London School of Economics and Political Science

Essays on Financial Economics

Marcela Andrea Valenzuela Bravo

Thesis submitted to the Department of Finance of the London School of Economics for the degree in Doctor of Philosophy.
London, August 2013

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

Statement of Cojoint Work

I confirm that the first two chapters of my dissertation, “Global Depth and Future Volatility” and “Competition, Signaling and Walking through the Book: Effects on Order Choice” are based on joint work with Ilknur Zer, also a PhD candidate from the London School of Economics.

To Wolfram and Oliver

Abstract

This thesis consists of an introductory chapter, three main chapters, and a concluding chapter. In Chapter 2, my co-author and I provide new empirical evidence that the distribution of liquidity has a strong in-sample and out-of-sample predictive power on intraday market volatility. To this end, we introduce a novel way of summarizing the relative depth provision in the whole limit order book. Our measure, *global depth*, considers the entire quoted depth and assigns weights decreasing with distance from the best quotes. We document that global depth outperforms alternative predictors of volatility, such as the bid-ask spread, standard depth variables, and measures of trading activity, in explaining the variations in market volatility.

The third chapter, forthcoming in the Journal of Banking and Finance, investigates the effects of competition and signaling in a pure order driven market and examines the trading patterns of agents when walking through the book is not allowed. We show that the variables capturing the cost of a large market order are not informative for an impatient trader under this market mechanism. We also document that the competition effect is not present only at the top of the book but persistent beyond the best quotes. Moreover, we show that institutional investors' order submission strategies are characterized by only a few pieces of the limit order book information.

The fourth chapter provides evidence that implied correlation is a significant in-

dicator of market-wide risk. From an aggregate perspective, I document that implied correlation explains an important fraction of the variation in market excess returns, with high implied correlation followed by an increase in subsequent market returns. The predictive power is stronger at a forecast of bimonthly, quarterly and semi-annually return horizons and robust to the inclusion of standard predictors. Moreover, I show that the information content of the correlation risk premium on market returns is fully driven by the implied correlation. My findings indicate that periods of high market-wide correlation produce a deterioration of the investment opportunity set and, as a consequence, an increase in the equilibrium expected return.

Acknowledgements

I am deeply indebted to my supervisor Dr. Christian Julliard for his patience, encouragement and guidance during the past years. I am also grateful to Dr. Jon Danielsson and Dr. Kevin James for their great co-authorships, for talks about finance and other philosophical questions, and for making my life at the LSE more enjoyable. Thanks to Prof. Thorsten Rheinlander and Prof. Piotr Fryzlewicz for their constant support and our fruitful discussions about research.

I am grateful to Dr. Andrea Vedolin and Prof. Richard Payne for agreeing to be my examiners, and their insightful comments for this thesis. Thanks Andrea for always having time to listen to my questions and for always giving me excellent suggestions. I would also like to thank Prof. Christopher Polk, Dr. Amil Dasgupta, Dr. Vicente Cunat, Dr. Michela Verardo and Prof. Daniel Ferreira for their invaluable advice. I am also thankful to Mary Comben and Osmana Raie for their administrative support during my PhD studies.

A very special thank you goes to Ilknur Zer, not only for our joint venture that resulted in two chapters of this dissertation, but also for her friendship and for making the years at the LSE such a pleasant period that I will always dearly remember. Thanks to my friends Nelson Camanho, Victoria Dobrynskaya, Christian von Drathen, Svetlana Bryzgalova and Sean Lew for being part in this adventure of working towards a PhD, and for sharing the good but also the difficult moments with me.

Finally, I would like to thank my family. First, my mother for always encouraging me to achieve more. You are the artificer of what I am now. Thanks to my son Oliver for giving me the motivation to finish my PhD studies. Lastly, and most especially, I'd like to thank my husband Wolfram for his constant support. Thank you for being part of my life, for always having time to listen and for helping me whenever I need it; for your advice on programming and for solving my LaTeX problems; for sharing the care of Oliver in those difficult periods...if I continued this list, I would never finish thanking you!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Information Content of a Limit Order Book

The information content of the limit order book has been addressed extensively in the theoretical literature of market microstructure. Being one of the first dynamic equilibrium models on limit order markets, Parlour (1998) analyses the "competition effect" on order choice. She suggests that an increase in the same–side thickness of the LOB "crowds out" limit orders on that side, since higher competition decreases the execution probability. Similarly, an increase in the opposite-side thickness is anticipated as a decreasing execution risk, hence encouraging more aggressive behavior. This crowding-out effect is symmetric for both sides of the book. In a dynamic equilibrium model, Foucault (1999) proposes that the order choice depends mainly on the asset volatility. When volatility increases, a limit order trader demands larger compensation for the risk of being picked off by posting higher ask and lower bid prices. This makes market orders more costly, which in turn increases the proportion of limit orders on the total order flow. Foucault, Kadan and Kandel (2005) consider the actual spread as a determinant of the order choice of the strategic liquidity traders that differ in their waiting costs. They conclude that for certain levels, high cost traders (impatient investors) will submit market orders, whereas others submit limit orders. However if spread increases over a cutoff level, all traders will supply liquidity to the market. Goettler, Parlour and Rajan (2005) solve numerically for the stationary Markov perfect equilibrium in a model in which traders endogenously choose whether to submit a market or a limit order and the order size. On the other hand, Rosu (2009), similar to Foucault et al. (2005), models a continuous-time market, but with a dynamic investor decision problem, i.e., an agent can modify her strategy decision continuously. Two of the very recent theoretical works allow asymmetric information for pure order driven markets; Goettler, Parlour and Rajan (2009) and Rosu (2012). In Goettler et al. (2009), informed traders are liquidity providers, i.e., they submit limit orders. However, in high volatility states, they switch their order choice to market orders to take advantage of the mispriced orders waiting in the queue. Rosu (2012), on the other hand, proposes that the informed traders can be patient or impatient based on how far the fundamental value is from the public price. That is; if the fundamental value of an informed trader is well above the public price plus a cutoff value, which is proportional to the volatility, then the agent will be aggressive and submit a market order to take advantage of her information instead of waiting to be compensated by a limit order.

Based on the aforementioned theoretical literature, Chapters 2 and 3 of this dissertation aim at understanding the effects of the information content of a limit order book in a pure order driven market. In "Global Depth and Future Volatility" we propose a new way of summarizing the distribution of liquidity in a limit order book and we examine whether this is informative about future volatility. On the other hand, "Competition, Signaling and Non-Walking Through the Book: Effects on Order Choice" investigates whether the state of a limit order book shapes investors' choice of order submission.

We use high frequency data of the Istanbul Stock Exchange (ISE). Our dataset consists of order and trade books of the biggest 30 stocks (ISE-30 index). By matching these two books, we reconstruct the complete limit order book dynamically. Hence, at any given time, we have a list of all orders waiting to be executed, whether they are buy or sell orders, their limit prices, and the volume accumulated at each quote. There are particular characteristics of our market that make it convenient for the analysis. Similar to other markets (Australian Stock Exchange and the Spanish Stock Exchange for instance), the ISE is an open limit order book market. Investors are able to observe outstanding and traded orders in real time. Moreover, ISE is a fully transparent market. The information is not limited to a certain quote, but investors can observe the information of the whole book. Another particular characteristic of ISE, similar to the Australian Stock Exchange, the Sao Paulo Stock Exchange (Bovespa), and the Stock Exchange of Hong Kong for instance, is that walking through the book is not allowed. When an investor submits a large market order, the unexecuted part is converted to a limit order at the quoted price instead of walking up or down the limit order book to be fully executed.

In Chapter 2, we focus on how the distribution of market liquidity is related to market realized volatility, as well as how the distribution of individual-stock liquidity is related to individual-stock volatility. To this end, we develop a new measure that summarizes the distribution of orders waiting at different price levels, i.e., it summarizes the limit order book distribution. Our measure, global depth, considers the entire quoted depth and assigns weights decreasing with distance from the best quotes. The construction of global depth is motivated from the current literature on order choice. We conclude from this literature that not only the volume of orders at the best quotes, but also the depth beyond the best quotes matters for developing a trading strategy. Hence, our proposed measure has two ingredients: it summarizes the volume distribution in the whole limit order book and weighs information based on the price distance from the best quotes.

In order to construct global depth we first sample a limit order book for every 15-minute trading intervals to obtain best bid and ask prices, submitted volume and the limit price of each order. We then calculate the (tick-adjusted) price distance of each limit order relative to the best limit price and the proportion of volume waiting up to each price level (cumulative volume distribution). Finally, we assign exponentially decaying weights with price distance, and thus we give the highest importance to the information at the best quotes, second highest to the second best quotes, and so forth.

The economic link between liquidity and future volatility follows the predictions of Goettler et al. (2005) and Goettler et al. (2009); if the volume of orders accumulated away from the best quotes increases, this may signal to the market that the current quotes are mispriced ("signaling effect"). In this case, large jumps are more likely, and thus we expect higher future volatility.

We employ a standard predictive regression model of the market volatility on aggregate global depth. We estimate the volatility of the value-weighted ISE-30 index employing the two-scales realized volatility estimator proposed by Ait-Sahalia, Mykland and Zhang (2011). The aggregate global depth for both sides of the market is obtained from the cross-sectional averages of individual stock's global depth. We document that global depth significantly and negatively predicts the intraday market volatility up to 150 minutes ahead. In a simple setting, global depth variables explain almost 25% of the variation in market volatility. Our measure is both economically and statistically the strongest in explaining future market volatility among other liquidity measures (e.g. bid-ask spread, Amihud illiquidity measure), standard predictors of volatility (e.g. trading activity variables), standard depth variables, and lag volatility. We also provide evidence of out-of-sample forecasting performance; global depth predicts market volatility up to 75 minutes ahead, and it explains over 14% of the out-of-sample variation in the 15-minutes-ahead market volatility. Finally, we document that the individual liquidity distribution–individual volatility relationship is negative and significant for 83% of the stocks in our sample.

Chapter 3 also explores the information content of a limit order book. This chapter directly examines the effects of the book information on the order choice of an agent. We specifically answer the following questions: (1) Does "competition effect" dominate "signaling effect" at every level of the depth? (2) How does the non-walking through the book market mechanism affect order decision of an impatient trader? (3) What is the difference in trading behavior between institutional traders and individual investors?

To this end, we model the order choice of an agent as a two-stage process. In the first stage, an investor decides whether to be impatient and submit a market order, or to be patient and submit a limit order. In the second stage, given that the agent is impatient, she decides whether to submit a large market order or a small market order based on the quantity to trade. On the other hand, if she is patient, she decides her limit price, i.e., whether to place an order at the best quotes, within the best quotes or beyond the best quotes. Hence, our empirical analysis relies on a two-stage sequential ordered probit (SOP) model.

To analyze competition and signaling effects beyond the best quotes, we mainly focus on the actions of patient traders. We document that for a patient trader, the competition effect overcomes the signaling effect for the depth variables closer to the best quotes whereas the signaling effect is stronger for the depth away from the best quotes. Particularly, we find that the volume up to the second best quotes has the strongest competition effect.

To examine the effects of non-walking through the book, we focus on the trading strategies of impatient traders. When walking through the book is allowed, price related variables such as spread, and price distances variables capture the cost of a large market order. As expected, we find that none of these variables affect the order choice of an impatient trader. On the other hand, a market order investor considers volatility, previous price trend, and volume accumulated beyond the best quotes on the opposite side of the book while deciding the quantity to trade. This result is also consistent with the non-walking through the book mechanism, since these variables affect the execution probability of the limit order-converted part of large market orders.

Finally, to study the trading patterns of institutional and individual investors, we focus on the first state of the SOP model. Institutional traders only take into account competition effect variables to decide whether to submit a market or a limit order. If they are informed traders (Chakravarty (2001)), our results suggest that institutions base their decisions more on their own private valuations than on the state of the limit order book.

1.2 Implied Correlation and Expected Returns

Chapter 4 examines the predictive power of aggregate implied correlation over market returns. The question of whether asset returns are predictable has a long history in finance. Since Kendall (1953) indicated that stock prices could follow a random walk, the literature has extended to provide evidence that stock returns are predictable. Campbell (1987) and Fama and French (1989), among others, document the predictive power of business cycle variables such as the term structure of interest rates and the default spread. When economic conditions are bad, returns are expected to be higher. As income is low, investors decrease their consumption to invest more only if returns are expected to be high in the future. Lettau and Ludvigson (2001) show that the short-term deviation between aggregate consumption, asset holdings, and labor (CAY) also predicts stock market returns. They argue that transitory movements induced by time variation in expected returns are intended to be "smoothed out". Hence, investors will increase their consumption over wealth (CAY increases) if subsequent excess returns are expected to be higher. On the other hand, financial ratios have also been found to contain information on future returns.

As Lewellen (2004) suggests, the negative relationship between the price-to-dividend ratio and subsequent returns can be given by a mispricing view. When the price-to-dividend ratio is high, prices are higher than fundamentals. In this case, returns should be lower in the future, since prices revert back to fundamentals.

I extend this literature and explore the predictive power of a new variable, the implied correlation, over future market returns. One of the first papers documenting that correlation changes over time is by Goetzmann, Li and Rouwenhorst (2001). In their study, they examine the correlation structure of the main world equity markets and find that correlation varies over time, which makes the diversification benefits also time-varying. Longin and Solnik (2001) present evidence that international correlation increases in bear markets, but not in bull markets. In the same line, Ang and Chen (2002) and Hong, Tu and Zhou (2007) study the stock correlations conditioning on extreme movements and find that the correlation is much higher for downside than upside moves. Hence, the diversification benefits decrease in times when they are most demanded. The time-varying nature of correlation and the reducing diversification benefits in bad times motivate the question of whether correlation is a risk factor, i.e., whether investors are willing to pay a premium for securities that pay off well in states of high correlation.

Krishnan, Petkova and Ritchken (2009) investigate this question by using a physical measure of correlation and document a negative price of correlation risk. Taking a different approach, Driessen, Maenhout and Vilkov (2009) decompose the market variance risk premium into correlation risk premium and individual variance risk premia using index options and individual stock options. They illustrate that the negative risk premium for the market variance is only consistent with a negative price of correlation risk. Mueller, Stathopoulos and Vedolin (2012), on the other hand, analyze the currency markets and show that correlation is priced in the cross-section of currency returns.

This chapter takes an aggregate perspective and investigates whether changes in implied correlation affect the expected market risk premium in the time-series. To this end, I use option data of the S&P100 index and its individual components to construct the aggregate implied correlation. I estimate risk-neutral expectations of variances from the strike of a simple variance swap introduced by Martin (2011). I rely on a regression model of the market excess return on the lagged implied correlation. The market risk premium is obtained as the CRSP NYSE/AMEX/NASDAQ value-weighted portfolio returns in excess of the one-month treasury bill rate.

My findings reveal that aggregate implied correlation is highly and positively related to subsequent market excess returns. I document that the predictive power is stronger for intermediate prediction horizons, and robust to the inclusion of different control variables such as the price-to-dividend and price-to-earnings ratios, consumption over wealth, default spread, term spread, relative risk-free rate, realized variance, implied variance, realized correlation, market variance risk premium, and the cross-sectional average of individual variance risk premia. The economic importance of implied correlation is the strongest compared to the rest of the variables; a one standard deviation increase in implied correlation translates into 1.31% increase in three-months-ahead monthly market returns.

The results presented in this chapter provide evidence that periods of high correlation indicate an increase in aggregate risk; when business conditions deteriorate, risk averse investors demand a higher risk premium to hold aggregate wealth, inducing an increase in the market risk premium.

Chapter 2

Global Depth and Future Volatility

Co-authored with Ilknur Zer (London School of Economics)

2.1 Introduction

This paper examines the link between two central concepts in financial markets: liquidity and volatility. Liquidity is essential for well-functioning financial markets. It is generally ample but occasionally evaporates very rapidly, signaling the start of a crisis. Therefore, it is crucial to understand the effects of liquidity provision on market dynamics. This has gained an increased attention from regulators, market participants, and academics alike. Nevertheless, we are still in the early stages of accurately defining and measuring liquidity, due to its unobservable and multidimensional nature. On the other hand, information on future volatility is one of the main ingredients in assessing risk-return trade-off for portfolio valuation, derivatives pricing models, and it affects the execution probability of a limit order. In this paper, we propose a novel way of summarizing the distribution of liquidity in a limit order book and examine whether liquidity dry-ups in equity markets anticipate spikes in volatility.

Our focus is to evaluate the role of the relative depth provision in future *market* volatility. Predicting market volatility, rather than an individual stock volatility, is important because it approximates the aggregate uncertainty. It is an indicator for policy makers of the vulnerability of financial markets, as changes in market volatility have systemic repercussions on the whole economy (see Schwert (1989) and Poon and Granger (2003) for further discussions). An individual stock volatility, on the other hand, may increase due to stock-specific news or announcements, and not necessarily due to systemic events, such as a sudden withdrawal of liquidity. We examine the volatility–liquidity relationship at an *intraday* level. Trading in financial markets nowadays mostly takes the form of electronic markets, where trading occurs fast. Hence, during stressed market conditions, liquidity may disappear very quickly. For example, the withdrawal of the high-frequency liquidity providers has contributed to the volatility present within the flash crash of 2010 within minutes. This makes it desirable to study the market dynamics at an intraday level. Nevertheless, little research has been undertaken to study the predictive power of *market* liquidity on *market* volatility at an *intraday* frequency.[1]

The high-frequency relationship between liquidity and subsequent volatility has important implications on traders' order choice strategies. There is extensive evidence, both theoretically and empirically that investors submit limit orders in high volatility states (see Foucault (1999) and Ranaldo (2004) for instance). When volatility is high, the risk of being picked off by an informed agent increases, inducing investors to submit less aggressive orders. Another explanation is given by the option-like feature of limit orders. Placing a buy (sell) limit order is equivalent to writing a free put (call) option to the market (Handa and Schwartz (1996)). The higher the volatility, the higher the option value of the limit order, as in this case the probability that the spot price hits the limit price increases. Hence, this paper presents a statistical model to predict volatility using available limit order book information, which can be employed by market participants to submit less aggressive orders when

[1] Relevant exceptions are Chordia, Roll and Subrahmanyam (2001) and Pastor and Stambaugh (2003). However, both studies focus on a contemporaneous relationship at a daily frequency.

volatility is expected to be high.

We provide new empirical evidence that the distribution of orders waiting to be traded strongly predicts market volatility. We measure the liquidity distribution by developing a short-run market measure, *global depth*. A stock's global depth is a weighted average of the volume of orders waiting at the entire limit order book, with weights decreasing with distance from the best bid/ask price. The aggregate level is the average of global depth of individual stocks. One natural motivation behind the weighting scheme comes from the execution probabilities. Limit orders submitted farther away from the best quotes have lower execution probabilities compared to the ones submitted closer to the best quotes. Hence, a trader gives higher weights to the information around the best quotes compared to the rest of the book.

There are several practical routes that one could take to construct a liquidity measure. With our approach, we aim to fill a gap that is left by the existing literature. Many studies focus on the volume of orders at the highest bid and the lowest ask prices (depth at the best quotes). Some others include the volume of orders waiting beyond the best quotes up to a specific price level. The main conclusion we extract from these studies is that, although both matter, depth at the best quotes is more informative.[2] Hence, a relevant proxy to capture the available liquidity needs two ingredients: it should consider the whole book and weigh the information in the book based on price distances.[3] In order to construct our measure, global depth, we first sample the limit order book in discrete trading intervals. Second, we consider the (tick-adjusted) price distance of each order relative to the best limit price. Then, by calculating the percentage of total volume supplied or demanded up to a given price distance, we obtain the empirical cumulative distribution function of the limit

[2]See Parlour (1998), Ahn, Bae and Chan (2001), Handa, Schwartz and Tiwari (2003), Ranaldo (2004), Bloomfield, O'Hara and Saar (2005), Foucault et al. (2005), Ellul, Holden, Jain and Jennings (2007), Cao, Hansch and Wang (2008), Cao, Hansch and Wang (2009), Pascual and Veredas (2009), Goettler et al. (2009) and Valenzuela and Zer (2013), among others.

[3]Price distance refers to the position of a given bid (ask) with respect to the best bid (ask) price.

order book. Finally, a stock's global depth is the weighted average of the distribution function, where weights are decreasing with price distances.

Compared to standard liquidity measures like spread, depth, and ratios based on both spread and depth, global depth provides a more complete picture of the liquidity provision by considering the whole book. Instead of focusing on the size of the orders waiting, our measure is based on the distribution of volume at a given time. That is, it measures the relative concentration of depth provision at each quote, which reveals information of the disagreement on the true price. As models of Goettler et al. (2005) and Goettler et al. (2009) predict, if orders waiting in a given book are accumulated at a quote farther away from the best prices, then this signals to the market that current quotes are mispriced. In this case, jumps are plausible, creating higher future volatility. On the other hand, higher liquidity provision around the best quotes relative to the rest of the book is associated with a consensus on the current price; therefore, we expect lower future volatility.

While conceptually this study could be conducted in any limit order market, there are certain market characteristics that are of particular benefit to address the liquidity–volatility relationship. It is definitely helpful if the data contains the entire order book. This is not the case for most data from the European and the US markets because of the multiple trading platforms and hybrid market structures. That makes the information flow fragmented. Furthermore, it is important for our analysis that the market provides high pre-trade transparency, i.e., the market participants can observe the whole book rather than being limited to the best five or ten quotes. One exchange that meets these criteria and is relatively large is the Istanbul Stock Exchange (ISE).[4] The order and trade books from ISE form the dataset that we use in this study. By matching these two books and removing the executed orders, we

[4] As of December 2011, ISE is the 20^{th} (8^{th}) biggest stock exchange in the world (Europe–Africa–Middle East region) in terms of value of share trading in electronic order book trades with a trading value of $405,136 million. See, the World Federation of Exchanges for details.

reconstruct the limit order book. That is, for a given time we have the best bid and ask prices, all of the orders waiting to be executed, their submitted prices and their corresponding volumes.

Our empirical results contribute to our understanding of the relationship between liquidity and future volatility of the efficient price. It is challenging to estimate intraday volatility of the true price because of the microstructure noise arising from several sources inherent in the trading process or high-frequency data, such as the informational effects, bid–ask bounces, or data recording errors. Ait-Sahalia et al. (2011) address this specific problem and provide the volatility proxy that we use in our study.

We provide new empirical evidence on both in-sample and out-of-sample informativeness of the liquidity distribution on market volatility of the efficient price at an intraday level. We show that global depth is both economically and statistically the strongest among standard liquidity and trading activity measures, on explaining the variations in market volatility. Out-of-sample forecasting tests provide formal evidence for substantial forecasting power of global depth. It predicts one-period-ahead market volatility with an out-of-sample R^2 of over 14%, where the forecasting power lasts up to 75 minutes ahead. Finally, we show that the time-series relation between global depth and market volatility is not driven by variations in a particular stock or industry, but rather that the relation is shared by the majority of the stocks. We find a negative and significant relationship between the individual stock level global depth and future volatility for 83% of the stocks in our sample.

The rest of the paper is organized as follows: the next section frames our work within the context of the existing literature. Section 2.3 describes data and the trading structure in our market. Section 2.4 explains the estimation of our measure in detail. Section 2.5 introduces the econometric methodology and variables included in the analysis. Estimation results, the out-of-sample forecasting evaluations, and robustness checks are presented in Section 2.6. Finally, Section 2.7 concludes.

2.2 Related Literature

This paper relates to recent literature that attempts to measure the liquidity provision considering the whole book. Domowitz, Hansch and Wang (2005) propose an illiquidity measure based on the supply and demand step functions for a given security. By using data from the Australian Stock Exchange, they conclude that not only the liquidity risk, but also the liquidity commonality, is priced in stock returns. In another related study, Naes and Skjeltorp (2006) examine the informativeness of the order book from the Oslo Stock Exchange. They introduce a new variable–the slope of the book–that describes the average elasticity across all price levels with the corresponding volumes, and show that it is negatively related to both trading volume and price volatility. Our contribution to this literature is twofold: first, we propose a new way of summarizing the state of the whole book, which considers the distribution of depth at different price levels. In addition, our proposed measure, global depth, weighs information provided by different quotes by assigning the highest weights to the best quotes and lower weights for the quotes that are farther away from the best prices. Second, by including several liquidity measures in our analysis, we run a horserace among them and evaluate their performances in explaining future volatility.

Our work also builds on the literature illustrating that limit orders are information driven. Foucault et al. (2005), Kaniel and Liu (2006), Rindi (2008), Goettler et al. (2009), and Rosu (2012) provide theoretical background explaining that informed traders may reveal their private information via limit order submissions. Foucault et al. (2005) show that if spread increases over a cutoff level, all traders submit limit orders. In the setting of Goettler et al. (2009), although informed traders are liquidity providers, they switch to market orders in order to benefit from the mispricing in high volatility states. In Rosu (2012)'s model, informed traders can submit limit or market orders based on how far the fundamental value is from the publicly available price. Kaniel and Liu (2006) show that informed traders are more likely to submit limit orders than market orders if the information is long lived. In the model of Rindi (2008), liquidity suppliers can be either uninformed or informed. She shows that the disclosure of traders' identity decreases the adverse selection, motivating informed traders to provide more liquidity. Bloomfield et al. (2005), Anand, Chakravarty and Martell (2005), and Menkhoff, Osler and Schmeling (2010) complement this literature by providing empirical evidence that informed traders submit limit orders. In this paper, we document evidence from an emerging country stock exchange that the limit order book contains information shaping agents' trading decisions. We show that several summary measures extracted from the limit order book have explanatory power on future volatility.

Finally, our paper is closely related to the literature that examines the predictive power of liquidity on volatility. In an early empirical work, Ahn et al. (2001) analyze the interactions between transitory volatility and order flow composition by using data from the Stock Exchange of Hong Kong. They show that an increase in transitory volatility is followed by an increase in the market depth, where the latter is measured by the total number of limit orders posted at the best quotes. Moreover, they show that although the depth at the best quotes explains future individual volatility, the depth beyond the best quotes does not have any explanatory power. Hence, they conclude that the transitory volatility arises mainly from the scarcity of limit orders at the best quotes. By employing cointegration analysis for the bid and ask quotes, Pascual and Veredas (2010) separate transitory volatility from informational volatility (volatility arising by the actions of informed agents) and show that trade size and quoted depth both at the best and away from the quotes have predictive power on individual volatility. Duong and Kalev (2008) investigate the forecasting power of the Naes and Skjeltorp (2006)'s definition of order book slope. They document a negative relation between future volatility and order book slope. Finally, by using data from the automated futures market, Coppejans, Domowitz and Madhavan (2001) study the dynamic relation between liquidity, return and volatility in a vector autoregressive framework. Consistent with the aforementioned studies, they find a negative relationship between liquidity and future volatility. We contribute to this literature in two ways: first, we extract a new measure from the limit order book, and second, we study the relationship between *market* liquidity and future *market* volatility.

2.3 The Market and Data

Our dataset comprises order and trade books of the individual constituents of the Istanbul Stock Exchange (ISE)–30 index for the period of June and July 2008.[5] The index corresponds to almost 75% of the total trading volume of the ISE for the sample period.

The ISE is a fully computerized as well as a fully centralized stock exchange, i.e., the trading of the listed stocks has to be executed in the ISE via electronic order submissions. Hence, our data fully captures the order flow. The information of a new order arrival or execution is updated instantaneously on traders' screens. All brokers are directly connected to the ISE system and have access to the full book. Prior to the submission of an order, they can see the quantity available at different prices, not limited to the best five or ten quotes.

The trading occurs between 09:30am to 5:00pm, with a lunch break. There are two opening call auctions: one for the morning session and one for the afternoon session. In contrast to the opening sessions, during the continuous double auction all of the orders submitted are either matched instantaneously, or booked until the corresponding match order arrives to the system based on the usual price and time

[5] We sincerely thank Recep Bildik, Ozkan Cevik, Ulkem Basdas, and Huseyin Eskici from Istanbul Stock Exchange for providing us the data and support for understanding the market mechanisms.

priorities.

All of the orders include the price and the quantity. Trade occurs if a matching order is submitted on the opposite side of the book. If an order is not fully executed, then the excess is converted into a limit order at the corresponding price instead of walking through the book.[6] Moreover, there are no hidden orders; the price and the quantity of all orders are fully displayed.

Order book data consists of information regarding the orders submitted for a given stock and date, whereas trade data records the executed orders. The order and trade ID numbers generated by the exchange system allow us to match orders in these two books and track any order through submission to (possible) execution or modification. Samples of the order and transaction data sets are presented in Tables A.1 and A.2 in Appendix A. By using the order and trade books, we first reconstruct the limit order book dynamically for each stock and obtain relevant information, such as the bid and ask prices and corresponding volumes at a given time. Hence, the reconstruction methodology enables us to obtain snapshots of a limit order book at any given time. In particular, we have the same information that a trader observes: the volume of orders waiting to be executed for the entire price range. We use this information to calculate the relative frequency of orders waiting in every price level. Sample of a constructed limit order book data is presented in Appendix A, Table A.3. To conserve space, only the information up to the 10 best prices is presented.

Table 2.1 presents the descriptive statistics for 30 stocks in our sample. We report the time-series averages of all the figures, except the market capitalization, for which the value at the beginning of the sample is presented.

The results reveal that one of the biggest stocks in our sample, GARAN, is 40 times more actively traded than the smallest stock, MIGRS. On average, the maximum

[6] This is similar to the Australian Stock Exchange, the Sao Paulo Stock Exchange (Bovespa), and the Stock Exchange of Hong Kong, for example.

Table 2.1: Descriptive Statistics for Each Stock
The table reports the summary statistics of ISE–30 stocks for June-July 2008. The first column presents the ticker of the securities in our sample. The market capitalization is the value at beginning of the sample period in million Turkish Liras (M TRY). Number of Orders (Trades) is the average of the total number of orders (trades) in a day. Ave. Trade Size is the daily average size of trades in number of shares. Spread is the tick-adjusted difference between the best ask and the best bid. Finally the last two columns report the average of the daily percentage of buy orders and limit orders, respectively.

	Mcap	**Number of Orders**	**Number of Trades**	**Ave. Trade Size**	**spread (tick adj.)**	**%Buy**	**%LO**
AKBNK	16650	2,609	1,643	5,376	1.04	46.79	68.56
AKGRT	1463	1,044	714	2,007	1.15	52.13	62.16
ARCLK	1664	1,003	576	1,234	1.25	45.50	71.04
ASYAB	1980	1,392	954	2,168	1.14	49.20	62.10
DOHOL	2160	2,438	1,546	7,676	1.06	44.11	68.74
DYHOL	1082	2,991	1,949	4,706	1.06	48.77	65.93
EREGL	9995	2,286	1,455	1,495	1.08	48.71	67.76
GARAN	14448	9,259	6,186	13,015	1.02	47.46	69.78
GSDHO	277	2,074	1,400	7,336	1.05	47.48	64.22
HALKB	7750	1,656	972	2,506	1.10	46.46	71.57
HURGZ	745	2,281	1,455	5,695	1.10	47.05	67.16
IHLAS	202	1,975	942	7,596	1.01	47.64	70.75
ISCTR	13165	7,332	4,732	6,777	1.03	49.48	69.81
ISGYO	459	700	367	3,448	1.11	44.94	71.81
KCHOL	7629	1,399	855	4,542	1.11	45.17	68.76
KRDMD	670	2,016	1,150	8,376	1.05	45.80	70.28
MIGRS	3614	346	152	3,040	1.03	38.90	70.28
PETKM	1024	1,156	688	1,537	1.14	46.81	70.54
PTOFS	2778	507	295	1,541	1.38	45.80	69.47
SAHOL	8676	1,103	713	3,076	1.15	48.54	66.25
SISE	1439	1,572	975	3,189	1.08	51.39	67.02
SKBNK	876	1,872	1,216	2,579	1.15	44.15	64.36
TCELL	17050	1,847	1,095	4,569	1.10	46.47	71.25
THYAO	919	1,252	787	2,040	1.10	50.52	68.10
TKFNK	2166	1,172	747	1,227	1.13	48.63	64.70
TSKB	490	707	448	3,840	1.06	48.98	63.23
TTKOM	14350	4,447	2,343	3,527	1.05	39.22	73.20
TUPRS	7387	1,413	761	1,036	1.07	48.45	73.68
VAKBN	4400	4,813	3,169	9,533	1.04	47.42	68.53
YKBNK	9999	2,939	1,911	7,562	1.04	48.33	67.08
Average	5184	2253	1406	4408	1.10	47.01	68.27
Median	2163	1752	973	3487	1.08	47.44	68.65
Min	202	346	152	1036	1.01	38.90	62.10
Max	17050	9259	6186	13015	1.38	52.13	73.68

trade size is over 13,000 units, with a median of 3,500 units. In terms of the number of orders submitted, GARAN is 5 times larger than the median, whereas MIGRS is 5 times smaller. The bid-ask spread is presented in column V. The results show that the inside spread of the ISE–30 constituents is narrow, with a tick-adjusted maximum of 1.38. Finally, about 68% of the submitted orders are limit orders and on average, the number of buy and sell orders are almost balanced.

2.4 Global Depth and the Limit Order Book Distribution

To evaluate the role of liquidity on future volatility, we first need a measure that summarizes the state of a given limit order book. We want our measure to capture the relative depth provision in the whole book to account for the liquidity supply beyond the inside quotes. Intuitively, one needs to consider the whole book, not only the information contained in the best quotes, because both price impact and execution probability of an order could depend on the depth beyond the best quotes.[7]

Latza and Payne (2013) investigate the forecasting power of market and limit order flows on high-frequency stock returns on a sample of traded stocks from the London Stock Exchange SETS system. They define the limit order flow as the difference between the weighted sums of the buy and sell limit order shares. The declining weights associated with each limit order capture the price positioning, hence the aggressiveness of a new limit order entry. Moreover, the extant literature documents that the information provided farther away from the best quotes is less informative compared to that from quotes closer to the best prices. One possible reason is that the execution probability of an order is a decreasing function of the price distance.

Hence, while considering the execution probability–price trade-off, it is natural for a trader to give higher importance to the information around the best quotes.

[7]For example, the execution probability of a limit order submitted, say at the second best quotes, depends on the accumulated volume of orders waiting at the best and the second best quotes.

These arguments bring the second ingredient of our measure: assigning weights to the information provided in different quotes based on price distances.

To construct our summary measure, global depth, we first consider the distribution of orders within different tick sizes along with their quoted volumes and calculate the limit order book probability density function (LOB–PDF). Second, we obtain the limit order book cumulative distribution function (LOB–CDF) by integrating the LOB–PDF over the different ranges of price distances. A stock's global depth is the weighted average of the cumulative distribution function of the limit order book.[8] Finally, the aggregate level of depth is approximated as the cross-sectional average of global depth measures of individual stocks.

2.4.1 Construction of global depth

We obtain the limit order book distribution and global depth by employing the following steps:[9]

1. For each security and each day, we sample the limit order books every 15 minutes, excluding the lunch break and the opening sessions.[10] The first snapshot of the book contains the unexecuted orders submitted until 10:00am, whereas the last one contains all of the unexecuted orders submitted until 17:00pm. Hereafter, the time subscript τ indexes these trading intervals.

2. We calculate the (tick-adjusted) price distance of each limit order relative to the best extant limit price at the end of each snapshot. In other words, for each

[8]One could easily find good arguments in favor of constructing global depth from the *probability* distribution function (LOB–PDF) instead of the the LOB–CDF. We repeat the analysis by using the LOB–PDF and obtain qualitatively similar but less strong results. Hence, the rest of our analysis depends on the measure calculated from the LOB–CDF.

[9]Appendix A, Section A.2, illustrates the steps with an example.

[10]We repeat the empirical analysis with 30–minute sampling frequencies. The results are presented in Section 2.6.5.

order i in the limit order book at τ, we define the price distance Δ as:

$$\Delta_{i,\tau}^{\text{buy}} = (p_\tau^B - p_i^{\text{buy}})/\text{tick},$$

$$\Delta_{i,\tau}^{\text{sell}} = (p_i^{\text{sell}} - p_\tau^A)/\text{tick},$$

where p_τ^B (p_τ^A) is the best bid (ask) price in interval τ and p_i^{buy} (p_i^{sell}) is the limit price of the i^{th} order.

