tinbergen institute

Sait R. Ozturk

Price Discovery and Liquidity in the High Frequency World

Erasmus University Rotterdam

Price Discovery and Liquidity in the High Frequency World

Sait R. Ozturk

665

This dissertation contains three papers investigating issues concerning price discovery and liquidity in the world of intraday trading across financial markets. We pay particular attention to intraday variations in market dynamics, usually employing a Kalman filtering approach combined with Maximum Likelihood methods as a tool to model them.

Chapter 2 proposes a novel econometric methodology to disentangle informative price moves, i.e. price discovery, from transitory noise. We analyze a recent data set of S&P 500 stocks and find statistical evidence for time-variation in intraday price discovery. Tighter quoted spreads attract informed trading from other exchanges. Exchange listing and industrial sector of a stock significantly affect the dominant venues of price discovery in different parts of the day and following macroeconomic news announcements.

Chapter 3 studies why a majority of trades still happen during the pit hours, i.e. when the trading pit is open, even after the pit ceased to be a liquid and informative venue. We examine the case of 30-year U.S. Treasury futures and find evidence for a feedback mechanism between trading activity, price informativeness, information asymmetry and price impact of trades.

Chapter 4 investigates a predatory tactic aimed at generating transitory extreme price moves, allegedly employed by high-frequency traders. Our data set contains years-long message-level NASDAQ data for 8,000 stocks. From late 2011 on, we observe a trend toward stronger positive relationships of the high-frequency trading activity before the move with the size of the price move, the degree of the subsequent reversal as well as with the deterioration in market quality measures.

Sait R. Ozturk obtained a BA degree in Economics from Sabanci University and an MA degree in Sociology from Bogazici University, both in Istanbul. Following this, he started the MPhil program in Economics at the Tinbergen Institute in 2010. Upon completing the MPhil program in 2012, he joined the PhD program at the Econometric Institute of Erasmus University Rotterdam.

PRICE DISCOVERY AND LIQUIDITY IN THE HIGH FREQUENCY WORLD

Sait R. Ozturk

ISBN 978 90 5170 811 0

This book is no. 665 of the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Price Discovery and Liquidity in the High Frequency World

Prijsvorming en Liquiditeit in de Hoge Frequentie Wereld

Doctoral dissertation

for the purpose of being admitted to the degree of Doctor at

Erasmus University Rotterdam

on the authority of the

Rector Magnificus

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board.

The public defense shall take place on

Friday the 16th of September 2016 at 11.30 hrs

by

Sait Ridvan Ozturk

born in Shumen, Bulgaria.

Erasmus University Rotterdam

Doctoral committee:

Promotor:	Prof.dr. D.J.C. van Dijk
Other members:	Prof.dr. F. de Jong
	Prof.dr. A. Lucas
	Dr. D. Bongaerts
Co-promotor:	Dr. M. van der Wel

Acknowledgements

I freshly remember how six years ago, while looking for a masters' thesis topic, I encountered the field of market microstructure, the study of financial trading mechanisms. As a student of economics with a newly gained interest in finance, it looked like an excellent starting point to understand how financial markets work in real life. This initial spark of interest opened a years-long chapter of my life, made me wander across countries and continents, enabled me to meet with dozens of wonderful and bright people and challenged me in many respects. This dissertation is merely one, even if the most tangible, product of this journey. I hope I passed the bigger test.

I owe the largest debt of gratitude to my caring and wise supervisors and co-authors, Michel and Dick. With his usual insightfulness, Michel gave my first and very warm welcome to this field of study. Dick has always been a source of extensive knowledge and experience. They meticulously tried to correct my errors both in logic and writing. I could not survive this journey without their guidance and assistance.

My third co-author, Bruce, taught me so much about the minutest details of a wide range of financial markets. Our frequent lunches in New Brunswick hosted many delightful discussions about finance, politics and, of course, food.

Elis was my closest companion throughout this journey. We enjoyed its best moments as well as endured its lowest points together. Her support and understanding enabled me to withstand arduous challenges. Bonds formed in such conditions may change form, but will always stay strong.

I cannot imagine how I could survive the formidable MPhil program and carry through the PhD without Taylan's support. He became a brotherly mentor and a source of idealistic optimism and assiduousness in my life.

I enjoyed some of my best house parties thanks to Lukas, Anghel, Lore and Max. Truth to be told, I was usually overwhelmed by their combined musical awesomeness. It's totally on me that I cannot yet play a single musical instrument properly, in spite of their amazing example. I promise, one day in the future I will also join the band.

With Uyanga, Tong, Violeta, Ilke, Alex and Mehmet, we shared not only the university, but many cinematic and gastronomical pleasures. Thanks for tolerating my occasional quirks, like my insistent screenings of Méliès' *Le Voyage dans la Lune*.

Piotr became my usual partner in crime at sailing through both Amsterdam and New York City. I look forward to our new adventures in London.

Thanks to his endless wit and sarcasm, my office mate Max, more than anyone else, compelled me to set my doctoral studies into perspective. I will recall those humbling lessons with immense joy.

I have always been delighted by Mustafa and Murat's excellent company, not just because they taught me many lessons about how to get things done.

David, Tomasz, Aras, Sevil, Alican, Erin, Ömer, Dilara and Lerby are among the many friends who eased my academic travels with their warm camaraderie and transformed them into homely adventures.

Among many others, Dennis, Bart, Sander, Francine, Gertjan, Tom, Bruno, Shawny, Myrthe, Martin, Eran, Pınar, Nalan, Laura, Pourya, Jonathan, Zara, Karolina, Lucy, Eszter, Gerge, Arturas and Erkki, made this journey so much more fun. I cannot take their warm presences for granted. I learned so much from you. To those who suffered, my deep apologies for being the clumsy elephant in the Lisa computer cluster again and again.

My journey has definitely been eased thanks to my encounter with the path of yoga. John's classes at the Erasmus Sport Center were instrumental at introducing me to the asana practice. The coincident discovery of my parents transformed yoga to a joyful family pastime. Çetin's guidance both deepened my practice and instilled it with a sense of holistic purpose. Thanks to them, I realized that willpower to overcome obstacles is not merely a quality of the mind, but is nurtured through the activity of the whole body.

Anneciğim, her ne kadar arada sırada ukala ukala bunu eleştirsem de, senin her zaman kendinden önce ailemizdeki herkesi düşünen anlayışlılığın ve ilgin kara bulutların çöktüğü anlarda en büyük güvenli limanım oldu. Babacığım, bizimle beraber öğrenmeye ve büyümeye cesaret eden açık fikirliliğin ve alçak gönüllüğünle herkesin sahip olmasını isteyeceğim bir baba oldun bize. Umarım sizin azminizden, coşkunuzdan ve iyiliğinizden nasibimizi almışızdır. Aysel, sadece kardeşim değil, en yakın dostum olacaksın her zaman. Hayat ağacının dallarında tırmanırken daha nice kirazlar yiyeceğiz.

Contents

1 Introduction **1**

2 Intraday Price Discovery in Fragmented Markets **7**

2.1 Introduction 8

2.2 Measuring price discovery 13

2.2.1 The Unobserved Components Model 13

2.2.2 De Jong-Schotman information shares 14

2.2.3 Intraday variation in information shares 18

2.2.4 Testing for intraday variation 20

2.3 Simulation study 22

2.3.1 Comparison of GMM and state space ML methods 22

2.3.2 Capturing time-variation with the state space ML method 23

2.3.3 Capturing time-variation with a misspecified model 31

2.4 Intraday variation of price discovery in S&P 500 stocks 34

2.4.1 Data and summary statistics 34

2.4.2 Estimation Results 39

2.4.3 What drives the intraday variation in information shares? 60

2.5 Conclusion 67

3 Why do the pit hours outlive the pit? **73**

3.1 Introduction 74

3.2 Hypotheses and Methodology 79

3.2.1 Hypotheses 79

3.2.2 Information Shares 82
3.2.3 Structural Model with Intraday Variation 85
3.3 Data and Descriptive Statistics 87
3.3.1 Data 87
3.3.2 Activity clustering 90
3.4 Why does the trading activity cluster around the pit hours? 99
3.4.1 Is pit trading informative? 99
3.4.2 What determines trading activity during the afterhours? 101
3.5 Conclusion 115

4 Momentum Ignition?
HFT Activity During Transitory Extreme Price Moves **119**
4.1 Introduction 120
4.2 Data and Summary Statistics 126
4.2.1 The ITCH data set 126
4.2.2 The Detection Algorithm for Extreme Price Moves 127
4.2.3 Summary Statistics of Move Events 129
4.3 Methodology 136
4.4 Analysis of Results 143
4.4.1 Predicting the Move 143
4.4.2 Predicting the Degree of Reversal 145
4.4.3 Effects on Market Quality 149
4.5 Conclusion 154

Bibliography **157**

Summary **165**

Samenvatting (Summary in Dutch) **169**

Chapter 1

Introduction

> Merchant: In this chaos of opinions, which one is the most prudent?
>
> Shareholder: To go in the direction of the waves and not fight against the powerful currents.
>
> *Joseph de la Vega, 1688*

The field of market microstructure studies trading mechanisms of financial securities, where latent demands are incorporated into asset prices (Madhavan, 2000; Biais, Glosten, and Spatt, 2005; Hasbrouck, 2007). Following the emergence of modern financial institutions, trading mechanisms became subjects of study for several authors, notably Joseph de la Vega's early work of the Amsterdam Stock Exchange in 1688. A modern scientific literature, on the other hand, developed in the last four decades, roughly since the term of market microstructure is coined for the first time by Garman (1976). This dissertation contributes to this collective endeavor with analyses on two of the main roles of financial markets, price discovery and the provision of liquidity, and on the effects of the technological developments and sophistication in financial markets.

Price discovery denotes the continual pricing of financial securities considering newly available information through the mechanism of demand and supply. Modern financial markets are typically organized around order books recording limit orders, i.e. the desires to buy or sell an amount of a security at a certain price. When all information is public, the classic Walrasian mechanism provides a straightforward relationship between individuals' consumption decisions and asset prices. However, this is hardly the case in the complex world of financial markets where transaction costs and information asymmetries are prevalent. The limit orders often scatter around a range of prices, where the highest volumes are usually quoted near the highest price to buy a security, the best bid, and the lowest price to sell a security, the best ask, the gap between which constitutes the bid-ask spread.

This dispersion of quotes and the bid-ask spread relate to another major role of financial markets, namely the provision of liquidity. Liquidity denotes the costs associated with trading, either in monetary units, if one is willing to cross the spread, or in terms of execution time, if one wants to wait for another party to cross the spread. Due to the broadness of this concept, a plethora of liquidity measures are proposed by the literature (Goyenko, Holden, and Trzcinka, 2009). Among the most straightforward ones are those related to the trade volume and the bid-ask spread of the concerning security.

The bid-ask spread has been attracting particular attention, because it captures both the uncertainty about the true value of a security and the amount of transaction costs a party needs to endure to buy or sell a security. A strand of the literature investigates the influence of the inventory management costs on the spread, including the aforementioned pioneering work of Garman (1976). Roll (1984) models the spread as unpredictable pricing errors stemming from order handling and other costs. A third branch points to adverse-selection costs caused by informed parties (Kyle, 1985; Glosten and Milgrom, 1985; Glosten, 1994). These main approaches have been extended by the inclusion other factors such as autocorrelation dynamics and remain open to cross-pollination, as shown by many studies (e.g., Lin, Sanger, and Booth, 1995; Huang and Stoll, 1997; Madhavan, Richardson, and Roomans, 1997).

The other aspect of liquidity, the trading activity measured by the trade volume, has a rather ambiguous relationship with informed trading. Two classical models in the literature point to opposite conclusions. In line with the positive effect of adverse selection on the bid-ask spread, Foster and Viswanathan (1990) finds that informed traders drive out uninformed ones, and thus decrease the overall trading activity, by imposing adverse selection costs on them. In contrast, Admati and Pfleiderer (1988) propose a model where uninformed traders choose to cluster with informed traders. In this model, the uninformed traders still face adverse selection costs, but they also benefit from the competition of informed traders sharing the same information.

Technological developments have been posing new questions about the efficacy of financial markets in these roles. Fully functioning electronic trading platforms started to emerge nearly three decades ago, the Chicago Mercantile Exchange's Globex platform, the subject of one of our studies, being one of the first in 1992. Although several studies demonstrate that the open limit order markets benefit informed as well as uninformed traders by the transparency, anonymity and control they provide and generate a liquidity externality by aggregating all orders in one venue (Glosten, 1994; Pagano and Röell, 1996; Biais, Foucault, and Salanié, 1998), the current success of these electronic markets have not been unanimously foretold (e.g., Venkataraman, 2001). Nowadays, trading pits survive only at a few markets, usually handling trades which are too large or too complicated for the electronic newcomers (De Jong, Nijman, and Röell, 1995; Sofianos and Werner, 2000).

The last decade saw the kindling of a new debate on the effects of technological change in the microstructure of financial markets. Firms specialized in high frequency trading (HFT) dramatically altered the market landscape with the use of automatized trading algorithms responding to shifts in the market dynamics within milliseconds. Many studies find HFT activity benefits price discovery (Carrion, 2013; Brogaard, Hendershott, and Riordan, 2014) with the downside that this advantage also imposes costs on the other parties, as shown by both theoretical (Hoffmann, 2014; Foucault, Hombert, and Roşu, 2016) and empirical (Brogaard et al., 2014) studies. However the main utilisers of the HFT technology do not seem to be aggressive traders, but passive yet smart liquidity providers acting on both sides of the market, buying from or selling to traders willing to cross the bid-ask spread (Menkveld, 2013; Hagstromer and Norden, 2013; Brogaard, Hagstromer, Norden, and Riordan, 2014). This finding resonates with the observed positive effects of

HFT activity on measures of liquidity (Hasbrouck and Saar, 2013).

The upcoming three chapters of this dissertation contribute these literatures. The first two explore intraday variation in the unobserved market variables, measuring price discovery and market liquidity. We achieve this investigation by exploiting the dynamic modelling capabilities of Kalman filtering methods (Menkveld, Koopman, and Lucas, 2007; Frijns and Schotman, 2009). These two chapters relate to the contemporary trends of market fragmentation and prevalence of electronic trading. The third one focuses on the other major debate on technological development in the market microstructure literature, namely the effects of HFT activity. We analyze HFT activity during extreme price movements, where it is a suspect of predatory trading.

The second chapter proposes a new methodology to examine price discovery in this high frequency world. We explore intraday variation in the contribution to price discovery across different exchanges, where the same stocks trade simultaneously. We estimate a structural model with time-varying parameters in state space form using Maximum Likelihood and produce measures of informativeness for each market, the so-called information shares (Hasbrouck, 1995; De Jong and Schotman, 2010). An extensive simulation study provides evidence for the accuracy of this methodology. We analyze data for 50 S&P 500 stocks in 2013 and find that the constancy of shares in price discovery, a frequently-held assumption of the existing literature, is rejected. Tighter quoted spreads attract informed trading from other exchanges. Exchange listing and industrial sector of a stock significantly affect the dominant venues of price discovery in different parts of the day and following macroeconomic news announcements.

The third chapter studies why a majority of trades still happen during the pit hours, i.e. when the trading pit is open, even after the pit ceased to be a liquid and informative venue. We investigate the case of 30-year U.S. Treasury futures using a ten-years-long intraday data set which contains the introduction of the CME Globex platform as an example of sophistication in electronic trading. We use a structural model to estimate the time-variation in potential factors of the clustering of trading activity around the pit hours, namely price informativeness, information asymmetry and price impact of trades. We find evidence for a feedback mechanism between trading activity and these factors.

Across the sample period, price informativeness during the afterhours is a consistently significant factor attracting trade activity. Information asymmetry has a negative effect on afterhours activity, particularly during the crisis years. The negative effect of price impact on afterhours activity ceases to be a significant factor from 2007 on, possibly due to improvements in order execution algorithms and electronic trading facilities.

Lastly, the fourth chapter investigates the effect of HFT activity around large price moves. Our data set covers message-level NASDAQ data for 8,000 CRSP stocks from July 2007 to December 2013. A monthly average 3.4 extreme moves happen for each stock. About half of the moves are transitory, i.e. they are followed by a reversal more than two-thirds of the move size, and nearly a quarter of them are permanent, i.e. they experience a reversal of less than one-third. We find HFT activity contributes significantly to price discovery, given that it predicts the move size and direction nearly as good as non-HFT trades, which comprise the bulk of the trade volume. HFT activity during the extreme events also reduces the market quality deterioration, which may be stemming from the relative eagerness of HFT firms to bet on mean-reversal and thus to soften the moves by trading against them. However, from October 2011 on, we observe for transitory moves a trend toward a stronger predictive power of premove HFT activity on the move size and a stronger positive relationship with the magnitude of price reversals as well as with the deterioration in three market quality measures, namely quoted spreads, market volatility and execution-to-cancellation ratios.

Chapter 2

Intraday Price Discovery in Fragmented Markets

This chapter is a joint project with Dr. Michel van der Wel and Prof. Dick van Dijk of Erasmus University Rotterdam.

2.1 Introduction

Financial markets incorporate new information into asset prices by matching buyers and sellers. They thereby facilitate the discovery of what the price of an asset should be. Nowadays this "price discovery" process can take place across multiple exchanges and instruments, as different securities and derivatives based on the same underlying asset may trade on several venues. In the case of such a multiplicity, there may be differences in the share with which each market's trades contribute to discovering the one true price of the underlying asset. Knowledge of these so-called information shares of different markets would benefit both investors concerned with price informativeness and adverse selection risk as well as policy makers investigating the determinants of price efficiency. Existing studies often assume the contributions of different markets to price discovery are constant at least over the course of the day. We analyze intraday variation in price discovery, and consider which factors may explain such variation.

The measurement of price discovery requires isolating informative price movements from noise. Observed price changes constitute the most obvious indicator of price discovery. However, they form an imperfect measure as observed prices are susceptible to transitory mispricing, caused by noise trading or temporary order imbalances, for example. In contrast, when security prices absorb new information due to informed trading, these price changes last permanently. Hasbrouck (1995) demonstrates that the above implies the existence of co-integration relationships between security prices and develops a framework exploiting these to distinguish permanent and transitory price changes. His work initiated a booming literature on price discovery measures and information shares.

Early studies, like Hasbrouck (1995), effectively assume the contributions of different trading venues to the efficient price innovations to be constant over time, or at least for the sample period used for estimation. However, changes in the characteristics of exchanges and securities - such as increases in trade volume and electronization of trading mechanisms - make this assumption implausible. Based on these motivations, the more recent literature mostly divides the sample into short sub-periods, and typically considers daily measurements of information shares (Chakravarty, Gulen, and Mayhew, 2004; Hasbrouck, 2003; Mizrach and Neely, 2008, among others).

In spite of providing a higher level of sophistication, measurements of information shares at the daily frequency are unable to keep up with the current pace of financial markets and available data. Current information share methodologies are not able to answer questions about differences in price discovery across different parts of the day or the digestion of public news, most of which happens in a matter of minutes, if not seconds. A growing body of studies infer intraday variation in informed trading indirectly from the dynamics in other market characteristics, such as liquidity, depth and volatility in limit order markets (Ahn, Bae, and Chan, 2001), asymmetric information proxies and trade volume before and after public announcements (Chae, 2005) or predictions of a model with informed and uninformed traders (Lei and Wu, 2005).

In this paper we consider the possibility of examining intraday variation in information shares directly. We propose a novel method to capture the intraday dynamics of price discovery based on the structural time series model proposed by Hasbrouck (1993). In this structural model, the observed security prices depend on a single underlying latent true price. Differences between the observed prices and the latent price consist of two components. On the one hand, these pricing errors are linked to the innovations to the latent true price capturing lagged adjustment or over-reaction to information. On the other hand, they stem from uncorrelated errors representing dynamics like noise trading. Following De Jong and Schotman (2010), information shares can be expressed as a function of the structural model parameters, including the variances of the latent price innovations and the uncorrelated errors. We extend this model by allowing the innovation and noise variances to vary throughout the trading day using a flexible Fourier form. This specification is appealing because the Fourier functions are able to capture a wide range of continuous patterns. This model with time-varying variances naturally leads to time-variation in the information shares, thus enabling us to capture intraday variation in the relative contributions of different trading venues to price discovery. We estimate this model with Maximum Likelihood (ML) using Kalman filtering.

We examine the usefulness of our modeling approach by means of a simulation study and an empirical application. In order to ensure the relevance of the simulation study for empirical research, we use the estimates from our empirical study as a benchmark and examine the effects of various modifications. The simulation study compares our ML estimates and GMM estimates following De Jong and Schotman (2010) for the case of no time-variation and evaluates the precision of ML estimates when information shares are time-varying. The simulation evidence demonstrates that our state space ML method generates accurate estimates for a wide range of settings with varying number of observations, venues and parameters of the time-varying information share model.

Our empirical study provides convincing evidence for intraday variation in informed trading. We examine 50 constituents of the S&P 500 index for the second half of 2013 using a 1-minute sampling frequency. Nearly all trade activity of these stocks occurs on four exchange groups, namely NYSE, NASDAQ, BATS and Direct Edge. The NYSE and the NASDAQ groups provide overall the largest contributions to price discovery with average information shares of 43.4% and 33.4%, respectively. The market open and close and macroeconomic news announcements lead to increases in the variance of the latent price process, capturing the overall informed trading activity, consistent with the U-shaped intraday pattern for informed trading documented in the literature (see, e.g., Admati and Pfleiderer, 1988; Foster and Viswanathan, 1993; Slezak, 1994). Such informative events also alter the shares of different exchange groups in price discovery: NYSE is on average 10.9% more informative during the first half hour of the day compared to the rest of the day, while NASDAQ has a 43.5% larger information share during the last hour. The FOMC announcements in the afternoon increase both the overall price discovery measured by the variance of latent innovations and NYSE's information share. Our statistical tests reject the hypothesis of constant price discovery during the midday, even when we exclude days with macroeconomic news.

We seek to explain the variation of information shares using intraday market and stock characteristics. Using a market share attraction model, we find that the number of trades, quoted spreads and volatility as well as market capitalization, exchange listing and industry have significant explanatory power for the dynamics of the relative information shares. Tighter quoted spreads and lower volatility in an exchange consistently attract informed trading activity from other exchanges. Although our sample of stocks trade in multiple venues and the two main primary listing markets are very competitive in their shares of trading activity, we find that the services provided by exchanges listing the stocks contribute significantly to price discovery. Exchange listing has a strong influence on the level of information shares: Being listed on NYSE instead of NASDAQ causes a 36.6% drop in the ratio of the NASDAQ information share to that of NYSE and even more dramatic drops of 59.4% and 63.4% for BATS and Direct Edge, respectively. The leading venues within stock groups in terms of industrial classification are mostly in line with the composition of the groups in terms of exchange listings. In particular, for financial stocks in our sample, all of which are listed on NYSE, non-NYSE exchange groups have about half of their usual information shares.

Our work is related to a number of studies investigating price discovery by means of state space methods. Upper and Werner (2007) estimate a reduced-form VECM representation in the state space framework, while Frijns and Schotman (2009) and Korenok, Mizrach, and Radchenko (2011) use directly the structural model of Hasbrouck (1993) in state space form, albeit not allowing for intraday variation in information shares. A closely related paper is Menkveld et al. (2007), who suggest a similar structural model in state space form that allows for time-variation in parameters throughout the day. Our set-up differs in three important respects. First, in their case the comparison is for overall variation in prices throughout the day for all markets an asset trades on, and not for price discovery across markets. A result is that they study variance ratios for different parts of the trading day (a time series aspect), and not price discovery measures across the various exchanges (a cross-sectional aspect) as we do. Second, their model is designed for lower intraday frequencies such as an hour, as they assume that the innovation in the latent efficient price is fully incorporated into the observed prices at each period (which is not plausible for higher intraday frequencies). Third, we study the higher-frequency change in structural model parameters using a flexible Fourier form, while they focus on step functions to model time-variation.

The information share methodology of De Jong and Schotman (2010) that we use has several advantages over other measures in the literature. Hasbrouck (1995) estimates the contributions of each security to the variance of innovations in the latent price. Comparative studies, such as Baillie, Booth, Tse, and Zabotina (2002), De Jong (2002) and Lehmann (2002), find this focus on variance more appropriate for price discovery measurement than the common factor decomposition of Gonzalo and Granger (1995), as Harris, McInish and Wood (1997; 2002) do. The proposal of De Jong and Schotman (2010) similarly works at the variance level, but resolves two main concerns about the Hasbrouck approach. Firstly, Hasbrouck information shares are not unique but they come in the form of a range, often with a substantial difference between the upper and lower boundaries.[1] Secondly, it relies on a reduced form estimation methodology which does not provide estimates of structural parameters.

The remainder of the paper is organized as follows. Section 2.2 introduces the unobserved components model and De Jong and Schotman (2010) information shares, followed by our extension to capture intraday variation. Sections 2.3 provides simulation evidence for our methodology. Section 2.4 reports the empirical results, including the analysis of the determinants of the estimated intraday variation in information shares. Section 2.5 concludes.

[1]Grammig and Peter (2013) provide identification restrictions using the distributional properties of financial price series to overcome the non-uniqueness problem.

2.2 Measuring price discovery

This section presents the methodology to measure the contributions of different trading venues to price discovery. Its four parts elaborate on the structural model of Hasbrouck (1993) and De Jong and Schotman (2010), the information shares suggested by De Jong and Schotman (2010), our novel implementation of intraday time-variation under the state space framework and testing for intraday variation, respectively.

2.2.1 The Unobserved Components Model

We use a version of the unobserved components model of Hasbrouck (1993) extended by De Jong and Schotman (2010) as our structural model. In this framework, all observed prices based on the same underlying asset (such as the observed prices on multiple exchanges of the same stock) are driven by one latent efficient price process (the unknown true price of that underlying stock). This latent price is defined as the end-of-period value of the asset conditional on all publicly available information at time t. Thus this price process satisfies the semi-strong form of market efficiency in line with the range of information it encompasses (Fama, 1970). Since all public information is impounded in this latent price, the best prediction for the asset price in period $t+1$ is the price at time t and therefore it is modeled as a random walk with stationary innovations r_t. The observed asset prices on different exchanges deviate from this latent price with a stationary error term as long-term or unbounded deviations are ruled out by arbitrage. These relations can be represented as an unobserved components model as

$$\begin{aligned} p_t &= \iota p_t^* + u_t, \\ p_t^* &= p_{t-1}^* + r_t, \end{aligned} \tag{2.1}$$

where p_t is an $N \times 1$ vector of log observed prices $p_{i,t}$, $i = 1, \ldots, N$, u_t is an $N \times 1$ vector of stationary disturbance terms $u_{i,t}$, p_t^* is the scalar latent efficient price, r_t is the innovation in the latent price with mean zero and variance σ_r^2 and ι is an $N \times 1$ vector of ones.

The error terms $u_{i,t}$ capture microstructure effects in the observed prices. It comprises two components distinguished by their correlation with the efficient price innovation r_t. First, $u_{i,t}$ has an information-correlated pricing error component $\alpha_i r_t$ that captures dynamics such as adverse selection. The second component $e_{i,t}$ is uncorrelated with information, but stems from factors such as noise trading or price discreteness. This idiosyncratic noise $e_{i,t}$ has mean zero and $N \times N$ covariance matrix Ω, allowing for correlation in this noise component across observed prices. With these two components, the specification for the disturbance terms $u_{i,t}$ is

$$u_t = \alpha r_t + e_t + \Psi e_{t-1}, \tag{2.2}$$

where α is an $N \times 1$ vector of α_i's, e_t is an $N \times 1$ vector of idiosyncratic noise $e_{i,t}$, and Ψ is an $N \times N$ coefficients matrix. De Jong and Schotman (2010) propose the inclusion of the lagged noise e_{t-1} in the observed price dynamics in order to capture serial correlation in high-frequency intraday returns. We provide a state space representation of the unobserved components model in Appendix A.

2.2.2 De Jong-Schotman information shares

De Jong and Schotman (2010) propose a price discovery measure quantifying the explanatory power of changes in each of the observed security prices for the innovations in the latent price. For this purpose, the total price innovation in period t is defined as

$$\nu_t = p_t - \iota p^*_{t-1} = (\iota + \alpha) r_t + e_t + \Psi e_{t-1}. \tag{2.3}$$

We may then consider the regression of the innovation in the latent price on the total innovations in individual prices, that is

$$r_t = \gamma' \nu_t + \eta_t, \tag{2.4}$$

where η_t is the innovation in the latent price unrelated to innovations in market prices.

The regression coefficient γ is given by

$$\gamma = \frac{cov(r_t, \nu_t)}{var(\nu_t)} = \Upsilon^{-1}(\iota + \alpha)\sigma_r^2. \tag{2.5}$$

where $cov(r_t, \nu_t) = (\iota + \alpha)\sigma_r^2$ follows from Eq. (2.3) and Υ denotes the covariance matrix of ν_t. From Eq. (2.3) we also have

$$\Upsilon = \sigma_r^2(\iota + \alpha)(\iota + \alpha)' + \Omega + \Psi\Omega\Psi', \tag{2.6}$$

Using Eq. (2.5), the goodness-of-fit of the regression in Eq. (2.4) can be expressed as

$$R^2 = 1 - \frac{\sigma_\eta^2}{\sigma_r^2} = \frac{\gamma'\Upsilon\gamma}{\sigma_r^2} = \gamma'(\iota + \alpha) = \sum_{i=1}^{N} \gamma_i(1 + \alpha_i).$$

This leads De Jong and Schotman (2010) to propose an information share for the price on the i-th market, denoted IS_i, with a partial R^2 interpretation, namely

$$IS_i = \gamma_i(1 + \alpha_i). \tag{2.7}$$

Assuming the diagonality of the idiosyncratic noise matrix Ω and the coefficients' matrix of lagged noise Ψ, the information shares in Eq. (2.7) can be expressed as

$$IS_i = \frac{(1 + \alpha_i)^2/(\omega_i^2 + \psi_i^2\omega_i^2)}{1/\sigma_r^2 + \sum_{j=1}^{N}(1 + \alpha_j)^2/(\omega_j^2 + \psi_j^2\omega_j^2)}, \tag{2.8}$$

where ω_i^2 and ψ_i are diagonal entries of respectively the Ω and the Ψ matrices. As this expression implies, the sum of these information shares, i.e. the R^2 of the regression, is not necessarily equal to one.

The information shares IS_i defined in Eq. (2.8) improve on Hasbrouck's approach by providing unique measures of price discovery estimated from a structural model, while keeping the focus on the variance of the latent innovations. Hasbrouck (1995) estimates the reduced form of the unobserved components model and the resulting information shares come in the form of a range between certain lower and upper bounds. The unique identification of Hasbrouck information shares requires a strong assumption like the diagonality of the residual covariance matrix, i.e. the shocks to the prices in the reduced form system should be uncorrelated, which is violated in any empirical application to a degree. Although the use of higher sampling frequencies reduces this correlation, Yan and Zivot (2010) point out that information share estimates based on high frequencies are more susceptible to distortions by transitory noise. The diagonality assumptions on the Ω and Ψ matrices in order to obtain Eq. (2.8), on the other hand, are both plausible and testable. The diagonality of Ω means that the idiosyncratic noise components of the price changes in different markets are uncorrelated. A diagonal Ψ matrix implies that the mispricing in one exchange is not influenced by the previous period's noise in another exchange. These two diagonality assumptions are much weaker and the GMM framework offers tests to evaluate their validity.[2]

Computing the information shares IS_i according to Eq. (2.8) obviously requires estimates of the parameters in the unobserved components model in Eq. (2.1) and Eq. (2.2). De Jong and Schotman (2010) present a GMM approach to obtain these. The auto-covariances of the observed returns provide the following moment conditions:

$$\Gamma_0 = E[\Delta p_t \Delta p_t'] = \sigma_r^2 \left((\iota + \alpha)(\iota + \alpha)' + \alpha\alpha'\right) + \Omega + (\Psi - I)\Omega(\Psi - I)' + \Psi\Omega\Psi', \quad (2.9)$$

$$\Gamma_1 = E[\Delta p_t \Delta p_{t-1}'] = -\sigma_r^2 \alpha(\iota + \alpha)' + (\Psi - I)\Omega - \Psi\Omega(\Psi - I), \quad (2.10)$$

[2]De Jong and Schotman (2010) provide an extensive discussion of this diagonality restriction compared to two other commonplace types of restrictions used to identify structural parameters of such models. The Beveridge–Nelson normalization excludes the noise process e_t and thus the noise covariance matrix Ω in Eq. (2.2). By contrast, the Watson normalization sets the correlation term of innovation and noise variances, α, to zero for one of the exchanges. Although more plausible, the Watson normalization requires an obvious market to be designated as the central market, which we lack due to the close competition between NYSE and NASDAQ. We use instead the diagonality restriction, as it also easily ensures the positive semi-definiteness of the time-varying noise covariance matrix.

$$\Gamma_2 = E[\Delta p_t \Delta p'_{t-2}] = -\Psi\Omega, \tag{2.11}$$

where $\Delta p_t = p_t - p_{t-1}$. These conditions identify the parameters required for the computation of information shares, namely σ_r^2, Ω and α.

Alternatively, the unobserved components model in Eq. (2.1) and Eq. (2.2) can be estimated by ML using the Kalman filtering. We consider this approach here, also because the extended model with time-varying information shares introduced in Section 2.2.3 cannot be estimated with GMM. As the latent price p_t^* follows a random walk and to account for overnight price changes, we re-initialize p_t^* every day with a diffuse prior and exclude a number of initial observations from the likelihood maximization as these may be unreliable due to the initial convergence of the Kalman filter.[3]

Kalman filtering enables the identification of a richer microstructure model. In particular, we allow for a higher lag order for the noise terms to have a more flexible structure for serial correlations. The error terms in Eq. (2.2) are redefined as

$$u_t = \alpha r_t + e_t + \sum_{j=1}^{L} \Psi_j e_{t-1-j}, \tag{2.12}$$

where L is the number of noise lags. In this model, the information shares are given by

$$IS_{i,t} = \frac{(1+\alpha_i)^2 / (\omega_{i,t}^2 (1 + \sum_{j=1}^{L} \psi_{i,j}^2))}{1/\sigma_{\epsilon,t}^2 + \sum_{j=1}^{N} (1+\alpha_j)^2 / (\omega_{j,t}^2 (1 + \sum_{j=1}^{L} \psi_{i,j}^2))}, \tag{2.13}$$

which collapses to Eq. (2.8) for $L = 1$.[4] We decide on the number of lags by comparing models differing in this aspect using the Schwarz Information Criterion.

[3]In our empirical study, we exclude the first three observations of every day in our state space ML estimation from the likelihood calculation. In our simulation study, on the other hand, we exclude the first 11 observations for the sake of consistency within the section: In Section 2.3.3 we use a step function with 10 steps as a benchmark and leaving 11 of 391 observations at the 1-minute frequency of the trading day gives 380, which is a multiple of 10.

[4]The derivation is given in Appendix B.

2.2.3 Intraday variation in information shares

Time-variation in the information shares IS_i in Eq. (2.8) can be introduced by considering a time-varying parameter extension of the unobserved components model as given by Eq. (2.1) and Eq. (2.2). This can be attained by making at least one of the parameter groups vary over time, namely α, Ψ, σ_r^2 or Ω. The latter two variance terms have the advantage of an established literature linking intraday volatility changes to changes in informed trading. Intraday volatility is documented to follow an inverted J-shape or a U-shape pattern during trading hours (Wood, McInish, and Ord, 1985; Lockwood and Linn, 1990). On the one hand, a number of asymmetric information models noted this pattern as an empirical prediction for markets with informed and uninformed traders (Admati and Pfleiderer, 1988; Foster and Viswanathan, 1993; Slezak, 1994). On the other hand, Hsieh and Kleidon (1996) document several dynamics unrelated to informed trading which aid to the formation of this intraday volatility pattern. The main area of contention lies on whether the start and the end of the day have higher levels of information absorption into prices and if this is accompanied with larger or lower amounts of noise.

Given these theoretical and empirical claims for the intraday variation of informed and noise trading, a natural way to model intraday variation in price discovery is making both the innovation and the noise variances time-varying. We implement time-variation using a combination of flexible Fourier trigonometric functions and a polynomial (Andersen, Bollerslev, and Das, 2001; Gallant, 1981). The variance entries $\zeta_{i,t}^2$ have the form

$$\zeta_{i,t}^2 = \exp\left(c_i + \sum_{p=1}^{P} \theta_{i,p} t^p + \sum_{q=1}^{Q}\left(\delta_{i,q}\cos\left(\frac{2\pi q t}{T}\right) + \phi_{i,q}\sin\left(\frac{2\pi q t}{T}\right)\right)\right), \tag{2.14}$$

where $\zeta_{i,t}^2$ represents the processes of σ_r^2 and ω_i^2's, i.e. the diagonal entries of the Ω matrix, $t = 1, \ldots, T$, with T being the number of observations per day, P the order of the polynomial part, and Q the total number of flexible Fourier sets. We use an exponential specification for the variances to facilitate an unconstrained maximization procedure given that trigonometric functions can have negative values. The flexible Fourier form can model complex dynamics and smooth transitions. However using solely the flexible Fourier part would impose equality of the variances at the start and end of the day. This is avoided by complementing it with the polynomial component. As in the case of the lag order L, the order of the polynomial P and the number of flexible Fourier sets Q can be decided using the Schwarz Information Criterion.

The flexible Fourier specification has several advantages over alternative specifications for capturing time-variation in parameters. A first, simpler, alternative would be to use step functions. A disadvantage of the step function approach is that it generates unlikely jumps between consecutive time periods. Moreover, it introduces the challenge of choosing the number of periods and optimizing period lengths, because assuming them to be equal in length may be too restrictive. A second alternative is to use spline functions instead of the flexible Fourier form. Also here a challenge is that of finding the right number of knots for the spline and the precise knot locations.

In this time-varying setting, we can evaluate the effect of changes in the innovation and noise variances on the information shares using Eq. (2.8). An increase in the innovation variance σ_r^2 boosts all information shares. Therefore both individual information shares and the total explanatory power of observed trading venues are amplified. By contrast, an increase in the noise variance $\omega_{i,t}^2$ of the asset's price on exchange i reduces the corresponding information share as well as the sum of all information shares, while increasing the shares of exchanges other than i.

Finally, introducing time-variation in the error variances as in Eq. (2.14) obviously implies that we can no longer use the GMM approach of De Jong and Schotman (2010) to estimate the model parameters. The model, however, still keeps its state space representation, albeit with time-varying variances, and as such we can obtain parameter estimates by means of ML combined with Kalman filtering.

2.2.4 Testing for intraday variation

In this section we provide a methodology to statistically test the existence of intraday variation in price discovery. We propose a Likelihood Ratio test where the constant price discovery model is a restricted version of the model with time-variation. In particular, the restricted model has a constant ratio of noise and innovation variances, while this ratio can vary in the unrestricted model. Lastly, we show how this test can be used to check for constancy of information shares during specific parts of the day by defining the intraday patterns of noise and innovation variances as a concatenation of multiple flexible Fourier patterns.

Given the definition of information shares in Eq. (2.8), the constancy of the information shares implies that the shares at time t are equal to the shares at time s, for all $s, t = 1, 2, \ldots, T$, that is, $IS_{i,t} = IS_{i,s}$, for all $i = 1, \ldots, N$. Cancelling out the constant terms this equality reduces to

$$\omega_{i,s}^2/\sigma_{\epsilon,s}^2 + \sum_{j=1}^{N} \left(\omega_{i,s}^2/\omega_{j,s}^2\right) \frac{(1+\alpha_j)^2}{1+\sum_{j=1}^{Q} \psi_{i,j}^2} = \omega_{i,t}^2/\sigma_{\epsilon,t}^2 + \sum_{j=1}^{N} \left(\omega_{i,t}^2/\omega_{j,t}^2\right) \frac{(1+\alpha_j)^2}{1+\sum_{j=1}^{Q} \psi_{i,j}^2}.$$

A straightforward and economically intuitive restriction in order to satisfy this equality is to make the ratio of all variance terms constant across time points. Then the equality will hold, since the only time-varying elements are the variance ratio terms $\omega_{i,\cdot}^2/\sigma_{\epsilon,\cdot}^2$ and $\omega_{i,\cdot}^2/\omega_{j,\cdot}^2$. This restriction can be imposed in the model by equating all polynomial and flexible Fourier parameters in Eq. (2.14), i.e. $\theta_{i,p}$, $\delta_{i,q}$ and $\phi_{i,q}$, to be identical across noise and innovation variances, while allowing the variances to differ by their constant terms c_i which provide the relative magnitudes. Thus the constant information shares case can be formulated as a restricted version of the time-varying case, allowing for a Likelihood Ratio test to assess this restriction.

As an extension of the above, we may also be interested in testing the constancy of information shares during a specific part of the day, while allowing for time-variation during other parts. For this purpose, we can specify the time-varying variance processes using a concatenation of multiple flexible Fourier forms, as follows. Suppose the trading day is divided into H parts, with the h^{th} part containing T_h time points. The specification in Eq. (2.14) can then be extended to define the variance i at time point $t_h = 1, \ldots, T_h$ as

$$\zeta^2_{i,h,t_h} = \exp\left(c_{i,h} + \sum_{p=1}^{P_{i,h}} \theta_{i,h,p}\left(\frac{t_h}{T_h}\right)^p + \sum_{q=1}^{Q_{i,h}}\left(\delta_{i,h,q}\cos\left(\frac{2\pi q t_h}{T_h}\right) + \phi_{i,h,q}\sin\left(\frac{2\pi q t_h}{T_h}\right)\right)\right) \tag{2.15}$$

and the whole process across intervals is defined as

$$\zeta^2_{i,t} = \zeta^2_{i,h,t_h} \text{ for } \sum_{j=0}^{h-1} T_j + 1 \leq t \leq \sum_{j=0}^{h} T_j \text{ and } t_h = t - \sum_{j=0}^{h-1} T_j,$$

where $t = 1, ..., \sum_{j=0}^{H} T_j$ and $T_0 = 0$. Note, however, that using multiple flexible Fourier patterns may result in undesired jumps in the variance at the points of concatenation. To ensure continuity, the value of the variance at the terminal time point of the interval k, $t_k = T_k$, should be equal to the value at the period before the start of period $k+1$, $t_{k+1} = 0$. Using that $\cos(2x\pi) = 1$ and $\sin(2x\pi) = 0$ for all integers x, these values are equal to

$$\zeta^2_{i,k,T_k} = \exp\left(c_{i,k} + \sum_{p=1}^{P_k} \theta_{i,k,p} + \sum_{q=1}^{Q_{i,k}} \delta_{i,k,q}\right)$$

and

$$\zeta^2_{i,k+1,0} = \exp\left(c_{i,k+1} + \sum_{q=1}^{Q_{i,k+1}} \delta_{i,k+1,q}\right).$$

Equating these two values we get an expression for $c_{i,k+1}$ in terms of the other parameters:

$$c_{i,k+1} = c_{i,k} + \sum_{p=1}^{P_{i,k}} \theta_{i,k,p} + \sum_{q=1}^{Q_{i,k}} \delta_{i,k,q} - \sum_{q=1}^{Q_{i,k+1}} \delta_{i,k+1,q}.$$

Thus making the constant terms after the first period a function of other parameters rather than parameters to be estimated separately ensures the continuity of the generated intraday variance patterns.

2.3 Simulation study

In this section we provide simulation evidence for the ability of the proposed modeling framework to capture intraday variation in price discovery. Section 2.3.1 compares GMM and state space ML results for the case without time-variation. In Section 2.3.2 we consider a data generating process (DGP) with time-varying information shares and examine to what extent our model is able to detect such time-variation.[5] Lastly, Section 2.3.3 explores various parameter configurations and the case where the DGP differs from the model that is actually estimated.

2.3.1 Comparison of GMM and state space ML methods

We design our simulations and choose parameter values in the DGP to mimic an empirical setting in order to demonstrate the relevance of our results for empirical work. As a benchmark case, we simulate observed prices of three trading venues and a latent price process over 100 days with 391 intraday observations using Eq. (2.1) and Eq. (2.2). This corresponds to data sampled at a 1-minute frequency for a trading day between 9:30h and 16:00h. We take the noise covariance matrix Ω and the matrix of lagged noise

[5]Unlike our empirical study, we use three instead of four trading venues in the benchmark case of the simulation study to alleviate the computational burden and present the results for the case with four venues as a variation. The benchmark case with two dominant venues and a third venue with a smaller share captures the close competition of NYSE and NASDAQ sidelining the other exchanges in our empirical study.

coefficients Ψ as diagonal. The innovation variance σ_r^2 is equal to 0.816, while the noise variances in Ω take considerably smaller values of 0.016, 0.012 and 0.107. The elements of the vector α have small negative magnitudes of -0.008, -0.022 and -0.006, such that the efficient price innovations are almost but not fully incorporated into the observed prices in each period. Lastly, the diagonal elements of the Ψ matrix are set to 0.172, 0.087, and 0.270, implying a modest degree of autocorrelation in observed price changes. We generate 100 replications of these three observed price series and apply both estimation methods using the true parameters as initial values.

Panel A of Table 2.1 compares the true information shares and the estimates obtained with both GMM and the state space ML methods. The parameter settings of the DGP imply that the second venue leads price discovery with a 53.9% information share. This is followed by the first venue with a share of 39.5%, while the third venue is much less important with a 5.8% information share. The results show that on average both the GMM and the state space ML methods provide fairly accurate estimates of the information shares. At the same time, the state space ML method performs quite a bit better. The mean estimates are close to the true values, with a maximum difference of 0.8% for the GMM and only 0.2% for the state space ML case. Likewise, the estimates do not show much variation across simulations, with the maximum standard deviation at 1.5% for GMM and 1% for the state space ML method. The same conclusion also follows from Panel B of Table 2.1, showing the average and standard deviations of the root mean squared error (RMSE) for the model parameters and the three information shares. While the average RMSEs are quite small for both methods, the state space ML approach shows superior performance with a mean RMSE of 0.7% compared to 1.2% for GMM.

2.3.2 Capturing time-variation with the state space ML method

We now advance to testing our state space ML approach in the measurement of intraday variation in price discovery. Following Section 2.2.3, we allow for variation in the

Table 2.1. Simulation Results of the GMM and State Space ML Methods in the Constant Case Benchmark

The table shows summary statistics of the simulation results for the GMM and state space ML methods with constant innovation and noise variances. Three stock series are generated for 100 days, each with 391 observations, using the unobserved components model of Eq. (2.1) and Eq. (2.2). Panel A reports the summary statistics for each information share. The first column denotes the information shares for each of the simulated stocks calculated using the data generating process (DGP) parameter values. For each of the information share estimates the mean and standard deviations over all simulations are given. The results are based on 100 simulations from the corresponding data generating process. Panel B provides a more concise summary of the information share results and also provides information for the parameter estimates. The presented data consists of the means (RMSE) and of the standard deviations (SD) of root mean squared errors of the parameter estimates and the information shares. The results are based on 100 simulations from the corresponding data generating process.

