1%

Panel B

			Dependent variable: $Stddev_{i,t}$		
	Full sample	Least active	Quartile 2	Quartile 3	Most active
$latency_{i,t}$	- 0.193x10^{-4}**	- 0.252x10^{-4}	- 0.269x10^{-4}*	- 0.279x10^{-4}	- 0.128x10^{-4}***
	(-1.94)	(-1.18)	(-1.91)	(-1.24)	(-9.10)
$Espread_{i,t}$	0.185x10^{-2}***	0.363x10^{-2}***	0.237x10^{-2}***	0.124x10^{-2}***	0.399x10^{-2}***
	(9.35)	(7.92)	(11.40)	(4.33)	(7.05)
$InvPri_{i,t}$	0.430x10^{-1}	0.289x10^{-2}	- 0.150**	0.109	0.102
	(0.80)	(0.04)	(-2.57)	(0.48)	(1.30)
$lnTV_{i,t}$	0.928x10^{-5}	0.343x10^{-4}	0.159x10^{-4}	- 0.298x10^{-4}	0.201x10^{-4}
	(0.52)	(0.91)	(0.62)	(-0.71)	(0.82)
$Depth_{i,t}$	- 0.665x10^{-7}*	- 0.370x10^{-8}	- 0.719x10^{-7}	0.281x10^{-7}	- 0.781x10^{-7}*
	(-1.73)	(-0.05)	(-0.93)	(0.14)	(-1.72)
$Momentum_{i,t}$	- 0.668x10^{-1}***	- 0.152***	- 0.618x10^{-1}***	- 0.179***	- 0.156x10^{-1}***
	(-94.41)	(-60.93)	(-32.67)	(-123.03)	(-16.83)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes
$\overline{R^2}$	17.8%	25.3%	23.5%	24.8%	22.9%

I present the results for the full sample and quartile estimations of Equation (4.2) in Table 4.6. Panels A and B show the results for the two stock price volatility proxies. Standard errors are robust to heteroscedasticity and autocorrelation. The estimates suggest a negative (positive) relationship between latency (speed) and volatility for both proxies. Specifically, the absolute value of price change and the standard deviation of stock returns decrease by 0.7 and 0.2bps respectively per unit increase (decrease) in latency (speed). $AbsCha_{i,t}$ and $Stddev_{i,t}$ are statistically significant at the 0.01 and 0.05 levels respectively. Economically what this means is that a decrease in latency from 4.2ms (fibre optic cable) to 2.3ms (microwave connection) is expected to increase standard deviation of stock returns by 1.9 * 0.2/13.32 = 2.8%. The estimates imply that an increase (decrease) in the latency (speed) of order transmission decreases volatility in stock prices. This may not necessarily be a negative effect on market quality if increased speed simply means that new information arrives at the market more often. If this is the case, I would expect to see more rapid changes in prices as investors revise their beliefs about the value of their holdings (see Madhavan et al., 1997). It is important to note that for the absolute value of price changes, the negative (positive) relation between latency (speed) and volatility holds for all quartiles (except Quartile 3) and the overall sample; however, the results for the standard deviation of stock returns suggest that this negative relation is mainly driven by the most active stocks, which indicates cross-sectional differences in the impact of latency on volatility. $\overline{R^2}$s for the full sample results are 42% and 18% respectively, again indicating that my model has a high explanatory power when the frequency of the estimation is considered.

The positive correlation between speed and volatility is consistent with the theory of Roşu (2016) and the empirical studies of Boehmer et al. (2015) and Shkilko and Sokolov (2016). Firstly, Roşu's (2016) model predicts that low latency will improve liquidity. As the market becomes more liquid, fast traders face an even lower price impact, and therefore trade more aggressively. Consequently, increments in aggressiveness in financial markets will increase stock price volatility (see also Collin-Dufresne and Fos, 2016). My finding is in line with this insight and shows that lower (higher) latency (speed) leads to higher volatility. All the estimated coefficients for the control variables are consistent with the literature.

4.4.3 Difference-in-difference estimation of the relationship between speed and market liquidity and volatility

In order to address potential endogeneity, specifically that an unobserved variable correlated with information latency might be driving liquidity/volatility or that there exists some reverse causality between market quality variables (i.e. liquidity and volatility in my set-up), I use a quasi-experimental setting studying two technological upgrades that improved latency on XSE. Specifically, I attempt to causally link the observed changes in liquidity and volatility to latency by employing a DiD framework.

On July 3, 2017 and April 9, 2018, XSE implemented upgrades to increase the exchange's speed (see Section 4.2.2 for details on the two upgrades). I compare the changes in the liquidity and volatility of stocks affected by the technological upgrades with those that are unaffected by estimating the following regression model:

$$DP_{i,d} = \alpha_i + \beta_d + \gamma_1 Event_d + \gamma_2 Treatment_i + \gamma_3 Event_d \times Treatment_i + \sum_{k=1}^{8} \delta_k C_{k,i,d} + \varepsilon_{i,d} \qquad (4.3)$$

where i denotes stocks and d denotes days. α_i and β_d are stock and time fixed effects. The dependent variable $DP_{i,d}$ corresponds to one of the liquidity and volatility proxies: quoted ($Qspread_{i,d}$) and effective ($Espread_{i,d}$) spreads for liquidity, and absolute value of price changes ($AbsCha_{i,d}$) and standard deviation of stock returns ($Stddev_{i,d}$) for volatility. $Qspread_{i,d}$ is the average of the differences between the ask and bid prices corresponding to each transaction, $Espread_{i,d}$ is a daily average, each intraday value is computed as twice the absolute value of the difference between a transaction's price and the prevailing bid-ask spread, $AbsCha_{i,d}$ measures the absolute difference between the last prices for stock i for days d and d-1, and $Stddev_{i,d}$ is the standard deviation of transaction prices' returns. Consistent with Equations (4.1) and (4.2), all variables are computed for CBOE.

$Event_d$ is a dummy taking the value 0 for the pre-upgrade period and 1 for the post-upgrade period. I employ a 4-month horizon to assess the impact; d comprises [-120; +120] days. It is important to note that my results are robust to different horizons: 1-, 2-, or 3-month periods before and after the upgrade. $Treatment_i$ is a dummy taking the value 1 for stocks that are affected by the upgrade and zero for stocks that are not. Specifically, my treatment group is the 100 stocks that are cross-listed on both XSE and CBOE. Hence, any XSE exchange latency upgrade will impact the *TL* of these stocks. My control group comprises of 100 stocks that are only listed on CBOE and not on XSE; thus, upgrades should not have any impact on them. In this framework, my treatment and control groups belong to different countries. However, this should not have a material impact on my results for at least two reasons. Firstly, the results are based on variations at frequencies less than one second; at these frequencies, microstructure effects are unlikely to be driven by regulatory regimes in the case of stocks trading in quite similar market structures. Secondly, all of the stocks in both groups are domiciled and traded within the jurisdiction of the European Securities Market Authority (ESMA), and are therefore covered by largely similar regulatory regimes. The approach of including stocks from different countries within the same DiD framework is consistent with the literature (see as an example Malceniece et al., 2018). Furthermore, in order to ensure that I compare like-for-like as much as possible, I employ the approach developed by Boulton and Braga-Alves (2010) to match each of the treatment stocks to a corresponding control stock; the matching variable is trading activity.

$C_{k,i,d}$ is a set of k control variables, which includes $Momentum_{i,d}$, $InvPri_{i,d}$, $Stddev_{i,d}$ (in the liquidity models), $Espread_{i,d}$ (in the volatility models), $lnTV_{i,d}$, $TimeT_{i,d}$, $Depth_{i,d}$, $Transactions_{i,d}$, and $Macro_{i,d}$. $Momentum_{i,d}$ is the first lag of daily return ($Momentum_{i,d}$ is the return of stock i on day d-1), $InvPri_{i,d}$ is the inverse of last transaction price, $lnTV_{i,d}$ is the natural logarithm of trading volume, $TimeT_{i,d}$ is a trend variable starting at zero at the beginning of the sample period and increasing by one every trading day d, $Depth_{i,d}$ is computed as the sum of ask and bid sizes, $Transactions_{i,d}$ is the number of transactions and $Macro_{i,d}$ is a dummy taking the value 1 for days with macroeconomic announcements, and zero otherwise. $Stddev_{i,d}$ and $Espread_{i,d}$ are as previously defined. γ_1 captures any common effects that might have impacted all stocks following the upgrade, γ_2 captures any pre-existing differences between the treatment and control groups. γ_3, the key coefficient, captures the interaction of $Event_d$ and $Treatment_i$ and thus estimates any incremental effect of the upgrades on the treatment group. The model is estimated with firm and time fixed effects, and standard errors are robust to heteroscedasticity and autocorrelation. Similar to Equations (4.1) and (4.2), I estimate the model for the full sample and quartiles. The DiD model is also estimated under various specifications, with and without the control variables.[39]

[39] I find that there is no material difference in the coefficients of interest between the two specifications. For parsimony, I present the results with control variables only.

Table 4. 7 Difference-in-difference estimation of the effects of latency on liquidity

This table examines the relationship between liquidity and latency by exploiting two technological upgrades on July 3, 2017 and April 9, 2018. Specifically, the table reports coefficient estimates from the following regression model, with observations sampled at the daily frequency:

$$DP_{i,d} = \alpha_i + \beta_d + \gamma_1 Event_d + \gamma_2 Treatment_i + \gamma_3 Event_d \times Treatment_i + \sum_{k=1}^{8} \delta_k C_{k,i,d} + \varepsilon_{i,d}$$

where $DP_{i,d}$ corresponds to one of two liquidity proxies: quoted ($Qspread_{i,d}$) and effective ($Espread_{i,d}$) spreads. $Qspread_{i,d}$ is the average of the differences between the ask and bid prices corresponding to each transaction, $Espread_{i,d}$ is a daily average, each intraday value is computed as twice the absolute value of the difference between a transaction's price and the prevailing bid-ask spread. $Event_d$ is a dummy taking the value zero for the pre-upgrade period and one for the post-upgrade period, and $Treatment_i$ is a dummy taking the value 1 for stocks impacted by the upgrade and zero for stocks not affected by the upgrade. The treatment group consists of the 100 stocks cross-listed on XSE and CBOE and the control group includes the 100 stocks listed on CBOE, but ***not*** cross-listed on XSE. $C_{k,i,d}$ is a set of *k* control variables, which includes $Momentum_{i,d}$, $InvPri_{i,d}$, $Stddev_{i,d}$, $lnTV_{i,d}$, $TimeT_{i,d}$, $Depth_{i,d}$, $Transactions_{i,d}$ and $Macro_{i,d}$. $Momentum_{i,d}$is the first lag of daily return for stock *i* on day *d* ($Momentum_{i,d}$ is the return of stock *i* on day *d-1*), $InvPri_{i,d}$is the inverse of last transaction price for stock *i* on day *d*, $Stddev_{i,d}$ is the standard deviation of transaction prices for stock *i* during day *d*, $lnTV_{i,d}$ is the natural logarithm of trading volume for stock *i* on day *d*, $TimeT_{i,d}$ is a trend variable for each stock *i* starting at zero at the beginning of the sample period and increasing by one every trading day *d*, $Depth_{i,d}$is computed as the sum of ask and bid sizes for stock *i* on day *d*, $Transactions_{i,d}$ is the number of transactions for stock *i* on day *d* and $Macro_{i,d}$ is a dummy for stock *i* and takes the value one for days *d*s with macroeconomic announcements and zero otherwise. Stocks are classified into quartiles using Euro trading volume. Firm and time fixed effects are employed, and standard errors are robust to heteroscedasticity and autocorrelation. t-statistics are reported in parentheses. The sample period covers [-4; +4 months] intervals around each upgrade. *, ** and *** correspond to statistical significance at the 0.10, 0.05 and 0.01 levels respectively.

