London School of Economics and Political Science

Essays on Financial Economics

Ilknur Zer Boudet

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).

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

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 Marcela Valenzuela, also a PhD candidate from the London School of Economics.

To Pierre and my family

Abstract

This thesis consists of five chapters. The first chapter previews the analysis of the following three chapters. In the second chapter, 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. My co-author and I 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 analyses the relationship between the firms' disclosure decisions and the market expected value of default probabilities. I use option prices to estimate the option implied probability of default, whereas the level of disclosure is measured by a self-constructed voluntary disclosure index for the largest 85 U.S. bank holding companies. I provide evidence that the enhanced disclosure is followed by reduced market implied default probabilities in the subsequent year. This evidence suggests that by mitigating the information asymmetries between the bank management and their depositors and regulators, disclosure affects investors' assessments of the riskiness of a bank. Finally, Chapter 5 sums up and points out directions for further work.

Acknowledgements

I cannot adequately express my gratitude to my supervisor Dr. Jon Danielsson for his undying belief in me. On the one hand, he left me all the freedom I wanted, on the other hand, whenever I needed help, he was there to share his knowledge and to guide me. I sincerely thank Jon for his support and especially for his friendship.

I greatly appreciate the support from Dr. Kevin James, Dr. Christian Julliard, Prof. Thorsten Rheinlander, and Prof. Piotr Fryzlewicz. They were always ready to listen to me, to give comments, and cheer me up. Their insightful comments and probing questions undoubtedly resulted in significant contributions to the development of this thesis.

A special thanks to my co-author and dearest friend Marcela Valenzuela. Without her, nothing would be the same. I would not come to the office with the same desire. I will miss every single minute we have spent together, all the fun we have had in the last five years, and even the very long working hours before the deadlines.

I am grateful to Dr. Dong Lou and Prof. Richard Payne for accepting to act as co-examiners, reading drafts of this dissertation and for their insightful comments.

I would like to thank Coskun Gunduz for his invaluable help, spending long days and nights with me for the reconstruction of the limit order book data set, which I used in the first two chapters of this thesis.

Thanks to Sean Lew, Nelson Camanho, Victoria Dobrynskaya, Svetlana Bryzgalova, Christian von Drathen, Dr. Wolfram Jahn, Dr. Vicente Cunat, Dr. Amil Dasgupta, Prof. Daniel Ferreira, Dr. Michela Verardo, Dr. Konstantinos Zachariadis, Mary Comben, and Osmana Raie for their friendship, help, and for making my life here at LSE better and easier.

I am grateful to Alison and Tim Frost for providing me the Cairn Capital scholarship award for four consecutive years. Without their support, studying for a PhD at LSE could only be a dream.

I would like to thank my parents for instilling in me a desire to achieve my goals and a commitment to finish what I started. They have always been proud of me for simply stepping up to take on a challenge. I would not even come to London without the support of my family - Mum, Dad, and Oznur. Last but not least, I would like to thank my husband, Pierre. Thanks to Pierre for being in my life, for his constant support, and for being there for me in those moments I was riddled with self-doubt. This thesis is dedicated with love to the amazing man who has shaped my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Informational Content of a Limit Order Book

The following two chapters of this thesis focus on the effects of the information content of a limit order book in a pure order driven market using high frequency data. In the first part, “Global Depth and Future Volatility”, we propose a new way of summarizing the liquidity distribution in a limit order book and further examine the predictive power of this distribution over future volatility. On the other hand, “Competition, Signaling and Non-walking Through the Book: Effects on Order Choice“, investigates how the state of a limit order book affects investors’ order submission strategies.

As of today, most of the equity and derivatives exchanges around the world are either pure order driven or at least allowing limit orders in addition to the on-floor market making. The role of limit orders in trading processes expanded progressively over the last decade. This shift in trading is followed by a tremendously increasing academic literature.

The state of the order book, as being the collection of all outstanding limit orders, shows the actual demand and supply for the underlying stock, it contains non-negligible information about the short-term price formation process, since the latter is a result of the state of the limit order book and the order flow. Moreover, understanding the dynamics of limit orders is crucial to understand their contribution to the liquidity. Besides academic purposes, studying the order book is interesting for practical purposes as well. Regulatory bodies want to set the rules of exchanges to produce fair trading that maximizes the flow. For market participants on the other hand, the choice of market or limit orders, or order aggressiveness in general, depends on the structure and dynamics of the limit order book since that affects the execution costs.

Even though that beneficial, construction of a limit order book is not an easy task and the limit order book data availability is restricted. Using the order flow and trade data provided by the Istanbul Stock Exchange, we reconstruct the limit order book dynamically and obtain useful variables to use in further financial analysis. Some stock exchanges or information companies supply five or ten best prices for the bid and ask side and corresponding volumes. However, by creating the whole data book, we will not be limited to the available periodicals and we can reach a very high frequency trading information. Both chapters 2 and 3 of this thesis use the reconstructed limit order book, order flow, and trade data from the Istanbul Stock Exchange for the period of June and July 2008. Our data consists of the biggest 30 stocks listed on the exchange, which corresponds to 75% of the total trading volume of the Istanbul Stock Exchange for the period under consideration.

Chapter 3 contributes to the vast literature that investigates the effects of a state of a limit order book and the characteristics of an asset, such as volatility, on order choice. First, we examine the trading patterns of agents given a specific market rule: non-walking through the book. In the Istanbul Stock Exchange, similar to the Australian Stock Exchange, the Sao Paulo Stock Exchange (Bovespa), and the Stock Exchange of Hong Kong, for example, walking through the book is not allowed. In this case, the unexecuted portion of a "large" market order is converted to a limit order rather than walk up or down the book to fully executed. This market mechanism affects the order choice of a trader since, in this case, the cost of a large market order is lower.