3. For each side of the book, day, and snapshot, we get the LOB–PDF by calculating the percentage of total volume supplied/demanded at a given Δ for $\Delta = 0, 1, 2, .., \Delta_c$.[11] Therefore, LOB–PDF summarizes both the relative magnitude of the depth provision and its price location.

4. By integrating the LOB–PDF of the buy (sell) side over the ranges of Δ, i.e., by calculating the cumulative frequencies, we obtain the LOB–CDF of the buy side (sell).

5. We define a stock's global depth as the weighted average of the LOB–CDF. That is, for stock s and trading interval τ,

$$\text{GD}_{s,\tau}^{\text{buy}} = \sum_{\Delta=0}^{\Delta_c} F_{s,\tau}^{\text{buy}}(\Delta) g(\lambda, \Delta), \tag{2.1}$$

where $F_{s,\tau}^{\text{buy}}(\Delta)$ is the buy side cumulative distribution function and $g(\lambda, \Delta)$ are the weights with

$$1 = \sum_{\Delta=0}^{\Delta_c} g(\lambda, \Delta)$$

for a constant decay parameter λ. A stock's global depth of the sell side is constructed analogously. Throughout the paper, we assume the following expo-

[11] To capture the whole book without missing any orders submitted farther away from the best quotes, we set $\Delta_c = 30$.

nential decaying weight function:[12]

$$g(\lambda, \Delta) = \frac{\exp(-\lambda\Delta)}{\sum_{\Delta=0}^{\Delta_c} \exp(-\lambda\Delta)}. \tag{2.2}$$

6. $g(\lambda, \Delta)$ is a non-linear function of the decay parameter λ, which can be exogenously given or estimated within a regression model. We obtain the "optimal" decay parameter by employing a non-linear panel regression of the form:

$$\sigma_{s,\tau+1} = b_0 + b_1\sigma_{s,\tau} + b_2 \sum_{\Delta=0}^{\Delta_c} F_{s,\tau}^{\text{buy}}(\Delta) g(\lambda, \Delta) + b_3 \sum_{\Delta=0}^{\Delta_c} F_{s,\tau}^{\text{sell}}(\Delta) g(\lambda, \Delta) \tag{2.3}$$
$$+ \sum_{k=1}^{20} b_k T_{k,\tau} + \sum_{s=1}^{30} c_s D_s + \varepsilon_{s,\tau},$$

where, for a given stock s in a trading interval τ, $\sigma_{s,\tau}$ is the mid-quote volatility, $F_{s,\tau}^{\text{buy}}$ ($F_{s,\tau}^{\text{sell}}$) is the cumulative limit order book distribution function for the buy (sell) side of the market, $g(\lambda, \Delta)$ is the weight function previously defined in equation (2.2), $T_{k,\tau}$ is the intraday dummy that equals to 1 if $k = \tau$, and finally, D_s are stock-specific dummy variables allowing for stock fixed effects.

7. For each stock s and interval τ, we evaluate global depth at the optimal decay parameter $\widehat{\lambda}$ and calculate $\text{GD}_{s,\tau}(\widehat{\lambda})$, as introduced in (2.1). Finally, the aggregated global depth measure is the cross-sectional average of individual stock global depth measures.

Global depth is the convolution of two functions: the LOB–CDF and the weight function. It is size-related and goes beyond the inside quotes. It aggregates all of the orders waiting on a given side of the market, and focuses on how the available liquidity is distributed across price levels. Thus, it provides a more complete picture of liquidity. It gives the flexibility of assigning different weights to different quotes based on price distances.[13]

[12]As a robustness, we use different weight functions. The discussion is presented in Section 2.6.5.
[13]Note that by setting $\lambda = 0$ one can assume equal weights for each of the quotes.

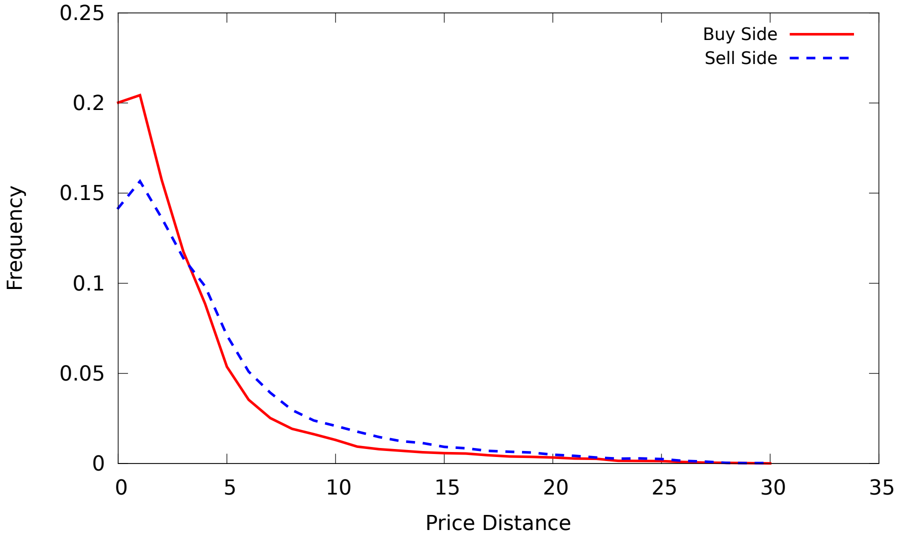
By definition, global depth is related to the standard "local" depth measures, i.e., the quoted depth up to a given threshold. An investigation of their relationship is presented in Appendix A, Section A.3. From this analysis, we conclude that although they are positively and significantly correlated, there is a non-trivial proportion of the variation of global depth that cannot be explained by the standard depth measures. Hence, global depth captures different information than that of standard depth measures.

2.4.2 Descriptive analysis

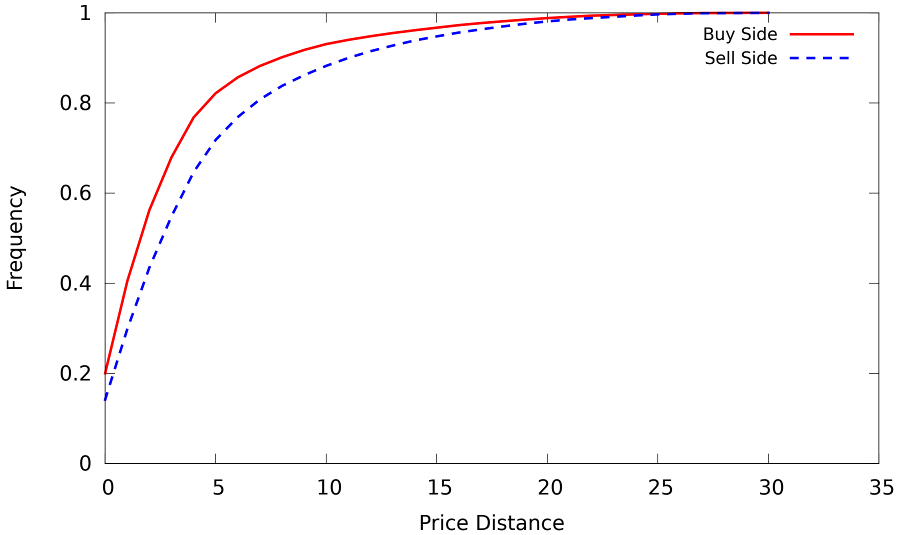
Figure 2.1, Panel A plots the limit order book probability density function (LOB–PDF) averaged across intervals, days, and stocks (average LOB–PDF), whereas Panel B plots the corresponding cumulative distribution (LOB–CDF). Panel A reveals that for both sides of the market, the frequency of the orders waiting at the second best quotes is the highest and the limit order book distribution is positively skewed. Similar to the findings of Bouchaud, Mezard and Potters (2002) for the analysis conducted on three stocks traded in Paris Bourse, the empirical densities of price distance Δ have a gamma-like shape.

Table 2.2 presents the descriptive statistics of the limit order book distribution. The first column reports the summary statistics of the average LOB–PDF. The last four columns report the statistics for four limit order book distributions at 10:00am (beginning of the day), 12:00pm (end of the morning session), 14:15pm (beginning of the afternoon session) and 17:00pm (end of the trading day).

The results reveal that the liquidity provision is concentrated closer to the best quotes for the buy side compared to the sell side, which can be observed by comparing either the mean or the skewness of the distribution. The mean of the distribution, for all of the time intervals, is higher for the sell side than the buy side, whereas the ranking is the opposite for skewness. This asymmetry of the volume distribution



(a) Panel A: LOB–PDF



(b) Panel B: LOB–CDF

Figure 2.1: Panel A plots the limit order book probability density function (LOB–PDF), averaged across stocks and trading intervals. Panel B plots the corresponding cumulative distribution functions.

can also be concluded from the cumulative frequencies of volumes for different price distances Δ. Around 40% and 30% of the depth is concentrated at the best or second best quotes ($\Delta = 0$ or $\Delta = 1$) for buy and sell sides, respectively. The frequency of orders waiting 5 or more ticks away from the quotes is 35% for the sell side, whereas it is only 23% for the buy side. Finally, the average variance of the sell side is 36% higher than the average variance of the buy side, indicating that the buy side is less dispersed.

Table 2.2: Summary Statistics: The Limit Order Book Distribution
For both sides of the market, this table presents the descriptive statistics for the empirical limit order book distributions. The mean, variance, skewness, and the fractions of number of shares accumulated up to a given price distance Δ are reported. The first column shows the summary statistics of the limit order book distribution which is obtained by averaging across intervals, days, and stocks. The last four columns report the statistics for four limit order book distributions (averaged across stocks) at 10:00am (beginning of the day), 12:00pm (end of the morning session), 14:15pm (beginning of the afternoon session) and 17:00pm (end of the trading day).

		uncond.	**10:00am**	**12:00pm**	**14:15pm**	**17:00pm**
Buy side	mean	3.43	3.64	3.32	3.41	3.42
	variance	18.42	20.06	17.67	17.83	17.52
	skewness	2.41	2.34	2.60	2.33	2.35
	up to 1 Δ	0.40	0.38	0.40	0.41	0.41
	up to 3 Δ	0.68	0.66	0.69	0.68	0.68
	up to 5 Δ	0.82	0.81	0.84	0.82	0.82
	up to 10 Δ	0.93	0.92	0.94	0.93	0.93
	up to 20 Δ	0.99	0.99	0.99	0.99	0.99
	up to 30 Δ	1.00	1.00	1.00	1.00	1.00
Sell side	mean	4.63	4.68	4.64	4.56	4.73
	variance	25.16	27.51	25.77	23.73	24.20
	skewness	1.84	1.83	1.89	1.77	1.74
	up to 1 Δ	0.30	0.31	0.29	0.30	0.28
	up to 3 Δ	0.55	0.56	0.55	0.55	0.53
	up to 5 Δ	0.72	0.73	0.72	0.72	0.70
	up to 10 Δ	0.88	0.87	0.88	0.89	0.88
	up to 20 Δ	0.98	0.98	0.98	0.98	0.98
	up to 30 Δ	1.00	1.00	1.00	1.00	1.00

2.5 Predicting Market Volatility

Examining the relationship between future market volatility and aggregate liquidity at an intraday level is the aim of this section. To this end, after sampling each trading day into twenty-one 15-minute intervals, we first calculate our proposed measure, global depth, for each stock and each interval. We then use the cross-sectional average of individual stock global depths for buy and sell sides of the market as main explanatory variables. The market volatility is defined as the volatility of the Istanbul

Stock Exchange–30 index. We employ the following predictive regression model:

$$\begin{aligned}\sigma^M_{\tau+1} = a_0 + a_1\sigma^M_\tau + a_2\overline{\mathrm{GD}}^{\mathrm{buy}}_\tau + a_3\overline{\mathrm{GD}}^{\mathrm{sell}}_\tau + \sum_{k=1}^{20} b_k T_{k,\tau} \\ + \text{controls} + \varepsilon_\tau,\end{aligned} \tag{2.4}$$

where for a given interval τ, σ^M_τ is the mid-quote-volatility of the value-weighted index, and $\overline{\mathrm{GD}}^{\mathrm{buy}}_\tau$ and $\overline{\mathrm{GD}}^{\mathrm{sell}}_\tau$ are global depth for buy and sell sides of the market, respectively. $T_{k,\tau}$ is the intraday dummy that equals to 1 if $k = \tau$.

We include the lagged volatility, σ^M_τ, and interval dummies in the set of explanatory variables to control the well-known systematic intraday patterns and clustering in volatility. Furthermore, we employ both the standard predictors of volatility and other liquidity measures as control variables. Similar to the construction of $\overline{\mathrm{GD}}$, the control variables are calculated as the equal-weighted cross-sectional average of the individual stock measures.[14]

The coefficients of interest, a_2 and a_3, are expected to be negative; the higher liquidity provision around the best quotes, the lower the future volatility. The first possible link follows from the price impact of an order. If the liquidity provision is concentrated near the best quotes, i.e., when global depth is high, then the price impact of an order is lower, leading to smaller future short-term volatility. The second link arises from the dispersed beliefs, based on the theoretical predictions of Goettler et al. (2005) and Goettler et al. (2009). They show that an increase in the frequency of orders waiting away from the best prices signals that the current quotes are mispriced. Hence, an increase in the dispersed beliefs about the true price of an asset may make large price jumps plausible, which in turn creates higher future volatility.

[14] As a robustness check, we repeat the analysis by calculating the value-weighted average of the explanatory variables to proxy the aggregated measures. The results are presented in Section 2.6.5. Our main results are also confirmed in these regressions.

2.5.1 Measuring volatility: the two scales realized volatility estimator (TSRV)

To explore the role of relative depth provision in explaining the volatility of the *true* price process rather than the *noise* component, we calculate the return volatility by employing the two scales realized volatility (TSRV) estimator proposed by Ait-Sahalia et al. (2011).

Let X denote the fundamental log-stock price process. In financial data, instead, we can only observe log-price Y, either in a form of transaction or quoted price, which is a linear combination of X and some noise ϵ:

$$Y_t = X_t + \epsilon_t,$$

where ϵ is assumed to be independent of the X process for identification purposes and X follows a geometric Brownian motion. The noise may be a result of many microstructure effects: frictions inherent in the trading process, temporary liquidity withdrawals, and measurement or data recording errors. In this paper, the market microstructure noise is assumed to be i.i.d., however no additional distributional assumption is imposed. In other words, we adopt the nonparametric case and let the diffusion term be an unrestricted stochastic process (see Ait-Sahalia et al. (2011) for further details).

Without the noise, the realized variance estimator, $[Y,Y]_T^{(all)} = \sum_{i=1}^{n} (Y_{t_{i+1}} - Y_{t_i})^2$ is a consistent and asymptotically unbiased estimator of the quadratic variation of the process X, $\langle X, X\rangle_T = \int_0^T \sigma_t^2 dt$, where T is any fixed time interval. However, in the presence of the microstructure noise, Ait-Sahalia, Mykland and Zhang (2005) and Zhang, Mykland and Ait-Sahalia (2005) show that the realized volatility (RV) is no longer a consistent and unbiased estimator of the volatility of the true value of an asset. It leads to an estimate of the volatility of the noise, instead of the true price of the underlying asset. As a solution, Ait-Sahalia et al. (2011) propose the two scales realized volatility estimator (TSRV), which enables the use of the full available sample data, and gives an unbiased and consistent estimate of the quadratic variation of the true price process.

The TSRV is defined as:

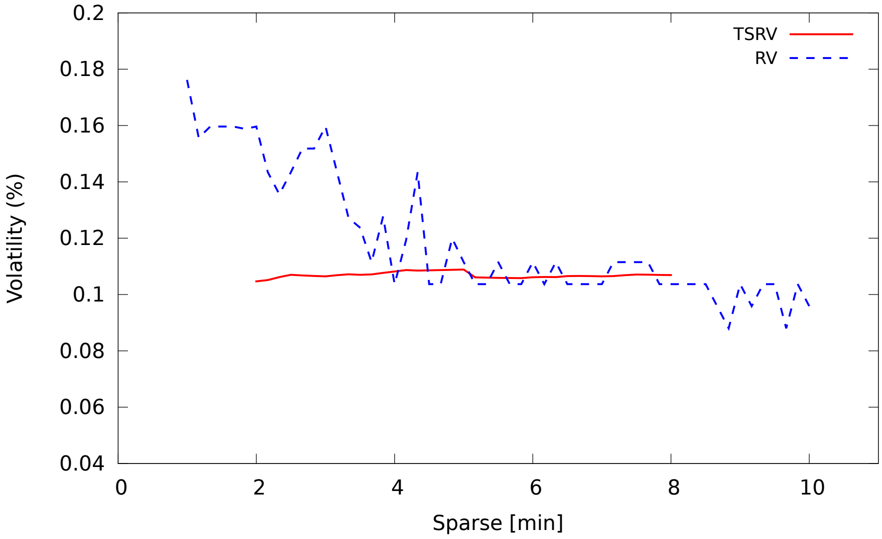
$$\langle X, X\rangle_T^{TSRV} = \sqrt{[Y,Y]_T^{ave} - \frac{1}{K}[Y,Y]_T^{(all)}}, \tag{2.5}$$

where $[Y,Y]_T^{(all)}$ is the realized variance calculated using the whole sample with size T and

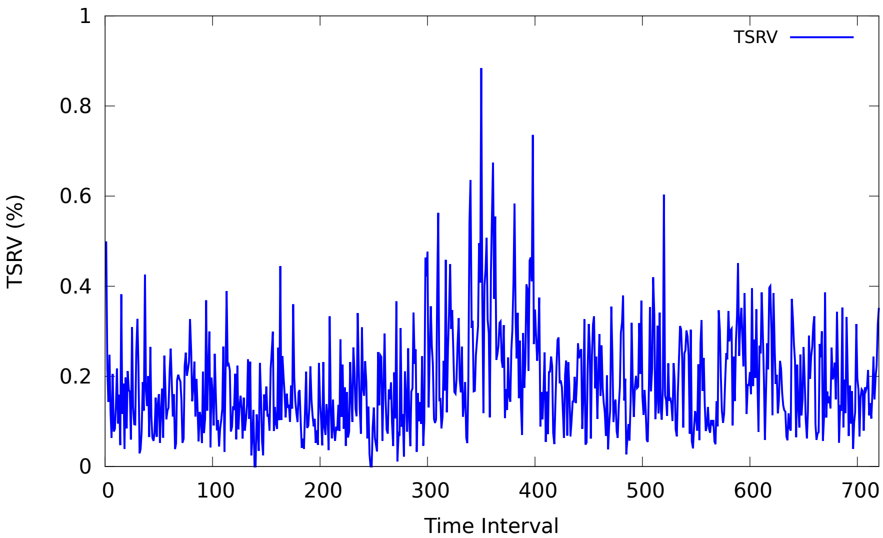
$$[Y,Y]_T^{ave} = \frac{1}{K}\sum_{k=1}^{K}[Y,Y]_T^{sparse,k}.$$

To obtain $[Y,Y]^{sparse,k}$, we first divide the whole sample into K number of moving window subsamples ($K = 5$ minutes) with a fixed length of N, where $N = T - K$. For example, the first subsample starts with the first and ends with the N^{th} observation, whereas, the second subsample starts with the second and ends with $(N+1)^{\text{th}}$ observation. Then, we sparse each subsample with one-minute frequency. So, $[Y,Y]^{sparse,k}$ is the realized variance estimator of the k^{th} one-minute-sparsed subsample of returns.

Figure 2.2, Panel A plots the RV and TSRV estimates of a stock in a day calculated for different sparse periods. Consistent with Ait-Sahalia et al. (2011), the TSRV is almost invariant to the choice of the sparse period, whereas the RV estimator changes dramatically, mainly due to noise embedded in the data. Panel B plots our dependent variable, the TSRV estimate of the mid-quotes for the value-weighted index calculated for each interval and day based on one-minute sparse periods (scaled by 100). There is substantial variability in the return volatility, with a standard deviation of 11%. Finally, the augmented Dickey-Fuller and Phillips-Perron tests suggest the stationarity of our dependent variable.



(a) Panel A: TSRV vs. RV



(b) Panel B: TSRV time-series

Figure 2.2: Panel A plots the realized volatility (RV) and the two scales realized volatility (TSRV) estimates calculated at different sparse periods. Panel B plots our dependent variable; the TSRV estimate of the mid-quotes for the value-weighted index calculated for each interval and day based on one-minute sparse periods (scaled by 100).

2.5.2 Estimation of the decay parameter

A stock's global depth is obtained by multiplying the cumulative limit order book distribution with a normalised weight function and then taking the area below the resulting curve. The weight function is a non-linear function of the decay parameter λ, which is estimated by using the first 3 days of data as a training period and running the non-linear regression model introduced in equation (2.3). The estimated decay parameter $\widehat{\lambda}$ is equal to 0.366, with a standard error of 0.173, suggesting a "moderate" decay on the information provided in each quotes.[15] Then, for the rest of the sample period, for each stock s and interval τ, we evaluate global depth at the optimal decay parameter $\widehat{\lambda}$, as introduced in equation (2.1), and calculate the cross-sectional average of $\mathrm{GD}_{s,\tau}(\widehat{\lambda})$.

Figure 2.3 presents the time-series plot of the aggregated optimal-decayed global depth measure for both sides of the market. We see that the depth provision around the best quotes on the buy side is higher compared to the sell side for most of the trading intervals, in line with the findings discussed in Section 2.4.2. These two variables are negatively correlated with a correlation coefficient of −25%. The average of global depth is 49% (40%), whereas it ranges from 30% (24%) to 62% (52%) for the buy (sell) side of the market. The augmented Dickey-Fuller and Phillips-Perron tests reject the unit-root in global depth variable for both sides of the market.

2.5.3 Control variables

Trade-related variables

Since both trading activity and volatility depend on the news arrival process, several studies have used trade-related variables to forecast price volatility. Consistent with Bollerslev and Domowitz (1993), Jones, Kaul and Lipson (1994), and Foucault, Moinas and Theissen (2007), the number of trades occurring in the interval τ, NT, and the average trade size, AQ, are included to capture the trading activity.

[15] Our empirical findings are robust to the different training periods chosen. We use 5 and 10 days of data as training period to estimate the decay parameter λ. The estimated parameter is equal to 0.304 and 0.335 when 5 and 10 days of data are used as training period, respectively. Hence, the optimal decay parameter does not change dramatically for different training periods employed.

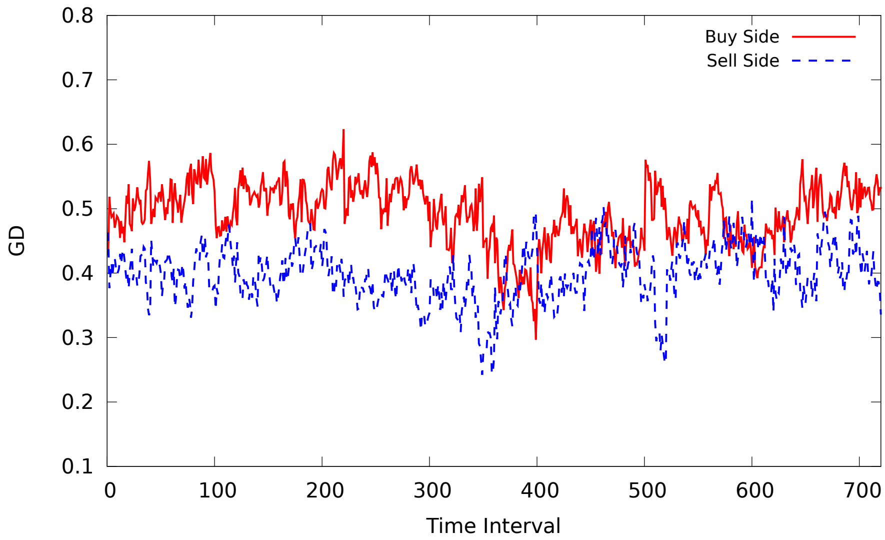


Figure 2.3: This figure plots the intraday estimates of global depth evaluated at the optimal decay parameter for buy and sell sides of the market. The estimation is based on the sampling of a trading day in 15-minute intervals.

Relative spread

In a related study, Foucault et al. (2007) show that the bid-ask spread is informative about future individual stock volatility. Hence, we also include the relative spread, relSPR_τ, which is calculated as the ratio of the bid-ask spread to the mid-quote prices for each interval.

Slope of the limit order book

Another measure extracted from the limit order book is "the slope of the order book" proposed by Naes and Skjeltorp (2006). The slope measures the sensitivity of the quantity supplied in the order book with respect to the prices. Furthermore, Duong and Kalev (2008) document evidence for the predictive power of the order book slope over price volatility. Following these studies, we consider the SLOPE as an explanatory variable, which is defined as:

$$\text{SLOPE}_{s,\tau} = \frac{DE_{s,\tau} + SE_{s,\tau}}{2}, \tag{2.6}$$

where $DE_{s,\tau}$ and $SE_{s,\tau}$ denote the slope for bid and ask sides, respectively, for stock s and interval τ, and calculated as follows:

$$DE_{s,\tau} = \frac{1}{N_B}\left[\frac{\nu_1^B}{p_1^B/p_0 - 1} + \sum_{k=1}^{N_B-1} \frac{\nu_{k+1}^B/\nu_k^B - 1}{|p_{k+1}^B/p_k^B - 1|}\right],$$

$$SE_{s,\tau} = \frac{1}{N_A}\left[\frac{\nu_1^A}{p_1^A/p_0 - 1} + \sum_{k=1}^{N_A-1} \frac{\nu_{k+1}^A/\nu_k^A - 1}{|p_{k+1}^A/p_k^A - 1|}\right],$$

where N_B (N_A) denotes the total number of bid (ask) prices. p_k is the quote at the tick level k. p_0 corresponds to the mid-quote at the end of interval τ. Finally, ν_t^B (ν_k^A) is the natural logarithm of the accumulated total volume up to the price level p_k^B (p_k^A).

In harmony with the findings of Duong and Kalev (2008), we expect the slope to be negatively related to the future volatility. The steeper the slope, the more concentrated the volumes in the order book are in a given time interval.

Standard depth measures

The "local" depth, defined as the total volume available to be traded at the best bid or ask prices, is one of the traditional measures of liquidity. We calculate DEPTH0$^{\text{buy}}$ and DEPTH0$^{\text{sell}}$ for the buy and sell sides of the market respectively.

Recent theoretical and empirical studies document that the volume at and farther away from the best quotes have a different impact on the order choice of a trader (see Goettler et al. (2005), Goettler et al. (2009), Cao et al. (2008), and Valenzuela and Zer (2013), among others). Moreover, Pascual and Veredas (2010) document that both at and away from the best quotes are informative about future individual stock volatilities. Hence, to capture the volume available beyond the best quotes, we include the cumulative depth from the second up to the five best quotes for the buy (DEPTH1_5$^{\text{buy}}$) and the sell (DEPTH1_5$^{\text{sell}}$) sides of the market in our analysis.

Amihud illiquidity measure

We employ the Amihud (2002)'s illiquidity measure, AMR, which is the ratio of absolute stock return to the turnover. For stock s and interval τ, it is calculated as

$$\mathrm{AMR}_{s,\tau} = \frac{|r_\tau|}{\sum_{i=1}^{\mathrm{NT}_\tau} p_i \cdot q_i}, \tag{2.7}$$

where NT is the number of trades in interval τ, r_τ is the return on mid-quotes between intervals τ and $\tau - 1$, q_i is the number of shares traded and p_i is the corresponding trade price for trade i.

Quote-slope

We include the log quote slope, logQS, introduced by Hasbrouck and Seppi (2001). The logQS aggregates the tightness and depth dimensions of liquidity. For each time interval τ, logQS is defined as follows:

$$\mathrm{logQS}_{s,\tau} = \frac{\ln \frac{p^A_\tau}{p^B_\tau}}{\ln\left(q^A_\tau \cdot q^B_\tau\right)}, \tag{2.8}$$

where q^A and q^B are the volume of orders waiting at the best ask price p^A and the best bid prices p^B, respectively. A decrease in the logQS means that the slope of the best quotes is flatter and the market is more liquid.

Domowitz-Hansch-Wang illiquidity measure

Finally, we consider the illiquidity measure proposed by Domowitz et al. (2005), DHW. This variable measures the cost of buying and selling Q shares of the stock, simultaneously. The higher the cost, the more illiquid the stock. For each time interval τ, DHW is calculated as follows,

$$\begin{aligned}\mathrm{DHW}_{s,\tau} = &\left[\sum_{k=1}^{m-1} q^A_k p^A_k + \left(Q_s - \sum_{k=1}^{m-1} q^A_k\right) p^A_m\right] \\ &- \left[\sum_{k=1}^{m'-1} q^B_k p^B_k + \left(Q_s - \sum_{k=1}^{m'-1} q^B_k\right) p^B_{m'}\right],\end{aligned} \tag{2.9}$$

where q_k^A and q_k^B are the volume of orders waiting at the k^{th} best ask price p_k^A and the k^{th} best bid price p_k^B, respectively. m and m' denote the position in which the last sell and buy orders are executed. Finally, for each stock s, Q_s is the median of the accumulated volume of orders waiting in the book.

2.6 Empirical Findings

As a natural first step in our empirical analysis, we compare the explanatory power of the optimal-decay-weighted global depth (GD evaluated at $\lambda = \widehat{\lambda}$), equal-weighted global depth (GD evaluated at $\lambda = 0$) and the standard "local" depth measures that take into account the depth provision up to a given threshold. We further investigate the in-sample predictive power of global depth on volatility by adding standard predictors of volatility and other liquidity measures in our analysis. Section 2.6.1 reports the results. Section 2.6.2 asks whether the global depth-volatility relationship holds for further horizons.

Our findings are based on regressions of the market volatility on lagged global depth measures and different sets of control variables. Market volatility is calculated as the two scales realized volatility of the mid-quote return of the value-weighted index. All of the specifications use 21 trading intervals on 36 days and include intraday dummies. To conserve space, we do not report the estimated coefficients of the dummy variables. t-statistics are calculated using Newey-West standard errors to capture possible autocorrelation in the residuals.

Finally, in Section 2.6.3, we examine whether the documented time-series relation between global depth and future market volatility is driven by a variation in a particular stock or industry. To this end, we shift our focus to the relation between the *individual* stock volatility and liquidity. We first run the regression model in a pooled data with stock fixed effects. t-statistics are based on cluster robust standard errors on stock level. The interval and stock dummies are jointly significant, but for the sake of brevity they are not reported. To take into account possible cross-sectional variations that cannot be captured by the stock fixed effects, we also run the predictive regressions for each of the stocks in our sample and report the summary of the individual regression results.

The discussion of the results is based on the estimated coefficients, their statistical significance, and the adjusted R^2s. To improve the ease of interpretation of the estimated coefficients, all of the explanatory variables are standardized to have a unit variance, and the dependent variable is presented in percentage terms.

2.6.1 One-period-ahead predictability regressions

Our first focus is to examine the predictive power of the optimal-decay-weighted and equal-weighted global depth measures. $\mathrm{GD}_\tau(\widehat{\lambda})$ is global depth evaluated at the optimal decay factor $\widehat{\lambda}$ and assigns exponential weights to the quotes based on price distances, as introduced in Section 2.4.1, whereas $\mathrm{GD}_\tau(\lambda = 0)$ is global depth evaluated at a decay factor 0, i.e., it assigns equal weights to each quotes. The dependent variable is the 15-minute-ahead market volatility, $\sigma^M_{\tau+1}$. Table 2.3 reports the results.

The results show a strong predictive power of global depth for both sides of the market over the one-period-ahead market volatility. Irrespective of the chosen decay factor λ, an increase in the average liquidity around the best quotes is followed by a lower level of volatility in the next period. The explanatory power of global depth evaluated at the optimal decay factor is higher compared to the one that assigns equal weights to each quote. This confirms that depth closer to the best quotes is more informative.

For all of the specifications, the economic importance of the buy side is higher than the sell side. This asymmetry is consistent with the extant literature, documenting that buy orders are more informative than sell orders (see, for instance, Burdett and

Table 2.3: Predictive Regressions–Global vs. Local Depth
This table reports the estimated coefficients of the regression model defined in equation (2.4). The dependent variable is the 15-minute-ahead market volatility, $\sigma^M_{\tau+1}$, calculated as the TSRV mid-quote volatility of the value-weighted index (multiplied by 100). $\overline{\text{GD}}^{\text{buy}}_{\tau}(\widehat{\lambda})$ and $\overline{\text{GD}}^{\text{buy}}_{\tau}(\lambda = 0)$ are the cross-sectional averages of global depth of individual stocks, $\text{GD}^{\text{buy}}_{s,\tau}$, evaluated at the optimal decay factor $\widehat{\lambda}$ and $\lambda = 0$, respectively, as outlined in Section 2.5.2. $\overline{\text{DEPTH0}}^{\text{buy}}_{\tau}$ ($\overline{\text{DEPTH0}}^{\text{sell}}_{\tau}$) is the cross-sectional average of volume at the best quotes for the buy (sell) side of the market, whereas $\overline{\text{DEPTH1\_5}}^{\text{buy}}_{\tau}$ and $\overline{\text{DEPTH1\_5}}^{\text{sell}}_{\tau}$ are the accumulated volume of orders from the second to the 5$^{\text{th}}$ best quotes for the buy and sell sides of the market, respectively. All of the explanatory variables are standardized. t-statistics are calculated using Newey-West standard errors to capture possible autocorrelation in the residuals and reported in parenthesis. For the sake of brevity, the estimated coefficients of the intraday dummies are omitted.

dep. var.: $\sigma^M_{\tau+15\text{min}}$	**I**	**II**	**III**	**IV**	**V**	**VI**
$\overline{\text{GD}}^{\text{buy}}_{\tau}(\widehat{\lambda})$			-0.034		-0.033	-0.036
			(-6.82)		(-6.51)	(-6.75)
$\overline{\text{GD}}^{\text{sell}}_{\tau}(\widehat{\lambda})$			-0.021		-0.020	-0.013
			(-3.51)		(-2.50)	(-1.41)
$\overline{\text{GD}}^{\text{buy}}_{\tau}(\lambda = 0)$		-0.026				
		(-5.49)				
$\overline{\text{GD}}^{\text{sell}}_{\tau}(\lambda = 0)$		-0.013				
		(-1.85)				
$\overline{\text{DEPTH0}}^{\text{buy}}_{\tau}$				-0.013	0.004	-0.010
				(-2.39)	(0.08)	(-1.14)
$\overline{\text{DEPTH0}}^{\text{sell}}_{\tau}$				-0.011	-0.002	0.002
				(-2.41)	(-0.54)	(0.40)
$\overline{\text{DEPTH1\_5}}^{\text{buy}}_{\tau}$						-0.019
						(-1.86)
$\overline{\text{DEPTH1\_5}}^{\text{sell}}_{\tau}$						0.001
						(0.19)
σ^M_{τ}	0.038	0.027	0.023	0.032	0.023	0.023
	(6.82)	(5.35)	(4.96)	(6.70)	(4.97)	(4.67)
constant	0.179	1.707	0.749	0.292	0.734	0.741
	(7.94)	(5.35)	(8.07)	(7.46)	(7.56)	(7.40)
adj. R^2(%)	16.94	22.87	24.62	19.65	24.44	24.79

O'Hara (1987), Griffiths, Smith, Turnbull and White (2000), and Duong and Kalev (2008), among others).

Second, the correlation between global depth and local depth measures reported in Appendix A may indicate that these variables share common information on future volatility. Hence, it is important to examine whether global depth is still significant in explaining subsequent volatility under the presence of standard depth variables.

To this end, we include both the volume of orders at the best quotes and total volume of orders from the second to the fifth best prices in our analysis. Table 2.3 columns IV-VI present the estimated coefficients and the corresponding t-statistics.

$\text{DEPTH0}^{\text{buy}}$ and $\text{DEPTH0}^{\text{sell}}$, the total volume of orders waiting at the best bid and ask prices, respectively, significantly explain the future market volatility. As expected, a decrease in the volume of orders at the best quotes creates higher future volatility. However, when global depth variables are included in the analysis, they are no longer significant. Finally, we see that including global depth to the regression significantly increases the adjusted R^2 from 16.9% to 24.6%, whereas including all of the local depth variables together with GD does not add any explanatory power. The adjusted R^2 increases slightly to 24.8%.