Panel A

	Dependent variable: $\boldsymbol{Qspread_{i,d}}$				
	Full sample	Least active	Quartile 2	Quartile 3	Most active
$Event_d$	0.103x10^{-2}***	0.147x10^{-3}	0.247x10^{-2}***	0.415x10^{-3}***	0.104x10^{-2}***
	(6.06)	(0.31)	(6.27)	(3.65)	(4.17)
$Treatment_i$	-0.209x10^{-2}***	0.399x10^{-3}	-0.135x10^{-2}***	-0.823x10^{-3}***	-0.293x10^{-2}***
	(-19.38)	(1.26)	(-5.45)	(-11.41)	(-16.23)
$Event_d \times Treatment_i$	-0.453x10^{-3}***	-0.189x10^{-3}	-0.184x10^{-2}***	-0.202x10^{-3}**	-0.252x10^{-3}***
	(-2.95)	(-0.44)	(-5.19)	(-1.98)	(-11.12)
$Momentum_{i,d}$	0.154x10^{-2}***	0.560x10^{-4}	0.191x10^{-2}	0.650x10^{-3}**	0.307x10^{-2}***
	(3.55)	(0.05)	(0.72)	(2.34)	(6.11)
$InvPri_{i,d}$	-0.159x10^{-1}***	-0.711x10^{-2}***	-0.313x10^{-1}***	-0.193x10^{-1}***	-0.225x10^{-1}***

	(-22.30)	(-4.63)	(-20.39)	(-17.72)	(-18.59)
$Stddev_{i,d}$	0.299x10^{-3}*** (3.84)	0.672x10^{-3}*** (3.25)	0.834x10^{-3}* (1.79)	0.212x10^{-3}*** (4.18)	0.669x10^{-4} (0.77)
$lnTV_{i,d}$	0.151x10^{-3}*** (13.98)	-0.122x10^{-3}*** (-3.53)	0.638x10^{-4}** (2.47)	0.154x10^{-3}*** (19.55)	0.204x10^{-3}*** (13.72)
$TimeT_{i,d}$	-0.486x10^{-5}*** (-3.15)	0.318x10^{-6} (0.07)	-0.695x10^{-5}* (-1.94)	-0.251x10^{-5}** (-2.45)	-0.788x10^{-5}*** (-3.50)
$Depth_{i,d}$	0.113x10^{-5}*** (13.80)	0.865x10^{-6}*** (3.23)	0.318x10^{-5}*** (17.48)	-0.651x10^{-5}*** (-6.92)	0.340x10^{-7} (0.33)
$Transactions_{i,d}$	0.566x10^{-6}*** (12.70)	0.396x10^{-5}*** (16.85)	0.173x10^{-5}*** (15.54)	0.190x10^{-6}*** (5.84)	0.745x10^{-6}*** (14.17)
$Macro_{i,d}$	-0.218x10^{-3}*** (-2.61)	-0.302x10^{-3} (-1.30)	-0.283x10^{-3} (-1.47)	-0.118x10^{-3}** (-2.12)	-0.239x10^{-3}* (-1.95)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes
$\overline{R^2}$	36.3%	35.8%	17.7%	38.6%	48.8%

Panel B

	Dependent variable: $\boldsymbol{Espread_{i,d}}$				
	Full sample	Least active	Quartile 2	Quartile 3	Most active
$Event_d$	0.190x10^{-2}*** (5.45)	0.809x10^{-3}* (1.95)	0.435x10^{-2}*** (4.40)	0.512x10^{-3}** (2.06)	0.178x10^{-2}** (2.47)
$Treatment_i$	-0.436x10^{-2}*** (-19.84)	0.284x10^{-2}*** (10.35)	0.748x10^{-3} (1.20)	-0.151x10^{-2}*** (-9.64)	-0.885x10^{-2}*** (-16.83)
$Event_d \times Treatment_i$	-0.977x10^{-3}*** (-3.12)	-0.745x10^{-3}** (-1.99)	-0.404x10^{-2}*** (-4.52)	-0.184x10^{-3} (-0.83)	0.243x10^{-3} (0.37)
$Momentum_{i,d}$	0.267x10^{-2}*** (3.02)	-0.181x10^{-3} (-0.19)	0.414x10^{-3} (0.06)	0.125x10^{-2}** (2.06)	0.518x10^{-2}*** (3.53)
$InvPri_{i,d}$	-0.371x10^{-1}*** (-25.44)	-0.160x10^{-1}*** (-11.99)	-0.847x10^{-1}*** (-21.98)	-0.346x10^{-1}*** (-14.50)	-0.595x10^{-1}*** (-16.87)

$Stddev_{i,d}$	$0.106\text{x}10^{-2}$*** (6.71)	$0.629\text{x}10^{-3}$*** (3.49)	$0.743\text{x}10^{-3}$ (0.63)	$0.242\text{x}10^{-2}$*** (21.85)	$0.454\text{x}10^{-3}$* (1.81)
$lnTV_{i,d}$	$0.143\text{x}10^{-3}$*** (6.52)	$-0.617\text{x}10^{-3}$*** (-20.53)	$-0.445\text{x}10^{-3}$*** (-6.85)	$0.249\text{x}10^{-3}$*** (14.48)	$0.377\text{x}10^{-3}$*** (8.71)
$TimeT_{i,d}$	$-0.713\text{x}10^{-5}$** (-2.26)	$-0.162\text{x}10^{-5}$ (-0.43)	$-0.422\text{x}10^{-5}$ (-0.47)	$-0.409\text{x}10^{-5}$* (-1.83)	$-0.157\text{x}10^{-4}$** (-2.39)
$Depth_{i,d}$	$0.228\text{x}10^{-5}$*** (13.69)	$0.298\text{x}10^{-5}$*** (12.79)	$0.610\text{x}10^{-5}$*** (13.36)	$-0.100\text{x}10^{-5}$*** (-4.87)	$0.872\text{x}10^{-7}$ (0.29)
$Transactions_{i,d}$	$0.305\text{x}10^{-5}$*** (33.56)	$0.134\text{x}10^{-4}$*** (65.47)	$0.918\text{x}10^{-5}$*** (32.93)	$0.563\text{x}10^{-6}$*** (7.91)	$0.315\text{x}10^{-5}$*** (20.60)
$Macro_{i,d}$	$-0.208\text{x}10^{-3}$ (-1.22)	$-0.401\text{x}10^{-3}$** (-1.98)	$-0.105\text{x}10^{-3}$ (-0.22)	$-0.140\text{x}10^{-3}$ (-1.15)	$-0.542\text{x}10^{-3}$ (-1.52)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes
$\overline{\boldsymbol{R^2}}$	30.3%	31.5%	21.6%	8.9%	9.2%

Table 4.7 reports the estimation results for when $DP_{i,d}$ in Equation (4.3) corresponds to either the quoted or effective spreads.

The interaction coefficients (γ_3) suggest that the technological upgrades are linked with decreases of about 4.5bps and 10bps in quoted and effective spreads respectively for the treated group of stocks, when compared to the control group. Both estimates are statistically significant at the 0.01 level. In order to put the economic significance of this result into some perspective, recall that the average latency reduction from the two upgrades, based on our analysis (see Panel C in Table 4.3 and Footnote 38), is about 2% or 0.08ms (2% * 4.39). Thus, a 2% (0.08ms) reduction in latency is estimated to decrease quoted (effective) spread by 4.5/454 = 1% (10/427 = 2.3%). This implies that, following the upgrade, liquidity increases and trading costs decrease more for my treatment group relative to the control group, and it further shows that the latency improvements are, over and above other controlled effects, driving stock market liquidity. Importantly, the fact that stocks that were expected to benefit from the technological upgrades see a significant improvement in liquidity allows me to establish a causal relationship between speed and liquidity, while ruling out endogeneity concerns. Therefore, the results are consistent with the earlier fixed effect models. The findings of the DiD frameworks are also consistent with the predictions of Hoffmann (2014) and Jovanovic and Menkveld (2016), and with the empirical findings of Menkveld (2013) and Hendershott et al. (2011), and suggest that speed is generally used by high-frequency market makers as a means of reducing adverse selection risk, thus leading to their provision of a higher level of liquidity. Similar to the earlier estimated fixed effects model for liquidity, while the positive relationship between speed improvements and quoted spread is driven by the most active stocks, the positive relationship between speed improvements and effective spread is driven by the least active stocks. The estimated coefficients of the control variables are generally consistent with the literature. The $\overline{R^2}$ for the quoted and effective spread models are 36% and 30%, respectively. These are substantial explanatory levels for daily frequency estimations.

Table 4. 8 Difference-in-difference estimation of the effects of latency on volatility

This table examines the relationship between volatility and latency around two technological upgrades on July 3, 2017 and April 9, 2018. Specifically, the table reports coefficient estimates from the following regression model using daily frequencies:

$$DP_{i,d} = \alpha_i + \beta_d + \gamma_1 Event_d + \gamma_2 Treatment_i + \gamma_3 Event_d \times Treatment_i + \sum_{k=1}^{8} \delta_k C_{k,i,d} + \varepsilon_{i,d}$$

where $DP_{i,d}$ corresponds to one of two volatility proxies: absolute value of price changes ($AbsCha_{i,d}$) and standard deviation of stock returns ($Stddev_{i,d}$). $AbsCha_{i,d}$ is the absolute difference between the last prices for stock *i* for days *d* and *d-1*, $Stddev_{i,d}$ is the standard deviation of transaction prices for stock *i* during day *d*. $Event_d$ is a dummy taking the value zero for the pre-upgrade period and one for the post-upgrade period, and $Treatment_i$ is a dummy taking the value one for stocks that are impacted by the upgrade and zero for stocks that are not. The treatment group consists of the 100 stocks cross-listed on XSE and CBOE and the control group includes the 100 stocks listed on CBOE, but ***not*** cross-listed on XSE. $C_{k,i,d}$ is a set of *k* control variables, which includes $Momentum_{i,d}$, $InvPri_{i,d}$, $Espread_{i,d}$, $lnTV_{i,d}$, $TimeT_{i,d}$, $Depth_{i,d}$, $Transactions_{i,d}$ and $Macro_{i,d}$. $Momentum_{i,d}$is the first lag of daily return for stock *i* on day *d* ($Momentum_{i,d}$ is the return of stock *i* on day *d-1*), $InvPri_{i,d}$is the inverse of last transaction price for stock *i* on day *d*. $Espread_{i,d}$ is a daily average, each intraday value is computed as twice the absolute value of the difference between a transaction's price and the prevailing bid-ask spread. $lnTV_{i,d}$ is the natural logarithm of trading volume for stock *i* on day *d*, $TimeT_{i,d}$ is a trend variable for each stock *i* starting at zero at the beginning of the sample period and incrementing by one every trading day *d*, $Depth_{i,d}$is computed as the sum of ask and bid sized for stock *i* on day *d*, $Transactions_{i,d}$ is the number of transactions for stock *i* on day *d*, and $Macro_{i,d}$ is a dummy for stock *i* taking the value one for days *d* with macroeconomic announcements and zero otherwise. Stocks are classified into quartiles using Euro trading volume. Firm and time fixed effects are employed, and standard errors are robust to heteroscedasticity and autocorrelation. t-statistics are reported in parenthesis. The sample period covers [-4; +4] intervals around each upgrade. *, ** and *** correspond to statistical significance at the 0.10, 0.05 and 0.01 levels respectively.