Second, the chapter tests whether competition or signaling effects, two theories that have been proposed in the existing literature by Parlour (1998) and Goettler, Parlour and Rajan (2005), Goettler, Parlour and Rajan (2009) respectively, dominate each other for depth beyond the best quotes. The signaling effect is based on an early dynamic equilibrium model of Parlour (1998). She analyses the effects of the depth of a limit order book and concludes that an increase in the same–side thickness of the limit order book crowds out the limit orders on that side due to increased competition. On the other hand, we borrow the term “signaling effect” from two recent theory papers of Goettler et al. (2005) and Goettler et al. (2009); the volume of orders accumulated away from the best quotes signals the mispricing in the market.

By employing a two-stage sequential ordered probit model, we show that none of the price information, either spread or price distance variables, matter for a market order trader in her order choice. We believe this is specific to markets in which walking through the book is not allowed. Also, we show that the competition effect is persistent beyond the best quotes and dominates the signaling effect at every level, being strongest for the volume of orders waiting at the second best quotes. Finally, we focus on the trading patterns of institutional vs. individual investors. We conclude that institutional investors consider only the level of competition while deciding to submit a market or a limit order, and none of the other state variables affect the order choice of institutions. This may suggest that institutional traders place their orders based on their own private valuations rather than the information provided by the limit order book.

In summary, “Competition, Signaling and Non-walking Through the Book: Effects on Order Choice“ concludes that a state of a limit order book provides essential information for a trader who wants to design an appropriate order submission strategy.

Moreover, existing theoretical microstructure literature illustrates that limit orders are information driven. In other words, informed traders may exploit their informational advantage by submitting limit orders (e.g. Foucault (1999), Foucault, Kadan and Kandel (2005), and Rosu (2012)). Chapter 2 of the thesis, "Global Depth and Future Volatility" builds upon these ideas. We specifically ask whether the orders waiting in a limit order book at a given time has any explanatory power over future volatility.

Volatility and liquidity are of particular interest in a wide spectrum of theories and applications in finance, for example option pricing or portfolio valuation. In microstructure content, volatility is important because the execution probability of limit order increases with volatility. Putting differently, the probability of the current price hits the pre-determined limit price increases when volatility is higher.

In order to measure how liquidity is distributed in a given limit order book, we develop a new measure, global depth. It has two ingredients: it aggregates the relative quoted depth provided by different price levels and it assigns weights decreasing with distance from the best bid or ask price. The economic link between the global depth and volatility follows from two recent theory papers of Goettler et al. (2005) and Goettler et al. (2009): when volume of orders are accumulated away from the best quotes increases, this signals to the market that the current quotes are mispriced. In this case, we argue that price jumps are more likely, which will in the end translate into higher price volatility.

We perform both in-sample and out-of-sample tests. In order to test the in-sample explanatory power of global depth on future market volatility, we employ a standard OLS regression model. The dependent variable is the mid-quote volatility of the ISE-30 index calculated by employing two-scales realized volatility (TSRV) estimator proposed by Ait-Sahalia, Mykland and Zhang (2011). The main explanatory variables are the aggregated global depth for buy and sell sides of the market, which are calculated as the cross-sectional averages of individual stocks' global depths. We document that global depth significantly and negatively predicts the intraday market volatility up to 150 minutes ahead. Moreover, 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. Similarly, the out-of-sample tests reveal 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.

1.2 Disclosure and Implied Probability of Default

Chapter 4 asks whether the level of disclosure chosen by the management affects the market assessment of the riskiness of a bank. Several commentators have pointed to the inadequacy of transparency of the financial institutions before the 2008 crisis and argued that the investors' limited information of the risks held by financial intermediaries amplified the crises. It is because in bad times, there is a flight to quality. Possible sudden and large amounts of selling pressure on "bad" may lead to withdrawals of liquidity, or even the breakdown of trading. In turn, the question of whether the severity of crises would be smaller if market participants have access to better information and sufficiently punish the banks which took excessive risks is not solely an academic interest.

Chapter 4 of this thesis, "Disclosure Practices and Option Implied Probability of Default" provides a disclosure template as a simple way of summarizing and assessing the extent of disclosure by financial institutions. The template is based on a scoring system constructed on the summary measures proposed by the December 2009 and 2011 Financial Stability Reports of the Bank of England. By using a sample of the largest publicly listed 85 U.S. bank holding companies for the period 1998–2011, I document a significant and negative relationship between the level of disclosure and various bank enterprise risks, with a focus on the market implied default risk, where the latter is estimated from the option prices based on the methodology proposed by Capuano (2008).

Economic theory suggests that rational investors interpret any piece of withheld information that can be credibly disclosed as conveying bad news (see for instance Grossman and Miller (1980), Grossman (1981), Milgrom (1981), and Verrecchia (2001)). The idea is based on the Akerlof (1970)'s asymmetric information argument, and that firms' disclosure can be effective in reducing information asymmetries between the managers, investors, and supervisory authorities.

This chapter makes two important contributions. First, to the best of my knowledge, it is the first study in the literature that formally investigates the relationship between the level of disclosure and market assessment of the default probability of an institution. Second, I propose a template to measure the level of voluntary disclosure, which is constructed by using publicly available 10–K and annual reports data. In contrast to the self-constructed disclosure indices proposed in the literature (e.g., Botosan (1997), Baumann and Nier (2003), and Francis, Nanda and Olsson (2008)), our disclosure index mainly focuses on the disclosure of the riskiness, rather than the profitability of an institution. Moreover, it considers the recent risk factors that threaten the financial system, like liquidity or spillover risk rather than the credit risk, which turn to be compulsory due to current Basel regulations. Finally, despite its data limitations, our validating experiments suggest its adequacy on measuring the level of management's decision of disclosure.