Overall, we conclude that the exponentially-weighted global depth has a superior in-sample predictive power compared to the standard depth measures and compared to global depth that assigns equal weights.

To confirm the robustness of the explanatory power of global depth on one-period-ahead market volatility, which is documented in “simple” regressions, we include several other control variables. Table 2.4 presents the estimated coefficients and the corresponding t-statistics.

The results reveal that global depth for the buy side strongly predicts market volatility. This result is remarkably robust to the inclusion of alternative liquidity measures and standard predictors of volatility. Besides global depth variables, the relative spread and the slope of the book are both positively and significantly correlated with the future volatility.

This result further extends the findings of Foucault et al. (2007), who document that the relative spread has explanatory power over future individual stock volatilities. We show that the aggregated measure has an explanatory power on the market volatility as well. Yet, the estimated (standardized) coefficients show that our ag-

Table 2.4: Predictive Regressions–Control Variables
This table reports the estimated coefficients of the regression model defined in equation (2.4). The dependent variable is the 15-minute-ahead market volatility, $\sigma^M_{\tau+1}$, calculated as the TSRV mid-quote volatility of the value-weighted index (multiplied by 100). $\overline{\mathrm{GD}}^{\mathrm{buy}}_{\tau}(\widehat{\lambda})$ is the cross-sectional average of global depth of individual stocks, $\mathrm{GD}^{\mathrm{buy}}_{s,\tau}$, evaluated at the optimal decay factor $\widehat{\lambda}$, as outlined in Section 2.5.2. All of the control variables are constructed analogously. SLOPE is the slope of the limit order book defined in equation (2.6), relSPR is the relative spread, NT is the number of trades and AQ is the average trade size. AMR is the Amihud (2002) illiquidity measure. The logQS is the log quote slope, introduced by Hasbrouck and Seppi (2001) and defined in equation (2.8). Finally, DHW is the Domowitz et al. (2005) illiquidity measure described in equation (2.9). All of the explanatory variables are standardized. t-statistics are calculated using Newey-West standard errors to capture possible autocorrelation in the residuals and reported in parenthesis. For the sake of brevity, the estimated coefficients of the intraday dummies are omitted.

dep. var.: $\sigma^M_{\tau+15\mathrm{min}}$	**I**	**II**	**III**	**IV**	**V**
$\overline{\mathrm{GD}}^{\mathrm{buy}}_{\tau}(\widehat{\lambda})$		-0.034	-0.030	-0.030	-0.028
		(-6.82)	(-6.83)	(-5.44)	(-5.23)
$\overline{\mathrm{GD}}^{\mathrm{sell}}_{\tau}(\widehat{\lambda})$		-0.021	-0.017	-0.011	-0.011
		(-3.51)	(-3.14)	(-1.50)	(-1.56)
$\overline{\mathrm{SLOPE}}_{\tau}$			0.013	0.017	0.015
			(1.87)	(2.79)	(2.26)
$\overline{\mathrm{relSPR}}_{\tau}$			0.028	0.022	0.021
			(5.41)	(2.28)	(2.12)
$\overline{\mathrm{NT}}_{\tau}$			0.008		0.008
			(1.34)		(1.24)
$\overline{\mathrm{AQ}}_{\tau}$			-1.17		0.001
			(-0.02)		(0.28)
$\overline{\mathrm{AMR}}_{\tau}$				0.002	0.002
				(0.47)	(0.58)
$\overline{\mathrm{logQS}}_{\tau}$				0.014	0.014
				(0.94)	(0.96)
$\overline{\mathrm{DHW}}_{\tau}$				0.003	0.003
				(0.62)	(0.60)
σ^M_{τ}	0.038	0.023	0.013	0.015	0.011
	(6.82)	(4.96)	(2.51)	(3.30)	(2.19)
constant	0.179	0.749	-0.796	-0.947	-0.889
	(7.94)	(8.07)	(-2.54)	(-3.14)	(-2.87)
adj. $R^2(\%)$	16.94	24.62	28.81	28.88	28.95

gregated global depth measure is both economically and statistically the strongest in explaining the variations in the market volatility.

The estimated coefficient of the slope has an unexpected sign. Naes and Skjeltorp (2006) and Duong and Kalev (2008) document that the slope is negatively related to the volatility. If the volume of orders is more concentrated in a given price, then the book has a higher slope, signaling the consensus about the true price. Therefore, a higher slope should be followed by lower future volatility. To investigate this further, we run the slope in a simple regression and see that it is negatively and significantly correlated with the future volatility at a 5% level as expected (not reported). Thus, we conclude that controling other liquidity measures changes the sign of the slope. This indicates that the relationship between the slope and volatility is not robust to the inclusion of other liquidity measures. Finally, we note that the adjusted R^2 increases significantly from 17% to 25% with the inclusion of GD variables, whereas we see a slight increase with the inclusion of further controls.

2.6.2 Predicting further horizons

In this section, we examine the informativeness of the limit order book distribution at time τ on multiple-period-ahead volatilities. Specifically, we run the same baseline regression model specified in equation (2.4), while we calculate the dependent variable as the mid-quote volatility of the index at time $\tau + h$, with $h = 1, 2, ..., 10$, where for example, $\tau + 2$ refers to the 30-minute-ahead volatility. The regression results are presented in Table 2.5.

In Panel A we report the "simple" regressions, whereas Panel B reports the results when all of the control variables are included in the regression equation. We see that the significance of the estimated coefficients as well as the predictive power of global depth is (almost) monotonically decreasing with the prediction horizon. Global depth is significant for all of the horizons, suggesting that the limit order book distribution is informative over the 150-minute-ahead volatility. When we add the other control variables, we see that the relative spread and the slope of the book significantly predicts longer term volatility as well.

Finally, although the illiquidity measure proposed by Hasbrouck and Seppi (2001), the quote-slope, does not significantly explain the 15-minute-ahead volatility, the relationship is significant for further horizons (up to 75 minutes ahead). Again, global

Table 2.5: Predictive Regressions–Further Horizons

This table reports the estimated coefficients of the regression model defined in equation (2.4). The dependent variable is the market volatility, $\sigma^M_{\tau+h}$ calculated as the TSRV mid-quote volatility of the value-weighted index (multiplied by 100) in period $\tau+h$ for $h = 1, 2, ..., 6$. $\overline{\text{GD}}^{\text{buy}}_{\tau}$ is the cross-sectional average of global depth of individual stocks, $\text{GD}^{\text{buy}}_{s,\tau}$, evaluated at the optimal decay factor $\widehat{\lambda}$, as outlined in Section 2.5.2. All of the control variables are constructed analogously. SLOPE is the slope of the limit order book defined in equation (2.6), relSPR is the relative spread, NT is the number of trades and AQ is the average trade size. AMR is the Amihud (2002) illiquidity measure. The logQS is the log quote slope, introduced by Hasbrouck and Seppi (2001) and defined in equation (2.8). Finally, DHW is the Domowitz et al. (2005) illiquidity measure described in equation (2.9). In Panel A for every time horizon, we report the "simple" regressions, whereas Panel B reports the results when all of the control variables are included in the regression equation. All of the explanatory variables are standardized. t-statistics are calculated using Newey-West standard errors to capture possible autocorrelation in the residuals and reported in parenthesis. For the sake of brevity, the estimated coefficients of the intraday dummies are omitted.

dep. var.: $\sigma^M_{\tau+h}$	**Panel A: "simple" regressions**							**Panel B: multiple regressions**						
	0–15	**15–30**	**30–45**	**.**	**105–120**	**120–135**	**135–150**	**0–15**	**15–30**	**30–45**	**.**	**105–120**	**120–135**	**135–150**
$\overline{\text{GD}}^{\text{buy}}_{\tau}(\widehat{\lambda})$	-0.034	-0.029	-0.027	.	-0.028	-0.027	-0.025	-0.028	-0.023	-0.023	.	-0.024	-0.025	-0.025
	(-6.82)	(-5.55)	(-4.60)	.	(-3.29)	(-3.30)	(-2.63)	(-5.23)	(-5.17)	(-4.29)	.	(-3.01)	(-2.93)	(-2.91)
$\overline{\text{GD}}^{\text{sell}}_{\tau}(\widehat{\lambda})$	-0.021	-0.023	-0.022	.	-0.020	-0.017	-0.016	-0.011	-0.011	-0.009	.	-0.002	0.002	-0.006
	(-3.51)	(-3.46)	(-3.07)	.	(-2.24)	(-1.93)	(-1.85)	(-1.56)	(-1.94)	(-1.34)	.	(-0.23)	(0.17)	(-0.60)
$\overline{\text{SLOPE}}_{\tau}$								0.015	0.023	0.020	.	0.017	0.011	0.010
								(2.26)	(2.69)	(2.33)	.	(1.82)	(1.31)	(1.01)
$\overline{\text{relSPR}}_{\tau}$								0.021	0.019	0.021	.	0.023	0.033	0.049
								(2.12)	(2.86)	(2.89)	.	(2.47)	(2.98)	(3.43)
$\overline{\text{NT}}_{\tau}$								0.008	0.013	0.006	.	0.010	0.005	0.009
								(1.24)	(1.86)	(1.01)	.	(1.09)	(0.45)	(0.94)
$\overline{\text{AQ}}_{\tau}$								0.001	0.003	0.003	.	0.006	-0.002	-0.013
								(0.28)	(0.68)	(0.75)	.	(1.15)	(-0.35)	(-2.13)
$\overline{\text{AMR}}_{\tau}$								0.002	0.009	-0.001	.	0.002	0.001	0.000
								(0.58)	(7.56)	(-1.12)	.	(0.36)	(0.27)	(-0.04)
$\overline{\text{logQS}}_{\tau}$								0.014	0.027	0.022	.	0.018	0.001	-0.023
								(0.96)	(2.80)	(2.40)	.	(1.62)	(0.09)	(-1.43)
$\overline{\text{DHW}}_{\tau}$								0.003	0.002	0.005	.	0.015	0.018	0.013
								(0.60)	(0.52)	(1.00)	.	(1.49)	(1.86)	(1.51)
σ^M_{τ}	0.023	0.016	0.012	.	0.000	0.004	0.010	0.011	-0.001	-0.002	.	-0.017	-0.006	0.004
	(4.96)	(3.40)	(2.03)	.	(-0.07)	(0.75)	(1.62)	(2.19)	(-0.16)	(-0.36)	.	(-2.29)	(-1.06)	(0.54)
constant	0.749	0.754	0.733	.	0.751	0.692	0.655	-0.889	-1.314	-1.200	.	-1.124	-1.038	-0.996
	(8.07)	(8.50)	(6.81)	.	(5.39)	(5.24)	(4.15)	(-2.87)	(-3.05)	(-2.98)	.	(-2.70)	(-2.43)	(-1.81)
				.							.			
adj. $R^2(\%)$	24.62	20.53	17.69	.	13.58	12.26	13.10	28.95	28.40	23.66	.	20.66	18.42	19.77

depth has a leading role in explaining longer horizon future volatility.

2.6.3 Predicting individual stock volatilities

This section examines the relationship between the limit order book distribution and the future volatility, if any, on an individual stock level. To this end, we run the following predictive regression:

$$\sigma_{s,\tau+1} = a_0 + a_1\sigma_{s,\tau} + a_2\mathrm{GD}^{\mathrm{buy}}_{s,\tau} + a_3\mathrm{GD}^{\mathrm{sell}}_{s,\tau} + \sum_{k=1}^{20} b_k T_{k,\tau} + \sum_{s=1}^{30} c_s D_s + \text{controls} + \varepsilon_{s,\tau}, \tag{2.10}$$

where, for stock s and interval τ, $\sigma_{s,\tau}$ is the mid-quote two scales realized volatility, $\mathrm{GD}^{\mathrm{buy}}_{s,\tau}$ and $\mathrm{GD}^{\mathrm{sell}}_{s,\tau}$ are global depth estimates for the buy and sell sides of the market, respectively. $T_{k,\tau}$ is the intraday dummy that equals to 1 if $k = \tau$, and D_s are stock-specific dummy variables allowing for stock fixed effects. We employ the same control variables introduced in Section 2.5.3.

We first run the regression model in a pooled data with stock fixed effects. Table 2.6 columns I to IV report the estimated coefficients for the pooled regression with the corresponding $t-$statistics. Second, we estimate individual regressions for all of the stocks in our sample to take into account the possible cross-sectional variations that cannot be captured by the stock fixed effects. The summary of these results are presented in columns V to VIII. We report the cross-sectional median of the estimated significant coefficients at a 5% level. In brackets, first, we report the percentage of the stocks with a significant coefficient at a 5% level, and second, we report the percentage of the positive estimates (given significant).

Our main result is confirmed in these individual volatility regressions. Global depth is negatively related to the future volatility for 83% of the stocks for the buy side of the market. We conclude that the time-series relation between the aggregate liquidity and market volatility is not driven by variations in a particular stock or

Table 2.6: Predictive Regressions–Individual Stocks

This table reports the estimated coefficients of the regression model defined in equation (2.10). $\text{GD}^{\text{buy}}_{s,\tau}$ and $\text{GD}^{\text{sell}}_{s,\tau}$ are the individual stock's global depth estimates for the buy and sell sides of the market, respectively, evaluated at the optimal decay factor, as outlined in Section 2.5.2. In a given trading interval τ, SLOPE is the slope of the limit order book defined in equation (2.6), relSPR is the relative spread, NT is the number of trades and AQ is the average trade size. AMR is the Amihud (2002) illiquidity measure. The logQS is the log quote slope, introduced by Hasbrouck and Seppi (2001) and defined in equation (2.8). Finally, DHW is the Domowitz et al. (2005) illiquidity measure described in equation (2.9). All of the explanatory variables are standardized. The dependent variable is $\sigma_{\tau+1}$, which is the TSRV volatility calculated using the mid-quotes of the orders originated in interval $\tau+1$ (multiplied by 100). Columns I to IV show the results for the pooled regression. t-statistics based on cluster robust standard errors on stock level are reported. Columns V to VIII summarize the results when the model is estimated for each stock separately. The cross-sectional median of the estimated significant coefficients at a 5% level is reported. In brackets, first, the percentage of the stocks with a significant coefficient at a 5% level and second, the percentage of the positive estimates, are reported. t-statistics are calculated using Newey-West standard errors to capture possible autocorrelation in the residuals. For the sake of brevity, the estimated coefficients of the intraday dummies and stock fixed effects are omitted.

	Pooled regression				**Summary of individual regressions**			
	I	II	III	IV	V	VI	VII	VIII
$\text{GD}^{\text{buy}}_{\tau}(\widehat{\lambda})$	-0.056	-0.059	-0.057	-0.054	-0.063	-0.059	-0.061	-0.055
	(-12.40)	(-12.88)	(-12.71)	(-12.35)	[83/0]	[87/0]	[77/0]	[77/0]
$\text{GD}^{\text{sell}}_{\tau}(\widehat{\lambda})$	-0.020	-0.028	-0.017	-0.018	-0.040	-0.049	-0.057	-0.055
	(-5.66)	(-6.63)	(-3.58)	(-4.02)	[27/0]	[40/0]	[13/0]	[13/0]
SLOPE_{τ}		-0.006	0.035	0.028		-0.046	0.058	0.047
		(-0.82)	(5.56)	(4.80)		[33/20]	[37/81]	[27/75]
relSPR_{τ}		0.051	0.014	0.009		0.083	0.056	-0.044
		(5.50)	(1.43)	(0.91)		[43/100]	[43/54]	[33/40]
NT_{τ}		0.0331		0.036		0.046		0.046
		(9.88)		(10.93)		[53/100]		[63/100]
AQ_{τ}		0.000		0.004		0.029		0.030
		(-0.02)		(0.98)		[23/71]		[23/71]
AMR_{τ}			-0.002	0.001			0.001	0.015
			(-0.80)	(0.29)			[20/50]	[17/60]
logQS_{τ}			0.077	0.080			0.115	0.108
			(9.80)	(10.22)			[40/100]	[43/100]
DHW_{τ}			0.013	0.014			0.069	0.071
			(2.67)	(2.85)			[30/100]	[27/100]
σ_{τ}	0.076	0.047	0.058	0.037	0.076	0.060	0.064	0.057
	(15.17)	(9.23)	(13.33)	(7.37)	[97/100]	[53/100]	[70/100]	[30/100]
constant	0.757	0.188	-0.192	-0.145	0.785	1.058	-0.102	0.692
	(20.19)	(1.25)	(-1.39)	(-1.09)	[100/100]	[60/94]	[33/50]	[30/67]
adj. R^2(%)	13.40	15.10	15.80	16.20	10.07	13.74	13.53	14.45

industry, but rather the relation is shared by the majority of the stocks.

The results reveal the asymmetry between the buy and sell sides of the market at the individual stock level as well. The sell side of the market is informative only for 27% of the stocks in the individual regressions. Although both sides of the market are significant in the pooled regression, the economic importance of the buy side is almost three times greater than the sell side.

Besides global depth, there are other pieces that contain information about future individual stock volatility. In line with the main prediction of Foucault et al. (2007), we find that a wider relative spread signals that the informed traders expect higher volatility in the future. Moreover, the number of trades and the (log) slope of the best quotes, logQS, are positively related to the future volatility.

In summary, we conclude that global depth on the buy side of the market has the leading explanatory power on one-period-ahead individual stock volatility compared to the standard predictors of volatility. This result is robust to the inclusion of the liquidity controls. Besides the standard predictors, we provide new evidence that the slope of the best quotes (an illiquidity measure proposed by Hasbrouck and Seppi (2001)) predicts volatility.

2.6.4 Out-of-sample tests

The results presented in Section 2.6 document that our proposed measure, global depth, significantly explains up to the 150-minute-ahead market volatility. Besides global depth, the slope of the order book and the relative spread contain information about the future market volatility. In this section, we assess the predictive ability of these three measures through out-of-sample forecasting experiments.

We evaluate the out-of-sample forecasting ability of each variable compared to its historical average. Specifically, for a subsample of observations up to a given time t, we compare the h-period-ahead squared forecast errors with the squared difference between the realized value at $t+h$ and the sample mean value up to time t. To do so, we split our data into two subsamples: T_{in} is the estimation period and T_{test} is the testing period with $T_{\text{in}} + T_{\text{test}} = T$. We then re-estimate the parameters of the model in which we use the variable of interest as the predictor. Recursive estimators of h-period-ahead forecasts are based on the sample starting from T_{in} up to $T-h$. We calculate the following error terms:

$$\varepsilon_{1,t+h} = \sigma^M_{t+h} - \widehat{\sigma^M_{t+h}},$$

$$\varepsilon_{2,t+h} = \sigma^M_{t+h} - \overline{\sigma^M_t},$$

where σ^M_{t+h} and $\widehat{\sigma^M_{t+h}}$ are the two scales realized and fitted market volatilities at time $t+h$ and $\overline{\sigma^M_t}$ is the mean value of the market volatility up to time t.

We evaluate the comparison by using two different metrics: the difference in mean-squared errors (ΔMSE) and the out-of-sample R^2. If the proposed measure has superior out-of-sample forecasting ability relative to the average of past data, then both of these measures will be positive. Finally, we employ the Diebold and Mariano (1995) predictive ability test (DM) to test the significance of ΔMSE. The difference in the mean-squared error and the out-of-sample R^2 are calculated as follows:

$$\Delta MSE = \frac{1}{T_{\text{test}} - h} \sum_{t=1}^{T_{\text{test}}-h} \varepsilon^2_{2,t+h} - \frac{1}{T_{\text{test}} - h} \sum_{t=1}^{T_{\text{test}}-h} \varepsilon^2_{1,t+h}, \tag{2.11}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{T_{\text{test}}-h} \varepsilon^2_{1,t+h}}{\sum_{t=1}^{T_{\text{test}}-h} \varepsilon^2_{2,t+h}}. \tag{2.12}$$

Panels A and B of Table 2.7 present the statistics when the estimation windows are 250 and 350 observations, respectively.

Our findings in Panel A reveal that the difference in mean-squared errors and out-of-sample R^2s are positive irrespective of the chosen forecasting variable. In other words, forecasts based either on global depth variables, slope or the relative spread increase the predictive power relative to forecasts based only on the sample mean of

Table 2.7: Out-of-Sample Forecasting Evaluation

The out-of-sample forecasting experiment results are reported in the table. The h-period-ahead forecast error is obtained as the difference between the realized volatility at $t+h$ and the fitted value of the predictive regression estimated up to time t, whereas the competing error is calculated from the sample mean volatility up to time t. The dependent variable is the 15-minutes market volatility, σ^M, calculated as the TSRV mid-quote volatility of the value-weighted index (multiplied by 100). $\overline{\text{GD}}_\tau^{\text{buy}}(\widehat{\lambda})$ is the cross-sectional average of global depth of individual stocks, $\text{GD}_{s,\tau}^{\text{buy}}$, evaluated at the optimal decay factor $\widehat{\lambda}$. Similarly, $\overline{\text{SLOPE}}_\tau$ and $\overline{\text{relSPR}}_\tau$ are the cross-sectional averages of the slope of the limit order book defined in equation (2.6) and the relative spread, respectively. The out-of-sample $R^2(\%)$ and the difference in mean-squared errors (ΔMSEx1000) are defined in equations (2.12) and (2.11), respectively. Finally, DM denotes the Diebold-Mariano predictive ability test. Panels A and B report the results when the estimation window is set to 250 and 350 observations, respectively.

Forecasting variable		**0–15min**	**15–30min**	**30–45min**	**45–60min**	**60–75min**	**75–90min**
Panel A: Estimation Window: 250 obs.							
$\overline{\text{GD}}_\tau^{\text{buy}}(\widehat{\lambda})$	out-of-sample $R^2(\%)$	11.24	8.64	7.55	7.05	5.70	4.10
	ΔMSE	1.82	1.40	1.22	1.14	0.93	0.67
	DM $t-$stat	2.76	2.54	2.49	2.33	2.05	1.52
$\overline{\text{GD}}_\tau^{\text{sell}}(\widehat{\lambda})$	out-of-sample $R^2(\%)$	2.51	3.83	3.17	3.33	3.76	3.91
	ΔMSE	0.41	0.62	0.51	0.54	0.61	0.64
	DM $t-$stat	0.79	1.00	0.95	0.96	1.15	1.19
$\overline{\text{SLOPE}}_\tau$	Out-of-sample $R^2(\%)$	3.06	1.37	1.68	2.90	3.52	4.10
	ΔMSE	0.50	0.22	0.27	0.47	0.57	0.67
	DM $t-$stat	1.63	0.79	1.38	1.75	1.55	1.26
$\overline{\text{relSPR}}_\tau$	out-of-sample $R^2(\%)$	15.39	13.93	13.54	13.43	13.12	13.64
	ΔMSE	2.49	2.25	2.19	2.18	2.13	2.23
	DM $t-$stat	3.22	3.14	3.04	2.99	2.89	3.21
Panel B: Estimation Window: 350 obs.							
$\overline{\text{GD}}_\tau^{\text{buy}}(\widehat{\lambda})$	out-of-sample $R^2(\%)$	14.48	11.96	11.33	10.00	8.12	6.15
	ΔMSE	2.00	1.64	1.55	1.36	1.10	0.81
	DM $t-$stat	2.65	2.52	2.73	2.32	2.15	1.54
$\overline{\text{GD}}_\tau^{\text{sell}}(\widehat{\lambda})$	out-of-sample $R^2(\%)$	0.31	1.15	1.56	0.92	0.43	-0.67
	ΔMSE	0.04	0.16	0.21	0.13	0.06	-0.09
	DM $t-$stat	0.08	0.22	0.34	0.21	0.10	-0.15
$\overline{\text{SLOPE}}_\tau$	Out-of-sample $R^2(\%)$	1.31	0.70	0.59	-0.07	-0.97	-0.93
	ΔMSE	0.18	0.10	0.08	-0.01	-0.13	-0.12
	DM $t-$stat	0.53	0.48	0.40	-0.03	-0.37	-0.22
$\overline{\text{relSPR}}_\tau$	out-of-sample $R^2(\%)$	9.98	6.71	7.09	6.47	5.19	5.68
	ΔMSE	1.38	0.92	0.97	0.88	0.70	0.75
	DM $t-$stat	1.77	1.44	1.45	1.35	1.04	1.21

past volatility. The Diebold-Mariano test shows that only global depth for the buy side of the market and the relative spread are the statistically significant predictors of market volatility, relative spread being stronger. Moreover, we see that the predictive power of both spread and global depth are decreasing almost monotonically with the prediction horizon.

Panel B, on the other hand, uncovers stronger results for $\overline{\mathrm{GD}}^{\mathrm{buy}}$. Our variable delivers impressive out-of-sample R^2's from 14.5% when forecasting one-period-ahead market volatility up to 6.2% when predicting 90-minutes-ahead market volatility. On the other hand, we observe that all of the statistics are worsened when we focus on the relative spread performance. The highest out-of-sample R^2 is 9.98% and reached when the forecast horizon is one-period-ahead. The statistical significance of the difference in mean-squared errors is also found to be the highest for the same prediction horizon, but only with a t-statistics of 1.77.

As a further analysis, we examine whether employing relative spread alone, or employing the buy side global depth along with the spread produces better forecasts. To do so, the first forecast errors are calculated from the model where $\overline{\mathrm{GD}}^{\mathrm{buy}}$ and $\overline{\mathrm{relSPR}}$ are the explanatory variables, whereas the second forecast errors are calculated from the model in which relative spread is the only explanatory variable. Similarly, we repeat the analysis for two different estimation window sizes; 250 and 350 observations. The results show that, when we set the estimation window size equal to 250, where both of the variables were found to have a good out-of-sample performance, including global depth into the analysis increases the out-of-sample R^2 by almost 7%. The difference in mean-squared errors is significant at 5% with a t-statistics of 2.66. When the estimation window is 350 observations, as expected, all of the statistics improve. The out-of-sample R^2 is increased to over 11% and ΔMSE is significant with a t-statistics of 3.20. Note that by construction, global depth does not include the bid-ask spread since the price distances are calculated as the position to the best quotes, rather than the mid-quotes. Thus global depth is related to the depth dimension of liquidity and can be thought as a complement of the tightness dimension. Hence, we conclude that capturing both the tightness and the depth dimension of liquidity significantly increases the out-of-sample forecasting power.

2.6.5 Robustness

We perform four sets of robustness tests. Our first set of robustness checks is on the specification of the weights to estimate global depth. We employ the following weight specification instead of exponential decaying factors:

$$\widetilde{g}(\lambda, \Delta) = \frac{\frac{1}{\Delta\lambda+1}}{\sum_{\Delta=0}^{30} \frac{1}{\Delta\lambda+1}}.$$

We re-estimate the optimal decaying factor λ via non-linear least squares as $\widehat{\lambda} = 1.318$ following the model outlined in equation (2.3). We then evaluate global depth at $\widehat{\lambda}$.

Second, instead of sampling the trading day using the 15-minute snapshots, we test the predictive power of the limit order book distribution on 30-minute intervals. Similarly, we first re-estimate the decay parameter for 30-minute intervals as 0.364 and then evaluate global depth at $\widehat{\lambda}$.

Third, we perform a robustness test on the specification of the regression model. We re-estimate the benchmark specification in equation (2.4) with the log-transformed variables to allow the left-hand side of the equation to include potentially both positive and negative numbers.

In our analysis, to proxy the aggregate level of liquidity, we first calculate global depth for each stock and get the cross-sectional average. Our final robustness check includes the re-calculation of the aggregated measures by using value-weighted cross-sectional averages.

Our results are presented in Table 2.8. Columns I and II repeat the results for the benchmark specification. Columns III and IV present the results for the first

Table 2.8: Robustness

This table reports the results for the robustness analysis. Columns I and II repeat the results reported in Table 2.4: the benchmark specification. Columns III and IV present the results for the first robustness check, i.e., when the linear decaying weight function introduced in equation (2.13) is used instead of the exponential decaying weights. The following two columns show the results when the sampling period is 30 minutes instead of 15 minutes. In columns VII and VIII, we report the estimated coefficients for the log-transformed variables. Finally, the last two columns report the results when the explanatory variables are aggregated via value-weighted cross-sectional averages instead of equal-weighted. All of the explanatory variables are standardized. In all of the specifications t-statistics are calculated using Newey-West standard errors to capture possible autocorrelation in the residuals and for the sake of brevity, the estimated coefficients of the intraday dummies are omitted. All of the variables are defined in Table 2.4.

	benchmark		linear-decaying weights		30–min. sampling		log-transform.		value-weighted	
	I	II	III	IV	V	VI	VII	VIII	IX	X
$\overline{\mathrm{GD}}_{\tau}^{\mathrm{buy}}(\widehat{\lambda})$	-0.034	-0.028	-0.031	-0.027	-0.043	-0.041	-0.034	-0.029	-0.031	-0.028
	(-6.82)	(-5.23)	(-6.55)	(-5.07)	(-5.41)	(-6.03)	(-6.86)	(-5.29)	(-6.20)	(-4.40)
$\overline{\mathrm{GD}}_{\tau}^{\mathrm{sell}}(\widehat{\lambda})$	-0.021	-0.011	-0.019	-0.009	-0.028	-0.009	-0.022	-0.011	-0.022	-0.016
	(-3.51)	(-1.56)	(-3.14)	(-1.27)	(-2.74)	(-1.03)	(-3.59)	(-1.55)	(-3.95)	(-2.29)
$\overline{\mathrm{SLOPE}}_{\tau}$		0.015		0.015		0.039		0.017		0.021
		(2.26)		(2.21)		(2.73)		(2.60)		(2.59)
$\overline{\mathrm{relSPR}}_{\tau}$		0.021		0.021		0.033		0.022		-0.001
		(2.12)		(2.09)		(3.10)		(2.24)		(-0.09)
$\overline{\mathrm{NT}}_{\tau}$		0.008		0.008		-0.002		0.005		0.009
		(1.24)		(1.27)		(-0.16)		(0.74)		(1.69)
$\overline{\mathrm{AQ}}_{\tau}$		0.001		0.002		0.012		-0.001		-0.004
		(0.28)		(0.33)		(1.45)		(-0.29)		(-0.99)
$\overline{\mathrm{AMR}}_{\tau}$		0.002		0.002		0.012		0.002		0.001
		(0.58)		(0.59)		(2.24)		(0.50)		(0.39)
$\overline{\mathrm{logQS}}_{\tau}$		0.014		0.015		0.045		0.014		0.034
		(0.96)		(1.00)		(2.93)		(0.93)		(2.30)
$\overline{\mathrm{DHW}}_{\tau}$		0.003		0.003		0.011		0.003		0.000
		(0.60)		(0.55)		(1.66)		(0.56)		(0.03)
σ_{τ}^{M}	0.023	0.011	0.024	0.012	0.047	0.02	0.023	0.01	0.025	0.01
	(4.96)	(2.19)	(5.02)	(2.23)	(4.87)	(2.01)	(4.86)	(2.60)	(5.39)	(2.66)
constant	0.749	-0.889	1.026	-0.704	1.016	-2.333	0.858	-3.220	0.626	-0.450
	(8.07)	(-2.87)	(7.10)	(-2.15)	(7.02)	(-4.05)	(7.94)	(-3.23)	(8.79)	(-1.63)
adj. $R^2(\%)$	24.62	28.95	24.31	28.68	33.23	43.02	24.84	28.83	24.26	27.29

robustness check, i.e., when the linear decaying optimal weights are employed instead of exponential decaying weights. The following two columns show the results when we use 30-minute sampling frequency instead of 15-minute sampling. In columns VII and VIII, we report the results for the log-transformed variables.

Finally, the last two columns report the results when the explanatory variables are aggregated via value-weighted cross-sectional averages. All of the regressions include the intraday dummy variables. The estimated coefficients are omitted for the sake of brevity. All of the explanatory variables are standardized.

The results for all of the robustness tests provide strong evidence for the informativeness of the buy side global depth on future volatility of the efficient price. The sell side of the market is significant only when the aggregated sell side global depth is approximated as the value-weighted average of the individual stocks. We observe an increase in the informativeness of global depth, specially in a multivariate setting, when the sampling period is 30 minutes instead of 15 minutes. All of the estimated coefficients and the adjusted R^2s are higher under the former frequency.

Overall, the results reveal that our findings for the informativeness of global depth over future efficient return volatility is robust to the weight functions, different model specifications, and the chosen sampling period.

2.7 Conclusion

In this paper, we evaluate the role of relative depth provision in future market volatility. To measure the former, we propose a novel way of summarizing the distribution of liquidity in a limit order book, while taking into account the relative magnitude and the location of the quoted depth. Our summary measure, global depth, considers how liquidity is distributed in the whole book and assigns weights to the information provided by different quotes.

By using high-frequency data from the Istanbul Stock Exchange, we document strong evidence that global depth is negatively correlated with one-period-ahead market and individual stock volatilities. It dominates the explanatory power of standard predictors of volatility. These results are remarkably robust to the inclusion of several liquidity measures. Besides global depth, we find evidence that the relative spread is informative, supporting the theoretical prediction of Foucault et al. (2007).

Out-of-sample forecasting experiments provide formal evidence of the predictive power of both global depth and the relative spread on future volatility. We conclude that including both measures in the analysis and thus capturing both the tightness and the depth dimension of liquidity, significantly increases the out-of-sample R^2.

We contribute to the existing empirical literature, which examines the informativeness of a limit order book on future volatility. However, this is the first study that examines the predictive power of aggregate liquidity on intraday market volatility. Moreover, we propose a new measure with a superior explanatory power compared to standard liquidity measures.

Chapter 3

Competition, Signaling and Non-Walking Through the Book: Effects on Order Choice

Co-authored with Ilknur Zer (London School of Economics)

3.1 Introduction

The limit order book and the characteristics of an asset, such as volatility, provide essential information for a trader who wants to design an appropriate order submission strategy. This in turn affects the price formation of an asset and the liquidity dynamics in the market. Following this, there has been a growing research interest on investors' choice of order submission over the last decade. By undertaking an empirical study of a pure order driven market, this paper aims to contribute to this literature. Our contribution is twofold: first, we examine the trading patterns of agents when walking through the book is not allowed, i.e., when orders that would otherwise walk through the book are converted into limit orders. Second, we test whether "competition" or "signaling" effects, two theories that have been proposed in the existing literature, dominate each other for depth beyond the best quotes. Both of these analyses are the first attempts in the literature.

In the Istanbul Stock Exchange, walking through the book is not allowed. That is, a "large" market order is first matched with the available volume at the best corresponding quote. Then, the remaining part is converted to a limit order at the quoted price instead of walking up or down the limit order book to be fully executed. This market rule obviously affects the cost of a market order. When walking down/up the book is allowed, the cost of execution of a large market order is higher since it matches with less favorable prices (Hamao and Hasbrouck (1995)). This in turn should affect the market order trader's submission strategy. By focusing on the order choice of an impatient (market order) trader, we analyze the informativeness of the price information contained in the book.

In an early work, Parlour (1998) suggests that an increase in the same-side thickness of the limit order book reveals high *competition*, which in turn increases the submission of more aggressive orders in order to jump the queue ("competition effect"). On the other hand, in their recent theoretical works, Goettler et al. (2005) and Goettler et al. (2009) argue that if the total volume of orders waiting beyond the best bid (ask) is "too high", then this *signals* to the market that the current quotes are mispriced and should decrease (increase) ("signaling effect"). By calculating the volume of orders waiting in the queue for the 10 best quotes, we analyze which effect dominates at every price level.

Our analysis requires considering the reaction of patient (limit order) and impatient (market order) traders separately to the changing market conditions. Hence, similar to Pascual and Veredas (2009), we employ a two-stage sequential ordered probit (SOP) model. Although our methodology coincides with their study, our research questions are different. In order to test whether competition effect is more persistent than the best quotes, we focus on the actions of patient traders. On the other hand, to analyze whether or how non-walking through the book affects the trading strategy of a market order trader, we focus on the trading strategies of impatient traders.