Panel A

	Dependent variable: $\boldsymbol{AbsCha_{i,d}}$				
	Full sample	Least active	Quartile 2	Quartile 3	Most active
$Event_d$	-0.987	-2.832	-1.552	1.389	-0.989
	(-0.69)	(-0.95)	(-0.57)	(0.39)	(-0.51)
$Treatment_i$	-2.600***	-0.814	-4.966***	-0.444	-2.827**
	(-2.88)	(-0.41)	(-2.90)	(-0.19)	(-2.00)
$Event_d \times Treatment_i$	2.550**	1.721	4.629*	0.459	3.174*
	(1.98)	(0.64)	(1.89)	(0.14)	(1.82)
$Momentum_{i,d}$	3.130	3.472	13.390	-0.496	3.408
	(0.86)	(0.51)	(0.73)	(-0.06)	(0.87)
$InvPri_{i,d}$	-14.183**	0.803	-32.249***	-50.965	-4.981

	(-2.36)	(0.08)	(-3.01)	(-1.47)	(-0.53)
$Espread_{i,d}$	61.664*** (3.84)	29.520 (0.52)	-6.597 (-0.31)	325.322*** (2.93)	166.247*** (8.00)
$lnTV_{i,d}$	0.117 (1.31)	-0.344 (-1.59)	0.365** (2.05)	0.172 (0.68)	0.101 (0.88)
$TimeT_{i,d}$	0.262x10^{-2} (0.20)	0.385x10^{-1} (1.44)	-0.413x10^{-2} (-0.17)	-0.159x10^{-1} (-0.49)	-0.479x10^{-2} (-0.27)
$Depth_{i,d}$	-0.862x10^{-4} (-0.13)	0.898x10^{-3} (0.57)	-0.503x10^{-3} (-0.40)	-0.716x10^{-3} (-0.24)	0.315x10^{-3} (0.39)
$Transactions_{i,d}$	-0.177x10^{-4} (-0.05)	-0.291x10^{-2}* (1.76)	-0.130x10^{-3} (-0.17)	-0.134x10^{-4} (-0.01)	-0.312x10^{-3} (-0.76)
$Macro_{i,d}$	-0.999x10^{-1} (-0.14)	-1.838 (-1.26)	0.133 (0.10)	0.560 (0.32)	0.708 (0.75)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes
$\overline{\boldsymbol{R}^{2}}$	25.9%	8.3%	10.6%	7.7%	46.9%

Panel B

	Dependent variable: $\boldsymbol{Stddev_{i,d}}$				
	Full sample	Least active	Quartile 2	Quartile 3	Most active
$Event_d$	0.114x10^{-1} (1.34)	0.152x10^{-1} (0.83)	0.839x10^{-2} (1.27)	0.402x10^{-3} (0.02)	0.206x10^{-1} (0.91)
$Treatment_i$	-0.225x10^{-1}*** (-4.15)	-0.250x10^{-1}** (-2.05)	-0.109x10^{-1}*** (-2.64)	-0.152x10^{-1} (-1.39)	-0.496x10^{-1}*** (-3.02)
$Event_d \times Treatment_i$	0.281x10^{-1}*** (3.65)	0.320x10^{-1}* (1.94)	0.144x10^{-1}** (2.42)	0.357x10^{-1}** (2.32)	0.364x10^{-1}* (1.80)
$Momentum_{i,d}$	-0.122x10^{-1} (-0.56)	-0.124x10^{-1} (-0.29)	-0.214x10^{-1} (-0.48)	-0.281x10^{-1} (-0.67)	-0.152x10^{-1} (-0.33)
$InvPri_{i,d}$	-0.116*** (-3.19)	-0.967x10^{-1} (-1.63)	-0.622x10^{-1}** (-2.39)	-0.516x10^{-1} (-0.31)	-0.246** (-2.23)

$Espread_{i,d}$	0.647***	1.227***	0.329×10^{-1}	11.643***	0.437*
	(6.71)	(3.49)	(0.63)	(21.85)	(1.81)
$lnTV_{i,d}$	0.525×10^{-2}***	0.611×10^{-2}***	0.307×10^{-2}***	0.438×10^{-2}***	0.632×10^{-2}***
	(9.69)	(4.54)	(7.11)	(3.64)	(4.69)
$TimeT_{i,d}$	-0.527×10^{-3}***	-0.615×10^{-3}***	-0.303×10^{-3}***	-0.484×10^{-3}***	-0.725×10^{-3}***
	(-6.80)	(-3.71)	(-5.08)	(-3.12)	(-3.56)
$Depth_{i,d}$	0.339×10^{-5}	0.113×10^{-4}	0.151×10^{-5}	0.582×10^{-5}	0.133×10^{-4}
	(0.82)	(1.09)	(0.50)	(0.41)	(1.42)
$Transactions_{i,d}$	-0.705×10^{-5}***	-0.132×10^{-4}	-0.382×10^{-5}**	-0.153×10^{-4}***	-0.420×10^{-5}
	(-3.13)	(-1.30)	(-2.00)	(-3.10)	(-0.87)
$Macro_{i,d}$	0.880×10^{-2}**	0.644×10^{-2}	0.380×10^{-2}	0.131×10^{-1}	0.116×10^{-1}
	(2.09)	(0.72)	(1.18)	(1.56)	(1.05)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes
$\overline{R^2}$	29.5%	35.2%	47.9%	31.1%	26.9%

Table 4.8 reports the estimation results for the volatility measures, i.e. the absolute value of price change and the standard deviation of intraday stock returns for stock *i* on day *d*. The interaction coefficients (γ_3) suggest that the technological upgrades are linked with increases in volatility. Both the absolute value of price change and the standard deviation of stock returns (volatility proxies) increase by 2.550 and 0.028 bps respectively for the treatment group of stocks in comparison to the control group; the changes are statistically significant at 0.05 and 0.01 levels respectively. The economic significance of these estimates is put into some perspective when we recall that the difference between the latencies of microwave and fibre optic cable is about 23 times higher than this reduction (1.9/0.08). Again, the results allow me to confirm the causal link between speed and volatility. Generally, the findings presented in Table 4.8 further support my earlier results and are consistent with the empirical findings of Shkilko and Sokolov (2016) and Boehmer et al. (2015). As already noted, the positive relationship between speed and volatility is related to increased aggressiveness in financial markets (for a more detailed discussion, see also Collin-Dufresne and Fos, 2016; Roşu, 2016). The $\overline{R^2}$ for the absolute value of price change and standard deviation of stock return models are 26% and 30%, respectively.

4.4.4 How does latency impact liquidity and volatility?

In this section's analysis, I focus on the two channels literature identifies as potential avenues through which speed impacts liquidity. The first is that speed aids adverse selection risk avoidance by liquidity providers; I call this the *adverse selection avoidance channel* (see as examples Hoffmann, 2014; Jovanovic and Menkveld, 2016). The second relates to speed helping faster traders to pick-off the limit orders of slower traders and hence end up decreasing liquidity (see as examples Foucault et al., 2016; Foucault et al., 2017); I call the channel the *picking-off channel* in line with Shkilko and Sokolov (2016).

The results suggest that fast traders generally use their speed advantage to avoid adverse selection rather than to pick-off limit orders of liquidity providers (see also Hagströmer and Nordén, 2013; Menkveld, 2013), or at least the effect of the former action dominates the effect of the latter action. In this section, I explore this issue further. If my argument about the adverse selection avoidance channel is valid, then latency improvements should be accompanied by smaller price impacts (see also Shkilko and Sokolov, 2016). This argument links to the stream of the market microstructure literature focusing on the links between price impact and liquidity. The argument is also anchored on the theoretical approach described in Menkveld (2013). Specifically, HFTs using their speed advantage to avoid adverse selection tend to follow market making strategies, basically working off a model based on their profits coming from the bid-ask spread. Adverse selection risk thus poses a risk to their strategy. It follows that if becoming quicker helps them to further decrease the risk of adverse selection, they would be more willing to provide more liquidity, which decreases a trade's price impact.

I follow Baron et al. (2018) in designing a test of the adverse selection avoidance channel. I use a fixed effects framework similar to Equations (4.1) and (4.2), the difference being that I now employ the price impact of each transaction as my new dependent variable. Price impact is computed as in Shkilko and Sokolov (2016); $PRIMP_{i,t} = 2q_t(mid_{t+1} - mid_t)$, where q_t is the direction of the trade,[40] and mid_t and mid_{t+1} are the prevailing midquotes for transactions *t* and *t+1* respectively. Specifically, I run the following model:

$$PRIMP_{i,t} = \alpha_i + \beta_t + \gamma latency_{i,t} + \sum_{k=1}^{6} \delta_k C_{k,i,t} + \varepsilon_{i,t} \qquad (4.4)$$

[40] I follow Lee and Ready (1991) algorithm to classify trades as sell and buy trades.

Table 4. 9 Price Impact and Latency: testing *adverse selection* channel

This table reports the coefficient estimates from the following regression model:

$$PRIMP_{i,t} = \alpha_i + \beta_t + \gamma latency_{i,t} + \sum_{k=1}^{6} \delta_k C_{k,i,t} + \varepsilon_{i,t}$$

where $PRIMP_{i,t}$ corresponds to the price impact for stock *i* and transaction *t*, α_i and β_t are stock and time fixed effects, $latency_{i,t}$ is information transmission latency between Frankfurt and London. $PRIMP_{i,t} = 2q_t(mid_{t+1} - mid_t)$, where q_t is the direction of trade, mid_t and mid_{t+1} are the mid-quotes for transaction *t* and *t+1*. $C_{k,i,t}$ is a set of *k* control variables, which includes the standard deviation of stock returns ($Stddev_{i,t}$) for stock *i* and transaction *t* as a proxy for volatility, the effective spread ($Espread_{i,t}$) for stock *i* and transaction *t* as a proxy for liquidity, the inverse of price ($InvPri_{i,t}$) for stock *i* and transaction *t*, the natural logarithm of trading volume ($lnTV_{i,t}$) for stock *i* and transaction *t*, market depth ($Depth_{i,t}$) for stock *i* and transaction *t*, and momentum ($Momentum_{i,t}$) for stock *i* and transaction *t*. $Stddev_{i,t}$ is calculated as the standard deviation of returns for contemporaneous and previous transactions (transactions at time *t* and *t-1*) for stock *i*, $Espread_{i,t}$ is measured as twice the absolute value of the difference between the transaction price and the prevailing bid-ask spread for stock *i* at time *t*, $InvPri_{i,t}$ is the inverse of the transaction price for stock *i* at time *t*, $lnTV_{i,t}$ is the natural logarithm of trading volume for stock *i* and transaction *t*, $Depth_{i,t}$ is the sum of prevailing bid and ask sizes for stock *i* corresponding to transaction *t* and $Momentum_{i,t}$ is the first lag of the stock return for stock *i* and transaction *t* (momentum for transaction *t* is the stock return transaction *t-1*). The sample consists of the 100 most active German stocks cross-listed on XSE and CBOE. All variables, except latency, are computed for the CBOE. Stocks are classified into quartiles using Euro trading volume. The sample period covers March 2017 to August 2018. Standard errors are robust to heteroscedasticity and autocorrelation and t-statistics are reported in parentheses. *, ** and *** correspond to statistical significance at the 0.10, 0.05 and 0.01 levels respectively.