I first employ a panel regression model with year and bank fixed effects and test the statistical relationship between the level of disclosure and the market implied default probability of a given bank holding company. The results confirm a negative and economically significant relationship; one standard deviation increase in the current level of disclosure is associated with a 19% and 27% decrease in the next year's and three months' probability of default, respectively. The results are robust to the inclusion of various bank characteristics, such as size, beta, capital buffers, or non-performing loans.

Next, I examine whether enhanced disclosure is associated with other enterprise risks. I consider four other risks: aggregate risk, measured by the stock return volatility, downside risk, proxied by the implied volatility estimated from the option prices, the systematic and the idiosyncratic risk of a bank holding company estimated from the CAPM model. I conclude that banks with higher level of disclosure this year benefit from lower enterprise risk in the following period. Finally, in a cross-sectional setting, I examine the determinants of disclosure. I find that bigger banks, banks that disclose more information last year, and less profitable institutions are more likely to disclose more.

Chapter 2

Global Depth and Future Volatility

Co-authored with Marcela Valenzuela (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$^{\text{th}}$ (8$^{\text{th}}$) 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 $\Delta = 0, 1, 2, .., \Delta_{\text{c}}$.[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 $\Delta_{\text{c}} = 30$.

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.

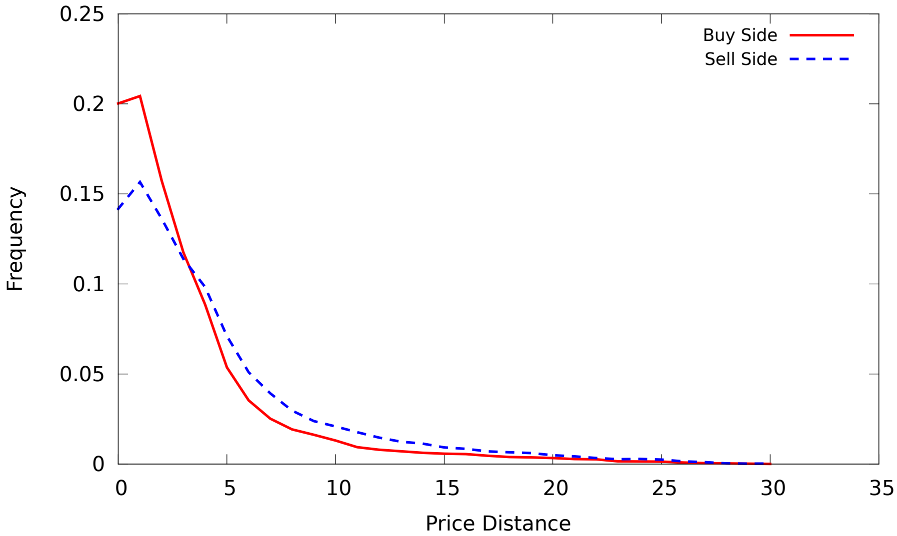
By definition, global depth is related to the standard "local" depth measures, i.e., the quoted depth up to a given threshold. An investigation of their relationship is presented in Appendix A, Section A.3. From this analysis, we conclude that although they are positively and significantly correlated, 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

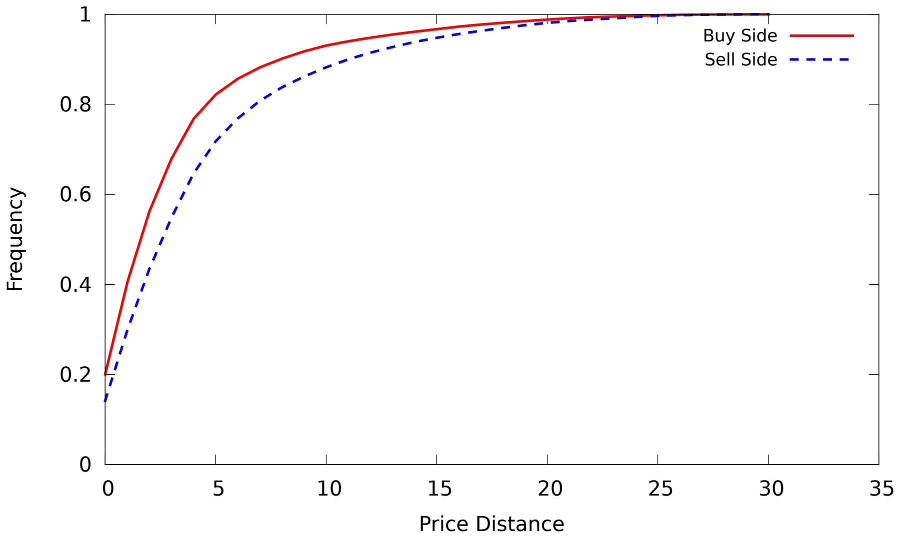
Figure 2.1, Panel A plots the limit order book probability density function (LOB–PDF) averaged across intervals, days, and stocks (average LOB–PDF), whereas Panel B plots the corresponding cumulative distribution (LOB–CDF). Panel A reveals that for both sides of the market, the frequency of the orders 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:

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

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:

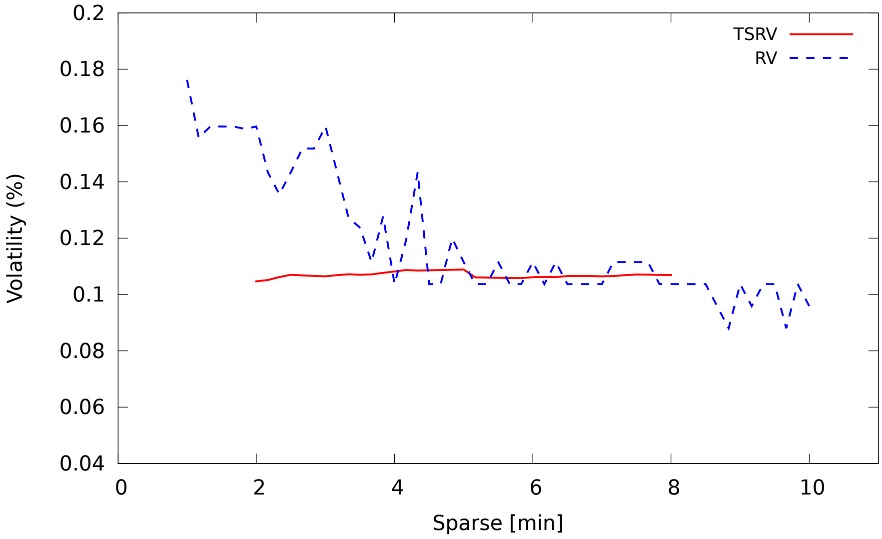
$$\langle X, X \rangle_T^{TSRV} = \sqrt{[Y, Y]_T^{ave} - \frac{1}{K}[Y, Y]_T^{(all)}}, \tag{2.5}$$

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

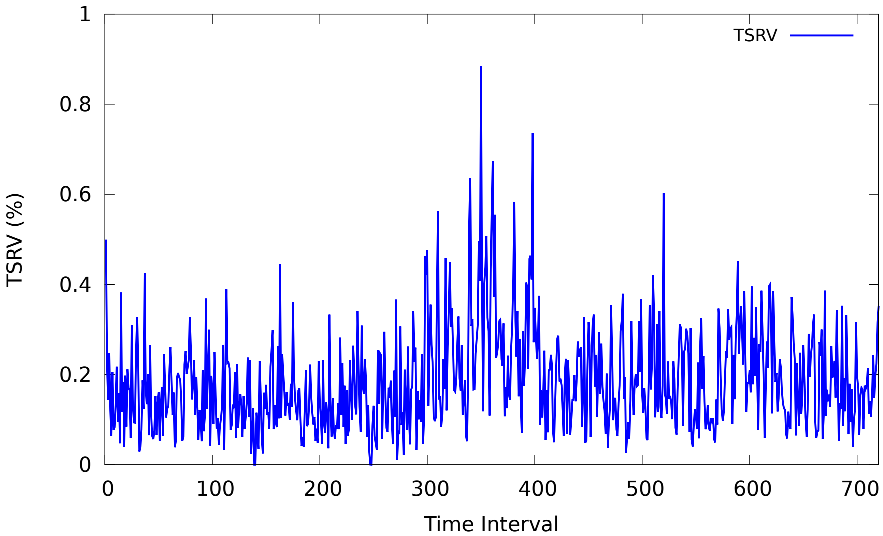
$$[Y, Y]_T^{ave} = \frac{1}{K} \sum_{k=1}^{K} [Y, Y]_T^{sparse,k}.$$

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.

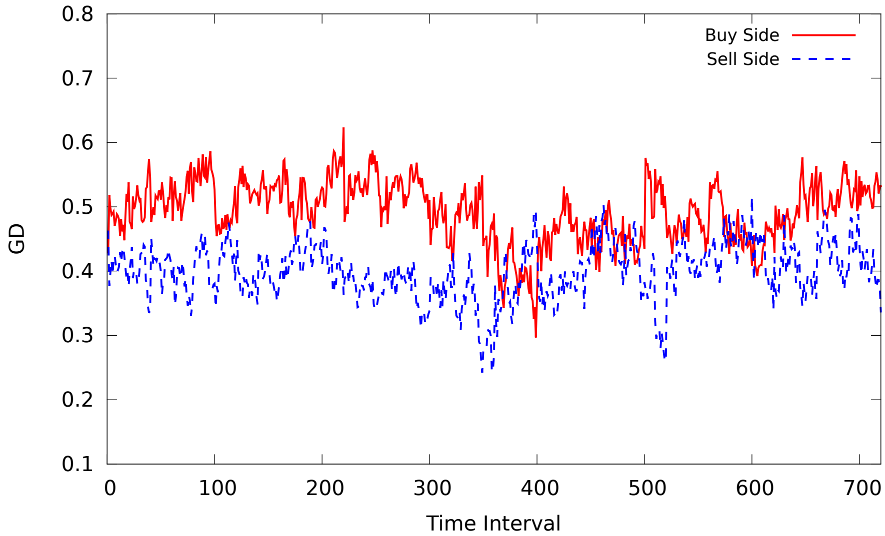


Figure 2.3: This figure plots the intraday estimates of global depth evaluated at the optimal decay parameter for buy and sell sides of the market. The estimation is based on the sampling of a trading day in 15-minute intervals.

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:

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

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:

$$\mathrm{logQS}_{s,\tau} = \frac{\ln \frac{p_\tau^A}{p_\tau^B}}{\ln\left(q_\tau^A \cdot q_\tau^B\right)}, \tag{2.8}$$

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,

$$\begin{aligned}\mathrm{DHW}_{s,\tau} = &\left[\sum_{k=1}^{m-1} q_k^A p_k^A + \left(Q_s - \sum_{k=1}^{m-1} q_k^A\right) p_m^A\right] \\ &- \left[\sum_{k=1}^{m'-1} q_k^B p_k^B + \left(Q_s - \sum_{k=1}^{m'-1} q_k^B\right) p_{m'}^B\right],\end{aligned} \tag{2.9}$$