Using the unprocessed order flow and trade data provided by the Istanbul Stock

Exchange (ISE), we first reconstruct the limit order book dynamically. We use the order flow, trade book and limit order book to analyze the effects of the information content of the books on the order choice of a trader on a sample of 30 stocks for the period of June and July 2008. Our data set has one major advantage compared to many studies: since the ISE is a fully computerized and centralized stock exchange (unlike NYSE, there is no specialist and unlike the London Stock Exchange for instance, there is a single trading platform in the ISE), the data generated *fully* captures the order flow and the execution process. Moreover, in our data set we can distinguish whether an order is initiated by an institutional or individual investor. By using this classification we examine whether the trading behavior is different for institutional traders compared to the individual ones.

There are several papers that provide a theoretical background that the state of the limit order book contains information that shapes agents' trading decisions.[1] Ahn et al. (2001), Ranaldo (2004), Beber and Caglio (2005), Ellul et al. (2007), Fong and Liu (2010), Menkhoff et al. (2010), among others, investigate the state of the book and its effects on order choice of an investor in an empirical framework. The aforementioned studies consider the informativeness of the limit order book only at the best quotes, as opposed to Cao et al. (2008), Cao et al. (2009), Pascual and Veredas (2009) and Lo and Sapp (2010).

Using data from the Australian Stock Exchange, Cao et al. (2008) show that the information contained at the best quotes affects order submissions, cancelations, and modifications. On the other hand, the rest of the book matters for order cancellations and modifications. Using the same data set, Cao et al. (2009) investigate whether the prices beyond the best bid and offer and their corresponding depths matter in price discovery. They conclude that the contribution of beyond the book to the price

[1] See Parlour (1998), Foucault (1999), Foucault et al. (2005), Goettler et al. (2005), Kaniel and Liu (2006), Goettler et al. (2009), Rosu (2009), among others.

discovery is 22%, whereas the remaining part comes from the current bid and ask prices as well as the transaction price. Using a two-stage sequential ordered probit model, Pascual and Veredas (2009) conclude that not only the best quotes, but the information beyond the best quotes matters in explaining the degree of patience of incoming orders. Moreover, they note that although the impatient traders strongly rely on the best quotes, for limit order traders, strategic decisions are primarily based on the state of the book beyond the best quotes. Lo and Sapp (2010) empirically show the trade-off between order aggressiveness and quantity. Using a simultaneous equations framework in a foreign exchange market, they conclude that order size tends to be smaller when an order is more aggressive. That is, by submitting smaller size market orders, traders avoid the higher execution costs. Our paper is the first study that investigates whether the volume of orders waiting at different price distances encourage agents to submit more aggressive orders and jump the queue, or rather signal them to submit less aggressive orders. Moreover, an atypical feature of our dataset enables us to examine the order choice of a trader when walking through the book is not allowed.[2]

Our main findings can be summarized as follows:

- The competition effect dominates the signaling effect for both sides of the market, in every stage.
- For a limit order agent, the competition effect is persistent beyond the best quotes. We show that for both sides of the market, the volume up to the second best quotes has the strongest competition effect.
- While fitting the size of her market order, for an impatient trader none of the price information, neither spread or price distance variables, matter in our

[2]There are other studies that use intraday data from the ISE. For instance, Bildik (2001) and Ekinci (2008) provide intraday descriptive analyses for the ISE. Bildik (2001) examines the intraday seasonality of the stock returns and volatilities, whereas Ekinci (2008) focuses on the intraday liquidity patterns.

market. This might be a result of the non-walking through the book, since under this mechanism, the spread and the price distance variables do not capture the cost of a large market order.

- We show that volatility, previous price trend and volume accumulated beyond the best quotes on the opposite side of the book affect the aggressiveness of market orders. This result might also be explained by the non-allowance of walking through the book, since these variables affect the execution probability of the unexecuted part of a large market order.

- Institutional investors consider only the competition effect variables while they decide to submit a market or a limit order. If they are informed traders as proposed by the existing literature, this may imply that institutions place orders based more on their own private valuations than the information provided by the limit order book.

The paper is organized as follows: Next section describes data and introduce the order aggressiveness categories. Section 3.3 presents the econometric methodology; the two-stage sequential ordered probit model. In Section 3.4, we list the explanatory variables and discuss the empirical questions. Section 3.5 presents the empirical findings and robustness checks. Finally section 3.6 concludes.

3.2 The Market and Data

3.2.1 Trading structure in the Istanbul Stock Exchange

The Istanbul Stock Exchange (ISE) is operating as a fully computerized pure order-driven market since November 1994. As of December 2012, the ISE index had a $358 billion value of shares traded year-to-date and $315 billion of market capitalization. The total value of shares trading and the market capitalization were

3% and 2% of NYSE respectively.[3] In terms of value of shares traded, it is the 20th largest stock exchange in the world and 5th within the emerging countries.[4]

Similar to all other major exchanges, a trading day starts with a call market matching mechanism to determine the opening price. For the rest of the day, a double auction continuous order matching mechanism is used for trading. Trading occurs in two sessions with a lunch break and every order is valid for a corresponding session or for a day. For the period under consideration, the double-continuous auction trading occurs between 9:45–12:00 in the morning session and 14:00–17:00 in the afternoon session. A given order is either matched, resulting in a trade, or queued up in a limit order book waiting to be executed based on the usual price and time priorities. The fully computerized system ensures the strict enforcement of those priority rules. The status of a given security is updated almost instantaneously on the traders' screens, whenever there is an order arrival, or execution.

Similar to the Australian Stock Exchange and the Spanish Stock Exchange for instance, the ISE is an open limit order book market. In this market, both individual and institutional investors are directly connected to the ISE system and they can observe the book in real time. On the other hand, the ISE offers more pre-trade transparency compared to many other exchanges. Upon arrival, traders can observe all of the orders submitted/traded, with the corresponding prices and volumes. The information is not truncated to any price step. Moreover, for the executed orders only, they can see the name of the corresponding party who initiated the trade.[5] The open book and pre-trade transparency properties are relevant for our study since we examine the "competition" and "signaling" effects beyond the best quotes up to the 10 best prices.

[3]Source: World Federation of Stock Exchanges.

[4]Emerging countries are classified based on the list of the International Monetary Fund July 16, 2012 report.

[5]The non-anonymity has changed by October 2010, but for the sample under consideration, traders can identify the name of the trading parties.

The other market mechanism worth to emphasize is that walking through the book is not allowed in the ISE, similar to the Australian Stock Exchange, the Sao Paulo Stock Exchange (Bovespa), and the Stock Exchange of Hong Kong, for example. Hence, the unexecuted portion of a marketable limit order[6] is converted to a limit order. If an investor wishes to buy (sell) shares by walking up (down) the book, she needs to use appropriate limit orders. This characteristic allows us to examine the effects of this particular market mechanism on the order choice of a market order trader.

3.2.2 Data and descriptive analysis

Our dataset contains the order and trade books for the period of June and July 2008 for the biggest 30 stocks listed on the Istanbul Stock Exchange (ISE–30 index). The 30 stocks in our sample correspond to 75% of the total trading volume of the ISE for the period under consideration. These data sets allow us to reconstruct the complete limit order book dynamically. The order book data consists of all submitted orders for a given stock and date, their corresponding prices and quantities, order submission times, an order identification number (order ID), buy/sell indicator, as well as whether the trader is an institutional or an individual one. On the other hand, the transaction data consists of all the executed orders, their corresponding prices and quantities, and execution times. These two books are linked to each other with order and trade ID numbers generated by the ISE system. Hence, our data enables us to track an order from submission to execution or modification (if any).

To reconstruct the limit order book, we incorporate every order according to the price and time priority rules and fill in the limit order book one by one. If the price of a new-coming buy (sell) order is higher (lower) than or equal to the ask (bid) price, we classify it as a market order. A market order is matched with the corresponding

[6] In this study, we call marketable limit orders as market orders following Payne (2003) and Hasbrouck and Saar (2009).

order(s) from the other side of the book and removed from the limit order book. Moreover, if an order revision (including the split) is submitted, the original order is removed from the limit order book. For a given limit order book snapshot, we have a list of orders submitted but not yet executed, whether they are buy or sell orders and originated by individual or institutional traders, price and volume information up to the 10^{th} best quotes. The volume available at the best, second best, and up to the 10^{th} best prices are calculated as the total volume of orders waiting at that price level. Hence, by reconstructing the limit order book, we have access to the information on both the length (price information) and the height (the corresponding volume information) of a limit order book, which is crucial for our analysis to understand how the information beyond the best quotes affects the order submission strategies of agents.

Table 3.1 reports the descriptive statistics of the order flow and trade book, averaged across the sample period. Besides the market capitalization, for which the value at the beginning of the sample period in million Turkish Liras (M TRY) is presented, all of the figures are obtained by averaging across trading days (excluding the opening sessions). The results show that, on average 2253 orders are submitted in a day, equivalent to 83 million TRY.[7] The highest number of orders is submitted and traded by Garanti Bankasi (GARAN) investors, whereas the smallest one is for Migros (MIGRS). In terms of volume of orders submitted, GARAN is 8 times bigger than the average, whereas MIGRS, is 9 times smaller than the average. Although our sample is composed by the 30 biggest stocks traded in the ISE, these results show a high degree of heterogeneity in the sample of study. On average around 1400 trades occur in a day with a total daily average trade size of 9 million shares. This corresponds to an average value traded of around 28 million TRY per day. The number of buy orders is slightly less than the number of sell orders, and the number of limit

[7] On 25th of July 2008, the exchange rate was 1.20USD/TRY.

Table 3.1: Descriptive Statistics for Each Stock

The table reports the summary statistics of ISE–30 stocks for June–July 2008. The first and the second columns present the ticker and names of the securities in our sample, respectively. The market capitalization is the value at beginning of the sample period in million Turkish Liras (M TRY). Number of Orders (Trades) is the average of the total number of orders (trades) in a day. Volume of Orders (Trades) is the average of the daily number of shares submitted (traded). Value of Orders (Trades) is the average of the daily value of orders (trades) (volume x price). Spread is the tick-adjusted difference between the best ask and the best bid. Finally the last two columns report the average of the daily percentage of buy orders and limit orders, respectively.

Company Ticker	**Company Name**	**Market Capitalization (M TRY)**	**Number of Orders**	**Volume of Orders (M shares)**	**Value of Orders (M TRY)**
AKBNK	Akbank	16650	2609	26	130.63
AKGRT	Aksigorta	1463	1044	4	18.35
ARCLK	Arcelik	1664	1003	2	10.51
ASYAB	Asya Katilim Bankasi	1980	1392	7	16.94
DOHOL	Dogan Holding	2160	2438	37	54.95
DYHOL	Dogan Yayin Holding	1082	2991	28	46.06
EREGL	Eregli Demir Celik	9995	2286	7	61.99
GARAN	Garanti Bankasi	14448	9259	221	749.10
GSDHO	Gsd Holding	277	2074	33	35.77
HALKB	Halk Bankasi	7750	1656	8	49.35
HURGZ	Hurriyet Gazetesi	745	2281	29	45.50
IHLAS	Ihlas Holding	202	1975	32	18.15
ISCTR	Is Bankasi	13165	7332	89	393.63
ISGYO	Is GMYO	459	700	5	4.94
KCHOL	Koc Holding	7629	1399	12	41.51
KRDMD	Kardemir	670	2016	34	38.73
MIGRS	Migros	3614	346	3	60.88
PETKM	Petkim	1024	1156	4	20.39
PTOFS	Petrol Ofisi	2778	507	2	8.47
SAHOL	Sabanci Holding	8676	1103	7	28.25
SISE	Sise Cam	1439	1572	10	14.73
SKBNK	Sekerbank	876	1872	10	21.47
TCELL	Turkcell	17050	1847	15	117.95
THYAO	Turk Hava Yollari	919	1252	5	26.83
TKFNK	Tekfen Holding	2166	1172	3	25.96
TSKB	TSKB	490	707	6	5.73
TTKOM	Turk Telekom	14350	4447	29	119.25
TUPRS	Tupras	7387	1413	3	75.11
VAKBN	Vakiflar Bankasi	4400	4813	86	151.08
YKBNK	Yapi ve Kredi Bankasi	9999	2939	42	106.19
Average		5184	2253	26.52	83.28
Median		2163	1752	10.04	40.12
Min		202	346	1.59	4.94
Max		17050	9259	221.13	749.10

Table 3.1: Descriptive Statistics for Each Stock (cont.)

Company Ticker	**Number of Trades**	**Volume Traded (M shares)**	**Value Traded (M TRY)**	**Spread (tick adj.)**	**%Buy**	**%LO**
AKBNK	1643	8.81	44.09	1.04	46.79	68.56
AKGRT	714	1.54	6.59	1.15	52.13	62.16
ARCLK	576	0.75	3.27	1.25	45.50	71.04
ASYAB	954	2.19	5.64	1.14	49.20	62.10
DOHOL	1546	12.37	18.45	1.06	44.11	68.74
DYHOL	1949	9.45	15.40	1.06	48.77	65.93
EREGL	1455	2.19	20.22	1.08	48.71	67.76
GARAN	6186	82.39	278.14	1.02	47.46	69.78
GSDHO	1400	10.91	11.78	1.05	47.48	64.22
HALKB	972	2.56	15.99	1.10	46.46	71.57
HURGZ	1455	9.53	15.09	1.10	47.05	67.16
IHLAS	942	7.63	4.30	1.01	47.64	70.75
ISCTR	4732	32.46	143.32	1.03	49.48	69.81
ISGYO	367	1.35	1.31	1.11	44.94	71.81
KCHOL	855	3.93	13.72	1.11	45.17	68.76
KRDMD	1150	9.91	11.39	1.05	45.80	70.28
MIGRS	152	0.48	9.84	1.03	38.90	70.28
PETKM	688	1.12	6.02	1.14	46.81	70.54
PTOFS	295	0.48	2.53	1.38	45.80	69.47
SAHOL	713	2.19	9.44	1.15	48.54	66.25
SISE	975	3.24	4.63	1.08	51.39	67.02
SKBNK	1216	3.23	7.06	1.15	44.15	64.36
TCELL	1095	5.05	40.15	1.10	46.47	71.25
THYAO	787	1.65	8.99	1.10	50.52	68.10
TKFNK	747	1.00	8.56	1.13	48.63	64.70
TSKB	448	1.72	1.62	1.06	48.98	63.23
TTKOM	2343	8.48	35.07	1.05	39.22	73.20
TUPRS	761	0.83	22.86	1.07	48.45	73.68
VAKBN	3169	31.17	54.61	1.04	47.42	68.53
YKBNK	1911	14.61	36.47	1.04	48.33	67.08
Average	1406	9.11	28.55	1.10	47.01	68.27
Median	973	3.24	11.59	1.08	47.44	68.65
Min	152	0.48	1.31	1.01	38.90	62.10
Max	6186	82.39	278.14	1.38	52.13	73.68

orders constitute about 68% of all the submitted orders. The average tick adjusted spread is quite narrow, being less than 2 for all of the stocks in our sample. This is similar to the findings of Griffiths et al. (2000) on the most liquid securities of the Toronto Stock Exchange, but lower than the spreads presented in Pascual and Veredas (2009)'s study of 36 stocks from the Spanish Stock Exchange.

Order aggressiveness

In order to analyze how the state of the book affects the order choice of an investor, we define order aggressiveness categories based on the classification of Biais, Hillion and Spatt (1995). The first two categories are related to the market order (MO) aggressiveness, whereas the rest is defined for the limit order (LO) aggressiveness based on the limit price position.

- Category 1 ("large MO buy"): $V_{\text{order}} \geq V_{\text{ask}}$ and $P_{\text{order}} \geq P_{\text{ask}}$.
- Category 2 ("small MO buy"): $V_{\text{order}} < V_{\text{ask}}$ and $P_{\text{order}} \geq P_{\text{ask}}$.
- Category 3 ("buy LO within the quotes"): $P_{\text{ask}} > P_{\text{order}} > P_{\text{bid}}$.
- Category 4 ("buy LO at the quote"): $P_{\text{ask}} > P_{\text{order}} = P_{\text{bid}}$.
- Category 5 ("buy LO away from the quote"): $P_{\text{order}} < P_{\text{bid}} < P_{\text{ask}}$.

where, V_{order} and P_{order} are the volume and the price of a buy order, respectively. V_{ask} is the total volume of orders waiting at the best ask price, P_{ask}. Finally, P_{bid} denotes the best bid price. Sell side is constructed analogously.

Table 3.2 presents the descriptive statistics of the order aggressiveness categories for both buy and sell sides of the market separately. The results suggest that for the buy side, the most frequent events are small buy market orders (category 2) followed by orders submitted at the quotes, whereas for the sell side the ones away from the best quotes (category 5) have the most frequent arrivals, contradicting the findings of Biais et al. (1995), Beber and Caglio (2005), and Griffiths et al. (2000) for the Paris Bourse, the NYSE and the Toronto Stock Exchange, respectively. Table 3.2 also shows a very low frequency of orders within the quotes (for both sides of the book), which can be explained by the small inside spread.The results regarding the execution rate, i.e., the proportion of orders executed, suggest that only around 20%

Table 3.2: Descriptive Statistics of the Order Aggressiveness Categories
This table presents the descriptive statistics of the order aggressiveness categories for both sides of the market. Orders are divided into five categories based on the limit price position following Biais et al. (1995). Category 1 ("large MO buy"): $V_{\text{order}} \geq V_{\text{ask}}$ and $P_{\text{order}} \geq P_{\text{ask}}$. Category 2 ("small MO buy"): $V_{\text{order}} < V_{\text{ask}}$ and $P_{\text{order}} \geq P_{\text{ask}}$. Category 3 ("buy LO within the quotes"): $P_{\text{ask}} > P_{\text{order}} > P_{\text{bid}}$. Category 4 ("buy LO at the quote"): $P_{\text{ask}} > P_{\text{order}} = P_{\text{bid}}$. Category 5 ("buy LO away from the quote"): $P_{\text{order}} < P_{\text{bid}} < P_{\text{ask}}$. V_{order} and P_{order} are the volume and the limit price of the buy order, respectively. V_{ask} is the accumulated volume of orders waiting at the best ask price, P_{ask}. Finally, P_{bid} denotes the best bid price. Sell side is constructed analogously. The first two columns report the proportion of orders and order sizes for each category. Execution rate is calculated as the proportion of orders executed in each category, whereas the last column presents the average execution time (in minutes) of orders in each category.

		Number of Orders (%)	Volume of orders (%)	Execution Rate (%)	Execution Time (min)
Buy Side					
	Category 1	3.77	14.82	98.05	3
	Category 2	33.24	24.31	99.77	0
	Category 3	0.98	1.90	86.88	18
	Category 4	32.14	34.79	67.33	24
	Category 5	29.87	24.17	21.31	88
Sell Side					
	Category 1	3.51	12.71	98.16	2
	Category 2	24.44	22.42	99.77	0
	Category 3	0.85	1.66	88.95	15
	Category 4	28.79	32.32	60.72	21
	Category 5	42.41	30.88	16.04	78

of orders away from the quotes are executed compared to 60% of execution rate for the orders at the quotes. That is, going from category 4 to 5; traders are facing a substantial non-execution risk. These figures are very similar to the study of Griffiths et al. (2000) conducted on the Toronto Stock Exchange. A similar conclusion can be derived from the average waiting times for execution.

3.3 Sequential Ordered Probit Regressions

We investigate how the information content of the limit order book affects the order choice of the investor, by considering the order choice as a two-stage process. As a first step in her order choice, observing the market dynamics and limit order book information, the agent is patient, i.e., submits a limit order, or she is impatient, i.e., submits a market order.[8] In the second stage, given the agent is patient, she decides the position of her limit price (decides to submit category 3, 4 or 5 order), whereas the impatient trader decides whether to submit a large or small market order (category 1 or 2 order). To allow this sequential decision, following Pascual and Veredas (2009), we employ a sequential ordered probit (SOP) model for the empirical investigation. The attractiveness of the SOP model, compared to the ordered probit model, is that the former enables us to compare the reaction of the patient and impatient trader to the changing market conditions separately.

3.3.1 First stage–arrival of a market or limit order trader

Let Y^* denote the degree of patience of an incoming agent in the first stage of the SOP model. Although Y^* is unobservable, we assume that it is a function of K observable (limit order book) variables, Xs. We consider volatility, price trend, volume and price distances as explanatory variables. A detailed explanation of the regressors is provided in the next section.

$$Y_t^* = \sum_{k=1}^{K} \beta_k X_{k,t-1} + \varepsilon_t, \tag{3.1}$$

$$Y_t = \begin{cases} 0 \text{ if } -\infty < Y_t^* \leq \delta \\ 1 \text{ if } \delta < Y_t^* < \infty \end{cases}, \tag{3.2}$$

where δ is the threshold and t refers to the transaction time, not the clock time. The first-stage-dependent variable is equal to 1 if the trader is impatient and submits a market order or 0 if the trader is patient and submits a limit order.

[8]One can argue that the degree of patience is based on a trader's information level, preferences or waiting costs, hence, exogenously determined. However, recent theoretical works suggest that market conditions and the state of the book affect the degree of patience. For example Goettler et al. (2009) claim that although a patient informed agent submits a limit order, when she observes high volatility, she switches to a market order to take advantage of the mispriced quotes. Similarly, in Foucault et al. (2005), if spread increases over a cutoff level, all traders, even the ones with high waiting costs, will submit limit orders. Moreover, Ranaldo (2004), Beber and Caglio (2005), among others, show empirically that a trader considers the state of the book while formulating her order strategies. Hence, we allow the arrival rate of patient and impatient agents to be influenced by the state of the book and market conditions.

Assuming that the error terms are normally distributed, the probability of the incoming trader being patient is:

$$\begin{aligned} P(Y_t = 0) &= P(-\infty < Y_t^* \leq \delta) \\ &= P(-\infty < \sum_{k=1}^{K} \beta_k X_{k,t-1} + \varepsilon_t \leq \delta) \\ &= \Phi(\delta - \sum_{k=1}^{K} \beta_k X_{k,t-1}), \end{aligned} \tag{3.3}$$

where Φ is the normal cumulative distribution function.

3.3.2 Second stage–patient trader

In the second stage, both patient and impatient traders choose their level of aggressiveness given the information content of the book. A patient trader has three choices: placing a limit order within, at or away from the best quotes. That is;

$$LO_t^* = \sum_{k=1}^{K} \theta_k X_{k,t-1}^{lo} + \varepsilon_t^{lo}, \tag{3.4}$$

$$LO_t = \begin{cases} 1 \text{ if } -\infty < LO_t^* \leq \delta_1^{lo} \\ 2 \text{ if } \delta_1^{lo} < LO_t^* \leq \delta_2^{lo} \\ 3 \text{ if } \delta_2^{lo} < LO_t^* < \infty \end{cases}, \tag{3.5}$$

where δ_1^{lo} and δ_2^{lo} are the thresholds.

The dependent variable is equal to 1 if a trader submits a limit order away from the best quotes (category 5), is equal to 2, if the order is submitted at the best quotes (category 4) or is equal to 3 if the order is submitted within the quotes (category 3). Hence, our dependent variable increases as aggressiveness increases.

Assuming that the error terms are normally distributed, the probability of the incoming patient trader being type $i = 1, 2, 3$ is:

$$P(\mathrm{LO}_t = i) = \Phi(\delta_i^{lo} - \sum_{k=1}^{K} \theta_k X_{k,t-1}) - \Phi(\delta_{i-1}^{lo} - \sum_{k=1}^{K} \theta_k X_{k,t-1}), \tag{3.6}$$

where $\delta_0^{lo} = -\infty$ and $\delta_3^{lo} = \infty$.

3.3.3 Second stage–impatient trader

Finally, the impatient trader decides the quantity she wants to trade; whether she submits an aggressive market order (category 1), or submits a small market order (category 2). The dependent variable is equal to 1 if a category 1 order is submitted, 0 otherwise.

$$MO_t^* = \sum_{k=1}^{K} \gamma_k X_{k,t-1}^{mo} + \varepsilon_t^{mo}, \tag{3.7}$$

$$MO_t = \begin{cases} 0 \text{ if } -\infty < MO_t^* \leq \delta_1^{mo} \\ 1 \text{ if } \delta_1^{mo} < MO_t^* < \infty \end{cases}, \tag{3.8}$$

where, δ_1^{mo} is the threshold.

As the coefficients of the sequential ordered probit measure the change in the latent variable with respect to a change in one of the explanatory variables, they are difficult to interpret. A direct interpretable measure is given by the marginal probabilities (marginal effects), which show how the probability of order choices is affected given a marginal change in any of the explanatory variables:

$$\begin{aligned}
\frac{\partial P(Y=0)}{\partial X_j} &= \frac{\partial \Phi(\delta - \sum_{k=1}^{K} \beta_k X_{k,t-1})}{\partial X_j} \\
&= -\phi(\delta - \sum_{k=1}^{K} \beta_k X_{k,t-1})\beta_j, &(3.9) \\
\frac{\partial P(LO=i)}{\partial X_j} &= \frac{\partial(\Phi(\delta_i^{lo} - \sum_{k=1}^{K} \theta_k X_{k,t-1}) - \Phi(\delta_{i-1}^{lo} - \sum_{k=1}^{K} \theta_k X_{k,t-1}))}{\partial X_j} \\
&= [\phi(\delta_{i-1}^{lo} - \sum_{k=1}^{K} \theta_k X_{k,t-1}) - \phi(\delta_i^{lo} - \sum_{k=1}^{K} \theta_k X_{k,t-1})]\theta_j, &(3.10) \\
\frac{\partial P(MO=0)}{\partial X_j} &= \frac{\partial \Phi(\delta_1^{mo} - \sum_{k=1}^{K} \gamma_k X_{k,t-1})}{\partial X_j} \\
&= \phi(\delta_1^{mo} - \sum_{k=1}^{K} \gamma_k X_{k,t-1})\gamma_j, &(3.11)
\end{aligned}$$

where $i = 1, 2, 3$ and $\delta_0^{lo} = -\infty$ and $\delta_3^{lo} = \infty$.

3.4 Empirical Analysis

Empirically we ask the following questions: whether "competition" or "signaling" effects dominate each other at every level of the depth, how/whether walking through the book affects the order decision of an impatient trader, and finally, whether the limit order book information affects the trading behavior institutional investors.

3.4.1 Covariates for the impact of depth at and beyond the best quotes

We test whether the competition and signaling effects, proposed by Parlour (1998) and Goettler et al. (2005), Goettler et al. (2009), respectively, dominate each other for depths beyond the best quotes. To do so, we calculate the volume of orders waiting in the queue for the 10 best prices. We define a proxy for the signaling and competition effects separately for every stage of the sequential ordered probit (SOP) model. In the first stage, when a trader decides whether to submit a market or a limit order, she considers only the increase of the volume at the best quotes (Vsame^1 and/or Vopp^1) as an increased competition. We therefore use the volume of orders waiting beyond the best quotes as a proxy for signaling effect. Given that the trader is impatient, in the second stage, she decides the size of her market order. In this case, since the order has the price priority, there will be no price competition and the volume of orders beyond the best quotes captures the signaling effect.

On the other hand, in the second stage, when a limit order trader decides her limit price, we consider two states: first, (tick-adjusted) inside spread greater than 1 and second, spread equal to 1. If an agent observes the inside spread greater than 1, then by submitting an order *within* the quotes (category 3 order) she can jump the queue. In this case, Vsame^1 and (possibly) depth beyond the best quotes captures the competition effect. However, if the spread is 1, then "mechanically" it is not possible to submit a category 3 order, i.e., a trader cannot gain priority over the orders already waiting at Vsame^1. In this case, while positioning her limit price, she may consider just the depth beyond the best quotes as an increased competition, at least up to some *cutoff* level, discarding the depth at the quotes as part of the competition effect. In order to determine the cutoff point, we run the SOP regressions with accumulated volume of orders from the second to the third, from the second to the fourth and from the second to the fifth best prices ($\text{Vsame}^{2\text{–}3}$, $\text{Vsame}^{2\text{–}4}$ and $\text{Vsame}^{2\text{–}5}$). The signaling effect will then be captured by $\text{Vsame}^{4\text{–}10}$, $\text{Vsame}^{5\text{–}10}$ and $\text{Vsame}^{6\text{–}10}$, respectively.

Table 3.3 reports the results. For both sides of the market, the volume up to the second best quotes has the strongest competition effect. That is, the competition effect persists beyond the best quotes. The marginal effects as well as the significance of the estimated coefficients are decreasing with the additional quotes added.[9] Moreover, at every price level, competition effect dominates the signaling effect. Finally, the results suggest an asymmetry between the sell and the buy side. The signaling effect is more persistent and stronger for the sell side.

As suggested, we pick the volume at the second best quote as the cutoff level. Hence, we define the competition effect, Vcomp and the signaling effect, Vsign as follows:

- Step 1– arrival rate of patient/impatient traders:

$$\begin{aligned} \text{Vcomp}_t &= \text{Vsame}_t^1, \qquad (3.12) \\ \text{Vsign}_t &= \text{Vsame}_t^2 + \text{Vsame}_t^3 + ... + \text{Vsame}_t^{10}. \end{aligned}$$

- Step 2– order choice of patient traders:

$$\begin{aligned} \text{Vcomp}_t &= \begin{cases} \text{Vsame}_t^2 & \text{if spread}_t = 1 \\ \text{Vsame}_t^1 + \text{Vsame}_t^2 & \text{if spread}_t > 1 \end{cases}, \qquad (3.13) \\ \text{Vsign}_t &= \text{Vsame}_t^3 + \text{Vsame}_t^4 + ... + \text{Vsame}_t^{10}. \end{aligned}$$

[9] For the sake of brevity we did not report the marginal effects, but only report the median coefficient for the statistically significant stocks. Note that the marginal effect of an order submitted at the quotes (category 4) is positively related to the coefficient reported.

Table 3.3: Analysis of Depth Beyond the Best Quotes
The table presents the results of the depth analysis using different cutoff values. Vcomp=Vsame^j + ...+Vsame$^{\text{cutoff}}$, where $j = 1$, if spread/tick> 1, $j = 2$ otherwise. Whereas, Vsign=Vsame$^{\text{cutoff}+1}$ + ...+Vsame^{10}. Vcompopp and Vsignopp are constructed analogously for the opposite side of the book. All of the volume variables are scaled by 1e-6. Vola is the EWMA volatility (multiplied by 1000), Trend is the previous price change of 60 observations (multiplied by 1000), SPR is the (tick adjusted) inside spread, calculated as the difference between the best ask and bid quotes. The median, the percentage of positive coefficients given that they are significant, and finally the percentage of stocks with a statistically significant slope at a 5% level are reported.

BUY		**Vola**	**Trend**	**SPR**	**Vcomp**	**Vcompopp**	**Vsign**	**Vsignopp**
cutoff=2	Median	0.04	0.82	0.78	0.67	0.08	-0.10	0.07
	Sig. (%)	50	73	97	80	27	60	37
	Pos. (%)	93	100	100	100	75	6	100
cutoff=3	Median	0.04	0.83	0.81	0.47	0.07	-0.14	0.05
	Sig. (%)	47	77	97	67	47	53	27
	Pos. (%)	93	100	100	95	79	0	88
cutoff=4	Median	0.04	0.82	0.84	0.39	0.09	-0.11	0.02
	Sig. (%)	47	77	97	57	43	43	40
	Pos. (%)	93	100	100	82	92	15	58
cutoff=5	Median	0.04	0.82	0.82	0.32	0.15	-0.09	0.06
	Sig. (%)	47	80	97	53	53	47	37
	Pos. (%)	93	100	100	75	100	29	64
SELL								
cutoff=2	Median	0.05	-0.62	0.66	1.30	0.11	-0.19	0.05
	Sig. (%)	43	60	100	83	27	83	47
	Pos. (%)	85	6	100	100	63	8	79
cutoff=3	Median	0.05	-0.63	0.68	0.68	0.08	-0.24	0.06
	Sig. (%)	47	57	97	77	50	77	43
	Pos. (%)	79	6	100	96	60	0	69
cutoff=4	Median	0.05	-0.64	0.65	0.21	0.10	-0.34	0.04
	Sig. (%)	43	57	97	67	43	67	47
	Pos. (%)	77	6	100	85	92	0	64
cutoff=5	Median	0.05	-0.65	0.65	0.12	0.07	-0.44	0.01
	Sig. (%)	43	57	97	57	57	60	40
	Pos. (%)	77	6	100	59	94	0	58

- Step 2– order choice of impatient traders:

$$\text{Vsign}_t = \text{Vsame}_t^2 + \text{Vsame}_t^3 + ... + \text{Vsame}_t^{10}. \tag{3.14}$$

where Vsame^i is the total volume of orders waiting at the i^{th} best quote. Competition and signaling effects for the opposite side of the book, Vcompopp and Vsignopp are constructed analogously.

3.4.2 Covariates for the impact of non-walking through the book

In markets where walking through the book is allowed, an aggressive (category 1) market order has to walk up or down the order book to be fully executed. For markets in which walking through the book is not allowed, any excess that cannot be executed at the pre-specified limit price joins the queue at the quoted price instead of walking through and executed with less favorable prices. By focusing on the order choice of a market order trader, we test the relevance of price information while fitting her order size when walking through the book is not allowed. In addition to the depth variables, we define the inside spread and the price distance variables.

i) The (tick adjusted) inside spread, calculated as the difference between the best ask and bid quotes.

ii)
- The (tick adjusted) price distance between the best and the second best quotes for the opposite and the same sides of the book.
- The (tick adjusted) price distance between the second best ask (bid) and the highest available ask (lowest available bid) quote for the opposite and the same sides of the book.

The spread and the price distance variables for the opposite side capture the (weighted) average execution price of an aggressive order for markets in which walking through is possible. Because, in that case, when a large buy (sell) market order is submitted, it will eat up all the available volume at the best ask (bid) and then move up (down) to the second best ask (bid), and if necessary move up to third after consuming the second, etc. Since the cost of a market order increases with $\text{Dopp}^{1\text{–}2}$ or/and $\text{Dopp}^{2\text{–}\max}$, this should lead to a submission of less aggressive market orders.

3.4.3 Additional explanatory variables

Besides our key explanatory variables discussed above, the current literature posits that the volatility and the previous price trend affect the order choice of an agent. We include these two variables in our analysis as explanatory variables.

Following Beber and Caglio (2005), we define the volatility as the exponential moving average of the last 60 mid-quote squared returns. The optimal decay factor λ is obtained via maximum likelihood estimation.[10]

$$\hat{\sigma}_t = \sqrt{\lambda \hat{\sigma}^2_{t-1} + (1-\lambda) r^2_{t-1}}. \tag{3.15}$$

Expected signs: While higher volatility implies a higher probability of execution, it also increases the adverse selection costs. Existing literature identifies a negative relationship between volatility and order aggressiveness. Foucault (1999), Wald and Horrigan (2005) and Goettler et al. (2009), among others, claim that in high volatility states, since the picking off risk increases, the aggressiveness of an incoming agent decreases.

An order submission strategy may also depend on recent movements in the price (Hall and Hautsch (2006)). We identify the previous price trend observed by the agents (Trend) as the change of the mid-quote prices for the last 60 observations at the time of the order arrival.

Expected signs: Given that a trader observes an increasing price trend upon arrival, this may indicate a possible future price increase as well. Since this movement will move the prices away from the current levels, a buy trader may interpret it as an increased non-execution risk of her limit order; hence, she prefers to submit a more aggressive order. This works opposite for the seller.

In all of the regressions, to control the seasonality on the arrival rate of orders,

[10] Riskmetric EWMA is a version of GARCH(1,1) where persistence parameters sum up to one and the constant term is equal to zero. In other words, the optimal decay parameter λ can be obtained by estimating the Integrated GARCH model.

we use time of the day dummy, indicating which half-an-hour of the day the order is submitted. Moreover, five previous lags of the dependent variables, determined by the Akaike information criterion (AIC) is included as control variables.[11]

3.5 Results

As mentioned in Section 3.2, the 30 stocks in our sample present a high degree of heterogeneity. Thus, we estimate the sequential ordered probit (SOP) regressions for each stock separately, for buy and sell sides of the market. All of the regressions include 5 lags of the dependent variable and the time-of-the day dummies. For the sake of brevity, those are not reported. We report the median, minimum, maximum and the 25$^{\text{th}}$ and the 75$^{\text{th}}$ percentile of the estimated coefficients, the percentage of statistically significant coefficients at 5% level, and the percentage of positive coefficients given that they are significant. Table 3.4, Table 3.5 and Table 3.6 present the results of the first stage, the second stage for a limit order trader, and the second stage for a market order trader of the SOP model, respectively. Table B.1 provides the description of the explanatory variables defined in Section 3.4 and Table B.2 provides a summary of the major findings.