Panel A: Summary Statistics for Information Share Estimates

	DGP	*GMM*		*ML*	
		Mean	*SD*	*Mean*	*SD*
IS1	39.5	38.7	1.5	39.3	0.9
IS2	53.9	54.6	1.5	54.1	1.0
IS3	5.8	5.9	0.1	5.8	0.1

Panel B: General Summary Statistics

	Parameters		*Information Shares*	
	RMSE	*SD*	*RMSE*	*SD*
ML	1.7	0.4	0.7	0.5
GMM	2.2	0.7	1.2	0.9

innovation and noise variances with the same pattern repeating each day. As before prices for three trading venues are simulated for 100 days with 391 intraday observations. The variances now fluctuate following a flexible Fourier form complemented with a polynomial function, as given in Eq. (2.14). In the benchmark DGP, each variance specification consists of 10 flexible Fourier sets and a polynomial of order 1. For brevity, we do not report all parameter settings of the polynomials and flexible Fourier sets, which have 94 parameters in total, displayed in Table 2.2. The mean of the innovation variance process $\sigma^2_{r,t}$ is 0.810 and the mean of noise variances in Ω_t have smaller values of 0.019, 0.009 and 0.103. We take the noise covariance matrix Ω_t and the matrix of lagged noise coefficients Ψ as diagonal like in the constant case. The diagonal elements of the Ψ matrix are 0.142, 0.122, and 0.210. Finally, the elements of the vector α again have small magnitudes of

-0.01, -0.02 and -0.005. We now consider 25 replications, due to the additional computation burden of the time-varying system with a large number of parameters and the great amount of variations in settings we consider.

Table 2.2. Parameter Values for the Simulation Study

The table shows the parameter values for the time-varying model consisting of Eq. (2.1), Eq. (2.2) and Eq. (2.14). The first three columns give the parameter estimates for the respective exchange groups and the fourth one refers to the process of the innovation variance σ_r^2. These are based on our estimations using the Expedia stock data from July 2 until December 28, 2007.

	NYSE	NASDAQ	REST	
α	-0.01029	-0.01961	-0.00508	
Ψ	0.14194	0.12194	0.20984	
	ω_1^2	ω_2^2	ω_3^2	σ_r^2
c	0.26570	-4.84778	-4.20136	-3.78099
θ_1	-1.04352	1.57777	-2.89436	2.14248
δ_1	0.48992	0.18142	-0.15738	-0.01820
ϕ_1	-0.15776	0.59546	-0.37510	0.39463
δ_2	0.12751	0.20959	-0.99779	0.30418
ϕ_2	-0.00771	0.40373	-0.48992	0.63969
δ_3	0.11698	0.33646	-1.15124	0.65764
ϕ_3	0.09574	0.36646	0.62914	0.58655
δ_4	0.10419	0.04929	-0.07339	0.21477
ϕ_4	0.00502	0.26494	-0.28529	0.17648
δ_5	0.04638	0.08667	-0.65573	0.48870
ϕ_5	0.00831	0.02683	0.82151	-0.25795
δ_6	0.01485	0.05297	-0.25810	-0.29911
ϕ_6	-0.02557	0.19182	-0.14621	0.33802
δ_7	0.01605	0.06739	-0.31784	0.26243
ϕ_7	0.02305	0.19266	0.02455	0.26765
δ_8	0.09063	-0.04767	-0.08683	0.24235
ϕ_8	-0.01191	0.11695	0.80992	-0.03020
δ_9	0.02632	0.01917	0.10254	0.17797
ϕ_9	0.01017	0.20337	-0.71649	0.28193
δ_{10}	0.02792	0.02218	-0.14725	0.03293
ϕ_{10}	-0.03748	0.16085	0.36835	0.08079

Figure 2.1 displays the true intraday information shares as implied by the parameter settings in the DGP (solid line), as well as the average estimates (dashed line), and minimum and maximum estimates (thin solid lines) across the 25 replications. The mean estimates are close to the true information shares throughout the entire day. Subtracting the mean estimates from the true information share values at each time point and

averaging the absolute values of these differences, we find a rather small mean absolute difference of 0.2%. The mean absolute difference of the lowest and highest estimates from the true values is also modest at 2.8%.

We evaluate a number of variations in the DGP settings, with results shown in Table 2.3. Specifically, we consider varying the number of days in the sample, the number of observations per day (the observation frequency), the number of series, the number of flexible Fourier (FF) sets, and the polynomial order. We mainly focus on lowering the number of available observations in terms of the number of days and intraday observations, because this is the direction where the results tend to worsen. Also an intraday pattern can be just temporary and we would like to capture it from as little observations as possible. In terms of the variance specifications we mostly investigate cases with more flexible Fourier sets and higher polynomial degrees, since this shows if the estimation procedure can handle a large number of parameters. The number of series under consideration reflects the usual amount of asset/exchange groups used in the literature. As in Panel B of Table 2.1, we present means and standard deviations of RMSE's for the parameter estimates and the information share estimates.

First consider the RMSE results of the benchmark case, corresponding to the information shares of Figure 2.1, to provide a context to evaluate the variations in Table 2.3. The information shares have a mean RMSE of 1.3% with a standard deviation of 1.0%. We observe an expected but limited decline of estimation accuracy compared to the constant case of Section 2.3.1, where the mean RMSE is 0.7% with a standard deviation of 0.5%.

Table 2.3 shows that a decrease in the amount of data has only a limited worsening effect on the information share estimates. Reducing the number of days from 100 to 10 increases the mean RMSE of the information shares from 1.3% to 5.8%. Likewise reducing the number of intraday observations from 771 to 71 raises the mean RMSE from 0.9% to 4.4%. These results suggest that our method can still effectively capture intraday patterns even with a limited amount of data. Similarly, estimation results improve with

Fig. 2.1. Simulation Results from the Benchmark Case of the Time-Varying Model

The figure shows summary statistics of the simulation results for information shares achieved by the state space ML method with time-varying innovation and noise variances in the flexible Fourier form. Three stock series are generated for 100 days, each with 391 intraday observations. Each figure displays for the corresponding simulated trading venue the true values, mean estimates and the upper and lower bounds containing all the estimates of information shares. The results are based on 25 simulations.

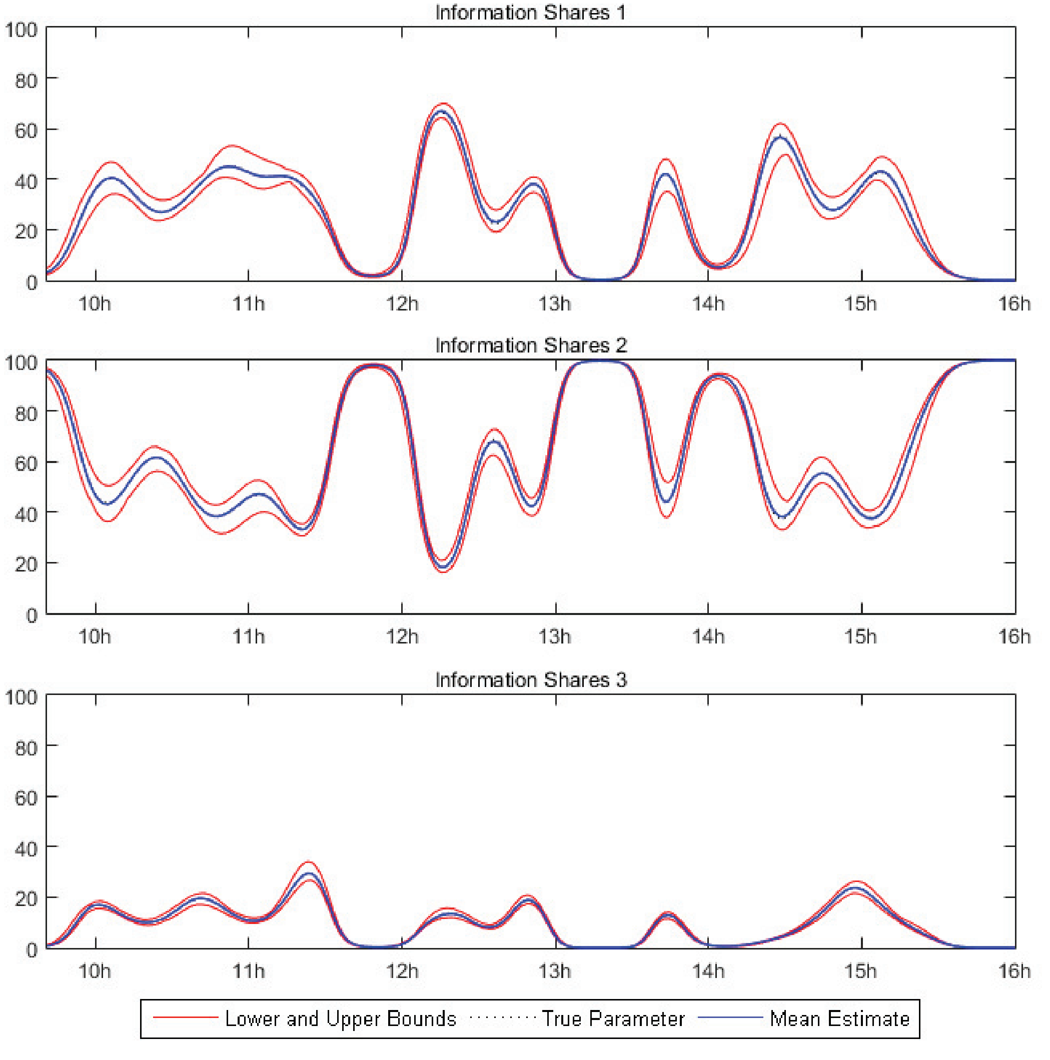


Table 2.3. Simulation Results of the Time-Varying Model

The table shows summary statistics of the simulation results from the state space ML method with time-varying innovation and noise variances. In the benchmark case, we consider a flexible Fourier type model on innovation and noise variances with 20 sets, a polynomial of order one, 3 trading venues, 100 days and with 391 observations. In the second column, the parameter setups corresponding to this benchmark case are emboldened. In the first column from top to bottom respectively the number of days, observations per period, trading venues, the flexible Fourier sets and the polynomial order are varied keeping others constant. The presented data consists of the means (RMSE) and of the standard deviations (SD) of root mean squared errors of the parameter estimates and the information shares. The results are based on 25 simulations from the corresponding data generating process.

		Parameters		*Information Shares*	
		RMSE	*SD*	*RMSE*	*SD*
Days	*10*	10.5	10.1	5.8	4.1
	25	4.6	3.7	3.1	2.2
	50	3.2	2.5	2.2	1.6
	100	1.5	0.6	1.3	1.0
Intraday Obs.	*71*	16.2	9.3	4.4	3.1
	191	3.0	2.0	2.4	1.7
	391	1.5	0.6	1.3	1.0
	771	1.2	0.8	0.9	0.6
Series	*2*	2.7	1.4	3.7	2.4
	3	1.5	0.6	1.3	1.0
	4	1.4	0.7	0.8	0.6
	5	1.5	0.9	0.6	0.5
FF Sets	*5*	1.7	0.9	0.9	0.6
	10	1.5	0.6	1.3	1.0
	15	2.0	1.7	1.4	1.2
	20	2.7	2.8	1.8	1.4
Poly. Order	***1***	1.5	0.6	1.3	1.0
	2	1.5	0.7	1.3	0.9
	3	1.9	1.0	1.6	1.5
	4	2.5	1.4	2.0	1.8

the number of observed price series. An increase from 2 to 5 series reduces the mean RMSE's from 3.7% to 0.6%.

Adding more flexible Fourier sets or increasing the polynomial order in the variance specifications increases both the complexity of the pattern to be estimated and the estimation uncertainty, but this has only a weakly worsening effect on estimation accuracy. The increase of the polynomial order from 1 to 4 adds 12 parameters, but the mean information share RMSE rises only from 1.3% to 2.0%. Similarly, the increase of the Fourier sets from 5 to 20 adds a far larger number of 120 parameters, yet the RMSE increases only from 0.9% to 1.8%.

Table 2.4 explores three types of parameter value setups under the benchmark Fourier model. In the first setup, we magnify or shrink the noise variances, keeping the innovation variances constant. In the benchmark case, the mean innovation variance is 18.3 times higher than the mean noise variance. While a 10-fold shrinkage of the noise leads to a nearly equal drop in the mean RMSE of the information shares, increase of noise has a far less dramatic effect. The mean RMSE remains nearly constant with 7.8% for the benchmark opposed to 6.3% and 8.7% for 10-fold and 100-fold magnifications, respectively. However the standard deviations of RMSE's increase from 5.6% to 20.3% and 26.7%, signaling a wider distribution of RMSE's across simulated data sets.

Table 2.4. Simulation Results with Parameter Variation

The table shows summary statistics of the simulation results for parameter variation using the flexible Fourier model as the data-generating and estimation model. We vary the ratio of the noise variances to the innovation variance, the values of the correlation vector α, and the degree of fluctuations in the intraday patterns. We use the estimated from the empirical study, unless stated otherwise. The variation in noise-innovation ratio is done by multiplying the empirical noise patterns by a given number. In the benchmark case, denoted in the table as "$\times 1$" the ratio of the mean noise to the mean innovation is 18.3. In the α variation case we report below the assigned values to respectively the NYSE, the NASDAQ and the REST group. In these two cases, the parameter setups corresponding to this benchmark case are emboldened. We play with the amount of fluctuations in the data by simulating the parameters of the flexible Fourier sets from uniform distribution with four different ranges. The flexible Fourier model we use has 10 sets, observed through 3 securities over 100 days, each with 391 observations. The summary statistics consist of the means (RMSE) and of the standard deviations (SD) of root mean squared errors of the parameter estimates and the information shares. The results are based on 25 simulations from the corresponding data generating process.

		Parameters		*Information Shares*	
		RMSE	*SD*	*RMSE*	*SD*
Noise	$\times 0.1$	0.396	0.226	0.006	0.064
	$\times \mathbf{1}$	0.118	0.268	0.078	0.056
	$\times 10$	1.075	0.245	0.063	0.203
	$\times 100$	43.918	1.325	0.087	0.267
α	**-0.01, -0.02, -0.05**	0.118	0.268	0.078	0.056
	-0.50, -0.50, -0.50	0.200	0.140	0.050	0.192
	-0.90, 0.00, -0.90	0.139	0.128	0.000	0.003
	-0.90, -0.90, 0.00	0.149	0.148	0.025	0.145
FF	U[-0.1,0.1]	0.131	0.165	0.024	0.138
	U[-0.5,0.5]	0.096	0.166	0.008	0.080
	U[-1,1]	7.909	0.528	0.028	0.131
	U[-2,2]	62.226	2.147	0.092	0.256

The second setup displayed in Table 2.4 looks at the correlation terms of innovation and noise variances, α. The slightly negative values of the α elements in the benchmark case mean that a little less than the full latent price innovation is incorporated into the observed security prices at each time point. We do not observe a sizable drop in accuracy, if half of the latent price innovation is hidden from observed prices by halving α values. To examine the interaction between noise and α values, we couple a high α with a low and then with a high noise level. In each case, the high α value is zero, while the others are set to -0.9. Both combinations lead to a dramatic improvement in the accuracy of information share estimates, the low noise being better than the high. However in both cases the parameter RMSPE statistics remain close to benchmark levels, signaling that the clear dominance these combinations assign to one market is the main driver behind the improvement. Even if the parameter estimates vary as much as in the benchmark case, all variations give the full information share to one market, leading to high accuracy and low variation in information share results. Therefore we can conclude that the parameter estimates are not affected by variations in α, although the resulting information share estimates tend to improve when one market attains a clear lead over others.

In the last panel of Table 2.4 we examine effect of the complexity of intraday variations. We use the flexible Fourier coefficients δ and ψ to play with the level of fluctuations. Generating the values of these coefficients used in the data generating process from a uniform distribution with a wider support increases the intraday variation. The coefficient values of the benchmark are close to results from a support of $[-0.5, 1]$. Higher complexity impairs the accuracy to a limited degree, as the widest support of $[-2, 2]$ has a mean RMSE of only 9.2% compared to 2.8% for $[-1, 1]$ and 7.8% of the benchmark. Estimates from the set-up with wider supports, which arguably lead also to unreasonable amounts of intraday variation, do not converge in most cases.

2.3.3 Capturing time-variation with a misspecified model

In this last part of our simulation study, we consider the effects of differences between specifications of the DGP and the model that is actually estimated. Firstly, we estimate DGP's with variances following a flexible Fourier form using models in the same form, but with correct and incorrect numbers of Fourier sets. Secondly, we introduce a state space model with a step function specification for the time-variation in variances and compare the estimation accuracy across models with and without time-varying variances.

Table 2.5 displays the mean RMSE's of the information shares under various cases where the DGP differs from the model. Panel A demonstrates that intraday variation in the error variances can be quite accurately captured as long as the number of Fourier sets in the model is at least as large as that of the DGP. We report nine setups with 5, 10 and 15 Fourier sets. For example, in the first row of the table we consider data generated using a flexible Fourier specification for the variances with 5 sets, and the columns represent the RMSE of the information shares in models with 5, 10 and 15 flexible Fourier sets, respectively. The mean RMSE's are below 8% as long as the estimation model uses an equal or larger number of Fourier sets compared to the DGP. The model with 15 Fourier sets has low mean RMSE's of 2.8%, 3.0%, and 5.6% for DGP's with respectively 5, 10 and 15 sets.

In practice, selecting the number of Fourier sets is an important part of the model specification. Here we examine the usefulness of the Schwarz Information Criterion (SIC) for this purpose. We find the SIC to be quite accurate, albeit having a small tendency towards overfitting. For each simulation with a DGP containing 5, 10 or 15 Fourier sets, we record the number of Fourier sets in the estimation model preferred by the Schwarz criterion and report its mean across simulations in the last column of Panel A of Table 2.5. The biggest difference between the number of parameters in the DGP and the estimation model occurs for the case with 5 Fourier sets where on average 6.1 sets are selected by the Schwarz criterion. The difference is even smaller for the other two cases: On average

Table 2.5. Simulation Results for Misspecified Models

The table shows summary statistics of the simulation results for cases where the data generating process (DGP) does not necessarily correspond to the estimation model. The settings of the DGP's are given in the leftmost column and those of the estimation models in the top row. In Panel A both the DGP and the estimation model are of flexible Fourier type, but have different numbers of Fourier sets. Panel B reports results for three DGP's, where variances are modeled as a constant, as a 20-period step function (SF), and as a Fourier model with 10 sets and a polynomial of order one (FF). These are estimated using the GMM method for the constant case and using ML for three state space models, i.e. constant, step function with 20 periods (SF) and Fourier with 10 sets and a polynomial of order one (FF). In each case 25 data sets are generated from the DGP and for each data sets the estimation model is started for 10 trials with random initial values. All simulated data sets span 100 days with 391 intraday observations for 3 price series. On the rightmost column in Panel A we report the mean of the number of Fourier sets preferred by the Schwarz criterion across simulations with DGPs of 5, 10 or 15 Fourier sets. The rest of the presented numbers are the means of root mean squared errors of the information shares.

Panel A: Flexible Fourier with different numbers of sets

		Estimated Model			Preferred
		5	*10*	*15*	FF Num.
DGP	*5*	1.4	3.5	2.8	6.1
	10	16.1	7.8	3.0	10.0
	15	26.9	22.0	5.6	15.4

Panel B: Estimations across models

		Estimated Model			
		Constant		*Time-Varying*	
		GMM	*ML*	*SF*	*FF*
DGP	*Constant*	31.9	18.5	17.3	16.1
	Time-Varying – SF	58.7	21.5	2.7	3.2
	Time-Varying – FF	52.3	31.5	15.0	7.8

15.4 sets are preferred for the DGP with 15 Fourier sets and in all simulations the correct number of parameters are chosen for the DGP with 10 sets.

Note that, in contrast to the previous sections we now use random parameter values instead of the true ones to initialize the numerical likelihood optimization, because no true initial values exist when the DGP and the model differ. In order to guard against the possibility of ending up in a local maximum of the likelihood function, we consider ten different sets of starting values for each replication.[6] The effect of using random initial values instead of true ones can be observed from the results for the case where both the

[6]The initial values are drawn from a uniform distribution. The support is [-1, 1] for the elements of α, [0, 1] for the elements of Ψ considering the positive autocorrelation of the data, and [-2, 2] for the parameters of the flexible Fourier form. The log innovation variances of the step function and constant models have a support of [0, 2] and the log noise variances [-5, 0].

DGP and the model have 10 Fourier sets, as this corresponds with the situation considered in Section 2.3.2. The mean RMSE of the information shares increases from 1.3% using true initial values to a substantially higher 7.8% for random initial values.

Panel B of Table 2.5 considers three different DGP's, namely a constant model, a step function model and the flexible Fourier model. For each DGP, we use GMM to estimate a model with constant information shares, and the state space ML method to estimate models with constant, step function and flexible Fourier specifications for the variances. We use the step function model as a simpler functional form for time-variation, where the noise and innovation variances stay at different constant levels for 20 periods per day.

The estimation models without time-variation display a low accuracy even for DGP's of their own kind. Whereas in Section 2.3.1 the GMM and the state space ML methods have similar levels of accuracy in the case of no time-variation, the use of random initial values gives a lead to the state space ML method with a mean information share RMSE of 18.5% to 31.9%. The step function and Fourier models have even better results at estimating the DGP without time variation giving mean information share RMSE's of 17.3% and 16.1%, respectively.

The performance difference between the GMM and the state space ML methods widens under time-varying DGP's, although both constant estimation models are incapable of fully capturing such patterns.[7] The Fourier model can capture time-varying patterns relatively better than the step function. Both the Fourier and the step function models have low mean RMSE's at estimating the step function DGP, with respectively 3.2% to 2.7%. However under the Fourier DGP, the Fourier model's 7.8% mean RMSE is nearly the half of the 15.0% mean RMSE of the step function.

[7]Part of the favorable performance of our methodology comes from the superiority of the ML estimation over GMM when the estimation model is correctly specified. However the constant state space ML estimates remain considerably more accurate than the GMM estimates even if the model is misspecified, i.e. the DGP has time-varying variances and the estimation model has constant ones.

2.4 Intraday variation of price discovery in S&P 500 stocks

We apply our methodology in an empirical setting involving prices of 50 S&P 500 stocks observed at different trading venues during the second half of 2013. Section 2.4.1 presents some key properties of the data. Section 2.4.2 discusses the results from the state space ML method. Section 2.4.3 investigates the stock- and exchange-specific determinants of the intraday variation of the information shares.

2.4.1 Data and summary statistics

We examine the prices of S&P 500 stocks over 124 trading days from July 1 until December 31, 2013. The stocks trade from 9:30h to 16:00h (New York time) at a number of exchanges. Our high-frequency data set consists of all quotes on all these exchanges, as obtained from the Trades And Quotes (TAQ) database. We use the midquote prices to avoid the bid-ask bounce present in transaction prices. The TAQ database time-stamps the quotes to the nearest second. We sample the data at the 1-minute frequency by using the midquote at the end of each minute.

We use a selection procedure based on market capitalization in order to obtain a representative sample of 50 stocks. We rank the constituents of the S&P 500 index on December 31^{st} 2013 according to their market capitalizations and select the stocks ranked 1^{st}, 11^{th}, 21^{st} etc. In case of problems related to data availability or estimation, we check the stock with the nearest market capitalization. This procedure provides us a representative sample of 50 stocks not only in terms of size, but also concerning other characteristics. Our sample has 38 NYSE-listed stocks, while the number of NYSE-listed stocks in the S&P 500 index is 39.1 when the total number of stocks in the index is scaled to 50. The sectoral distribution of firms in our sample is similarly close to that of all index constituents for the seven sectors in our sample: Manufacturing (22 in our sample vs 20.0 in the index), Utilities (8 vs 6.4), Finance (6 vs 8.3), Services (6 vs 5.7), Trade (5 vs 5.5), Mining (2 vs 3.0) and Construction (1 vs 0.7). Table 2.6 reports the market capitalization, industrial classification and the exchange listing of the 50 stocks we have chosen from the S&P 500 index.

Next, we arrange the data into groups according to quote origin. This aggregation of individual exchanges into groups aids both the model parsimony in estimation and the interpretation of the results. We consider the four biggest exchange groups and the remaining regional exchanges at the time: NYSE (TAQ codes A, N and P), NASDAQ (TAQ codes B, Q, T and X), BATS (TAQ codes Y and Z), Direct Edge (TAQ codes J and K) and the remaining exchanges (TAQ codes C, I, M and W). We generate the midquote sequences at the 1-minute frequency for each exchange and pick the one with the smallest bid-ask spread at the end of the sampling interval as the midquote of the exchange group. For example in the case of the NYSE group, we generate the midquote sequence of each of the three exchanges with codes A, N and P and pick the midquote of the exchange with the smallest spread at the end of each minute as the midquote of the NYSE exchange group.[8] The last exchange group consisting of small regional exchanges hosts only 2.1% of the total trade volume for our sample, as can be seen in Tables 2.7 and 2.8, and in preliminary information share analyses generated expectedly dismal results even below their share in the trade activity. Therefore we excluded them from the study.

Table 2.7 reports the averages of the number of trades and trade volumes for each of the 50 stocks and Table 2.8 provides these statistics for various stock groups. For all stocks, the leading market share in the number of trades and trade volumes belongs to either NYSE or NASDAQ. In terms of trade volumes the exchange with the highest market share is always the exchange where the stock is listed. However NASDAQ has a higher number of trades than NYSE for 16 out of 38 NYSE-listed stocks, while leading

[8] An alternative method is picking the highest best bid and the lowest best ask of the group exchanges and computing their midquote as the group midquote. However in our study this method led to a considerable amount of cases where the best bid is higher than the best ask due to the price staleness at some of the exchanges. Therefore we opted to pick the midquote corresponding to the smallest spread, which ensures a non-negative spread and usually belongs the more liquid and thus more frequently updated exchanges. Both the smallness of the spread and proximity of the most recent quote update enhance the precision of the midquote itself.

Table 2.6. Descriptive Statistics

This table reports descriptive statistics for 50 constituents of the S&P 500 index on December 31^{st}, 2013.

Ticker Symbol	Company Name	Market Cap	Listing	Industry
AAPL	Apple Inc	500,680,634	NASDAQ	Manufacturing
ADBE	Adobe Systems Inc	29,715,612	NASDAQ	Services
ADT	A D T Corp	8,165,389	NYSE	Services
AEE	Ameren Corp	8,773,682	NYSE	Utilities
AIG	American International Group Inc	75,163,263	NYSE	Finance, Insurance and Real Estate
AKAM	Akamai Technologies Inc	8,425,452	NASDAQ	Services
ATI	Allegheny Technologies	3,847,399	NYSE	Manufacturing
BK	Bank of New York Mellon Corp	40,129,359	NYSE	Finance, Insurance and Real Estate
CA	C A Inc	14,912,704	NASDAQ	Services
CL	Colgate Palmolive Co	60,332,879	NYSE	Manufacturing
CMI	Cummins Inc	26,412,703	NYSE	Manufacturing
COP	Conocophillips	86,553,244	NYSE	Manufacturing
CTL	Centurylink Inc	18,825,611	NYSE	Utilities
DO	Diamond Offshore Drilling Inc	7,913,872	NYSE	Mining
DPS	Dr Pepper Snapple Group Inc	9,775,327	NYSE	Manufacturing
DTE	D T E Energy Co	11,737,553	NYSE	Utilities
FDO	Family Dollar Stores Inc	7,391,052	NYSE	Retail Trade
FDX	Fedex Corp	44,889,020	NYSE	Utilities
FIS	Fidelity National Info Svcs Inc	15,628,019	NYSE	Services
FLIR	Flir Systems Inc	4,237,538	NASDAQ	Manufacturing
GT	Goodyear Tire & Rubber Co	5,914,800	NASDAQ	Manufacturing
IP	International Paper Co	21,750,836	NYSE	Manufacturing
IVZ	Invesco Ltd	16,135,392	NYSE	Finance, Insurance and Real Estate
JCI	Johnson Controls Inc	34,738,103	NYSE	Construction
JPM	JP Morgan Chase & Co	219,837,373	NYSE	Finance, Insurance and Real Estate
KMX	Carmax Inc	10,496,181	NYSE	Retail Trade
KR	Kroger Company	20,418,273	NYSE	Retail Trade
LOW	Lowes Companies Inc	51,820,827	NYSE	Retail Trade
MJN	Mead Johnson Nutrition Co	16,918,264	NYSE	Manufacturing
MOS	Mosaic Company New	14,049,637	NYSE	Manufacturing
MRO	Marathon Oil Corp	24,591,180	NYSE	Manufacturing
MSI	Motorola Solutions Inc	17,463,128	NYSE	Manufacturing
MU	Micron Technology Inc	23,011,500	NASDAQ	Manufacturing
MWV	Meadwestvaco Corp	6,569,108	NYSE	Manufacturing
NKE	Nike Inc	55,959,988	NYSE	Manufacturing
NRG	N R G Energy Inc	9,288,508	NYSE	Utilities
OKE	Oneok Inc New	12,826,926	NYSE	Utilities
ORLY	O Reilly Automotive Inc New	13,635,538	NASDAQ	Retail Trade
PM	Philip Morris International Inc	139,596,724	NYSE	Manufacturing
POM	Pepco Holdings Inc	4,777,851	NYSE	Utilities
QEP	Q E P Resources Inc	5,494,963	NYSE	Mining
SIAL	Sigma Aldrich Corp	11,187,190	NASDAQ	Manufacturing
SYMC	Symantec Corp	16,306,843	NASDAQ	Services
TRV	Travelers Companies Inc	32,962,717	NYSE	Finance, Insurance and Real Estate
TWC	Time Warner Cable Inc	38,196,095	NYSE	Utilities
UTX	United Technologies Corp	104,420,832	NYSE	Manufacturing
VLO	Valero Energy Corp New	27,193,925	NYSE	Manufacturing
WY	Weyerhaeuser Co	18,397,607	NYSE	Finance, Insurance and Real Estate
XLNX	Xilinx Inc	12,253,982	NASDAQ	Manufacturing
XRAY	Dentsply International Inc New	6,898,704	NASDAQ	Manufacturing

Table 2.7. Summary Statistics of Trading Activity

This table reports summary statistics of the trading activity in five exchange groups for the period of July-December 2013. We provide the mean number of trades and trade volumes and their distribution across the exchange groups for 50 S&P 500 stocks.

Ticker	Ann.	Number of Trades						Trade Volume					
Symbol	Return	Total.	NYSE	NSDQ	BATS	DE	Rest	Total.	NYSE	NSDQ	BATS	DE	Rest
AAPL	86.0	49,530	21.5	43.6	14.6	19.4	0.9	7,007,650	20.4	42.5	12.6	21.2	3.4
ADBE	58.1	16,121	32.1	30.5	19.2	17.2	1.0	2,298,914	42.5	26.2	15.4	14.8	1.2
ADT	49.8	16,057	26.5	30.4	27.1	14.7	1.3	3,031,996	43.1	24.1	20.2	11.5	1.0
AEE	25.9	15,235	29.3	34.0	18.9	16.2	1.6	2,578,035	43.1	27.9	13.9	13.6	1.6
AIG	23.7	52,741	24.2	36.5	22.7	15.3	1.2	11,134,156	37.8	30.2	15.9	13.5	2.6
AKAM	-0.9	18,427	37.4	30.2	14.2	17.3	1.0	2,875,779	49.2	24.1	10.8	14.9	1.1
ATI	43.4	13,010	33.6	30.3	13.9	21.5	0.7	1,785,250	43.3	26.1	11.3	18.4	0.8
BK	26.2	12,503	38.1	31.4	14.3	15.0	1.2	1,878,909	49.9	25.7	10.7	12.7	1.0
CA	108.9	11,512	37.0	33.6	12.8	15.9	0.7	1,532,132	42.1	29.9	11.1	15.3	1.6
CL	48.1	10,746	37.6	32.9	12.5	16.2	0.8	1,410,731	41.7	29.3	10.7	15.0	3.2
CMI	129.5	70,916	15.5	34.9	25.7	22.6	1.3	25,138,135	14.9	39.8	23.0	19.5	2.7
COP	66.4	16,236	17.0	45.7	18.9	16.9	1.5	2,303,065	15.5	50.5	16.4	16.0	1.6
CTL	29.5	13,450	29.9	31.3	22.0	15.2	1.5	2,272,877	45.6	24.6	16.1	12.5	1.2
DO	61.9	7,145	41.4	28.1	11.8	18.0	0.7	899,407	46.3	24.5	10.1	18.1	0.9
DPS	-17.6	15,324	25.6	33.9	21.8	17.0	1.7	2,744,866	39.8	27.9	15.4	15.3	1.6
DTE	11.0	8,364	38.0	30.8	12.9	17.3	0.9	1,062,110	43.7	27.8	11.4	16.0	1.1
FDO	30.8	14,246	32.3	32.4	20.2	13.5	1.7	2,281,533	44.0	27.6	15.7	11.3	1.3
FDX	-21.2	20,077	31.1	30.0	14.6	23.0	1.3	3,345,721	40.0	25.4	11.0	20.0	3.6
FIS	35.0	7,977	35.5	31.3	15.7	16.2	1.3	1,098,794	43.7	26.8	12.6	14.6	2.4
FLIR	10.1	5,503	30.9	32.1	12.5	22.6	1.9	717,349	38.9	28.5	10.5	20.5	1.6
GT	1.4	6,163	36.9	32.4	12.6	16.4	1.8	785,354	43.4	28.2	10.7	15.4	2.2
IP	127.5	7,128	37.2	31.8	13.1	16.6	1.3	943,138	43.2	27.3	11.0	15.3	3.1
IVZ	13.2	7,245	28.6	35.3	15.9	19.0	1.2	976,684	37.9	30.4	13.3	17.3	1.1
JCI	19.1	12,703	28.2	35.9	20.9	13.8	1.2	1,985,964	38.7	31.1	16.6	12.4	1.1
JPM	38.0	3,936	15.7	51.1	17.2	14.5	1.4	486,487	14.9	55.0	15.3	13.6	1.2
KMX	0.2	11,258	31.6	32.9	15.2	18.3	1.9	1,642,795	38.4	29.3	12.3	17.2	2.8
KR	20.6	5,064	16.3	49.4	16.8	16.1	1.4	645,865	15.9	52.1	14.8	15.8	1.3
LOW	-31.1	6,006	40.7	29.5	12.4	16.3	1.1	767,383	46.1	25.8	10.6	16.4	1.0
MJN	32.6	18,823	32.2	28.5	22.3	16.0	1.0	2,883,464	44.1	23.0	16.8	14.9	1.3
MOS	27.7	33,219	24.7	33.8	24.2	16.1	1.3	6,088,594	34.4	28.0	18.1	16.3	3.3
MRO	44.7	26,485	26.6	35.1	20.9	15.8	1.6	4,352,269	37.1	30.7	16.2	14.3	1.7
MSI	101.9	15,131	30.5	36.9	15.7	15.5	1.4	2,465,498	44.5	29.2	11.8	13.2	1.2
MU	24.2	10,076	35.6	24.1	19.3	20.2	0.7	1,302,009	43.6	21.5	16.3	17.9	0.7
MWV	2.8	20,731	28.8	31.0	22.7	15.9	1.6	3,402,445	42.3	25.2	17.2	13.6	1.7
NKE	17.6	15,033	30.1	35.5	17.1	15.8	1.5	2,241,117	39.8	30.7	13.8	14.2	1.5
NRG	9.8	23,405	14.9	38.2	25.9	19.7	1.3	4,278,350	14.2	43.6	22.9	17.8	1.5
OKE	41.4	11,576	15.5	43.2	26.4	13.6	1.3	1,743,323	14.2	49.3	22.8	12.5	1.2
ORLY	33.6	3,105	16.0	51.7	13.1	18.5	0.7	370,316	15.0	55.2	11.7	17.2	0.9
PM	20.1	9,720	17.0	45.6	18.7	16.9	1.8	1,270,706	16.4	47.7	16.4	17.6	2.0
POM	5.1	6,384	37.7	32.5	12.1	16.6	1.1	832,845	41.7	28.5	10.2	15.4	4.2
QEP	12.8	3,815	38.4	30.8	14.7	14.3	1.7	525,301	48.2	25.9	11.8	12.5	1.6
SIAL	134.3	15,445	19.8	40.4	20.2	18.6	1.1	2,542,583	18.5	44.2	17.1	18.5	1.7
SYMC	82.1	6,953	32.5	30.5	17.0	18.7	1.4	922,754	40.7	26.6	14.0	17.6	1.2
TRV	113.9	28,847	25.7	33.6	21.0	18.2	1.5	4,652,865	35.9	28.7	16.4	17.1	2.0
TWC	25.3	3,681	19.6	52.6	13.3	13.6	0.9	431,667	18.6	54.2	12.0	13.5	1.7
UTX	31.6	14,086	16.9	44.3	23.0	14.5	1.3	1,897,997	16.3	47.9	19.8	14.6	1.4
VLO	1.8	5,297	40.2	30.4	14.1	14.6	0.7	632,983	46.0	26.9	12.2	14.2	0.7
WY	18.1	8,367	34.9	30.5	14.5	18.4	1.7	1,172,442	42.5	25.9	11.6	16.3	3.7
XLNX	51.8	5,007	38.5	29.1	15.0	15.5	1.9	687,741	48.2	24.4	11.9	14.1	1.5
XRAY	-6.8	7,044	29.8	28.2	26.8	14.0	1.2	1,216,200	44.9	22.2	19.9	12.0	1.1

Table 2.8. Summary Statistics

This table reports summary statistics of the trading activity in five exchange groups for the period of July-December 2013. We provide the mean number of trades and trade volumes and their distribution across the exchange groups for various aggregations of our data set of 50 S&P 500 stocks. The second column reports the number of stocks used for each aggregation group.

Group Name	#	Number of Trades Total.	NYSE	NSDQ	BATS	DE	Rest	Trade Volume Total.	NYSE	NSDQ	BATS	DE	Rest
All	50	14,937	26.6	35.1	19.6	17.5	1.3	2,630,891	31.9	33.1	16.7	16.3	2.1
by Listing:													
NYSE-Listed	38	13,688	30.7	32.5	18.8	16.7	1.3	2,195,484	41.0	27.5	14.7	14.9	1.9
NASDAQ-Listed	12	18,892	17.4	40.9	21.3	19.2	1.2	4,009,679	16.0	42.7	20.2	18.7	2.4
by Market Cap:													
Market Cap 1	12	23,265	28.1	34.9	19.1	16.7	1.1	3,856,654	37.3	30.0	15.0	15.5	2.2
Market Cap 2	12	19,675	24.6	34.2	21.3	18.6	1.4	4,258,259	27.2	34.5	19.0	17.2	2.1
Market Cap 3	12	10,178	26.1	36.2	19.1	17.2	1.3	1,538,353	29.7	36.1	16.5	15.9	1.9
Market Cap 4	14	7,817	28.0	36.0	17.4	17.1	1.4	1,121,811	33.6	33.9	14.6	16.1	1.9
by Industry:													
Manufacturing	22	16,610	25.9	35.1	18.9	18.8	1.2	3,121,782	27.9	34.7	17.3	17.8	2.2
Utilities	8	9,625	32.2	33.2	17.6	15.9	1.2	1,430,337	41.2	28.4	14.2	14.5	1.7
Finance	6	23,596	26.8	33.6	22.6	15.7	1.3	4,402,720	39.0	28.1	16.7	13.9	2.3
Services	6	12,867	19.6	40.1	21.3	17.4	1.5	1,987,663	20.0	43.1	18.8	16.4	1.7
Trade	5	11,233	29.3	34.7	18.7	15.7	1.5	1,735,002	39.4	29.8	14.9	14.0	1.9
Mining	2	7,187	37.4	30.1	13.6	17.5	1.4	969,913	43.9	25.9	11.2	16.3	2.7
Construction	1	15,131	30.5	36.9	15.7	15.5	1.4	2,465,498	44.5	29.2	11.8	13.2	1.2

in all of its 12 listed stocks. While NASDAQ's 33.1% share in the trade volumes is only slightly above the 31.9% share of the NYSE, it has a 8.5% lead in terms of the number of trades.

Figure 2.2 shows the intraday distribution of four market quality measures averaged across trading days for each exchange group: The number of trades, trade volume, quoted spreads and volatility measured by the square root of the mean squared midquote change at each minute. The number of trades and trade volume show the familiar U-shaped pattern that is well-documented in the literature. NASDAQ has the largest number of trades, followed by the NYSE. The NYSE is comparable to NASDAQ in terms of trade volume, implying larger average trade sizes in the NYSE compared to NASDAQ. By contrast the quoted spreads and volatility follow an inverted J-shaped pattern, peaking at the market open but not increasing towards the market close. The NYSE has the largest spreads, followed by BATS and NASDAQ which are quite close to each other. The NYSE also has a larger volatility at the start of the day until about 11:00h. We observe two types of jumps in the intraday patterns of these variables. All share the jumps at 10:00h due to macroeconomic announcements and at 14:00h following FOMC announcements.[9] Another type of jumps specific to the number of trades and trade volumes occur at hour and half-hour transitions.

2.4.2 Estimation Results

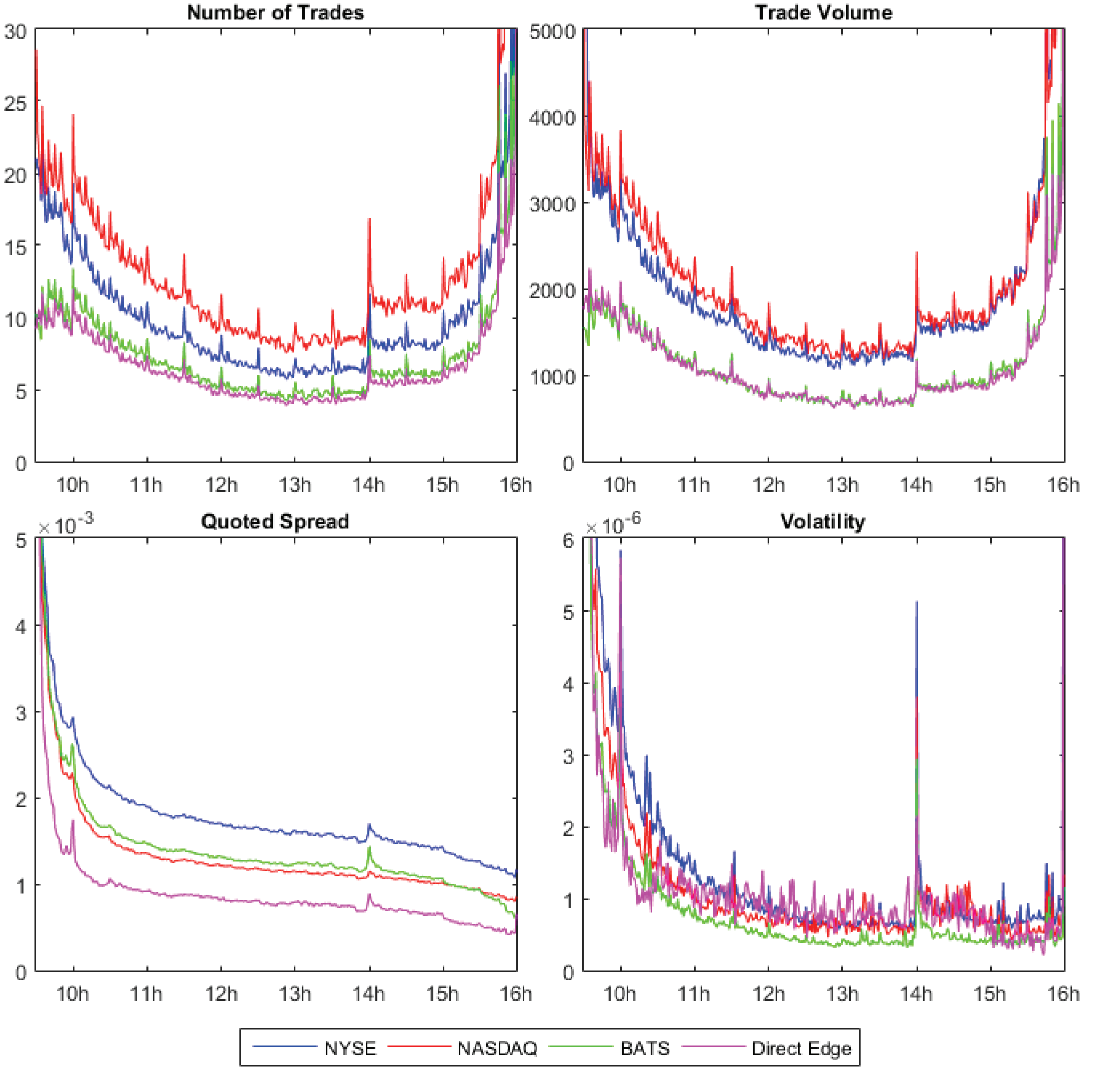
We estimate the unobserved components model given by Eq. (2.1) and Eq. (2.2) using the GMM method and the version allowing for higher degrees of serial correlation by substituting Eq. (2.2) with Eq. (2.12) using the state space ML method. In the state space case we define three intraday periods implementing the concatenation presented in Eq. (2.15): the market open (9:30h-10:30h), the midday (10:30h-15:00h) and the market close (15:00h-16:00h). Although the GMM method does not provide time-varying estimates and the simulation study finds the state space ML estimation to be more accurate, the GMM method still presents a valuable point of comparison in an empirical setting, especially due to its weak distributional assumptions. These methods have different data requirements and treatment of overnight returns. The covariance matrices that serve as the input for the GMM method are computed via log price differences and exclude overnight returns. In contrast, the state space ML estimation uses directly the log prices and excludes not only overnight returns but also the first 3 of 390 intraday observations when calculating the likelihood.

We select the model specification (in terms of the number of lags in Eq. (2.12) and the polynomial order and number of Fourier terms in Eq. (2.14)) based on both statistical tests for model selection and the dependence of the final results on the model. We start with the most basic setup without any polynomial and flexible Fourier terms and noise

[9] In our data set, Treasury budget announcements are also released at 14:00h. However, as will analyze in more detail in Section 2.4.2, FOMC announcements generate probably all of the market volatility at this time. Therefore we use FOMC announcements and the 14:00h news interchangeably.

Fig. 2.2. Market Quality Variables

The figure displays intraday averages of four market quality measures for each exchange group. The averages are computed using the data for 50 S&P 500 stocks in the period of July-December 2013.



lags, i.e. a model where L is zero in Eq. (2.12) and P and Q are zero in Eq. (2.14). We use 10 estimation trials with random initial values and choose the estimates with the largest likelihood. We then increase L, P and Q, while checking if the direction of increase leads to an improvement in the Schwarz Information Criterion. All variances in the same intraday interval share the same P and Q, which decreases the amount of possible estimations and is also required for our testing framework presented in Section 2.2.4. We find that for most stocks in our sample the estimated variance patterns tend to stabilize when the number of Fourier sets reaches $Q = 5$. Including more Fourier terms leads to negligible changes in the information share estimates.[10] Therefore we limit Q to be at most 10 to avoid overfitting[11] and to alleviate the computational burden.[12] For the cases reaching this maximum we check the effect of further increasing the number of Fourier terms on the information shares to rule out any significant dependence of the results on the specifications of the model.

Table 2.9 reports the parameter and information share estimates from both methods, providing daily means for time-varying variables. The elements of α and the diagonals of the Ω and of Ψ matrices have similar qualitative rankings and close magnitudes across the two estimation methods. The estimates point to several interesting properties of the price discovery process. The small estimates of α suggest that price innovations are almost fully incorporated within a minute. Also the noise in observed price changes is relatively small compared to genuine innovations, with the innovation variance being about 20 times larger than the mean noise variance. Lastly, the estimates of the diagonal elements of Ψ imply only a modest level of autocorrelation in the one-minute returns.