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. $\text{GD}_\tau(\widehat{\lambda})$ 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 $\text{GD}_\tau(\lambda = 0)$ 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). $\overline{\mathrm{GD}}^{\mathrm{buy}}_{\tau}(\widehat{\lambda})$ and $\overline{\mathrm{GD}}^{\mathrm{buy}}_{\tau}(\lambda=0)$ are the cross-sectional averages of global depth of individual stocks, $\mathrm{GD}^{\mathrm{buy}}_{s,\tau}$, evaluated at the optimal decay factor $\widehat{\lambda}$ and $\lambda = 0$, respectively, as outlined in Section 2.5.2. $\overline{\mathrm{DEPTH0}}^{\mathrm{buy}}_{\tau}$ ($\overline{\mathrm{DEPTH0}}^{\mathrm{sell}}_{\tau}$) is the cross-sectional average of volume at the best quotes for the buy (sell) side of the market, whereas $\overline{\mathrm{DEPTH1\_5}}^{\mathrm{buy}}_{\tau}$ and $\overline{\mathrm{DEPTH1\_5}}^{\mathrm{sell}}_{\tau}$ are the accumulated volume of orders from the second to the 5$^{\mathrm{th}}$ 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.: $\sigma^M_{\tau+15\mathrm{min}}$	**I**	**II**	**III**	**IV**	**V**	**VI**
$\overline{\mathrm{GD}}^{\mathrm{buy}}_{\tau}(\widehat{\lambda})$			-0.034		-0.033	-0.036
			(-6.82)		(-6.51)	(-6.75)
$\overline{\mathrm{GD}}^{\mathrm{sell}}_{\tau}(\widehat{\lambda})$			-0.021		-0.020	-0.013
			(-3.51)		(-2.50)	(-1.41)
$\overline{\mathrm{GD}}^{\mathrm{buy}}_{\tau}(\lambda=0)$		-0.026				
		(-5.49)				
$\overline{\mathrm{GD}}^{\mathrm{sell}}_{\tau}(\lambda=0)$		-0.013				
		(-1.85)				
$\overline{\mathrm{DEPTH0}}^{\mathrm{buy}}_{\tau}$				-0.013	0.004	-0.010
				(-2.39)	(0.08)	(-1.14)
$\overline{\mathrm{DEPTH0}}^{\mathrm{sell}}_{\tau}$				-0.011	-0.002	0.002
				(-2.41)	(-0.54)	(0.40)
$\overline{\mathrm{DEPTH1\_5}}^{\mathrm{buy}}_{\tau}$						-0.019
						(-1.86)
$\overline{\mathrm{DEPTH1\_5}}^{\mathrm{sell}}_{\tau}$						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). $\overline{\text{GD}}^{\text{buy}}_{\tau}(\widehat{\lambda})$ 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). 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**
$\overline{\text{GD}}^{\text{buy}}_{\tau}(\widehat{\lambda})$		-0.034	-0.030	-0.030	-0.028
		(-6.82)	(-6.83)	(-5.44)	(-5.23)
$\overline{\text{GD}}^{\text{sell}}_{\tau}(\widehat{\lambda})$		-0.021	-0.017	-0.011	-0.011
		(-3.51)	(-3.14)	(-1.50)	(-1.56)
$\overline{\text{SLOPE}}_{\tau}$			0.013	0.017	0.015
			(1.87)	(2.79)	(2.26)
$\overline{\text{relSPR}}_{\tau}$			0.028	0.022	0.021
			(5.41)	(2.28)	(2.12)
$\overline{\text{NT}}_{\tau}$			0.008		0.008
			(1.34)		(1.24)
$\overline{\text{AQ}}_{\tau}$			-1.17		0.001
			(-0.02)		(0.28)
$\overline{\text{AMR}}_{\tau}$				0.002	0.002
				(0.47)	(0.58)
$\overline{\text{logQS}}_{\tau}$				0.014	0.014
				(0.94)	(0.96)
$\overline{\text{DHW}}_{\tau}$				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). $GD^{buy}_{s,\tau}$ and $GD^{sell}_{s,\tau}$ 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
$GD^{buy}_{\tau}(\widehat{\lambda})$	-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]
$GD^{sell}_{\tau}(\widehat{\lambda})$	-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]
$SLOPE_{\tau}$		-0.006	0.035	0.028		-0.046	0.058	0.047
		(-0.82)	(5.56)	(4.80)		[33/20]	[37/81]	[27/75]
$relSPR_{\tau}$		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]
$logQS_{\tau}$			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). $\overline{\mathrm{GD}}_{\tau}^{\mathrm{buy}}(\widehat{\lambda})$ is the cross-sectional average of global depth of individual stocks, $\mathrm{GD}_{s,\tau}^{\mathrm{buy}}$, evaluated at the optimal decay factor $\widehat{\lambda}$. Similarly, $\overline{\mathrm{SLOPE}}_{\tau}$ and $\overline{\mathrm{relSPR}}_{\tau}$ 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.							
$\overline{\mathrm{GD}}_{\tau}^{\mathrm{buy}}(\widehat{\lambda})$	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
$\overline{\mathrm{GD}}_{\tau}^{\mathrm{sell}}(\widehat{\lambda})$	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
$\overline{\mathrm{SLOPE}}_{\tau}$	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
$\overline{\mathrm{relSPR}}_{\tau}$	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.							
$\overline{\mathrm{GD}}_{\tau}^{\mathrm{buy}}(\widehat{\lambda})$	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
$\overline{\mathrm{GD}}_{\tau}^{\mathrm{sell}}(\widehat{\lambda})$	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
$\overline{\mathrm{SLOPE}}_{\tau}$	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
$\overline{\mathrm{relSPR}}_{\tau}$	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 $\overline{\text{GD}}^{\text{buy}}$. 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 $\overline{\text{GD}}^{\text{buy}}$ and $\overline{\text{relSPR}}$ 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 Marcela Valenzuela (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 20$^{\text{th}}$ largest stock exchange in the world and 5$^{\text{th}}$ 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:

$$\begin{aligned} \frac{\partial P(Y=0)}{\partial X_j} &= \frac{\partial \Phi(\delta - \sum_{k=1}^{K} \beta_k X_{k,t-1})}{\partial X_j} \\ &= -\phi(\delta - \sum_{k=1}^{K} \beta_k X_{k,t-1})\beta_j, \end{aligned} \tag{3.9}$$