3.5.1 Impact of depth at and beyond the best quotes

Table 3.4 reveals that an increase in the depth at the best quotes (Vcomp) is perceived as an increased competition and encourages traders to submit more market orders for both sides of the market. On the other hand, when competition on the opposite side of the book (Vcompopp) increases, agents predict that the market order arrivals increases on the opposite side of the book, implying an increased probability of execution for their limit orders, hence they submit more limit orders. These results

[11] While the Bayesian information criterion (BIC) chooses 5 as the optimal lag in the first stage and the second stage–limit order trader, it chooses 2 as the optimal lag in the second stage–market order trader. We perform a robustness analysis with optimal lags chosen by the BIC and conclude that the results are similar.

are consistent with the findings of Ranaldo (2004), Beber and Caglio (2005), and Pascual and Veredas (2009). Our results suggest that an increase in the volume of orders waiting beyond the best quotes (Vsign) is perceived as a disagreement on the current price and discourages the market order submissions. This signaling effect is more pronounced on the sell side of the book compared to the buy side. This contradicts with the results of Pascual and Veredas (2009) who find a positive relationship between the accumulated number of orders waiting from the second to the fifth best quotes and the arrival rate of market order traders. They conclude that this finding supports the "crowding-out" hypothesis of Parlour (1998).

Table 3.5 presents the regression results for a patient trader. It suggests that only the same side of the book matters for both, buyer and seller. Vcomp and Vsign has expected signs. An increase in the competition leads to a submission of aggressive limit orders to jump the queue, whereas an increase on the same side depth away from the quotes (Vsign) is perceived as a possible mispricing of the best quotes as Goettler et al. (2005) and Goettler et al. (2009) predict and lead to a submission of less aggressive limit orders.

Marginal effects regarding the depth variables reveal that the volume at the best quotes is particularly emphasized while determining the degree of patience of the incoming trader compared to depth beyond the best quotes. Furthermore, the competition effect is stronger compared to the signaling effect for both sides of the market in all stages of the SOP.

3.5.2 Impact of non-walking through the book

Table 3.6 shows that, while fitting the size of her market order for an impatient trader, none of the price information, neither spread nor price distance variables, matter. This is intuitive, since when walking through the book is not allowed, the spread and the price distance variables for the opposite side do not alter the execution

Table 3.4: First Stage Sequential Ordered Probit

The table presents the results of the first stage of the two-stage sequential ordered probit model. The dependent variable is equal to 1 if the incoming trader is impatient (submits a market order, MO), and 0 otherwise. Vola is the EWMA volatility (multiplied by 1000), Trend is the price change of the last 60 observations (multiplied by 1000), SPR is the (tick adjusted) inside spread, calculated as the difference between the best ask and bid quotes, Vcomp (Vcompopp) is the volume accounting for the competition effect on the same (opposite) side of the book, Vsign (Vsignopp) is the volume accounting for the signaling effect on the same (opposite) side of the book as defined in equation (3.12). All of the volume variables are scaled by 1e-6. Dsame$^{1-2}$ is the price distance between the best and the second best quotes, whereas Dsame^{2-max} is the price distance between the second best ask (bid) and the highest available ask (lowest available bid) quote for the same side of the book. Dopp$^{1-2}$ and Dopp^{2-max} are constructed analogously for the opposite side of the book. All of the regressions include 5 lags of the dependent variable and the time-of-the day dummies. For the sake of brevity, those are not reported. The median, minimum, maximum and the 25th and the 75th percentile of the estimated coefficients, the percentage of positive coefficients given that they are significant, and finally the percentage of stocks with a statistically significant slope at a 5% level are reported. The cross sectional median of marginal effects (scaled by 1e3) is also reported.

Buy	**Vola**	**Trend**	**SPR**	**Vcomp**	**Vcompopp**	**Vsign**	**Vsignopp**	**Dsame$^{1-2}$**	**Dsame^{2-max}**	**Dopp$^{1-2}$**	**Dopp^{2-max}**
Median	-0.02	-1.08	-0.31	1.64	-1.95	-0.02	0.00	-0.14	-0.01	-0.14	-0.02
Min.	-0.09	-5.67	-4.46	0.08	-8.03	-0.35	-0.56	-0.83	-0.04	-0.96	-0.75
P25	-0.04	-1.55	-0.40	0.80	-3.76	-0.06	-0.08	-0.40	-0.02	-0.36	-0.04
P75	0.01	-0.62	-0.23	3.53	-1.03	0.04	0.05	-0.09	0.00	-0.04	0.00
Max.	0.05	0.87	-0.04	7.40	-0.16	0.67	0.39	0.18	0.02	0.50	0.05
Sig. (%)	63	83	80	100	100	60	53	40	43	30	73
Pos. (%)	16	0	0	100	0	33	56	8	15	22	32
Marginal Effects–median											
MO	-9.28	-406.55	-121.38	650.20	-746.44	-6.21	-0.56	-54.55	-1.99	-54.80	-6.28
Sell											
Median	-0.03	1.02	-0.37	1.76	-1.77	-0.14	0.01	-0.15	0.00	-0.25	-0.01
Min.	-0.10	-0.14	-1.26	0.14	-7.89	-0.93	-0.51	-1.02	-0.64	-0.75	-0.12
P25	-0.04	0.59	-0.41	0.73	-3.32	-0.28	-0.12	-0.41	-0.02	-0.35	-0.02
P75	-0.02	1.36	-0.24	4.15	-0.77	-0.05	0.07	-0.02	0.00	0.04	0.00
Max.	0.05	4.85	-0.05	9.97	-0.14	0.05	0.39	0.73	0.03	0.26	0.02
Sig. (%)	67	80	83	100	100	77	70	40	30	43	50
Pos. (%)	5	100	0	100	0	4	57	25	44	0	20
Marginal Effects–median											
MO	-8.76	360.78	-126.33	624.97	-620.37	-50.47	2.59	-45.69	-0.88	-55.83	-2.97

Table 3.5: Second Stage Sequential Probit–Patient Traders

The table presents the results of the second stage of the two-stage sequential ordered probit model for patient traders. Given the trader is patient, the dependent variable is equal to 1, 2 or 3 if the trader submits a category 5, category 4 or category 3 order (limit price within, at or away from the best quotes), respectively. Vcomp (Vcompopp) is the volume accounting for the competition effect on the same (opposite) side of the book, Vsign (Vsignopp) is the volume accounting for the signaling effect on the same (opposite) side of the book as defined in equation (3.13). They are scaled by 1e-6. The rest of the explanatory variables are defined in Table 3.4. All of the regressions include 5 lags of the dependent variable and the time-of-the day dummies. For the sake of brevity, those are not reported. The median, minimum, maximum and the 25$^{\text{th}}$ and the 75$^{\text{th}}$ percentile of the estimated coefficients, the percentage of positive coefficients given that they are significant, and finally the percentage of stocks with a statistically significant slope at a 5% level are reported. The cross sectional median of marginal effects (scaled by 1e3) is also reported.

Buy	**Vola**	**Trend**	**SPR**	**Vcomp**	**Vcompopp**	**Vsign**	**Vsignopp**	**Dsame^{1_2}**	**Dsame^{2_max}**	**Dopp^{1_2}**	**Dopp^{2_max}**
Median	0.02	0.67	0.78	0.52	0.08	-0.07	0.01	0.03	0.00	0.03	0.00
Min.	-0.10	-0.14	0.08	-0.07	-0.52	-0.72	-0.20	-1.26	-0.04	-2.40	-0.06
P25	-0.01	0.42	0.56	0.20	-0.04	-0.24	-0.01	-0.31	0.00	-0.20	-0.01
P75	0.04	1.16	0.86	1.43	0.30	-0.01	0.11	0.47	0.02	0.22	0.02
Max.	0.14	2.10	1.79	4.17	2.23	0.54	0.56	1.91	0.04	5.39	1.59
Sig. (%)	50	73	97	80	27	60	37	60	50	27	43
Pos. (%)	93	100	100	100	75	6	100	44	53	50	54
Marginal Effects–median											
LO–Above	-8.13	-259.55	-303.69	-204.17	-32.90	27.73	-5.44	-11.87	-1.46	-10.25	-0.22
LO–At	7.57	250.95	287.37	194.83	32.33	-26.76	5.33	11.31	1.41	9.85	0.21
LO–Within	0.25	9.09	12.02	5.96	0.56	-0.65	0.14	0.28	0.07	0.50	0.01
Sell											
Median	0.02	-0.42	0.66	0.58	0.02	-0.08	0.02	0.00	0.00	0.26	0.00
Min.	-0.05	-2.49	0.20	-0.03	-1.28	-1.33	-0.59	-1.00	-0.05	-0.73	-0.06
P25	0.00	-0.94	0.53	0.19	-0.03	-0.36	-0.01	-0.41	-0.01	-0.07	-0.01
P75	0.05	-0.10	0.84	1.88	0.11	-0.03	0.11	0.54	0.02	0.41	0.00
Max.	0.22	0.50	1.18	5.56	0.96	0.65	0.53	1.39	1.67	5.00	0.04
Sig. (%)	43	60	100	83	27	83	47	40	37	40	30
Pos. (%)	85	6	100	100	63	8	79	42	55	75	56
Marginal Effects–median											
LO–Above	-6.45	164.39	-264.32	-229.53	-8.61	32.53	-8.54	-0.31	-0.39	-101.07	-0.38
LO–At	6.26	-158.65	256.27	221.10	8.01	-32.08	8.45	0.31	0.38	99.31	0.35
LO–Within	0.14	-4.70	8.52	5.93	0.11	-0.89	0.08	0.00	0.01	1.80	0.02

price of a large market order compared to a small one. To analyze this further, we first test the joint significance of these variables and second, we use a different proxy to capture the price and volume information contained beyond the best quotes.

For the majority of the stocks, we cannot reject the null hypothesis $\gamma_{\mathrm{SPR}} = \gamma_{\mathrm{Dopp^1\_2}} = \gamma_{\mathrm{Dopp^2\_max}} = 0$ with a median $\chi^2 = 4.63$ (p-val=0.1759) and $\chi^2 = 2.88$ (p-val=0.4112) for buy and sell sides, respectively, where γ is defined in equation (3.7). This suggests that the price information contained in the limit order book is even jointly uninformative for a market order trader. As a different proxy, we fit a second degree polynomial for the total volume available at each price and the corresponding quotes. Then the coefficients of the quadratic term for both sell and buy sides of the book are used in the SOP regressions. As expected, the fit of the quadratic trend for the same and the opposite sides of the book are insignificant at 5% level.

Our results suggest that a market order trader only considers volatility, previous price trend, and volume accumulated beyond the best quotes on the opposite side of the book. In high volatility states an impatient trader submits more aggressive market orders. This can be explained by two: first, an impatient trader may benefit from a high volatility state since it increases the probability of fully execution of large size orders. This is due to the fact that the excess is converted to a limit order and the execution probability of a limit order increases with volatility.[12] This result is consistent with findings of Hall and Hautsch (2006). In their analysis conducted on Australian Stock Exchange, another market with non-walking through the book, they focus only on the aggressive market and limit orders. Their results suggest that high volatility states increase the arrival rate of aggressive market orders. Second, given that the trader submits a market order in a high volatility state, it is more likely that

[12]For example Cho and Nelling (2000) show that execution probability of limit orders are increasing with volatility.

Table 3.6: Second Stage Sequential Probit Regressions–Impatient Traders

The table presents the results of the second stage of the two-stage sequential ordered probit model. Given the trader is impatient, the dependent variable is equal to 0 if she submits a small market order (MO) (category 2 order) or equal to 1 if she submits a large MO (category 1 order). Vcomp (Vcompopp) is the volume accounting for the competition effect on the same (opposite) side of the book, Vsign (Vsignopp) is the volume accounting for the signaling effect on the same (opposite) side of the book as defined in equation (3.14). They are scaled by 1e-6. The rest of the explanatory variables are defined in Table 3.4. All of the regressions include 5 lags of the dependent variable and the time-of-the day dummies. For the sake of brevity, those are not reported. The median, minimum, maximum and the 25th and the 75th percentile of the estimated coefficients, the percentage of positive coefficients given that they are significant, and finally the percentage of stocks with a statistically significant slope at a 5% level are reported. The cross sectional median of marginal effects (scaled by 1e3) is also reported.

Buy	**Vola**	**Trend**	**SPR**	**Vcomp**	**Vsign**	**Vsignopp**	**Dsame^{1}–^{2}**	**Dsame^{2}–^{max}**	**Dopp^{1}–^{2}**	**Dopp^{2}–^{max}**
Median	0.18	-1.01	-0.12	-0.14	-0.05	-0.09	-0.04	0.01	-0.16	0.00
Min.	0.10	-5.57	-0.89	-2.54	-0.98	-1.62	-0.97	-0.05	-0.86	-0.65
P25	0.15	-1.52	-0.22	-0.31	-0.14	-0.69	-0.16	0.00	-0.35	-0.05
P75	0.22	-0.41	0.01	0.03	0.03	-0.04	0.08	0.04	0.11	0.02
Max.	0.41	0.56	0.46	2.79	0.26	0.46	0.53	0.09	0.56	0.11
Sig. (%)	100	67	3	27	33	70	10	27	23	47
Pos. (%)	100	0	0	13	50	5	0	88	29	43
Marginal Effects–median										
Large MO	23.29	-134.84	-14.43	-8.13	-5.55	-12.14	-5.81	1.44	-12.59	0.60
Sell										
Median	0.19	1.20	-0.01	-0.51	-0.08	-0.09	0.01	0.01	-0.10	-0.01
Min.	0.10	-0.18	-0.61	-4.22	-1.37	-1.07	-0.81	-0.17	-1.56	-0.09
P25	0.17	0.44	-0.13	-0.98	-0.27	-0.27	-0.15	-0.01	-0.41	-0.03
P75	0.24	2.10	0.12	-0.12	-0.05	0.00	0.26	0.02	0.12	0.01
Max.	0.63	5.00	0.81	0.32	0.02	0.74	0.57	0.05	0.80	0.07
Sig. (%)	100	67	7	63	53	50	13	10	37	13
Pos. (%)	100	100	50	0	0	7	0	33	18	25
Marginal Effects–median										
Large MO	31.36	181.68	-2.50	-71.17	-12.70	-19.65	1.68	1.02	-8.24	-0.93

she is informed as Goettler et al. (2009) predict. She would like to take advantage of the mispricing at the quotes, which makes her to submit an aggressive market order.

The accumulated volume of orders on the opposite side of the book (Vsignopp) and the change of the mid-quote prices for the last 60 observations (Trend) are negatively related with the buy market order aggressiveness. In other words, an impatient buyer splits her orders into several small quantities rather than submitting a large market order when Vsignopp or Trend increases. Because, an increase in Vsignopp or Trend signals a possible future price increase, increasing the non-execution risk for the limit-order-converted-part of the aggressive market order. The opposite is true for the seller.

In comparison to the study of Pascual and Veredas (2009), which is conducted on the Spanish Stock Exchange, we have different results. The authors show that the spread and the price distances on the opposite side of the market matters for an impatient trader's decision. In addition, in his study on the Swiss Stock Exchange, Ranaldo (2004) demonstrates that the sensitivity of a large market order with respect to volatility is more negative compared to a small one. Thus, in high volatility states an impatient trader prefers to submit a small market order, which contradicts our finding. One plausible explanation of the discrepancy in the results could be the walking through the book mechanism, which is allowed in both of the markets.[13]

3.5.3 Effects of the additional variables

In line with the existing literature, we find that the probability of an incoming agent being patient increases with volatility, since the picking off risk increases in high volatility states. On the other hand, Table 3.5 shows that, given that the agent is patient and submits a limit order, she prefers to submit more aggressive limit orders

[13]Non-walking through the book is not the only difference between the ISE and the other markets mentioned. Hence, we can only conjecture that the findings might be driven by non-walking through the book.

when volatility is higher since submitting orders away from the quotes decreases the execution probability significantly.[14] This result is weak for both sides of the market.

Our results suggest that, when the previous price trend increases, a buyer submits more limit orders whereas a seller submits more market orders. This contradicts the expected sign proposed. One possible interpretation is the expectation of mean reversion in the prices. If a seller, for instance, believes that prices will revert back, she would submit an aggressive market order to take advantage of this "mispricing", instead of waiting and to be compensated by a limit order.

Consistent with the majority of the literature, the first stage SOP regressions show that when spread is wider the arrival rate of patient traders increases. On the other hand, Table 3.5 shows that, the importance of the inside spread is more pronounced for the limit order trader while positioning their limit price. We find that a wider spread persuades patient traders to compete more heavily to jump the queue when spreads are wide, which confirms the predictions of Foucault et al. (2005) and Goettler et al. (2005).

3.5.4 Trading behavior of institutions

The current literature points out that individual and institutional investors may differ in their level of information implying that institutions are informed traders (Lo and MacKinlay (1990), Cornell and Sirri (1992), Koski and Scruggs (1998), and Chakravarty (2001)). In our data we can distinguish whether an order is initiated by an institutional or individual investor, with a limitation however. Due to internal regulations, some of foreign institutional investors are classified as individual instead of institution. Thus, whenever it is marked as an institutional investor in our data set, it is an institutional investor for sure. However, individual traders are pooled with

[14]For instance, Table 3.2 suggests that submitting an order away from the quotes instead of at the quotes decreases the execution probability from 60% to 20%.

foreign institutions.[15] This in turn reduces our sample size significantly, but does not affect the conclusions we derived. In our sample, on average 3.7% of all orders are initiated by institutional investors.

In order to formally test whether we can separate the sample as individual and institutional trading, we run the following two-stage sequential ordered probit (SOP) regression for both buy and sell sides of the market and test the null hypothesis $\mu = \gamma_1 = \gamma_2 = ... = \gamma_K = 0$.

$$Y_t^* = \alpha + \mu D_{t-1}^{\text{INS}} + \sum_{k=1}^{K} \beta_k X_{k,t-1} + \sum_{k=1}^{K} \gamma_k D_{t-1}^{\text{INS}} X_{k,t-1} + \varepsilon_t, \tag{3.16}$$

where Xs are the observable (limit order book) variables defined in Section 3.4, and $Y_{s,t}^*$ is the dependent variable introduced in equation (3.2). We define a dummy variable, D^{INS} which takes the value 1 if the order is initiated by an institutional trader.[16] The hypothesis is rejected at 5% of significance level with a median $\chi^2 = 46.65$ (p-val=0.0009) for 76% of the stocks for the sell side of the market. Similar conclusion holds for the buy side of the market. The joint hypothesis is rejected for the 83% of the stocks with a median $\chi^2 = 41.49$ (p-val=0.0000). These reveal that the information contained in the limit order book affects the trading behavior of institutions and individuals differently.

Following this, we separate the sample into two groups: orders initiated by institutional investors and by individual investors and re-run the first stage SOP regressions introduced in equation (3.2) for each of the groups separately. The results for the sell side of the market are presented in Table 3.7. Buy side results are qualitatively similar. The same explanatory variables, introduced in Section 3.4, are employed as in the analysis using the whole sample. The dependent variable is equal to 1 if the

[15] According to the information provided on the web page of the ISE, for the June and July 2008, on average, 10% of the trading value is originated by foreign investors. The maximum and minimum ratios are around 30% and 2%, respectively.

[16] It is not possible to run this regression for one of the stocks in our sample (IHLAS) due to limited number of observations. Hence, we excluded that stock from our analysis in this section.

Table 3.7: First Stage Sequential Ordered Probit–Institutional vs Individual Investors

The table presents the results of the first stage of the two-stage sequential ordered probit model for institutional (INS) and individual (IND) investors for the sell side of the market. For both set of regressions, the dependent variable is equal to 1 if the incoming trader is impatient (submits a market order, MO), and 0 otherwise. All of the explanatory variables are defined in Table 3.4. All of the regressions include 5 lags of the dependent variable and the time-of-the day dummies. For the sake of brevity, those are not reported. The median, minimum, maximum and the 25th and the 75th percentile of the estimated coefficients, the percentage of positive coefficients given that they are significant, and finally the percentage of stocks with a statistically significant slope at a 5% level are reported. The cross sectional median of marginal effects (scaled by 1e3) is also reported.

INS	**Vola**	**Trend**	**SPR**	**Vcomp**	**Vcompopp**	**Vsign**	**Vsignopp**	**Dsame$^{1-2}$**	**Dsame^{2-max}**	**Dopp$^{1-2}$**	**Dopp^{2-max}**
Median	-0.05	1.06	-0.41	2.51	-2.78	0.04	0.02	-0.06	-0.01	0.19	0.01
Min.	-0.17	-2.02	-5.39	0.40	-13.45	-0.93	-1.24	-1.66	-2.09	-1.63	-0.20
P25	-0.09	0.04	-0.82	1.44	-4.54	-0.11	-0.18	-0.56	-0.05	-0.88	-0.03
P75	0.06	1.72	0.10	5.14	-1.63	0.17	0.10	0.17	0.03	0.57	0.05
Max.	0.12	5.59	4.08	14.84	-0.48	1.97	0.69	0.76	0.15	2.01	0.24
Sig. (%)	10	24	3	93	83	14	28	7	21	10	38
Pos. (%)	33	100	0	100	0	75	38	0	17	0	55
Marginal Effects–median											
MO	-17.80	403.50	-137.50	962.00	-929.50	10.36	7.70	-23.70	-2.67	72.90	4.19
IND											
Median	-0.03	1.05	-0.35	1.78	-1.74	-0.15	0.01	-0.12	0.00	-0.17	-0.01
Min.	-0.11	-0.13	-1.22	0.14	-8.03	-0.99	-0.52	-1.03	-0.60	-0.73	-0.12
P25	-0.05	0.59	-0.42	0.70	-3.51	-0.37	-0.12	-0.34	-0.02	-0.33	-0.02
P75	-0.02	1.37	-0.26	4.09	-0.75	-0.05	0.07	-0.01	0.01	0.02	0.00
Max.	0.05	4.80	-0.07	10.02	-0.14	0.04	0.40	0.73	0.03	0.26	0.02
Sig. (%)	70	87	77	100	100	83	70	37	40	43	50
Pos. (%)	5	100	0	100	0	4	57	27	42	0	20
Marginal Effects–median											
MO	-9.96	367.50	-125.00	632.50	-610.00	-52.00	1.93	-41.50	-0.26	-57.80	-3.11

incoming trader is impatient (submits a market order) and 0 if she submits a limit order.

When we examine the results for the sample of individual investors, we see that volatility, the previous price trend, the inside spread, the competition variables, and the signaling variables are highly significant at a 5% level. On the other hand, the regression results for institutions reveal that only the volume at the same and at the opposite side of the book, (Vcomp and Vcompopp), are significant for institutional investors. The joint hypothesis $\beta^{\mathrm{INS}}_{\mathrm{Vcomp}} = \beta^{\mathrm{INS}}_{\mathrm{Vcompopp}} = 0$ is rejected with a median $\chi^2 = 51.07$ (p-val=0.0000) for all of the stocks except one. In other words, competition matters in their decision to submit a limit or a market order. Other features of the results presented in Table 3.7 are worth to underline. Volatility is not informative for an informed agent. This may suggest that institutional traders do not face the picking off risk that drives them to submit more limit orders rather than a market order in high volatility states. Similarly, the signaling variables (Vsign and Vsignopp) are not informative as expected. Informed agents do not rely on the signaling on the current prices provided by the market. Finally, the coefficients on volatility, price trend, spread, signaling variables, and price distance variables are jointly insignificant for 62% of the stocks with a median $\chi^2 = 13.42$ (p-val=0.0967).

To sum up, we conclude that, similar to the individual investors, institutional investors consider the information provided by the limit order book while designing their trading strategies. However, their decision to submit a market or a limit order is based on only a few pieces of the limit order book information. They take into account other traders' actions only for competition. This suggests that institutional investors' order submission strategies are based on their own private valuations rather than the state of the book.

3.5.5 Robustness

We provide several robustness checks to conclude that our findings are not driven by an arbitrary choice. The first robustness check is related to the model specification. Instead of estimating the model with ordered probit, we use ordered logit. The second robustness checks are on the definitions of the transient volatility and the price trend. Throughout the paper, we proxy the price fluctuations by using the exponential-weighted moving average (EWMA) volatility and the price trend as the percentage change in the mid-quote prices for the last 60 observations. First, we re-estimate the optimal decay parameter λ by using 100 mid-quote returns instead of 60. Similarly, as a robustness check for the price trend, we employ different window sizes of 100 and 120. Moreover, we re-estimate the two-stage sequential ordered probit model with different transient volatility measures, namely the standard deviation and absolute value of the mid-quote changes of the previous 60, 100 and 120 orders prior to the order submission.

Table 3.8 presents the robustness test results for the first stage and second stage patient trader, whereas Table 3.9 reports the results for the second stage impatient trader for the sell side of the market. For the sake of brevity, buy side is not reported since the results are qualitatively similar. All of the results are qualitatively robust, except for the volatility in the second stage–limit order trader. To sum up, we conclude that all of our main findings are remarkably robust to different proxies.

Table 3.8: Robustness: First Stage and the Second Stage–Limit Order Trader
This table reports the results for the robustness analysis for the sell side of the market for the first stage and the second stage–limit order (LO) trader. The first three rows repeat the results for the benchmark model, whereas the following three rows present the results for the logistic regression (Logit). The robustness analyses on the definition of volatility (Vola_std60, Vola_abs60) and on the previous trend (Trend100) are provided. Vola_std60 is the standard deviation of the last 60 mid-quote returns. Vola_abs60 is the absolute change in the last 60 mid-quote prices and Trend100 is the previous price change of the last 100 observations. All of the regressions include 5 lags of the dependent variable and the time-of-the day dummies. For the sake of brevity, those are not reported. The median, the percentage of positive coefficients given that they are significant, and finally the percentage of stocks with a statistically significant slope at a 5% level are reported.

1st stage		**Vola**	**Trend**	**SPR**	**Vcomp**	**Vcompopp**	**Vsign**	**Vsignopp**
Benchmark	Median	-0.03	1.02	-0.37	1.76	-1.77	-0.14	0.01
	Sig. (%)	67	80	83	100	100	77	70
	Pos. (%)	5	100	0	100	0	4	57
Logit	Median	-0.05	1.67	-0.65	3.08	-3.08	-0.24	0.02
	Sig. (%)	67	80	83	100	100	77	67
	Pos. (%)	5	100	0	100	0	4	60
Vola_std60	Median	-0.06	1.02	-0.36	1.80	-1.78	-0.14	0.01
	Sig. (%)	77	80	83	100	100	80	70
	Pos. (%)	0	100	0	100	0	4	57
Vola_abs60	Median	-0.01	1.05	-0.36	1.79	-1.74	-0.13	0.01
	Sig. (%)	83	83	83	100	100	77	70
	Pos. (%)	0	100	0	100	0	4	57
Trend100	Median	-0.03	0.43	-0.36	1.71	-1.67	-0.13	0.01
	Sig. (%)	63	67	80	100	100	80	70
	Pos. (%)	5	100	0	100	0	8	62
2nd stage LO								
Benchmark	Median	0.02	-0.42	0.66	0.58	0.02	-0.08	0.02
	Sig. (%)	43	60	100	83	27	83	47
	Pos. (%)	85	6	100	100	63	8	79
Logit	Median	0.03	-0.68	1.02	0.94	0.04	-0.14	0.04
	Sig. (%)	43	57	93	83	30	83	47
	Pos. (%)	85	6	100	100	44	8	79
Vola_std60	Median	0.02	-0.40	0.67	0.58	0.01	-0.08	0.02
	Sig. (%)	33	60	97	83	33	83	47
	Pos. (%)	90	6	100	100	50	8	79
Vola_abs60	Median	0.00	-0.41	0.66	0.56	0.00	-0.09	0.02
	Sig. (%)	37	60	97	83	30	83	47
	Pos. (%)	64	6	100	100	44	8	79
Trend100	Median	0.02	-0.14	0.67	0.63	0.02	-0.08	0.02
	Sig. (%)	43	47	100	90	27	80	47
	Pos. (%)	85	29	100	100	38	8	79

Table 3.9: Robustness: Second Stage–Market Order Trader

This table reports the results for the robustness analysis for the sell side for the second stage–market order (MO) trader. The first three rows repeat the results for the benchmark model, whereas the following three rows present the results for the logistic regression (Logit). The robustness analyses on the definition of volatility (Vola_100, Vola_std60, Vola_std100, Vola_abs60, Vola_abs100) and on the previous trend (Trend100) are provided. Vola_100 is the exponential moving average of the previous 100 squared returns with optimal decay parameter. Vola_std60 (Vola_std100) is the standard deviation of the last 60 (100) mid-quote returns. Vola_abs60 (Vola_abs100) is the absolute change in the last 60 (100) mid-quote prices and Trend100 is the previous price change of the last 100 observations. All of the regressions include 5 lags of the dependent variable and the time-of-the day dummies. For the sake of brevity, those are not reported. The median of the estimated coefficients, the percentage of statistically significant coefficients at 5% level, and the percentage of positive coefficients given that they are significant are provided.

2nd stage MO		**Vola**	**Trend**	**SPR**	**Vcomp**	**Vsign**	**Vsignopp**	**Dsame^{1}_^{2}**	**Dsame^{2}_^{max}**	**Dopp^{1}_^{2}**	**Dopp^{2}_^{max}**
Benchmark	Median	0.19	1.20	-0.01	-0.51	-0.08	-0.09	0.01	0.01	-0.07	-0.01
	Sig. (%)	100	67	7	63	53	50	13	10	37	13
	Pos. (%)	100	100	50	0	0	7	0	33	18	25
Logit	Median	0.37	2.06	-0.05	-1.00	-0.17	-0.17	0.05	0.01	-0.15	-0.01
	Sig. (%)	100	67	7	57	50	50	13	20	43	10
	Pos. (%)	100	100	50	0	0	7	0	50	23	33
Vola_100	Median	0.23	1.20	0.00	-0.32	-0.09	-0.10	0.02	0.01	-0.08	-0.01
	Sig. (%)	100	67	3	47	50	50	10	20	30	13
	Pos. (%)	100	100	0	0	0	7	0	50	22	25
Vola_std60	Median	0.22	1.09	-0.02	-0.41	-0.08	-0.09	0.01	0.01	-0.07	-0.01
	Sig. (%)	100	67	7	43	50	50	13	23	37	13
	Pos. (%)	100	100	50	0	0	7	0	43	18	25
Vola_std100	Median	0.22	1.21	0.00	-0.40	-0.09	-0.09	0.02	0.01	-0.07	-0.01
	Sig. (%)	100	67	7	47	47	43	7	20	30	10
	Pos. (%)	100	100	50	0	0	0	0	50	22	33
Vola_abs60	Median	0.02	1.18	0.00	-0.55	-0.12	-0.13	0.07	0.01	-0.03	-0.01
	Sig. (%)	77	67	7	70	63	50	10	13	30	17
	Pos. (%)	100	100	50	0	0	7	0	25	22	20
Vola_abs100	Median	0.01	1.37	0.01	-0.54	-0.12	-0.15	0.10	0.01	-0.04	-0.01
	Sig. (%)	70	67	7	70	57	53	7	13	23	30
	Pos. (%)	100	100	50	0	0	13	0	25	29	11
Trend100	Median	0.20	0.62	-0.01	-0.36	-0.12	-0.11	0.03	0.01	-0.08	-0.01
	Sig. (%)	100	57	7	57	53	50	10	17	30	17
	Pos. (%)	100	94	50	0	0	13	0	20	22	20

3.6 Conclusion

This paper investigates how the information content of a limit order book affects the order choice of an investor. By employing a two-stage sequential ordered probit model, we first answer whether the competition or signaling effects dominate each other. Second, we examine the order decision of a trader under the non-walking through the book mechanism. Finally, we study the trading behavior of institutional and individual investors separately.

By reconstructing the limit order book for the Istanbul Stock Exchange, we show that the competition effect is present only at the best quotes while determining the arrival rate of a market or a limit order. On the other hand, a patient trader perceives an increase in the depth up to the second best quotes as an increased competition and submits a more aggressive limit order. An increase in the same-side-depth behind the top of the book is perceived as a signal of a possible mispricing of the current quotes and encourages agents to submit less aggressive orders. This is consistent with the predictions of Goettler et al. (2005) and Goettler et al. (2009). We show that, at every stage, the competition effect is stronger than the signaling effect.

In our market, in her decision to submit a “large” or “small” market order, only volatility, previous price trend and volume accumulated on the opposite side of the book matter for an impatient trader. In other words, none of the price information affects the order choice of an impatient trader. This result might be explained by the non-walking through the book property of our market. Because under this mechanism, the spread and the price distance variables do not capture the execution price of an aggressive market order.

Finally, the results show that institutional investors trading strategies are affected by fewer pieces of the limit order book information compared to individual investors. An institutional investor considers other traders’ actions only for competition and signaling does not influence her order choice. Moreover, since they have informational advantages over individual investors, they do not face the picking off risk that makes the market order trading more costly in high volatility states.

Chapter 4

Implied Correlation and Expected Returns

4.1 Introduction

Correlation is one of the most important concepts in an extensive variety of theories and applications in finance. Over the last two decades, empirical research has documented evidence that correlation among assets changes over time and rises during periods of market downturn, diminishing the portfolio diversification benefits in times they are most needed.[1] Hence, it is natural to examine whether correlation is an important indicator of market-wide risk. The volatility of the market is a key determinant of aggregate risk; an increase in market volatility leads to an increase in the market expected return, since risk averse investors demand a higher risk premium to hold the market portfolio when systematic risk rises. As changes in correlation induce changes in market volatility, it is expected that they also affect the time variation of the market equity premium.

In this paper, I investigate whether changes in implied correlation induces changes in the market equity premium. Using option data of the S&P100 index and its individual constituents, I extract information of second moments following Martin (2011)'s approach to construct the aggregate implied correlation. The motivation of employing

[1] Relevant studies are Roll (1988), Bollerslev, Engle and Wooldridge (1988), Ho, Stapleton and Subrahmanyam (1995), Longin and Solnik (2001), Ang and Chen (2002), Moskowitz (2003), Brandt and Diebold (2006), and Hong et al. (2007), among others.

a forward-looking measure relies on the advantages of extracting information about equilibrium stock prices from option data. Bates (1991) suggests that option prices reflect the market participants' expectations by giving a direct indication of the aggregate subjective distributions of investors. Moreover, they provide information that is not fully captured by historical prices. As Buss and Vilkov (2011) point out, option prices update faster to new market conditions, since historical data have some inertia incorporated. Furthermore, when estimating risk-neutral expectations of higher moments using options, we do not face the trade-off between using long time-series of data to obtain precise estimates and short windows to produce conditional instead of unconditional estimates.[2]

I show that aggregate implied correlation is highly and positively related to both monthly and cumulative subsequent market returns. The relationship is stronger for intermediate prediction horizons, particularly at bimonthly, quarterly and semi-annually return horizons, and robust to the inclusion of standard predictors such as valuation ratios, business cycle variables, and second moments of the return distributions. The economic importance of implied correlation[3] is the highest compared to the other variables: a one standard deviation increase in implied correlation translates into a 1.31 percent increase in three-months-ahead monthly market returns. Furthermore, I find that these results are not driven by the recent financial crisis. My findings may indicate that periods of high market-wide correlation produce a deterioration of the investment opportunity set and, as a consequence, an increase in the equilibrium expected return.