[10] Details on the chosen model specification for each stock can be found in Table 2.10.

[11] The simulation study concerning model misspecification in Section 2.3.3 finds a small tendency of the Schwarz Information Criterion to choose overfitted models, though such overfitted models also tend to have a higher estimation accuracy.

[12] The estimation has considerable computational requirements. Providing an average figure for computation time across all estimations was not possible, because we used a computer cluster consisting of many machines with varying specifications. However we estimated a smaller sample on a machine with an Intel Core i5-2410M (2.30GHz) processor to provide a representative figure. Each estimation trial of a model with three intraday periods using 10 flexible Fourier terms and a polynomial of order one in their variance specifications took 23.5 hours on average.

Table 2.9. Estimation Results

This table reports the mean values for the GMM and state space ML estimation results for various groups of 50 S&P 500 stocks. The estimates belong to the information shares, the innovation and noise variances, the under/over-reaction coefficients α and the lag coefficients ψ. We report the averages of these estimates for various stock groups. Panels A and B report these group averages for the daily means of the time-varying ML estimates and the GMM estimates, respectively.

Panel A: Averages of Time-varying State Space ML Estimates

Group	Information Shares				Variance					α				ψ			
Name	NYSE	NSDQ	BATS	DE	Inn.	NYSE	NSDQ	BATS	DE	NYSE	NSDQ	BATS	DE	NYSE	NSDQ	BATS	DE
All	43.4	33.4	14.5	7.2	0.330	0.011	0.044	0.047	0.103	-0.004	-0.005	-0.001	-0.002	0.032	0.074	0.062	0.061
NYSE-Listed	49.7	31.4	12.2	5.6	0.317	0.007	0.043	0.051	0.107	-0.002	-0.002	0.000	0.001	0.018	0.068	0.064	0.061
NASDAQ-Listed	23.3	39.7	22.0	12.3	0.371	0.026	0.046	0.032	0.088	-0.007	-0.012	-0.003	-0.011	0.075	0.092	0.056	0.060
Market Cap 1	45.4	33.7	14.2	6.0	0.263	0.004	0.018	0.033	0.094	0.004	0.000	0.006	0.003	0.019	0.042	0.049	0.051
Market Cap 2	45.3	31.1	13.4	9.5	0.377	0.004	0.021	0.027	0.044	-0.002	-0.002	0.000	0.000	0.012	0.062	0.063	0.047
Market Cap 3	36.8	40.7	13.7	6.4	0.296	0.019	0.032	0.062	0.144	-0.006	-0.006	-0.003	-0.007	0.060	0.077	0.064	0.077
Market Cap 4	45.6	28.8	16.6	6.9	0.377	0.016	0.095	0.062	0.126	-0.010	-0.009	-0.004	-0.003	0.035	0.108	0.069	0.069
Manufacturing	43.5	31.7	14.6	8.4	0.355	0.013	0.055	0.042	0.086	-0.003	-0.005	0.001	-0.001	0.034	0.071	0.063	0.060
Utilities	46.9	32.1	14.8	4.8	0.327	0.009	0.050	0.083	0.242	-0.003	-0.005	-0.004	-0.009	0.024	0.079	0.066	0.081
Finance	45.0	31.3	17.8	5.5	0.272	0.002	0.014	0.009	0.030	0.000	0.000	0.003	0.004	0.010	0.058	0.044	0.041
Services	29.6	44.0	16.3	9.3	0.302	0.010	0.010	0.032	0.052	-0.007	-0.008	-0.004	-0.002	0.063	0.050	0.063	0.055
Trade	42.6	37.9	11.4	5.2	0.284	0.021	0.051	0.060	0.113	-0.007	-0.005	-0.004	-0.006	0.036	0.102	0.069	0.071
Mining	62.7	22.2	7.4	6.2	0.447	0.010	0.065	0.091	0.118	-0.007	-0.005	-0.003	0.004	0.015	0.100	0.074	0.066
Construction	51.8	28.0	10.8	9.0	0.317	0.003	0.068	0.014	0.021	-0.003	-0.001	0.002	0.004	-0.011	0.126	0.033	0.034

Panel B: GMM Estimates Assuming Constant Information Shares

Group	Information Shares				Variance					α				ψ			
Name	NYSE	NSDQ	BATS	DE	Inn.	NYSE	NSDQ	BATS	DE	NYSE	NSDQ	BATS	DE	NYSE	NSDQ	BATS	DE
All	53.9	23.9	14.9	5.7	0.319	0.009	0.028	0.037	0.213	-0.004	-0.009	0.000	0.024	0.004	0.073	0.045	0.135
NYSE-Listed	61.9	20.3	12.2	4.3	0.305	0.006	0.029	0.040	0.205	0.000	0.000	0.003	0.030	-0.016	0.061	0.047	0.146
NASDAQ-Listed	28.6	35.3	23.2	10.5	0.362	0.018	0.025	0.026	0.240	-0.015	-0.036	-0.008	0.006	0.069	0.112	0.037	0.103
Market Cap 1	55.3	24.2	13.8	6.0	0.249	0.004	0.015	0.022	0.087	0.015	0.001	0.016	0.030	0.018	0.029	0.045	0.243
Market Cap 2	55.6	21.5	15.2	6.9	0.366	0.004	0.013	0.019	0.069	0.003	0.002	0.006	0.027	-0.069	0.053	0.011	0.054
Market Cap 3	44.8	32.3	15.8	4.9	0.286	0.015	0.021	0.051	0.465	-0.013	-0.014	-0.007	0.026	0.042	0.082	0.043	0.127
Market Cap 4	59.0	18.6	14.7	5.3	0.365	0.013	0.058	0.052	0.229	-0.018	-0.021	-0.012	0.017	0.024	0.121	0.076	0.120
Manufacturing	52.0	24.6	14.3	7.3	0.347	0.010	0.036	0.034	0.187	-0.004	-0.013	0.003	0.032	0.020	0.059	0.066	0.095
Utilities	64.4	18.2	12.8	2.8	0.304	0.008	0.029	0.060	0.546	-0.001	-0.005	-0.010	0.040	-0.003	0.129	0.052	0.120
Finance	59.4	19.4	17.9	2.8	0.263	0.002	0.008	0.008	0.064	0.010	0.009	0.019	0.031	-0.051	-0.019	0.033	0.424
Services	34.9	36.3	19.7	8.2	0.294	0.007	0.009	0.024	0.046	-0.014	-0.016	-0.008	0.000	0.055	0.073	-0.031	0.075
Trade	51.5	27.2	14.0	4.5	0.275	0.017	0.031	0.049	0.237	-0.008	-0.011	-0.008	-0.010	-0.007	0.120	0.059	0.115
Mining	67.8	15.9	8.7	5.9	0.423	0.010	0.043	0.075	0.139	-0.006	0.000	0.003	0.019	0.035	0.076	0.045	0.110
Construction	76.8	6.1	13.1	3.6	0.313	0.002	0.025	0.012	0.047	-0.007	0.003	0.011	0.025	-0.242	0.244	-0.014	-0.070

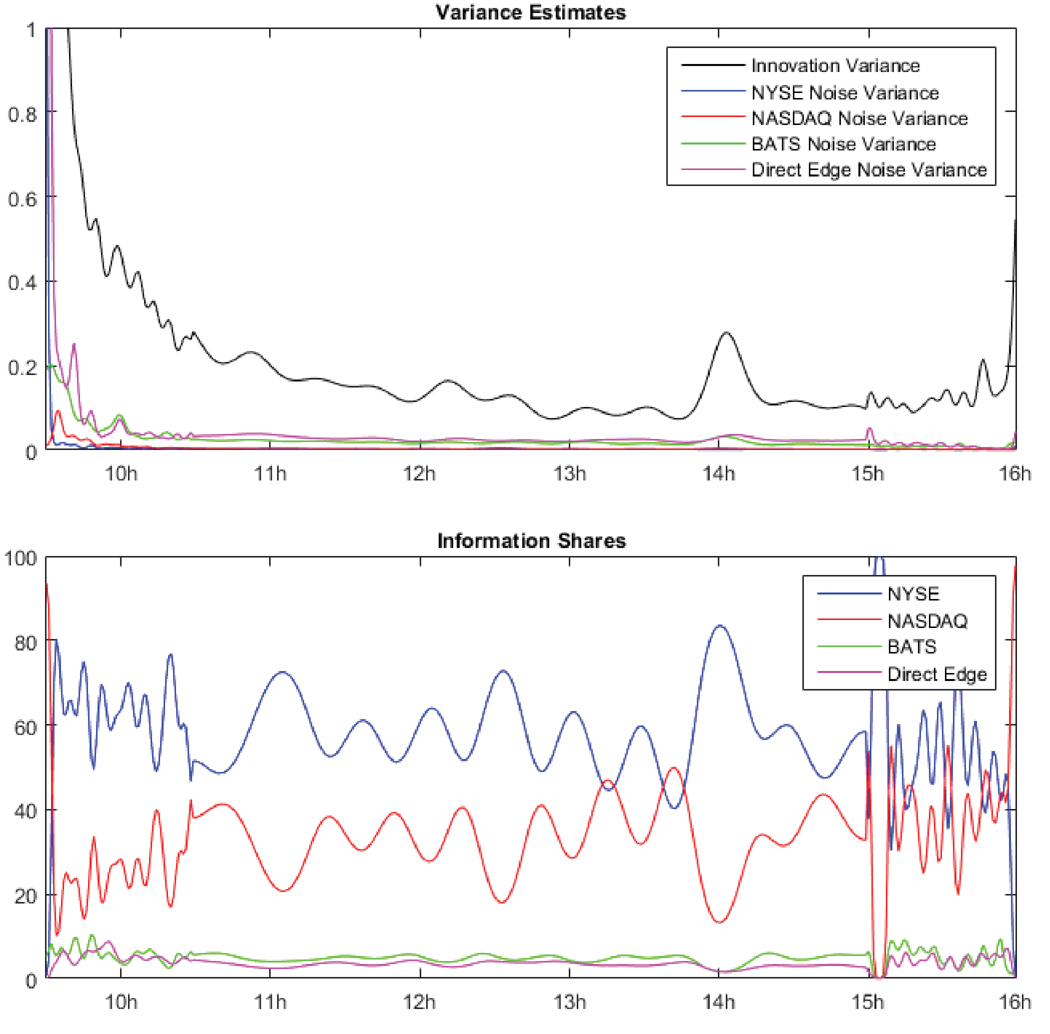
The differences in the information shares across the estimation methods stem mainly from the estimates of the idiosyncratic error variances in Ω. The negligible effect of the differences in the estimates of the innovation variance σ_r^2 can be observed from the sum of information shares, which is very similar across methods. As Eq. (2.8) shows, elements of α are summed with one in the expression for the information shares. Thus the differences of α estimates across methods have a negligible influence on the information shares, because they are very close to each other and to zero. Similarly, the diagonal elements of the Ψ matrix are squared to compute the information shares, reducing their overall impact to almost nil. Therefore we observe that the (differences in) information shares are inversely proportional to the (differences in) noise variance estimates.

We present the results for the Philip Morris International (PM) stock as an example in Figure 2.3. The NYSE acts as the primary listing exchange for PM and, similar to 37 of the other 49 stocks, the listing exchange leads the daily average contributions to price discovery. The intraday information share panel in Figure 2.3 shows that the NYSE has a considerable lead in price discovery: The average information share of the NYSE group is 58.2% compared to NASDAQ's 32.8%. Our GMM estimates give a similar lead to the NYSE with 62.1% versus 28.2%. The NYSE achieves this large lead although it does not dominate the trade activity of PM shares. Unlike most other cases its trading activity is not dominated by the exchange of its listing: 47.7% of its trade volume occurs on NASDAQ, while the rest is almost equally distributed between the other three exchange groups.

The influence of listing in contributions to price discovery and the remarkable informativeness of the NYSE, in spite of its lower share in trade activity, are part of a general pattern in our data set. As reported in Table 2.9, the NYSE accounts for 49.7% of the price discovery for stocks listed on it, while NASDAQ leads with a 39.7% information share among its listing group. This causes an overall lead of the NYSE across the whole sample with 43.4% over NASDAQ's 33.4%. This follows from the lower noise variances

Fig. 2.3. Variance and Information Share Estimates for the PM Stock

The figure displays innovation and noise variance estimates and computed information shares for the PM stock.



of exchanges when a stock is listed in it, as depicted in Figure 2.4. The figures across different industry groups follow from their compositions in terms of exchange listings. The GMM method gives qualitatively similar results with a 5–10% difference in estimates.

We observe considerable fluctuations in the estimates of variances and information shares, but several level differences are worth highlighting. The innovation and noise variances of the PM stock follow a U-shaped pattern. This pattern has a sound theoretical basis, as the informed trading literature documents such a U-shaped intraday innovation variance with a large peak at the start of the day (Admati and Pfleiderer, 1988; Foster and Viswanathan, 1993; Slezak, 1994). This pattern experiences expected information-related bumps at announcement times for macroeconomic news: the considerable jump at the FOMC announcements at 14:00h and a smaller jump at 10:00h which is more visible for other stocks as depicted in Figure 2.4. The noise variance of NASDAQ rises after 14:00h for both listing groups to levels even above BATS and Direct Edge, showing the inability for NASDAQ to remain informative at pricing FOMC news. The resulting information share estimates presented in Figure 2.5 display an increase in the NYSE informativeness at 14:00h for both listing-groups, albeit at different levels.

For the PM stock, the NYSE has a clear leadership at the market open, while the market close has a tighter competition between the two leading exchanges. Across all stocks, the NYSE group has a 48.5% information share during the first half hour of the trading day compared to NASDAQ's 30.7% share, while during the last half hour NASDAQ leads with 46.0% over 36.0%. This pattern is again highly dependent on the listing, as can be seen in Figure 2.5: NASDAQ dominates both the market open and close periods within its listing group and the NYSE's larger share stems from its competitive position in the afternoon even for the NASDAQ-listed stocks and its dominance among the NYSE-listed stocks.

In order to test whether these intraday variations in price discovery are statistically significant, we use the methodology outlined in Section 2.2.4. We examine particularly the

Fig. 2.4. Variance Estimates for All Stocks and Listing Groups

The figure displays mean intraday innovation and noise variance estimates for all stocks and for two listing groups. Out of 50 stocks, 38 are NYSE-listed and 12 are NASDAQ-listed. For each exchange group we average the minute-level intraday estimates across stocks.

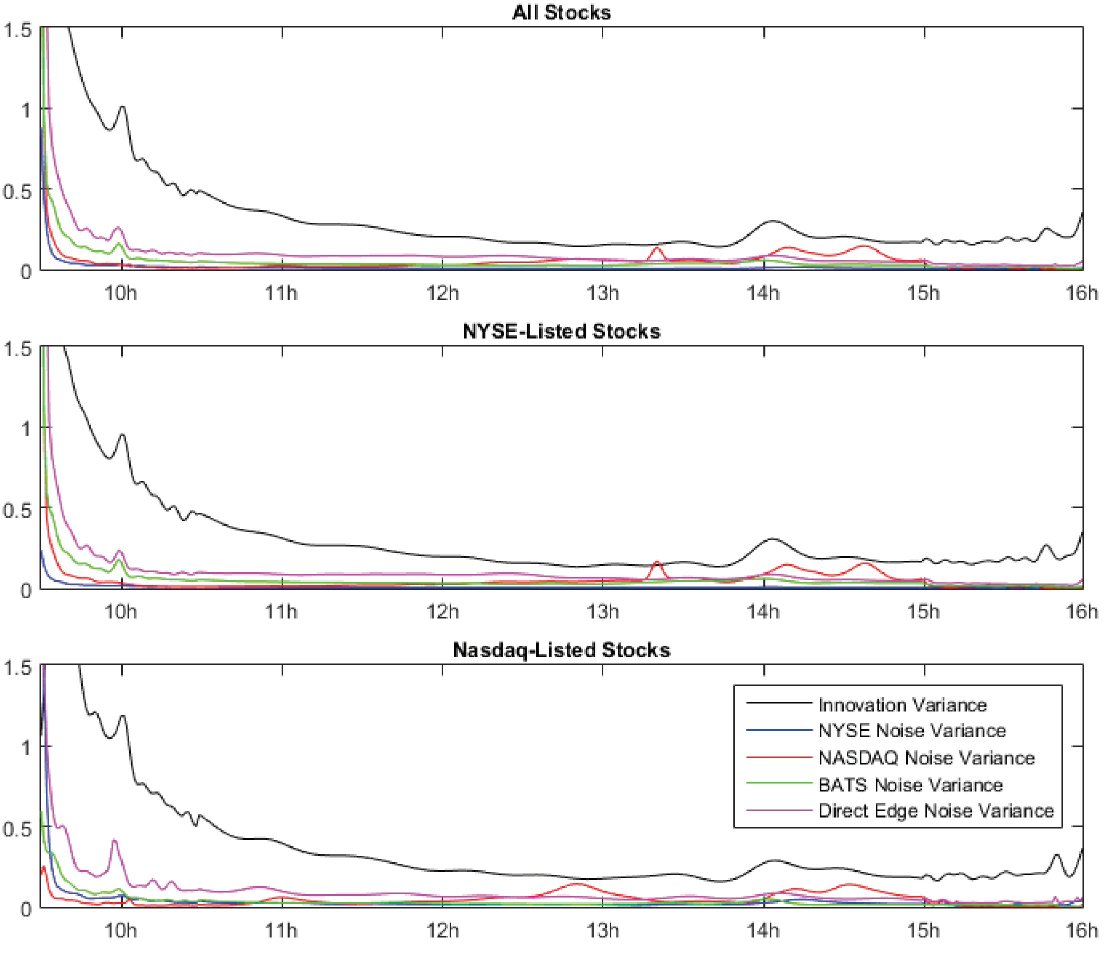
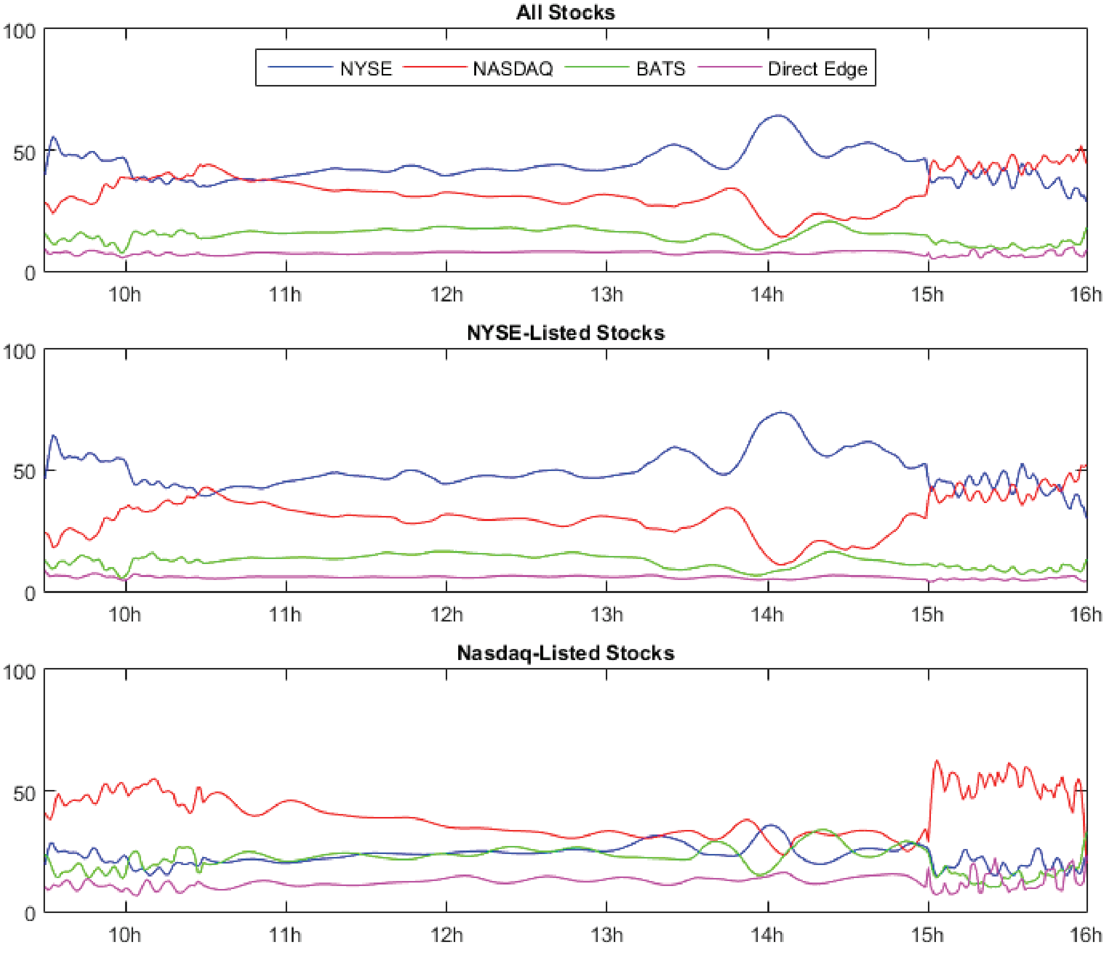


Fig. 2.5. Information Share Estimates for All Stocks and Listing Groups

The figure displays mean intraday information share estimates for all stocks and for two listing groups. Out of 50 stocks, 38 are NYSE-listed and 12 are NASDAQ-listed. For each exchange group we average the minute-level intraday estimates across stocks.



midday period (10:30h-15:00h), because as the literature suggests and as we have shown in this study market open and close experience the pricing of a larger magnitude of information compared to the rest of the day. Thus, the midday can be hypothesized to be a calmer and less competitive period. Table 2.10 reports the Lagrange Ratio test statistics for evaluating the hypothesis of constant information shares in the midday (10:30h-15:00h). For each security we report the number of flexible Fourier sets at the three parts of the day, the log-likelihood of the non-constant model estimation, the Lagrange Ratio test statistic, the number of parameters in the non-constant model and the number of restrictions in the constant model. In each case the hypothesis of constancy is rejected at 5% significance level. Since our results indicate a strong influence of macroeconomic news announcements on the shifts of informativeness across exchanges, we also run this test excluding all days with news announcements during the midday period. Even in that case, the hypothesis of constant price discovery during the midday is rejected.

The influence of macroeconomic news is worth further elaboration, given the coincidence of spikes in the innovation variance with their usual release times. We look into the case of the 38 NYSE-listed stocks due to the demonstrated capabilities of the NYSE at pricing new macroeconomic information. For this subset of stocks, we estimate the same structural model representing one latent price process observed in four stock exchanges with noise terms. The data set is divided into three subsamples based on the existence of major macroeconomic news announcements: 20 days with news at 10:00h, 4 days with news at 14:00h and 100 days without news at neither 10:00h nor 14:00h.[13] While a wide range of news items are released on the 20 days with 10:00h news, all of the four cases of 14:00h news belong to FOMC meeting announcements. This focus on major indicators makes our results comparable to the literature on macroeconomic news and provides a starting point for examining the effects of these major indicators as well

[13]An alternative setup to estimate the effect of news announcements would allow for discontinuities in the variance patterns. We do not employ such a methodology in this study, but the concatenation of multiple flexible Fourier patterns presented in Section 2.2.4 provides a straightforward approach to incorporate discontinuities at predetermined times.

Table 2.10. Summary Statistics for the Lagrange Ratio tests

This table presents the summary statistics for the Lagrange Ratio test for the constancy of information shares in the midday (10:30h-15:00h). For each security we report the number of flexible Fourier sets at the three parts of the day, the log-likelihood of the non-constant model estimate, the test statistic, the number of parameters in the non-constant model and the number of restrictions in the constant model.

Ticker	Num. of FF sets			Log	LR-test	Num. of	Num. of
Symbol	Open	Midday	Close	Likelihood	Statistic	Parameters	Restrictions
AAPL	10	8	2	23427.3	934.3	228	68
ADBE	10	9	5	121350.1	839.0	268	76
ADT	10	9	9	187025.1	467.6	308	76
AEE	10	10	10	121037.6	571.3	328	84
AIG	10	10	8	259864.0	503.4	308	84
AKAM	10	8	7	101741.1	873.6	278	68
ATI	10	10	3	113051.6	851.3	258	84
BK	10	9	5	119181.8	537.9	268	76
CA	9	8	5	-5453.0	791.8	248	68
CL	10	10	8	-65247.3	1062.3	308	84
CMI	10	6	3	134016.1	862.8	218	52
COP	10	9	5	121929.1	813.2	268	76
CTL	9	10	7	125175.7	226.4	288	84
DO	10	10	5	-25635.3	507.0	278	84
DPS	9	7	6	204732.6	679.3	248	60
DTE	10	10	5	-8476.4	576.4	278	84
FDO	10	10	2	108984.5	167.4	248	84
FDX	10	9	8	32372.4	336.3	298	76
FIS	10	10	10	55914.8	951.2	328	84
FLIR	9	10	4	42180.5	132.7	258	84
GT	8	9	3	-39149.6	971.3	228	76
IP	10	9	10	-74609.8	185.7	318	76
IVZ	10	9	3	58300.0	776.6	248	76
JCI	10	10	8	44112.9	617.3	308	84
JPM	10	9	9	-33025.4	325.6	308	76
KMX	10	9	4	18876.1	679.2	258	76
KR	8	10	3	-2862.1	239.2	238	84
LOW	9	6	5	-11754.5	749.8	228	52
MJN	9	6	3	178319.3	678.2	208	52
MOS	9	10	4	173453.0	173.0	258	84
MRO	9	9	4	162460.2	163.9	248	76
MSI	10	10	4	113215.1	269.6	268	84
MU	9	10	2	99902.3	136.7	238	84
MWV	8	8	2	148352.8	533.2	208	68
NKE	9	9	2	96540.2	939.8	228	76
NRG	10	8	2	204572.3	715.4	228	68
OKE	10	10	5	151923.3	637.2	278	84
ORLY	10	10	4	-18807.9	980.9	268	84
PM	10	9	3	6131.4	205.3	248	76
POM	9	8	9	-21848.6	1006.1	288	68
QEP	8	9	5	-18615.8	215.6	248	76
SIAL	9	10	9	20779.0	634.1	308	84
SYMC	10	10	2	169184.9	260.2	248	84
TRV	9	10	4	-28656.6	676.4	258	84
TWC	10	10	5	69107.2	121.6	278	84
UTX	10	8	3	74038.5	864.7	238	68
VLO	9	10	4	-47454.2	965.8	258	84
WY	8	8	7	108469.7	1014.8	258	68
XLNX	10	10	2	-14690.7	1104.0	248	84
XRAY	10	10	3	-18114.1	622.2	258	84

as other macroeconomic news items in more detail.[14]

Figures 2.6 and 2.7 display the intraday patterns of the noise and innovation variances, and information shares, respectively, for these three subsamples. The 14:00h bump is fully explained by the FOMC announcement, while the 10:00h bump endures even after excluding days with major macroeconomic announcements at that time. After excluding FOMC announcements, the third subsample without major news releases still contains all of the Treasury Budget announcements during our sample period. However, we observe a very flat pattern at 14:00h, which indicates that the Treasury Budget announcements do not have much impact on the stock markets.[15] By contrast, the estimates for the third subsample still exhibit a bump at 10:00h, demonstrating that the exclusion of major macroeconomic news announcements does not fully smooth out the variance bump around the release time. As in the case of the Treasury Budget, some of such minor indicators may not be influential. However, others have a sizable effect on the market dynamics, shown by the persistence of the 10:00h variance bump in spite of the exclusion of most well-known news releases. This result highlights the potential importance of even minor macroeconomic indicators and the necessity of carefully conditioning the analysis on expected news events.

For a more detailed examination of the effects of major macroeconomic news releases, we use the estimates from the third subsample without news as a benchmark. The relative difference of the estimates using the data sets with news set against this benchmark presents a straightforward descriptive statistic that gives insight into the effect of the macroeconomic announcements at shaping intraday volatility patterns. Note that this statistic measures how much larger the effect of major indicators is relative to the re-

[14] We use the major news releases categorized as 'market moving indicators' by the Econoday website. This category consists of 21 indicators, including GDP, FOMC decisions, industrial production, inflation, the employment situation and home sales. A detailed list is available at:
http://mam.econoday.com/resource_ctr_why.asp

[15] Note that this is not a result of the Treasury Budget announcements being in line with market expectations. We have data about the range of the market forecasts and the median forecast in two of the five Treasury Budget announcements and, in both cases, the realization is not only away from the median, but also outside the forecast range.

Fig. 2.6. Macroeconomic News and Variance Estimates of NYSE-listed Stocks

The figure displays mean intraday innovation and noise variance estimates for 38 NYSE-listed stocks using three subsamples based on the existence of macroeconomic news announcements: 20 days with news at 10:00h, 4 days with news at 14:00h and 100 days without news at 10:00h and 14:00h. For each exchange group we average the minute-level intraday estimates across stocks.

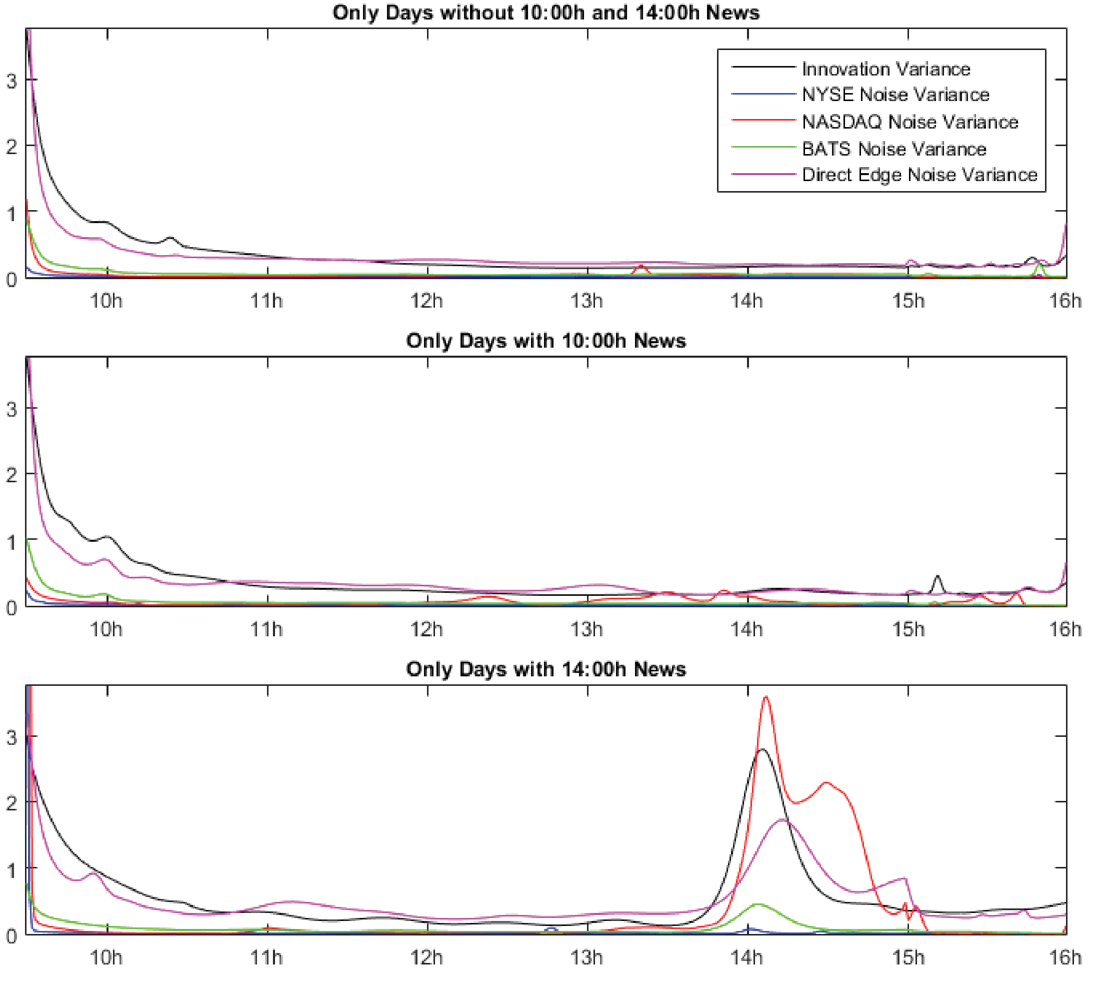
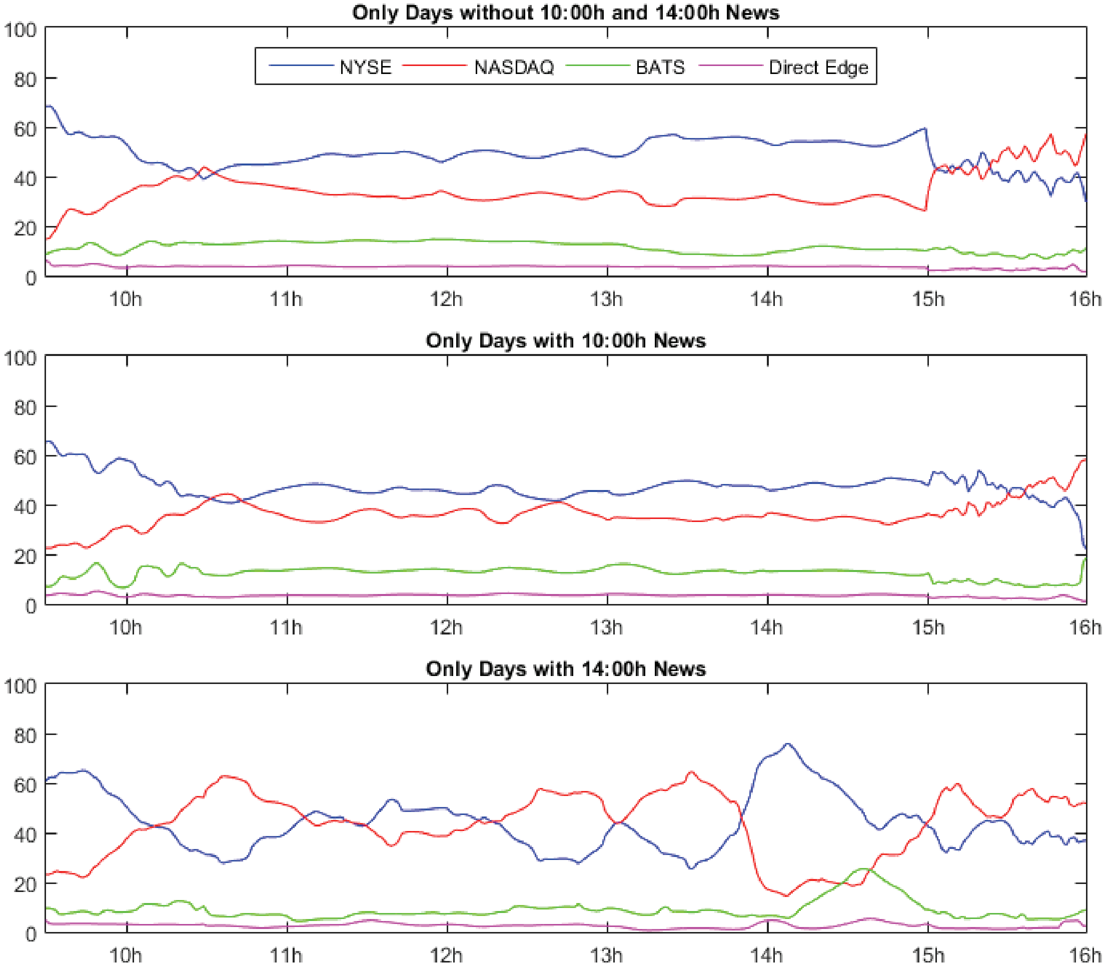


Fig. 2.7. Macroeconomic News and Information Share Estimates of NYSE-listed Stocks

The figure displays mean intraday information share estimates for 38 NYSE-listed stocks using three subsamples based on the existence of macroeconomic news announcements: 20 days with news at 10:00h, 4 days with news at 14:00h and 100 days without news at 10:00h and 14:00h. For each exchange group we average the minute-level intraday estimates across stocks.



maining indicators released at the same time. The effect of these major news releases on the market can be simply measured by comparing pre- and post-release volatility and information share figures within the results of each subsample.

News announcements at 10:00h considerably increase variances as well as the NYSE's share in price discovery around the release time. The innovation and noise variances are respectively 25.4% and 31.5% to 48.3% larger than the benchmark at the release time. The peak point of this relative difference for the innovation variance during the first hour of the trading day is also reached at 10:00h, while the peaks of the noise variances occur in the minutes after the release time. The existence of 10:00h news interrupts the trend towards the equalization of the NYSE and NASDAQ information shares after the dominance of the NYSE during the market open. The peak of the relative difference for information shares during the first hour is reached by the NYSE at 10:03h with 15.1%, which coincides with a trend of underperformance for price discovery in NASDAQ bottoming at 10:06h with -22.1%.

The FOMC announcements generate larger fluctuations in variances and, to a lesser degree, in information shares, benefitting again the NYSE to the detriment of NASDAQ. At 14:00h innovation and noise variances are respectively 14.5 times and 5.0 to 47.6 times larger than the benchmark without news announcements and all of them keep rising in the following minutes. The NYSE gains an even larger lead in price discovery compared to the 10:00h news: At 14:00h the NYSE information share is 32.3% larger than the benchmark, whereas the NASDAQ information share is 46.8% smaller. The NYSE tends to expand its share in the following minutes at the expense of NASDAQ. Although the estimation error due to the small sample size may have played a role in their magnitude, the variance estimates during FOMC announcements remain comparable to the large innovation variances during the market open, while dwarfing the variances across the trading day, and they reflect the market sensitivity to FOMC announcements especially during the taper tantrum[16] of summer 2013 and its aftermath.

We now turn to the results by industry groups. The intraday variance estimates for the four biggest industry groups, namely Manufacturing, Utilities, Finance and Services, are depicted in Figure 2.8.[17] All confirm the U-shaped pattern with 10:00h and 14:00h jumps related to macroeconomic news. The biggest 14:00h jump occurs in the Finance group: The maximum innovation variance reached within the half hour after 14:00h is 178.6% larger than the prior average between 12:30h and 13:30h. This is followed by a 114.0% jump in the Utilities group. Parallel to the increase in innovation variance, all exchanges across all industry groups experience hikes in noise variances around the two news release times. In all groups except for Manufacturing NASDAQ has the biggest relative increases in noise at 14:00h. The NYSE has the lowest relative increases only for the two smaller groups of Finance and Services, but in absolute terms it has the smallest maximum noise around 14:00h.

As can be seen in Table 2.9, there are considerable differences across industry groups in terms of the mean information shares, but qualitatively all groups except Services share the lead of the NYSE over NASDAQ followed by BATS and Direct Edge. The leadership of NASDAQ in the Services group may be driven by the listing effect, given that 4 out of 6 stocks in Services are listed on NASDAQ. The intraday patterns of information shares displayed in Figure 2.9 also resemble the listing-based differences in Figure 2.5 in terms of the dominant exchanges at the open, midday and close and in terms of the effect of macroeconomic news to the informativeness of exchanges. In particular, the low NYSE noise variance around 14:00h leads to a remarkable increase of the NYSE information shares around this time.

[16]The taper tantrum refers to the period of market uncertainty, following Fed chairman Ben Bernanke's comments in May 2013 about a possible reduction in the rate of its bond purchases, a part of the Fed's quantitative easing program.

[17]The association of stocks with industry groups is achieved using their Standard Industrial Classification (SIC) codes. Codes from 1000 to 1499 are grouped under Mining, from 1500 to 1799 under Construction, from 2000 to 3999 under Manufacturing, from 4000 to 4999 under Utilities, from 5000 to 5999 under Trade, from 6000 to 6799 under Finance, and from 7000 to 8999 under Services.

Fig. 2.8. Variance Estimates for Industry Groups

The figure displays mean intraday innovation and noise variance estimates for four industry groups with the highest numbers of stocks: Manufacturing (22), Utilities (8), Finance (6) and Services (6). For each exchange group we average the minute-level intraday estimates across stocks.

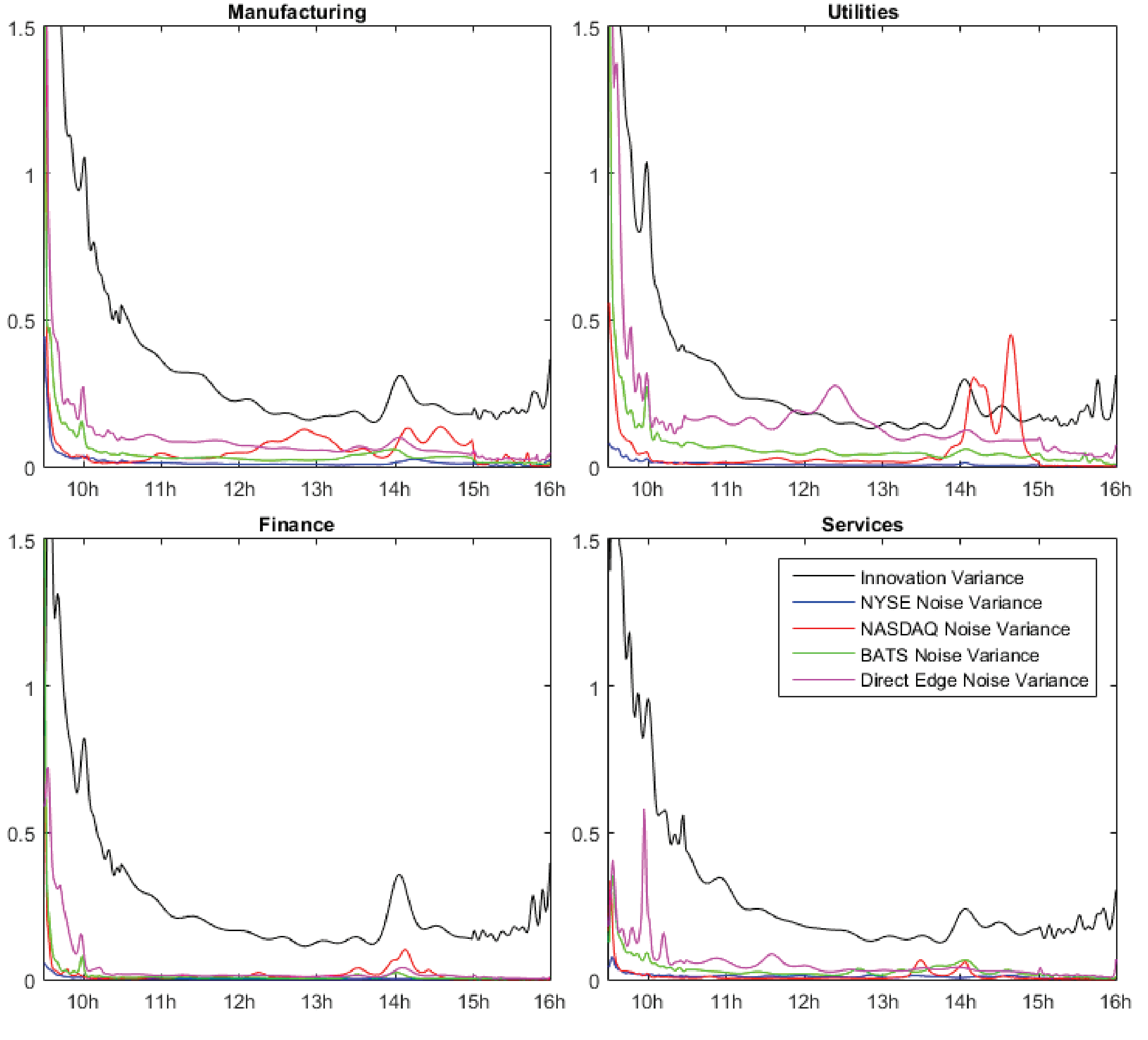
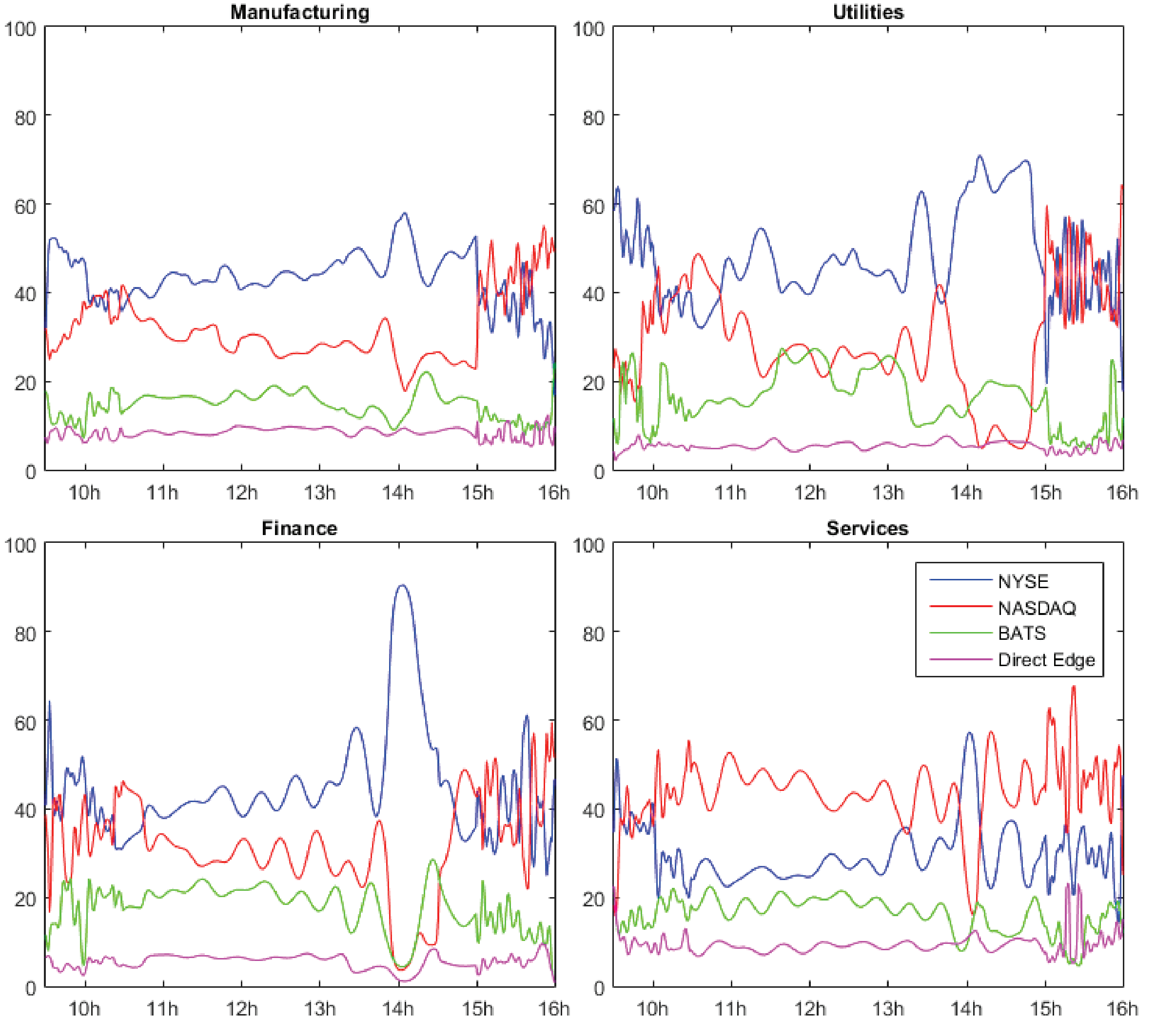


Fig. 2.9. Information Share Estimates for Industry Groups

The figure displays mean intraday information share estimates for four industry groups with the highest numbers of stocks: Manufacturing (22), Utilities (8), Finance (6) and Services (6). For each exchange group we average the minute-level intraday estimates across stocks.