$$\begin{aligned} \frac{\partial P(LO=i)}{\partial X_j} &= \frac{\partial (\Phi(\delta_i^{lo} - \sum_{k=1}^{K} \theta_k X_{k,t-1}) - \Phi(\delta_{i-1}^{lo} - \sum_{k=1}^{K} \theta_k X_{k,t-1}))}{\partial X_j} \\ &= [\phi(\delta_{i-1}^{lo} - \sum_{k=1}^{K} \theta_k X_{k,t-1}) - \phi(\delta_i^{lo} - \sum_{k=1}^{K} \theta_k X_{k,t-1})]\theta_j, \end{aligned} \tag{3.10}$$

$$\begin{aligned} \frac{\partial P(MO=0)}{\partial X_j} &= \frac{\partial \Phi(\delta_1^{mo} - \sum_{k=1}^{K} \gamma_k X_{k,t-1})}{\partial X_j} \\ &= \phi(\delta_1^{mo} - \sum_{k=1}^{K} \gamma_k X_{k,t-1})\gamma_j, \end{aligned} \tag{3.11}$$

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 (Vsame^{2-3}, Vsame^{2-4} and Vsame^{2-5}). The signaling effect will then be captured by Vsame^{4-10}, Vsame^{5-10} and Vsame^{6-10}, 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:

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

- Step 2– order choice of patient traders:

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

[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]

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

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^{th} and the 75^{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 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	-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^{th} and the 75^{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 $\gamma_{\text{SPR}} = \gamma_{\text{Dopp}^1\_^2} = \gamma_{\text{Dopp}^2\_\text{max}} = 0$ 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}_$^{\mathbf{max}}$**	**Dopp^{1}_^{2}**	**Dopp^{2}_$^{\mathbf{max}}$**
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

Disclosure Practices and Option Implied Probability of Default

4.1 Introduction

Investors' limited information of the risks held by financial intermediaries is argued to amplify both phases of the recent credit cycle. Opacity of banks, coupled with ever increasing complexity, contributed to the general mispricing of risks as investors badly misunderstood the risks inherent in structured products. Reliable, timely and granular information disclosure can go some way towards alleviating these problems. Hence, efforts to promote transparency through greater disclosure have been an important theme of the discussions of the current banking reform proposals and reports of regulatory authorities. The Basel II, Pillar 3 recognizes disclosure as a way to impose strong incentives on banks to perform less risky activities. December 2009 and 2011 Financial Stability Reports of Bank of England underline the level of the transparency and enhanced disclosure as a toll to mitigate the informational frictions especially in stress times. Thus the relationship between disclosure and riskiness of a financial institution is of key importance to investors, banks, and regulators. Despite of its importance, few empirical studies adress this. This paper aims to fill this gap. We create a measure of voluntary disclosure with public data and then use this to formally investigate the relationship between the level of disclosure and market assessment of the riskiness of a bank.

Given the level of balance sheet risk, if sufficient transparency imposes incentives on banks to hold less risky positions through the monitoring of the investors, then banks that disclose more information should choose less risky activities. In other words, investors or debt holders may exercise a direct market discipline allowing a reduction in bank's default probability.[1] Even if the bank does not choose to perform less risky activities, rational investors can interpret the absence of disclosure as a negative signal about the firm value, since less informed party presumes that withheld information is a less favorable information (e.g., Grossman and Miller (1980), Grossman (1981), Milgrom (1981), and Verrecchia (2001)). Putting differently, higher disclosure reduces the information asymmetries between the bank management and their depositors and regulators. This in turn may affect investor's assessment of the riskiness of the firm or reduce the heterogeneity of beliefs about the true value of the firm (Lambert, Leuz and Verrecchia, 2007). Motivated by the aforementioned theoretical papers, this study analyses the relationship between the firms' disclosure decisions and the market expected value of default probabilities.

Our main hypothesis is that banks with higher level of disclosure in the current year benefit from lower market implied default probabilities in the following year. We test this hypothesis under the alternative that the level of disclosure does not have any real impact on investors' assessment on the default probabilities of banks. This may be because of the failure of market discipline, a market mechanism in which investors have sufficient information to assess and incentives to monitor risk taking behavior of banks (Crockett, 2002). Although increased transparency is a necessary condition for investors to reach informed judgements, it is not sufficient. Investors

[1] In a cross-country study, Tadesse (2006) shows highly regulated disclosures lead to lower financing costs and lower risk profile. Nier and Baumann (2006) show that banks that disclose more information on their risk profile are subject to stronger degree of market discipline and choose to hold higher capital buffers to limit their probability of default.

only price the risks which they actually bear. If market participants are insured then their incentives to monitor and punish the risks are reduced.

This paper makes two important contributions. First, it is the first study in the literature that formally investigates the statistical relationship between the level of disclosure and market assessment of the credit riskiness of a bank. Second, we propose a template to measure the level of voluntary disclosure, which is constructed by using publicly available data. Despite its data limitations, our validating experiments suggest its adequacy on measuring the level of management's decision of disclosure.

In order to measure an institution's default probability, we employ the methodology proposed by Capuano (2008) and further developed by Vilsmeier (2011) and use option prices to estimate the implied probability of default (IPoD). By using techniques of maximum entropy, the IPoD model extracts market-based default probabilities. As options are forward-looking instruments, using option prices brings us the advantage of extracting information on market participants' expectations. Moreover, option prices suffer less from the structural breaks by updating faster to new market conditions compared to historical data (Buss and Vilkov, 2011). Then we would expect our default measure to have an advantage in unstable periods.