	Dependent variable: $\boldsymbol{PRIMP_{i,t}}$				
	Full sample	Least active	Quartile 2	Quartile 3	Most active
$latency_{i,t}$	0.971x10^{-3}**	0.243x10^{-2}	0.108x10^{-2}*	-0.429x10^{-3}	0.333x10^{-2}***
	(2.18)	(0.89)	(1.73)	(-1.26)	(3.10)
$Momentum_{i,t}$	0.502***	-8.426***	-1.059***	0.191x10^{-1}	1.358***
	(15.68)	(-26.33)	(-3.65)	(0.84)	(19.17)
$InvPri_{i,t}$	1.241	-18.942**	7.948	3.971**	-10.334
	(0.51)	(-2.27)	(0.89)	(2.08)	(-0.95)
$Espread_{i,t}$	-0.964x10^{-1}***	-0.325x10^{-1}	-0.115***	-0.159***	-0.470x10^{-1}***
	(-10.86)	(-0.56)	(-3.60)	(-11.52)	(-3.44)
$Stddev_{i,t}$	9.111***	-14.958	-22.525***	12.895***	4.677***
	(277.62)	(-41.40)	(-69.73)	(564.21)	(62.65)
$lnTV_{i,t}$	0.382x10^{-2}***	0.194x10^{-2}	0.571x10^{-2}	0.349x10^{-2}***	0.495x10^{-2}**

	(4.82)	(0.41)	(1.46)	(5.83)	(2.45)
$Depth_{i,t}$	0.435x10^{-6}	0.591x10^{-6}	-0.491x10^{-5}	0.191x10^{-5}*	-0.143x10^{-4}
	(0.25)	(0.06)	(-0.42)	(1.73)	(-1.54)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes
$\overline{\boldsymbol{R^2}}$	14.1%	27.7%	7.4%	22.9%	14.1%

where α_i and β_t are stock and time fixed effects respectively, and $latency_{i,t}$is the TL between XSE and CBOE for transaction t and stock i. $C_{k,i,t}$ is a set of previously defined k control variables, which includes $Stddev_{i,t}$, $Espread_{i,t}$, $InvPri_{i,t}$, $lnTV_{i,t}$, $Depth_{i,t}$, and $Momentum_{i,t}$.

The results obtained from the estimation of Equation (4.4) are presented in Table 4.9. The standard errors are robust to heteroscedasticity and autocorrelation. The estimated latency coefficient in Table 4.9 is positive and statistically significant at the 0.05 level. This implies that, consistent with the previous sets of results, the price impact increases (decreases) by 10bps per unit increase in latency (speed). The magnitude of the effect is also economically meaningful; a 1ms decrease in latency is expected to decrease price impact by 4% (10/254). The results are therefore in line with the adverse selection avoidance channel argument, and suggest that speed incentivizes liquidity providers to trade more as it helps them to avoid being adversely selected (see also Hoffmann, 2014; Jovanovic and Menkveld, 2016). The result does not hold for the least active stocks, which might be explained by the concentration of HFTs in the most active stocks. The $\overline{R^2}$ for the full sample is 14%.

As previously demonstrated, volatility also increases with speed. Therefore, I next investigate the channel(s) through which speed impacts volatility. A potential channel is described in the theoretical model presented by Roşu (2016). Specifically, the model shows that as market liquid improves (which, as shown above, is a consequence of speed), fast traders face a lower price impact, and therefore trade even more aggressively. Consequently, increments in aggressiveness in financial markets will increase stock price volatility (see also Collin-Dufresne and Fos, 2016 for the relationship between aggressiveness and volatility). I call this channel the *aggressiveness channel*. A series of estimations already show that speed decreases price impact and increases liquidity (see Tables 4.5, 4.7 and 4.9 for details) and form the first part of the test of the veracity of the *aggressiveness channel*. In order to execute the second part of the test of the *aggressiveness channel* argument, i.e. the role of latency on aggressiveness, I estimate the logit regression model in Equation (4.5). If my intuition is valid, then latency (speed) should decrease (increase) aggressiveness since Roşu (2016) predicts that speed can increase volatility through its impact on aggressiveness.

$$Aggressiveness_{i,t} = \alpha_i + \beta_t + \gamma latency_{i,t} + \sum_{k=1}^{6} \delta_k C_{k,i,t} + \varepsilon_{i,t} \tag{4.5}$$

where $Aggressiveness_{i,t}$ is a binary dependent variable for stock *i* and transaction *t*. Specifically, $Aggressiveness_{i,t}$ equals one for the aggressive trades and zero otherwise. In order to classify trades according to their aggressiveness, I employ the modified version of the approach proposed by Barber et al. (2009) and Kelley and Tetlock (2013). I start by determining the direction of each transaction in the spirit of Lee and Ready (1991). Then, I compare the transaction price with the prevailing best bid (ask) price for sell (buy) transactions. If the transaction price is below (above) or equal to the prevailing best bid (ask) price, I classify this sell (buy) transaction as an aggressive trade. α_i and β_t are stock and time fixed effects, $latency_{i,t}$is the TL between XSE and CBOE. $C_{k,i,t}$ is a set of previously defined k control variables, which includes $Stddev_{i,t}$, $Espread_{i,t}$, $InvPri_{i,t}$, $lnTV_{i,t}$, $Depth_{i,t}$, and $Momentum_{i,t}$.

Table 4. 10 Trade Aggressiveness and Latency: testing *aggressiveness* channel

This table reports the coefficient estimates from the following logit regression model:

$$Aggressiveness_{i,t} = \alpha_i + \beta_t + \gamma latency_{i,t} + \sum_{k=1}^{6} \delta_k C_{k,i,t} + \varepsilon_{i,t}$$

where $Aggressiveness_{i,t}$ is a binary dependent variable for stock *i* and transaction *t*. Specifically, $Aggressiveness_{i,t}$ equals one for aggressive trades and zero otherwise. In order to delineate trades as aggressive or non-aggressive, I first classify trades on the basis of trade direction (buy or sell) using Lee and Ready's (1991) algorithm. I then compare the transaction prices with the prevailing best bid (ask) price for sell (buy) transactions. If a transaction price is below (above) or equal to the prevailing best bid (ask) price I classify the sell (buy) transaction as an aggressive trade. α_i and β_t are stock and time fixed effects, $latency_{i,t}$ is the key variable in the model and the information transmission latency between Frankfurt and London. $C_{k,i,t}$ is a set of *k* control variables, which includes the standard deviation of stock returns ($Stddev_{i,t}$) for stock *i* and transaction *t* as a proxy for volatility, the effective spread ($Espread_{i,t}$) for stock *i* and transaction *t* as a proxy for liquidity, the inverse of price ($InvPri_{i,t}$) for stock *i* and transaction *t*, the natural logarithm of trading volume ($lnTV_{i,t}$) for stock *i* and transaction *t*, market depth ($Depth_{i,t}$) for stock *i* and transaction *t*, and momentum ($Momentum_{i,t}$) for stock *i* and transaction *t*. $Stddev_{i,t}$ is calculated as the standard deviation of returns for contemporaneous and previous transactions (transactions *t* and *t-1*) for stock *i*, $Espread_{i,t}$ is measured as twice the absolute value of the difference between the transaction price and the prevailing bid-ask spread for stock *i* and transaction *t*, $InvPri_{i,t}$ is the inverse of the transaction price for stock *i* and transaction *t*, $lnTV_{i,t}$ is the natural logarithm of trading volume for stock *i* and transaction *t*, $Depth_{i,t}$ is the sum of prevailing bid and ask sizes for stock *i* corresponding to transaction *t* and $Momentum_{i,t}$ is the first lag of the stock return for stock *i* and transaction *t* (momentum for transaction *t* is the stock return for transaction *t-1*). The sample consists of 100 most active German stocks that cross-listed in XSE and CBOE. All variables, except latency, are computed for the CBOE. Stocks are classified into quartiles using Euro trading volume. Marginal effects are reported in brackets and they are computed as the mean of marginal effects across stocks. The sample period covers March 2017 to August 2018. Standard errors are robust to heteroscedasticity and autocorrelation and t-statistics are reported in parenthesis. *, ** and *** correspond to statistical significance at the 0.10, 0.05 and 0.01 levels respectively.

	Dependent variable: $\boldsymbol{Aggressiveness_{i,t}}$				
	Full sample	Least active	Quartile 2	Quartile 3	Most active
$latency_{i,t}$	-0.186x10^{-1}***	-0.398x10^{-1}***	-0.276x10^{-1}***	-0.219x10^{-1}***	-0.123x10^{-1}***
	[-0.284x10^{-2}]	[-0.640x10^{-2}]	[-0.427x10^{-2}]	[-0.332x10^{-2}]	[-0.188x10^{-2}]
	(-19.09)	(-11.09)	(-10.10)	(-10.90)	(-9.48)
$Momentum_{i,t}$	0.105	0.651x10^{-1}	-0.378	0.309	0.924
	[0.161x10^{-1}]	[0.105x10^{-1}]	[-0.584x10^{-1}]	[0.469x10^{-1}]	[0.141]
	(0.84)	(0.28)	(-1.23)	(0.56)	(1.26)
$InvPri_{i,t}$	1.691***	0.726***	1.192***	4.165***	3.862***
	[0.259]	[0.117]	[0.184]	[0.632]	[0.588]

	(25.73)	(2.95)	(10.59)	(10.66)	(39.52)
$Espread_{i,t}$	1.196*** [0.183] (35.53)	1.028*** [0.164] (8.83)	1.176*** [0.182] (14.97)	0.786*** [0.119] (19.47)	4.799*** [0.730] (49.25)
$Stddev_{i,t}$	-0.107* [-0.164x10^{-1}] (-1.95)	1.037 [0.166] (1.42)	-0.322 [-0.498x10^{-1}] (-1.03)	-0.147 [-0.224x10^{-1}] (-1.43)	-0.113 [-0.171x10^{-1}] (-1.60)
$lnTV_{i,t}$	-0.617x10^{-1}*** [-0.945x10^{-2}] (-36.44)	-0.625x10^{-1}*** [-0.100x10^{-1}] (-9.90)	-0.977x10^{-1}*** [-0.151x10^{-1}] (-20.24)	-0.849x10^{-1}*** [-0.129x10^{-1}] (-22.02)	-0.548x10^{-1}*** [-0.834x10^{-2}] (-24.38)
$Depth_{i,t}$	-0.605x10^{-4}*** [-0.926x10^{-5}] (-22.66)	-0.463x10^{-4}*** [-0.742x10^{-5}] (-14.56)	-0.371x10^{-4}*** [-0.575x10^{-5}] (-16.28)	-0.167x10^{-4}*** [-0.254x10^{-4}] (-17.55)	-0.827x10^{-4}*** [-0.125x10^{-4}] (-19.22)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes
McFadden R^2	27.2%	31.1%	14.6%	28.2%	25.7%

Table 4.10 reports the estimation results for the logit model. The results are qualitatively similar for the overall sample and quartiles. I also report marginal effects in parentheses, which show an increase in the probability of aggressive trades if the explanatory variable increases by one standard deviation, conditional on all other explanatory variables being at their unconditional means. My results show that the $latency_{i,t}$ coefficient is negative and statistically significant at 0.01 level, which implies that indeed increments (decrements) in latency (speed) decrease the probability of aggressive trades. Based on the marginal effects, traders are 0.3% less (more) likely to aggressively trade subsequent to increasing latency (speed). Overall, I conclude that trader improvements in the speed of order execution ultimately drives increased trading aggressiveness, a conclusion consistent with the aggressiveness channel hypothesis and Roşu (2016). The *McFadden* R^2 for the full sample is 27%, a substantial explanatory level for an estimation based on an intraday estimation frequency.