This paper is part of a vast literature on the predictability of the market risk premium. Considerable effort has been dedicated to provide evidence that valuation ratios such as price-to-dividend and price-to-earnings ratios, and different business

[2]For further discussion see for instance Chang, Christoffersen and Jacobs (2013) and Conrad, Dittmar and Ghysels (2013)

[3]Implied correlation is used interchangeable with aggregate implied correlation.

cycle variables such as consumption over wealth, default spread, and term spread, among others, have predictive power for market returns (e.g. Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988), Campbell and Shiller (1989), Fama and French (1988), Fama and French (1989), Lettau and Ludvigson (2001), Lamont (1998), Lewellen (2004)). The forecasting power of second moments has also received attention in the academic literature. In a recent paper, Bollerslev, Tauchen and Zhou (2009) show that a high variance risk premium predicts high future stock market returns, especially at a quarterly return horizon. Their results depend decisively on the method employed: the "model-free" rather than the "Black-Scholes" options implied volatility, and the use of intraday data instead of daily frequency data to estimate realized volatilities. Following Bollerslev et al. (2009), and motivated by Driessen et al. (2009) who show that the pricing of index variance risk depends on the pricing of individual variance risk and correlation risk, Cosemans (2011) presents evidence that the predictive power of the market variance risk premium is mainly driven by the correlation risk premium and the systematic component of the average variance risk premium in individual constituents. The question of how changes in aggregate realized correlation affect expected returns on the market has been addressed by Pollet and Wilson (2010). They show that the variance of market returns is approximately equal to the product of average variance of individual stock returns and the average realized correlation. Their findings reveal that the physical correlation strongly predicts future market excess returns, whereas the average variance does not.

This paper contributes to this literature in different ways. First, I provide a new variable, the aggregate implied correlation, which contains information on future market returns that cannot be explained by the aforementioned predictors. Second, I document further evidence for Pollet and Wilson (2010)'s findings by using a forward-looking measure of correlation: I show that the relationship between the market variance and the product of individual variances and pairwise correlations also holds when using forward-looking estimates, indicating that changes in implied correlation induce changes in the market implied variance. Third, I document that the forecasting power of realized correlation is only present when the recent financial crisis is not included in the analysis, whereas that of implied correlation does not depend on the sample period. I show that the predictive power of realized correlation is no longer significant under the presence of aggregate implied correlation, suggesting that a forward-looking measure reveals more information on the expected market equity premium than its physical counterpart. Fourth, I extend Cosemans (2011)'s study by showing that the forecasting power of the correlation risk premium is driven by the information content of implied correlation.

This study also builds on the research on correlation of asset returns. Existing theoretical studies have considered the time-varying nature of correlation and its relationship with asset returns. Ang and Bekaert (2002) solve a dynamic portfolio choice problem under the presence of two i.i.d. regime switches. They identify a "bear" regime with lower conditional means, higher correlations, and higher volatilities and thus reproducing the asymmetric exceedence on correlations. Buraschi, Porchia and Trojani (2010) also examine the effect of stochastic volatility and correlations on optimal portfolio choice. By assuming that the covariance matrix follows a Wishart process, their estimations reveal that the hedging demands are considerably larger compared to models that consider constant correlations. Recent papers have addressed the dynamics of correlation in an endowment economy. In a two-trees Lucas (1978) economy Cochrane, Longstaff and Santa-Clara (2008) examine a simple equilibrium model with homogenous agents. They obtain closed-form solutions that allows to examine the implications of correlation among stocks. In a similar economy, Buraschi, Trojani and Vedolin (2011) analyze a model with heterogenous agents. They document that the larger the belief disagreement, the larger the correlation risk premium. Martin (2013), on the other hand, considers multiple assets

(Lucas Orchard) to analyze the behavior of asset prices allowing for rare disasters. His model solution leads to correlations that arise endogenously and increase in times of disasters. In this study I provide empirical evidence that changes in correlation affect stock returns in an aggregate perspective. I find a positive and significant relationship between implied correlation and future market excess returns in the time-series, which is consistent with risk averse investors demanding a higher risk premium to hold aggregate wealth in periods of high correlation.

Finally, this paper is also related to studies focusing on the risk-return trade-off of risk-neutral measures of second moments. Ang, Hodrick, Xing and Zhang (2006) examine the pricing of volatility risk by using the implied volatility index VIX and find a negative price of risk. Conrad et al. (2013) find that the risk-neutral volatility, skewness and kurtosis of individual assets are highly connected to future returns. Chang et al. (2013), on the other hand, document that the market skewness is negatively priced in the cross-section of stock returns, whereas the positive pricing of market kurtosis highly depends on the test methodology. My paper extends this literature by also examining how risk-neutral expectations of second moments affect the time-series of market returns.

The rest of this paper is organized as follows. Section 4.2 presents the construction of implied correlation, the predictive methodology, and data description. The main findings of the predictive regressions, analysis of the pre-crisis period, out-of-sample experiments, and the predictive power of correlation risk premium are reported in Section 4.3. Finally, Section 4.4 concludes.

4.2 Empirical Methodology and Data

One of the aims of this study is to examine the predictive power of implied correlation on market returns. Pollet and Wilson (2010) show that the realized correlation between stocks provides more information on true aggregate risk than the market

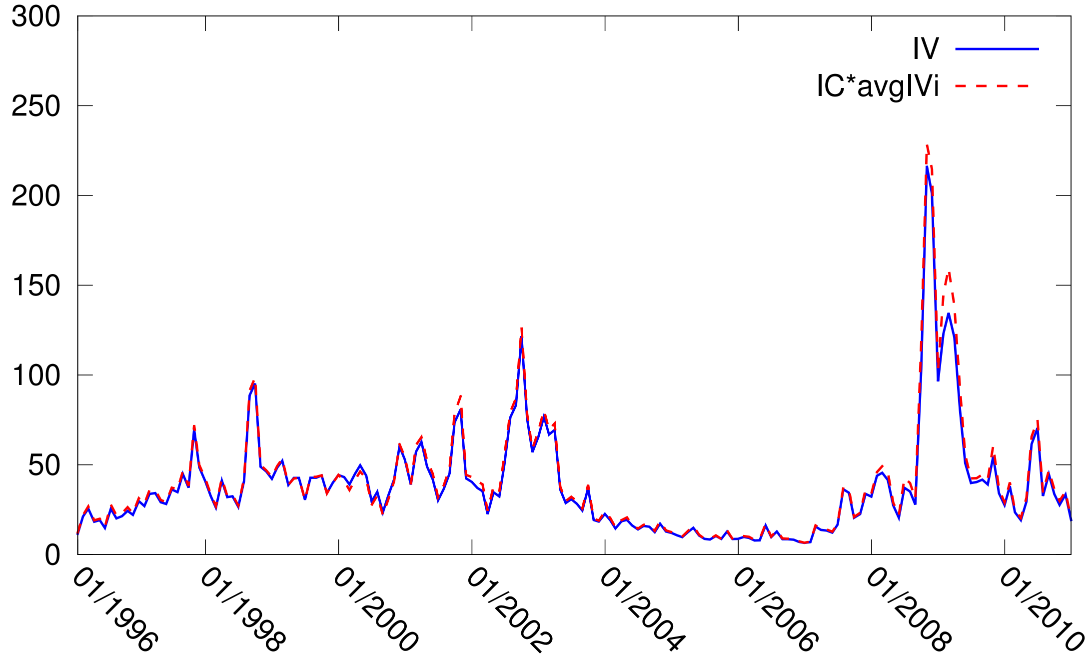


Figure 4.1: The figure presents monthly time-series estimates of the implied variance of the S&P100 index (IV) and the product between the implied correlation (IC) and the average implied variance of the S&P100 individual components (avgIVi), for the period from January 1996 to December 2010. Implied variances for the index and for all of the individual components are calculated as the risk-neutral expectations of the 30-days-ahead variance from equation (4.5). The implied correlation is obtained from equation (4.4) by using the implied variance estimates. Variances are reported in percentages squared

variance. They find that the realized variance of the S&P500 index is almost equal to the product between the average realized variance of the individual components and the average of pairwise realized correlations. In Figure 4.1, I show that this relation also holds almost perfectly when using risk-neutral measures. Changes in implied correlation induce changes in the market implied variance, also indicating that the option-implied correlation may reveal information on aggregate risk (the details of the construction of implied variances and correlation are provided below). Moreover, as forward-looking measures incorporate new market conditions quicker than historical estimates, I expect that implied correlation performs better in explaining future market risk premia compared to its physical counterpart.

This section presents the construction of aggregate implied correlation, the predictive regression methodology to forecast stock market returns, and a detailed description of the data used in the analysis. Summary statistics of the explanatory variables are also provided.

4.2.1 Construction of Implied Correlation

The instantaneous variance of the index at a given time t, σ^2_{It}, is a function of the instantaneous variance of individual constituents, σ^2_{it}, and the correlation between pairs of stock returns, ρ_{ijt},

$$\sigma^2_{It} = \sum_{i=1}^{N} w_i^2 \sigma^2_{it} + \sum_{i=1}^{N} \sum_{i \neq j} w_i w_j \sigma_{it} \sigma_{jt} \rho_{ijt}, \tag{4.1}$$

where w_{it} denotes the market weight of i^{th} component. From this equation, I can obtain an expression for the expected integrated variance under the risk-neutral probability measure $\mathcal{Q}$ over an interval of length $T - t$,

$$E_t^{\mathcal{Q}} \left[\int_t^T \sigma^2_{I\tau} \mathrm{d}\tau \right] = E_t^{\mathcal{Q}} \left[\int_t^T \sum_{i=1}^{N} w_i^2 \sigma^2_{i\tau} \mathrm{d}\tau \right] + E_t^{\mathcal{Q}} \left[\int_t^T \sum_{i=1}^{N} \sum_{i \neq j} w_i w_j \sigma_{i\tau} \sigma_{j\tau} \rho_{ij\tau} \mathrm{d}\tau \right]. \tag{4.2}$$

By assuming equal pairwise implied correlations between all of the pair stock returns, $\rho_{ij\tau} = \rho_\tau$, and given that it is not possible to estimate the second term of the previous equation, we can use the following approximation,

$$\begin{aligned} & E_t^{\mathcal{Q}} \left[\int_t^T \sum_{i=1}^{N} \sum_{i \neq j} w_i w_j \sigma_{i\tau} \sigma_{j\tau} \rho_{ij\tau} \mathrm{d}\tau \right] \\ & \approx \sum_{i=1}^{N} \sum_{i \neq j} w_i w_j \sqrt{E_t^{\mathcal{Q}} \left[\int_t^T \sigma^2_{i\tau} \mathrm{d}\tau \right]} \sqrt{E_t^{\mathcal{Q}} \left[\int_t^T \sigma^2_{j\tau} \mathrm{d}\tau \right]} E_t^{\mathcal{Q}} \left[\int_t^T \rho_\tau \mathrm{d}\tau \right]. \end{aligned} \tag{4.3}$$

Then, it is straightforward to derive the expression for aggregate implied correlation $\mathrm{IC}_t = E_t^{\mathcal{Q}} \left[\int_t^T \rho_\tau \mathrm{d}\tau \right]$ by rearranging the equations above,

$$\mathrm{IC}_t = \frac{E_t^{\mathcal{Q}} \left[\int_t^T \sigma^2_{I\tau} \mathrm{d}\tau \right] - \sum_{i=1}^{N} w_i^2 E_t^{\mathcal{Q}} \left[\int_t^T \sigma^2_{i\tau} \mathrm{d}\tau \right]}{\sum_{i=1}^{N} \sum_{i \neq j} w_i w_j \sqrt{E_t^{\mathcal{Q}} \left[\int_t^T \sigma^2_{i\tau} \mathrm{d}\tau \right]} \sqrt{E_t^{\mathcal{Q}} \left[\int_t^T \sigma^2_{j\tau} \mathrm{d}\tau \right]}}. \tag{4.4}$$

IC_t represents the market's expectation of future market-wide correlation, implied by option prices of the index and prices of options on its components. It summarizes the pairwise correlations among all the individual components. An increase in IC_t is associated with a deterioration of the market's expectations of the portfolio diversification benefits.

Estimation of IC_t

To calculate the implied variance of the index and implied variance of the index components, I use the risk-neutral variance of simple returns which can be estimated from the strike of a simple variance swap. Martin (2011) introduces this financial contract with different properties than a standard variance swap. For instance, simple variance swaps can be hedged in the presence of jumps and they measure the risk-neutral variance of simple returns. According to Martin (2011), it also provides a natural way to calculate implied correlations, since the decomposition of the index variance (equation (4.1)) refers to simple returns, not log returns.

Following this approach, the risk-neutral expectation of the integrated variance, $E_t^{\mathcal{Q}}\left[\int_t^T \sigma_\tau^2 \mathrm{d}\tau\right]$, is approximated as the strike of a simple variance swap defined as,

$$V(0,T) \equiv \frac{2\exp^{rT}}{F_T^2}\left(\int_0^{F_T} \mathrm{put}_T(K)\mathrm{d}K + \int_{F_T}^{\infty} \mathrm{call}_T(K)\mathrm{d}K\right), \tag{4.5}$$

where F_T denotes the underlying asset's forward price to time T at time 0, and $\mathrm{put}_T(K)$ and $\mathrm{call}_T(K)$ are the put and call prices with maturity date T and strike price K, respectively. The integral is defined over an infinite set of strike prices. By assuming that the available strike prices of the put options belong to the interval $[K^P_{min}, K^P_{max}]$ where $0 < K^P_{min} < K^P_{max} < +\infty$, I solve the integral numerically using the trapezoidal method. Thus, the first term on the right hand side of equation (4.5) is approximated as follows,

$$\frac{2}{F_T^2}\left(\int_{K^P_{min}}^{K^P_{max}} \mathrm{put}_T(K)\mathrm{d}K\right) \approx \frac{K^P_{max} - K^P_{min}}{m}\sum_{k=1}^{m}\left(\frac{\mathrm{put}_T(K_i)}{F_T^2} + \frac{\mathrm{put}_T(K_{i-1})}{F_T^2}\right). \tag{4.6}$$

In a similar manner, I numerically approximate the second term of the right hand side of equation (4.5) to finally obtain the estimates of the implied variance for the index and individual stocks.

I use daily option data provided by OptionMetrics for the S&P100 index and its constituents from January 1996 to December 2010. I collect the following information: expiration dates, strike prices, highest closing bid and lowest closing ask for both put and call options. Daily closing stock prices are obtained from the CRSP database. I filter the data based on the following criteria: First, I delete all double entries. Second, I remove all the observations with empty implied volatility, since this case corresponds to options with non-standard settlements, and third, I remove all entries with highest closing bid equal to zero. Similar to Martin (2011), I approximate the forward price to the spot price. The implied variance is estimated for different maturities and by interpolating I construct daily estimates with 30 days of time-to-maturity. Monthly time-series are given by the estimates at the end of each month. Once I approximate the implied variance for the index and its constituents, I finally obtain the aggregate implied correlation from equation (4.4).

4.2.2 Predictive regression methodology

The empirical methodology relies on a standard regression model of monthly market risk premium on the lagged implied correlation, and standard control predictors for different return horizons h,

$$rx^m_{t+h} = \alpha_1(h) + \alpha_2(h)\mathrm{IC}_t + controls + \varepsilon_t, \tag{4.7}$$

where rx^m_{t+h} is the monthly market return in excess of the monthly risk free rate at $t+h$, and *controls* is a set of control predictors described below. The coefficient of interest $\alpha_2(h)$ is expected to be positive. This is consistent with risk averse investors perceiving states of high market-wide correlation as an increase in aggregate risk, which induces future market returns to rise.

4.2.3 Data description

In order to construct the market risk premium, I use (log) market returns based on the CRSP NYSE/AMEX/NASDAQ value-weighted portfolio in excess of the one-month treasury bill (log) rate obtained from Kenneth French's online data library.[4]

Following the extant literature on stock market returns predictability, I consider three groups of control predictors: (i) second moments and variance risk premia, (ii) valuation ratios, and (iii) business cycle variables.

Second moments and variance risk premia

Following Pollet and Wilson (2010), I include the average realized correlation as a control variable. I calculate the sample correlation for each pair of constituents i and j of the S&P100 index, each month, as follows,

$$\hat{\rho}_{ijt} = \frac{\hat{\sigma}_{ijt}}{\hat{\sigma}_{it}\hat{\sigma}_{jt}}, \tag{4.8}$$

where $\hat{\sigma}_{it}$ is the realized volatility of stock i and $\hat{\sigma}_{ijt}$ is the covariance between stocks i and j. I compute the realized volatility for each month as the squared root of the realized variance,

$$\mathrm{RV}_{it} = \sum_{d=1}^{D_t} \left((1 + R_{id}) - \frac{1}{D_t} \sum_{d=1}^{D_t} (1 + R_{id}) \right)^2, \tag{4.9}$$

where R_{id} is the return of a trading day d, and D_t is the number of trading days in month t. Finally, I obtain the average realized correlation as the sum of the value-weighted pairwise correlations,

$$\mathrm{RC}_t = \sum_{i=1}^{N} \sum_{i \neq j} w_{it} w_{jt} \hat{\rho}_{ijt}. \tag{4.10}$$

Bollerslev et al. (2009) provide evidence of forecasting power of the market variance risk premium (VRP_t) on market returns. Following their study, I compute VRP_t

[4]Similar results are obtained when using the S&P500 and the S&P100 returns as proxies for aggregate market returns.

as the difference between the risk-neutral expectation and physical expectation of the market variance,

$$\mathrm{VRP}_t = E_t^{\mathcal{Q}}\left(\int_t^T \sigma_\tau^2 \mathrm{d}\tau\right) - E_t^{\mathcal{P}}\left(\int_t^T \sigma_\tau^2 \mathrm{d}\tau\right), \tag{4.11}$$

where the time horizon $T - t$ is 30 days. The implied variance is estimated by numerically solving the risk-neutral expectation of the simple return variance defined in equation (4.5) using the trapezoidal rule described in equation (4.6). The physical expectation of the market variance is approximated as the realized variation of the index from $t-1$ to t described in equation (4.9).[5]

Furthermore, Cosemans (2011) documents that the cross-sectional average of the variance risk premia ($\overline{\mathrm{VRPi}}_t$) of the S&P100 constituents is also highly related to the future market risk premium. In accordance with his study, I also include $\overline{\mathrm{VRPi}}_t$ as the value-weighted average of the variance premia on all the index constituents.

Valuation ratios

The price-to-dividend ($\mathrm{P}_t/\mathrm{D}_t$) and price-to-earnings ($\mathrm{P}_t/\mathrm{E}_t$) ratios have been widely recognized as predictors of market returns (see for instance Campbell and Shiller (1988), Campbell and Shiller (1989), Fama and French (1988), Lamont (1998), and Lewellen (2004)). Following these studies, I consider the log of $\mathrm{P}_t/\mathrm{D}_t$ and the log of $\mathrm{P}_t/\mathrm{E}_t$, obtained as the S&P500 price at the end of each month divided by its dividends per share and earnings per share accumulated over the last twelve months, respectively.

Business cycle variables

Lettau and Ludvigson (2001) show that fluctuations in the consumption–wealth ratio (CAY_t) are strong predictors of market excess returns. I construct monthly time-

[5] I repeat the analysis using intraday frequency data to estimate the realized variance of the market. The results are presented in Appendix C.

series on CAY_t using the most recently available observations. Quarterly estimates of the consumption–wealth ratio are obtained from Lettau's website.

The aggregate variation in stock market returns is also explained by variables widely used in bond returns (see for instance Keim and Stambaugh (1986), Campbell (1987), Fama and French (1989), among others). Following these studies, I include the default spread (DS_t), the term spread (TS_t), and the relative risk-free rate (RREL_t). DS_t is measured as the difference between BAA and AAA corporate bond spreads from the Federal Reserve Bank of St Louis. TS_t is calculated as the difference between the ten-year Treasury bond and the three-month Treasury bill yields obtained from the Federal Reserve Bank of St Louis. Finally, RREL_t is constructed as the one-month T-bill rate minus its trailing twelve month moving averages.

4.2.4 Descriptive analysis

Figure 4.2 presents implied and realized correlation estimates. Panels A and B of Table 4.1 report summary statistics and the unconditional correlation matrix, respectively. All of the variables are constructed in a monthly basis for the period from January 1996 to December 2010.

Figure 4.2 shows that the implied correlation is higher than the realized correlation for most of the sample period, which indicates a positive correlation risk premium. Consistent with other studies (for instance Cosemans (2011) and Driessen, Maenhout and Vilkov (2012)), the figure also reveals that correlation increases at stress times or during periods of market uncertainty. I observe that some of the peaks for both measures of correlation take place at events such as the LTCM default and Russian crisis in 1998, the Iraq war in 2003, and the recent financial crisis, with sharp increases around the collapse of Northern Rock in August 2007 and Lehmann Brothers in September 2008. In these events, I also observe that the difference between these two measures decreases, making the correlation risk premium negative in some periods.

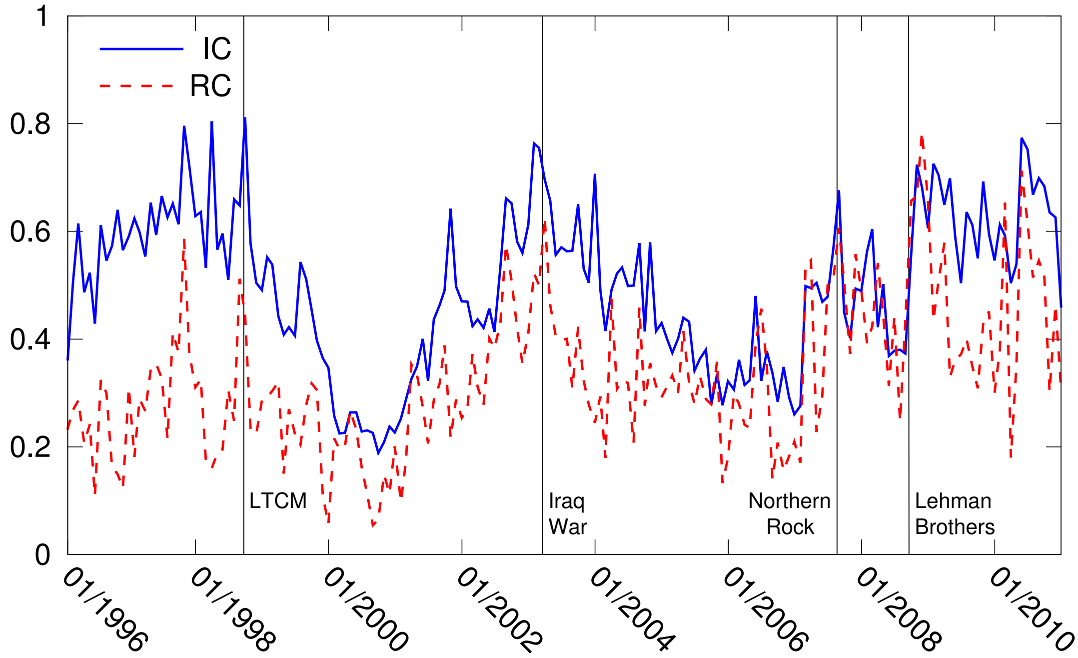


Figure 4.2: The figure presents monthly time-series estimates of the implied correlation (IC) and the realized correlation (RC) for S&P100 index, for the period from January 1996 to December 2010. The implied correlation is obtained at the end of each month using implied variances for the index and for all of the individual components, calculated as the risk-neutral expectations of the 30-days-ahead variance from equation (4.5). The realized correlation is obtained as the cross-sectional average of the pairwise return correlations of the index components over a one-month window as illustrated in equation (4.10).

Table 4.1 reports a monthly mean of 0.5, with value ranging from 0.19 to 0.81 and standard deviation of 0.14 for IC_t. The mean of the realized correlation is 0.33, which indicates a positive average correlation risk premium of approximately 17% for the sample period. All of the variables have positive autocorrelation coefficients, except for the second coefficient of the market excess return. IC_t, $\log(\mathrm{P}_t/\mathrm{D}_t)$, $\log(\mathrm{P}_t/\mathrm{E}_t)$, CAY_t, DS_t, TS_t, and RREL_t display a first autocorrelation coefficient of more than 0.8. In unreported results, the augmented Dickey-Fuller and Phillips-Perron Unit Root tests strongly reject the null of a unit root for IC_t. However, this is not the case for the valuation ratios, consumption over wealth, default spread and term spread. This indicates that the implied correlation exhibits faster mean reversion than these variables. Panel B shows that IC_t is correlated with DS_t, TS_t, and RREL_t, with correlation coefficients of 0.26, 0.47, and -0.31, respectively. This suggests that changes in implied correlation partly vary over the business cycle. Moreover, Panel B reveals that IC_t and $\log(\mathrm{P}_t/\mathrm{D}_t)$ are negatively correlated, which may indicate that they have common information about the variation of expected returns.

4.3 Predictive Regression Results

This section presents the results of the baseline predictive regression (4.7) for the period from January 1996 to December 2010. In Section 4.3.1, I investigate the in-sample forecasting power of implied correlation on market excess returns including second moments of the return distributions, valuation ratios and business cycle variables as control set. In Section 4.3.2, I explore whether the in-sample findings are driven by the recent financial crisis by focusing on a subsample period up to December 2006. Furthermore, recent literature has provided evidence for the predictive power of the correlation risk premium. Thus, I examine in detail whether the information content of implied correlation on future returns is captured by the correlation risk premium. Section 4.3.3 reports the results.

All of the findings are based on regressions of the CRSP value-weighted portfolio excess return on the lagged implied correlation (IC_t) and control variables. I use monthly observations in all the cases. The reported t-statistics are calculated using Newey-West standard errors to account for possible serial correlations in the residuals. The discussion of the results is based on the estimated coefficients, their statistical significance and the adjusted R^2's. To help the interpretation of estimated coefficients, all of the explanatory variables are standardized to have mean zero and unit variance, and the market risk premium is in percentage terms.

Finally, in Section 4.3.4, I investigate whether implied correlation has a better out-of-sample performance than historical average returns. The forecast experiments are undertaken on both monthly and cumulative returns as dependent variables, and

Table 4.1: Summary Statistics

The table reports the summary statistics and the correlation matrix of the implied correlation (IC_t), the realized correlation (RC_t), the variance risk premium (VRP_t), the value-weighted average of variance premia on all the index constituents ($\overline{\mathrm{VRPi}}_t$), the log of price-dividend ratio ($\log(\mathrm{P}_t/\mathrm{D}_t)$), the log of price-earnings ratio ($\log(\mathrm{P}_t/\mathrm{E}_t)$), the consumption–wealth ratio (CAY_t), the default spread (DS_t), the term spread (TS_t), the relative risk-free rate (RREL_t), and the market excess return ($\mathrm{R}_{\mathrm{m}t} - \mathrm{R}_{\mathrm{f}t}$). All of the variables are monthly estimates from January 1996 to December 2010. The implied correlation is obtained at the end of each month using implied variances for the index and for all of the individual components, calculated as risk-neutral expectations of the 30-days-ahead variance from equation (4.5). VRP_t is the difference between the market implied variance, calculated using equation (4.5) and the market realized variance, calculated as the sum of squared daily index return deviations over a one-month window. $\log(\mathrm{P}_t/\mathrm{D}_t)$ is obtained as the ratio of the S&P500 index price and its dividends per share aggregated over the last twelve months. $\log(\mathrm{P}_t/\mathrm{E}_t)$ is constructed as the price of the S&P500 index divided by its accumulated earnings over the last twelve months. Monthly time-series on CAY_t are obtained by using the most recently available quarterly observations. DS_t is calculated as the difference between BAA and AAA corporate bond spreads. TS_t is constructed as the difference between the ten-year Treasury bond and the three-month Treasury bill yields. RREL_t is calculated as the one-month T-bill rate minus its most recent twelve month moving averages. Finally, $\mathrm{R}_{\mathrm{m}t} - \mathrm{R}_{\mathrm{f}t}$ is the difference of the CRSP value-weighted portfolio (log) returns and the one-month treasury bill (log) rate. VRP_t and $\overline{\mathrm{VRPi}}_t$ are reported in percentages squared, and RREL_t and $\mathrm{R}_{\mathrm{m}t} - \mathrm{R}_{\mathrm{f}t}$ are reported in percentage terms.

	$\mathbf{IC}_t$	$\mathbf{RC}_t$	$\mathbf{VRP}_t$	$\overline{\mathbf{VRPi}}$	$\log(\mathbf{P}_t/\mathbf{D}_t)$	$\log(\mathbf{P}_t/\mathbf{E}_t)$	$\mathbf{CAY}_t$	$\mathbf{DS}_t$	$\mathbf{TS}_t$	$\mathbf{RREL}_t$	$\mathbf{R}_{\mathbf{m}t} - \mathbf{R}_{\mathbf{f}t}$
Panel A: Summary Statistics											
mean	0.50	0.33	2.39	-25.69	3.99	3.24	-0.59	1.01	1.68	-1.63	0.33
min	0.19	0.05	-325.93	-603.13	3.16	2.71	-3.73	0.55	-0.53	-23.17	-20.49
max	0.81	0.78	48.62	37.44	4.41	4.83	3.14	3.38	3.70	12.83	10.37
st. dev.	0.14	0.14	33.82	66.14	0.24	0.42	2.00	0.49	1.21	7.24	4.99
AC(1)	0.81	0.61	0.51	0.46	0.97	0.93	0.98	0.96	0.98	0.85	0.15
AC(2)	0.73	0.46	0.26	0.30	0.94	0.86	0.95	0.89	0.94	0.82	-0.05
Panel B: Unconditional Correlation Matrix											
IC_t	1.00										
RC_t	0.57	1.00									
VRP_t	-0.01	-0.39	1.00								
$\overline{\mathrm{VRPi}}$	-0.05	-0.24	0.90	1.00							
$\log(\mathrm{P}_t/\mathrm{D}_t)$	-0.49	-0.53	0.23	0.13	1.00						
$\log(\mathrm{P}_t/\mathrm{E}_t)$	0.16	0.04	0.04	-0.12	-0.13	1.00					
CAY_t	0.18	-0.25	0.04	-0.14	0.19	0.12	1.00				
DS_t	0.26	0.54	-0.51	-0.51	-0.65	0.41	-0.20	1.00			
TS_t	0.47	0.39	-0.06	-0.02	-0.40	0.39	-0.08	0.42	1.00		
RREL_t	-0.31	-0.27	0.08	0.14	0.22	-0.39	-0.21	-0.45	-0.47	1.00	
$\mathrm{R}_{\mathrm{m}t} - \mathrm{R}_{\mathrm{f}t}$	-0.02	-0.29	0.32	0.28	0.10	0.12	0.00	-0.23	0.01	0.13	1.00

the predictive ability is assessed by the out-of-sample R^2.

4.3.1 In-sample predictive power

Table 4.2 presents the simple regression results of monthly market excess returns on the lagged IC_t at different forecast horizons. Panel A reveals that the implied correlation is strongly related to subsequent monthly market excess returns, particularly at return horizons from one to six months. The significance is at a 5% level and the highest adjusted R^2s are 6.81%, 6.23% and 4.77% when explaining monthly market returns at two, three and six months ahead. For the same horizons, a one standard deviation increase in IC_t translates into 1.36%, 1.31% and 1.17% increase in market excess returns, respectively.

The results indicate a positive and economically significant relationship between IC_t and the future market risk premium. When examining my findings after the inclusion of traditional predictors, I find that this relationship is particularly strong for market returns at two, three and six months in the future. Table 4.3 reports the regression results when the forecast horizon is two months ahead. The first five columns present the simple setting results on control variables.[6] The figures reveal that the variance risk premium (VRP_t) and the value-weighted cross-sectional average of individual stocks' variance risk premium ($\overline{\mathrm{VRPi}}_t$) have significant predictive power for monthly market risk premium, two periods ahead. These results are consistent with Bollerslev et al. (2009) and Cosemans (2011), who document a significant relationship between the VRP_t and $\overline{\mathrm{VRPi}}_t$ on subsequent cumulative market excess returns, respectively. Both of these studies show that a higher difference between the model free implied variance (MFIV) and the realized variance obtained from intraday frequency observations is positively connected to the future market risk premium. My findings provide further evidence of a significant relationship when using the im-

[6]For the sake of brevity, I only report the coefficients that are found to be significant in at least one of the following prediction horizons: two, three and six months.

Table 4.2: Predictive Regressions of Monthly Market Risk Premium

Panel A presents the results of the following predictive regression: $rx^m_{t+h} = \alpha_1(h) + \alpha_2(h)\text{IC}_t + \epsilon_t$, where the dependent variable is monthly market excess return at $t+h$, h the forecast horizon in months, and IC_t the implied correlation defined in Table 4.1. The regression is based on monthly data from January 1996 to December 2010. The coefficients are estimated with ordinary least squares and t-statistics are calculated using Newey-West standard errors to account for possible serial correlation in the residuals. The implied correlation is standardized to have mean zero and standard deviation equal to one, and the dependent variable is in percentage terms.

h	**1**	**2**	**3**	**4**	**5**	**6**	**9**	**12**	**15**	**18**	**24**	**36**
Panel A												
IC_t	0.94	1.36	1.31	1.04	0.91	1.17	0.58	0.70	0.54	0.52	1.13	-0.42
	(2.84)	(4.64)	(4.16)	(3.20)	(2.32)	(3.63)	(1.61)	(2.03)	(1.56)	(1.46)	(2.80)	(-0.80)
constant	0.32	0.32	0.31	0.30	0.29	0.30	0.30	0.27	0.28	0.20	0.16	0.06
	(0.79)	(0.79)	(0.81)	(0.79)	(0.76)	(0.78)	(0.71)	(0.64)	(0.64)	(0.46)	(0.36)	(0.13)
adj. R^2 (%)	3.01	6.81	6.23	3.67	2.65	4.77	0.73	1.30	0.49	0.41	4.19	0.01

plied variance approximated as the strike of a simple variance swap and the realized variance estimated with daily frequency data.

When I compare the simple regression results with the second column of Table 4.2, I observe that the economic importance of IC_t dominates all of the explanatory variables. The statistically most important control variable, VRP_t, is expected to produce an increase of 0.94% of monthly market risk premium, two months ahead. The implied correlation, on the other hand, translates into 1.36% increase in market excess returns. Moreover, IC_t explains a higher proportion of the market risk premium variation compared to the other variables. Once more, VRP_t is the only control variable that appears to explain a fair part of the variation in subsequent returns. However its adjusted R^2 is only 2.98%, which is less than half than that of the implied correlation.

The last ten columns report the regression results when IC_t is included along with different control variables. To conserve space, I only report the results that represent the general findings.[7] Columns VI to XI present different combinations including IC_t and second moments/variance risk premia as regressors. The figures reveal that the slope of the implied correlation remains statistically significant at a 5% level. Not surprisingly, combining IC_t with VRP_t produces an adjusted R^2 of 10.02%. Moreover, the slope of the realized variance (RV_t) becomes negative and statistically significant along with IC_t. However, column IX suggests that in the presence of IC_t, RV_t contains information on subsequent returns that is given by VRP_t. Focusing on $\overline{\mathrm{VRPi}}_t$, columns X and XI indicate that the information content of this variable is also captured by VRP_t.

In the last set of columns (XII to XIV), I show the results when the log of price-to-dividend ratio ($\log(\mathrm{P}_t/\mathrm{D}_t)$) and business cycle variables are included in the multiple

[7] I also analyze many other different specifications (not reported) reaching qualitatively the same results. This is also the case when analyzing return predictions at three and six months ahead.

Table 4.3: Predictive Regressions: Two Months Ahead

The table presents the regression results of monthly market excess returns $R_{mt} - R_{ft}$ on the lagged implied correlation IC_t and different set of lagged control predictors at a prediction horizon of two months ahead. The regressions are based on monthly data from January 1996 to December 2010. The coefficients are estimated with ordinary least squares and t-statistics are calculated using Newey-West standard errors to account for possible serial correlation in the residuals. IV_t is the implied variance of the S&P100 index calculated as the risk-neutral expectations of the 30-days-ahead variance from equation (4.5) at the end of each month. RV_t is the realized variance of the S&P100 index estimated from equation (4.9). The rest of the variables are defined in Table 4.1. All the explanatory variables are standardized to have mean zero and standard deviation equal to one, and the dependent variable is in percentage terms.