Besides the default probability, in order to investigate our main hypothesis, we need to measure the level of disclosure. Our empirical proxy for disclosure is a self-constructed voluntary disclosure index, mainly based on the summary measures proposed in December 2009 and 2011 Financial Stability Reports of the Bank of England.[2] Our index gauges the level of disclosure provided on four main categories: liquidity risk profiles of the companies, risk positions of key group affiliates and subgroups, intra-annual information, and finally exposures between financial institutions and exposures to hidden risks.

[2]We sincerely thank Christian Castro from the Bank of Spain, Rhiannon Sowerbutts, and Peter Zimmerman from the Bank of England for insightful comments and suggestions for the creation of the disclosure template.

We hand-collected data to construct the disclosure index. Data collection and validation requires some effort. Hence, we restrict ourselves to the publicly open largest 85 U.S. bank holding companies in terms of asset value as of December 2007 for the period 1998–2011. This accounts for the 76% of the total assets of the U.S. banking system. We select our sample based on 2007 because we want to include the actually defaulted bank holding companies in the 2008 crises. Our focus on bank holding companies is motivated by three. First, they file periodic reports to the Securities and Exchange Commission (SEC), from which we are able to obtain 10–K and proxy statements to construct the disclosure index. Second, U.S. bank holding companies are regulated by the Federal Reserve and the FDIC. Hence, they are subject to uniform requirements for compulsory disclosures, which is important to identify voluntary disclosures. Finally, a typical bank holding company has a complex structure. It is comprised of several independent subsidiaries and involved in a wide range of financial activities. This may enhance the importance of granular financial disclosure for investors to identify correctly the risk taking behavior.

The main hypothesis is tested by employing a panel regression model. First, we measure our dependent variable as the annual average of the implied default probability estimates between two 10–K statement disclosure dates. Thus we test whether the revealed information has an all-year-long effect. Second, we use the three-months averages of the implied default probability estimates following the announcement of a 10–K statement. It is compulsory for a bank holding company to fill quarterly 10–Q reports to the SEC. Although those reports are not as comprehensive as the annual 10–K statements, they still provide a continuing view of a company's financial position. Hence, it is likely that the informativeness of an annual report decreases with the releases of 10–Q statements, i.e., after three months of the release of an annual report. Results confirm our hypothesis; a higher level of disclosure is associated with lower levels of market implied default probability both in the following three months and in the following year, where the relationship is stronger for the former. The documented association is economically significant: one standard deviation increase in the current level of disclosure is associated with a 19% and 27% decrease in the next year's and three months' probability of default, respectively. In all of the specifications, we include year and bank fixed effects to capture for any time-invariant heterogeneities across bank holding companies. The results are robust to the inclusion of various bank characteristics, such as size, beta, capital buffers, or non-performing loans.

The underlying assumption under the baseline panel regression model is that the level of disclosure is exogenous after controlling for the market risk, bank holding company characteristics, year, and bank fixed effects. However, a bank holding company's past level of market implied default risk can affect both the current level of default risk and the decision on the level of disclosure. Hence, the model we propose has a potential endogeneity issue, which may affect the choice of estimation procedure. Generally one can either employ a panel regression model, which ignores endogeneity or adopt a procedure that incorporates endogeneity, such as the Arellano and Bond (1991) dynamic panel GMM estimator. Although we focus our attention on the panel approach, which is simpler and the one that literature generally opt to do, in order to see whether endogeneity matters, we also test our main hypothesis by employing the GMM model. The coefficient on aggregate disclosure score continues to be negative and significant and the results confirm the joint validity of our instruments.

Next, we examine whether enhanced disclosure is associated with other bank holding company enterprise risks. In line with the findings of Bushee and Noe (2000), Baumann and Nier (2004), and Kothari, Li and Short (2009), our results provide evidence that banks with higher level of disclosure benefit from lower stock volatility. Economic theory suggests a number of plausible explanations. First, by mitigating uncertainty, higher disclosure may reduce the magnitude of the impact of news, which in turn may reduce realised volatility. Second, reduced information asymmetries in turn may reduce the adverse price impact of a large trade (Diamond and Verrecchia, 1991). Then we focus on the implied volatility estimated from the option prices as a measure of downside risk. Similar to the findings of Ederington and Lee (1996) and Rogers, Skinner and Van Buskirk (2009) we find that implied volatility, the market's participants risk-neutral expectations of volatility, declines following information releases. Finally, we document that enhanced disclosure in this year significantly decreases both the systematic and the idiosyncratic risk of a bank holding company in the following year, whereas the association is stronger for the latter. Hence, we argue that disclosure helps to alleviate informational frictions and lead to a more efficient allocation of risk.

Finally, in a cross-sectional setting, we examine the determinants of disclosure. We find that banks, which disclose more information last year, continue to provide higher levels of disclosure compared to its peers this year. In line with the existing literature, we also find that bigger and less profitable institutions are more likely to disclose more.[3] Moreover, higher levels of observed systematic risk encourage managers to disclose more information in the subsequent year. This could be argued as an evidence of market discipline; increased public information helps investors to assess the risk taking behavior of a bank, and changes in market dynamics influence the decisions made by bank management.

Our paper is related to the literature that investigates the consequences of corporate disclosure on capital markets. We contribute to the literature by examining the link between the voluntary disclosures and various firm enterprise risks. There is an extensive empirical literature that studies the link between the accounting information and the cost of capital, liquidity, and stock return volatility. Using the analyst disclosure ratings provided by the AIMR Reports, Healy, Hutton and Palepu

[3] See for example Lang and Lundholm (1993), Skinner (1994), Botosan and Plumlee (2002), Bujaki and McConomy (2002), Ho and Taylor (2007), and Francis et al. (2008).