4.5 Economic implications: the trade-off between higher (lower liquidity/volatility) and lower (higher liquidity/volatility) latency

In Section 4.4, I find that, as argued by various regulators and investors,[41] lower (transmission) latency between financial markets leads to better liquidity and higher volatility. In the market microstructure literature, liquidity and volatility are considered to be two important market quality metrics (see as examples, Hendershott et al., 2011; Malceniece et al., 2018). Specifically, higher liquidity is perceived as good whereas higher volatility might be perceived as less beneficial. Thus, my main empirical finding, i.e. lower latency improves liquidity and increases volatility, is unable to show whether speed is beneficial or harmful for financial markets overall; more explicitly, my analysis does not allow me to show the (net)

[41] https://www.reuters.com/article/us-highfrequency-microwave/lasers-microwave-deployed-in-high-speed-trading-arms-race-idUSBRE9400L920130501

economic implication of latency. Nevertheless, my analysis suggests that there is a trade-off, or at least an inflection point at which the liquidity enhancing benefits of speed are offset by its volatility increasing effects. Therefore, in this section, I examine the relative impact of liquidity, volatility, and latency on expected return by interacting liquidity/volatility with latency. This approach allows me to attempt an estimation of the economic implication of latency, and to investigate the trade-off between higher (lower liquidity/volatility) and lower latency (higher liquidity/volatility). Specifically, I investigate the impacts of volatility and liquidity on expected return during regular trading periods and higher latency periods, and then compare them. I employ expected return as a key speed-impacting variable for two reasons. Firstly, to an investor, expected return serves as an indicator of profits relative to risk; hence it holds significant economic implications. Secondly, making valid a comparison between high and low latency in this study requires that I employ a variable impacted by both liquidity and volatility. The literature shows that, indeed, expected return is a direct measure satisfying this criterion. For example, Holmström and Jean (2001) and Acharya and Pedersen (2005) propose asset pricing models in which expected return is positively correlated with liquidity risk, and Pástor and Stambaugh (2003) empirically test this relationship and find that indeed, expected stock returns are positively related to fluctuations in aggregate liquidity. Poterba and Summers (1986) explain the theoretical (positive) relationship between expected return and volatility, and French et al. (1987) empirically show the positive relationship between expected return and volatility (see also Pindyck, 1984). In addition to the well-established literature about the relationship between liquidity/volatility and expected return, Malceniece et al. (2018) and Brogaard et al. (2014b) show the potential relationship between latency and the cost of capital/market efficiency, i.e. the efficiency of capital allocation. The overwhelming view in the literature is therefore that expected return is impacted by volatility, liquidity, and latency.

Developing a framework estimating the marginal impacts of latency-interacted liquidity and volatility proxies is thus a valid approach. My framework includes the following specification:

$$ER_{i,t} = \alpha_i + \beta_t + \beta_1 Stddev_{i,t} + \beta_2 Espread_{i,t} + \beta_3 latency_{i,t} + \beta_4 Stddev_{i,t} * D_{latency,i,t} + \beta_5 Espread_{i,t} * D_{latency,i,t} + \sum_{k=1}^{3} \delta_k C_{k,i,t} + \varepsilon_{i,t} \quad (4.6)$$

where $ER_{i,t}$ is the expected return for stock *i* at interval *t* and computed as the mean of returns for the previous 60 transaction intervals. $D_{latency,i,t}$ is a dummy equalling one during periods of high (low) latency (speed); a transaction interval is designated as a high (low) latency (speed) transaction interval if the latency for that interval is one standard deviation higher than the mean for surrounding -60, +60 corresponding transaction intervals.[42] α_i and β_t are stock and time fixed effects, and $latency_{i,t}$ is the *TL* between XSE and CBOE. $C_{k,i,t}$ is a set of previously defined *k* control variables, which includes $Depth_{i,t}$, $InvPri_{i,t}$ and $lnTV_{i,t}$.

I run two variants of Equation (4.6). In the first specification, I run the model with only two explanatory variables, $Stddev_{i,t}$ and $Espread_{i,t}$. In this analysis, my main aim is to show if expected return is impacted by both volatility and liquidity, as suggested by the literature (see Acharya and Pedersen, 2005; Poterba and Summers, 1986). This analysis is particularly important since my motivation for using expected return as a main variable is based on the suggestion in the literature of a relationship between liquidity/volatility and expected return. In the second specification, I run the complete model as specified in Equation (4.6). As noted, I aim to examine the relative impact of liquidity and volatility on expected return, and therefore in the second specification, I standardize all variables to compare the size of coefficients on a comparable scale.[43]

[42] For robustness, I designate the transaction interval as an interval of high (low) latency (speed) transaction interval if latency for that interval is two standard deviation higher than the mean for the surrounding -60, +60 corresponding transaction intervals. The results obtained are qualitatively similar to the reported results.

[43] For robustness, I compute standardize coefficients based on un-standardized variables within the regression model as well. The results obtained are qualitatively similar with the ones I present in the chapter.

Table 4. 11 Expected return and trade-off between higher (lower liquidity/volatility) and lower latency (higher liquidity/volatility)

This table reports the coefficient estimates of two specifications of the following regression model:

$$ER_{i,t} = \alpha_i + \beta_t + \gamma_1 Stddev_{i,t} + \gamma_2 Espread_{i,t} + \gamma_3 Stddev_{i,t} * D_{latency,i,t} + \gamma_4 Espread_{i,t} * D_{latency,i,t} + \gamma_5 latency_{i,t} + \sum_{k=1}^{3} \delta_k C_{k,i,t} + \varepsilon_{i,t}$$

where $ER_{i,t}$ is the expected return for stock *i* and transaction *t*, α_i and β_t are stock and time fixed effects, $Stddev_{i,t}$ is the standard deviation of returns for stock *i* and transaction *t*, $Espread_{i,t}$ is effective spread for stock *i* and transaction *t*, $D_{latency,i,t}$ is a dummy equalling one during periods of high (low) latency (speed) for stock *i*, $latency_{i,t}$ is the information transmission latency between Frankfurt and London, and $C_{k,i,t}$ is a set of *k* control variables, which includes the market depth ($Depth_{i,t}$) for stock *i* and transaction, the inverse of price ($InvPri_{i,t}$) for stock *i* and transaction *t*, and the natural logarithm of trading volume ($lnTV_{i,t}$) for stock *i* and transaction *t*. $ER_{i,t}$ is computed as the mean of the previous 60 transaction intervals (t) returns for stock i, $Stddev_{i,t}$ is calculated as the standard deviation of returns for the contemporaneous and previous transactions (transactions *t* and *t-1*) for stock *i*, $Espread_{i,t}$ is measured as twice the absolute value of the difference between the transaction price and the prevailing bid-ask spread for stock *i* and transaction *t*, $D_{latency,i,t}$ is designated as an interval of high (low) latency (speed) interval (t) if latency for that interval is one standard deviation higher than the mean for the surrounding -60, +60 corresponding transaction intervals for stock *i*, $Depth_{i,t}$ is the sum of prevailing bid and ask sizes for stock *i* corresponding to transaction *t*, $InvPri_{i,t}$ is the inverse of the price for stock *i* and transaction *t*, and $lnTV_{i,t}$ is the natural logarithm of trading volume for stock *i* and transaction *t*. In Panel A reports a parsimous model estimation, while Panel B reports results for the full model estimation. The sample consists of the 100 most active German stocks that cross-listed on XSE and CBOE. All variables, except $latency_{i,t}$, are computed for the CBOE. The sample period covers March 2017 to August 2018. Standard errors are robust to heteroscedasticity and autocorrelation and t-statistics are reported in parentheses. *, ** and *** correspond to statistical significance at the 0.10, 0.05 and 0.01 levels respectively.

Panel A

	Dependent variable: $\boldsymbol{ER_{i,t}}$
	Full sample
$Stddev_{i,t}$	0.343*** (22.46)
$Espread_{i,t}$	0.659x10^{-2}*** (2.60)
Stock fixed effects	Yes
Time fixed effects	Yes
$\overline{R^2}$	22.8%

Panel B

Dependent variable: $\boldsymbol{ER_{i,t}}$

	Full sample	Least active	Quartile 2	Quartile 3	Most active
$Stddev_{i,t}$	$0.655\text{x}10^{-1}$***	0.211***	$-0.666\text{x}10^{-2}$***	$0.614\text{x}10^{-1}$***	$-0.246\text{x}10^{-1}$***
	(38.89)	(63.39)	(-2.14)	(21.80)	(-5.12)
$Espread_{i,t}$	$-0.137\text{x}10^{-2}$	$-0.103\text{x}10^{-2}$	$-0.239\text{x}10^{-2}$	$0.105\text{x}10^{-2}$	$-0.293\text{x}10^{-2}$
	(-1.54)	(-0.52)	(-1.42)	(0.63)	(-1.55)
$Stddev_{i,t} * D_{latency,i,t}$	$-0.419\text{x}10^{-2}$***	$-0.139\text{x}10^{-1}$***	$0.137\text{x}10^{-1}$***	$-0.147\text{x}10^{-1}$***	$-0.347\text{x}10^{-2}$**
	(-6.18)	(-10.45)	(10.57)	(-11.48)	(-2.27)
$Espread_{i,t} * D_{latency,i,t}$	$0.420\text{x}10^{-2}$***	$0.733\text{x}10^{-2}$***	$-0.876\text{x}10^{-3}$	$0.471\text{x}10^{-2}$***	$0.584\text{x}10^{-2}$***
	(5.61)	(4.58)	(-0.62)	(3.29)	(3.67)
$latency_{i,t}$	$-0.151\text{x}10^{-2}$	$-0.324\text{x}10^{-3}$	$-0.112\text{x}10^{-1}$***	$0.534\text{x}10^{-2}$	$0.266\text{x}10^{-3}$
	(-0.88)	(-0.10)	(-3.37)	(1.62)	(0.07)
$Depth_{i,t}$	$-0.132\text{x}10^{-2}$	$-0.467\text{x}10^{-2}$**	$0.291\text{x}10^{-2}$	$-0.454\text{x}10^{-2}$**	$0.378\text{x}10^{-3}$
	(-1.38)	(-2.37)	(1.61)	(-2.45)	(0.18)
$InvPri_{i,t}$	-4.151***	-2.051***	-4.297***	-7.005***	-3.754***
	(-50.46)	(-12.78)	(-25.15)	(-42.44)	(-22.79)
$lnTV_{i,t}$	$0.684\text{x}10^{-2}$***	$0.164\text{x}10^{-1}$***	$0.150\text{x}10^{-1}$***	$-0.230\text{x}10^{-2}$	$0.224\text{x}10^{-3}$***
	(3.21)	(3.72)	(3.59)	(-0.56)	(0.05)
Stock fixed effects	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes
$\overline{\boldsymbol{R}^2}$	42.3%	50.1%	40.9%	38.1%	40.1%

Table 4.11 reports the estimation results for Equation (4.6). Panel A reports the coefficient estimations with two explanatory variables, i.e. proxies for volatility ($Stddev_{i,t}$) and liquidity ($Espread_{i,t}$). The results show that both $Stddev_{i,t}$ and $Espread_{i,t}$ are individually positively and significantly related with expected return. This result is consistent with predictions of the theoretical models developed by Acharya and Pedersen (2005) and Poterba and Summers (1986). The estimates show that volatility and liquidity risk are indeed priced, and therefore higher volatility and lower liquidity leads to higher expected return (see French et al., 1987; Pástor and Stambaugh, 2003 for empirical consistency). Panel B shows the estimation results for the complete form of Equation (4.6). The coefficient of $Stddev_{i,t}$ is positive and statistically significant; $Stddev_{i,t}$ is associated with a 0.065 standard deviation increment in expected return. It implies that volatility is priced and consistent with results reported in Panel A; higher volatility is linked to higher expected return. Although the coefficient for the volatility proxy is statistically significant, $Espread_{i,t}$, the liquidity proxy, is not statistically significant, which is inconsistent with the results reported in Panel A; I believe this linked to the addition of the latency interacted $Espread_{i,t}$ in the specification.[44] Notwithstanding, the main focus for this estimation are the interaction variables' coefficients. I observe that when $Stddev_{i,t}$ and $Espread_{i,t}$ are interacted with the latency dummy ($D_{latency,i,t}$), both variables become highly statistically significant. $Stddev_{i,t} * D_{latency,i,t}$ is negatively related with expected return, which implies that, while on average volatility leads to higher expected return (see the coefficient estimates of $Stddev_{i,t}$, 0.065), increased latency has an ameliorating effect on volatility, leading to reduced compensation since the risk presented by volatility reduces. $Espread_{i,t} * D_{latency,i,t}$ is positively related with expected