	I	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**	**XIII**	**XIV**	**XV**
IC_t						1.49	1.55	1.37	1.16	1.40	1.34	1.39	1.66	1.79	1.65
						(4.25)	(4.57)	(4.24)	(3.54)	(4.40)	(4.17)	(2.96)	(5.76)	(4.44)	(5.00)
IV_t	0.36					-0.30									
	(0.79)					(-0.67)									
RV_t		-0.38					-0.76		0.91						
		(-1.36)					(-2.75)		(1.21)						
VRP_t			0.94					0.96	1.77		1.66				0.90
			(2.53)					(3.21)	(2.29)		(2.34)				(2.86)
$\overline{VRPi}_t$				0.65						0.73	-0.77				
				(2.00)						(2.18)	(-0.95)				
$\log(P_t/D_t)$					-0.62							0.06			
					(-1.11)							(0.08)			
CAY_t														0.10	
														(0.35)	
DS_t														-0.29	
														(-0.43)	
TS_t														-0.28	
														(-0.68)	
$RREL_t$													0.97	0.77	0.90
													(2.36)	(1.90)	(2.49)
constant	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
	(0.72)	(0.77)	(0.84)	(0.80)	(0.74)	(0.82)	(0.88)	(0.90)	(0.88)	(0.86)	(0.90)	(0.79)	(0.90)	(0.90)	(1.01)
adj. R^2 (%)	-0.05	0.00	2.98	1.12	0.99	6.57	8.47	10.02	10.02	8.40	9.95	6.29	9.70	8.86	12.41

regression. The descriptive analysis in Panel B of Table 4.1 reveals that IC_t and $\log(\text{P}_t/\text{D}_t)$ may possibly share information on future returns. The correlation matrix also suggests that part of the variation of IC_t relates to the business cycle. Hence, it is extremely important to examine whether the forecasting power of implied correlation on future market risk premium remains positive and significant after controling for these variables.

Overall, IC_t is highly significant in all of the specifications examined. The lowest t-statistics is 2.96 when combining the implied correlation with the log of price-to-dividend ratio. Having these two variables together actually reduces the adjusted R^2 of the model. Moreover, I find that the relative risk-free rate (RREL_t) becomes statistically significant in conjunction with the implied correlation. The adjusted R^2 increases to almost 10%, which is higher than the fraction of the variation explained when all of the business cycle variables are included in the multiple setting (columns XIII and XIV). Finally, as expected, IC_t along with VRP_t and RREL_t explain an important proportion of the variation of monthly market excess returns two months in the future.

I also provide evidence for the forecasting power of implied correlation when the focus horizon is three months ahead (Table 4.4). Similar conclusions are obtained: The economic importance of IC_t (1.31%) is also the highest compared to the rest of the variables in a simple regression setting and the explanatory power of IC_t is strongly robust to the inclusion of different predictors. I again observe that IC_t and RREL_t perform well in a bivariate setting; the economic importance and t-statistics of both variables increase substantially, and in conjunction they reach an adjusted R^2 of 11.34%.

Once more, the story is similar for a prediction horizon of six months. The new findings of Table 4.5 show that the coefficients of the implied variance (IV_t) and realized variance (RV_t) become statistically significant in explaining future excess returns

Table 4.4: Predictive Regressions: Three Months Ahead

The table presents the regression results of monthly market excess returns $\mathrm{R}_{\mathrm{m}t} - \mathrm{R}_{\mathrm{f}t}$ on the lagged implied correlation IC_t and different set of lagged control predictors at a prediction horizon of three months ahead. The regressions are based on monthly data from January 1996 to December 2010. The coefficients are estimated with ordinary least squares and t-statistics are calculated using Newey-West standard errors to account for possible serial correlation in the residuals. IV_t is the implied variance of the S&P100 index calculated as the risk-neutral expectations of the 30-days-ahead variance from equation (4.5) at the end of each month. RV_t is the realized variance of the S&P100 index estimated from equation (4.9). The rest of the variables are defined in Table 4.1. All the explanatory variables are standardized to have mean zero and standard deviation equal to one, and the dependent variable is in percentage terms.

	I	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**	**XIII**	**XIV**	**XV**
IC_t						1.63	1.51	1.32	1.48	1.34	1.29	1.35	1.70	1.69	1.70
						(4.00)	(4.18)	(4.07)	(3.69)	(4.11)	(4.02)	(2.78)	(5.06)	(4.43)	(4.79)
IV_t	0.00					-0.73									
	(0.00)					(-1.56)									
RV_t		-0.44					-0.82		-0.67						
		(-1.26)					(-2.48)		(-0.67)						
VRP_t			0.73					0.75	0.15		1.39				0.66
			(2.47)					(2.03)	(0.15)		(1.48)				(1.98)
$\overline{\mathrm{VRPi}}_t$				0.47						0.55	-0.70				
				(1.10)						(1.03)	(-0.70)				
$\log(\mathrm{P}_t/\mathrm{D}_t)$					-0.59							0.08			
					(-1.14)							(0.12)			
CAY_t														0.30	
														(0.98)	
DS_t														(0.12	
														(0.21)	
TS_t														(-0.14	
														(-0.38)	
RREL_t													1.25	1.30	1.20
													(2.84)	(3.01)	(2.98)
constant	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31
	(0.73)	(0.75)	(0.78)	(0.76)	(0.75)	(0.84)	(0.87)	(0.87)	(0.87)	(0.85)	(0.87)	(0.81)	(0.95)	(0.95)	(1.01)
adj. R^2 (%)	-0.57	0.20	1.52	0.30	0.79	7.38	8.21	7.95	7.69	6.88	7.78	5.71	11.34	10.19	12.57

Table 4.5: Predictive Regressions: Six Months Ahead

The table presents the regression results of monthly market excess returns $\mathrm{R}_{\mathrm{m}t} - \mathrm{R}_{\mathrm{f}t}$ on the lagged implied correlation IC_t and different set of lagged control predictors at a prediction horizon of six months ahead. The regressions are based on monthly data from January 1996 to December 2010. The coefficients are estimated with ordinary least squares and t-statistics are calculated using Newey-West standard errors to account for possible serial correlation in the residuals. IV_t is the implied variance of the S&P100 index calculated as the risk-neutral expectations of the 30 days-ahead-variance from equation (4.5) at the end of each month. RV_t is the realized variance of the S&P100 index estimated from equation (4.9). The rest of the variables are defined in Table 4.1. All the explanatory variables are standardized to have mean zero and standard deviation equal to one, and the dependent variable is in percentage terms.

	I	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**	**XIII**	**XIV**	**XV**
IC_t						0.96	1.01	1.15	1.13	1.12	1.12	1.06	1.54	1.19	1.54
						(3.15)	(3.52)	(4.06)	(3.56)	(3.83)	(3.84)	(2.63)	(4.01)	(3.31)	(4.27)
IV_t	0.89					0.45									
	(2.34)					(0.96)									
RV_t		0.88					0.62		0.11						
		(3.02)					(1.93)		(0.14)						
VRP_t			-0.67					-0.64	-0.54		-0.05				-0.72
			(-1.84)					(-2.48)	(-0.79)		(-0.07)				(-2.50)
$\overline{\mathrm{VRPi}}_t$				-0.78						-0.70	-0.65				
				(-1.98)						(-2.30)	(-0.83)				
$\log(\mathrm{P}_t/\mathrm{D}_t)$					-0.74							-0.22			
					(-1.87)							(-0.48)			
CAY_t														0.69	
														(2.01)	
DS_t														0.93	
														(2.35)	
TS_t														0.32	
														(1.02)	
RREL_t													1.13	1.74	1.19
													(2.19)	(2.84)	(2.31)
constant	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30
	(0.71)	(0.70)	(0.69)	(0.69)	(0.73)	(0.77)	(0.77)	(0.78)	(0.78)	(0.77)	(0.77)	(0.78)	(0.91)	(1.01)	(0.94)
adj. R^2 (%)	2.52	2.47	1.17	1.78	1.54	4.86	5.63	5.81	5.26	6.12	5.57	4.36	8.74	10.29	10.25

at a 5% level in a simple regression setting. However, looking at the bivariate regressions, I observe that the explanatory power of these variables is reduced including IC_t. Furthermore, $\log(P_t/D_t)$ is found to be weakly significant per se, but it looses its significance after the inclusion of IC_t.

In summary, my findings reveal that the in-sample forecasting power of implied correlation is robust to the inclusion of traditional predictors of market risk premium, particularly at prediction horizons of two, three and six months ahead. First, implied correlation contains information on future market returns that cannot be captured by second moments and variance risk premia. Second, valuation ratios do not perform well for the sample period of this study, and combining $\log(P_t/D_t)$ along with IC_t actually reduces the fit of the model. Third, when controling for all of the business cycle variables, IC_t is still highly significant, indicating that there is extra information of implied correlation that is not captured by these variables. Fourth, $RREL_t$ does not have a forecasting power alone, however it significantly explains future market excess returns in the presence of IC_t. Both variables together explain a high fraction of the market returns variation.

Finally, I perform two extra empirical exercises as robustness checks: first, the market realized variance is estimated from high-frequency–as opposed to daily–observations[8] and second, I use the cumulative returns as dependent variable. In both cases I obtain qualitatively the same findings previously discussed. Detailed explanations of this analysis along with the main results are reported in Appendix C.

4.3.2 Pre-crisis period

In this section, I investigate whether the relationship between implied correlation and future market returns is driven by the crisis period. To this end, I repeat the analysis but only considering a subsample from January 1996 to December 2006.

[8]Estimates of the market realized variance from high frequency data are obtained from Hao Zhou's webpage

Table 4.6: Predictive Regressions in the Pre-crisis Period

This table presents the regression results for the pre-crisis period. Panel A reports the results of the following predictive regression: $\mathrm{R}_{\mathrm{m}t} - \mathrm{R}_{\mathrm{f}t} = \alpha_1(h) + \alpha_2(h)\mathrm{IC}_t + \epsilon_t$, where the dependent variable is monthly market excess return at $t+h$, h the forecast horizon in months, and IC_t the implied correlation defined in Table 4.1. Panel B reports the multivariate predictive regression results with a forecast horizon of three months ahead. The regressions are based on monthly data from January 1996 to December 2006. The coefficients are estimated with ordinary least squares and t-statistics are calculated using Newey-West standard errors to account for possible serial correlation in the residuals. RV_t is the realized variance of the S&P100 index estimated from equation (4.9). The rest of the variables are defined in Table 4.1. All the explanatory variables are standardized to have mean zero and standard deviation equal to one, and the dependent variable is in percentage terms.

h	**1**	**2**	**3**	**4**	**5**	**6**	**9**	**12**	**15**	**18**	**24**	**36**
Panel A												
IC_t	0.97	1.19	1.28	0.88	0.63	0.99	0.83	0.79	0.79	0.81	0.55	-1.01
	(2.90)	(3.72)	(4.06)	(2.52)	(1.41)	(3.30)	(1.85)	(1.93)	(2.02)	(1.97)	(1.34)	(-2.55)
constant	0.47	0.46	0.46	0.45	0.43	0.45	0.45	0.41	0.43	0.32	0.26	0.13
	(1.30)	(1.33)	(1.35)	(1.31)	(1.22)	(1.27)	(1.21)	(1.09)	(1.08)	(0.82)	(0.61)	(0.32)
adj. R^2 (%)	3.84	6.04	7.09	2.92	1.09	3.81	2.43	2.05	2.00	2.12	0.46	4.37

Table 4.6: Predictive Regressions in the Pre-crisis Period (cont.)

Panel B	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**
IC_t	1.28				0.97	1.48	1.06	1.34	1.24	1.69	1.72
	(4.06)				(2.36)	(4.68)	(3.16)	(4.36)	(2.59)	(3.92)	(5.10)
RC_t		1.10			0.59				0.72		
		(3.11)			(1.29)				(1.66)		
IV_t											
RV_t											
VRP_t						-0.81					-0.70
						(-2.09)					(-1.61)
$\overline{\text{VRPi}}_t$											
$\log(\text{P}_t/\text{D}_t)$			-0.97				-0.50		0.56		
			(-2.87)				(-1.43)		(1.06)		
$\log(\text{P}_t/\text{E}_t)$				-0.68				-0.79	-1.16		-0.31
				(-1.94)				(-2.55)	(-2.39)		(-0.69)
CAY_t										-0.08	
										(-0.17)	
DS_t										0.02	
										(0.05)	
TS_t										-0.15	
										(-0.32)	
RREL_t										0.96	0.73
										(1.76)	(1.48)
constant	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
	(1.35)	(1.26)	(1.30)	(1.23)	(1.35)	(1.31)	(1.40)	(1.50)	(1.55)	(1.51)	(1.48)
adj. R^2 (%)	7.09	5.07	3.76	1.43	7.56	9.32	7.30	9.33	9.73	9.03	12.03

Table 4.6 presents the results.

The simple regression results reported in Panel A indicate that IC_t also predicts monthly market risk premium in the pre-crisis period, similarly at short and intermediate forecast horizons. The highest adjusted R^2 is around 7% when predicting returns three months ahead. Moreover, the economic importance and t-statistics are also the highest for this forecast horizon. To explore in more detail the explanatory power of IC_t in a multiple setting, I report the predictive regression results with a forecast horizon of three months in Panel B. The first set of columns shows the simple regression results of the control variables with a significant coefficient. New interesting findings compared to the main results are reached in the pre-crisis analysis. First, the realized correlation (RC_t) presents the evidence of forecasting power reported by Pollet and Wilson (2010). However, the slope of RC_t is not longer significant when IC_t is also included in the specification (column V). This result may indicate that a forward-looking measure contains more information than a physical measure of correlation on market risk premium.

Second, valuation ratios are significantly related to the future equity premium when the financial crisis is not included in the analysis. Unreported results reveal that the $\log(\text{P}_t/\text{D}_t)$ is significant in a simple setting for return horizons from one to almost two years. The price-to-earnings ratio, on the other hand, also presents a higher predictive power compared to the full sample results, but not as strong as $\log(\text{P}_t/\text{D}_t)$. However, when examining the multivariate setting, I observe that the explanatory power of $\log(\text{P}_t/\text{D}_t)$ is highly reduced under the presence of IC_t. In summary, the multivariate specifications illustrate that the explanatory power of IC_t is remarkably robust to the inclusion of different control variables, which indicates that the information content of IC_t on future market returns is also not captured by second moments, variance risk premia, valuation ratios, and business cycle variables in the pre-crisis period.

4.3.3 Analysis of the correlation risk premium

Cosemans (2011) shows that the correlation risk premium (CRP_t) is significantly related to the future market risk premium, particularly at short forecast horizons. Moreover, he documents that the forecasting power of the variance risk premium is partly driven by CRP_t. In this line, I examine whether the predictive power of IC_t is also captured by the correlation risk premium.

To this end, I compute CRP_t as the difference between the risk-neutral and physical expectations of correlation, $\mathrm{IC}_t - \mathrm{RC}_t$, and conduct two different analyses. First, I run the baseline regression (4.7) considering both monthly and cumulative market returns as dependent variables, and IC_t and CRP_t as predictors.[9],[10] Second, I analyze whether predicting market returns combining IC_t and RC_t gives a better fit than considering only $\mathrm{IC}_t - \mathrm{RC}_t$. Note that the second specification is nested in the first one, therefore I employ a likelihood ratio test (*LR* test) with the following restricted and unrestricted models:

$$\text{R model}: rx_{t+h} = \beta_0 + \beta_3(\mathrm{IC}_t - \mathrm{RC}_t) + \varepsilon_t, \tag{4.12}$$

$$\text{U model}: rx_{t+h} = \beta_0 + \beta_1\mathrm{IC}_t + \beta_2\mathrm{RC}_t + \epsilon_t, \tag{4.13}$$

where the null and alternative hypotheses are as follows:

$$\mathrm{H}_0 : \beta_2 = -\beta_1,$$

$$\mathrm{H}_A : \beta_2 \neq -\beta_1,$$

I calculate the log likelihood under the null and alternative hypotheses, LL_U and LL_0, respectively, to finally perform the *LR* test, where $LR = 2(\mathrm{LL}_\mathrm{U} - \mathrm{LL}_0)$ distributes $\chi^2(1)$.

Table 4.7 reports the results with a focus return horizon from one up to six months ahead. Each header row specifies the name of the indepedent variable considered in

[9] Cosemans (2011) performs his analysis by using cumulative market returns.

[10] The correlation coefficient between IC_t and CRP_t is 0.49 in this sample period.

Table 4.7: Analysis of the Correlation Risk Premium

The first three set of rows of the table present the predictive regression results of monthly market returns at $t+h$ and cumulative market returns from $t+1$ up to $t+h$ on the implied correlation (IC_t), the realized correlation (RC_t), and the difference $\mathrm{IC}_t{-}\mathrm{RC}_t$, i.e., the correlation risk premium (CRP_t). When predicting monthly market returns, t-statistics are calculated using Newey-West standard errors. When the dependent variable is cumulative market returns, I employ Britten-Jones et al. (2011)'s standard errors to correct for the overlapping problem. The table also reports the likelihood ratio test for the following unrestricted (U) and restricted (R) models: U model : $rx_{t+h} = \beta_0 + \beta_1 \mathrm{IC} + \beta_2 \mathrm{RC}_t + \epsilon_t$ and R model : $rx_{t+h} = \beta_0 + \beta_3 \mathrm{CRP}_t + \varepsilon_t$. The null and alternative hypotheses are as follows: $\mathrm{H}_0 : \beta_2 = -\beta_1$ and $\mathrm{H}_A : \beta_2 \neq -\beta_1$. The regressions are based on monthly data from January 1996 to December 2010. All the explanatory variables are standardized to have mean zero and standard deviation equal to one, and the dependent variable is in percentage terms.

Dep. Variable	**Monthly Market Excess Returns**						**Cumulative Market Excess Returns**				
h	**1**	**2**	**3**	**4**	**5**	**6**	**2**	**3**	**4**	**5**	**6**
I											
$\mathrm{IC}_t - \mathrm{RC}_t$	0.71	1.08	1.04	0.72	0.37	0.85	0.90	0.93	0.88	0.78	0.80
	(1.63)	(2.30)	(2.54)	(1.70)	(0.70)	(2.10)	(2.75)	(3.04)	(2.93)	(2.65)	(2.74)
adj. R^2 (%)	1.46	4.09	3.72	1.45	-0.03	2.24	5.06	8.40	9.65	9.14	11.18
II											
$\mathrm{IC}_t - \mathrm{RC}_t$	0.32	0.54	0.52	0.27	-0.10	0.35	0.44	0.46	0.42	0.33	0.35
	(0.50)	(0.88)	(1.06)	(0.54)	(-0.17)	(0.67)	(1.21)	(1.37)	(1.30)	(1.06)	(1.13)
IC_t	0.78	1.09	1.05	0.90	0.96	0.99	0.92	0.96	0.93	0.90	0.90
	(1.47)	(2.82)	(3.00)	(2.35)	(2.36)	(2.17)	(2.34)	(2.56)	(2.55)	(2.53)	(2.58)
adj. R^2 (%)	2.78	7.16	6.52	3.33	2.12	4.58	9.01	15.13	17.79	18.41	22.08
III											
RC_t	-0.34	-0.57	-0.55	-0.29	0.11	-0.37	-0.47	-0.49	-0.44	-0.35	-0.36
	(-0.50)	(-0.88)	(-1.06)	(-0.54)	(0.17)	(-0.67)	(-1.21)	(-1.37)	(-1.3)	(-1.06)	(-1.13)
IC_t	1.14	1.68	1.62	1.20	0.85	1.37	1.40	1.46	1.39	1.26	1.27
	(2.72)	(3.23)	(3.32)	(2.51)	(1.44)	(3.33)	(3.44)	(3.73)	(3.60)	(3.32)	(3.40)
adj. R^2 (%)	2.78	7.16	6.52	3.33	2.12	4.58	9.01	15.13	17.79	18.41	22.08
$\chi^2(1)$	3.42	6.81	6.23	4.41	4.81	5.24	8.58	14.52	17.64	19.84	23.79
p_value	0.06	0.01	0.01	0.04	0.03	0.02	0.00	0.00	0.00	0.00	0.00

the specifications. When the variable to forecast is cumulative market returns, the regression involves overlapping observations inducing serial correlation in the residuals. I correct for this problem by using Britten-Jones, Neuberger and Nolte (2011)'s standard errors which provide a simple way to correct for this problem; transformation of the original regression into an equivalent representation with a non-overlapping dependent variable. The results reveal that when running simple regressions (specification I), CRP_t highly predicts market excess returns. The explanatory power is strong when forecasting cumulative as opposed to monthly market returns. However, when comparing these findings to the findings obtained considering only IC_t (Table 4.2 and Panel A of Table C.2), I observe that for all of the forecast horizons examined, the economic importance and adjusted R^2s of the implied correlation are higher.

The bivariate regression results (Specification II) indicate that the predictive power of IC_t is not driven by CRP_t. IC_t is highly significant in predicting market returns under the presence of CRP_t at all of the return horizons, except at one month ahead. The interesting result is that the slope of CRP_t is no longer significant and its economic importance is highly reduced. The likelihood ratio test–in the last two rows of the table–supports this finding. Specification III presents the results for the unrestricted model. I observe that for all of the return horizons, except for one month ahead, I reject the null hypothesis with a p-value lower than 5%. This indicates that forecasting market returns with IC_t and RC_t as predictive variables instead of using the difference $\mathrm{IC}_t - \mathrm{RC}_t$, significantly improves the fit of the model.

4.3.4 Out-of-sample tests

The main findings indicate that implied correlation has a strong in-sample predictive power for the equity premium. However, one of Welch and Goyal (2008)'s critique is that predictive models often have a poor out-of-sample performance. They show that historical average returns have a better forecasting power for market excess returns than models with "popular" predictor variables. Keeping this critique in mind, I explore whether implied correlation also exhibits out-of-sample forecasting power in the sense that it beats the historical average return. I particularly focus on predictions of two, three and six months ahead, where the implied correlation was found to have the strongest in-sample predictive power. To this end, I compute the following error terms,

$$\varepsilon_{1,t+h} = rx^m_{t+h} - \widehat{rx^m_{t+h}}, \tag{4.14}$$

$$\varepsilon_{2,t+h} = rx^m_{t+h} - \overline{rx^m_t}, \tag{4.15}$$

where rx^m_{t+h} is the realized return at $t+h$, $\widehat{rx^m_{t+h}}$ is the fitted value obtained from estimating the predictive model up to time t, and $\overline{rx^m_t}$ is the historical average return calculated up to time t. I use six years of estimation window to obtain the parameters of the predictive regressions and I recursively calculate the error terms over a testing window (T_{test}) of nine years. Then I evaluate the out-of-sample performance by comparing the mean-squared errors (MSEs) and the statistical significance of their difference by employing the test of Diebold and Mariano (1995) (DM). The implied correlation has a superior forecasting ability if the MSE of the predictive regression is lower than that of the historical average return. This is equivalent to observing a positive out-of-sample R^2,

$$R^2_{\text{out}} = 1 - \frac{\sum_{t=1}^{T_{\text{test}}-h} \varepsilon^2_{1,t+h}}{\sum_{t=1}^{T_{\text{test}}-h} \varepsilon^2_{2,t+h}}. \tag{4.16}$$

Table 4.8 presents the results. In the first set of three rows, I begin discussing h-periods-ahead monthly market return forecasts. Consistent with the in-sample results, I find that the implied correlation delivers positive out-of-sample R^2s of more than 5%, 3% and 4% when the forecast horizons are two, three and six months ahead, respectively. However, DM tests reveal that the difference in mean-squared errors is not statistically different from zero. Taking into account the short sample period

Table 4.8: Out-of-Sample Tests

The table reports the out-of-sample forecasting tests for prediction horizons of two, three and six months for the period from January 1996 to December 2010. The market excess return is based on the CRSP NYSE/AMEX/NASDAQ value-weighted portfolio in excess of the one-month treasury bill (log) rate. The forecast error at period $t+h$ is obtained as the difference between the market excess return at $t+h$ and the fitted value from estimating the predictive model up to time t with the implied correlation as predictor variable. The benchmark error is calculated by employing the historical average return up to time t. Six years of estimation window are used to obtain the parameters of the predictive regressions and the error terms are recursively calculated over nine years of testing window. The out-of-sample R^2 is defined in equation (4.16). ΔMSE is the difference in mean-squared errors, reported in percentages squared, and their statistical difference is assessed by employing the Diebold-Mariano test (DM t-stat). The first three rows report the statistics of h-months-ahead forecasts of monthly market excess returns, whereas the last set of rows presents the one-step-ahead forecasts of h-month-cumulative market excess returns.

h	**2**	**3**	**6**
dep. var: Monthly market excess returns			
R^2_{out} (%)	5.12	3.20	4.25
ΔMSE	1.25	0.79	1.03
DM t-stat	(1.35)	(0.83)	(1.58)
dep. var: Cumulative market excess returns			
R^2_{out} (%)	6.97	7.61	11.32
ΔMSE	1.09	0.89	0.80
DM t-stat	(1.64)	(1.68)	(2.83)

used to perform out-of-sample experiments, this finding provides weak evidence for monthly market returns predictability at the aforementioned forecast horizons.

The last set of rows reports the out-of-sample predictive power of the implied correlation on one-step-ahead cumulative returns. For this purpose, I use the fitted values obtained from the previous monthly predictions to construct cumulative fitted returns. If the aim is to predict the next-period h-month-cumulative excess returns, I train the linear predictive model over the estimation window to forecast h-periods-ahead monthly returns. Then I use the parameters estimated from the trained model up to time t to obtain the following monthly fitted values: $\widehat{rx^m_{t+1}},...,\widehat{rx^m_{t+h}}$, and by summing them up, I construct the fitted cumulative excess return from $t+1$ to $t+h$. The forecast errors are calculated as the difference between the realized and fitted cumulative returns and finally compared to the benchmark errors of equation (4.15). In overall, the statistics indicate that the implied correlation has a good out-of-sample performance on predicting cumulative returns. On the one hand, when the return horizon is two and three months, R^2_{out}'s are around 7% and the predictive regression has a statistically significant lower mean-squared prediction error. On the other hand, the statistics are stronger when the return horizon is 6 months, reaching an out-of-sample R^2 of approximately 11%.

In summary, these findings document consistent evidence with the in-sample results. The implied correlation exhibits out-of-sample predictive power for both monthly and cumulative excess returns for the period from January 1996 to December 2010.

4.4 Conclusion

This paper shows that aggregate implied correlation may be an important indicator of market-wide risk. I estimate risk-neutral expectations of second moments from option prices of the S&P100 index and of the S&P100 individual components using a sample period from January 1996 to December 2010, and examine whether correlation provides information on the time-series of expected market returns and/or affects the risk-return trade-off of stock returns.

I document that implied correlation has strong forecasting power for market excess returns. The degree of predictability is not only present for cumulative returns, but also for monthly returns, and it is particularly strong at prediction horizons of two, three and six months. The implied correlation displays the highest economic importance, and it explains the largest proportion in the variation of future aggregate returns, among alternative variables. The results are robust to different specifications, to the inclusion of standard predictors, and are not driven by the recent financial crisis. Moreover, this paper provides further evidence for Pollet and Wilson (2010)'s results. The physical correlation is highly and positively related to three-months-ahead market returns, in the pre-crisis period. However, its forecasting power is no longer significant under the presence of implied correlation, suggesting that a risk-neutral expectation of correlation reveals more information on the future equity premium. Similarly, my findings also support the results of Cosemans (2011); the correlation risk premium has a strong in-sample forecasting power for cumulative market returns at intermediate return horizons, but its information content is found to be fully captured by the implied correlation.

Chapter 5

Conclusion

The results presented in chapters 2 and 3 indicate that the limit order book contains information on trader's order choice and on the short-term price movements.

The Chapter "Global depth and future volatility" proposes a summary measure of the limit order book distribution (global depth) and provides empirical evidence that global depth has a strong in-sample and out-of-sample predictive power over market volatility. The information content of global depth cannot be captured by other liquidity measures and trading related variables. Moreover, we document that the predictive power is not only at the aggregate level. For most of the stocks in our sample, individual stock global depth significantly forecasts individual stock volatility.

Chaper 2 provides market participants a summary measure of the limit order book distribution with predictive power over volatility. One interesting extension of this study is to use the volatility forecasts to design a trading strategy. In high volatility states the risk of being picked off by an informed trader increases. Hence, If investors are able to anticipate an increase in volatility, they can submit less aggressive orders and reduce their execution costs. A different route to follow for further research is to explore the predictive power of our summary measure over returns. As global depth gives the relative price position of the quoted depth, it may also contain information on the direction of the price. Unreported results reveal that neither global depth nor other variables such as trading related variables and liquidity measures are able to predict intraday returns for the sample period considered in this study. However, when we focus on the predictability regressions at a daily frequency, we find that global depth significantly forecasts daily returns at a 10% level. Considering that we use two months of data, this result motivates the extension of this analysis by investigating the global depth–future returns relationship either in another exchange or by increasing the sample size.

In “Competition, signaling and non-walking through the book: effects on order choice”, we show that the competition effect is stronger than the signaling effect for both sides of the market in every stage of the trader’s decision process. For a limit order agent, the competition effect is the strongest for the volume at the second best quotes for both, buyers and sellers. On the other hand, as walking through the book is not allowed, for an impatient trader only volatility, price trend, and signaling variables on the opposite side affect her order choice decision. Finally, in comparison to individual traders, we document that institutional traders’ order submission strategies are less affected by the state of the limit order book.

One novel feature of this chapter is that it analyzes the effect of the limit order book information when walking through the book is not allowed. However, the main focus of Chapter 3 is on its impact on order choice. One possible extension could be to analyze the effects of this market mechanism on liquidity. When walking through the book is allowed, large market orders walk up the book until they are fully executed. Hence, the transitory price impact of such an order is higher under this market mechanism. It would be interesting to examine if this is also the case on the permanent price impact. This would give important information to stock exchanges for designing their market mechanism rules.

Chapter 4, on the other hand, addresses a question in the area of asset pricing, and provides empirical evidence on the informativeness of the implied correlation on both monthly and cumulative market excess returns. I particularly find that the predictive power of implied correlation is stronger at intermediate forecast horizons. It is robust to the inclusion of standard predictors of market returns and robust to different specifications. The economic importance and the proportion of the time-series variation in market risk premium explained by the implied correlation are the strongest among the rest of the variables.

The analysis presented in Chapter 4 points out different directions for further research. As Fama and French (1989) documented, expected bond and stock returns move together, and the variation in both of them is related to the business conditions. Hence, an extension of this analysis is to investigate whether implied correlation has any forecasting power over bond returns. Furthermore, the findings presented could be reconciled to the literature on cross-sectional asset pricing. Some studies have examined whether correlation affects the risk-return trade-off of asset returns. Mueller et al. (2012) use the implied correlation to analyze the cross-section of currency returns. Krishnan et al. (2009), on the other hand, focus on the equity market and a physical measure of correlation as risk factor. The results of this chapter reveal that implied correlation is more informative about future market returns than the realized correlation. Hence, an open question is whether this is also the case in the cross-section of equity returns.

Appendix **A**

Appendix A

A.1 Data Samples

Tables A.1, A.2 and A.3 present samples of the order data, trade data and limit order book data for one of the stocks in our sample for July 1, 2008, respectively. Table A.1 provides the identity number of an order (OrderID), the number of shares submitted (Volume), the corresponding limit price in Turkish Lira (Price), and time (Time). In addition, order data includes identifiers showing whether an order is valid for one session or for the whole day (TIF), whether the order is submitted by an individual or an institutional client (TraderType), whether the order is an immediate or cancel order (KTR) order, and finally the identity number of the split order (SplitID).

Table A.2 reports the transaction time (Time), the traded price in Turkish Lira (Price), and the number of shares traded (Volume). The identity numbers of buy and sell orders for a given trade are also presented (BuyerID and SellerID, respectively).

Finally, Table A.3 presents the best bid and ask prices (B1 and A1, respectively), the inside spread A1−B1 (SPR), and the number of shares waiting at the best bid and ask prices (VB1 and VA1). Prices and number of shares beyond the best quotes are also provided. To conserve space, only the information up to the tenth position is reported.

Table A.1: Order Data

Table reports a sample of the order data for Akbank (AKBNK) for July 1, 2008. OrderID is the identity number of the submitted order assigned by the Exchange, Volume is the number of shares to be bought or sold, Price is the limit price (in Turkish Lira), TIF is Time-in-force (0 if the order is valid for one session, 1 if it is valid for the whole day), Time is the order submission time, TraderType takes value IND or INS if the order is submitted by an individual client or an institutional client, respectively. KTR takes value E if an order is an immediate or cancel order. Finally SplitID is the ID number of the order which is split into several orders.

OrderID	**Ticker**	**OrderType**	**Volume**	**Price**	**TIF**	**Time**	**TraderType**	**KTR**	**SplitID**
107200800181205	AKBNK	Buy	50000	4.02	0	15:30:35	IND		
107200800181222	AKBNK	Buy	25000	4.02	1	15:30:37	IND		
107200800181254	AKBNK	Buy	527	4.02	0	15:30:39	IND		
107200800181275	AKBNK	Sell Modification	24425	4.04	0	15:30:40	INS	E	
107200800181304	AKBNK	Sell	10000	4.04	0	15:30:41	IND		
107200800181309	AKBNK	Sell	1000	4.04	0	15:30:42	IND		
107200800181363	AKBNK	Buy	50	4.04	0	15:30:47	IND		
107200800165524	AKBNK	Buy Modification	5	4.02	0	15:30:50	IND	E	
107200800181427	AKBNK	Buy	1	4.08	0	15:30:53	IND		
107200800181431	AKBNK	Sell Modification	5000	4.04	0	15:30:53	IND		
107200800181452	AKBNK	Buy	1000	4.04	0	15:30:55	IND		
107200800181479	AKBNK	Buy	100	4.02	0	15:30:57	IND		
107200800173629	AKBNK	Short Sell	5000	4.04	0	15:31:00	IND		
107200800181717	AKBNK	Sell	100	4.04	0	15:31:27	IND		
107200800181844	AKBNK	Buy	100	3.94	1	15:31:40	IND		
107200800181888	AKBNK	Buy	5000	4	0	15:31:44	INS		
107200800182186	AKBNK	Sell	15000	4.02	0	15:32:23	IND		
107200800182191	AKBNK	Buy	1	4.08	0	15:32:24	IND		
107200800182195	AKBNK	Sell	25000	4.02	1	15:32:25	IND		
107200800181304	AKBNK	Short Sell	10000	4.02	0	15:32:26	IND		
107200800173629	AKBNK	Short Sell	5000	4.02	0	15:32:28	IND		
107200800182223	AKBNK	Buy	500	4	0	15:32:28	IND		
107200800182230	AKBNK	Sell	700	4.02	0	15:32:40	IND		
107200800182346	AKBNK	Buy	100	4.02	1	15:32:40	IND		
107200800178541	AKBNK	Buy Modification	2000	4.02	0	15:32:47	IND		
107200800182411	AKBNK	Sell Split	1000	4.06	0	15:32:52	IND		107200800181194

Table A.2: Trade Data

The table reports a sample of the trade data for Akbank (AKBNK) for July 1, 2008. Time is the transaction time, Price is the traded price (in Turkish Lira) and Volume gives the number of orders traded. BuyerID and SellerID are the identity numbers of the matching buy and sell orders for a given trade.