Lastly, we analyze intraday price discovery in terms of market capitalization. We divide the 50 stocks into four groups, where the first three groups have 12 stocks each and the fourth group with the smallest firms has 14 stocks. Figure 2.10 displays the average variance patterns for the resulting groups. Noise variances tend to be higher for smaller stocks, while the innovation variances are much more similar. The main exception to this tendency among the two biggest exchange groups is the peak of NYSE noise variance for the third market capitalization group. As can be seen in Figure 2.11, this leads to a higher NASDAQ share in price discovery for this group. As in the case of the Services industry group, this may be a consequence of listing decisions, because the third group has the highest share of NASDAQ-listed stocks (5 out of 12).

Taken all together, we observe a significant amount of intraday variation in the informativeness of the four exchange groups. The averages of our time-varying information share estimates coincide with the GMM estimates of constant information shares, putting the NYSE in the lead for all stocks and NASDAQ (the NYSE) for the stocks listed on NASDAQ (the NYSE). The NYSE's contribution to price discovery increases considerably during major macroeconomic news releases at 10:00h and 14:00h, mainly to the detriment of NASDAQ's. The exclusion of days with these prominent news items does not necessarily flatten the innovation variance patterns at release times, possibly signalling the importance of some of the minor macroeconomic indicators. NASDAQ information shares tend to be larger within its listing group at the market open and close. The NYSE also leads the price discovery at the market open within its listing group, but its share at the close does not surpass NASDAQ's. Lastly, our tests reject the constancy of information shares during the midday even when we exclude the days with macroeconomic announcements.

Fig. 2.10. Variance Estimates for Market Capitalization Groups

The figure displays mean intraday innovation and noise variance estimates for four market capitalization groups. We rank our stocks according to their market capitalizations of December 31^{st} 2013. We divide them into four groups, where the first three groups have 12 and the 4^{th} group with the smallest firms has 14 stocks. For each exchange group we average the minute-level intraday estimates across stocks.

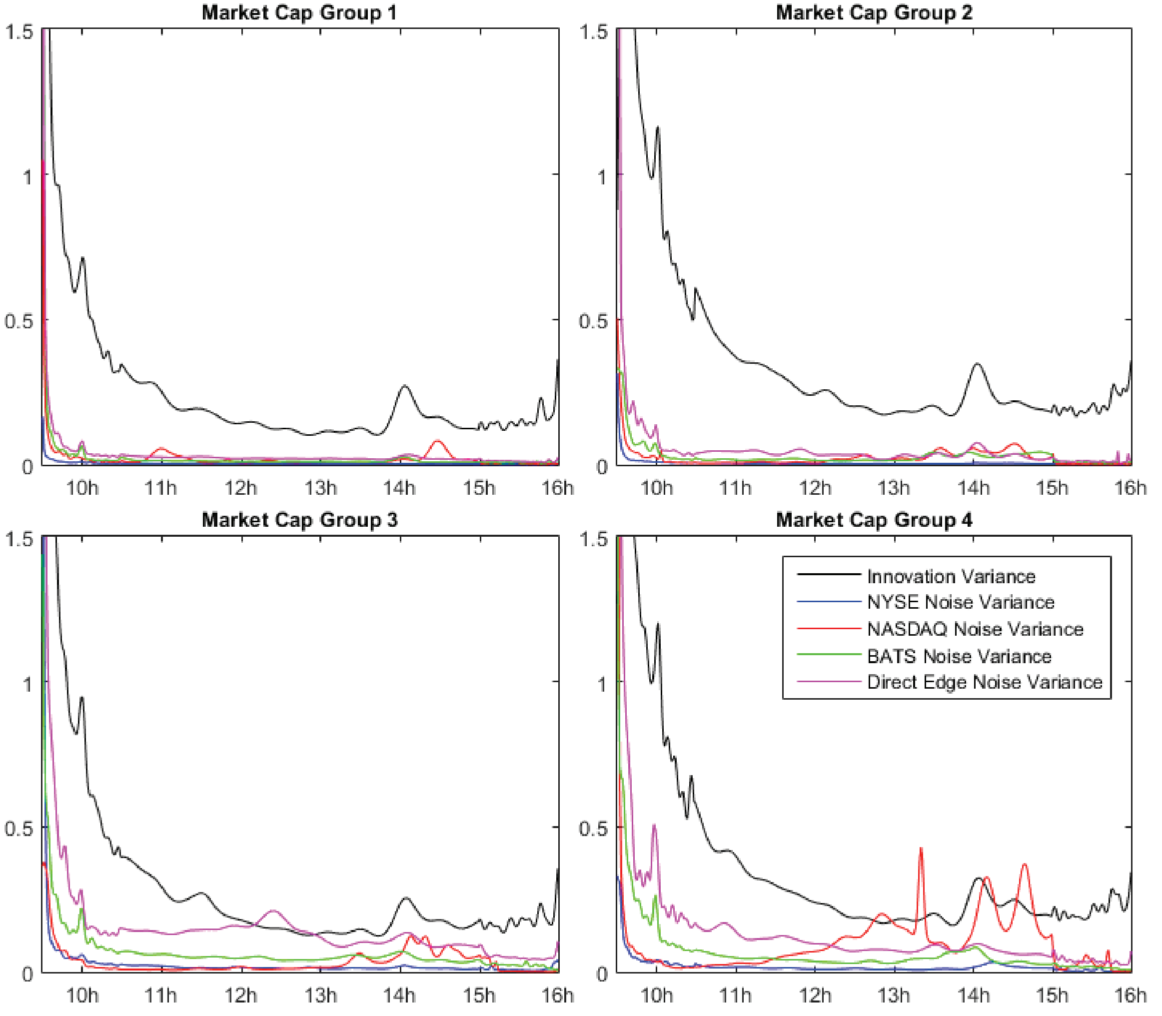
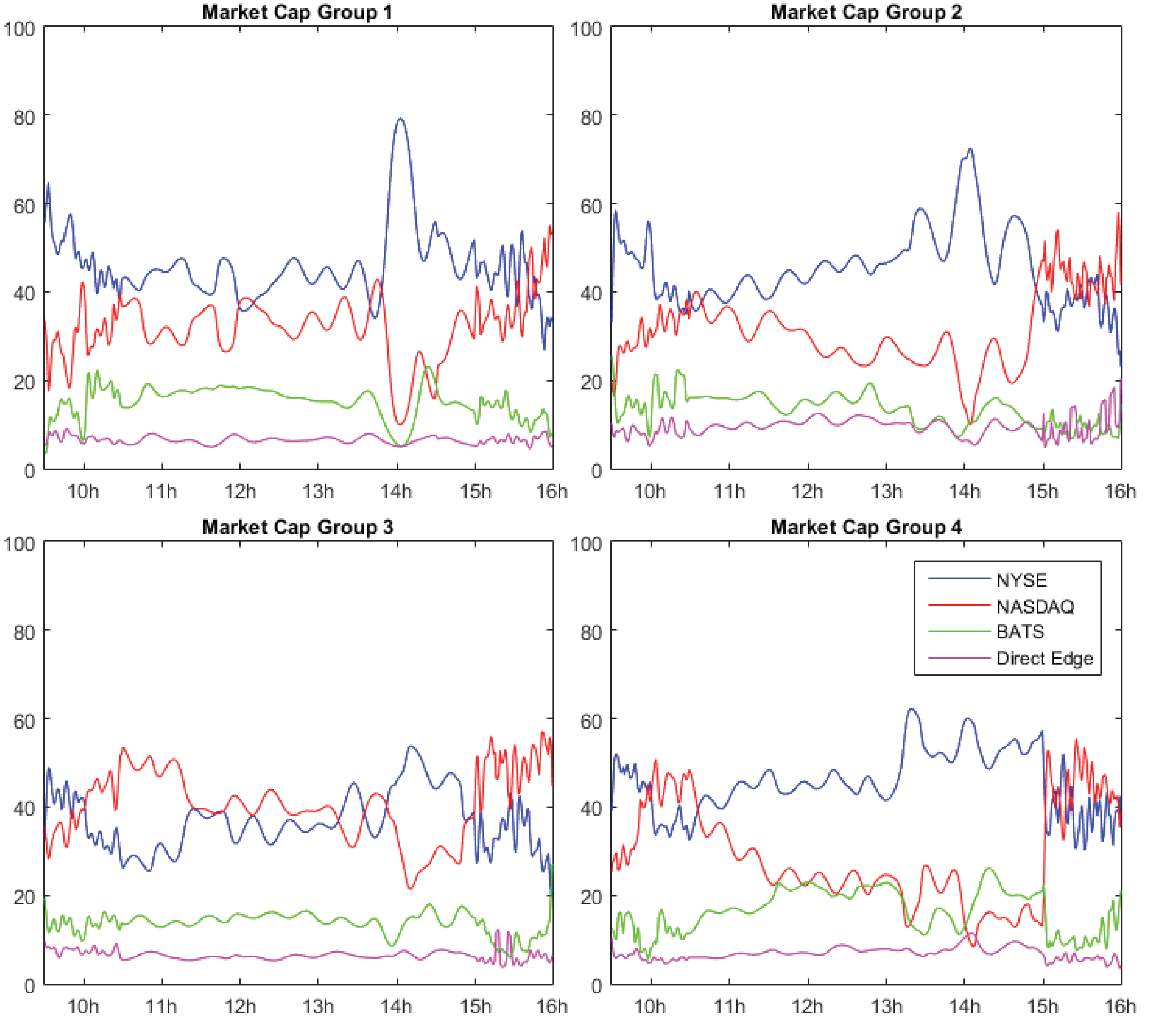


Fig. 2.11. Information Share Estimates for Market Capitalization Groups

The figure displays mean intraday information share estimates for four market capitalization groups. We rank our stocks according to their market capitalizations of December 31st 2013. We divide them into four groups, where the first three groups have 12 and the 4th group with the smallest firms has 14 stocks. For each exchange group we average the minute-level intraday estimates across stocks.



2.4.3 What drives the intraday variation in information shares?

We conclude our empirical analysis by exploring which market features can explain the intraday variation of the information shares of the four exchange groups. This analysis provides more detailed statistical inference on the effects of the stock characteristics we examined in the previous section by grouping the 50 S&P 500 stocks in various dimensions. We use a market share attraction (MSA) model (Cooper and Nakanishi, 1988) to relate the dynamics of the estimated information shares to the number of trades, quoted spreads, volatility, and market capitalization, controlling for dummies of exchange listing, industry and time-of-the-day. For each of the first three variables we consider the average value in minute t across all days in the sample. For the quoted spread we take the average difference between the bid and ask within the minute. For volatility we use the square root of the mean of squared 1-minute returns.

MSA models provide an effective framework to investigate the relationships between information shares and market variables. Just like the changes of market shares in relation to the ability of companies to attract customers, the ability of a trading venue to attract informed trading shapes its information share. MSA models also offer a gain in efficiency over separate regressions, as we can consider the determinants of the information shares for all exchanges simultaneously and in a consistent manner, as explained below. Note however that this method estimates elasticities between variables rather than causal relationships and therefore some care needs to be taken for interpreting the findings.

MSA models are based on the notion of a latent 'attraction' of a brand or company or, in our case, a trading venue. The attraction of exchange i at time t is defined as

$$A_{i,t} = \exp\left(\mu_i + \varepsilon_i\right) \prod_{j=1}^{I} \prod_{k=1}^{K} f(x_{k,j,t})^{\beta_{k,j,i}} \tag{2.16}$$

where $x_{k,j,t}$ is the value of the k^{th} explanatory variable of exchange j at time t, $\beta_{k,j,i}$ denotes the effect of this explanatory variable on the attraction of exchange i, I is the number of exchanges and K is the number of explanatory variables for each exchange. The function $f(\cdot)$ denotes a particular transformation of the explanatory variable $x_{k,j,t}$. We return to this below. The model is completed by setting the observed information share of exchange i at time t equal to its relative attraction, that is,[18]

$$IS_{i,t} = \frac{A_{i,t}}{\sum_{j=1}^{I} A_{j,t}}. \tag{2.17}$$

We limit ourselves to the two most popular specifications of the function $f(\cdot)$ in Eq. (2.16). These are the identity function, i.e. $f(x_{k,j,t}) = x_{k,j,t}$, and the exponential transformation, i.e. $f(x_{k,j,t}) = \exp(x_{k,j,t})$. The first specification results in the so-called Multiplicative Competitive Interaction (MCI) specification, while the latter leads to the Multinomial Logit (MNL) specification. These two specifications differ in terms of the implied elasticities of the information shares with respect to the explanatory variables. Under the MCI specification the elasticity of information share i to the k^{th} explanatory variable of exchange j at the minute t is given by

$$e_{k,j,i,t}^{MCI} = \frac{\partial IS_{i,t}}{\partial x_{k,j,t}} \frac{x_{k,j,t}}{IS_{i,t}} = \beta_{k,j,i} - \sum_{r=1}^{I} IS_{r,t}\beta_{k,j,r}, \tag{2.18}$$

while the elasticity for the MNL specification is

$$e_{k,j,i,t}^{MNL} = \frac{\partial IS_{i,t}}{\partial x_{k,j,t}} \frac{x_{k,j,t}}{IS_{i,t}} = \left(\beta_{k,j,i} - \sum_{r=1}^{I} IS_{r,t}\beta_{k,j,r}\right) x_{k,j,t}. \tag{2.19}$$

Due to the multiplication with $x_{k,j,t}$ in Eq. (2.19), the MNL specification is restricted to have small elasticities for values of the explanatory variable close to zero, whereas the MCI specification does not impose such a constraint.[19]

[18]Note that this implies that the information shares always sum to unity at each point in time. This need not necessarily hold for the information shares of De Jong and Schotman (2010) as defined in Eq. (2.8), although in our empirical application the mean of the sum of information shares is very close to unity with 98.5%. We therefore normalize the information share estimates for the remainder of the analysis.

[19]Further elaborations on these two models can be found in Cooper and Nakanishi (1988) and Cooper

In order to estimate the model parameters, we may take one of the exchanges as the benchmark (labeled "b"). We then rewrite the model that results from combining Eq. (2.16) and Eq. (2.17) as

$$\log(IS_{i,t}) - \log(IS_{b,t}) = \tilde{\mu}_i + \sum_{j=1}^{I}\sum_{k=1}^{l}\tilde{\beta}_{k,j,i}x_{k,j,t} + \sum_{j=1}^{I}\sum_{k=l+1}^{K}\tilde{\beta}_{k,j,i}\log(x_{k,j,t}) + \tilde{\varepsilon}_i, \qquad (2.20)$$

where $\tilde{\mu}_i = \mu_i - \mu_b$, $\tilde{\beta}_{k,j,i} = \beta_{k,j,i} - \beta_{k,j,b}$ and $\tilde{\varepsilon}_i = \varepsilon_i - \varepsilon_b$, and we have assumed that the first l explanatory variables enter the model with an MNL specification and the remaining $K-l$ explanatory variables are included with an MCI specification. The complete system has $(I-1)\cdot(1+I\cdot K)$ parameters, which can be estimated by OLS (Fok, Franses, and Paap, 2002). Note that only differences of the coefficients $\beta_{k,j,i}$ can be identified. As shown by Cooper and Nakanishi (1988), this is however sufficient to completely identify the elasticities given in Eq. (2.18) and Eq. (2.19).

As we have noted, the MNL specification implies a restriction on the elasticity to be small for explanatory variables values close to zero. Although it may be more restrictive in terms of this aspect, the MNL specification also has the advantage of allowing zero values of explanatory variables, because we do not use the natural logarithm of the MNL variable in the estimation equation (2.20) unlike the MCI case. The number of trades theoretically fits well with the implications of the MNL specification, because changes in the number of trades, when it has a small value, probably would not cause a considerable change in the information share of the relevant exchange. By contrast, small levels of spreads can be caused by heightened inter-exchange competition and changes around those levels can have major effects on the contributions to price discovery. Similarly, as we discussed in the previous section, changes in small noise variances can drive the major shifts in the information shares. For the modelling of the market capitalization, at least the necessity for having small elasticities at smaller values is not self-evident. Therefore we use the

(1993).

MNL specification for the number of trades and the more general MCI specification for the quoted spreads, volatility and market capitalization.[20]

We analyze the effects of the listing, industry and time-of-the-day dummies directly from this regression. The estimated dummy coefficients $\tilde{\beta}_l$ capture effects on the log ratio of an exchange to the benchmark exchange. Thus the effect on the relative information share of the exchanges $\hat{\beta}_l$ can be computed as

$$\hat{\beta}_l = \exp \tilde{\beta}_l - \exp 0 = \exp \tilde{\beta}_l - 1. \tag{2.21}$$

We consider the NYSE as a natural benchmark, especially for estimating the effect of exchange listing, given its prominent position in price discovery.

This methodology provides elasticities for each minute of the day, as can be inferred from the t subscript in the elasticity formulas Eq. (2.18) and Eq. (2.19). The intraday variation is a function of the level of information shares. We present the daily means in our analysis. We test for the significance of the mean elasticities by approximating their distribution using 10,000 simulations from the asymptotic distribution of the regression coefficients $\tilde{\beta}_{k,j,i}$.

The elasticities of information shares to quoted spreads display the most robust results, as can be seen in Panel A of Table 2.11. Each exchange's information share has a significantly negative elasticity to its own quoted spread, while having (often significant) positive elasticities to other exchanges. A well-documented response to signs of informed trading would be the widening of the spread in order to avoid adverse selection costs (e.g., Neal and Wheatley, 1998), which would imply that information share of an exchange has a positive elasticity to the spreads in that exchange. Given that we consistently find opposite signs, we can hypothesize that this effect is overwhelmed by tighter quoted spreads reducing transaction costs and attracting informed trading. While there is an almost one-

[20]The use of the MNL specification for the market capitalization does not generate qualitative changes in our results.

to-one effect for the NYSE and NASDAQ, the information shares of BATS and Direct Edge increase by about twice the percentage drop in the spreads. This may be related to the smaller shares of these two exchange groups in price discovery, leaving considerable room for improvement, i.e. marginal returns on market quality improvements are larger.

By controlling for volatilities of other exchanges in the MSA regression, the volatility variables mainly capture the idiosyncratic noise of each exchange. Thus the structure of volatility elasticities resembles the quoted spread results, but is weaker both in terms of statistical significance and robustness. Each exchange's information share has a significantly negative elasticity to its own volatility, except for the insignificant figure for Direct Edge. The cross-effects tend to be positive, when they are significant, except for the negative elasticity of BATS information share to Direct Edge volatility.

The number of trades fail to provide a coherent pattern. Further analysis of the number of trades by disaggregating various trade size groups, available upon request, does not provide a clear pattern either. We have observed in the previous section that the dominance of NASDAQ in the number and to a weaker extent in the volume of trades does not directly correspond to a large share in price discovery. The ability to execute orders in smaller slices using algorithmic trading could have made inferences on information based on trade sizes unreliable.

Table 2.11. Results of the Market Share Attraction Analysis

The table shows the results of the market share attraction analysis. We estimate the coefficients of the market share attraction model via OLS using the set of Eq. (2.20). In the case of the non-dummy variables we compute the daily elasticity series in Eq. (2.18) and Eq. (2.19) with the coefficient estimates, use 10,000 simulations of the OLS coefficients to approximate the distribution the mean elasticities and report their significance. Panel A has elasticity of information shares to the number of trades, the quoted spreads, volatility and market capitalization. In the case of the dummy variables we report the effect on the relative information share as in Eq. (2.21). Panel B reports the estimated change in relative information shares for listing and industry. Panel C reports the estimated change in relative information shares for each half-hour. The superscript *** marks significance at level 1%, ** at level 5%, and * at level 10%. We have 387 observations per stock for each of the four exchanges and by taking one exchange as the benchmark as in Eq. (2.20), we obtain 58,050 usable observations.

Panel A: Microstructure Variables

	Num. of Trades				*Spread*				*Volatility*				Market
	NYSE	NSDQ	BATS	DE	NYSE	NSDQ	BATS	DE	NYSE	NSDQ	BATS	DE	Cap
NYSE	−0.02	0.23***	−0.47***	0.16***	−1.01***	0.19***	0.59***	0.07	−0.66***	0.90***	−0.20***	0.00	−0.07***
NSDQ	0.01	−0.26***	0.65***	−0.37***	0.13	−0.69***	0.06	0.27***	0.91***	−1.57***	0.66***	0.17***	0.11***
BATS	0.10*	0.00	−0.20***	0.18**	2.05***	0.66***	−2.20***	0.23**	−0.10	0.71***	−0.78***	−0.33***	−0.09***
DE	−0.11*	0.04	−0.20***	0.29***	1.34***	0.68***	0.60***	−2.11***	−0.02	0.37***	−0.23*	−0.08**	0.08***

Panel B: Listing and Industry Dummies

	NYSE-List	Min	Cons	Manu	Util	Trd	Fin	Serv
$\Delta(NSDQ/NYSE)$	−0.36	−0.29	−0.39	−0.32*	−0.08	0.69**	−0.51**	3.10***
$\Delta(BATS/NYSE)$	−0.59***	0.00	0.51***	−0.13	0.71**	0.17	−0.51***	0.01
$\Delta(DE/NYSE)$	−0.63***	0.16	1.12***	−0.17	0.06	0.50***	−0.48***	0.37***

Panel C: Time Dummies

	9:30-9:45	9:45-10:15	10:15-10:45	10:45-11:15	11:15-11:45	11:45-12:15	12:15-12:45	12:45-13:15	13:15-13:45	13:45-14:15	14:15-14:45	14:45-15:15	15:15-15:45	15:45-16:15
$\Delta(NSDQ/NYSE)$	−0.40*	0.05	0.87***	0.15	−0.13	0.04	−0.05	0.01	−0.45	−0.72***	0.07	0.54	0.64**	2.34**
$\Delta(BATS/NYSE)$	−0.05	0.04	0.37***	0.39***	0.09	0.94**	0.34***	0.63**	−0.42**	−0.55***	−0.26	−0.09	−0.32*	−0.24
$\Delta(DE/NYSE)$	−0.15	−0.13	0.14**	0.25***	−0.04	0.28***	0.29***	0.31**	−0.15	−0.37**	−0.25	0.07	−0.13	0.13

We observe in Panel B of Table 2.11 that listing on the NYSE instead of NASDAQ has a considerable negative effect on the information shares of all other exchanges relative to the NYSE. Surprisingly, however, this effect is not significant at 5% level for the relative information share of NASDAQ. This result stems from the inclusion of the Services dummy, because in all other variations of the regression the effect turns out to be highly significant. Four of six stocks within this industry group are listed on NASDAQ and they observe some of the biggest leads of NASDAQ over the NYSE in terms of price discovery (on average 29.6%). This is reflected by the strikingly large positive effect of the Services dummy on the relative NASDAQ informativeness, indicating a tripling of NASDAQ information share relative to the NYSE under NASDAQ-listing. Thus controlling for this affiliation captures a sizable proportion of the listing effect.

For the industry (time-of-the-day) dummies, we report the estimates from regressions where all other industry (time-of-the-day) dummies are excluded. This allows us to measure the effect of the concerning industrial affiliation (time of the day) compared to all others, instead of taking an arbitrary industry (time-of-the-day) as a benchmark. The results are already transformed from the regression estimates using Eq. (2.21).

Exchange groups tend to be significantly more informative for specific industries. Out of 21 dummy coefficients for relative information shares for seven industries 12 turn out to be statistically significant at 5% level. Only Mining, the second smallest group, has no significant relationships. For each industry, significant effects on relative information shares also have the same sign. As the benchmark is the information share of the NYSE, this implies that other exchanges may share industrial specializations, while the NYSE has no real competitor for its industrial specializations.

The NYSE particularly excels at incorporating innovations in financial stocks. All of the six financial stocks are listed in the NYSE and for all non-NYSE exchanges the change in relative information shares is statistically significant. We estimate about 50% lower information shares for non-NYSE exchanges relative to the NYSE in this sector.

This probably goes hand-in-hand with the NYSE's aptitude at pricing macroeconomic news, considering that especially the 14:00h news causes the biggest increases for financial stocks.[21]

Macroeconomic news announcements generate sizable shifts in terms of both overall market volatility and the shares of exchange groups in price discovery. Jumps at 10:00h and 14:00h disturb the relatively smooth inverted J- or U-shaped intraday pattern for most of our variables. All of our market quality variables measuring the trading activity, spreads and volatility, displayed in Figure 2.2, and the intraday innovation variance estimates in Figure 2.4 experience these two jumps.

We use time dummies to capture the changes in price discovery at announcement times. Panel C of Table 2.11 points to significant time-related changes in the structure of price discovery even after controlling for other intraday and stock-related factors. While around 10:00h news we do not find a significant change from the average pattern giving the NYSE a lead over NASDAQ, the news at 14:00h causes a significant shift of price discovery towards the NYSE from all other exchanges.

We see also distinct patterns for the market open, close and the midday. The market open does not significantly differ from the average allocation of information shares. From the end of early morning news until the afternoon news, price discovery shifts from the NYSE particularly to BATS and Direct Edge, with significant positive changes in relative information shares for all periods from 10:15h to 13:15h except for 11:15h to 11:45h. And lastly, towards the market close we observe a significant shift from the NYSE towards NASDAQ.

2.5 Conclusion

This paper proposes a novel approach to measure the contribution to price discovery made by different trading venues, with an explicit focus on capturing intraday dynamics

[21] Note that time dummies control for influences of the two main intraday periods with macroeconomic news releases, while estimating the effects of the industry groups.

in information shares. We use a state space representation of the unobserved components model of Hasbrouck (1993) and De Jong and Schotman (2010). We introduce intraday time-variation in De Jong and Schotman (2010) information shares by allowing for time-varying volatilities of the efficient price innovations and idiosyncratic noise, using flexible Fourier specifications.

Our simulation study displays the capability of our method in capturing intraday dynamics of price discovery for typical data sets used in the market microstructure literature. Across a wide range of settings and parameter configurations it consistently provides accurate estimates of the model's structural parameters and the associated information shares.

In our empirical analysis we examine 50 S&P 500 stocks during the second half of 2013 using a 1-minute sampling frequency. We gather the exchanges on which these stocks are traded in four groups by quote origin: NYSE, NASDAQ, BATS and Direct Edge. We observe that most of the new information is incorporated into prices via the first two groups, particularly by the NYSE. However events such as the opening and the closure of the market and macroeconomic news announcements lead to major shifts in the contributions of each exchange to price discovery. In particular, the NYSE's share in price discovery widens during major macroeconomic news releases at 10:00h and 14:00h. Statistical tests show that the contributions to price discovery in the midday also do not remain constant. Among various market quality measures, we find narrower quoted spreads and smaller volatility in an exchange group lead to more informed trading. Exchange listing also turns out to be a major influence on information shares favoring the exchange where the stock is listed.

Our state space ML methodology advances the information shares literature to the investigation of intraday dynamics. In present-day financial markets the incorporation of news into prices takes minutes, if not seconds, and also access to high frequency data gets easier. These factors provide a fertile ground for the application of our methodology to contemporary issues in price discovery. Due to its ability in estimating structural parameters, this method can easily be extended to incorporate richer market microstructure dynamics and applied to various settings where a number of observed security prices share an underlying asset.

Appendix A: The state space representation of the unobserved components model

In the state space form, the unobserved components model given in Equations (1) and (2) can be represented by these two equations:

$$p_t = [\iota_{N\times 1}\ \alpha\ \Psi] \begin{bmatrix} p_t^* \\ r_t \\ e_{t-1} \end{bmatrix} + G\varepsilon_t, \text{ where } G = [0_{N\times 1}\ I_N] \text{ and } \varepsilon_t = \begin{bmatrix} r_{t+1} \\ e_t \end{bmatrix}, \tag{2.22}$$

$$\begin{bmatrix} p_{t+1}^* \\ r_{t+1} \\ e_t \end{bmatrix} = \begin{bmatrix} 1 & 0_{1\times N+1} \\ 0_{N+1\times 1} & 0_{N+1\times N+1} \end{bmatrix} \begin{bmatrix} p_t^* \\ r_t \\ e_{t-1} \end{bmatrix} + H\varepsilon_t, \text{ where } H = \begin{bmatrix} \iota_{2\times 1} & 0_{2\times N} \\ 0_{N\times 1} & I_N \end{bmatrix}, \tag{2.23}$$

with $\iota_{n\times m}$ an $n \times m$ vector of ones, $0_{n\times m}$ an $n \times m$ matrix of zeros, Ψ is an $N \times N$ matrix and I_N is an $N \times N$ identity matrix.[22] The variance parameters are uniquely identified using the covariance matrix of the stacked disturbances $\begin{bmatrix} H \\ G \end{bmatrix} \varepsilon_t$, which comprises the innovation and noise variances:

$$E\left[\begin{bmatrix} H \\ G \end{bmatrix} \varepsilon_t \varepsilon_t' \begin{bmatrix} H \\ G \end{bmatrix}'\right] = \begin{bmatrix} \sigma_r^2 \iota_{2\times 2} & 0_{2\times N} & 0_{2\times N} \\ 0_{N\times 2} & \Omega & \Omega \\ 0_{N\times 2} & \Omega & \Omega \end{bmatrix}.$$

[22] As the noise terms e_t and the innovation r_t are in different equations, we could have avoided combining them under ε_t. However this formulation is in line with the model entry requirements of the SsfPack by Koopman, Shephard and Doornik (1998) used in this paper. Oftentimes a model bears more than one equivalent state space representation.

Appendix B: Proof of the information share formula

We define $\Theta = \Omega + \sum_{j=1}^{L} \Psi_j \Omega \Psi_j'$ and assuming the diagonality of Ω and Ψ_j matrices Θ is also a diagonal matrix with diagonal elements $\omega_i^2 \left(1 + \sum_{j=1}^{L} \psi_{i,j}^2\right)$. To compute the inverse of $\Upsilon = \sigma_r^2(\iota+\alpha)(\iota+\alpha)' + \Omega + \sum_{j=1}^{L} \Psi_j \Omega \Psi_j'$, we use the Sherman-Morrison formula

$$\Upsilon^{-1} = \Theta^{-1} - \frac{\Theta^{-1}\sigma_r^2(\iota+\alpha)(\iota+\alpha)'\Theta^{-1}}{1+\sigma_r^2(\iota+\alpha)'\Theta^{-1}(\iota+\alpha)}.$$

Then γ is defined as

$$\gamma = \Upsilon^{-1}(\iota+\alpha)\sigma_r^2 = \frac{\Theta^{-1}(\iota+\alpha)\sigma_r^2}{1+\sigma_r^2(\iota+\alpha)'\Theta^{-1}(\iota+\alpha)}.$$

And the information shares are defined as

$$IS_i = \gamma_i(1+\alpha_i) = \frac{(\Theta^{-1}(\iota+\alpha))_i}{1/\sigma_r^2 + (\iota+\alpha)'\Theta^{-1}(\iota+\alpha)}(\iota+\alpha)_i, \; i = 1\ldots N,$$

which can be rewritten in the form given in the text, because Θ^{-1} is a diagonal matrix with diagonal elements $1/\left(\omega_i^2 \left(1 + \sum_{j=1}^{L} \psi_{i,j}^2\right)\right)$.

Chapter 3

Why do the pit hours outlive the pit?

This chapter is a joint project with Dr. Michel van der Wel and Prof. Dick van Dijk of Erasmus University Rotterdam.

3.1 Introduction

Over the last two decades, electronic communication networks (ECNs) have evolved from auxiliaries of the trading pit to the dominant venues of trading. Besides cutting trading costs associated with human intermediaries, ECNs also bring the prospect of round-the-clock trading in a single venue and thereby further opening financial markets to global participation. However one aspect of pit trading, namely high trading activity during the traditional pit trading hours, remains surprisingly resilient. Although ECNs have extended potential trading hours to nearly the whole day, the pit hours still enjoy most of the trade volume as well as price the bulk of daily information. We investigate the factors behind the persistence of the high trading activity during the pit hours in a long and recent data set of 30-year U.S. Treasury futures.[1]

This concentration of trading activity during the pit hours may be driven by the informativeness and liquidity provided by pit traders or releases of domestic macroeconomic news, which happen almost always within the pit hours and constitute the main source of information in the U.S. Treasury market (Andersen, Bollerslev, Diebold, and Vega, 2003; Green, 2004). However, this phenomenon may also be the result of a liquidity externality during the pit hours or, conversely, an illiquidity externality during the rest of the day, referred to as 'afterhours' in the following. The concentration of trading activity on a portion of the day diminishes adverse selection and search costs and generates an externality as additional trading activity within this portion boosts the benefits for all parties (Pagano, 1989; Hendershott and Mendelson, 2000). The traditional pit hours serve as a natural focal point for such an activity clustering. Nevertheless, this effect also leads to a market failure for the ECN during the afterhours, because its low liquidity leaves afterhours traders vulnerable to larger transaction and adverse selection costs. As afterhours traders mainly comprise of non-domestic parties with less discretion on the timing of their

[1]Trading in the 30-year U.S. Treasury bond is possible for the whole day, rotating between venues in New York, Tokyo and London. Pit trading for the futures contract occurs during the interval of 08:20h-15:00h EST, while the electronic market is always open, except for a progressively declining amount of evening hours.

trades, this unintended discrimination harms the efficiency of global risk management and may generate biases in the transmission of global risks into the pricing of U.S. Treasury securities.

Our data set, spanning from 2004 to 2013, allows for examining a number of aspects of the activity clustering around the pit hours. The sample period contains the recent upsurge in electronic trading systems with the implementation of the Reg NMS framework[2] in the U.S. and the parallel sophistication of ECN platforms, in particular the introduction of the Treasury futures to the CME Globex platform in 2008. This marks the transformation of the trading pit from a significant provider of price discovery to a redundant venue with little trading activity, leading to the CME decision to close the trading pits for U.S. Treasury futures by July 2015. In spite of this dramatic reduction in pit activity, the share of the pit hours in ECN trade volume has declined only from 87.6% in 2004 to 73.4% in 2013. The large jump in trading activity - downwards for the pit and upwards for the ECNs - at the start of 2008 allows us to examine the characteristics and determinants of afterhours trading under two different settings.

We test two alternative hypotheses to explain the activity clustering around the pit hours. We firstly examine the effect of macroeconomic news announcements and the benefits of trading during the pit hours. Trading activity increases rapidly with the pit open and similarly drops after the pit close. This pattern is even more amplified on days without macroeconomic new releases near these times, discounting these news releases as an explanatory factor for the clustering. If pit trades are relatively informative or provide a significant amount of liquidity, market participants may prefer to trade during the pit hours at more informative prices and with lower trading costs. Although the trading pit hosts only a small portion of the daily trading activity, our analysis shows that it has a considerable share in price discovery, especially during the early years of our sample.

[2]The regulatory agency for U.S. exchanges, the Securities and Exchange Commission, implemented Regulation National Market System (Reg NMS) in 2007 to enhance the linkage across exchanges and to improve trade efficiency. This regulation is widely regarded as a cornerstone in the proliferation of electronic trading systems.

In 2004 pit trades account for 32.0% of the variation in permanent price innovations, while only 10.0% of the trades during the pit hours are executed in the pit itself. With the introduction of the Globex Platform in 2008, the informativeness of the pit drops to 11.5% and pit activity declines so rapidly that proper inference is not possible for later years. As this fast decline of the pit does not affect the activity clustering around the pit hours, we reject informativeness of pit trading as a significant factor for the persistence of this clustering.

As an alternative explanation, we postulate a feedback mechanism between trading activity on the one hand and price informativeness and trading costs on the other hand. Using a modelling framework capturing the intraday variation in price informativeness, information asymmetry and price impact of trades, we examine the existence and evolution of this feedback mechanism. Although imputed adverse selection costs disincentivize trading with a better-informed party, Admati and Pfleiderer (1988) show that liquidity traders can cluster with informed traders as long as their adverse selection costs are outweighed by their benefits from informative prices generated by the competition between informed traders. Barclay and Hendershott (2004) present empirical evidence for a mutually reinforcing relationship between the low afterhours trading activity and the two factors related to trading costs, namely adverse selection and price impact, which we also evaluate in this study. Taken all together, under this mechanism price informativeness has a mutually reinforcing relationship with liquidity, whereas information asymmetry and price impact of trades are reduced by and also repel liquidity.

We examine the dynamics of the afterhours market for such a feedback mechanism. Investigating the afterhours provides insights on whether and to what degree the aforementioned factors prevent or support the low trading activity during the afterhours and thus the activity clustering around pit hours. We find statistically and economically significant effects in line with the postulated mechanism. Percentage changes in price informativeness have a positive and nearly one-to-one relationship with changes in trading activity. By contrast, changes in information asymmetry and price impact have a negative relationship with afterhours trading activity. We observe a strengthening in the negative effect of information asymmetry during the crisis years, accompanying a temporary reversal of the small trend towards the reduction of the share of pit hours in daily trade volume. The reduction in trading activity for each one percent increase in information asymmetry moves from a pre-crisis average of -2.6% to an average of -20.8% from the second half of 2008 to 2010 and drops to -1.2% afterwards.

The results concerning the effect of the price impact of trades alleviates some of the main concerns related to this illiquidity externality during the afterhours. We estimate that, after a rise during the crisis years, from 2012 on the price impact starts dropping below its pre-Globex low in 2006-2007 and in the meantime the magnitude gap between the night hours and the highly liquid hours just before the market open also diminishes. Similarly, price impact loses its significant effect on trading activity clustering from the second half of 2007 on. Therefore market participants trading outside pit hours, especially during the night hours, would face less price impact and the incentives to postpone trading to the pit hours would be reduced. The spread of algorithmic trading with the introduction of the Globex Platform is a likely explanation for this change. Although not affecting the magnitude of the clustering, the introduction of the Globex Platform more than halves the average trade size and leads to a more even distribution of the trade size across the day. This reduction in average trade size signals the use of algorithmic trading methods to execute big orders with a series of small trades.

In terms of econometric methodology, we provide a unified framework to estimate structural parameters as time-varying processes. We assume a latent price process, in which price changes originate either from the surprises in the flow of buyer- or seller-initiated trades, namely the order flow, or from price innovations unrelated to the trading process. The surprises encompass the pricing of private information signals, measuring the information asymmetry (Madhavan et al., 1997). As we control for the price changes generated through the private information signals with the surprise term, the innovation process captures the pricing of public information. The latent price is observed with noise. We distinguish the price impact of trades from the rest of noise sources. We model the innovation and noise variances as well as the information asymmetry and the price impact of trades as time-varying processes. This framework improves upon estimation methods for different aspects of afterhours trading by various reduced-form methods and accounts for the time-variation in the structural parameters.

Several studies use structural models to analyze afterhours trading processes. Aside from differences across specific models, these generally assume the structural parameters to be constant over a time frame, while our methodology allows for the estimation of parameter variation at the frequency of the data inputs as in Ozturk, Van der Wel, and van Dijk (2014). Barclay and Hendershott (2004) investigate the activity clustering during the afterhours by decomposing the effective spread into adverse selection and fixed components with the Lin et al. (1995) model and find the trading costs generated by low trading activity during the afterhours as a major factor for the persistence of pit hours trading. He, Lin, Wang, and Wu (2009) analyse dynamics of round-the-clock price discovery in the U.S. Treasury futures market with the Madhavan et al. (1997) model and find that the information asymmetry as well as price informativeness peaks in the preopen. We also use a version of Madhavan et al. (1997), as it is more amenable to using only trade data.

Our estimation methodology is closely related to a growing literature in the application state space modelling of market microstructure issues. Frijns and Schotman (2009) and Korenok et al. (2011) estimate the Hasbrouck (1993) model, which we use to measure price discovery during the pit hours, using Kalman filtering. Korenok et al. (2011) also use the Madhavan et al. (1997) model to incorporate order flow dynamics. Hautsch, Hess, and Veredas (2011) examine the effect of macroeconomics news on innovation and noise components of volatility. In contrast to these models using intraday data, Hendershott and Menkveld (2014) investigate price pressures with a state space model using daily data, thereby avoiding modelling issues like adverse selection based on short-term information. As an alternative methodology to these applications of Kalman filtering, Jondeau, Lahaye, and Rockinger (2015) estimate intraday variation in the model parameters using the particle filter.

The remainder of the paper is organized as follows. Section 3.2 presents the methods we use and formalizes our hypotheses. Section 3.3 shows descriptive statistics of our 30-year U.S. Treasury futures data set. Section 3.4 investigates the determinants of the activity clustering around the pit hours by measuring pit informativeness and the dynamics of the afterhours. Section 3.5 concludes.

3.2 Hypotheses and Methodology

In this section we firstly present our hypotheses. In the second subsection, we outline the information share methodology that we use. In the last subsection, we present a structural microstructure model with time-varying parameters, estimated using state space methods.

3.2.1 Hypotheses

To structure our analysis, we present two hypotheses regarding the activity clustering around the pit hours. The first hypothesis poses the informativeness of pit trading as a cause for the preference of trading during the pit hours. The second hypothesis relates to the feedback loop between trading activity and variables related to informational and transaction costs and price informativeness.

Hypothesis 1: The trading pit constitutes an important venue for price discovery. Trading activity clusters around the pit hours, because traders benefit from more informed prices emerging from the trading pit.

The relative informativeness of the prices in the trading pit can pose a straightforward reason why the pit hours still attract the bulk of trade volume. A large literature shows that pit traders manage to avoid adverse selection problems through longstanding relationships and reputation and attract mainly uninformed and liquidity-oriented traders (Seppi, 1990; Benveniste, Marcus, and Wilhelm, 1992; Madhavan and Cheng, 1997; Battalio, Ellul, and Jennings, 2007). Contrasting with this largely uninformative order flow into the trading pit, the executions of pit traders seem to propagate price discovery. Sofianos and Werner (2000) note that the floor brokers act like "a smart order book" cutting the order into pieces and executing it strategically over an extended period of time. They condition their trades on the limit order book as well as the hidden liquidity arriving directly to the trading pit and benefit from order imbalances (Grossman, 1992; Madhavan and Smidt, 1993; Barclay, Hendershott, and Kotz, 2006). This quality infuses the pit order flow with a high predictive power on the asset's future price (Hasbrouck and Sofianos, 1993; Madhavan and Sofianos, 1998; Kavajecz, 1999; Handa, Schwartz, and Tiwari, 2006).

We analyse the contribution of the trading pit to price discovery using the information share methodology of De Jong and Schotman (2010), which we'll detail in the next subsection. This method relies on a structural model related to our model for afterhours trading. The first hypothesis implies a considerable amount of price discovery in the trading pit and a positive relationship between the trading activity during the pit hours and the information share of the trading pit.

Hypothesis 2: Trading activity has a positive relationship with price informativeness and negative ones with trading costs related to information asymmetry and price impact. The activity clustering persists due to the benefits of trading at already-liquid times of the day, namely more informed prices and less adverse selection and price impact costs.

A mutually-reinforcing relationship between trading activity and other microstructure factors during the afterhours may result in a persistent trading activity difference between parts of the day. The model of Admati and Pfleiderer (1988) implies that the clustering of liquidity and informed traders may be sustained by benefitting both parties: While the informed traders enjoy the reduced impact of their trades, the liquidity traders benefit from the competition between traders with similar information. Information shared by market participants, like public news announcements, would not require trading to be priced, as all parties would agree on a new price level for the security in the light of the new information. By contrast, in order to avoid alarming non-informed traders, information possessed by fewer traders would be priced over time, leaving a trace of unexpected price innovations in the order flow. Another implication of this clustering of informed and liquidity traders is the reduction of the price impact of trades for all parties. In the case of the low liquidity levels persistent in the afterhours, transaction costs generated by price impact of trades would drive away trading activity and would be further increased due to reduced trading activity. Following the theoretical results of Admati and Pfleiderer (1988), we hypothesize that trading activity has a positive relationship with the magnitude of price informativeness and negative ones with the degree of information asymmetry and the price impact of trades.

Several empirical studies support the postulated relationships of trading activity toward price informativeness and information asymmetry. News announcements provide a strong evidence for a positive relationship between trading activity and new public information (Chordia, Roll, and Subrahmanyam, 2001; Hautsch et al., 2011). By contrast, the aversion of non-informed market participants to trades with informed parties where they may endure adverse selection costs constitutes one of the main dynamics of trading models at least since the 'no-trade theorem' of Milgrom and Stokey (1982). Barclay and Hendershott (2004) propose the higher information asymmetry during the afterhours as an explanation for the lower liquidity during this time period.

The relationship between trading activity and price impact of trades remains rather contentious. Foster and Viswanathan (1990) and Hasbrouck (1991) provide an early rebuttal to Admati and Pfleiderer (1988) by showing that the highly-liquid market open period also exhibits larger price impact for trades. Extending on the methodology of

Hasbrouck (1991), Dufour and Engle (2000) and Chung, Li, and McInish (2005) also find that trading activity measured by the interval between trades increases price impact. On the other side, Barclay and Hendershott (2004) posit the higher impact of trades due to the low liquidity during the afterhours as one of the main pillars of an illiquidity externality. Analyzing intraday data, Jondeau et al. (2015) demonstrate that more liquid stocks experience less price impact.

We test the second hypothesis with time-varying estimates of price informativeness and trading costs, coming from the model presented in the third subsection. We divide our ten-years-long data into half-yearly series, generating 20 subsamples, each with a 5-minute sampling frequency for observations. For each half-year, the regression of trade volume on the measures of the price informativeness, the information asymmetry and the price impact gives the effects of these three variables on trading activity. Our hypothesis implies the regression coefficient to be positive for the price informativeness and negative for the other two variables. A cross-sectional regression for each 5-minute period during the afterhours across years constitutes another test for the significance of these relationships across the afterhours. We test this mechanism for the afterhours, because we want to examine the factors sustaining the low afterhours trading activity.

3.2.2 Information Shares

The first hypothesis requires the measurement of informativeness of trading in the pit compared to the ECN. We accomplish this using the information share methodology of De Jong and Schotman (2010). In this framework, the observed prices of an asset in different venues, in this case in the trading pit and the ECN, are driven by a latent efficient price process. This latent price is modeled as a random walk with stationary innovations. The observed asset prices deviate from this latent price with a set of stationary error terms as long-term or unbounded deviations are ruled out by arbitrage. The error terms capture microstructure effects in the observed prices. They comprise two components distinguished by their correlation with the efficient price innovation. First, the information-correlated pricing error component captures dynamics such as adverse selection. The second component is uncorrelated with information, but stems from factors such as noise trading or price discreteness.