[44] To support this argument, i.e. insignificance of $Espread_{i,t}$ in Panel B is sourced by adding the interaction coefficient, I run the model without $Espread_{i,t} * D_{latency,i,t}$ and find that indeed $Espread_{i,t}$ is statistically significant in this case. This shows that the positive relationship between illiquidity and expected return presented in Panel A, is eliminated because of the added interacted variable. For parsimony, I do not show this result, however they are available upon request.

return which shows that illiquidity leads to higher expected return when latency is high. This is in line with the expectation that the widening of the spread implies an increase in adverse selection, which needs to be priced. Furthermore, the fact that once the interaction variable $Espread_{i,t} * D_{latency,i,t}$ is added to the model the significance of $Espread_{i,t}$'s coefficient disappears, while $Espread_{i,t} * D_{latency,i,t}$'s coefficient becomes statistically significant, implies that latency is a pivotal determinant of the relationship between liquidity and expected return. It thus appears that when liquidity is impaired, I would expect to see a higher level of adverse selection risk, which leads to investors demanding higher returns as compensation for the adverse selection-induced larger spread they are forced to trade with. The results reported in Panel B have several important implications. Firstly, transmission latency – the combination of traders' execution latency, exchange latency, and connection latency – is one of the most important determinants of the relationship between volatility/liquidity and expected return. Therefore, it plays a vital role in today's financial markets and the economy. This insight is consistent with recent empirical findings in the literature, for example, the literature on the potential relationship between HFT and the cost of capital (see as an example, Malceniece et al., 2018), and the economic importance of market fragmentation in the efficiency of modern financial markets (see as an example, O'Hara and Ye, 2011). Secondly, the magnitude (absolute value) of $Stddev_{i,t}$'s coefficient at 0.065 is about 48 times higher than the magnitude of $Espread_{i,t}$'s coefficient at 0.0014. This implies that the risk presented by volatility is the more important driver of expected return; this result is further underscored by the lack of statistical significance of $Espread_{i,t}$'s coefficient in Panel B. However, when both proxies are interacted with $D_{latency,i,t}$, the magnitude of the impact of interacted liquidity on expected return, 0.00420, is slightly higher than the magnitude (absolute value) of the impact of interacted volatility on expected return, 0.00419, demonstrating the unmistakable effect of latency on expected return. Thus, investors may view the risk of trading in slow markets as being as high as the risk of trading in markets where price volatility is driven by increased speed, perhaps even seeing the former risk as being higher than the latter. The implication of this finding is that the net effect of low latency is the enhancement of market quality. While latency influences the effects of both liquidity and volatility on expected return, the effect is more defining and stronger for liquidity. This finding is consistent with that of Aït- Sahalia and Saglam (2013), who show that the speed advantage of HFTs improves the welfare of all traders, i.e. both HFTs and low frequency traders, in financial markets, and hence the benefits of high speed trumps its risks. The $\overline{R^2}$ for the full sample is 42%, which shows that my model explains a substantial part of the variation in expected return at the intraday level. For comparison, return predictability models typically explain single percentage digits (see as examples, Chordia et al., 2008; Rzayev and Ibikunle, 2019).

4.6 Conclusion

In this study, I examine the role of latency on market quality by focusing on liquidity and volatility proxies; my findings are four-fold.

By estimating latency between Frankfurt and London from transaction-level data, I provide empirical evidence that prices in London respond to price changes in Frankfurt within 3-5ms. This result is consistent with the latencies claimed by the providers of microwave and fiber optic connections between London and Frankfurt, and thus demonstrates the empirical relevance of my information transmission latency estimation method.

Secondly, I report that decreases in the information transmission latency between the home and satellite markets increases liquidity and volatility in the satellite market; the results are robust to alternative liquidity and volatility proxies and more importantly, economically meaningful. In order to address potential endogeneity concerns I employ a difference-in-difference framework and test the role of technological upgrades in the home market on the liquidity and volatility in the satellite market, by examining cross-listed stocks. I find that, indeed, liquidity and price volatility in the satellite market increases significantly more for stocks directly impacted by the technological innovations in the home market. This allows me to establish a causal relationship between speed on the one hand and liquidity and volatility on the other, thus ruling out endogeneity concerns.

Thirdly, I examine the potential channels through which latency impacts liquidity and volatility. I provide empirical evidence consistent with the predictions of theoretical market microstructure models, suggesting that fast traders use their speed to avoid adverse selection risk/cost, a component of the bid-ask spread. This ability to avoid adverse selection risk leads to a narrowing of the spread/increased liquidity, which in turn reduces the price impact of trades. Faced with lower adverse selection risk, fast traders are more likely to trade even more readily, leading to increased aggressive trading and higher price volatility.

The positive effect of speed on market quality through the enhancement of liquidity and its adverse effect on market quality through its increasing of volatility implies a trade-off between speed's positive and negative effects. Therefore, I investigate the relative impact of liquidity, volatility, and latency on expected return; the latter is driven by the other three. I show that latency is an important determinant for the relationship between volatility/liquidity and expected return, and more importantly, while high latency can improve market quality by reducing volatility, its liquidity deterioration effect dominates its volatility reduction effect. This implies that the net effect of low latency is the enhancement of market quality.

5. Summary

5.1 Summary of findings

5.1.1 Order aggressiveness and flash crashes

In Chapter 2, I investigate the contribution of aggressive orders to flash crashes by developing a new framework; the framework is the extension of the theoretical approach described in Menkveld (2013). This chapter contributes to the literature on flash crashes by drawing the link between order aggressiveness and flash crashes, making no assumptions regarding liquidity constraints in the market, and explaining the economic motivation of aggressive trading during flash crashes. The framework is a three-stage trading strategy. At the first stage, a hypothetical trader submits an (excessive) aggressive sell order, leading to stock prices going down. Thereafter, the trader submits an aggressive buy limit order and thereby, generates increasing pressure on prices. During the last stage, the trader sells his/her securities and leaves the market. I show that through the adoption of the noted trading strategy, the hypothetical aggressive trader can obtain higher profit under a few necessary conditions, i.e. when the trader can avoid being adversely selected. The theoretical framework raises three arguments: (1) contemporaneous aggressive orders contribute to flash crashes, (2) the build-up of order aggressiveness is inextricably linked to flash crashes, and (3) aggressive orders are more profitable during flash crashes. Thereafter, I use ultra-high frequency data from 53 S&P 500 stocks, affected by the May 6 2010 flash crash, to test the arguments motivated by the framework.

This chapter reveals three major findings motivated by the predictions of the framework. First, I show that there is an excessive order aggressiveness at the sell side prior to and during the first half of the May 6 2010 flash crash, and thereafter, the buy side became more aggressive. Second, I find that the build-up of order aggressiveness may contribute to the onset of flash crashes; the number of aggressive orders prior to the flash crash is positively and significantly related to the flash crash. Third, I find that traders behave more aggressively during flash crashes, and the economic motivation of this excess aggressiveness is the profitability of aggressive orders during these periods. Specifically, I estimate that for the stocks in my sample, an informed investor during the flash crash could achieve an additional 1,482 bps return on his portfolio.

5.1.2 A state space modelling of the information content of trading volume

In Chapter 3, I propose the state space modelling approach to decompose high frequency trading volume into informed and uninformed parts. This chapter contributes to the literature on the information content of the trading volume, by proposing a more efficient way to decompose the trading volume into liquidity- and information-driven components and by examining the role of informed trading in market toxicity, and eliminating arbitrage opportunities.

This chapter acknowledges three major findings. First, the state space approach is an empirically relevant and more efficient method to decompose high frequency trading volume into its components, i.e. uninformed and informed components. Second, I find that informed trading reduces volatility, illiquidity and toxicity in financial markets. Third, my findings suggest that informed HFTs eliminate arbitrage opportunities and reduce the return predictability window.

This chapter has important implications for financial markets. Specifically, by using this approach, stock exchanges may further understand the evolution of high frequency trading volume/information in financial markets, as it allows for the direct estimation of both motivations of trading from transactions level data.

5.1.3 Need for Speed? International transmission latency, liquidity and volatility

In Chapter 4, I examine the role of latency on market quality by developing a new proxy for latency, i.e. information transmission latency (TL), in fragmented financial markets. This chapter contributes to the literature on the HFT by being the first to empirically estimate TL between the two biggest European financial centres, Frankfurt and London, and analysing the combined role of the traders' execution latency, exchange latency and connection latency (microwave or fibre connections) between exchanges on the liquidity and volatility of financial markets. This implies that the latency measure I use, TL, is more relevant when measuring the impact of speed on market quality in a fragmented trading space; thus, it has further economic insights. Furthermore, the study proposes a new method to investigate the net economic impact of high speed on financial markets.

This chapter reveals four major findings. First, my findings suggest that the constructed latency metric is empirically relevant, and the information transmission latency between the two biggest European financial markets, Frankfurt and London, is 3–5ms. Second, I show that higher transmission speed leads to higher liquidity and volatility and that these relationships are causal. For this, I employ a difference-in-difference analysis to show the causality. Third, I provide empirical evidence that the channels – *adverse selection avoidance* and *aggressiveness* – proposed by various theoretical models can explain the study's finding of the relationship between speed and market quality. Finally, I show that transmission latency between financial markets is an important determinant of the relationship between volatility/liquidity and the expected returns, and more importantly, liquidity deterioration impact of high latency dominates its volatility reducing effect. It thus implies that the net effect of high transmission speed is the enhancement of market quality.

5.2 Limitations and future research suggestions

Six main financial markets microstructure issues (flash crashes, trading volume, liquidity, volatility, price efficiency and high-frequency trading) are investigated in this thesis. My studies deliver the view that the evolution of market structure is not necessarily beneficial/harmful for financial markets. Although the studies in this thesis investigate research questions in detail, nevertheless there are some limitations in it mainly because of data availability.

Chapter 2 provides an important insights into the links between aggressive orders and flash crashes. In this study the aggressive orders and the profit related to these orders are estimated from order-level data. While the empirical methods to estimate the number of aggressive orders and profits of aggressive traders are well-established and widely accepted in the literature, the data employed in this thesis does not have the identifier of traders and thus, does not allow me to directly compute the number of aggressive orders and profit of aggressive traders for HFTs. Future research can more accurately investigate the predictions of the framework described in Chapter 2 employing the data with traders' identifier. Furthermore, as noted in Chapter 2, while my sample selection criteria is motivated by the literature, it may lead to sample selection bias.

Chapter 3 proposes new model, state-space modelling approach, to decompose trading volume into liquidity- and information-driven components. I did not include the impact of liquidity shocks to the model because of two reasons. Firstly, I aimed to keep model as tractable as possible. Secondly, examining the liquidity shocks is outside of the scope of this study. Future research can further decompose liquidity-motivated trading volume into general liquidity and liquidity shocks components.

Another limitation can be found in Chapter 4. In this chapter I estimate the information transmission speed between Frankfurt and London by using TRTH data. TRTH provides exchange timestamps of transactions up to millisecond frequency and therefore, this study is not able to estimate the latency beyond millisecond basis. It is clear that some information can be lost in the method described in this chapter as a result. The use of microseconds rather than milliseconds is therefore ideal for limiting bias. More empirical studies can estimate more accurate latency between financial markets by using data with microsecond frequency.