Ticker	**Time**	**Price**	**Volume**	**BuyerID**	**SellerID**
AKBNK	15:30:35	4.02	11501	107200800181205	107200800181191
AKBNK	15:30:47	4.04	50	107200800181363	107200800173428
AKBNK	15:30:53	4.04	1	107200800181427	107200800173428
AKBNK	15:30:55	4.04	1000	107200800181452	107200800173428
AKBNK	15:32:23	4.02	15000	107200800181205	107200800182186
AKBNK	15:32:24	4.04	1	107200800182191	107200800173428
AKBNK	15:32:25	4.02	23499	107200800181205	107200800182195
AKBNK	15:32:25	4.02	1501	107200800181222	107200800182195
AKBNK	15:32:26	4.02	10000	107200800181222	107200800181304
AKBNK	15:32:28	4.02	5000	107200800181222	107200800173629
AKBNK	15:32:29	4.02	700	107200800181222	107200800182230
AKBNK	15:33:01	4.04	1	107200800182498	107200800173428
AKBNK	15:33:25	4.02	7799	107200800181222	107200800182673
AKBNK	15:33:25	4.02	527	107200800181254	107200800182673
AKBNK	15:33:25	4.02	5	107200800165524	107200800182673
AKBNK	15:33:25	4.02	100	107200800181479	107200800182673
AKBNK	15:33:25	4.02	100	107200800182346	107200800182673
AKBNK	15:33:25	4.02	2000	107200800178541	107200800182673
AKBNK	15:33:25	4.02	1000	107200800182428	107200800182673
AKBNK	15:33:25	4.02	5000	107200800181888	107200800182673
AKBNK	15:33:58	4	10	107200800163849	107200800182976
AKBNK	15:34:09	4.02	15469	107200800183084	107200800182678
AKBNK	15:34:09	4.02	965	107200800183084	107200800161924
AKBNK	15:34:09	4.02	50000	107200800183084	107200800182805
AKBNK	15:34:09	4.02	10000	107200800183084	107200800117710
AKBNK	15:34:09	4.02	5000	107200800183084	107200800182940

Table A.3: Limit Order Book Data

The table reports a sample of the limit order book data for Akbank (AKBNK) for July 1, 2008. B1 and A1 are the best bid and ask prices respectively, whereas SPR is the inside spread calculated as A1−B1. VB1 and VA1 give the number of shares waiting at the best bid and ask prices, respectively. Similarly, B2 (B10) and A2 (A10) are the second (tenth) best prices and VB2 (VB10) and VA2 (VA10) are the corresponding number of shares.

Time	**B1**	**A1**	**SPR**	**VB1**	**VA1**	**B2**	**A2**	**VB2**	**VA2**	**.**	**B10**	**A10**	**VB10**	**VA10**
15:30:35	4.00	4.02	0.02	112816	11501	3.98	4.04	215352	51426	.	3.82	4.2	25154	66670
15:30:37	4.02	4.04	0.02	38499	51426	4	4.06	112816	160316	.	3.84	4.22	25204	32000
15:30:39	4.02	4.04	0.02	63499	51426	4	4.06	112816	160316	.	3.84	4.22	25204	32000
15:30:40	4.02	4.04	0.02	64026	51426	4	4.06	112816	160316	.	3.84	4.22	25204	32000
15:30:41	4.02	4.04	0.02	64026	75851	4	4.06	112816	160316	.	3.84	4.22	25204	32000
15:30:42	4.02	4.04	0.02	64026	85851	4	4.06	112816	160316	.	3.84	4.22	25204	32000
15:30:47	4.02	4.04	0.02	64026	86851	4	4.06	112816	160316	.	3.84	4.22	25204	32000
15:30:50	4.02	4.04	0.02	64026	86801	4	4.06	112816	160316	.	3.84	4.22	25204	32000
15:30:53	4.02	4.04	0.02	64031	86801	4	4.06	112811	160316	.	3.84	4.22	25204	32000
15:30:53	4.02	4.04	0.02	64031	86801	4	4.06	112811	160316	.	3.84	4.22	25204	32000
15:30:55	4.02	4.04	0.02	64031	91800	4	4.06	112811	160316	.	3.84	4.22	25204	32000
15:30:57	4.02	4.04	0.02	64031	90800	4	4.06	112811	160316	.	3.84	4.22	25204	32000
15:31:00	4.02	4.04	0.02	64131	90800	4	4.06	112811	160316	.	3.84	4.22	25204	32000
15:31:27	4.02	4.04	0.02	64131	95800	4	4.06	112811	155316	.	3.84	4.22	25204	32000
15:31:40	4.02	4.04	0.02	64131	95900	4	4.06	112811	155316	.	3.84	4.22	25204	32000
15:31:44	4.02	4.04	0.02	64131	95900	4	4.06	112811	155316	.	3.84	4.22	25204	32000
15:32:23	4.02	4.04	0.02	64131	95900	4	4.06	117811	155316	.	3.84	4.22	25204	32000
15:32:24	4.02	4.04	0.02	49131	95900	4	4.06	117811	155316	.	3.84	4.22	25204	32000
15:32:25	4.02	4.04	0.02	49131	95899	4	4.06	117811	155316	.	3.84	4.22	25204	32000
15:32:26	4.02	4.04	0.02	24131	95899	4	4.06	117811	155316	.	3.84	4.22	25204	32000
15:32:28	4.02	4.04	0.02	14131	85899	4	4.06	117811	155316	.	3.84	4.22	25204	32000
15:32:28	4.02	4.04	0.02	14131	85899	4	4.06	117811	155316	.	3.84	4.22	25204	32000
15:32:40	4.02	4.04	0.02	9131	80899	4	4.06	118311	155316	.	3.84	4.22	25204	32000
15:32:40	4.02	4.04	0.02	8431	80899	4	4.06	118311	155316	.	3.84	4.22	25204	32000
15:32:47	4.02	4.04	0.02	8531	80899	4	4.06	118311	155316	.	3.84	4.22	25204	32000
15:32:52	4.02	4.04	0.02	10531	80899	4	4.06	116311	155316	.	3.84	4.22	25204	32000

A.2 Calculation of Global Depth

Suppose that the limit order book for stock X at 11:00am is as follows:

Order type	Volume	Limit price	Time	Best Bid	Best Ask
Sell	50000	8.3	09:30:00	-	8.2
Buy	10000	7.9	09:30:01	7.9	8.2
Sell	1800	8.3	09:30:02	7.9	8.2
.					
.					
.					
Sell	3334	8.05	10:58:17	8	8.05
Buy	25000	8	10:58:20	8	8.05
Buy	50000	8	10:58:38	8	8.05
Sell	1	8.1	10:58:50	8	8.05

The first step in the calculation of global depth involves the calculation of the tick-adjusted price distance Δ of each limit order in the given book relative to the best extant limit price:

$$\Delta_{i,\tau}^{\text{buy}} = (p_\tau^B - p_i^{\text{buy}})/\text{tick},$$

$$\Delta_{i,\tau}^{\text{sell}} = (p_i^{\text{sell}} - p_\tau^A)/\text{tick},$$

where p_τ^B (p_τ^A) is the best bid (ask) price in interval τ and p_i^{buy} (p_i^{sell}) is the limit price of the i^{th} order.

Say the tick size is 0.05. Then we have the following price distances for the orders:

Order type	Volume	Limit Price	Time	Best Bid	Best Ask	Δ
Sell	50000	8.3	09:30:00	-	8.2	5
Buy	10000	7.9	09:30:01	7.9	8.2	2
Sell	1800	8.3	09:30:02	7.9	8.2	5
.					.	.
.					.	.
.					.	.
Sell	3334	8.05	10:58:17	8	8.05	0
Buy	25000	8	10:58:20	8	8.05	0
Buy	50000	8	10:58:38	8	8.05	0
Sell	1	8.1	10:58:50	8	8.05	1

Next, we obtain of the percentage of total volume supplied/demanded at a given Δ for $\Delta = 0, 1, 2, .., 30$. This way, we reach the limit order book probability density function (LOB–PDF). By integrating the LOB–PDF of the each side of the market over the ranges of Δ, i.e., by calculating the cumulative frequencies, we obtain the limit order book cumulative distribution function (LOB–CDF). That is:

	Buy side			**Sell side**		
Δ	Total Volume	Frequency	Cum. Frequency	Total Volume	Frequency	Cum. Frequency
0	78500	0.270	0.270	68400	0.186	0.186
1	52575	0.181	0.450	71602	0.194	0.380
2	58440	0.201	0.651	54588	0.148	0.528
3	45579	0.156	0.807	62068	0.168	0.697
.		.				
.		.				
.		.				
29	0	0.000	1.000	0	0.000	1.000
30	0	0.000	1.000	0	0.000	1.000

Global depth for each stock is the weighted average of the LOB–CDF, where the weight function is given in Equation (2.2). For the estimated decay parameter $\widehat{\lambda} = 0.366$, we have the following weights and the resulting global depth (GD):

		Buy side			**Sell side**		
Δ	weights ($\lambda = \widehat{\lambda}$)	Cum. Freq.	weight*cum.freq	GD	Cum. Freq.	weight*cum.freq	GD
0	0.307	0.270	0.083		0.186	0.057	
1	0.213	0.450	0.096		0.380	0.081	
2	0.147	0.651	0.096		0.528	0.078	
3	0.102	0.807	0.083		0.697	0.071	
.		.			.		
.		.			.		
.		.			.		
29	0.000	1.000	0.000		1.000		
30	0.000	1.000	0.000	0.576	1.000		0.485

Finally, the aggregated global depth for a given interval τ is calculated as the cross-sectional average of the individual stock global depth measures.

A.3 Global Depth vs. "Local" Depth

By definition, global depth is related to the standard "local" depth measures, i.e., the quoted depth up to a given threshold. To investigate their relationship, we estimate the following regressions for buy and sell sides of the market separately:

$$\mathrm{GD}(\Delta, \lambda)_{s,\tau} = b_{0,s} + b_{1,s}\mathrm{DEPTH}_{s,\tau} + \epsilon_{s,\tau}.$$

For a given limit order book at time τ and for each stock s, DEPTH denotes the "local" depth measure calculated at different thresholds. First, we calculate the volume available at the best quotes, DEPTH0. Second, to capture the volume available beyond the best quotes, we calculate the cumulative volume of orders up to the three best quotes, DEPTH03, and five best quotes, DEPTH05.[1] We evaluate global depth given in (2.1) and (2.2) at three exogenously given decaying factors, $\lambda = 0, 0.5$ and 1. Hence, we first assign equal weights for each of the quotes, then we allow for exponential decaying weights with a lower and higher decay factors.

The table below presents the results. To conserve space, only the analysis for the buy side of the market is reported. Results for the sell side are qualitatively similar. The main conclusion from this analysis is that, irrespective of the chosen decay factor, the local depth at the best quotes is strongly and positively related to our summary measure, global depth. As expected, the explanatory power of DEPTH0 over global depth is increasing with the decay factor; a higher λ makes global depth more closely related to the depth at the best quotes. However, as the results suggest, even when $\lambda = 1$, DEPTH0 can explain only 42% (33%) of the variation in global depth for the buy (sell) side. The explanatory power of the depth variables over global depth increases up to 59% when we include all of the depth variables.

From this analysis, we conclude that although they are positively and significantly correlated, there is a non-trivial proportion of the variation of global depth that cannot

[1]As an additional robustness, we employ different depth variables measured not only at the first three or five quotes, but also at different thresholds. The results are qualitatively similar.

be explained by the standard depth measures. This analysis suggests that our variable provides different information than that of the standard depth measures.

Table A.4: Global Depth and "Local" Depth Variables

PANEL A			$\text{GD}(\lambda = 0)^{\text{buy}}_{\tau}$		
DEPTH0_{τ}	0.15			0.14	0.17
	[100/100]			[93/100]	[100/100]
DEPTH03_{τ}		0.05		0.01	
		[87/87]		[57/53]	
DEPTH05_{τ}			0.03		-0.02
			[60/89]		[60/28]
constant	0.88	0.88	0.88	0.88	0.88
	[100/100]	[100/100]	[100/100]	[100/100]	[100/100]
adj. R^2	26.28	28.86	22.99	31.99	29.37
PANEL B			$\text{GD}(\lambda = 0.5)^{\text{buy}}_{\tau}$		
DEPTH0_{τ}	0.70			1.10	1.13
	[100/100]			[100/100]	[100/100]
DEPTH03_{τ}		0.14		-0.16	
		[67/80]		[87/0]	
DEPTH05_{τ}			0.05		-0.20
			[37/55]		[100/0]
constant	0.34	0.39	0.42	0.38	0.41
	[100/100]	[100/100]	[100/100]	[100/100]	[100/100]
adj. R^2	39.74	14.16	6.23	41.01	45.61
PANEL C			$\text{GD}(\lambda = 1)^{\text{buy}}_{\tau}$		
DEPTH0_{τ}	0.78			1.38	1.36
	[100/100]			[100/100]	[100/100]
DEPTH03_{τ}		0.24		-0.31	
		[57/63]		[100/0]	
DEPTH05_{τ}			-0.01		-0.26
			[37/45]		[100/0]
constant	0.21	0.26	0.30	0.27	0.28
	[100/100]	[100/100]	[100/100]	[100/100]	[100/100]
adj. R^2	41.87	6.77	1.86	50.86	58.71

Appendix **B**

Appendix B

This appendix consists of two tables related to Chapter 3, Competition, Signaling and Non-Walking Through the Book: Effects on Order Choice. Table B.1 presents the definitions of the key variables used throughout the analysis in the chapter. Table B.2 on the other hand, summarizes our main findings, expressed qualitatively. For each finding, we as well include a pointer to the supporting empirical evidence (the corresponding table), a pointer to similar results in the empirical/theoretical literature (if it exists) and a statement of whether the result is in agreement or contrast with previous empirical/theoretical literature.

Table B.1: Definitions of Explanatory Variables

The table presents the description of the independent variables used in the two-stage sequential ordered probit model.

Regressors	**Definition**
Covariates for the depth at and beyond the best quotes	
Vcomp	The total volume of orders *at the best quote* in the first stage and second stage–MO and the accumulated volume of orders *up to the second best quotes* in the second stage–LO for the same side of the book.
Vcompopp	The total volume of orders *at the best quote* in the first stage and the accumulated volume of orders *up to the second best quotes* in the second stage–LO for the opposite side of the book.
Vsignal	The accumulated volume of orders *from the second up to the tenth best quotes* in the first stage and second stage–MO and the accumulated volume of orders *from the third up to the tenth best quotes* in the second stage–LO for the same side of the book.
Vsignalopp	The accumulated volume of orders *from the second up to the tenth best quotes* in the first stage and second stage–MO and the accumulated volume of orders *from the third up to the tenth best quotes* in the second stage–LO for the opposite side of the book.
Covariates for walking through the book	
SPR	The (tick size adjusted) difference between the best ask and bid quotes.
Dsame$^{1-2}$	The price distance between the best and the second best quotes for the same side of the book.
Dsame^{2-max}	The price distance between the second best ask (bid) and the highest available ask (lowest available bid) quote for the same side of the book.
Dopp$^{1-2}$	The price distance between the best and the second best quotes for the opposite side of the book.
Dopp^{2-max}	The price distance between the second best ask (bid) and the highest available ask (lowest available bid) quote for the opposite side of the book.
Additional variables	
Vola	The exponential moving average of the last 60 mid-quote squared returns.
Trend	The change of the mid-quote prices for the last 60 observations

Table B.2: Summary of the Main Findings

This table presents the summary of our main findings along with the corresponding table. All the variables are defined in Table B.1.

Regressors	Main Findings	Table	Consistent with	Inconsistent with
Covariates for the depth at and beyond the best quotes				
Vcomp	Encourages market orders.	3.4	Parlour (1998), Ranaldo (2004), Beber and Caglio (2005), Pascual and Veredas (2009)	
	It persists beyond the best quotes and it is the strongest up to the 2^{nd} best quote.	3.3		
Vcompopp	Discourages market orders.	3.4	Parlour (1998), Ranaldo (2004), Pascual and Veredas (2009)	
Vsign	(weakly) discourages market orders.	3.4	Goettler et al. (2005), and Goettler et al. (2009)	Pascual and Veredas (2009)
	Discourages the limit order aggressiveness.	3.5	Goettler et al. (2005), Goettler et al. (2009), and Pascual and Veredas (2009)	
Vcomp/ Vsign	The competition effect is stronger compared to the signaling effect.	3.4, 3.5, 3.6		

Table B.2: Summary of the Main Findings (cont.)

Regressors	Main Findings	Table	Consistent with	Inconsistent with
Covariates for walking through the book				
SPR	Discourages MOs.	3.4	Ranaldo (2004), Beber and Caglio (2005), Ellul, Holden, Jain, and Jennings (2007) Cao et al. (2008), and Pascual and Veredas (2009)	
	Encourages aggressive LOs.	3.5	Ellul, Holden, Jain, and Jennings (2007) and Pascual and Veredas (2009)	
	No significant effect on the market order aggressiveness.	3.6		Pascual and Veredas (2009)
$\mathrm{Dopp}^{1}\_{}^{2}$/ $\mathrm{Dopp}^{2}\_{}^{\max}$	No significant effect on MOs.	3.4	Cao et al. (2008)	Pascual and Veredas (2009)
	No significant effect on the market order aggressiveness.	3.6		Pascual and Veredas (2009)
Additional variables				
Vola	Discourages MOs.	3.4	Foucault (1999), Ahn et al. (2001), Ranaldo (2004), Beber and Caglio (2005),	
	Encourages aggressive MOs.	3.6	Hall and Hautsch (2006)	Ranaldo (2004)
Trend	Discourages (encourages) buy (sell) MOs.	3.4		Beber and Caglio (2005), Cao et al. (2008)
	Encourages (discourages) aggressive buy (sell) LOs.	3.5		
	Discourages (encourages) aggressive buy (sell) MOs.	3.6		

Appendix **C**

Appendix C

C.1 High-Frequency Realized Variance

The appendix examines the predictive power of implied correlation when controlling for the market realized variance estimated from intraday data. Bollerslev et al. (2009) estimate the realized variation from high-frequency observations to analyze the predictive power of the variance risk premium on market returns. According with this study, I use the market realized variance estimated from high-frequency observations obtained from Hao Zhou's webpage. I compute the implied variance of the market (IV_t) following the methodology described in equations (4.5) and (4.6) in Section 4.2.1. Then I obtain the market variance risk premium, VRPh_t, as the difference between IV_t and RVh_t. Table C.1 presents the results of the predictive regressions when the dependent variable is three-months-ahead market excess returns for the period from January 1996 to December 2010. Results for predictions at two and six months ahead are qualitatively similar.

Table C.1: Robustness: High-Frequency Market Realized Variance
This table reports the regression results of the three-months-ahead monthly market excess returns on the implied correlation (IC_t), the high-frequency market realized variance (RVh_t), the high-frequency variance risk premium (VRPh_t), and other control predictors defined in Section 4.2.3. The regression is based on monthly data for the period from January 1996 to December 2010. t-statistics are calculated using Newey-West standard errors. The explanatory variables are standardized to have mean zero and standard deviation equal to one, and the dependent variable is in percentage terms.

	I	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VII**
IC_t				1.54	1.25	1.47	1.47	1.57
				(4.27)	(3.88)	(3.77)	(2.94)	(3.91)
RVh_t	-0.50		0.39	-0.90		-0.70		
	(-1.57)		(0.76)	(-3.34)		(-1.19)		
VRPh_t		0.88	1.18		0.79	0.24	0.84	0.91
		(3.65)	(2.15)		(2.61)	(0.40)	(2.85)	(2.37)
$\log(\mathrm{P}_t/\mathrm{D}_t)$							-0.33	
							(-0.61)	
CAY_t								0.25
								(0.79)
DS_t								0.56
								(1.08)
TS_t								-0.17
								(-0.45)
RREL_t							1.25	1.43
							(3.13)	(3.01)
constant	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31
	(0.75)	(0.78)	(0.78)	(0.86)	(0.85)	(0.87)	(1.02)	(1.02)
adj. R^2 (%)	0.42	2.52	2.22	8.72	8.15	8.27	12.85	12.32

C.2 Predicting Cumulative Market Excess Returns

This section shows that implied correlation is strongly related to future cumulative returns. Table C.2 presents the results of the following predictive regression:

$$\frac{1}{h}\sum_{i=1}^{h} r_{t+i} = \beta_1(h) + \beta_2(h)\mathrm{IC}_t + \mathit{controls} + \epsilon_t \tag{C.1}$$

where r_{t+i} is the log of the monthly market excess return at time $t+i$. Hence, the dependent variable, $\frac{1}{h}\sum_{i=1}^{h} r_{t+i}$, is the (scaled by the horizon h) cumulative market excess return from $t+1$ up to $t+h$. The explanatory variables include monthly estimates for the implied correlation IC_t and a set of control predictors, all previously defined in Section 4.2.3, for the period from January 1996 to December 2010.

The regressors are standardized to have mean zero and standard deviation equal to one, and the dependent variable is in percentage terms. As the regression involves overlapping observations inducing serial correlation in the residuals, I employ Britten-Jones et al. (2011)'s standard errors. Panel A presents the results at different forecast horizons when only IC_t is included as a regressor. The figures indicates that the relationship between IC_t and future market risk premium is statistically significant at a 5% level for all of the forecast horizons examined. Particularly, I find that a one standard deviation increase in the IC_t produces 1.19% increase in the subsequent quarterly market risk premium. The highest adjusted R^2 is 40.87% for a bi-annually horizon. However, this result must be interpreted carefully as Boudoukh, Richardson and Whitelaw (2008) and Cochrane (2005) argue. They point out that the adjusted R^2 with overlapping observations increases with the return horizon.

When examining my findings in a multivariate context, I observe that the results are qualitatively the same as the one obtained when the dependent variable is monthly market returns. For the sake of brevity, Panel B only reports the prediction results for quarterly returns. In summary, the forecasting power of implied correlation is remarkably robust to the inclusion of different control variables; it is not captured by second moments, variance risk premia, valuation ratios and business cycle variables. Furthermore, I also observe that both IC_t and RREL_t perform well in a bivariate setting.

Table C.2: Predictive Regressions of Cumulative Market Excess Returns

The table presents the results of the following predictability regression: $\frac{1}{h}\sum_{i=1}^{h} r_{t+i} = \alpha_1(h) + \alpha_2(h)\mathrm{IC}_t + controls + \epsilon_t$, where the dependent variable is the (scaled by the horizon) cumulative market excess return from $t+1$ up to $t+h$, and h is the forecast horizon in months. IC_t is the implied correlation described in Section 4.2.1. *controls* is a set of control predictors defined in Section 4.2.3. The regression is based on monthly data from January 1996 to December 2010. The coefficients are estimated with ordinary least squares. Britten-Jones et al. (2011) t-statistics to correct for overlapping observations are reported in parenthesis. All of the variables are standardized to have mean zero and standard deviation equal to one.

h	**1**	**2**	**3**	**4**	**5**	**6**	**9**	**12**	**15**	**18**	**24**	**36**
Panel A												
IC_t	0.94	1.14	1.19	1.14	1.07	1.08	0.95	0.86	0.81	0.77	0.76	0.57
	(2.84)	(3.27)	(3.52)	(3.43)	(3.27)	(3.34)	(3.01)	(2.80)	(2.70)	(2.65)	(2.82)	(2.50)
constant	0.32	0.30	0.29	0.28	0.26	0.26	0.25	0.24	0.22	0.20	0.12	0.10
	(0.79)	(0.84)	(0.84)	(0.83)	(0.79)	(0.80)	(0.82)	(0.79)	(0.76)	(0.72)	(0.60)	(0.48)
adj. R^2 (%)	3.01	8.50	13.95	16.48	17.54	20.87	23.60	24.67	26.52	28.64	40.87	39.02

Table C.2: Predictive Regressions of Cumulative Market Excess Returns (cont.)

Panel C	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**
IC_t	1.44	1.41	1.20	1.17	1.24	1.20	1.20	1.51	1.59	1.07	1.50
	(3.80)	(4.10)	(3.67)	(3.08)	(3.73)	(3.62)	(3.14)	(4.41)	(4.22)	(2.89)	(4.49)
IV_t	-0.57										
	(-1.52)										
RV_t		-0.89		0.15							
		(-2.76)		(0.19)							
VRP_t			0.96	1.10		1.12				1.03	0.89
			(3.46)	(1.58)		(1.78)				(3.67)	(3.27)
$\overline{VRPi}_t$					0.84	-0.17					
					(3.00)	(-0.27)					
$\log(P_t/D_t)$							0.02			-0.28	
							(0.05)			(-0.69)	
CAY_t									0.13		
									(0.33)		
DS_t									-0.31		
									(-0.73)		
TS_t									-0.17		
									(-0.38)		
$RREL_t$								1.01	0.84		0.94
								(2.89)	(2.02)		(2.74)
constant	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29
	(0.86)	(0.87)	(0.88)	(0.88)	(0.87)	(0.88)	(0.84)	(0.88)	(0.86)	(0.88)	(0.91)
adj. R^2 (%)	16.15	21.07	23.09	22.69	20.73	22.70	13.46	23.00	23.33	23.22	30.82

Bibliography

Ahn, H., Bae, K. and Chan, K. (2001). Limit orders, depth, and volatility: Evidence from the stock exchange of hong kong, *Journal of Finance* **56**: 767–788.

Ait-Sahalia, Y., Mykland, P. and Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise, *Review of Financial Studies* **18**: 351–416.

Ait-Sahalia, Y., Mykland, P. and Zhang, L. (2011). Ultra high frequency volatility estimation with dependent microstructure noise, *Journal of Econometrics* **160**: 160–175.

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

Anand, A., Chakravarty, S. and Martell, T. (2005). Empirical evidence on the evolution of liquidity: Choice of market versus limit orders by informed and uninformed traders, *Journal of Financial Markets* **8**: 289–309.

Ang, A. and Bekaert, G. (2002). International asset allocation with regime shifts, *Review of Financial Studies* **15**: 1137–1187.

Ang, A. and Chen, J. (2002). Asymmetric correlations of equity portfolios, *Journal of Financial Economics* **63**: 443–494.

Ang, A., Hodrick, R., Xing, Y. and Zhang, X. (2006). The cross-section of volatility and expected returns, *Journal of Finance* **61**: 259–299.

Bates, D. (1991). The crash of '87: Was it expected? the evidence from option markets, *Journal of Finance* **46**: 1009–1044.

Beber, A. and Caglio, C. (2005). Order submission strategies and information: Empirical evidence from the nyse. *FAME Research Paper Series*, 146, International Center for Financial Asset Management and Engineering.

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

Bildik, R. (2001). Intra–day seasonalities on stock returns: Evidence from the turkish stock market, *Emerging Markets Review* **2**: 387–417.

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

Bollerslev, T. and Domowitz, I. (1993). Trading patterns and prices in the interbank foreign exchange market., *Journal of Finance* **48**: 1421–1443.

Bollerslev, T., Engle, R. and Wooldridge, J. (1988). A capital asset pricing model with time–varying covariances, *Journal of Political Economy* **96**: 116–131.

Bollerslev, T., Tauchen, G. and Zhou, H. (2009). Expected stock returns and variance risk premia, *Review of Financial Studies* **22**: 4463–4492.

Bouchaud, J., Mezard, M. and Potters, M. (2002). Statistical properties of stock order books: empirical results and models, *Quantitative Finance* **2**: 251–256.

Boudoukh, J., Richardson, M. and Whitelaw, R. (2008). The myth of long–horizon predictability, *Review of Financial Studies* **21**: 1577–1605.

Brandt, M. W. and Diebold, F. (2006). A no–arbitrage approach to range–based estimation of return covariances and correlations, *Journal of Business* **69**: 61–74.

Britten-Jones, M., Neuberger, A. and Nolte, I. (2011). Improved inference in regression with overlapping observations, *Journal of Business Finance & Accounting* **38**: 657–683.

Buraschi, A., Porchia, P. and Trojani, F. (2010). Correlation risk and optimal portfolio choice, *Journal of Finance* **65**: 393–420.

Buraschi, A., Trojani, F. and Vedolin, A. (2011). When uncertainty blows in the orchard: comovement and equilibrium volatility risk premia, *Journal of Finance* **Forthcoming**.

Burdett, K. and O'Hara, M. (1987). Building blocks: An introduction to block trading, *Journal of Banking and Finance* **11**: 193–212.

Buss, A. and Vilkov, G. (2011). Option–implied correlation and factor betas revisited. Working Paper, Goethe University.

Campbell, J. (1987). Stock returns and the term structure, *Journal of Financial Economics* **18**: 373–399.

Campbell, J. and Shiller, R. (1988). Stock prices, earnings, and expected dividends, *Journal of Finance* **43**: 661–676.

Campbell, J. and Shiller, R. (1989). The dividend–price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* **1**: 195–228.

Cao, C., Hansch, O. and Wang, X. (2008). Order placement strategies in a pure limit order book market., *The Journal of Financial Research* **31**: 113–140.

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

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

Chang, B., Christoffersen, P. and Jacobs, K. (2013). Market skewness risk and the cross-section of stock returns, *Journal of Financial Economics* **107**: 46–68.

Cho, J. and Nelling, E. (2000). The probability of limit order execution, *Financial Analysts Journal* **56**: 28–33.

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

Cochrane, J. H. (2005). *Asset Pricing*, Princeton University Press.

Cochrane, J., Longstaff, F. and Santa-Clara, P. (2008). Two trees, *Review of Financial Studies* **21**: 347–385.

Conrad, J., Dittmar, R. and Ghysels, E. (2013). Ex ante skewness and expected stock returns, *Journal of Finance* **68**: 85–124.

Coppejans, M., Domowitz, I. and Madhavan, A. (2001). Liquidity in an automated auction. Working paper, BlackRock; Barclays Global Investors.

Cornell, B. and Sirri, E. (1992). The reaction of investors and stock prices to insider trading, *Journal of Finance* **47**: 1031–1059.

Cosemans, M. (2011). The pricing of long and short run variance and correlation risk in stock returns. Working paper, Erasmus University.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy, *Journal of Business and Economic Statistics* **13**: 253–263.

Domowitz, I., Hansch, O. and Wang, X. (2005). Liquidity commonality and return co-movement, *Journal of Financial Markets* **8**: 351–376.

Driessen, J., Maenhout, P. and Vilkov, G. (2009). The price of correlation risk: Evidence from equity options, *Journal of Finance* **64**: 1377–1406.

Driessen, J., Maenhout, P. and Vilkov, G. (2012). Option–implied correlations and the price of correlation risk. Working paper, INSEAD.

Duong, H. and Kalev, P. (2008). Order book slope and price volatility. Working paper, Monash University.

Ekinci, C. (2008). *Intraday Liquidity in the Istanbul Stock Exchange,* , John Wiley & Sons,, chapter 6, pp. 77–94.

Ellul, A., Holden, C., Jain, P. and Jennings, R. (2007). Order dynamics: Recent evidence from the nyse, *Journal of Empirical Finance* **14**: 636–661.

Fama, E. and French, K. (1988). Dividend tields and expected stock returns, *Journal of Financial Economics* **22**: 3–25.

Fama, E. and French, K. (1989). Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* **25**: 23–49.

Fong, K. Y. and Liu, W. (2010). Limit order revisions, *Journal of Banking and Finance* **34**: 1873–1885.

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

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

Foucault, T., Moinas, S. and Theissen, E. (2007). Does anonymity matter in electronic limit order markets?, *Review of Financial Studies* **20**: 1707–1747.

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

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

Goetzmann, W., Li, L. and Rouwenhorst, K. G. (2001). Long–term global market correlations. Yale ICF Working Paper No. 08–04.

Griffiths, M., Smith, B., Turnbull, A. and White, R. (2000). The costs and determinants of order aggressiveness, *Journal of Financial Economics* **56**: 65–88.

Hall, A. and Hautsch, N. (2006). Order aggressiveness and order book dynamics, *Empirical Economics* **30**: 973–1005.

Hamao, Y. and Hasbrouck, J. (1995). Securities trading in the absence of dealers: Trades and quotes on the tokyo stock exchange, *Review of Financial Studies* **8**: 849–878.

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

Handa, P., Schwartz, R. and Tiwari, A. (2003). Quote setting and price formation in an order driven market, *Journal of Financial Markets* **6**: 461–489.

Hasbrouck, J. and Saar, G. (2009). Technology and liquidity provision: The blurring of traditional definitions., *Journal of Financial Markets* **12**: 143–172.

Hasbrouck, J. and Seppi, D., J. (2001). Common factors in prices, order flows, and liquidity, *Journal of Financial Economics* **59**: 383–411.

Ho, T., Stapleton, R. and Subrahmanyam, M. (1995). Multivariate binomial approximations for asset prices with non–stationary variance and covariance characteristics, *Review of Financial Studies* **8**: 1125–1152.

Hong, Y., Tu, J. and Zhou, G. (2007). Asymmetries in stock returns: Statistical tests and economic evaluation, *Review of Financial Studies* **20**: 1547–1581.

Jones, C., Kaul, G. and Lipson, M. (1994). Transactions, volume, and volatility, *Review of Financial Studies* **7**: 631–651.

Kaniel, R. and Liu, H. (2006). So what orders do informed traders use?, *Journal of Business* **79**: 1867–1913.

Keim, D. B. and Stambaugh, R. F. (1986). Predicting returns in the stock and bond markets, *Journal of Financial Economics* **17**: 357–390.

Kendall, M. (1953). The analysis of economic time series, part I: prices, *Journal of the Royal Statistical Society* **96**: 11–25.

Koski, J. and Scruggs, J. (1998). Who trades around the ex-dividend day?, *Financial Management* **27**: 58–72.

Krishnan, C., Petkova, R. and Ritchken, P. (2009). Correlation risk, *Journal of Empirical Finance* **16**: 353–367.

Lamont, O. (1998). Earnings and expected returns, *Journal of Finance* **53**: 1563–1587.

Latza, T. and Payne, R. (2013). Limit order flows, market order flows and high–frequency return forecasting. Working paper, Cass Business School.

Lettau, M. and Ludvigson, S. (2001). Consumption, aggregate wealth and expected stock returns, *Journal of Finance* **56**: 815–849.

Lewellen, J. (2004). Predicting returns with financial ratios, *Journal of Financial Economics* **74**: 209–235.

Lo, A. and MacKinlay, A. (1990). When are contrarian profits due to stock market overreaction?, *Review of Financial Studies* **3**: 175–206.

Lo, I. and Sapp (2010). Order aggressiveness and quantity: How are they determined in a limit order market, *Journal of International Financial Markets, Institutions & Money* **20**: 213–237.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets, *Journal of Finance* **56**: 649–676.

Martin, I. (2011). Simple variance swaps. NBER Working Paper No. 16884.

Martin, I. (2013). The lucas orchard, *Econometrica* **81**(1): 55–111.

Menkhoff, L., Osler, C. and Schmeling, M. (2010). Limit-order submission strategies under asymmetric information, *Journal of Banking and Finance* **34**: 2665–2677.

Moskowitz, T. (2003). An analysis of covariance risk and pricing anomalies, *Review of Financial Studies* **16**: 417–457.

Mueller, P., Stathopoulos, A. and Vedolin, A. (2012). International correlation risk. Working paper, London School of Economics.

Naes, R. and Skjeltorp, J. (2006). Order book characteristics and the volume-volatility relation: Empirical evidence from a limit order market, *Journal of Financial Markets* **9**: 408–432.

Parlour, C. (1998). Price dynamics in limit order markets, *Review of Financial Studies* **11**(4): 789–816.

Pascual, R. and Veredas, D. (2009). What pieces of limit order book information matter in explaining order choice by patient and impatient traders, *Quantitative Finance* **9**: 527–545.

Pascual, R. and Veredas, D. (2010). Does the open limit order book matter in explaining informational volatility?, *Journal of Financial Econometrics* **8**: 57–87.

Pastor, L. and Stambaugh, R. (2003). Liquidity risk and expected stock returns, *Journal of Political Economy* **111**: 642–685.

Payne, R. (2003). Informed trade in spot foreign exchange markets: an empirical investigation, *Journal of International Economics* **61**: 307–329.

Pollet, J. and Wilson, M. (2010). Average correlation and stock market returns, *Journal of Financial Economics* **96**: 364–380.

Poon, S. and Granger, C. W. J. (2003). Forecasting volatility in financial markets: A review, *Journal of Economic Literature* **16**: 478–539.

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

Rindi, B. (2008). Informed traders as liquidity providers: Anonymity, liquidity and price formation, *Review of Finance* **12**: 497–532.

Roll, R. (1988). The international crash of october 1987, *Financial Analysts Journal* **44**: 19–35.

Rosu, I. (2009). A dynamic model of the limit order book, *Review of Financial Studies* **22**: 4601–4641.

Rosu, I. (2012). Liquidity and information in order driven markets. Working paper, HEC Paris.

Schwert, G. W. (1989). Why does stock market volatility change over time, *Journal of Finance* **44**: 1115–1153.

Valenzuela, M. and Zer, I. (2013). Competition, signaling and non–walking through the book: Effects on order choice, *Journal of Banking and Finance* . Forthcoming.

Wald, J. K. and Horrigan, H. T. (2005). Optimal limit order choice, *The Journal of Business* **78**: 597–620.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* **21**: 1455–1508.

Zhang, L., Mykland, P. and Ait-Sahalia, Y. (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data, *Journal of the American Statistical Association* **100**: 1394–1411.