These relations can be represented as

$$\begin{aligned} p_{i,t} &= p_t^* + \alpha_i \epsilon_t + e_{i,t} + \psi_i e_{i,t-1}, \quad i = ECN, Pit, \\ p_t^* &= p_{t-1}^* + \epsilon_t, \end{aligned} \tag{3.1}$$

where $p_{i,t}$ is the log observed price, p_t^* is the latent efficient price, ϵ_t is the innovation in the latent price with mean zero and variance σ_ϵ^2, the coefficients α_i capture over/underreaction to the innovations ϵ_t, $e_{i,t}$ are the zero-mean noise disturbances with covariance matrix Ω and are uncorrelated with the innovations in the latent price as well as other noise $e_{j,t}$, $i \neq j$, and the coefficients ψ_i capture serial correlation in the noise. This would simplify to the Roll (1984) model under no over/underreaction in prices to information ($\alpha_i = 0$), the exclusion of the lagged noise terms ($\psi_i = 0$) and the replacement of the noise terms e_t with the effective spread.

De Jong and Schotman (2010) propose a price discovery measure quantifying the explanatory power of changes in each of the observed security prices for the innovations in the latent price. The total price innovation of each venue in period t is defined as

$$\nu_{i,t} = p_{i,t} - p_{t-1}^* = (1 + \alpha i)\epsilon_t + e_{i,t} + \psi i, t e_{i,t-1}, \quad i = ECN, Pit. \tag{3.2}$$

To measure informativeness of the venues, we may then consider the regression of the innovation in the latent price on the total innovations in individual prices, that is

$$\epsilon_t = \gamma' \nu_t + \eta_t, \tag{3.3}$$

where η_t is the innovation in the latent price unrelated to innovations in market prices.

De Jong and Schotman (2010) decompose the goodness-of-fit of this regression into information shares which show how much of the price innovations is explained by the total innovations in each market. Ozturk, Van der Wel and Van Dijk (2014) show that the information shares are defined as

$$IS_i = \frac{(1+\alpha_i)^2/(\omega_i^2(1+\psi_i^2))}{1/\sigma^2 + \sum_{j=1}^{N}(1+\alpha_j)^2/(\omega_j^2(1+\psi_i^2))} \tag{3.4}$$

and their sum gives the goodness-of-fit of the regression, which does not necessarily equal to one. We use the GMM methodology to estimate the model parameters required for the computation of these information shares.[3]

We evaluate Hypothesis 1 by measuring annual information shares of the trading pit and the ECN in order to examine the pattern of price discovery over time. This allows a comparison with the time series pattern of the pit activity as well as the clustering of the ECN activity around the pit hours and an examination of the strength of their interrelationships. The informativeness of the trading pit should be to be proportional to the amount of pit activity, through which information is compounded into prices. However, given the literature surveyed in the previous subsection, the pit is expected to have a larger share in price discovery compared to its share in the total trade activity. Hypothesis 1 requires the informativeness of the trading pit, as measured by the information shares of (3.4), and the persistence of the activity clustering around the pit hours to be closely related.

[3]Ozturk et al. (2014) estimate this model both via GMM, assuming the constancy of model parameters, and via Maximum Likelihood with Kalman filtering, as we do for the more sophisticated model of the following section, which allows for time-variation in the model parameters and thereby produces time-varying information shares. The GMM result is comparable to the average of the time-varying information shares. Given the similarity of the resulting information shares and that in this analysis intraday variation in the information shares is not a point of interest, we use the GMM methodology due to its lower computational requirements.

3.2.3 Structural Model with Intraday Variation

In order to estimate the variables related to the second hypothesis, we extend the model presented in the previous section with richer market microstructure dynamics. In the latent price process, we differentiate price innovations incorporated with and without trading. Changes in the beliefs of market participants about asset prices can emerge either from public news or through the signals in the order flow indicating information asymmetry. We express this difference in the structural model by including the surprise in the order flow as a determinant of the changes in the latent price, constraining innovations to changes based on commonly shared information which does not require trading to be priced. Thus the latent price process is defined as

$$p_t^* = p_{t-1}^* + \theta_t \left(q_t - E[q_t|q_{t-1}] \right) + \epsilon_t \tag{3.5}$$

where p_t^* is the latent efficient price, $q_{i,t}$ is the order flow, $(q_t - E[q_t|q_{t-1}])$ is the surprise in the order flow, θ_t is the unexpected order flow coefficient and ϵ_t is the public information component of the price innovation with mean zero and time-varying variance σ_t^2.

As we estimate this model only in the afterhours, as per the focus of Hypothesis 2, the observed price process consists of solely the ECN data and the model undergoes three major modifications. The main change is the introduction of a measure of price impact of trades. In particular, we replace α_i, the under/over-reaction coefficient of observed prices to latent price innovations, with the price impact coefficient δ. This new variable captures the reaction of observed prices to the whole order flow rather than just its informative component. Following prior empirical studies, we model the effect of trade volume as a concave function rather than a linear one (e.g., Kempf and Korn (1999)). As a minor modification, we allow for more lags of noise to capture the serial correlation in the data created by transitory noise. Thus observed prices follow the process

$$p_t = p_t^* + \delta_t q_t + e_t + \sum_{j=1}^{J} \psi_j e_{t-j} \tag{3.6}$$

where J is the number of noise lags, e_t are noise terms with mean zero and time-varying variance ω_t^2 and δ_t is price impact coefficient. Lastly the order flow is modelled as an autoregressive function

$$q_t = \sum_{j=1}^{R} \rho_j q_{t-j} + \eta_t, \tag{3.7}$$

where R is set to six using information criteria results.

We estimate Eq. (3.5) and (3.6) by Maximum Likelihood using Kalman filtering.[4] This estimation method also allows for incorporating more complex dynamics into the model. We model the variance of public price innovation σ_t^2, the noise variance ω_t^2 as well as the coefficients of unexpected order flow θ_t and of the price impact δ_t as time-varying processes. As in Ozturk et al. (2014), we implement time-variation using a combination of flexible Fourier trigonometric functions and a polynomial function. The time-varying parameters have the form

$$c + \sum_{p=1}^{P} \kappa_p (t \pmod N)^p + \sum_{q=1}^{Q} \left(\xi_q \cos\left(\frac{2\pi q t}{N}\right) + \zeta_q \sin\left(\frac{2\pi q t}{N}\right)\right), \tag{3.8}$$

where t denotes time with $t = 1, \ldots, T$, T being the number of all observations, N is the number of observations per day, P the order of the polynomial part, and Q the total number of flexible Fourier sets. We use the exponent of this specification for the variances to facilitate an unconstrained maximization procedure given that trigonometric functions can have negative values. The flexible Fourier form can model complex dynamics and smooth transitions. However using solely the flexible Fourier part would impose equality of the variances at the start and end of the day. We avoid this by complementing it with

[4]The state space representation of the model is given in the appendix.

the polynomial component.

This model provides estimates of three variables, which we require in order to evaluate Hypothesis 2: Price informativeness measured by the public innovation variance σ_t^2, information asymmetry measured by the coefficients of unexpected order flow θ_t and the price impact coefficient δ_t. We use these variables to evaluate the existence of the postulated feedback mechanism relating them to trading activity measured by trade volume. Hypothesis 2 requires these variables to have significant explanatory power on the changes in trade activity, measured by trade volume.

3.3 Data and Descriptive Statistics

In this section we firstly introduce some summary statistics of our data set and then provide evidence for the activity clustering around the pit hours.

3.3.1 Data

We employ a data set of intraday transaction prices and volumes of 30-year U.S. Treasury bond futures contracts spanning a 10-years-long period from 2004 to 2013.[5] The trades are time-stamped at the second level. We sign the trades using the tick test.[6] Trading takes place both in the trading pit and in the ECN. The original trading venue of the contract, the Chicago Board of Trade (CBOT), introduced electronic trading in mid-2003, which moved in January 2008 to the Chicago Mercantile Exchange (CME) Globex platform after the merger of CBOT and CME in July 2007.

The extended trading hours of the ECN allows for the simultaneous incorporation of price movements in the underlying security, the 30-year U.S. Treasury bond, to the

[5] The 30-year Treasury bond was discontinued from February 2002 to February 2006 and the futures contract was priced using the substitutes provided by the U.S. Treasury: The Long-Term Average Rate until June 2004 and afterwards an extrapolation factor to compute an estimated 30-year rate using the 20-year Constant Maturity rate.

[6] We did robustness checks for a number of time intervals using the considerably more computationally intensive method of Hasbrouck (2004). This provided very similar results for the 5-minute aggregates we use in the estimation of the state space model.

futures price. The bond itself trades round-the-clock in Tokyo (19:30h-03:00h EST), London (03:00h-07:30h EST) and New York (07:30h-17:30h EST). The trading pit of the futures contract is open from 08:20-15:00h EST. The electronic market for the futures closes during our sample period at 17:00h EST, but its opening time has moved from 20:00h EST in 2004 to 18:00h EST in 2013.

Figure 3.1 shows the movement of the 30-Year U.S. Treasury futures price during our sample period. The contract price experiences dramatic changes during the financial crisis. The first big jump in December 2008 corresponds to the reduction of the federal funds target rate by the Fed. In the first half of 2009, the price returns to its level before the jump in December 2008. In mid-2010 the European debt crisis and stock market volatility lead to a flight to safety causing another appreciation in the Treasury futures price. While this movement starts to reverse in late 2010, S&P's downgrade of U.S. debt in August 2011 results in even more demand for U.S. debt, triggering new highs for the last years of our sample.

The introduction of the Globex Platform in 2008 transforms the relationship between electronic and pit trading fundamentally and irreversibly. Table 3.1 presents yearly summary statistics for our data set. The pit hours statistics in Panel A show that from the start of the ECN on, the trading pit has a relatively small share in the number of trades with 1,232 pit trades compared to 11,124 ECN trades in 2004. Pit trading practically disappears after the introduction of the Globex in 2008, reducing the annual average of the daily number of pit trades from 310 in 2007 to 69 in 2008. By 2013, merely 6 trades per day occur in the pit. Figure 3.2 displays the number of trades executed in the trading pit compared to the ECN over time in more detail. In early 2008, a discrete drop in pit trading accompanies a jump in ECN activity. Given the lack of a parallel increase in ECN trade volume, the main driver behind the surge in amount of trades in the ECN seems to be the rise of algorithms to cut larger trades in smaller pieces and disperse the execution of these pieces across the day. As Table 3.1 shows, the ECN trade volume during the pit

Fig. 3.1. The Movements of the 30-Year U.S. Treasury Futures Price

The figure shows the value of the 30-Year U.S. Treasury Futures over the 4108 trading days from 2004 to 2013. The left axis is the price in USD and the bottom axis gives the days. The data is sampled at 5-minute frequency.

hours plummets from the peak of 286,336 in 2007 to 148,358 in 2009 and rises back to 240,779 in 2013.

3.3.2 Activity clustering

The intraday distribution of the trading activity in the ECN, presented in Figure 3.3, highlights the significant influence of the pit hours. A 10-minute interval during the pit hours hosts on average to 2.0% of the daily trade volume. This average activity figure is 10 times lower for a similar interval outside the pit hours.

The trading activity tends to rapidly rise near the pit open and drop after the pit close. The share in the daily trading activity doubles from 1.1% to 2.2%, comparing the 10-minute intervals before and after the pit open at 8:20h. Similarly, the pit close at 15:00h leads to a reduction in activity shares from 2.2% to 1.5%. In order to exclude the effect of nearby macroeconomic news announcements on these activity changes at the pit open and close times, Figure 3.3 also documents the trading activity shares excluding days with an announcement 30 minutes before or after the open and close times. In both cases we observe a more salient jump. The share of a 10-minute interval in the daily trading activity rises from 1.2% to 2.7% with the pit open and drops from 2.2% to 1.5% after the pit close.

This substantial share of the pit hours in the trading activity survives the severe decline in the number of pit trades, especially after 2007. Figure 3.4 shows that the pit hours consistently attract a plurality of the ECN trade volume over the years. The share of the pit hours in ECN trade volume declines only from 87.6% in 2004 to 73.4% in 2013. This modest trend of trading activity diffusion to the afterhours stops during the early years of the financial crisis, restarting again from 2010 on. This short interlude during the crisis years may be the result of the increasing importance of macroeconomic announcements made during the pit hours and a stronger preference for trading in more liquid times of the day.

Fig. 3.2. Number of Trades in the Pit and Electronic Markets

The figure shows the 22-day moving average of the daily number of trades of the 30-Year U.S. Treasury Futures in the pit and the electronic markets from 2004 to 2013. The left and right axes give the number of trades for the trading pit and the ECN, respectively, and the bottom axis gives the days.

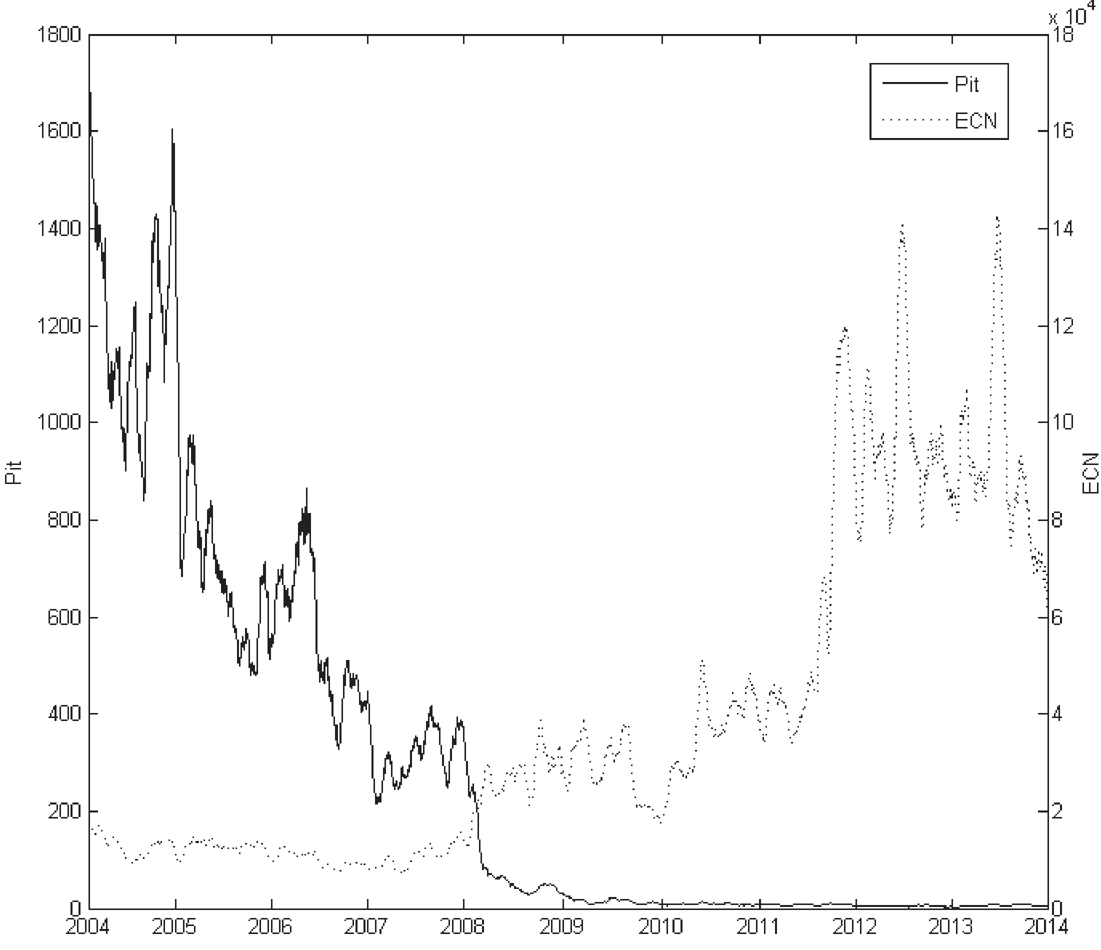


Fig. 3.3. Intraday Distribution of Trade Volume in the ECN

The black line shows the percentage share of the total trade volume for 10-minute-long intervals across the trading day. The red (green) line displays how much the share of each 10-minute interval differs from the overall average presented with the black line for days without macroeconomic news announcements less than 30 minutes before or after the pit open (close) time of 8:20h (15:00h). The pit hours (8:20h-15:00h) are marked wit a grey background.

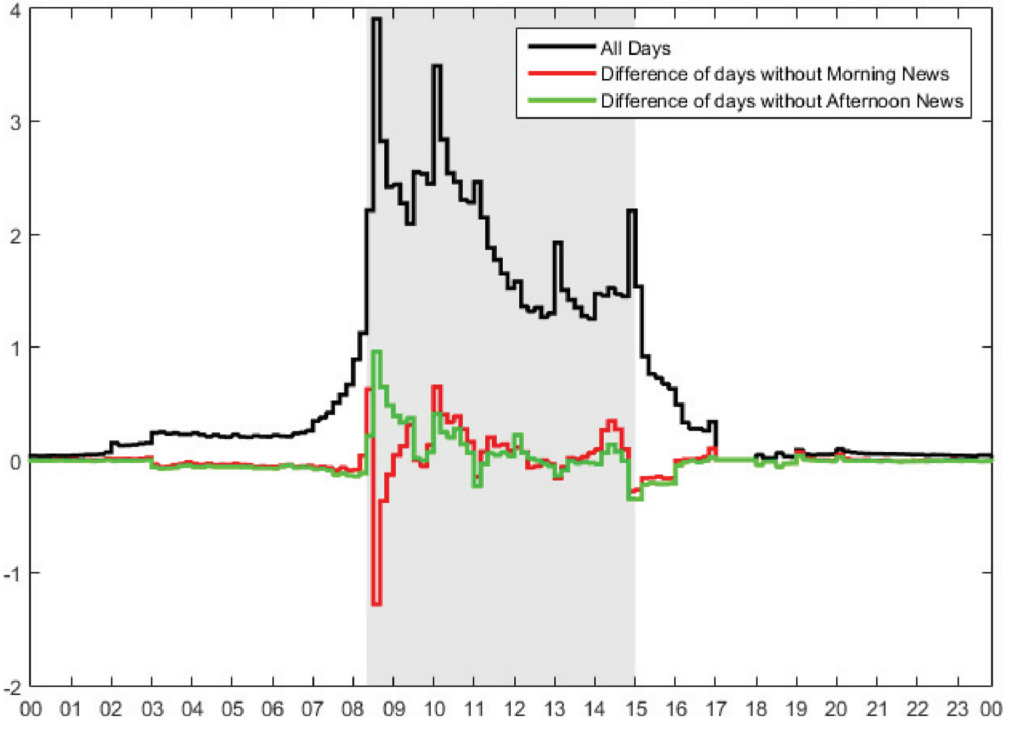


Fig. 3.4. The Distribution of ECN Trade Volume across Intraday Intervals

The figure shows the intraday distribution of the ECN trade volume for the 30-Year U.S. Treasury Futures from 2004 to 2013. The ratios of intraday periods are computed using 22-day moving averages of the trade volume figures.

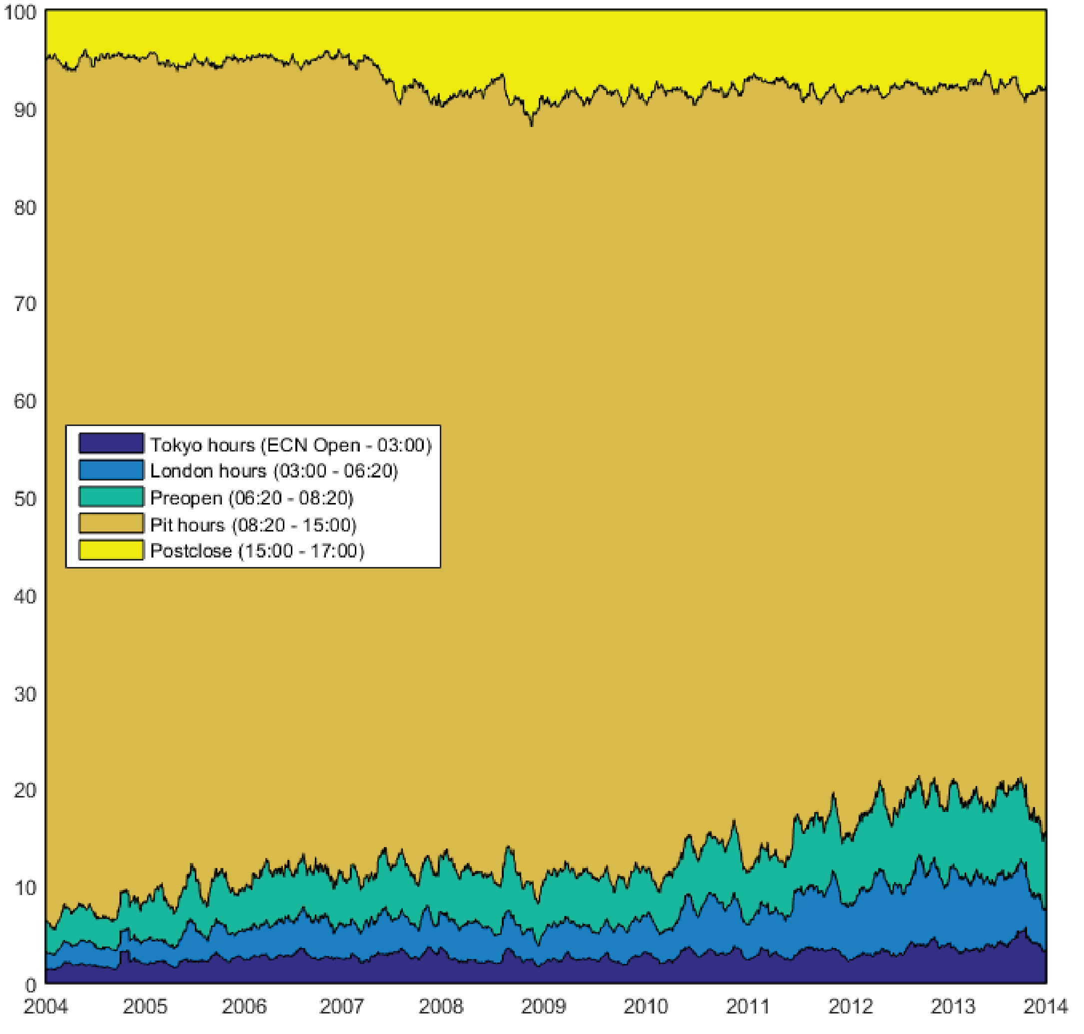


Table 3.1. Summary Statistics

We report for various intraday intervals the mean number of trades and trade volumes as well as the means and the standard deviations of 5-minute returns in basis points.

Panel A: Pit and ECN Statistics for the Pit Hours (8:20h-15:00h EST)

	Pit				ECN			
	Tr. Vol.[a]	Num. of Tr.	Mean	Std. Dev.	Tr. Vol.	Num. of Tr.	Mean	Std. Dev.
2004		1232	-0.129	11.622	191019	11124	0.015	6.173
2005		671	0.103	7.530	235934	10655	0.012	4.792
2006		559	-0.100	6.063	256834	8167	-0.002	4.270
2007		310	-0.318	6.815	286336	8471	0.014	4.799
2008		69	0.381	14.554	220149	20509	0.045	7.704
2009		14	-0.617	15.698	148358	21836	-0.022	8.351
2010		9	0.189	12.742	208236	27794	0.024	6.097
2011		7	-0.281	14.626	234472	45446	0.117	6.664
2012		5	-0.134	12.265	225187	68412	0.017	4.976
2013		6	-0.392	12.403	240779	63771	-0.036	5.048

Panel B: ECN Statistics for the Afterhours

	Tokyo hours (ECN Open-03:00h EST)				London hours (03:00h-6:20h EST)			
	Tr. Vol.	Num. of Tr.	Mean	Std. Dev.	Tr. Vol.	Num. of Tr.	Mean	Std. Dev.
2004	4245	389	-0.038	2.537	4439	342	0.027	2.493
2005	6249	407	-0.011	1.874	7443	437	0.009	2.471
2006	8397	401	0.012	1.734	10754	432	-0.036	2.287
2007	10076	512	0.008	2.011	12836	532	0.003	2.649
2008	7089	1126	0.018	2.845	9862	1188	0.076	3.725
2009	4454	1093	-0.006	2.585	6354	1254	0.000	3.771
2010	7912	1772	0.017	2.507	11852	2138	0.034	3.511
2011	9383	2491	0.005	2.813	16177	4161	-0.029	4.114
2012	10268	3459	0.000	2.181	21501	7794	0.037	3.453
2013	12707	3968	0.022	2.273	22273	6868	-0.022	3.150

	Preopen (6:20h-8:20h EST)				Postclose (15:00h-17:00h EST)			
	Tr. Vol.	Num. of Tr.	Mean	Std. Dev.	Tr. Vol.	Num. of Tr.	Mean	Std. Dev.
2004	7424	524	0.004	3.025	11028	618	0.102	2.876
2005	12528	663	-0.030	3.002	14706	623	0.069	2.476
2006	15469	576	-0.013	2.676	15451	471	0.005	2.180
2007	18458	693	0.068	3.355	24186	737	-0.046	3.133
2008	15337	1800	0.003	5.212	24418	2291	0.099	5.427
2009	9660	1723	-0.083	5.044	16369	2220	-0.032	4.930
2010	15379	2451	-0.079	4.336	22376	2579	0.016	3.406
2011	19388	4388	0.025	4.882	24141	4321	0.015	4.564
2012	24063	7733	-0.032	3.887	24796	6437	0.012	2.935
2013	26840	7026	-0.074	3.672	25521	6222	-0.010	3.075

[a]We do not have trade size data for pit trades and therefore we are not able to compute the trade volume statistics for the trading pit.

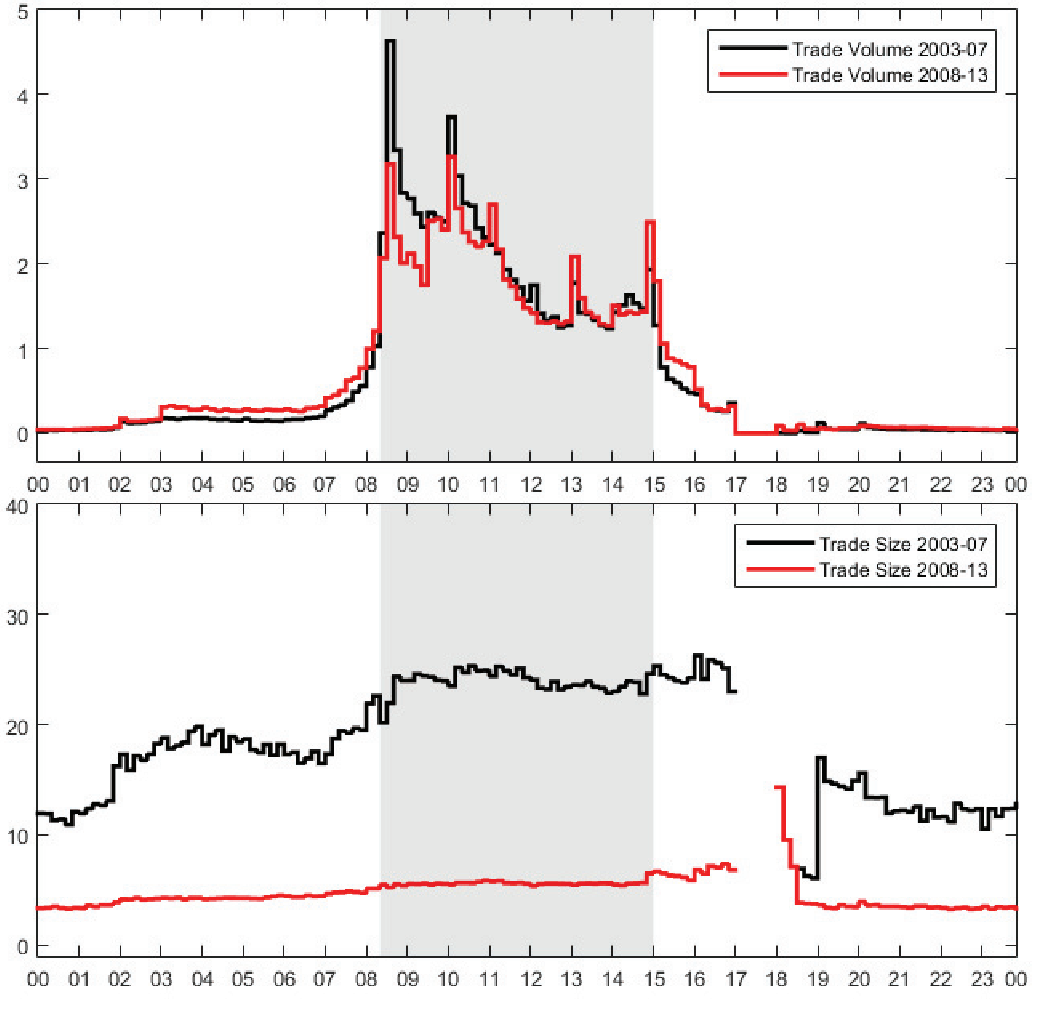
This mild diffusion of trading activity to the afterhours happens unevenly across afterhours periods. Panel B of Table 3.1 reports summary statistics of four afterhours periods: Tokyo hours (ECN Open-03:00h EST), London hours (03:00h-6:20h EST), preopen (6:20h-8:20h EST) and postclose (15:00h-17:00h EST). Compared to other parts of the day, the ECN trade volume during the pit hours has the lowest relative increase during our 10-years-long sample period. While the pit hours volume grows by 26.0%, the trade volumes in the Tokyo hours, the London hours, the preopen and the postclose increase by 199.3%, 401.7%, 261.5% and 131.4%, respectively. However the changes in absolute values amount to less spectacular 21.0% more increase during the afterhours compared to the pit hours: The total trade volume increase during the afterhours from 2004 to 2013 is 60,204 contracts, compared to a magnitude of 49,760 for the pit hours.

Building on the premise that the introduction of the Globex platform in January 2008 seals the fate of pit trading, as shown in Figure 3.2, we split the sample period into pre-Globex (2004-2007) and Globex (2008-2013). Figure 3.5 contrasts the intraday distribution of ECN trade volume and trade size for these two periods. Although the clustering of the trade volume during the pit hours is virtually the same in the two periods, the average trade size is more than halved during the Globex period and gets more even across the day. The reduction in average trade size signals the introduction of aforementioned algorithmic trading facilities to execute big orders with a series of small trades. The small increases at both 2:00h EST and 3:00h EST relate to the changing hours of the London market open for a few days in each year due to daylight saving time differences.

Lastly, we use modified versions of two frequently-used measures of liquidity and price discovery to compare the pre-Globex and Globex periods. The Amihud (2002) illiquidity measure (AIL) and weighted price contributions (WPC) are adapted to measure illiquidity

Fig. 3.5. Intraday Distribution of Trade Volume and Trade Size in the ECN

The figure shows the percentage share of the total trade volume and mean trade sizes for 10-minute-long intervals across the trading day for two time intervals: from 2004 to 2007 and from 2008 to 2013. The pit hours (8:20h-15:00h) are marked wit a grey background.



and price discovery in 10-minute intervals across the day. They are defined as

$$\begin{aligned} AIL_t &= \frac{1}{N}\sum_{n=1}^{N}\frac{\Delta p_{n,t}}{p_{n,t}Vol_{n,t}}, \\ WPC_t &= \sum_{n=1}^{N}\frac{|\Delta p_n|}{\sum_{n=1}^{N}|\Delta p_n|}\frac{\Delta p_{n,t}}{\Delta p_n}, \end{aligned} \tag{3.9}$$

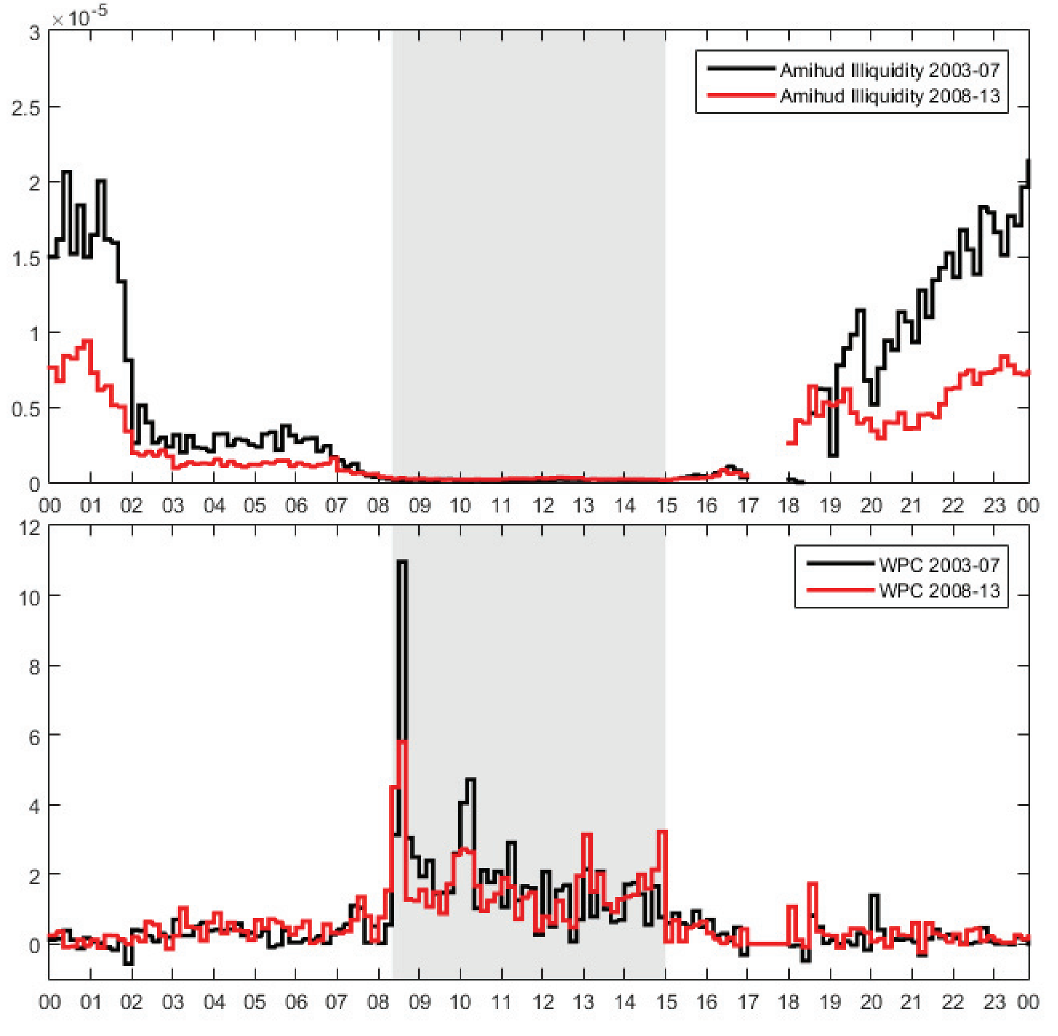
where $\Delta p_{n,t} = p_{n,t} - p_{n,t-1}$, $p_{n,t}$ is the price at intraday time t of day n with $t = 1, ..., T$ and $n = 1, ..., N$, $\Delta p_n = p_n - p_{n-1}$, p_n is the price at the pit close (15:00h EST) of day n and $Vol_{n,t}$ is the ECN trade volume for the intraday interval from time $t-1$ to t at day n.

Figure 3.6 compares the intraday distribution of liquidity and price discovery with these preliminary indicators. In line with the mild diffusion of trade volume to the afterhours in Figure 3.5, the Amihud illiquidity measure drops for the afterhours, but remains far larger than the pit hours illiquidity. The average illiquidity of the pit hours measured by the average price change generated by the same amount of dollar-volume increases by 61.5% from pre-Globex to Globex years, mainly due to a surge at the start of the financial crisis. However the pit hours remain 5.7 times more liquid than the London and preopen hours (a drop from 18.2 times in the pre-Globex years), 1.9 times more liquid than the postclose hours (a decline from 3.7 times) and 27.5 times more liquid than the Tokyo hours (a drop from 96.2 times).

The WPC statistics in Figure 3.6 indicate a small trend towards the dispersion of price discovery to the afterhours in line with the changes in trading activity. The total contributions of the pit hours decrease from the 77.0% share of the pre-Globex years to 67.4% for the Globex years. The total contributions at the London and preopen hours increase from 12.5% to 18.3% and that of the Tokyo hours from 5.0% to 10.0%. Unlike other afterhours periods, the share of postclose hours in price discovery experiences a drop from 5.5% to 3.5%. We also note a slight shift of informativeness from day open to day close parallel to the shift in trade volumes in Figure 3.5.

Fig. 3.6. Intraday Amihud Illiquidity and Weighted Price Contributions of the ECN

The figure shows the mean Amihud illiquidity measures and weighted price contributions for 10-minute-long intervals across the trading day for two time intervals: from 2004 to 2007 and from 2008 to 2013. The pit hours (8:20h-15:00h) are marked wit a grey background.



All in all, we document that a large portion of trades happen during pit hours, while there is only a modest trend towards the dispersion of trading activity to the afterhours. The introduction of the Globex platform certainly improved the conditions of the afterhours for trading. However this improvement mostly remains an amelioration over the past conditions of the afterhours rather than catching up with the advantages of the pit hours. In particular, relative to the pre-Globex period the same total dollar-volume generates significantly less price change during the afterhours, but this price change remains still at least double of that generated during the pit hours. Afterhours prices become mildly more informative as the share of the daily price change generated by the afterhours increases from a quarter to a third. Only in the average trade sizes we find an equalization across different times of the day, which reduces the ability to trade with relatively bigger sizes in the pit hours without signalling one's trading objectives.

3.4 Why does the trading activity cluster around the pit hours?

In this section we test the two hypotheses outlined in Section 3.2.1. We firstly check the informativeness of pit trading over time, as trading with more informative prices may be a reason for ECN participants to choose to trade during the pit hours. Secondly, we investigate how price informativeness and costs related to adverse selection and price impact affect trading activity during the afterhours.

3.4.1 Is pit trading informative?

In this section we examine Hypothesis 1 by measuring the share of the trading pit in price discovery using the methodology presented in Section 3.2.2 and relating its changes to the shifts in activity clustering. Table 3.2 reports the information share results for the pit and the ECN over the years. We estimate the parameters of the structural model represented by Eq. (3.1) using GMM and compute the information shares of both venues as in Eq. (3.4). The trading pit accounts for 32.0% of price discovery in 2004 compared to 65.1% for the ECN. This figure drops to 21.8% already in 2007, while the share of the ECN increases to 73.9%. In line with the reduction of the number of trades in the pit, in 2008 the information share of the pit further declines to 11.5%, whereas the ECN sustains its dominance with a share of 69.8%.[7] From 2009 on, the price staleness in the pit due to the pit hosting only 14 trades per day, as shown in Table 3.1, impedes inferences on price discovery.[8]

Table 3.2. Yearly Information Shares for the Pit and the ECN

We report the yearly information shares estimated using data sampled at 5 minute frequency. The information share estimates are in percentages.

	Information Shares	
	ECN	Pit
2004	65.1	32.0
2005	71.2	25.0
2006	65.5	28.1
2007	73.9	21.8
2008	69.8	11.5

These results attest to the informativeness of pit trading compared to its share in the number of trades as well as its rapid demise with the sophistication in electronic trading. From 2004 to 2007, in spite of the shift in the ratio of the number of trades during the pit hours from seven-fold to 22-fold in favor of the ECN, the share of the pit in price discovery is reduced relatively milder, from one-third to one-fifth. Even when the more dramatic change caused by the introduction of the Globex platform in 2008 increases the ratio of the number of trades to 243-fold in favor the Globex, the pit still retains an information

[7]In a simulation study, available upon request, we tested the effect of price staleness on the measurement of price discovery. This study implies that the increase of the price staleness in the pit data would decrease estimation accuracy, in particular by underestimating the pit information share. Thus the sizable information share of the trading pit even after the reduction in pit trades starting with 2008 is probably not an overestimate.

[8]The measurement of information shares requires the cointegration of the price series of the pit and the ECN, whereby the common stochastic trend represents the underlying efficient price process. Due to the severe reduction of pit activity with the introduction of the Globex Platform in 2008, the pit data exhibit an increasing amount of staleness, i.e. lack of price change due to the lack of trade price updates. This leads to the rejection of the hypothesis of cointegration between the price series of the pit and the ECN for an increasing majority of days. A similar problem emerges with the statistical testing of the validity of the model for the data. From 2009 on, Hansen's J-test starts rejecting the null hypothesis of model validity for the data at the 5% significance level, even after excluding days which fail the cointegration test.

share of more than 10%. However, for the later years even if the few remaining pit trades continue to be very informative in line with the past evidence, their rarity should bring down the contribution of the trading pit to price discovery. Therefore the informativeness of the trading pit should have substantially declined over the second half of the sample, even if not proportional to the drop in the number of pit trades.

The substantial reduction in pit activity and price discovery from 2008 on does not cause a similar shift in trading activity clustering around the pit hours. As shown in Figure 3.4, the modest trend towards the diffusion of trading activity to the afterhours actually halts in 2008. The financial crisis may have acted as a strong counterforce by increasing the importance of macroeconomic news announcements during the pit hours. However when this possible effect subsides during the recovery of the later years and the trading pit is reduced to a symbolic venue with less than 10 trades per day during the last four years of our sample, we still do not observe a decline of the same order of magnitude for the activity clustering.

In summary, we reject Hypothesis 1 that market participants' preference for trading during the pit hours is due to the informativeness of pit trading. The pit has a considerable share in price discovery compared to its share in trading activity at least in the first half of our sample. However the rapid shrinkage of pit activity after the implementation of the Globex ECN is not accompanied by a shift in the activity clustering.

3.4.2 What determines trading activity during the afterhours?

In this section we use the estimates from the structural model presented in Section 3.2.3 to evaluate whether the low trade activity during the afterhours is determined by a feedback mechanism postulated in Hypothesis 2. Firstly we present the intraday estimates of price informativeness, information asymmetry and price impact of trades. Then we provide a preliminary analysis by comparing the dynamics of the variable estimates and trading activity across four afterhours periods. Lastly we test whether the effects of the three variables on trading activity conform with our second hypothesis.

Figure 3.7 presents the dynamics of trade volume and the estimates of the public information variance from the structural model presented in Eq. (3.5) and (3.6). We estimate the model 20 times for separate six-months sub-periods. Trade volumes follow a similar pattern across years, as in Figure 3.5: Trading is very limited until the opening of the London market at 3:00h EST (or 2:00h EST depending on the day-light saving time differences) and shows a dramatic rise from 7:00h EST on. The level of this pattern changes in line with the annual averages of trade volumes presented in Table 3.1. The estimates of public innovation variance σ_t^2, which we use as a proxy for the price informativeness, follow a similar pattern, but with different levels. While trade volume peaks during the first halves of 2007 and 2013, innovation variance peaks in the second halves of 2008 and 2011. Thus trading activity peaks at the end of stable periods like the last half-year before the Quant Meltdown of August 2007 or the U.S. recovery after the financial crisis. In contrast, the innovation variance climaxes during the heights of crises like the dramatic reduction of the target interest rate by the Fed in December 2008 or the downgrade of U.S. debt by Standard and Poor's in August 2011. Therefore, in a given day, the amount of price innovation tends to be proportional to the amount trade activity; however, trades become more informative during volatile crisis periods.

Fig. 3.7. Trade Volume Figures and Public Innovation Variance Estimates

The figure shows, from top to bottom, the intraday variation in trade volume and the estimates of the public innovation variance σ_t^2 from the model presented in Section 3.2.3 for 20 half-yearly intervals from 2004 to 2013. Both the estimations and the trade volumes use data sampled at 5-minutes frequency. Note that the hour axis flows from right to left for the plots on the right, to ease their presentation.

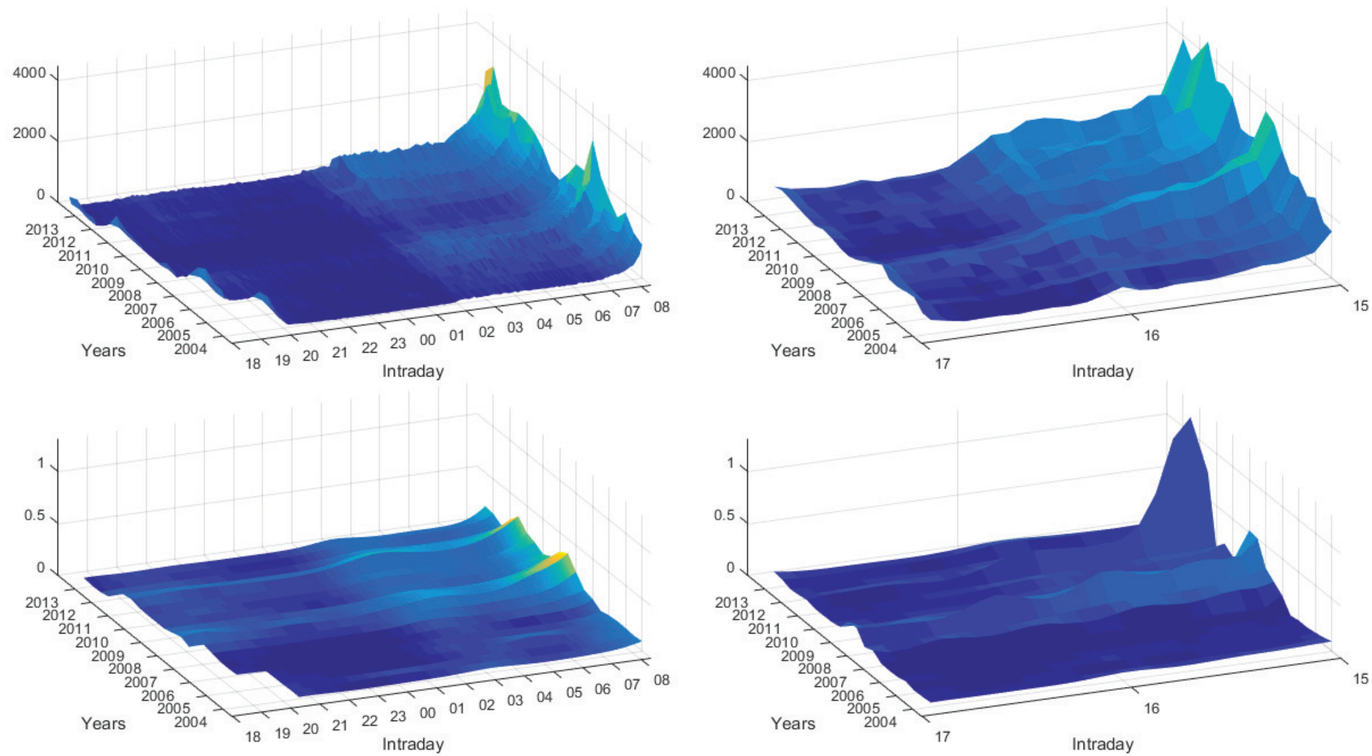
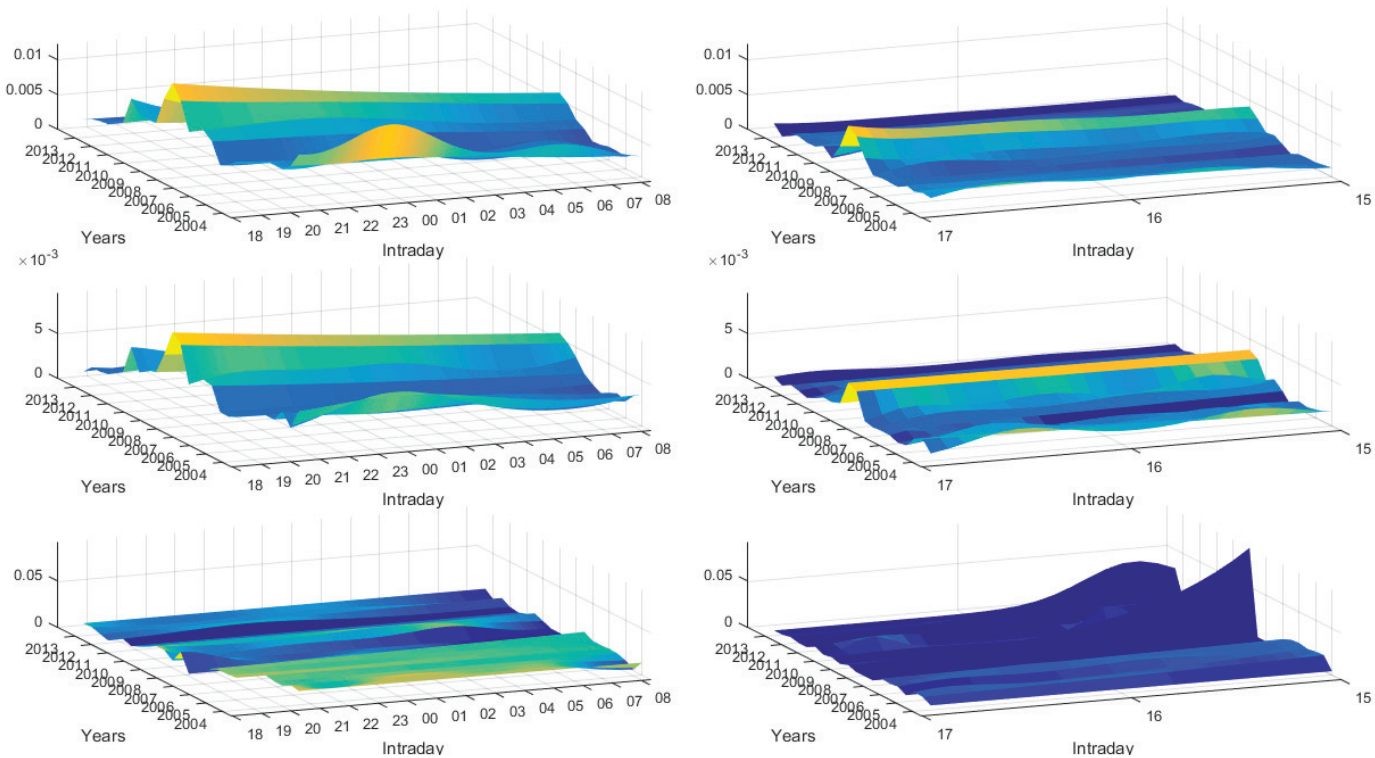


Figure 3.8 displays the other three time-varying parameter estimates. The levels of the coefficients of the unexpected order flow θ_t and the price impact δ_t move in line with the innovation variance σ_t^2, peaking at the second halves of 2008 and 2011. However, in contrast to the innovation variance σ_t^2, the noise variance ω_t^2 and the coefficients of the unexpected order flow θ_t and price impact δ_t follow a flatter pattern. The two coefficients display a clear downwards slope towards the market open in line with the negative relationship with the rising trade volume, as predicted in Hypothesis 2.