Appendices

APPENDIX 2.A. Framework variables and definitions

VARIABLE	DEFINITION
P_t^{a}	Ask Price at time *t*.
P_t^{b}	Bid Price at time *t*.
P_t^{mp}	Mid-Price at time *t*.
$P_t^{ag.a}$	Ask Price at time *t* under aggressive trading strategy.
$P_t^{ag.b}$	Bid Price at time *t* under aggressive trading strategy.
$P_t^{ag.mp}$	Mid-Price at time *t* under aggressive trading strategy.
$P_t^{mm.a}$	Ask Price at time *t* under market-making strategy.
$P_t^{mm.b}$	Bid Price at time *t* under market-making strategy.
$P_t^{mm.mp}$	Mid-Price at time *t* under market-making strategy.
π_t^{ag}	Profit at time *t* under aggressive trading strategy.
$\pi_t^{ag.ba}$	Profit from bid-ask spread at time *t* under aggressive trading strategy.
$\pi_t^{ag.p}$	Profit from position at time *t* under aggressive trading strategy.
U_t^{st}	Attractiveness or fitness function of each type of strategy at time *t*.
η	Memory parameter.
ξ	Intensity of switching parameter.
φ_t^{ag}	Relative weight of aggressive trading strategy at time *t*.

APPENDIX 2.B. List of the sample stocks

ISIN CODE	RIC	Security name
US0378331005	AAPL.OQ	Apple Inc.
US03073E1055	ABC.N	AmerisourceBergen Corp.
IE00B4BNMY34	ACN.N	Accenture plc
US0530151036	ADP.OQ	Automatic Data Processing Inc.
US0236081024	AEE.N	Ameren Corp.
US0015471081	AKS.N	AK Steel Holding Corp.
US0200021014	ALL.N	Allstate Corp.
US0231351067	AMZN.OQ	Amazon.com Inc.
US0325111070	APC.N	Anadarko Petroleum Corp.
US1101221083	BMY.N	Bristol-Myers Squibb Co.
US0846707026	BRKb.N	Berkshire Hathaway Inc.
US2058871029	CAG.N	ConAgra Brands Inc.
US1491231015	CAT.N	Caterpillar Inc.
US1651671075	CHK.N	Chesapeake Energy Corp.
US1567001060	CTL.N	CenturyLink Inc.
US1667641005	CVX.N	Chevron Corp.
US2635341090	DD.N	E I du Pont de Nemours and Co.
US2479162081	DNR.N	Denbury Resources
US2605431038	DOW.N	Dow Chemical Co.
US2786421030	EBAY.OQ	eBay Inc.
US2686481027	EMC.N	EMC Corp.
US30219G1085	ESRX.OQ	Express Scripts Holding Co.
US2971781057	ESS.N	Essex Property Trust Inc.
US3453708600	F.N	Ford Motor Co.
US3696041033	GE.N	General Electric Co.
US38259P7069	GOOG.OQ	Alphabet Inc. (Google Inc. Class C)
US4370761029	HD.N	Home Depot Inc.
US4282361033	HPQ.N	Hewlett-Packard Inc.
US4592001014	IBM.N	International Business Machines Corp.
US4581401001	INTC.OQ	Intel Corp.
US9255501051	JDSU.OQ	JDS Uniphase Corp.
US4781601046	JNJ.N	Johnson & Johnson
US1912161007	KO.N	The Coca Cola Co.
US5260571048	LEN.N	Lennar Corp.
US58155Q1031	MCK.N	McKesson Corp.
IE00BTN1Y115	MDT.N	Medtronic Plc.
US88579Y1010	MMM.N	3M Co.
US02209S1033	MO.N	Altria Group Inc.
US5949181045	MSFT.OQ	Microsoft Corp.
US68389X1054	ORCL.OQ	Oracle Corp.
US7134481081	PEP.N	PepsiCo Inc.
US7170811035	PFE.N	Pfizer Inc.
US7427181091	PG.N	Procter & Gamble Co.
US7181721090	PM.N	Philip Morris International Inc.
US7132911022	POM.N	Pepco Holdings Inc.
US8454671095	SWN.N	Southwestern Energy Co.

US8835561023	TMO.N	Thermo Fisher Scientific Inc.
US8825081040	TXN.N	Texas Instruments Inc.
US91324P1021	UNH.N	United Health Group Inc.
US9497461015	WFC.N	Wells Fargo & Co.
US9311421039	WMT.N	Wal-Mart Stores Inc.
US30231G1022	XOM.N	Exxon Mobil Corp.
US9884981013	YUM.N	Yum! Brands Inc.

References

Acharya VV, Pedersen LH. Asset pricing with liquidity risk. Journal of Financial Economics 2005;77; 375-410

Admati AR, Pfleiderer P. A Theory of Intraday Patterns: Volume and Price Variability. The Review of Financial Studies 1988;1; 3-40

Aït- Sahalia Y, Saglam M. High Frequency Traders: Taking Advantage of Speed. 2013;

Aitken M, Frino A. Execution costs associated with institutional trades on the Australian Stock Exchange. Pacific-Basin Finance Journal 1996;4; 45-58

Akerlof GA. The Market for "Lemons": Quality Uncertainty and the Market Mechanism. The Quarterly Journal of Economics 1970;84; 488-500

Aldridge I. High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems Wiley; 2010;

Amihud Y. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 2002;5; 31-56

Andersen T, Bondarenko O. VPIN and the flash crash. Journal of Financial Markets 2014;17; 1-46

Avramov D, Chordia T, Goyal A. The Impact of Trades on Daily Volatility. The Review of Financial Studies 2006;19; 1241-1277

Barber BM, Lee Y, Liu Y, Odean T. Just How Much Do Individual Investors Lose by Trading? The Review of Financial Studies 2009;22; 609-632

Barclay MJ, Hendershott T. Price Discovery and Trading After Hours. The Review of Financial Studies 2003;16; 1041-1073

Barclay MJ, Hendershott T. A comparison of trading and non-trading mechanisms for price discovery. Journal of Empirical Finance 2008;15; 839-849

Baron M, Brogaard J, Hagströmer B, Kirilenko AA. Risk and Return in High-Frequency Trading. Journal of Financial and Quantitative Analysis 2018;Forthcoming;

Battalio RH. Third Market Broker-Dealers: Cost Competitors or Cream Skimmers? The Journal of Finance 2012;52; 341-352

Bennett P, Wei L. Market structure, fragmentation, and market quality. Journal of Financial Markets 2006;9; 49-78

Biais B, Foucault T, Moinas S. Equilibrium fast trading. Journal of Financial Economics 2015;116; 292-313

Biais B, Hillion P, Spatt C. An empirical analysis of the limit-order book and order flow in the Paris Bourse. The Journal of Finance 1995;50; 1655-1689

Biais B, Hillion P, Spatt C. Price Discovery and Learning during the Preopening Period in the Paris Bourse. Journal of Political Economy 1999;107; 1218-1248

Boehmer E, Fong KYL, Wu J. International evidence on algorithmic trading. American Finance Association (AFA) Annual Meeting. San Diego; 2015.

Bollen NPB, Smith T, Whaley RE. Modeling the bid/askspread: measuring the inventory-holding premium. Journal of Financial Economics 2004;72; 97-141

Borkovec M, Domowitz I, Serbin V, Yegerman H. Liquidity and Price Discovery in Exchange-Traded Funds: One of Several Possible Lessons from the Flash Crash. The Journal of Index Investing 2010;1; 24-42

Boulton TJ, Braga-Alves MV. The skinny on the 2008 naked short-sale restrictions. Journal of Financial Markets 2010;13; 397-421

Branch B, Freed W. Bid-Asked Spreads on the AMEX and the Big Board. The Journal of Finance 1977;32; 159-163

Brogaard J, Carrion A, Riordan R, Shkilko A, Sokolov K. High Frequency Trading and Extreme Price Movements. Journal of Financial Economics 2014a;128; 253-265

Brogaard J, Hendershott T, Riordan R. High-Frequency Trading and Price Discovery. The Review of Financial Studies 2014b;27; 2267-2306

Brunnermeier MK. Asset Pricing Under Asymmetric Information: Bubbles, Crashes, Technical Analysis, and Herding. Oxford University Press, Oxford; 2001;

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

Buti S, Rindi B, Werner IM. Diving into Dark Pools. 2011;

Caminschi A, Heaney R. Fixing a leaky fixing: short-term market reactions to the London pm gold price fixing. Journal of Futures Markets 2014;34; 1003-1039

Chaboud AP, Chiquoine B, Hjalmarsson E, Vega C. Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market. The Journal of Finance 2014;69; 2045-2084

Chakrabarty B, Moulton P, Shkilko A. Evaluating Trade Classification Algorithms: Bulk Volume Classification versus the Tick Rule and the Lee-Ready Algorithm. Journal of Financial Markets 2015;25; 52-79

Chordia T, Roll R, Subrahmanyam A. Market Liquidity and Trading Activity. The Journal of Finance 2001;56; 501-530

Chordia T, Roll R, Subrahmanyam A. Order imbalance, liquidity, and market returns. Journal of Financial Economics 2002;65; 111-130

Chordia T, Roll R, Subrahmanyam A. Evidence on the speed of convergence to market efficiency. Journal of Financial Economics 2005;76; 271-292

Chordia T, Roll R, Subrahmanyam A. Liquidity and market efficieny. Journal of Financial Economics 2008;87; 249-268

Chung D, Hrazdil K. Liquidity and market efficiency: A large sample study. Journal of Banking & Finance 2010;34; 2346-2357

Clark PK. A subordinated stochastic process model with finite variance for speculative prices. Econometrica 1973;41; 135-155

Collin-Dufresne P, Fos V. Do Prices Reveal the Presence of Informed Trading? The Journal of Finance 2015;70; 1555-1582

Collin-Dufresne P, Fos V. Insider Trading, Stochastic Liquidity, and Equilibrium prices. Econometrica 2016;84; 1441-1475

Comerton-Forde C, Putniņš TJ. Dark trading and price discovery. Journal of Financial Economics 2015;118; 70-92

Copeland TE. A Model of Asset Trading Under the Assumption of Sequential Information Arrival. The Journal of Finance 1976;31; 1149-1168

Cornell B. The Relationship between Volume and Price Variability in Futures Markets. Journal of Futures Markets 1981;1; 303-316

Daigler RT, Wiley MK. The Impact of Trader Type on the Futures Volatility-Volume Relation. The Journal of Finance 1999;54; 2297-2316

David SH, Peter RH. Behavioural travel modelling. Croom Helm, London; 1979;

Degryse H, de Jong F, van Kervel V. Does order splitting signal uninformed order flow? 2013;

Degryse H, Jong F, Ravenswaaij M, Wuyts G. Aggressive Orders and the Resiliency of a Limit Order Market. Review of Finance 2005;9; 201-242

Dufour A, Engle RF. Time and the impact of a trade. The Journal of Finance 2000;55; 2467–2498

Durbin J, Koopman S. Time Series Analysis by State Space Models. Oxford University Press, Oxford, UK; 2012;

Easley D, De Prado M, O'Hara M. The microstructure of the "flash crash": flow toxicity, liquidity crashes, and the probability of informed trading. Journal of Portfolio Management 2011;37; 118-129

Easley D, De Prado M, O'Hara M. Flow Toxicity and Liquidity in a High-frequency World. The Review of Financial Studies 2012;25; 1457-1493

Easley D, Kiefer N, O'Hara M, Paperman J. Liquidity, Information, and Infrequently Traded Stocks. The Journal of Finance 1996;51; 1405-1436

Easley D, Kiefer NM, O'Hara M. One Day in the Life of a Very Common Stock. The Review of Financial Studies 1997;10; 805-835

Easley D, O'Hara M. Price, trade size, and information in securities markets. Journal of Financial Economics 1987;19; 69-90

Ederington LH, Lee JH. The Short-Run Dynamics of the Price Adjustment to New Information. the Journal of Financial and Quantitative Analysis 1995;30; 117-134

Engle RF. The econometrics of ultra-high frequency data. Econometrica 2000;68; 1–22