Fig. 3.8. Information Asymmetry, Price Impact and Noise Variance Estimates

The figure shows, from top to bottom, the estimates of the information asymmetry coefficient θ_t and price impact coefficient δ_t and the noise variance ω_t^2 from the model presented in Section 3.2.3 for 20 half-yearly intervals from 2004 to 2013. Both the estimations use data sampled at 5-minutes frequency. Note that the hour axis flows from right to left for the plots on the right, to ease their presentation.



The shares of the innovation and noise variances in the total price variance highly depend on the time of the day. As we have mentioned, the innovation variance σ_t^2 measures the contribution of public information because we control for the effect of private information signals in the evolution of the latent price process in Eq. (3.5). The estimated public information variance, presented in the bottom plots of Figure 3.7, accounts for 30.8% of the realized volatility of 5-minute returns in the night hours, 12.1% in the preopen and 66.0% in the postclose. The increase of the contribution of public information from the preopen to the postclose is in line with most models on market microstructure indicating such a decline in information asymmetry over the trading period (Kyle, 1985; Glosten and Milgrom, 1985; Foster and Viswanathan, 1990; Easley and O'Hara, 1992). By contrast, the proportion of the noise variance, presented in the bottom plots of Figure 3.8, to the total variation, measured by the realized volatility of 5-minute returns, stays at a similar level during the preopen and the postclose, respectively 7.2% and 7.3%, but nearly doubles its share during the night hours with 13.2%. With the introduction of the Globex Platform in 2008, these shares move almost proportionally across all intraday time periods: The portion of the public information in the total price variation increases on average 29.0% across the day, whereas that of the noise variance drops by 75.6%.

The price changes stemming from the price impact of trades and the private information signalled by the unexpected order flow account for the remainder of the price volatility, as can be seen in Eq. (3.5) and (3.6): Taking the public information and noise variances into consideration, 55.9% of the price variation during the night hours, 81.0% of the preopen and 26.7% of the postclose variation occurs due to these two factors. The full effect of the the price impact (private information signals), i.e. the product of the estimated coefficient with the order flow (surprise) data, causes on average a price change of 0.25 (0.18) standard deviations during the night hours, 0.26 (0.24) during the premove and 0.18 (0.12) during the postclose.

For a preliminary analysis of the afterhours dynamics, we take the averages of the parameter estimates for the four afterhours periods defined in the previous section: Tokyo hours (ECN Open-03:00h EST), London hours (03:00h-6:20h EST), preopen (6:20h-8:20h EST) and postclose (15:00h-17:00h EST). The ratios of noise variances to innovation variances, exhibited in Figure 3.9, indicate that prices during each of these afterhours periods become more informative around the implementation of Reg NMS and the introduction of the Globex platform. In all six-month subperiods the noisiness of prices decreases monotonically from the ECN open to the pit open due to the rise of innovation variances. During the Tokyo hours the noise variance even exceeds the innovation variance for five out of eight pre-Globex half-years. We find the preopen period to be almost always more informative than the postclose period as in Barclay and Hendershott (2003) and He et al. (2009).

Figure 3.10 presents the afterhours averages for the unexpected order flow coefficient, which is a measure of information asymmetry. By contrast to the noise-to-innovation ratios, the financial crisis has a far bigger effect compared to the concurrent advances in electronic trading. Although information asymmetry reaches its lowest levels at the end of our sample period, it peaks instead of diminishing in 2008 and 2009. The intraday picture exhibits a recurring pattern of monotonic decline in information asymmetry from the ECN open to the ECN close for the afterhours. The first half of 2007 and the second half of 2011 constitute the sole exceptions with small increases in the preopen compared to the London hours. This finding provides further support to the decline in information asymmetry over the trading period, combined with the aforementioned increase of the contribution of public information to the total price variation from the preopen to the postclose.

The price impact of trades constitutes a major obstacle for the proliferation of trading. Therefore the sustainability of activity clustering during the pit hours relies on the resilience of the price impact during the afterhours. The mean price impact coefficients during afterhours periods, shown in Figure 3.11, peak in 2008 and 2009 and return to

Fig. 3.9. Noise-to-Innovation Ratios During Afterhours Periods

The figure shows the mean noise-to-innovation ratios in four afterhours periods. The ratios are computed by dividing the noise variance ω_t^2 to the innovation variance σ_t^2 and taking their averages for each afterhours period. The parameters are estimated using the model presented in Section 3.2.3 for 20 half-yearly intervals from 2004 to 2013.

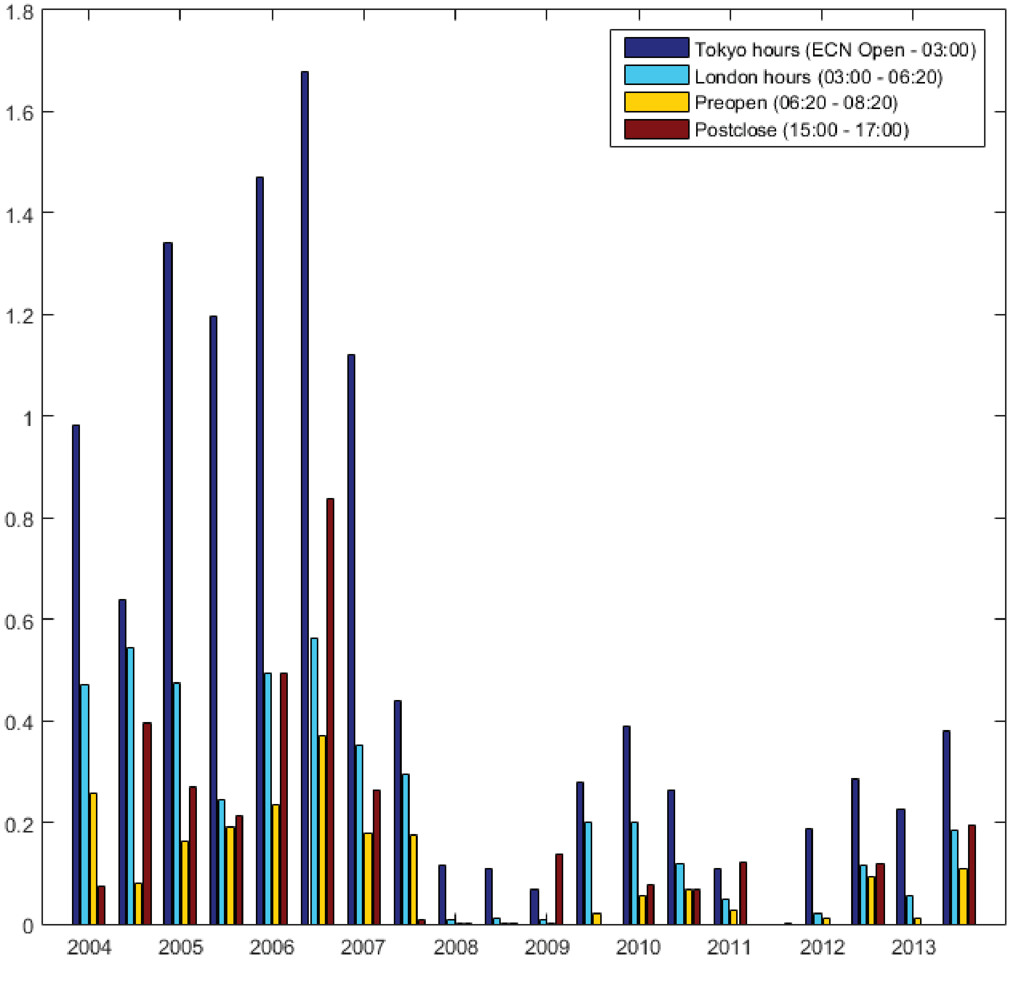
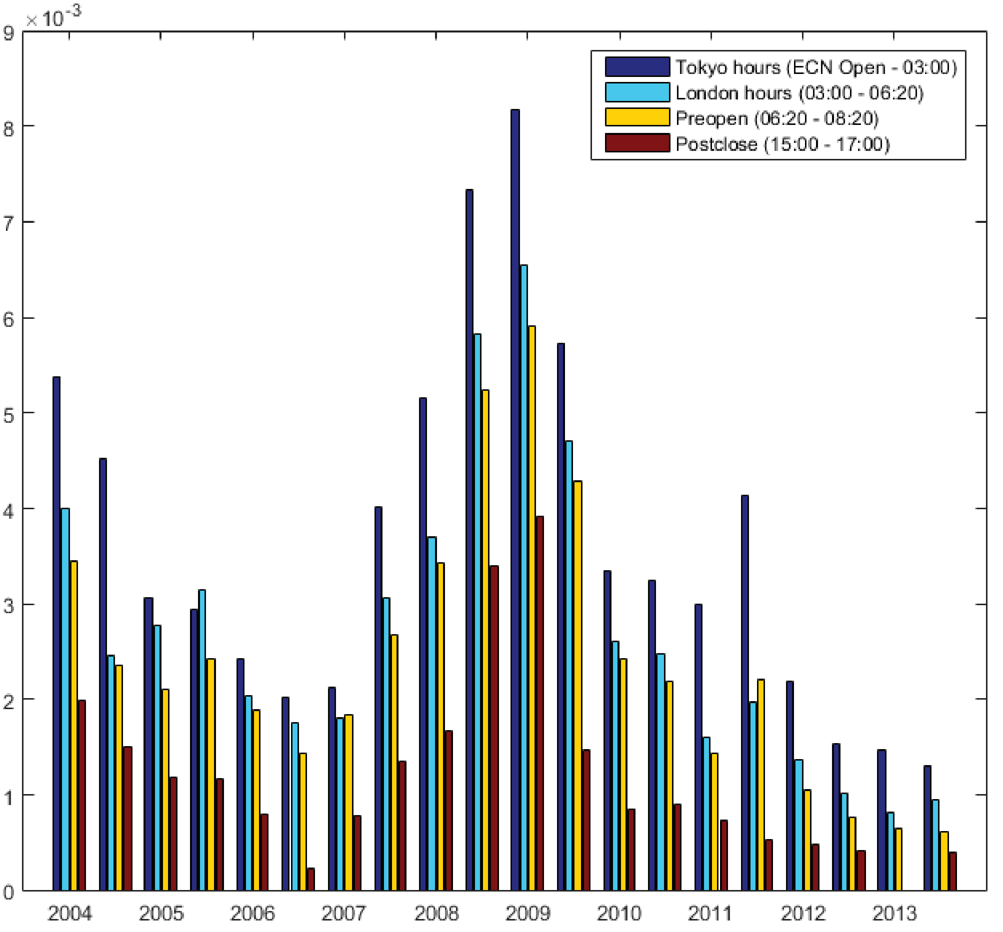


Fig. 3.10. The Magnitudes of Unexpected Order Flow Coefficients During Afterhours Periods

The figure shows the mean of the unexpected order flow coefficient θ_t in four afterhours periods. The parameters are estimated using the model presented in Section 3.2.3 for 20 half-yearly intervals from 2004 to 2013.



their prior level in 2010. We see a downward shift in price impact only from 2012 on. In the first half of 2012, the price impact coefficient drops again to its pre-Globex low in 2006-2007 and reduces even further from the second half of 2012 on. The gap between night hours and preopen values also decreases in the last years. The overall decrease in price impact and the decrease of the differences across afterhours periods constitutes some evidence for the decline of the price impact barrier against trading during the afterhours.

The feedback mechanism postulated in the second hypothesis requires the three market variables to have a significant explanatory power on the trade activity. In order to test this, we regress the changes in the afterhours trade volume on the changes in the proxies of price informativeness, information asymmetry and price impact. As the regressors are estimates themselves, we use the Murphy and Topel (1985) method to calculate correct standard errors. Table 3.3 reports for each half-year the results of separate regressions for each variable and a regression using all three variables. Price informativeness emerges as the most significant factor explaining intraday changes in trading activity. It has a strong positive relationship with trading activity as postulated and its effect is significant at 5% level in 13 out of 20 regressions. The coefficient displays little fluctuation across years and indicates that a percentage change in price informativeness relates to a percentage change of the same magnitude for trading activity.

The results for information asymmetry and price impact express a weaker relationship with trading activity. Out of the 20 regressions, the effect of information asymmetry is insignificant at 5% level for four separate regressions and 11 combined ones. Price impact has an insignificant effect for 11 regressions of both types. In the significant cases the signs of the coefficients are almost always in line with the negative effect posited by the second hypothesis. The few exceptions are either very small in terms of magnitude, as is the case for two positive coefficients of information asymmetry in separate regressions, or emerge due to the interaction between price impact and information asymmetry in combined regressions. Including the coefficients of the unexpected order flow (information

Fig. 3.11. The Magnitudes of Price Impact Coefficients During Afterhours Periods

The figure shows the mean of the price impact coefficient δ_t in four afterhours periods. The parameters are estimated using the model presented in Section 3.2.3 for 20 half-yearly intervals from 2004 to 2013.

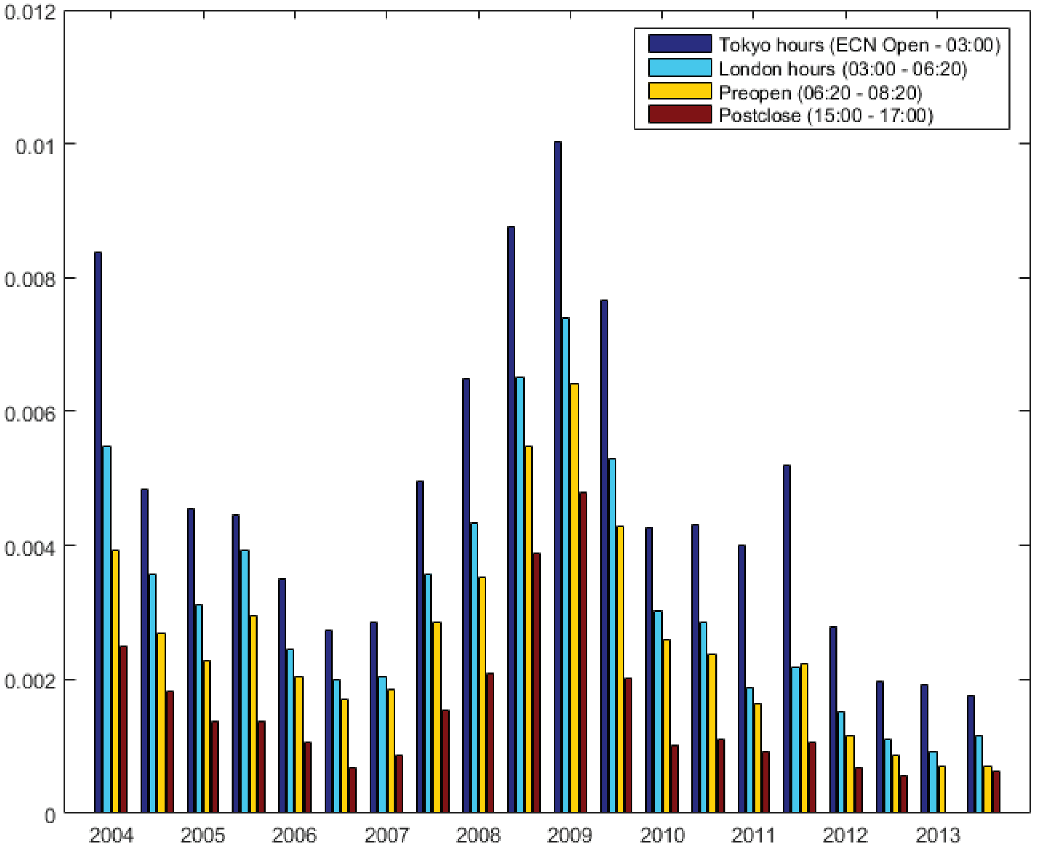


Table 3.3. Regressions to Test for activity clustering

We report the coefficient estimates for the regressions of trade volume on each of the variable estimates separately. Both the independent and dependent variables are in log differences. The Time column gives the half-year period the trade volume data and the time-varying estimates are based on. The IV, UOF and PI columns give coefficient estimates of public innovation variance, unexpected order flow coefficient and price impact coefficients. These are followed by the mean adjusted R^2 of the separate regressions and the adjusted R^2 of the combined regression. The superscript *** marks significance at level 1%, ** at level 5%, and * at level 10%. As the regressors are estimates themselves, we use the Murphy and Topel (1985) method to calculate correct standard errors.

	Separate Regressions				Combined Regression				
Time	*IV*	*UOF*	*PI*	adj. R^2	*IV*	*UOF*	*PI*	adj. R^2	N
2004a	1.4***	−0.6***	−1.3***	4.0	1.1***	−0.2	−0.6*	5.9	164
2004b	0.3	−7.8**	−4.2***	1.1	−0.1	2.2	−4.8***	2.7	164
2005a	1.2***	−3.9**	−3.6***	4.0	0.8**	5.3**	−5.2***	5.9	164
2005b	0.8**	−0.6***	−2.0**	2.6	0.3	−0.6***	−2.0***	5.1	164
2006a	0.8***	−0.4**	−6.0**	2.5	0.8***	0.0	−4.9**	5.6	176
2006b	0.6**	0.0	−2.2**	0.5	0.5**	0.0	−1.8**	1.0	176
2007a	0.4**	−4.4**	−1.3**	0.5	0.2	−4.8***	−1.4***	1.0	176
2007b	0.4***	−0.1	−0.2	1.1	0.3**	0.8	−1.0	3.4	176
2008a	1.6***	−0.3**	−0.9	2.2	1.7***	−0.2**	0.4	5.2	182
2008b	0.6**	−25.3***	−2.8	2.2	0.4	−19.5**	−0.2	3.2	182
2009a	0.6***	−31.4***	−1.3	3.4	0.5***	−30.0***	1.5*	8.4	182
2009b	1.2***	−20.8***	0.2	3.9	0.9***	−21.2***	−2.9***	10.3	182
2010a	1.3***	−12.6***	−0.4*	4.8	1.2***	−4.0	0.0	11.5	182
2010b	1.0***	−13.7***	−1.0	5.0	1.0***	2.5	−1.4	11.6	182
2011a	0.7***	0.4	0.3	2.1	1.0***	1.1	−0.2	9.1	182
2011b	0.1	0.0**	0.1	−0.1	0.0	−0.2**	0.9*	0.3	182
2012a	1.4***	−0.7**	0.0	5.6	1.4***	−0.1	0.4	14.1	188
2012b	1.1***	−4.2**	−0.3	4.2	1.0***	−5.0***	1.6**	12.8	188
2013a	1.6***	0.0**	0.2**	5.6	1.5***	0.0	0.0	11.9	188
2013b	1.0***	0.1	−0.9	3.8	1.0***	−0.2	−0.7**	11.5	188

asymmetry) and the total order flow (price impact) in one regression may not be very sensible, because they share the effect of the adverse selection in prices: Information asymmetry measures the informative component of adverse selection and price impact includes the under- and overreaction to this informative component. One may very well capture the effect of the other, if the uninformative component of the price impact is relatively small, making the coefficient estimates difficult to interpret.

The effects of information asymmetry and price impact display a strong time-dependence. The negative relationship of information asymmetry with trading activity peaks during the crisis years, from the second half of 2008 on. In the first half of 2009 a one percent change in information asymmetry causes a trading activity drop with a magnitude more than 30%. This increased negative effect subsides starting with 2011. In the case of the price impact the significance rather than the magnitude of the relationship shifts over time. Starting with the second half of 2007 the significant influence of price impact on trading activity disappears with a few exceptions. This can again be related to the rise of algorithmic trade execution systems allowing the execution of large orders in small batches, which reduces the price impact costs for electronic trading. Thus price impact of trades ceases to be a significant reason for not trading in relatively illiquid periods.

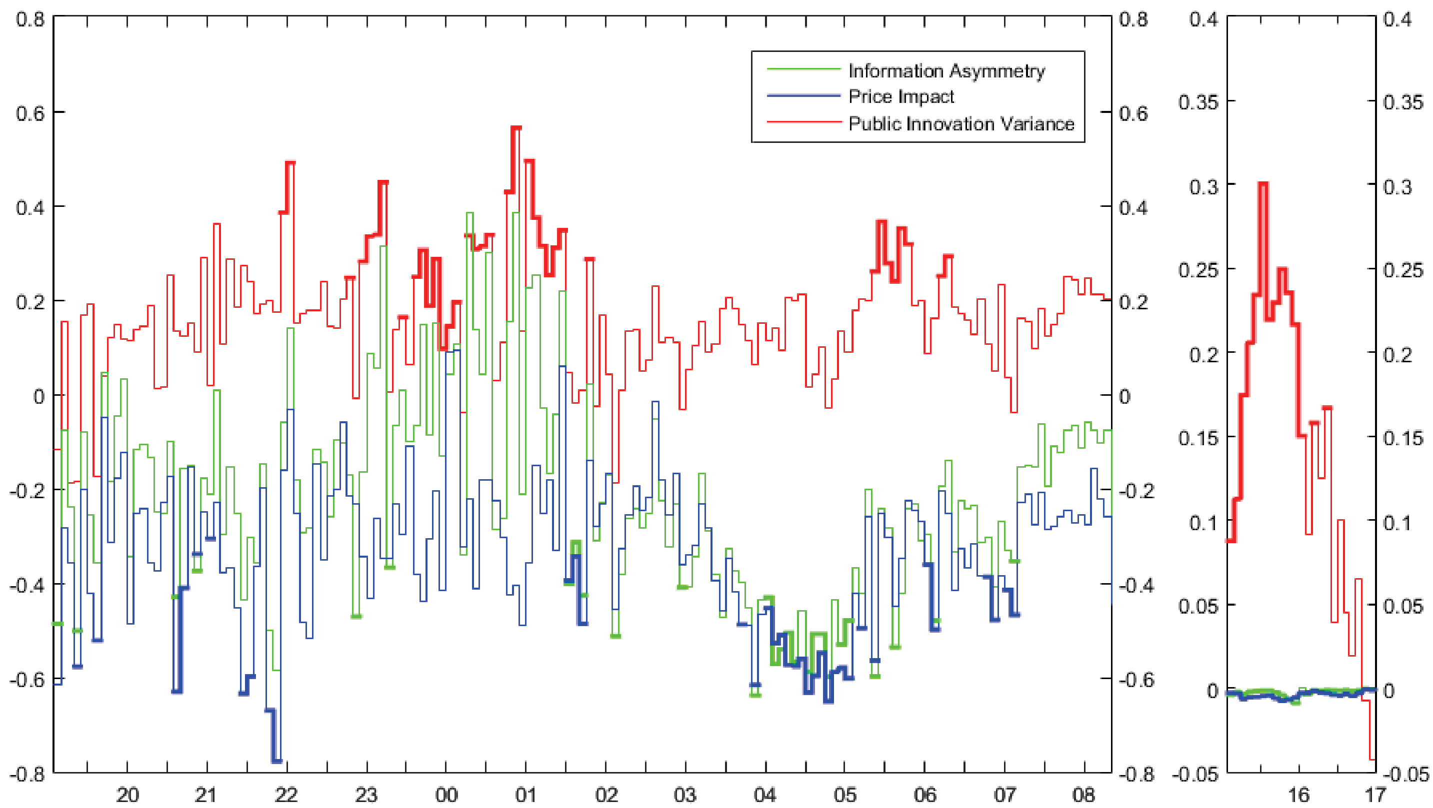
We can test these relationships also cross-sectionally for each 5-minute interval of the afterhours. Figure 3.12 reports the coefficient estimates generated by regressing the changes in trade volume on the contemporaneous changes in the estimates of public innovation variance, unexpected order flow coefficient and price impact coefficient. Although we have a small number of observations per regression (19 at best), for 51, 50 and 61 out of the 184 5-minute intervals we find the regression coefficients to be significant at 5% level for the price informativeness, information asymmetry and price impact regressions, respectively.

The signs of the statistically significant regression coefficients are in line with the time series regressions and the second hypothesis. In all cases we observe a less than one-to-one effect of percentage changes in the three variables on trading activity. Although the postclose period has the highest share of significant cases, the effects of information asymmetry and price impact are negligibly small for the year-to-year changes, mainly due to the large increases they experience during the crisis years.

In summary, we find nuanced but positive evidence for the second hypothesis. Price informativeness during the afterhours has a significant positive relationship with trading activity. The negative relation between information asymmetry and trading activity has risen particularly during the financial crisis. Price impact requires a more refined interpretation, because its negative relation with trading activity loses its significance from the second half of 2007 on.

Fig. 3.12. The Coefficient Estimates for the Cross-Sectional Regressions

The figure shows the coefficient estimates generated by regressing the changes in trade volume on the estimates of public innovation variance, unexpected order flow coefficient and price impact coefficients for each 5-minute interval during the afterhours. Results significant at 5% level are emboldened. As the regressors are estimates themselves, we use the Murphy and Topel (1985) method to calculate correct standard errors.



3.5 Conclusion

In this paper, we investigate the factors behind the concentration of trading around the pit hours. For the case of the U.S. Treasury futures, about three quarters of trades happen during this 400 minutes of a trading day. We document only a mild secular trend towards the erosion of this clustering. We find the largest trade volume increases during the trade hours of the London market for U.S. Treasury futures and the largest drops in Amihud (2002) illiquidity during the Tokyo market hours. The recent financial crisis stops this trend at least for a while, probably because of the increasing importance of macroeconomic announcements made during the pit hours and a stronger preference for trading in more liquid times of the day.

We find the informativeness of pit trades to be an unsatisfactory explanation for clustering of trading activity during pit hours. The trading pit indeed has a sizable share in price discovery before the introduction of the Globex Platform in 2008. However the substantial reduction of pit trading after 2008 does not cause a significant change in the mild erosion trend of the trading activity share of the pit hours. The last few years make this conclusion clearer as the dynamics generated by the financial crisis subside.

We use a structural model estimated with state space methods to analyze afterhours trading. Public information variance displays a strong time-varying pattern similar to trade volumes, increasing near open and close times of different markets ad towards the pit hours. We observe flatter patterns, but considerable differences across afterhours periods for the coefficients of the order flow and the unexpected order flow, measuring price impact and information asymmetry respectively. The preopen stands out as the most informative afterhours period, while the postclose has the least information asymmetry. Discounting for the effects of the financial crisis, we observe a progressive decrease during the afterhours for information asymmetry, price impact of trades and the amount of noise in prices, attributable to the improvements in electronic trading.

Our findings confirm price informativeness and costs related to information asymmetry and price impact as significant explanatory factors for activity clustering. Price informativeness during the afterhours has a stable and strong positive relationship with the distribution of trading activity. Information asymmetry generates adverse selection costs pushing liquidity traders away and its effect is particularly strong in the crisis period. Price impact costs, on the other hand, have a negative effect on trading activity until the second half of 2007, but cease to be a significant factor afterwards. We attribute this change to the improvements in algorithmic execution systems in the same period which have a documented diminishing effect on trade sizes with the introduction of the Globex Platform.

Appendix: The state space representation of the structural model

In the state space form, the structural model of Section 3.2.3 model given in Eq. (3.5) and (3.6) can be represented by these two equations:

$$p_t = \delta_t q_t + [1\ 1\ 1\ \psi_{1,j}] \begin{bmatrix} p^*_{t-1} \\ \epsilon_{t-1} \\ e_{t-1,j} \end{bmatrix} + G\varepsilon_t, \text{ where } G = [1\ 1] \text{ and } \varepsilon_t = \begin{bmatrix} \epsilon_t \\ e_t \end{bmatrix},$$

$$\begin{bmatrix} p^*_t \\ \epsilon_t \\ e_{t,j} \end{bmatrix} = \begin{bmatrix} \theta_t\left(q_t - E[q_t|q_{t-1}]\right) \\ 0_{(1+J)\times 1} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0_{1\times J} & 0 \\ 0 & 0 & 0_{1\times J} & 0 \\ 0 & 0 & 0_{1\times J} & 0 \\ 0_{J\times 1} & 0_{J\times 1} & I_J & 0_{J\times 1} \end{bmatrix} \begin{bmatrix} p^*_{t-1} \\ \epsilon_{t-1} \\ e_{t-1,j} \end{bmatrix} + H\varepsilon_t,$$

$$\text{where } H = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0_{J\times 1} & 0_{J\times 1} \end{bmatrix},$$

with $\psi_{n,m}$ a stacked row vector of ψ_j coefficients from the n^{th} to the m^{th}, $e_{t-n,m}$ a stacked column vector of disturbances e_j from time $t-n$ to time $t-n-m$, $0_{n\times m}$ an $n\times m$ matrix of zeros, I_n is an $n\times n$ identity matrix. The variance parameters are uniquely identified using the covariance matrix of the stacked disturbances $\begin{bmatrix} H \\ G \end{bmatrix}\varepsilon_t$, which comprises the innovation and noise variances:

$$E\left[\begin{bmatrix} H \\ G \end{bmatrix}\varepsilon_t\varepsilon_t'\begin{bmatrix} H \\ G \end{bmatrix}'\right] = \begin{bmatrix} \sigma_t^2 & \sigma_t^2 & 0 & 0_{1\times J} & \sigma_t^2 \\ \sigma_t^2 & \sigma_t^2 & 0 & 0_{1\times J} & \sigma_t^2 \\ 0 & 0 & \omega_t^2 & 0_{1\times J} & \omega_t^2 \\ 0_{J\times 1} & 0_{J\times 1} & 0_{J\times 1} & 0_{J\times J} & 0_{J\times 1} \\ \sigma_t^2 & \sigma_t^2 & \omega_t^2 & 0_{1\times J} & \sigma_t^2+\omega_t^2 \end{bmatrix}.$$

Chapter 4

Momentum Ignition? HFT Activity During Transitory Extreme Price Moves

This chapter is a joint project with Prof. Bruce Mizrach of Rutgers University.

4.1 Introduction

The increasing role of technology in the contemporary financial markets heightens concerns about the predatory use of speed via high frequency trading (HFT). Large price moves, especially when they revert substantially in a matter of minutes if not seconds, frequently attract comments concerning the use of the so-called momentum ignition strategy. The Concept Release on Equity Market Structure by the U.S. Securities and Exchange Commission (SEC), which has been open for comments since 2010, mentions momentum ignition as one of the two main "directional strategies that may present serious problems in today's market structure" and defines it as initiating "a series of orders and trades [...] in an attempt to ignite a rapid price move either up or down" (SEC, 2010). Similarly, the recent European Commission Regulation No 596/2014 on market abuse explicitly mentions this strategy in its supplementary annexes and uses a comparable definition based on the use of trade activity to instigate a price change.

Only a small portion of the overall HFT activity is generated by firms mainly employing liquidity-taking tactics required for the application of directional strategies such as momentum ignition. Smart liquidity providers, called as 'the new market makers', comprise the majority of the HFT activity, which directly relates to the observed positive effects of HFT firms on market quality measures such as liquidity (Menkveld, 2013; Hagstromer and Norden, 2013; Brogaard et al., 2014). However, the liquidity provided by market-making HFT firms is more likely to be consumed by opportunistic HFT firms instead of traditional liquidity traders, imposing higher adverse selection costs and spreads on the market (Brogaard, Hendershott, and Riordan, 2015; Menkveld and Zoican, 2014). This aggressive minority of HFT firms specialized in liquidity-taking also enjoys a larger share in profits than their passive counterparts dedicated to market-making, the profitability of which may even depend on the fee rebates provided by stock exchanges for their liquidity provision (Baron, Brogaard, and Kirilenko, 2012; Brogaard et al., 2014).

A recent literature of studies looks into the sources of this success by investigating specific strategies employed by HFT firms and evaluating their effects on market quality. Particularly, the practice of placing and then immediately cancelling a large number of orders, namely quote stuffing, has been attracting a variety of studies (Conrad, Wahal, and Xiang, 2015; Egginton, van Ness, and van Ness, 2013; Gai, Yao, and Ye, 2013). The ongoing debate over the explanation of this phenomenon questions whether it constitutes one part of a larger aggressive HFT tactic or can be explained by the risk management of liquidity providers, experimentations with HFT machinery and algorithms or competitive Edgeworth cycles (Baruch and Glosten, 2013; Gao and Mizrach, 2015; Hasbrouck, 2015). More directly related to liquidity-taking strategies, Clark-Joseph (2013) documents the practice of exploratory trading, the execution of small-sized trades at a loss in order to discern information on the market conditions. Similarly, Hirschey (2013) argues that the frequent lead of HFT sales and purchases over similar non-HFT trades can best be explained by the use of order anticipation strategies.

The strategy of momentum ignition has drawn the attention of academics as well as practitioners. A large but temporary price change comprises the characteristic feature of this strategy, although a consensus is lacking over its details. Such changes are usual suspects for market fraud since the use of 'pump-and-dump' tactics by lower frequency traders. Hitting one side on the market with trades for a short period constitutes a frequently mentioned method to generate such temporary price moves. Theoretically, the HFT technology can aid the precise timing and the optimal execution of such a wave of market orders. The size of the move also makes it economically sensible to be exploited by HFT firms specialized in directional trading. Unlike the HFT market-making strategies benefitting from the wideness of the bid-ask spread, liquidity-taking strategies need to cross the spread and still remain profitable, which is easier to ensure for large moves. Also, regardless of the profits some parties may enjoy during these events, large moves tend to deteriorate market quality harming unprepared market participants and raising concerns of regulators.

In the literature, potential cases of momentum ignition have been captured using two main approaches: filters for various components of the strategy and jump-detection methods. Tse, Lin, and Vincent (2012) adopt a multi-stage filter to capture momentum ignition patterns and define the tactic by a premove period of high trade volume with at most small price fluctuations and a following rapid move and reversal. Sokolov (2014) and the related study of Brogaard, Carrion, Moyaert, Riordan, Shkilko, and Sokolov (2015) employ jump-detection methods to find extreme move patterns and examine the activity of HFT firms during these events. We use a series of filters ensuring the existence of trade activity during the event and looking for a premove period with stable prices followed by an extreme price move relative to the past volatility of the stock. Compared to jump-detection methods, these filters can capture events of a variety of time lengths for the move and reversal periods. The use of jump-detection methods also generate a large amount of data, marking each unit interval for the existence or the absence a jump, which is infeasible to process for estimation in large data sets.

We analyze a message-level data set of 8,000 stocks trading on NASDAQ from July 2007 to December 2013. The resulting data set of extreme move events consists of 1,675,100 observations, corresponding to about 3.40 events per month for each stock.[1] About half of these moves end with a reversal of more than two-thirds of the magnitude of the initial move, which we deem as transitory and focus as suspects for the application of the momentum ignition strategy. Moves with less than one-third reversal and those with one-third to two-thirds reversal each constitute about a quarter of the total.

We adopt the speed of order changes as a filter for the HFT activity. This choice suits to our focus on market orders immediately resulting in trades, because many studies consider the execution or cancellation of a new addition to the order book within a very

[1]As a point of comparison, Brogaard et al. (2015) find an average of 4.75 extreme move events using the jump-detection method of Lee and Mykland (2012). Tse et al. (2012) discover 1.6 momentum ignition patterns with an average move size of 38 basis points per day for each stock of the STOXX600 Index. By contrast, we set a lower bound of 100 basis points for the price moves to ensure the economic significance of the captured events, giving an average move size of 207 basis points.

small amount of time, usually in 50-100 milliseconds to be precise, as a sign of the HFT activity (Scholtus, van Dijk, and Frijns, 2014; Hasbrouck, 2015). By contrast, among the limit orders managed by HFT firms, only the ones targeted with rapid replacement messages can be identified with this time-based filter, e.g., Hasbrouck and Saar (2013). The ability to use extensive data sets constitutes another advantage of this identification choice. So far only a few publicly available data sets, generally limited in terms of their stock and time coverage, provide explicit identification of the HFT activity.

We formalize the accounts of momentum ignition into two testable hypotheses. Firstly, the HFT activity in the premove, i.e., the minute before the move, should be in line with the move direction, implying both a correct positioning and potentially an instigative power over the move. And secondly, the premove HFT activity should relate positively with the magnitude of the reversal after the move event. Besides testing for these hypotheses, we also consider the effect of the HFT activity before and during the event on three market quality measures: the quoted spreads, market volatility and the execution-to-cancellation ratios. To test these relationships, we use a series of regressions of the move size, the degree of reversal and the market quality change on the HFT trade volume and the HFT share in the trade volume, scaled by their past averages.

We also control for the two other market conditions which can cause a temporary extreme move: fleeting liquidity imbalances and the overreactions driven by non-HFT parties. In both cases, the price move would revert considerably, once the market attracts arbitrageurs bringing in liquidity or when the overreaction subsides. Moreover, the HFT activity can precede the price moves in both events giving a false sign of causation: Due to their ability to swiftly generate trading signals from data on market conditions and asset fundamentals, HFT firms would be the first to react to and exploit liquidity imbalances and news announcements. We use the premove depth on both sides of the market relative to its past average to control for the effects of liquidity imbalances. Non-HFT trade activity is added to account for events initiated and driven by non-HFT parties, including market overreaction.

We estimate fixed effects models with an unbalanced panel of data and a three-way error component, which controls for the variation across firms, months and three intraday periods of open, midday and close. Due to the high dimensionality of our data set comprising 8,000 stocks observed over 78 months, we implement the Within estimator of Davis (2002), who extended on the one- and two-way error component models for unbalanced data developed by, respectively, Baltagi (1985) and Wansbeek and Kapteyn (1989). In order to avoid endogeneity issues caused by stock-specific events or cross-stock trading strategies based on stocks of the same industry or index, we adopt the instrumental variable approach of Hasbrouck and Saar (2013) to our variables capturing the HFT activity. We test for structural breaks in the estimated coefficients using the poolability test of Han and Park (1989).

Controlling for the premove HFT activity, we find that the HFT activity during the move and its reversal alleviates the post-event deterioration of market quality. We interpret this as the effect of HFT market-makers, given the premove HFT activity controls for the influence of HFTs using predatory tactics like momentum ignition. Hasbrouck and Saar (2013) also find a positive relationship between the HFT activity and market quality measures such as spreads, depth and volatility, even at times of falling prices and anxiety in the market. Using a smaller data set with explicit identification for HFT liquidity provision, Brogaard et al. (2015) show that HFT firms act as net liquidity suppliers around jumps, whereby they absorb the order flow and lead to smaller and more durable price moves. We find that even the liquidity-taking HFT activity during the move and the reversal is far less aligned with the move or the reversal compared to the non-HFT activity, signalling a higher prevalence to bet on mean-reversal strategies. Thus the HFT activity during the move events may tend to soften the magnitude of extreme moves.

The regression results affirm our main hypotheses concerning the use of the momentum ignition strategy. Firstly, the premove HFT activity has a statistically significant predictive power over transitory extreme moves. A 1% increase of the HFT share in trade volume in line with (against) the move relative to its past average causes a 0.49 (0.46) basis points increase (decrease) in the move size. The effect sizes for the changes in HFT trade volume itself are about ten-times smaller, but they still nearly equal that of the percentage changes in the far larger non-HFT activity. The non-HFT trade volume in the premove has been on average more than 10-times higher than the HFT trade volume, although the share of the HFT activity in the premove trade volume more than doubles from 2007 to 2013 following a monotonic trend across the years.

Secondly, the premove HFT activity on both sides on the market are associated with larger reversals, especially during the subinterval from October 2011 to December 2013. Increases in the HFT activity on both sides of the market lead to stronger reversals: A 1% change in the HFT trade volume against (in line with) the move direction increases the degree of reversal by 0.44 (0.46) basis points.

Thirdly, again during the subinterval from October 2011 to December 2013, we find some evidence for a negative relationship between the premove HFT activity and the market quality change caused by the extreme move event. We estimate a statistically significant negative effect for half of the 12 coefficients belonging to the premove HFT activity, whereas only one in the remaining half captures a significant positive effect. On average a 1% increase in the HFT trade volume (in the HFT share in the trade volume) during the premove deteriorates the market quality change by 0.20%, 0.02% and 0.10% (0.48%, 0.56% and 0.06%) for quoted spreads, market volatility and execution-to-cancellation ratios, respectively.

The subinterval with significant supportive evidence for the use of the momentum ignition strategy is situated in the aftermath of the peaks of the financial crisis. This coincides with the reported drops in HFT profitability due to a number of factors such as the reduction of especially market-making profits due to lower market volatility and the overcrowding of the HFT field. This timing also casts doubt on the external validity of the frequently-used NASDAQ data set of 2008-2009, in which the trading activity of HFT firms are explicitly identified.

The remainder of the paper is organized as follows. Section 4.2 introduces the data set and our methodology for the detection of extreme move events. Section 4.3 presents our hypotheses concerning the existence of the momentum ignition strategy and how to test them. Section 4.4 evaluates the effect of the HFT activity on the move size, the degree of reversal and the change in market quality.

4.2 Data and Summary Statistics

In this section, we firstly introduce how we use the NASDAQ data set. Then, we elaborate on our detection algorithm for extreme price moves. Lastly, we provide summary statistics for the move events.

4.2.1 The ITCH data set

We construct the full order book using data from the ITCH message feed of the NASDAQ TotalView database. The data span from July 2007 to December 2013 and consist of 8,000 randomly selected stocks out of 11,341 available ones.[2] Each message in the TotalView-ITCH data feed is time-stamped to the nanosecond and messages have three main categories: additions of displayed orders, order cancellations and order executions.

The message data contain two types of new quote additions to the order book. Market participants can choose whether to enter their limit orders into the book with their market participant ID or to trade anonymously. The additions of anonymous traders have the code A, while new limit orders with a market participant ID have the code F.

Messages can leave the order book in five ways. An E message indicates the partial or full execution of a limit order at the quoted price. If the execution happens at a price other than the original quoted price, the data set records this as a C message. Three other

[2]The use of just a subset of the whole stock universe is a temporary measure taken due to the momentary unavailability of the whole data set.

types of messages denote the cancellation of an order: Orders that are partially cancelled are designated with an X and full cancellations are reported with a D message. Lastly, a market participant can cancel an order and replace it with another order and these are denoted with a U message.[3]

We aggregate these messages in 1-second-long intervals to generate three market variables: new quote volume (A and F messages), cancellation volume (D, U and X messages) and trade volume (C and E messages). We distinguish the HFT component in the latter two variables by filtering new quote additions cancelled or executed within 50 milliseconds after their placement. Given that the average reaction time of a human to visual stimuli is around 180-200 ms (Brebner and Welford, 1980), cut-off values of 50 ms or 100 ms are frequently employed by researchers as a filter for the HFT activity (e.g., Hasbrouck and Saar, 2013; Scholtus et al., 2014; Hasbrouck, 2015).

4.2.2 The Detection Algorithm for Extreme Price Moves

The official documents as well as the accounts of practitioners generally define the momentum ignition strategy in broad strokes, without providing much specificity. As we have mentioned in the introduction, a common feature of these definitions is the use of trades to generate a rapid and large price move. As this price move is not based on new information or actual demand and supply dynamics, popular accounts also add a reversal soon after the move. We aim to capture constituents of the universe of extreme price moves to evaluate whether the trade activity of HFT firms indeed leads to the expected consequences of a momentum ignition strategy.

We detect extreme price moves by processing the order book data with a set of filters and a volatility benchmark based on the past data. As a measure of volatility, we use the

[3]For a more complete description, we refer to the documentation with the releases of the Nasdaq ITCH Total View data set available at: http://www.nasdaqtrader.com/Trader.aspx?id=itch.

high-low range for each minute, defined as

$$HL_{i,t} = \frac{p_{i,t}^{high} - p_{i,t}^{low}}{p_{i,t}^{high} + p_{i,t}^{low}}, \tag{4.1}$$

where p_t^{high} and p_t^{low} are the highest and the lowest midquote of stock i at minute t, respectively.

We use three filters to find extreme price moves:

1. Premove stability filter: The high-low range in the premove, i.e. the minute before the start of the move, should be at most the average high-low range of the same minute in the prior 22 days. This filter makes sure that the starting point is not deep into an ongoing price move, but the terminal point of a period of stable prices.
2. Large move filter: The move size, scaled by the square root of move length, should be at least 1% of the stock price and also at least 10 times the average high-low range of the same minute in the prior 22 days.
3. Trade activity filter: There should be at least one trade during the move and the reversal. This filter aims to choose only the events, during which a market participant may have profitted from the changed prices.

We restrict our study to the 380 minutes from 9:35 to 15:55, in order to avoid the high volatility and other problems associated with the market open and close, which may distort our measures. At each second, we look into the following 3 minutes in 5 second increments for a move length satisfying the large move filter. After finding the move length with the largest change scaled by time, we look into the 3 minutes following the peak/bottom in 5 second increments for the largest reversal scaled by the square root of reversal length.

This algorithm differs from the filter-based methodology of Tse et al. (2012) in two main aspects: the volatility benchmark and the premove trade activity. Tse et al. (2012) use the volatility in the recent minutes as a benchmark to evaluate the stability in the premove and the relative magnitude of the move. However intraday volatility does not necessarily follow a smooth pattern: A temporary increase in volatility may disqualify an upcoming large move, while a drop in volatility may make a following small move look like an extreme price change. We use the 22-day past average of volatility during the corresponding minute as a more reliable benchmark. Secondly, Tse et al. (2012) look for short-term increases in trading activity during the premove. We find this too restrictive as it does not represent a feature commonly mentioned in the public accounts of momentum ignition patterns.

This multi-stage filter enables us to capture a wide range of extreme move patterns. Unlike many jump-detection methods, we can pick up moves and reversals with variable time lengths. Our benchmark volatility, the minutely moving average of past high-low ranges, is by construction sensitive to the intraday volatility patterns, which usually follow a U-shape. This benchmark also relies on the past data instead of considering the whole data set and thus the recognized events coincide with those cases, which the market participants would consider as extreme moves at their occurrence.

4.2.3 Summary Statistics of Move Events

Our detection algorithm captures 1,675,100 extreme move events, corresponding to about 3.40 events per month for each stock. We differentiate three degrees of reversal after a move: The transitory moves revert more than two-thirds of the move, while the permanent ones revert less than one-third of the move. The intermediary category of reversals between one-third and two-thirds comprises the mid moves. Figure 4.1 displays the distribution of events across time. The number of extreme moves tends to decrease over time, probably due to the drop of volatility relative to the peaks of the crisis. The period of 2008-2010 has about 2.5 times more events than the period of 2011-2013.

The volatility swings during the crisis also lead to dramatic drops in the number of recognized events, such as the drops to zero for March 2008 and the period from October

Fig. 4.1. The Distribution of Extreme Move Events Across Time

The figures shows the monthly distribution of extreme move events disaggregated according to the degree of reversal. The upper figure shows the absolute numbers, while the lower plot displays the percentage shares in the total number of events. The transitory moves revert back more than two thirds of the move. The permanent moves revert back less than a third of the move. The mid moves revert back a third to two thirds.

2008 to March 2009. Especially the second period contains a sizable upwards swing in volatility followed by a substantial drop in the aftermath of the bankruptcy of Lehman Brothers. Such swings trigger the first two filters disqualifying most of the potential cases: Large volatility increases make the premove periods more volatile than the past average, thereby no potential premove period seems to have stable prices, and large volatility drops make the move sizes small compared to the past average, recorded during higher volatility levels.

Figure 4.2 displays the distribution of the extreme move events across the trading day, disaggregated by the degree of their reversals. Although the volatility benchmark depends on the time of the day, the more volatile periods of the market open and close exhibit larger numbers of extreme move events. The relative shares of the three groups corresponding to different degrees of reversal remain relatively stable, just like in their distribution across time depicted in Figure 4.1. 53.6% of the extreme moves are transitory, compared to the 22.8% share of the permanent moves.

The move sizes as a percentage of the stock price tend to increase over time, as shown in Figure 4.3. For the transitory (permanent) moves, the share of the smallest move size group containing sizes of 1% to 2% decreases from 71.9% (63.7%) in the period of 2008-2010 to 64.9% (56.7%) in the period of 2011-2013, while all other groups increase their shares. This move size increase coincides with the increasing share of stocks with lower market capitalization and lower trade volumes, as depicted in Figures 4.4 and 4.5. Thus the decline of market volatility reduces especially the number of extreme move events at large-cap stocks, which experience smaller moves due to their higher liquidity.

Figure 4.6 depicts the swings in the trade volume during five periods of an event: the premove, the two halves of the move and those of the reversal. The trading activity in the direction of the move, i.e., the volume of buyer/seller-initiated trades for up/down moves, tends to surpass the trades with the opposite direction at the premove. This asymmetry heightens during both halves of the move. During the reversal, we observe the mirror

Fig. 4.2. The Distribution of Extreme Move Events Within the Day

The figure shows the number of extreme move events per 5-minute interval disaggregated according to the degree of reversal. The moves are allocated to each interval according to their stating points. The upper figure shows the absolute numbers, while the lower plot displays the percentage shares in the total number of events. The transitory moves revert back more than two-thirds of the move. The permanent moves revert back less than one-third of the move. The mid moves revert back one-third to two-thirds.

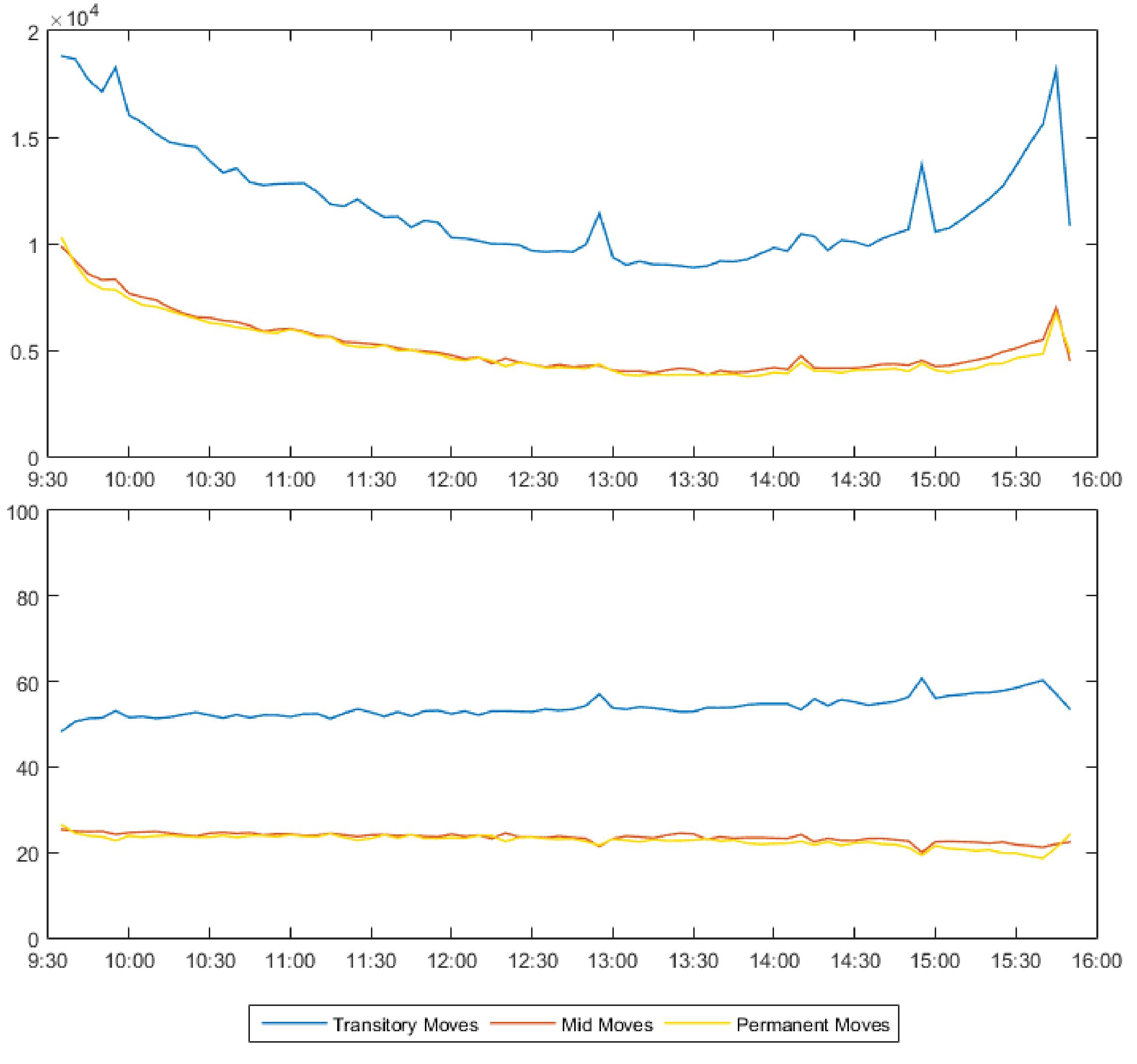


Fig. 4.3. The Move Size Distribution of Extreme Move Events Across Time

The figure shows the monthly shares of 7 move size groups in the number of extreme move events. The upper (lower) figure shows the distribution of the transitory (permanent) moves. Move size increases from blue to yellow. The groups consist of move size intervals of 1% to 2%, 2% to 3%, 3% to 4%, 4% to 5%, 5% to 10% and 10% to 50%.

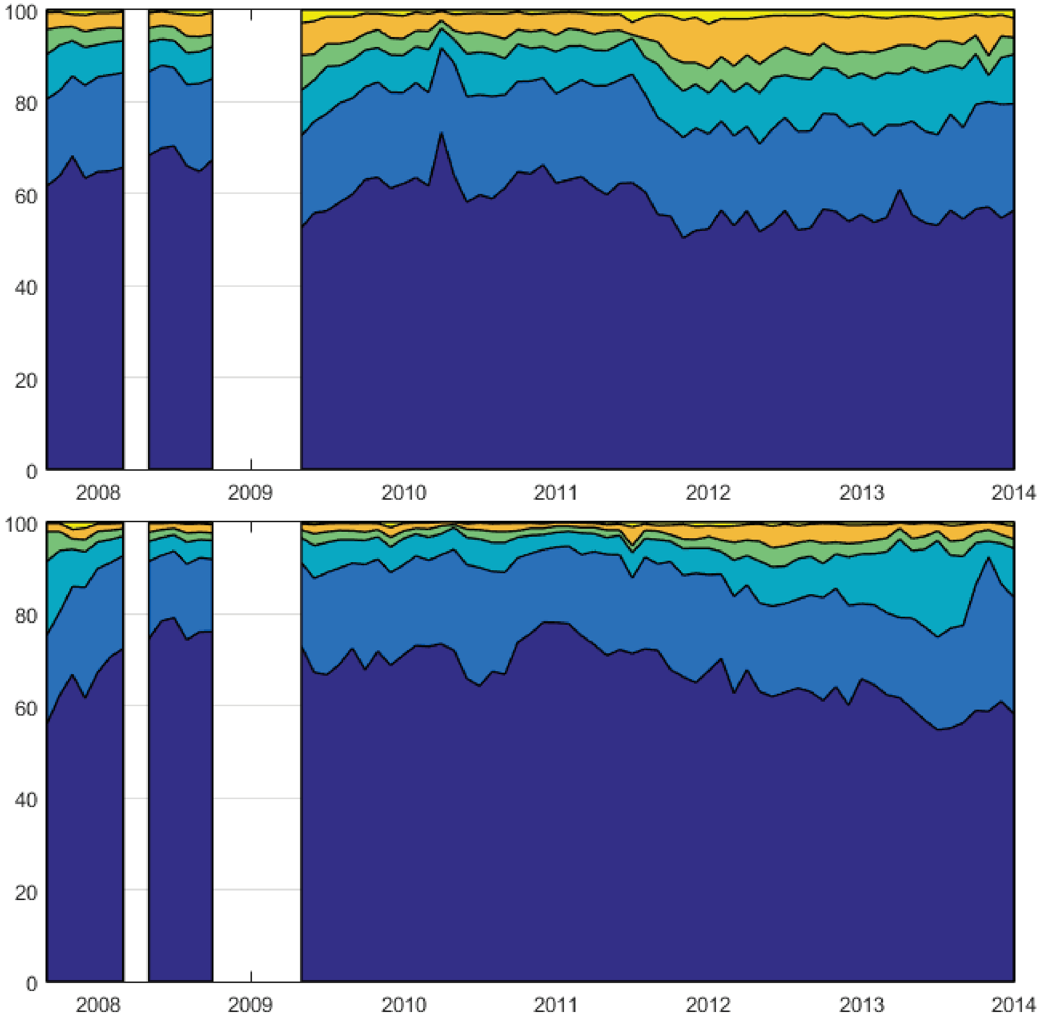


Fig. 4.4. The Market Cap Distribution of Extreme Move Events Across Time

The figure shows the monthly distribution of extreme move events disaggregated according to the market capitalization decile of the stock. The upper (lower) figure shows the distribution of the transitory (permanent) moves. Market capitalization decreases from yellow to blue.

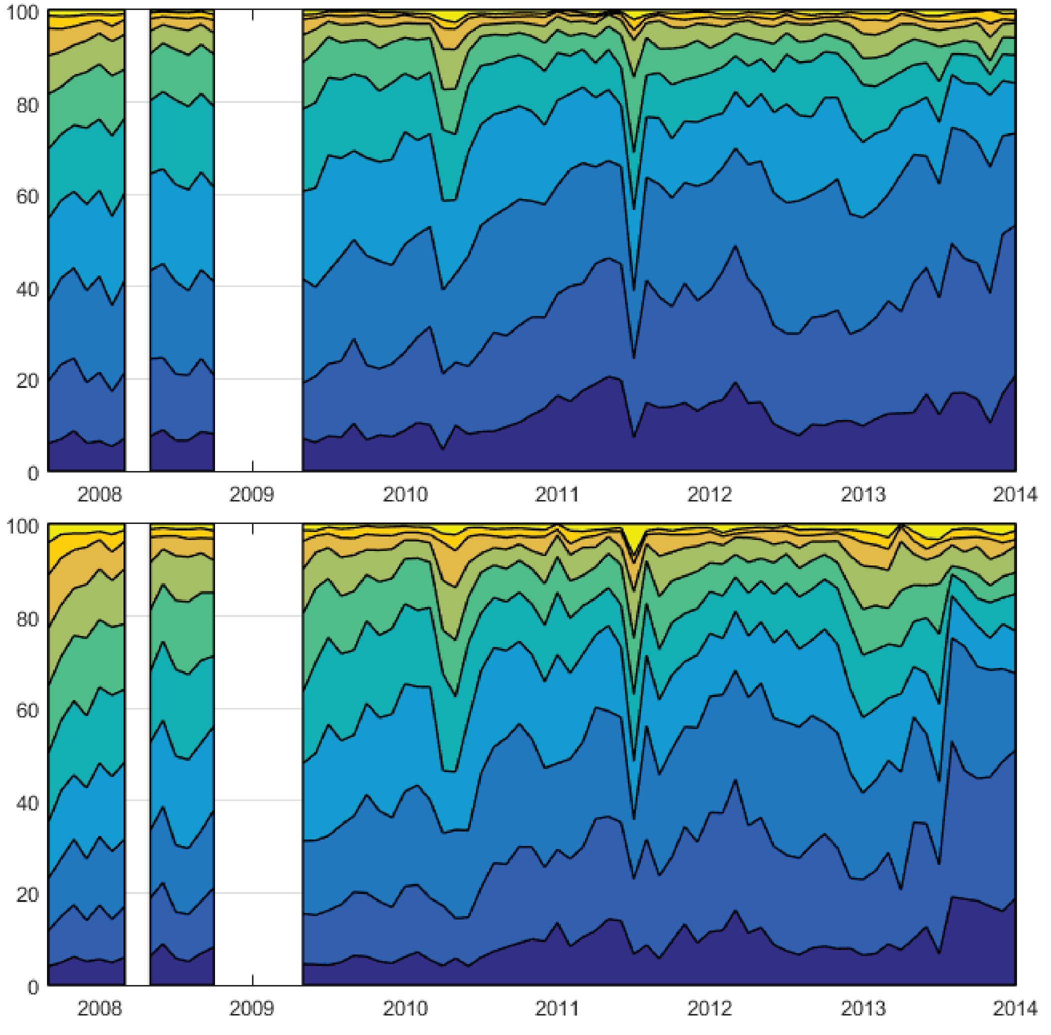


Fig. 4.5. The Trade Volume Distribution of Extreme Move Events Across Time

The figure shows the monthly distribution of extreme move events disaggregated according to the trade volume decile of the stock. The upper (lower) figure shows the distribution of the transitory (permanent) moves. Market capitalization decreases from yellow to blue.

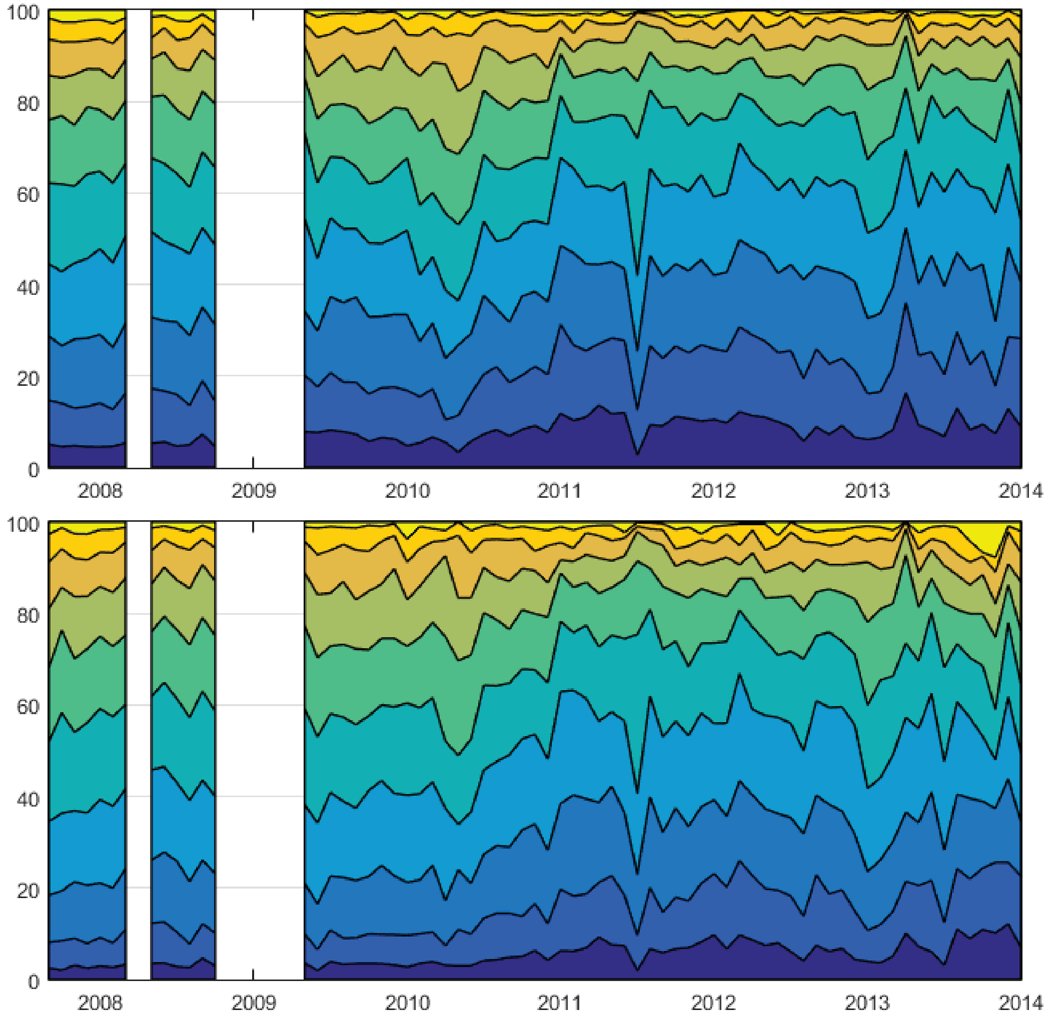


image of this asymmetry: The trade volume in the opposite direction outnumbers that in the move direction. As expected, the transitory moves which experience a higher degree of reversal also experience a far larger trade volume in the reversal direction.

HFT trade activity also tends to follow the aforementioned asymmetries throughout the event periods. However, Figure 4.7 shows that the share of the HFT activity in the total trade volume is larger in the direction opposite to the market move: During the move, HFT trades constitute a clearly larger share of trades against the move direction than those in line with it and, to a weaker degree, the HFT share against the reversal direction increases during the reversal. This shows that although the net HFT trade volume is in line with the move (reversal) during the move (reversal), during the move HFT firms bet in favor of the reversal of the move stronger than the non-HFT market participants.

4.3 Methodology

In this section, we present hypotheses for the existence of the momentum ignition strategy and the estimation methods to test their validity. These center around the predictive power of the premove HFT activity on the move and its reversal. We also look into the effects of the HFT activity on market quality. Our analysis focusses on the universe of transitory moves, because they are the primary suspects of the momentum ignition strategy in public accounts. Inclusion of other move types would introduce different dynamics, potentially disassociating our hypotheses from identification of the usage of this strategy. For example, the inclusion of permanent moves would bring in a clear price discovery role to the relation between the premove activity and the move size, weakening the possibility to interpret this relation as the instigation of a non-informative and thus temporary price change.

The official documents as well as the accounts of practitioners generally define the momentum ignition strategy in broad strokes, without providing much specificity. As we

Fig. 4.6. The Mean Trade Volume During the Events

The figure shows the mean HFT and non-HFT trade volumes at five event periods: the premove, and the first and second halves of the move and reversal. The two upper figures report the statistics for the permanent moves and the two lower figures for the transitory moves. The figures on the left report the volume of trades in the direction of the move, i.e. buyer/seller-initiated trades for up/down moves, and the figures on the right report the trade volume in the opposite direction.

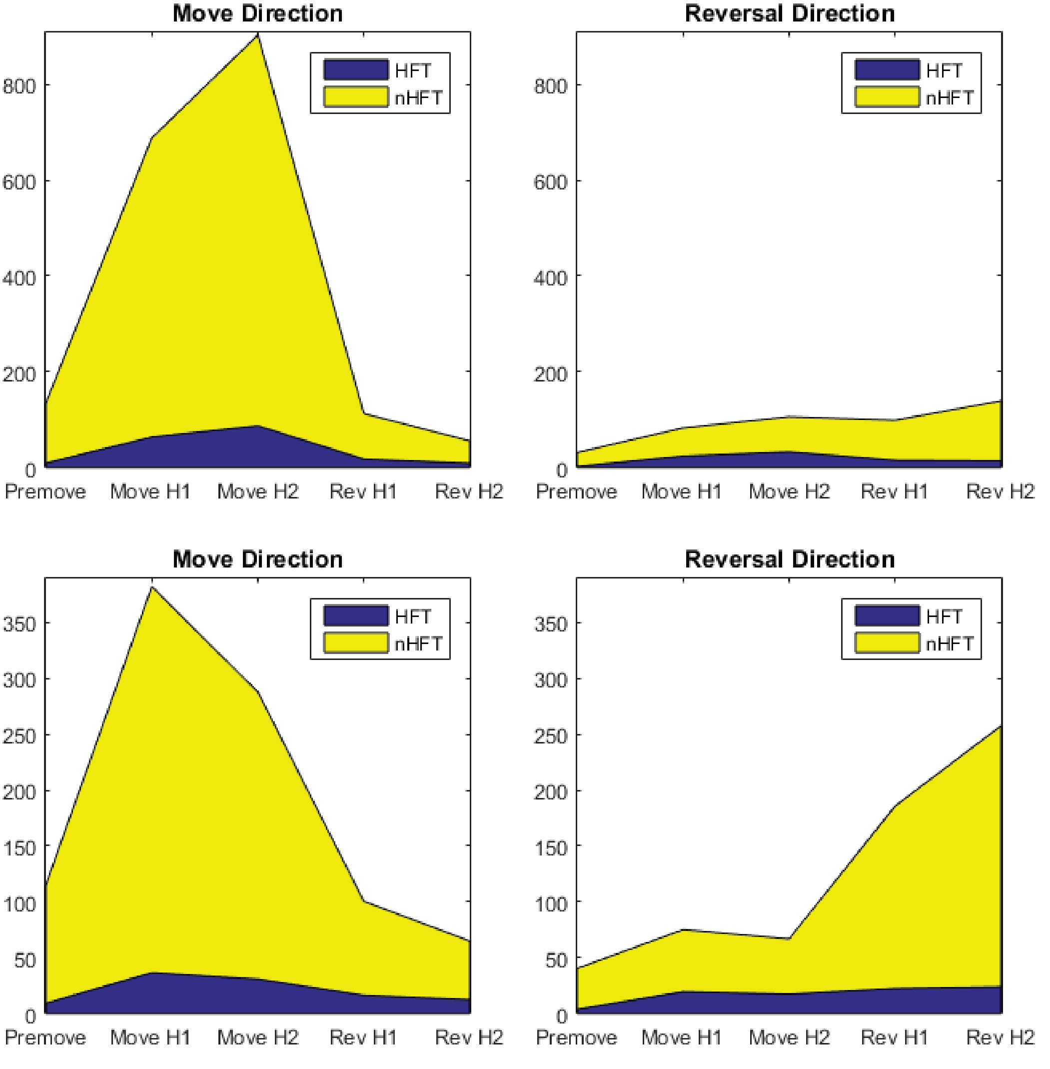
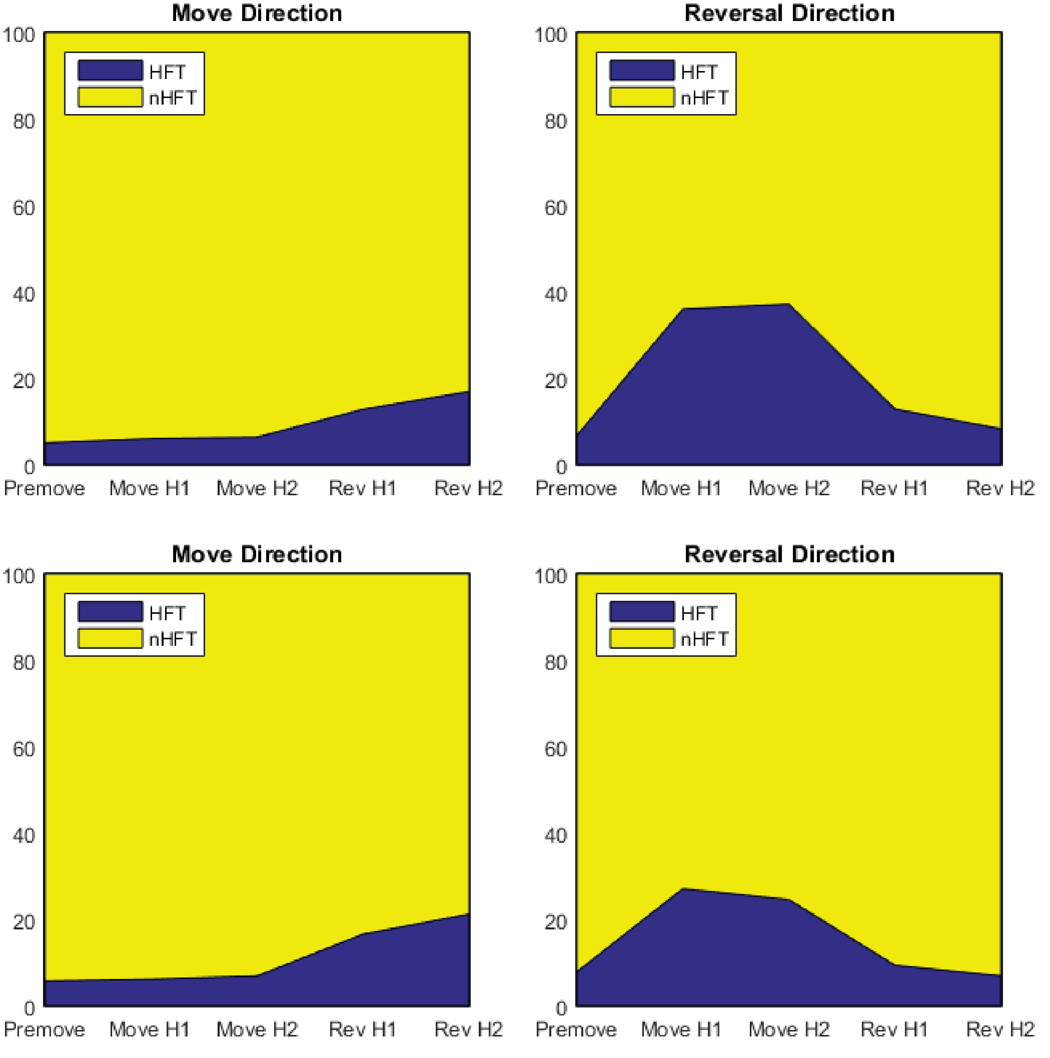


Fig. 4.7. The HFT Share in the Trade Volume During the Events

The figure shows the mean HFT and non-HFT trade volumes at five event periods: the premove, and the first and second halves of the move and reversal. The two upper figures report the statistics for the permanent moves and the two lower figures for the transitory moves. The figures on the left report the volume of trades in the direction of the move, i.e. buyer/seller-initiated trades for up/down moves, and the figures on the right report the trade volume in the opposite direction.



have mentioned in the introduction, a common feature of these definitions is the use of trades to generate a rapid and large price move. As this price move is not based on new information or actual demand and supply dynamics, popular accounts also add a reversal soon after the move. Based on these claims, we formulate two hypotheses required for the existence of the momentum ignition strategy:

1. The HFT trading in the premove has predictive power over the move. The HFT activity before the move event should be able to take a correct position before the move, potentially also leading the initial price movement.

2. The HFT trading in the premove has predictive power over the degree of reversal. Increasing HFT activity, especially in line with the move direction should be associated with a larger degree of reversal.

Testing for these hypotheses should also include controls for observationally equivalent alternatives to momentum ignition, which can generate these relationships. In particular we control for two alternative market scenarios: fleeting liquidity imbalances and the price movements driven by non-HFT parties. The HFT activity can precede the price moves in both types of events giving a false sign of causation. Due to their ability to swiftly generate trading signals from data on market conditions and asset fundamentals, HFT firms would be the first to react and to exploit liquidity imbalances and news announcements. Also in both cases, the initial price move would revert considerably, once the market attracts arbitrageurs bringing in liquidity or correcting the overreaction. We use the premove depth on both sides of the market to control for the effects of liquidity imbalances. The non-HFT trade activity is added to account for events initiated and driven by non-HFT parties.

We examine the effects of the HFT and non-HFT trading activity on three types of variables, denoted as the dependent variable $Change_{i,p,t}$ in the following estimation equations (4.2) and (4.3): The signed move size in percentages, the degree of reversal relative to the move size in percentages and the change in the three market quality variables in percentages. The subscripts denote that the dependent variable belongs to stock i observed in the p^{th} intraday period of month t. We divide the trading day to three intraday periods, namely the open (9:30h-11:00h), the midday (11:00h-14:30h) and the close (14:30h-15:00h), and assign each event to these periods depending on the starting time of the move.

As the main explanatory variables related to the testing of the hypotheses, we use the percentage difference of the HFT trade volume and the HFT share in the trade volume from their 22-day averages, in order to make these figures comparable across stocks. To account for the observationally equivalent alternatives, we use two sets of control variables. We add the change in the premove new quote additions on both sides of the market relative to their 22-day averages as a control for the first scenario concerning liquidity imbalances. To control for the second scenario of events driven by the non-HFT activity, we include the percentage difference of the non-HFT trade volume from its 22-day average.

We define the estimation equation for the effect of the HFT trade volume as

$$\begin{aligned} Change_{i,p,t} =& \alpha + \sum_{h}^{nHFT,HFT} \sum_{j=1}^{J} \sum_{s}^{Pro,Anti} \beta_{h,j,s} TrVol_{h,j,s,i,p,t} + \sum_{s}^{Pro,Anti} \gamma_s QVolPM_{s,i,p,t} \\ &+ \sum_{k=1}^{K} \gamma_k x_{i,p,t} + \varepsilon_{i,p,t}, \end{aligned} \tag{4.2}$$

where $TrVol_{h,j,s,i,p,t}$ is the percentage difference of the trade volume from its 22-day average for the same minute of the day, belonging to the market participant type h, $h = nHFT, HFT$, at the j^{th} subinterval of the event i on the market side s, $s = Pro, Anti$, HL_i is the 22-day average of the high-low range at the minute of the move start, $QVolPM_{s,i,p,t}$ is the change in the new quote additions in the premove of the event i on the market side s, $s = Pro, Anti$, relative to their 22-day average for the same minute of the day and $x_{i,p,t}$ are K explanatory variables depending on the choice of $Change_{i,p,t}$: For the market quality variables it includes the 22-day average level of the concerning variable and, for all variables, it includes the past high-low range to capture the effect of the recent volatility of the stock. The new quote additions on the bid (ask) side and buyer-initiated (seller-initiated) trades are denoted as Pro ($Anti$) for up and $Anti$ (Pro) for down moves.

The effect of the HFT share in the trade volume is evaluated by the similar regression equation of

$$Change_{i,p,t} = \alpha + \sum_{j=1}^{J} \sum_{s}^{Pro,Anti} \beta_{j,s} HFTshare_{j,s,i,p,t} + \sum_{s}^{Pro,Anti} \gamma_s QVolPM_{s,i,p,t} + \sum_{k=1}^{K} \gamma_k x_{i,p,t} + \varepsilon_{i,p,t}, \tag{4.3}$$

where $HFTshare_{j,s,i,p,t}$ is the HFT share in the trade volume, scaled by its 22-day average for the same minute of the day, at the j^{th} subinterval of the event i on the market side s, $s = Pro, Anti$.

We model the disturbances $\varepsilon_{i,p,t}$ as a three-way error component

$$\varepsilon_{i,p,t} = \mu_i + \zeta_p + \lambda_t + \epsilon_{i,p,t}, \tag{4.4}$$

where μ_i captures the stock-specific effect, ζ_p the effect specific to the intraday period, λ_t the month-specific effect and $\epsilon_{i,p,t}$ acts as ordinary disturbances. Direct estimation of this fixed effects model would require the addition of a huge amount of dummies for each stock, month and intraday period. We circumvent this problem by a Within transformation wiping out the variation specific to the stock, month and intraday period from the regressors. Baltagi (1985) and Wansbeek and Kapteyn (1989) provide the Within estimators of, respectively, one- and two-way error component models for unbalanced data and we use the generalization of their work to multi-way error components by Davis (2002).

The possibility of omitted variables leading to changes in both the HFT activity and the dependent variables causes endogeneity concerns. Such omitted variables may be stock-specific as well as spread-out across a series of closely related stocks such as cross-stock trading strategies taking positions on the components of the same industry or the same stock index. Hasbrouck and Saar (2013) propose the market-wide component of the HFT activity excluding the activity in the possibly related stocks as an instrument. This market-wide component would correlate with the HFT activity but not with the omitted variables specific to a stock or a group of stocks.

The market-wide components consist of the averages of the relevant variables, excluding the stock itself, the stocks in the same industry defined by its SIC code and the stocks in the S&P 500 or NASDAQ 100, if the concerning stock is listed in either of them. In Eq. (4.2), we instrument the HFT (non-HFT) variables in each period of the event with the market-wide average of the HFT (non-HFT) variables of the same event period for both trade directions. In Eq. (4.3), the instruments for the HFT share variables in each period of the event similarly consist of the market-wide averages of the HFT share variables in the same period.

As a final consideration, we examine the possible changes in the relationship between the HFT activity and the dependent variables over time. Our data set exhibits at least the vast transformation extending from the depths of the financial crisis to the ongoing recovery and HFT firms have to evolve according to these changing market conditions. As our hypotheses are mainly concerned with the predictive power of the premove HFT activity, we focus on the changes of its effect. The poolability test of Han and Park (1989) provides a method to look for structural breaks in the estimated coefficients. This amounts to an F-test for the difference of the coefficient estimates between two subsamples of the data.

We can reject the hypothesis of a constant effect of the premove HFT activity for a number of cases. These structural breaks cluster around the second half of 2011. Although the most likely break points are not identical, we report the results of a second set of regressions with a reduced data set, starting with October 2011. All regressions admitting a break also cannot reject the hypothesis of a break in October 2011, although another time in the second half of 2011 may be more probable.

The regression results are provided under different statistical significance levels and we interpret them using the 5% significance level. We do not differentiate between the two halves of the move and reversal periods, as we did in Section 4.2.3, in order to ease the presentation of the results. Thus the number of event periods J can be at most three.

4.4 Analysis of Results

In this section, we investigate the effects of the HFT trading activity on transitory extreme price moves, both in terms of the existence of the momentum ignition strategy and in terms of market quality. We examine the trading activity of the HFT firms in connection with their explanatory power on the price move and on the magnitude of the reversal following the move. And lastly, we use three market quality measures, namely volatility measured by the high-low range, quoted spreads and the execution-to-cancellation ratios, and estimate how they are affected by the HFT activity before and during the move events.

4.4.1 Predicting the Move

We measure the predictive power of HFT trading on the move direction and size by estimating Eq. (4.2) and Eq. (4.3) using the trade activity data for only the premove period, i.e. $J = 1$. Table 4.1 reports the coefficient estimates.

The HFT and non-HFT trade activity in the premove display a comparable predictive power on the move size. For both of them, increases in the selling (buying) activity rise the chance of an upcoming down (up) move. Although the effects are statistically significant, they remain economically small. A 1% increase in the HFT (non-HFT) trade volume in line with the move relative to its past average causes a 0.04 (0.03 to 0.04) basis points rise in the move size. The magnitudes are similar for the change in the HFT and non-HFT activity against the move direction. The similar sizes of these effects are particularly striking considering that a 1% increase in the HFT trade volume compared

Table 4.1. The Move Size

The table reports how changes in the trade volume ('ΔHFT') and the HFT share in the trade volume ('$\Delta HFT\%$') relative to their 22-day average affect the move size. The estimation equations are given in Eq. (4.2) and Eq. (4.3). Besides trade volumes, the regression also includes the change in the new quote additions in the premove relative to their 22-day average ('QVol') and the 22-day average of the high-low range at the minute of the move start ('HL'). The effect of the activity in line with the move direction is reported under 'Pro' and the effect of the activity against the move direction under 'Anti'. The move size is in basis points, while the changes in trade volumes and the shares in the trade volume are in decimal points to ease the exposition of the coefficient estimates. The estimation is done using both the whole data set spanning from July 2007 to December 2013 and the subinterval of October 2011 to December 2013. The superscript *** marks statistical significance at 1%, ** at 5% and * at 10% level using the Driscoll and Kraay (1998) extension of the Newey-West estimator.

			Jul 2007 - Dec 2013		Oct 2011 - Dec 2013	
			ΔHFT	$\Delta HFT\%$	ΔHFT	$\Delta HFT\%$
PM	nHFT	Anti	-4.8^{***}	-4.3^{***}	-3.3^{***}	-2.9^{***}
		Pro	3.9^{***}	3.1^{***}	3.2^{***}	2.3^{***}
	HFT	Anti	-4.8^{***}	-45.8^{***}	-5.4^{***}	-36.2^{***}
		Pro	4.1^{***}	49.2^{***}	4.3^{**}	35.9^{**}
	QVol	Anti	-5.8^{*}	-5.3^{**}	-3.9^{*}	-4.7^{*}
		Pro	4.3^{**}	5.1^{**}	4.1^{**}	4.8^{**}
HL			6.9^{***}	7.3^{***}	4.3^{***}	5.0^{***}
N.obs.			853,838	859,267	172,438	173,233

to its past average corresponds to a far smaller magnitude of the trade volume compared to the same increase in the non-HFT trade volume. HFT trades amount to only 7.8% of the premove trade volume.[4]

The regressions estimating the effect of the HFT share in the trade volume provide further evidence for the predictive power of the HFT activity. Increases of the HFT share in the buying (selling) volume tends to move prices up (down), checking the signs of the coefficient estimates. The effect is statistically significant and economically larger than the effect of changes in the trade volume, albeit remaining relatively small. A 1% increase of the HFT share in trade volume in line with (against) the move relative to its past average causes a 0.49 (0.46) basis points increase (decrease) in the move size.

Figure 4.8 displays the plots of the cumulative sums of errors between forecasts and realizations (CUSUM) and squared errors (CUSUMSQ) for regressions. The sum of forecast errors for the regressions with the change in the HFT trade volume as well as the

[4]To give a context for this figure, the annual averages of the share of HFT trades in the premove trade volume increase monotonically from 4.9% in 2007 to 10.3% in 2013.

change in the HFT share of the trade volume stay within the 5% significance bounds and diverge from the zero line far less than the other regressions. However, as with the other plots, the second half of 2011 exhibits a trend for divergence. The sum of squared errors crosses the 5% significance line multiple times with all other regressions mainly during the three peaks of the crisis, signalling a larger standard deviation at these episodes: the aftermath of the bankruptcy of Lehman Brothers in 2008-2009, the European debt crisis in mid-2010 and the S&P downgrade of U.S. debt in August 2011. Although these plots provide some evidence for a structural break, we cannot reject the hypothesis of a constant effect of the premove HFT activity across the sample using the methodology of Han and Park (1989).

The coefficient estimates for the subsample of October 2011 - December 2013 qualitatively resemble the aforementioned results for the whole sample, except for a weakening of the effects for most cases possibly due to the smaller move sizes caused by the reduced market volatility. Increases in the premove HFT and non-HFT activity as well as the HFT share in the premove trade volume against (in line with) the move direction lead to smaller (larger) price moves.

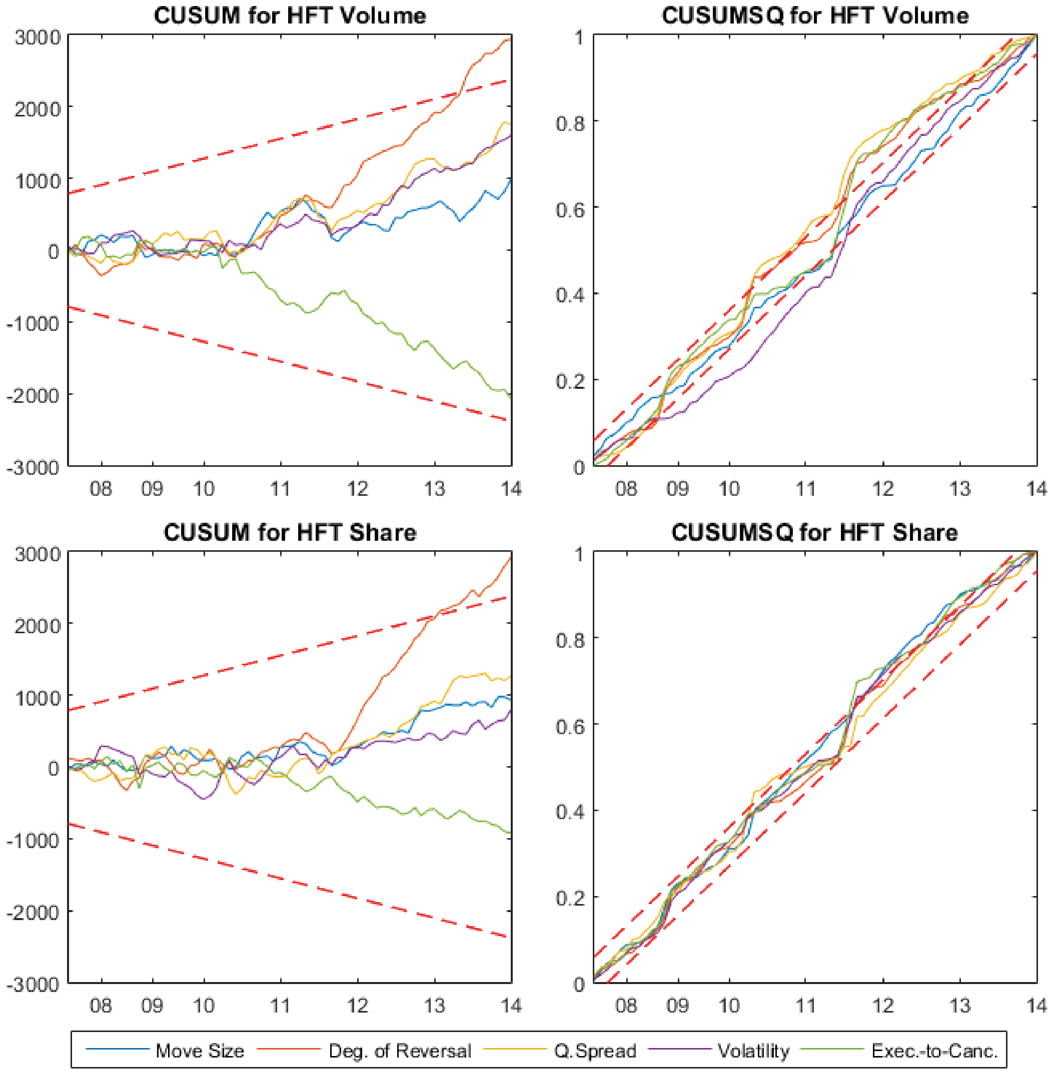
This regression analysis confirms our first hypothesis postulating a positive relationship between net HFT trade activity before the event and the magnitude of the transitory extreme move, even after controlling for the effect of the non-HFT trade volume and liquidity imbalances. The effect size of the changes in the premove HFT activity even rivals that of the non-HFT activity, which accounts for a far larger magnitude in the trade volume. We cannot reject the hypothesis of a constant effect across the whole sample and find qualitative similar results for the subsample of October 2011 - December 2013.

4.4.2 Predicting the Degree of Reversal

In this section, we examine the second hypothesis proposing a relationship between the HFT trade activity before the extreme move events and the degree of reversal following

Fig. 4.8. CUSUM and CUSUMSQ Plots for the Effect of the Premove HFT Activity

The figure displays the cumulative sums of forecast errors (CUSUM) and squared errors (CUSUMSQ) for the regression equations (4.2) and Eq. (4.3) using the July 2007 - December 2013 data. Each regression equation is estimated for five dependent variables: The move size ('Move Size'), the degree of reversal ('Deg. of Reversal'), the quoted spread ('Q.Spread'), market volatility ('Volatility') and the execution-to-cancellation ratio ('Exec.-to-Canc.'). In order to present five different regressions with different amounts of total number of observations and different amounts of monthly data, we scaled the cumulative sums to a data set of 70 months, each with 100,000 observations.



the moves. We estimate Eq. (4.2) and Eq. (4.3) using the trade activity data for both the premove and the move, i.e. $J = 2$. Table 4.2 reports the estimation results.

Table 4.2. The Degree of Reversal

The table reports how changes in the trade volume ('ΔHFT') and the HFT share in the trade volume ('$\Delta HFT\%$') relative to their 22-day average affect the degree of reversal. The estimation equations are given in Eq. (4.2) and Eq. (4.3). Besides trade volumes, the regression also includes the change in the new quote additions in the premove relative to their 22-day average ('QVol') and the 22-day average of the high-low range at the minute of the move start ('HL'). The effect of the activity in line with the move direction is reported under 'Pro' and the effect of the activity against the move direction under 'Anti'. The degree of reversal is in basis points, while the changes in trade volumes and the shares in the trade volume are in decimal points to ease the exposition of the coefficient estimates. The estimation is done using both the whole data set spanning from July 2007 to December 2013 and the subinterval of October 2011 to December 2013. The superscript *** marks statistical significance at 1%, ** at 5% and * at 10% level using the Driscoll and Kraay (1998) extension of the Newey-West estimator.