Engle RF, Russell JR. Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. Econometrica 1998;66; 1127-1162

Epps TW, Epps ML. The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture-of-Distributions Hypothesis. Econometrica 1976;44; 305-321

Fama E. Efficient capital markets: a review of theory and empirical work. The Journal of Finance 1970;25; 383-417

Foucault T. Order flow composition and trading costs in a dynamic limit order market. Journal of Financial Markets 1999;2; 99-134

Foucault T, Hombert J, Rosu I. News trading and speed. The Journal of Finance 2016;71; 335-382

Foucault T, Kozhan R, Tham WW. Toxic arbitrage. Review of Financial Studies 2017;30; 1053-1094

French KR, Schwert GW, Stambaugh RF. Expected Stock Returns and Volatility. Journal of Financial Economics 1987;19; 3-29

Frino A, Ibikunle G, Mollica V, Steffen T. The impact of commodity benchmark on derivatives markets: The case of the dated Brent assessment and Brent futures. Journal of Banking & Finance 2017; 1-17

Glosten L, Milgrom P. Bid, ask, and transaction prices in a specialist market with heterogeneously informed agents. Journal of Financial Economics 1985;14; 71-100

Golub A, Keane J, Poon S. High Frequency Trading and Mini Flash Crashes. 2012;

Gonzalo J, Granger CWJ. Estimation of Common Long-Memory Components in Cointegrated Systems. Journal of Business & Economic Statistics 1995;13; 27-35

Grammig J, Melvin M, Schlag C. Internationally cross-listed stock prices during overlapping trading hours: price discovery and exchange rate effects". Journal of Empirical Finance 2005;12; 139-164

Griffiths M, Smith B, Turnbull A, White R. The costs and determinants of order aggressiveness. Journal of Financial Economics 2000;56; 65-88

Hagströmer B, Nordén L. The diversity of high-frequency traders. Journal of Financial Markets 2013;16; 741-770

Hagströmer B, Nordén L, Zhang D. How Aggressive Are High-Frequency Traders? The Financial Review 2014;49; 395-419

Harris L. Cross-security Tests of the Mixture of Distributions Hypothesis. Journal of Financial and Quantitative Analysis 1986;21; 39-46

Harris L. Transactions Data Tests of the Mixture of Distributions Hypothesis. Journal of Financial and Quantitative Analysis 1987;22; 127-141

Hasbrouck J. Measuring the Information Content of Stock Trades. The Journal of Finance 1991;46; 179-207

Hasbrouck J. One security, many markets: Determining the contributions to price discovery. The Journal of Finance 1995;50; 1175-1199

Hellwig MF. On the Aggregation of Information in Competitive Markets. Journal of Economic Theory 1980;22; 477-498

Hendershott T, Jones CM, Menkveld AJ. Does Algorithmic Trading Improve Liquidity? The Journal of Finance 2011;66; 1-33

Hendershott T, Menkveld AJ. Price pressures. Journal of Financial Economics 2014;114; 405-423

Hendershott T, Moulton PC. Automation, speed, and stock market quality: The NYSE's hybrid. Journal of Financial Markets 2011;14; 568-604

Hiemstra C, Jones JD. Testing for Linear and Nonlinear Granger Causality in the Stock Price-Volume Relation. The Journal of Finance 1994;49; 1639-1664

Hirschey N. Do High-Frequency Traders Anticipate Buying and Selling Pressure? 2017;

Hoffmann P. A dynamic limit order market with fast and slow traders. Journal of Financial Economics 2014;112; 156-170

Holmström B, Jean T. LAPM: A Liquidity-Based Asset Pricing Model. The Journal of Finance 2001;56; 1837-1867

Huang RD, Stoll HR. The Components of the Bid-Ask Spread: A General Approach. The Review of Financial Studies 1997;10; 995-1034

Ibikunle G. Opening and closing price efficiency: Do financial markets need the call auction? Journal of International Financial Markets, Institutions & Money 2015;34; 208-227

Ibikunle G. Trading places: Price leadership and the competition for order flow. Journal of Empirical Finance 2018;49; 178-200

Jacob Leal S, Napoletano M, Roventini A, Fagiolo G. Rock around the clock: An agent-based model of low- and high-frequency trading. Journal of Evolutionary Economics 2016;26; 49-76

Jennings RH, Starks LT, Fellingham JC. An Equilibrium Model of Asset Trading with Sequential Information Arrival. The Journal of Finance 1981;36; 143-161

Jovanovic B, Menkveld A. Middleman in limit order markets. 2016;

Justiniano A, Primiceri GE. The Time-Varying Volatility of Macroeconomic Fluctuations. American Economic Review 2008;98; 604-641

Kalotychou E, Staikouras SK. Stock Market Volatility. Taylor & Francis Group; 2009;

Kaniel R, Liu H. So What Orders Do Informed Traders Use? The Journal of Business 2006;79; 1867-1913

Karpoff JM. The Relation Between Price Changes and Trading Volume: A Survey. Journal of Financial and Quantitative Analysis 1987;22; 109-126

Kelley EK, Tetlock PC. How Wise Are Crowds? Insights from Retail Orders and Stock Returns. The Journal of Finance 2013;68; 1229-1265

Kirilenko A, Kyle A, Samadi M, Tuzun T. The Flash Crash: High Frequency Trading in an Electronic Market. The Journal of Finance 2017;72; 967-998

Kyle AS. An Equilibrium Model of Speculation and Hedging. 1981;

Kyle AS. Market Structure, Information, Futures Markets, and Price Formation. International Agricultural Trade. Westview Press: Colorado; 1984;

Kyle AS. Continuous Auctions and Insider Trading. Econometrica 1985;53; 1315-1335

Kyle AS. Informed Speculation with Imperfect Competition. The Review of Economic Studies 1989;56; 317-355

Kyle AS, Obizhaeva AA. Large Bets and Stock Market Crashes. 2016;

Lamoureux CG, Lastrapes WD. Heteroskedasticity in Stock Return Data: Volume versus GARCH effects. The Journal of Finance 1990;45; 221-229

Laughlin G, Aguirre A, Grundfest J. Information Transmission between Financial Markets in Chicago and New York. The Financial Review 2014;49; 283-312

Laumonier A. HFT in the banana land. 2016.

Lee BS, Rui OM. The dynamic relationship between stock returns and trading volume: Domestic and cross-country evidence. Journal of Banking & Finance 2002;26; 51-78

Lee C, Ready M. Inferring Trade Direction from Intraday Data. The Journal of Finance 1991;46; 733-746

Madhavan A. Market microstructure: A survey. Journal of Financial Markets 2000;3; 205-258

Madhavan A. Exchange-Traded Funds, Market Structure and the Flash Crash. Financial Analysts Journal 2012;68; 20-35

Madhavan A, Richardson M, Roomans M. Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks. The Review of Financial Studies 1997;10; 1035-1064

Malceniece L, Malcenieks K, Putniņš TJ. High Frequency Trading and Co-Movement in Financial Markets. Journal of Financial Economics 2018;Forthcoming;

McCarthy J, Najand M. State Space Modeling of Price and Volume Dependence: Evidence from Currency Futures. Journal of Futures Markets 1993;13; 335-344

Mcinish T, Upson J, Wood R. The Flash Crash: Trading Aggressiveness, Liquidity Supply, and the Impact of Intermarket Sweep Orders. Financial Review 2014;49; 481-509

Menkveld AJ. High frequency trading and the new market makers. Journal of Financial Markets 2013;16; 712-741

Menkveld AJ. The economics of high-frequency trading: Taking stock. Annual Review of Financial Economics 2016;8; 1-24

Menkveld AJ, Koopman SJ, Lucas A. Modeling Around-the-Clock Price Discovery for Cross-Listed Stocks Using State Space Methods. Journal of Business & Economic Statistics 2007;25; 213-225

Menkveld AJ, Yueshen B. The Flash Crash: A Cautionary Tale about Highly Fragmented Markets. Management Science 2017;Forthcoming;

Menkveld AJ, Zoican MA. Need for Speed? Exchange Latency and Liquidity. The Review of Financial Studies 2017;30; 1188-1228

Morris S. Trade with Heterogeneous Prior Beliefs and Asymmetric Information. Econometrica 1994;62; 1327-1347

O'Hara M. Market Microstructure Theory. Blackwell Publishers, Cambrigde, Massachusetts; Oxford, England.; 1995;

O'Hara M. Presidential Address: Liquidity and Price Discovery. The Journal of Finance 2003;58; 1335-1354

O'Hara M. High frequency market microstructure. Journal of Financial Economics 2015;116; 257-270

O'Hara M, Ye M. Is market fragmentation harming market quality? Journal of Financial Economics 2011;100; 459-474

Pacurar M. Autoregressive conditional duration models in finance: A survey of the theoretical and empirical literature. Journal of Economic Surveys 2008;22; 711-751

Park B. Surprising information, the MDH, and the relationship between volatility and trading volume. Journal of Financial Markets 2010;13; 344-366

Pástor L, Stambaugh RF. Liquidity Risk and Expected Stock Returns. Journal of Political Economy 2003;111; 642-685

Petersen MA, Fialkowski D. Posted versus effective spreads: Good prices or bad quotes? Journal of Financial Economics 1994;35; 269-292

Pindyck RS. Risk, inflation, and the stock market. American Economic Review 1984;74;

Poterba JM, Summers LH. The Persistence of Volatility and Stock Market Fluctuations. American Economic Review 1986;76; 1142-1151

Putniņš TJ. What do price discovery metrics really measure? . Journal of Empirical Finance 2013;23; 68-83

Roşu I. Fast and flow informed trading. 2016;

Russell JR, Engle RF. A discrete-state continous-time model of financial transactions prices and times: the autoregressive conditional multinomialautoregressive conditional duration model. Journal of Business and Economic Statistics 2005;23; 166–180.

Rzayev K, Ibikunle G. A high frequency analysis of the information content of trading volume. 2019;

Schwert GW. Why Does Stock Market Volatility Change Over Time? The Journal of Finance 1989;44; 1115-1153

SEC. Regulation NMS. Securities and Exchange Commission: Washington, DC; 2005.

SEC. Findings Regarding the Market Events of May 6, 2010. 2010;

Shkilko A, Sokolov K. Every Cloud Has a Silver Lining: Fast Trading, Microwave Connectivity and Trading Costs. 2016;

Smirlock M, Starks L. A Transactions Approach to Testing Information Arrival Models. Washington University: Washington; 1984;

Sofianos G. Specialist gross trading revenues at the New York Stock Exchange. 1995;

Sun Y, Ibikunle G. Informed trading and the price impact of block trades : A high frequency trading analysis. International Review of Financial Analysis 2016;54; 114-129

Suominen M. Trading Volume and Information Revelation in Stock Markets. the Journal of Financial and Quantitative Analysis 2001;36; 545-565

Van Ness BF, Van Ness RA, Yildiz S. The role of HFTs in order flow toxicity and stock price variance, and predicting changes in HFTs' liquidity provisions. Journal of Economics and Finance 2016;41; 739-762

Wang J. A Model of Intertemporal Asset Prices Under Asymmetric Information. The Review of Economic Studies 1993;60; 249-282

Wu K, Bethel W, Leinweber D, Rübel J, Gu M. A big data approach to analyzing market volatility. Algorithmic Finance 2013;2; 241-267

Wuyts G. The impact of aggressive orders in an order-driven market: a simulation approach. The European Journal of Finance 2011;18; 1015-1038

Yan B, Zivot E. A structural analysis of price discovery measures. Journal of Financial Markets 2010;13; 1-19

Zhang MY, Russell JR, Tsay RS. A nonlinear autoregressive conditional duration model with applications to financial transaction data. Journal of Econometrics 2001;104; 179–207