			Jul 2007 - Dec 2013		Oct 2011 - Dec 2013	
			ΔHFT	$\Delta HFT\%$	ΔHFT	$\Delta HFT\%$
PM	nHFT	Anti	-82.9^{**}	-89.6^{*}	-34.1^{**}	-41.0^{**}
		Pro	91.2^{**}	94.7^{**}	56.2^{**}	53.9^{**}
	HFT	Anti	-102.8^{*}	-1034.3	43.5^{**}	634.3^{*}
		Pro	95.2^{***}	1098.2	45.9^{***}	464.1^{**}
	QVol	Anti	-18.4^{*}	-19.2^{*}	-9.1^{*}	-10.2^{**}
		Pro	23.1^{**}	19.6^{**}	12.3^{***}	10.4^{***}
Move	nHFT	Anti	209.3^{***}	193.4^{***}	239.1^{**}	231.9^{***}
		Pro	-302.7^{***}	-287.2^{***}	-123.9^{**}	-198.2^{**}
	HFT	Anti	199.0^{**}	842.2^{**}	158.9^{***}	582.0^{***}
		Pro	-234.9^{**}	-908.2^{*}	-206.4^{**}	-683.5^{**}
HL			18.9^{***}	20.3^{***}	18.1^{***}	17.4^{***}
N.obs.			631,309	748,533	127,942	150,128

The most statistically and economically significant effects belong to the variables of the move period, most probably due to the temporal proximity of this period to the upcoming reversal. All non-HFT and three out of four HFT coefficients are statistically significant for this period. A 1% increase in the non-HFT trades against (in line with) the move direction reduce (increase) the degree of reversal by 1.93 to 2.09 (2.87 to 3.03) basis points. For the HFT activity, the effect of changes in the HFT selling and buying trade volume relative to their past average are statistically significant, although only the change of the HFT share against the move direction is similarly significant. A 1% increase in the HFT trades against (in line with) the move direction reduce (increase) the degree of reversal by 1.99 (2.35) basis points. As in the effect on move sizes, the same increase in the HFT share of the trade volume against the move direction causes a larger increase of 8.42 basis points.

As for the premove period, only the non-HFT activity contains coefficients with consistent signs and a high degree of statistical significance. Three out of four non-HFT coefficients are statistically significant with effect sizes about one-third to one-fifth of the move variables. The non-HFT trade activity against (in line with) the move direction increases (reduces) the degree of reversal. We estimate the premove HFT activity to be statistically insignificant in most cases.

The CUSUM and CUSUMSQ analyses presented in Figure 4.8 provides the strongest evidence for a structural break in the case of the regressions on the degree of reversal. For both the HFT trade volume and the HFT share regressions, the sum of forecast errors crosses the 5% significance line in early 2013. The trend of divergence from the zero line starts arguably as early as in the last months of 2010, but the second half of 2011 presents a more robust starting point, especially for the HFT share regressions. Indeed, the hypothesis of constant coefficients is rejected for both regressions given the alternative of a structural break in October 2011.

The regressions for the subsample of October 2011 - December 2013, presented in Table 4.2, estimate a stronger positive relationship between HFT trade activity during the premove and the degree of reversal. Whereas only one out of four coefficients is statistically significant for the regressions using the whole sample period, the subsample hosts three significant coefficients for the premove HFT activity. The effect of changes in non-HFT trade activity differ by trade direction, where a 1% increase in the trades against (in line with) the move direction reduce (increase) the degree of reversal by 0.34 (0.41) basis points. By contrast, increases in the HFT activity on both sides lead to stronger reversals, although only one of the two coefficients for trades against the move direction is statistically significant: A 1% change in the HFT trade volume against (in line with) the move direction increases the degree of reversal by 0.44 (0.46) basis points.

The move period continues to be the most economically and statistically significant factor determining the degree of reversal during the subsample. A 1% increase in the non-HFT trades against (in line with) the move direction reduces (increases) the degree of reversal by 2.32 to 2.39 (1.24 to 1.98) basis points and the same figure for HFT trades is 1.59 (2.06).

These regressions of the degree of reversal on the HFT activity, among other explanatory variables, provide evidence for the validity of the second hypothesis concerning the momentum ignition strategy. Especially for the subsample from October 2011 to December 2013, we find a significant positive relationship between the premove HFT activity and the magnitude of the reversal following the move. It is telling that we observe a strong statistically significant relationship only in the aftermath of the peaks of the financial crisis. This coincides with the reported drops in HFT profitability due to a number of factors such as the reduction of especially market-making profits due to lower market volatility and the overcrowding of the HFT field.

4.4.3 Effects on Market Quality

We measure the effects of these move and reversal events on market quality with three variables: quoted spreads, volatility measured by the high-low range and the execution-to-cancellation ratios. We present summary statistics for the change in these market quality variables. Afterwards we discern on the relationship between the HFT trade activity and the market quality change after extreme moves by estimating Eq. (4.2) and Eq. (4.3) using the market activity data for all of the three event subperiods, i.e. $J = 3$.

The transitory extreme move events lead to dramatic drops in market quality. Comparing their level at the end of the minute after the end of the reversal to their average level in the five minutes before the premove, quoted spreads have increased by 233.4%, volatility by 664.5% and the execution-to-cancellation ratio has dropped by 33.5%. After five minutes, the volatility increase and the drop in the execution-to-cancellation ratio partially recovers to 343.3% above and 7.6% below the level before the premove, respectively. By contrast, quoted spreads further widen to 241.3% of their initial level.

Tables 4.3, 4.4 and 4.5 present the estimates of Eq. (4.2) and Eq. (4.3) for quoted spreads, volatility and execution-to-cancellation ratios, respectively. The trade activity during the move and reversal periods creates a distinct pattern in its relationship with the market quality change: Almost always, increasing trade volume against (in line with) the ongoing price direction, i.e. selling (buying) during the move and buying (selling) during the reversal, alleviates (exacerbates) the market quality deterioration. Across move types, out of 36 coefficients of changes in HFT and non-HFT trade volumes during the move and reversal, 27 are statistically significant and out of those only 3 do not follow this pattern. Their effect is also sizable, probably owing to the magnitude of the market quality deteriorations: On average a 1% increase in the non-HFT trade volume against (in line with) the ongoing move direction improves (deteriorates) the market quality change by 1.20%, 2.60% and 0.10% (0.84%, 2.65% and 0.15%) for quoted spreads, market volatility and execution-to-cancellation ratios, respectively. The same figures for the effect of HFT trade volume are respectively 0.54%, 1.47% and 0.12% (0.69%, 1.95% and 0.13%).

The presence of a larger HFT share in the trade volume improves the market quality change, regardless of on which side of the market it happens. Nearly all coefficient estimates related to changes in the HFT share are statistically significant and all of the significant ones imply an improvement on market quality with increasing values. On average a 1% increase of the HFT share in the trade volume improves the market quality change by 2.56%, 5.58% and 0.44% for quoted spreads, market volatility and execution-to-cancellation ratios, respectively. This may be related to the higher share of activity against the direction of the price move for the HFT firms compared to the non-HFT market participants, depicted in Figure 4.7, which can be interpreted as a stronger tendency to follow a mean-reversion strategy leading to softer price movements.

The results for the premove activity are statistically weaker and less significant in terms

Table 4.3. The Change in the Quoted Spreads

The table reports how percentage changes in the trade volume ('ΔHFT') and the HFT share in the trade volume ('$\Delta HFT\%$') relative to their 22-day average affect the percentage change of the quoted spreads before and after the extreme move event. The estimation equations are given in Eq. (4.2) and Eq. (4.3). Besides trade volumes, the regression also includes the change in the new quote additions in the premove relative to their 22-day average ('QVol'), the 22-day averages of the high-low range ('HL') and the quoted spread ('Level') at the minute of the move start. The effect of the activity in line with the move direction is reported under 'Pro' and the effect of the activity against the move direction under 'Anti'. The estimation is done using both the whole data set spanning from July 2007 to December 2013 and the subinterval of October 2011 to December 2013. The superscript *** marks statistical significance at 1%, ** at 5% and * at 10% level using the Driscoll and Kraay (1998) extension of the Newey-West estimator.

			Jul 2007 - Dec 2013		Oct 2011 - Dec 2013	
			ΔHFT	$\Delta HFT\%$	ΔHFT	$\Delta HFT\%$
PM	nHFT	Anti	-0.47^{*}	0.26	-0.25	0.61^{*}
		Pro	0.28^{**}	0.01^{*}	0.14^{**}	0.10^{*}
	HFT	Anti	0.41^{*}	-0.39	0.38^{**}	-0.06
		Pro	0.43^{**}	0.38^{*}	0.02^{**}	0.48^{**}
	QVol	Anti	-0.39^{**}	-0.01^{*}	0.02^{*}	-0.08^{*}
		Pro	-0.44^{**}	-0.26^{*}	-0.41^{**}	-0.07^{**}
Move	nHFT	Anti	-0.97^{**}	-0.48^{**}	-0.53^{*}	-0.64^{**}
		Pro	0.67^{***}	0.03^{***}	1.04^{**}	1.26^{**}
	HFT	Anti	-0.04^{***}	-4.34^{**}	-0.22^{*}	-3.29^{**}
		Pro	0.10^{**}	-0.59^{**}	0.54^{*}	-2.98^{**}
Rev.	nHFT	Anti	1.14^{***}	0.73^{***}	1.19^{**}	0.09^{**}
		Pro	-1.22^{**}	-1.15^{**}	-0.84^{**}	-0.41^{**}
	HFT	Anti	0.60^{**}	-3.03^{***}	0.82^{**}	-4.15^{**}
		Pro	-0.41^{***}	-0.16^{***}	-0.88^{**}	-3.80^{**}
Level			1.56^{***}	1.17^{***}	1.40^{**}	1.37^{**}
HL			2.28^{**}	2.56^{**}	2.09^{**}	2.09^{**}
N.obs.			606,100	747,045	123,937	148,743

of the effect size. Each of the regression sets for three market quality variables contains in total 8 coefficients for the HFT and non-HFT premove activity and, in sum, only four coefficients of the non-HFT premove activity and three coefficients of the HFT premove activity are statistically significant. The effect sizes across market quality variables are also one-fourth to one-half of those during the move and the reversal.

The CUSUM analysis depicted in Figure 4.8 presents some evidence for a structural break in the last months of 2011. Although none of the market quality variables crosses the 5% significance line, all tend to diverge from the zero line especially in the last two years of the sample. Given the alternative hypothesis of a structural break on October

Table 4.4. The Change in the Market Volatility

The table reports how percentage changes in the trade volume ('ΔHFT') and the HFT share in the trade volume ('$\Delta HFT\%$') relative to their 22-day average affect the percentage change of the market volatility, measured by the high-low range defined in Eq. (4.1), before and after the extreme move event. The estimation equations are given in Eq. (4.2) and Eq. (4.3). Besides trade volumes, the regression also includes the change in the new quote additions in the premove relative to their 22-day average ('QVol') and the 22-day average of the high-low range at the minute of the move start ('Level'). The effect of the activity in line with the move direction is reported under 'Pro' and the effect of the activity against the move direction under 'Anti'. The estimation is done using both the whole data set spanning from July 2007 to December 2013 and the subinterval of October 2011 to December 2013. The superscript *** marks statistical significance at 1%, ** at 5% and * at 10% level using the Driscoll and Kraay (1998) extension of the Newey-West estimator.

			Jul 2007 - Dec 2013		Oct 2011 - Dec 2013	
			ΔHFT	$\Delta HFT\%$	ΔHFT	$\Delta HFT\%$
PM	nHFT	Anti	-0.68^{*}	-1.95	-0.39^{**}	-3.00
		Pro	2.05^{**}	1.02^{*}	3.26	1.61^{**}
	HFT	Anti	-0.91^{*}	-1.60	0.80^{*}	0.56^{**}
		Pro	0.49^{*}	0.57^{**}	0.29^{**}	0.56^{**}
	QVol	Anti	-0.68^{**}	-1.76	-0.57^{**}	-0.18^{**}
		Pro	-0.70^{**}	-1.94^{*}	-0.47	-0.40^{*}
Move	nHFT	Anti	-1.68^{***}	-4.47^{**}	-2.09^{**}	-2.07^{**}
		Pro	4.46^{*}	0.47^{**}	6.01^{**}	3.10^{**}
	HFT	Anti	-0.58^{*}	-2.76^{***}	-2.31^{**}	-4.68^{**}
		Pro	4.18^{*}	-9.43^{***}	-0.72^{**}	-0.03^{**}
Rev.	nHFT	Anti	2.96^{**}	5.20^{**}	0.96^{**}	5.30^{**}
		Pro	-2.38^{***}	-3.92^{**}	-0.43^{**}	-2.45^{**}
	HFT	Anti	0.75^{*}	-10.47^{***}	3.79^{**}	-0.34^{**}
		Pro	-2.21^{**}	-0.75^{***}	-4.32^{**}	-15.85^{**}
Level			6.01^{***}	5.39^{***}	5.36^{**}	5.91^{**}
N.obs.			419,947	529,526	85,238	105,838

2011, The poolability test of Han and Park (1989) rejects the hypothesis of a constant effect of the premove HFT activity on quoted spreads and market volatility, but not for the execution-to-cancellation ratios.

We find a relatively stronger and generally adverse effect of the premove HFT activity on the market quality change for the subsample of October 2011 - December 2013. 7 out of 12 coefficients of the premove HFT activity on both sides of the order book are statistically significant and all but one indicate a negative effect on the market quality change. On average, a 1% increase in the HFT trade volume (in the HFT share in the trade volume) during the premove deteriorates the market quality change by 0.20%, 0.02% and 0.10% (0.48%, 0.56% and 0.06%) for quoted spreads, market volatility and

Table 4.5. The Change in the Execution-to-Cancellation Ratios

The table reports how percentage changes in the trade volume ('ΔHFT') and the HFT share in the trade volume ('$\Delta HFT\%$') relative to their 22-day average affect the percentage change of the execution-to-cancellation ratios before and after the extreme move event. The estimation equations are given in Eq. (4.2) and Eq. (4.3). Besides trade volumes, the regression also includes the change in the new quote additions in the premove relative to their 22-day average ('QVol'), the 22-day averages of the high-low range ('HL') and the execution-to-cancellation ratio ('Level') at the minute of the move start. The effect of the activity in line with the move direction is reported under 'Pro' and the effect of the activity against the move direction under 'Anti'. The estimation is done using both the whole data set spanning from July 2007 to December 2013 and the subinterval of October 2011 to December 2013. The superscript *** marks statistical significance at 1%, ** at 5% and * at 10% level using the Driscoll and Kraay (1998) extension of the Newey-West estimator.

			Jul 2007 - Dec 2013		Oct 2011 - Dec 2013	
			ΔHFT	$\Delta HFT\%$	ΔHFT	$\Delta HFT\%$
PM	nHFT	Anti	0.14	0.11^{*}	0.05^{*}	0.14
		Pro	-0.06^{**}	-0.02^{**}	-0.13^{**}	-0.11^{*}
	HFT	Anti	0.08^{**}	0.03^{*}	0.06^{**}	0.09^{*}
		Pro	-0.13^{*}	-0.15^{*}	-0.10^{**}	-0.06^{**}
	QVol	Anti	-0.08^{***}	-0.10^{*}	-0.04^{**}	0.08
		Pro	0.07^{**}	0.04^{*}	0.06^{**}	0.01^{**}
Move	nHFT	Anti	0.08^{*}	0.09^{*}	0.13^{**}	0.02^{**}
		Pro	0.25^{***}	-0.20^{*}	-0.23^{**}	-0.30^{**}
	HFT	Anti	0.08^{*}	0.71^{**}	0.00^{**}	0.74^{**}
		Pro	-0.07^{*}	0.02^{**}	-0.18^{**}	-0.09
Rev.	nHFT	Anti	-0.04^{***}	-0.28^{**}	-0.12^{**}	-0.14^{**}
		Pro	0.00^{**}	0.18^{**}	0.24^{**}	0.17^{**}
	HFT	Anti	-0.21^{***}	1.03^{**}	-0.22^{**}	0.19^{**}
		Pro	0.21^{***}	0.13^{***}	0.12^{**}	0.11^{**}
Level			0.26^{***}	0.64^{**}	0.55^{**}	0.47^{**}
HL			0.64^{**}	0.84^{**}	0.39^{**}	0.63^{**}
N.obs.			489,546	674,865	98,893	132,397

execution-to-cancellation ratios, respectively.

We reach two seemingly contradictory results for the effect of the HFT activity on the market quality change caused by transitory extreme move events. While higher shares of HFT trades in the total trade volume during the move and the reversal have a strong tendency to soften the drop in market quality, we find some evidence for the premove HFT activity to have a negative effect, especially in the later years of our sample. The users of the HFT technology considerably vary in terms of their market strategies, the usual contrast being between passive market makers and aggressive directional traders (Baron et al., 2012; Brogaard et al., 2014; Menkveld and Zoican, 2014). We would expect the

HFT activity within the legal bounds and related to market making activity to improve market quality and contribute to price discovery, given the evidence in the literature (Brogaard et al., 2014; Carrion, 2013; Hasbrouck and Saar, 2013). Taking the confirmation of the hypotheses relating the HFT activity before transitory extreme move events to the momentum ignition strategy into consideration, the higher premove HFT activity may be capturing such incidents of market fraud, whereas the HFT activity during the event is dominated by the market makers, which comprise the bulk of the HFT activity in financial markets (Menkveld, 2013; Hagstromer and Norden, 2013; Brogaard et al., 2014). Thus, we capture the effects of two different types of HFT strategies on market quality.

4.5 Conclusion

We investigate an extensive NASDAQ data set of 8,000 stocks spanning from July 2007 to December 2013 for extreme move events. About half of 1,675,100 such events are followed by a reversal more than two-thirds of the move size, which we categorize as the transitory moves. The remaining half is about equally divided between those with less than one-third reversal and those reverting from one-third to two-thirds. Although the number of events are in a trend of reduction since the financial crisis years, the ratio between these three degrees of reversal remains quite stable.

We formalize the use of the momentum ignition strategy with two hypotheses relating the HFT activity before transitory extreme move events to the magnitudes of the move and the reversal. We use two main types of regressions to evaluate the effect of the HFT activity on the move, the degree of reversal and the market quality changes caused by an event. The first regression uses the percentage difference of the volumes of HFT and non-HFT trades from their 22-day average and the second regression computes the same percentage difference for the HFT share in the total trade volume. These transformations as well as the implementation of a fixed effects model controlling for factors specific for each stock, month and intraday period make the HFT activity variables more comparable across stocks. We also control for alternative market conditions which can generate rapid price moves with significant reversals, namely price movements initiated and driven by non-HFT trade activity such as market overreaction and temporary liquidity imbalances attracting both HFT and non-HFT parties.

The regressions of the move size on market variables find evidence for the predictive power of the premove HFT activity on the upcoming extreme move, confirming our first hypothesis. The trade activity of HFT firms nearly match the predictive power of non-HFT market participants. Non-HFT trades constitute the vast majority of the trade volume, i.e., 92.2% of the premove trade volume, and therefore are able to move the midquote far easily. In spite of this asymmetry, changes in the HFT trade volume relative to its past level have only a 13.3% smaller effect on the move size compared to those for the non-HFT market participants.

Particularly the October 2011 - December 2013 subsample presents strong evidence for our second hypothesis relating the premove HFT activity to the magnitude of the reversal following the move. Although the evidence is stronger for HFT trades in line with the upcoming move, increases in the HFT activity on both sides lead to stronger reversals: A 1% change in the HFT trade volume increases the degree of reversal by 0.45 basis points. This timing coincides with the public reports about the profitability concerns of HFT firms due to lower market volatility and tougher competition in a more crowded market.

We also examine the effect of transitory extreme move events on market quality and how it is influenced by the HFT activity. Across our three market quality measures, we find that extreme price moves affect market quality negatively. Comparing their magnitudes one minute after the end of the event to their levels before the premove, quoted spreads and volatility have risen on average by 233.4% and 664.5% respectively and the execution-to-cancellation ratios have dropped by 33.5%. During the move and the reversal, higher HFT share in the trade activity on both sides of the order book improves particularly the change in quoted spreads and volatility. The premove HFT activity, by contrast, gives some evidence for a negative effect on market quality, particularly during the subsample spanning from October 2011 to December 2013.

Bibliography

Admati, A. R. and P. Pfleiderer (1988). A theory of intraday patterns: Volume and price variability. *Review of Financial Studies 1*(1), 3–40.

Ahn, H., H. Bae, and K. Chan (2001). Limit orders, depth and volatility: Evidence from the stock exchange of hong kong. *Journal of Finance 56*(2), 769–790.

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

Andersen, T., T. Bollerslev, and A. Das (2001). Variance-ratio statistics and high-frequency data: Testing for changes in intraday volatility patterns. *Journal of Finance 56*(1), 305–327.

Andersen, T., T. Bollerslev, F. X. Diebold, and C. Vega (2003). Micro effects of macro announcements: Real-time price discovery in foreign exchange. *The American Economic Review 93*(1), 38–62.

Baillie, R., G. Booth, Y. Tse, and T. Zabotina (2002). Price discovery and common factor models. *Journal of Financial Markets 5*(3), 309–321.

Baltagi, B. (1985). Pooling cross-sections with unequal time series lengths. *Economics Letters 18*, 133–136.

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

Barclay, M. J. and T. Hendershott (2004). Liquidity externalities and adverse selection: Evidence from trading after hours. *Journal of Finance 59*(2), 681–710.

Barclay, M. J., T. Hendershott, and K. Kotz (2006). Automation versus intermediation: Evidence from treasuries going off the run. *Journal of Finance 61*(5), 2395–2414.

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

Baruch, S. and L. R. Glosten (2013). Fleeting orders.

Battalio, R., A. Ellul, and R. Jennings (2007). Reputation effects in trading on the New York Stock Exchange. *Journal of Finance 62*(3), 1243–1271.

Benveniste, L. M., A. J. Marcus, and W. J. Wilhelm (1992). What's special about the specialist? *Journal of Financial Economics 32*(1), 61–86.

Biais, B., T. Foucault, and F. Salanié (1998). Floors, dealer markets and limit order markets. *Journal of Financial Markets 1*(2), 253–284.

Biais, B., L. Glosten, and C. Spatt (2005). Market microstructure: A survey of microfoundations, empirical results, and policy implications. *Journal of Financial Markets 8*, 217–264.

Brebner, J. T. and A. T. Welford (1980). Introduction: an historical background sketch. In A. T. Welford (Ed.), *Reaction Times*, pp. 1–23. New York: Academic Press.

Brogaard, J., A. Carrion, T. Moyaert, R. Riordan, A. Shkilko, and K. Sokolov (2015). High-frequency trading and extreme price movements.

Brogaard, J., B. Hagstromer, L. L. Norden, and R. Riordan (2014). Trading fast and slow: Colocation and market quality. *Review of Financial Studies 27*(8), 2267–2306.

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

Brogaard, J., T. Hendershott, and R. Riordan (2015). High frequency trading and the 2008 short sale ban.

Carrion, A. (2013). Very fast money: High-frequency trading on the NASDAQ. *Journal of Financial Markets 16*, 680–711.

Chae, J. (2005). Trading volume, information asymmetry, and timing information. *Journal of Finance 60*(1), 413–442.

Chakravarty, S., H. Gulen, and S. Mayhew (2004). Informed trading in stock and option markets. *Journal of Finance 59*(3), 1235–1256.

Chordia, T., R. Roll, and A. Subrahmanyam (2001). Market liquidity and trading activity. *Journal of Finance 56*(2), 501–530.

Chung, K. H., M. Li, and T. H. McInish (2005). Information-based trading, price impact of trades, and trade autocorrelation. *Journal of Banking and Finance 29*(7), 1645–1669.

Clark-Joseph, A. D. (2013). Exploratory trading.

Conrad, J., S. Wahal, and J. Xiang (2015). High-frequency quoting, trading, and the efficiency of prices. *Journal of Financial Economics* (116), 271–291.

Cooper, L. (1993). Market share models. In J. Eliasberg and G. Lilien (Eds.), *Handbook of Operations Research and Management Science Vol. 5: Marketing*, pp. 259–314. New York: North-Holland.

Cooper, L. and M. Nakanishi (1988). *Market Share Analysis*. Boston: Kluwer.

Davis, P. (2002). Estimating multi-way error components models with unbalanced data structures. *Journal of Econometrics* (106), 67–95.

De Jong, F. (2002). Measures of contributions to price discovery: A comparison. *Journal of Financial Markets 5*(3), 323–327.

De Jong, F., T. Nijman, and A. Röell (1995). A comparison of the cost of trading French shares on the Paris Bourse and on SEAQ International. *European Economic Review 39*(7), 1277–1301.

De Jong, F. and P. Schotman (2010). Price discovery in fragmented markets. *Journal of Financial Econometrics 8*(1), 1–28.

Driscoll, J. and A. Kraay (1998). Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics 80*(4), 549–560.

Dufour, A. and R. Engle (2000). Time and the price impact of a trade. *Journal of Finance 55*, 2467–2498.

Easley, D. and M. O'Hara (1992). Time and the process of security price adjustment. *Journal of Finance 47*(2), 577–605.

Egginton, J., B. F. van Ness, and R. A. van Ness (2013). Quote stuffing.

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance 25*(2), 383–417.

Fok, D., P. Franses, and R. Paap (2002). Econometric analysis of the market share attraction model. In P. Franses and A. L. Montgomery (Eds.), *Advanced in Econometrics Vol. 16: Econometric Models in Marketing*, pp. 223–256. New York: JAI/Elsevier.

Foster, F. and S. Viswanathan (1993). Variations in trading volume returns, volatility and trading costs: evidence on recent price formation models. *Journal of Finance 48*(1), 187–211.

Foster, F. D. and S. Viswanathan (1990). A theory of the interday variations in volume, variance, and trading costs in securities markets. *Review of Financial Studies 3*(4), 593–624.

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

Frijns, B. and P. Schotman (2009). Price discovery in tick time. *Journal of Empirical Finance 16*(5), 759–776.

Gai, J., C. Yao, and M. Ye (2013). The externalities of high-frequency trading.

Gallant, A. (1981). On the bias in flexible functional forms and an essentially unbiased form: The fourier flexible form. *Journal of Econometrics 15*(2), 211–245.

Gao, C. and B. Mizrach (2015). Quote stuffing and market quality.

Garman, M. (1976). Market microstructure. *Journal of Financial Economics 3*, 257–275.

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

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

Gonzalo, J. and C. Granger (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business and Economic Statistics 13*(1), 27–36.

Goyenko, R. Y., C. W. Holden, and C. A. Trzcinka (2009). Do liquidity measures measure liquidity? *Journal of Financial Economics 92*, 153–181.

Grammig, J. and F. Peter (2013). Telltale tails: A new approach to estimating unique market information shares. *Journal of Financial and Quantitative Analysis 48*(2), 459–488.

Green, T. C. (2004). Economic news and the impact of trading on bond prices. *The Journal of Finance 59*(3), 1201–1233.

Grossman, S. J. (1992). The informational role of upstairs and downstairs trading. *Journal of Business 65*(4), 509–528.

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

Han, A. and D. Park (1989). Testing for structural change in panel data: Application to a study of U.S. foreign trade in manufacturing goods. *Review of Economics and Statistics 71*, 135–142.

Handa, P., R. Schwartz, and A. Tiwari (2006). The economic value of a trading floor: Evidence from the American Stock Exchange. In R. A. Schwartz, J. A. Byrne, and A. Colaninno (Eds.), *Electronic vs. Floor Based Trading*, pp. 121–151. US: Springer.

Harris, F., T. McInish, and R. Wood (1997). Common long-memory components for intraday stock prices: A measure of price discovery.

Harris, F.H.deB., M. T. and R. Wood (2002). Security price adjustment across exchanges: an investigation of common factor components for Dow stocks. *Journal of Financial Markets 5*(3), 277–308.

Hasbrouck, J. (1991). The summary of informativeness of stock trades: An econometric analysis. *Review of Financial Studies 1*, 571–595.

Hasbrouck, J. (1993). Assessing the quality of a security market: a new approach to transaction-cost measurement. *Review of Financial Studies 6*(1), 191–212.

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

Hasbrouck, J. (2003). Intraday price formation in U.S. equity index markets. *Journal of Finance 58*(6), 2375–2399.

Hasbrouck, J. (2004). Liquidity in the futures pits: Inferring market dynamics from incomplete data. *Journal of Financial and Quantitative Analysis 39*(2), 305–326.

Hasbrouck, J. (2007). *Empirical market microstructure: the institutions, economics, and econometrics of securities trading.* New York: Oxford University Press.

Hasbrouck, J. (2015). High frequency quoting: Short-term volatility in bids and offers.

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

Hasbrouck, J. and G. Sofianos (1993). The trades of market makers: An empirical analysis of NYSE specialists. *Journal of Finance 48*(5), 1565–1593.

Hautsch, N., D. Hess, and D. Veredas (2011). The impact of macroeconomic news on quote adjustments, noise, and informational volatility. *Journal of Banking and Finance 35*(10), 2733–2746.

He, Y., H. Lin, J. Wang, and C. Wu (2009). Price discovery in the round-the-clock U.S. treasury market. *Journal of Financial Intermediation 18*(3), 464–490.

Hendershott, T. and H. Mendelson (2000). Crossing networks and dealer markets: Competition and performance. *Journal of Finance 55*(5), 2071–2115.

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

Hirschey, N. (2013). Do high-frequency traders anticipate buying and selling pressure?

Hoffmann, P. (2014). A dynamic limit order market with fast and slow traders. *Journal of Financial Markets 113*(1), 156–169.

Hsieh, D. and A. Kleidon (1996). Bid-ask spreads in foreign exchange markets: Implications for models of asymmetric information. In J. Frankel, G. Galli, and A. Giovannini (Eds.), *The Microstructure of Foreign Exchange Markets*, pp. 41–65. Chicago: University of Chicago Press.

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

Jondeau, E., J. Lahaye, and M. Rockinger (2015). Estimating the price impact of trades in a high-frequency microstructure model with jumps. *Journal of Banking and Finance 61*(2), 205–224.

Kavajecz, K. A. (1999). A specialist's quoted depth and the limit order book. *Journal of Finance 54*(2), 747–771.

Kempf, A. and O. Korn (1999). Market depth and order size. *Journal of Financial Markets 2*(1), 29–48.

Korenok, O., B. Mizrach, and S. Radchenko (2011). A structural approach to information shares.

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

Lee, S. S. and P. A. Mykland (2012). Jumps in equilibrium prices and market microstructure noise. *Journal of Econometrics 168*, 396–406.

Lehmann, B. N. (2002). Some desiderata for the measurement of price discovery across markets. *Journal of Financial Markets 5*(3), 259–276.

Lei, Q. and G. Wu (2005). Time-varying informed and uninformed trading activities. *Journal of Financial Markets 8*(2), 153–81.

Lin, J.-C., G. Sanger, and G. G. Booth (1995). Trade size and components of the bid-ask spread,. *Review of Financial Studies 8*(4), 1153–1183.

Lockwood, L. J. and S. C. Linn (1990). An examination of stock market return volatility during overnight and intraday periods, 1964-1989. *Journal of Finance 45*(2), 591–601.

Madhavan, A. (2000). Market microstructure: A survey. *Journal of Financial Markets 3*, 205–258.

Madhavan, A. and M. Cheng (1997). In search of liquidity: Block trades in the upstairs and downstairs markets. *Review of Financial Studies 10*(1), 175–203.

Madhavan, A., M. Richardson, and M. Roomans (1997). Why do security prices change? a transaction-level analysis of NYSE stocks. *Review of Financial Studies 10*(4), 1035–1064.

Madhavan, A. and S. Smidt (1993). An analysis of changes in specialist inventories and quotations. *Journal of Finance 48*(5), 1595–1628.

Madhavan, A. and G. Sofianos (1998). An empirical analysis of NYSE specialist trading. *Journal of Financial Economics 48*(2), 189–210.

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

Menkveld, A., S. Koopman, and A. Lucas (2007). Modeling round-the-clock price discovery for cross-listed stocks using state space methods. *Journal of Business and Economic Statistics 25*(2), 213–225.

Menkveld, J. A. and M. A. Zoican (2014). Need for speed? exchange latency and market quality.

Milgrom, P. and N. Stokey (1982). Information, trade and common knowledge. *Journal of Economic Theory 26*(1), 17–27.

Mizrach, B. and C. Neely (2008). Information shares in the U.S. Treasury market. *Journal of Banking and Finance 32*(7), 1221–1233.

Murphy, K. M. and R. H. Topel (1985). Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics 3*(4), 370–379.

Neal, R. and S. M. Wheatley (1998). Adverse selection and bid-ask spreads: Evidence from closed-end funds. *Journal of Financial Markets 1*(1), 121–149.

Ozturk, S. R., M. Van der Wel, and D. J. van Dijk (2014). Intraday price discovery in fragmented markets.

Pagano, M. (1989). Trading volume and asset liquidity. *The Quarterly Journal of Economics 104*(2), 255–274.

Pagano, M. and A. Röell (1996). Transparency and liquidity: a comparison of auction and dealer markets with informed trading. *Journal of Finance 51*(2), 579–611.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance 39*(4), 1127–1139.

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

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

Seppi, D. J. (1990). Equilibrium block trading and asymmetric information. *Journal of Finance 45*(1), 73–94.

Slezak, S. (1994). A theory of the dynamics of security returns around market closures. *Journal of Finance 49*(4), 1163–1211.

Sofianos, G. and I. M. Werner (2000). The trades of NYSE floor brokers. *Journal of Financial Markets 3*(2), 139–176.

Sokolov, K. (2014). High-frequency trading, jumps, and momentum ignition.

Tse, J., X. Lin, and D. Vincent (2012). High frequency trading - measurement, detection and response.

Upper, C. and T. Werner (2007). The tail wags the dog: Time-varying information shares in the Bund market.

Venkataraman, K. (2001). Automated versus floor trading: An analysis of execution costs on the Paris and New York exchanges. *Journal of Finance 56*(4), 1445–1485.

Wansbeek, T. and A. Kapteyn (1989). Estimation of the error-components model with incomplete panels. *Journal of Econometrics 41*(3), 341–361.

Wood, R., T. McInish, and J. Ord (1985). An investigation of transactions data for NYSE stocks. *Journal of Finance 40*(3), 723–39.

Yan, B. and E. Zivot (2010). A structural analysis of price discovery measures. *Journal of Financial Markets 13*(1), 1–19.

Summary

The last decade has seen a series of dramatic transformations in the financial markets. Electronic exchanges now dominate the trading processes of a vast array of financial instruments, while trading pits, if they still exist, are transformed to specialist shops or TV studios for financial news, as reminders of a bygone era. Securities do not merely trade in various exchanges scattered across different countries, but also in many of them simultaneously in the same country. The temporary price differences generated by this market fragmentation and countless other market inefficiencies are arbitraged out by a new set of market participants, well-adjusted to the new electronic trading mediums: high frequency traders. Their secretive and growing arsenal of automatized algorithmic trading strategies will surely continue to provide much food for thought and concern for regulators, researchers as well as the general public.

This dissertation address a number of open questions in this rapidly changing world of intraday trading. The first project proposes a novel methodology to measure intraday changes in price informativeness for financial instruments trading simultaneously in multiple venues. We estimate a structural model based on Hasbrouck (1993) and De Jong and Schotman (2010) using Kalman filtering. This framework decomposes the price variation into permanent price innovations of the underlying efficient price process and transitory noise. We use the flexible Fourier form to model time-variation in these two variances. We provide simulation evidence for the accuracy of this methodology over a wide range of parameter configurations and examine 50 S&P 500 stocks during the second half of

2013. These stocks trade simultaneously in four main exchange groups, namely NYSE, NASDAQ, BATS and Direct Edge. Although NYSE and NASDAQ tend to be the most informative venues, we find statistically significant variation across the trading day depending on the changes in variables such as trade volume, bid-ask spreads, volatility and time-of-the-day.

The second project examines why the trading in electronic markets continues to be concentrated in the hours when the trading pit is open, even after the heyday of pits have passed. We examine two hypotheses providing potential explanations for this activity clustering in the context of a data set of U.S. Treasury futures spanning from 2004 to 2013. The first hypothesis argues that the pit hours attract trading due to the price informativeness and liquidity provided by the trading pit. We find that the contribution of the trading pit to price discovery indeed surpasses its share in the trading activity. However, after the introduction of a new electronic trading platform called Globex in 2008, both the number of trades and the informativeness of the trading pit plummet and this does not cause much change in the activity clustering around the pit hours.

The second hypothesis postulates a feedback mechanism between trading activity, price informativeness, information asymmetry and price impact of trades, which keeps afterhours trading at a low level. We estimate these market variables using a structural market microstructure model and find statistically and economically significant relationships between these variables across the sample period. From 2008 on we observe that the effect of the price impact of trades becomes statistically insignificant and average trade sizes both considerable decrease and become more similar across the trading day. We attribute this to the sophistication of algorithmic trade execution strategies, facilitated by the introduction of the Globex Platform.

The third project investigates the existence and prevalence of a predatory trading tactic aimed at generating transitory price trends, allegedly employed by high frequency traders. We find 1,675,100 extreme price move events in a NASDAQ data set of 8,000 stocks from July 2007 to December 2013. This figure corresponds to a monthly average of 3.4 extreme moves for each stock. About half of these events are transitory, i.e. the price move is followed by a reversal of more than two-thirds, and nearly a quarter are permanent, i.e. the price reverts by less than one-third of the initial move.

We use a fixed effects model controlling for factors specific for each stock, month and intraday period. We also consider other scenarios which may lead to an extreme move and reversal pattern such as temporary liquidity imbalances and the overreaction of non-HFT parties. In line with the documented positive effects of HFT market making, the HFT activity during the move and the reversal tend to alleviate the market quality drop caused by the extreme move event. However the HFT activity before the event shows signs of market manipulation from October 2011 on: The premove HFT activity during the transitory moves becomes more predictive of the move size and direction, correlates with larger subsequent price reversals and exacerbates the deterioration in market quality variables namely quoted spreads, market volatility and execution-to-cancellation ratios.

Samenvatting (Summary in Dutch)

Tijdens het laatste decennium hebben er een reeks dramatische transformaties plaatsgevonden binnen de financiële markten. Elektronische beurzen zijn het handelsproces van een breed scala aan financiële instrumenten gaan domineren. Handelsvloeren, voor zover ze nog bestaan, zijn veranderd zijn in speciaalzaken of televisie studio's die gebruikt worden voor het uitzenden van het financiële nieuws, ter nagedachtenis aan een vroeger tijdperk. Effecten worden niet alleen verhandeld op verschillende beurzen verspreidt over verschillende landen, maar ook op verschillende beurzen tegelijk in hetzelfde land. De tijdelijke prijsverschillen die gegenereerd worden door de marktfragmentatie en talloze andere markt inefficiënties, verdwijnen snel door een nieuwe groep marktdeelnemers, flitshandeleren, die goed aangepast zijn aan de nieuwe elektronische handelsmechanismen. Hun geheime en groeiende arsenaal aan geautomatiseerde, algoritmische handelsstrategieën leiden tot nadenken en bezorgdheid bij de marktautoriteiten, onderzoekers en de maatschappij in het algemeen.

Deze dissertatie beschouwt een aantal open vragen in de snel veranderende wereld van handelen op financiële markten. Het eerste essay stelt een nieuwe methodologie voor om veranderingen in het informatiegehalte van effectenprijzen op verschillende locaties te meten. Dit model schat een structureel model gebaseerd op Hasbrouck (1993) en De Jong en Schotman (2010), door middel van het Kalman filter. Dit raamwerk ontbindt de prijsvariatie in permanente prijsveranderingen van het onderliggende efficiënte prijs proces en tijdelijke ruis. Er wordt gebruik gemaakt van de flexibele Fourier vorm om de tijdsvariatie te modelleren in deze twee varianties. Door middel van simulaties wordt de nauwkeurigheid van de methodologie gegeven voor een wijde reeks aan parameter instellingen, en als toepassing worden 50 S&P 500 aandelen tijdens de tweede helft van 2013 bekeken die op vier beurzen tegelijkertijd verhandeld worden (namelijk NYSE, NASDAQ, BATS en Direct Edge). Alhoewel de NYSE en NASDAQ over het algemeen het meest informatief zijn, vinden we statistisch significante variatie over de handelsdag afhankelijk van veranderingen in variabelen zoals handelsvolume, het verschil tussen bied- en laatprijzen ('spread'), volatiliteit en tijd van de dag.

Het tweede essay bestudeert waarom het handelen in elektronische beurzen vooral plaatsvindt in de uren dat de handelsvloer open is, ook na de hoogtijdagen van handelsvloeren. Ik onderzoek twee hypotheses die een verklaring zouden kunnen geven voor deze clustering van activiteit in de context van een dataset van U.S. Treasury futures met data van 2004 tot en met 2013. De eerste hypothese stelt dat de openingstijden van handelsvloeren handelsactiviteit aantrekken door de informatiewaarde van de prijs en liquiditeit verstrekt door de handelsvloer. De resultaten geven aan dat de contributie van de handelsvloer tot prijsontdekking inderdaad sterker is dan het aandeel in handelsactiviteit. Echter, na de introductie van het nieuwe elektronische handelsplatform genaamd Globex in 2008, daalde zowel het aantal transacties als de informatiewaarde van de handelsvloer sterk en dit verandert weinig aan het clusteren van de handelsactiviteit tijdens de handelsvloer uren.De tweede hypothese postuleert een terugkoppelingsmechanisme tussen handelsactiviteit, informatiewaarde van de prijs, informatie asymmetrie en invloed van transacties op de prijs, die handel buiten de openingstijden van de handelsvloer laag houdt. Deze marktvariabelen worden geschat met behulp van een structureel model, en de resultaten tonen statistisch en economisch significante relaties aan tussen deze variabelen over de schattingsperiode. Vanaf 2008 is te zien dat het effect van invloed van transacties op de prijs statistisch insignificant wordt en gemiddelde transactie volumes zowel kleiner worden als meer op elkaar gaan lijken over de handelsdag. Dit valt toe te schrijven aan het raffinement van de op algoritmes gebaseerde handelsstrategieën, gefaciliteerd door de introductie van het Globex Platform.

Het derde essay onderzoekt het bestaan en de invloed van roofzuchtige tactieken gericht op het genereren van kortstondige prijstrends, die naar verluid worden toegepast door flitshandelaren. In een NASDAQ dataset van 8.000 aandelen van juli 2007 tot en met december 2013 zijn er 1.675.100 extreme prijsveranderingen. Dit komt neer op een maandelijks gemiddelde van 3,4 extreme veranderingen voor elk aandeel. Ongeveer de helft van de veranderingen is tijdelijk, d.w.z. de prijsverandering is gevolgd door een terugval van meer dan twee derde, en bijna een kwart is permanent, d.w.z. de prijs valt voor minder dan een derde terug ten opzichte van de initiële verandering. Het derde essay maaktgebruik van een fixed effects model zodat gecontroleerd kan worden voor factoren die specifiek zijn aan het aandeel, de maand en periode van de dag. Ookook andere scenario's die kunnen leiden tot een extreme prijsverandering en terugval worden beschouwd, zoals een tijdelijke afwijking van liquiditeit en een overreactie van niet-flitshandelaren. In lijn met de gedocumenteerde positieve effecten van flitshandelaren market making, heeft de activiteit van flitshandelaren tijdens prijsverandering en terugval de neiging om de verslechtering van de marktkwaliteit door de extreme prijsverandering te verminderen. De activiteit van flitshandelaren voor de prijsverandering vertoont tekenen van marktmanipulatie vanaf oktober 2011: De grootte en richting van de prijsverandering valt nog tijdens deze verandering te voorspellen met activiteit van flitshandelaren, correleert met een grotere opvolgende prijs terugval en verergert de verslechtering van markt kwaliteit variabelen, namelijk spread, markt volatiliteit en ratio's van uitgevoerde tot gecancelde transacties.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

625 J.J. DE VRIES, Estimation of Alonso's Theory of Movements for Commuting

626 M. POPŁAWSKA, Essays on Insurance and Health Economics

627 X. CAI, Essays in Labor and Product Market Search

628 L. ZHAO, Making Real Options Credible: Incomplete Markets, Dynamics, and Model Ambiguity

629 K. BEL, Multivariate Extensions to Discrete Choice Modeling

630 Y. ZENG, Topics in Trans-boundary River sharing Problems and Economic Theory

631 M.G. WEBER, Behavioral Economics and the Public Sector

632 E. CZIBOR, Heterogeneity in Response to Incentives: Evidence from Field Data

633 A. JUODIS, Essays in Panel Data Modelling

634 F. ZHOU, Essays on Mismeasurement and Misallocation on Transition Economies

635 P. MULLER, Labor Market Policies and Job Search

636 N. KETEL, Empirical Studies in Labor and Education Economics

637 T.E. YENILMEZ, Three Essays in International Trade and Development

638 L.P. DE BRUIJN, Essays on Forecasting and Latent Values

639 S. VRIEND, Profiling, Auditing and Public Policy: Applications in Labor and Health Economics

640 M.L. ERGUN, Fat Tails in Financial Markets

641 T. HOMAR, Intervention in Systemic Banking Crises

642 R. LIT, Time Varying Parameter Models for Discrete Valued Time Series

643 R.H. KLEIJN, Essays on Bayesian Model Averaging using Economic Time Series

644 S. MUNS, Essays on Systemic Risk

645 B.M. SADABA, Essays on the Empirics of International Financial Markets

646 H. KOC, Essays on Preventive Care and Health Behaviors

647 V.V.M. MISHEVA, The Long Run Effects of a Bad Start

648 W. LI, Essays on Empirical Monetary Policy

649 J.P. HUANG, Topics on Social and Economic Networks

650 K.A. RYSZKA, Resource Extraction and the Green Paradox: Accounting for Political Economy Issues and Climate Policies in a Heterogeneous World

651 J.R. ZWEERINK, Retirement Decisions, Job Loss and Mortality

652 M. K. KAGAN, Issues in Climate Change Economics: Uncertainty, Renewable Energy Innovation and Fossil Fuel Scarcity

653 T.V. WANG, The Rich Domain of Decision Making Explored: The Non-Triviality of the Choosing Process

654 D.A.R. BONAM, The Curse of Sovereign Debt and Implications for Fiscal Policy

655 Z. SHARIF, Essays on Strategic Communication

656 B. RAVESTEIJN, Measuring the Impact of Public Policies on Socioeconomic Disparities in Health

657 M. KOUDSTAAL, Common Wisdom versus Facts; How Entrepreneurs Differ in Their Behavioral Traits From Others

658 N. PETER, Essays in Empirical Microeconomics

659 Z. WANG, People on the Move: Barriers of Culture, Networks, and Language

660 Z. HUANG, Decision Making under Uncertainty-An Investigation from Economic and Psychological Perspective

661 J. CIZEL, Essays in Credit Risk, Banking, and Financial Regulation

662 I. MIKOLAJUN, Empirical Essays in International Economics

663 J. BAKENS, Economic Impacts of Immigrants and Ethnic Diversity on Cities

664 I. BARRA, Bayesian Analysis of Latent Variable Models in Finance