(1999) show that increased disclosure rating is associated with increased stock liquidity, analyst coverage and higher stock returns. Leuz and Verrecchia (2000) document a positive association between the disclosure and higher stock liquidity and a negative relationship between the firm's cost of capital and disclosure. Heflin, Shaw and Wild (2005) and Balakrishnan, Billings, Kelly and Ljungqvist (2013) show that disclosure has a sizeable and beneficial effect on liquidity. Botosan (1997), Botosan and Plumlee (2002), and Barth, Konchitchki and Landsman (2013) document supporting evidence of the negative relationship between transparency and cost of capital. Bushee and Noe (2000), Baumann and Nier (2004), and Kothari et al. (2009) document a negative and significant association of disclosure with stock return volatility. Baumann and Nier (2003) show that the bank's book leverage is decreasing with the level of disclosure, which can be interpreted as lower default probability after controlling for the risk.

Our paper also contributes to a number of self-constructed disclosure indices in the current literature. Being one of the earliest studies, using the annual reports of 122 firms in 1990, Botosan (1997) produces a cross-sectional ranking of disclosure levels based on five categories of voluntary disclosure: background information, summary of historical results, key non-financial statistics, projected information, and management discussion and analysis. Francis et al. (2008) further develop Botosan (1997)'s disclosure index for a sample of 677 firms in 2001 to investigate the relationship among the voluntary disclosure scores, earnings quality, and cost of capital. Lang and Lundholm (2000) examine disclosure activity around seasoned equity offerings with a sample of 41 "offering" and matched "non-offering" firms. The level of disclosure is measured by the score associated to main four groups of announcements identified in the Dow Jones News Retrieval and then Lexis/Nexis news databases: performance, management spin, forward-looking, and others, such as management changes and stock related information. Baumann and Nier (2003)'s disclosure index records whether the particular category is disclosed in BankScope database or not. Finally, in a recent study, Cheung, Jiang and Tan (2010) create a transparency index based on the OECD Principles of Corporate Governance for 100 major Chinese listed companies for the period 2004–2007.

We contribute to this literature by considering several dimensions of voluntary disclosure. Some of these refer to the riskiness of the bank's assets, for example the information of the unencumbered collateral, whereas others focus on the funding vulnerabilities. In contrast to the index of Botosan (1997) and Francis et al. (2008), for instance, our disclosure index mainly focuses on the disclosure of the riskiness, rather than the profitability of an institution. For example, we issue a score of 1 if an institution discloses information on the level or ratio of liquid assets, but we do not consider whether asset turnover or return on assets is disclosed. Similar to the study of Baumann and Nier (2003), we look at the maturity and type of funding. On the other hand, instead of focusing in the risk factors that turn to be compulsory due to current Basel regulations, for example credit risk, our index focuses on more recent risk factors that threaten the financial system, like liquidity or spillover risk. Finally, in addition to the aforementioned disclosure templates, we consider disclosures on the structure of the banking group, to test whether investors place value on information about intra-group exposures.

The paper is organized as follows: next section describes the sample and data sources. Moreover, the construction of the disclosure index, validation of the metric, and details for the estimation of the option implied probability of default are provided. In Section 4.3, we introduce our empirical methodology along with a preliminary analysis. Section 4.4 presents the results and discussions. Finally Section 4.5 concludes.

4.2 Data and Empirical Proxies

In Section 4.2.1, we introduce our sample and data sources. Section 4.2.2 describes in detail the disclosure index, provides evidence supporting its reliability, and descriptive analysis conducted on the index. In Section 4.2.3 we introduce the methodology and empirical implementation of the option implied default probabilities and provide preliminary analysis.

4.2.1 Sample selection and data sources

We use several sources to construct our data set. The information related to the disclosure index is obtained from the bank holding companies' (BHCs) 10–K statements, proxy statements as well as the annual reports from the SEC–Edgar system. We hand-collect data for the largest publicly listed 85 U.S. BHCs, in terms of the total book value of assets as of December 2007, which accounts for the 76% of the total assets of the U.S. banking system as of December 2007.[4] Our sample spans the time period of 1998–2011. Table C.1 in Appendix C lists the sample BHCs with the corresponding identifiers. Moreover, we use the SEC–Edgar system to extract the dates when the 10–K reports of a given BHC is available to public (released at the web page).

In order to estimate the option implied default probability (IPoD) for a BHC for a given date, we use both the CRSP and the OptionMetrics Standardized Options datasets. We get the daily stock prices from CRSP database, whereas all of the information regarding the call options; bid and ask prices, trading volumes, open interests and the corresponding strike prices are from OptionMetrics Standardized Options Dataset. From our sample, we eliminate a day if the trading volume is 0 for all of the options traded. Moreover, we consider only the options with time to expiry

[4]Our sample does not contain of some of the financial institutions that were not a BHC, but became a BHC after 2008, such as Goldman Sachs, Ally Financial, and American Express.

are greater than 6 months. After these filtrations, the sample reduces to 80 BHCs.

We obtain data on daily stock returns, market capitalizations, and as well the bid and ask prices of the equity of each BHC from CRSP. We use this data to estimate the return volatility of a BHC company. We use CRSP value-weighted portfolio returns as market return to estimate the beta of each stock. Market returns and the risk free rates are obtained from Kenneth French's online data library. Finally, FR Y-9C reports from the Federal Reserve Bank of Chicago are used to for the consolidated financial statement data of the BHCs.

4.2.2 Measuring disclosure

Today's executives must communicate complex business structures to a broader range of investors with varying levels of sophistication. Moreover, disclosure demanded by investors is outside the bounds of standard reports required by the regulatory authorities. December 2009 and 2011 Financial Stability Reports of the Bank of England provide possible areas for improved disclosure and summary measures to assess the quantitative information provided by a financial institution. We further work on this assessment and propose a hand-collected index of voluntary disclosures. Our index consists of 14 sub-indices of voluntary disclosures, forming four main categories: liquidity risk profiles of the companies, risk positions of key group affiliates and sub-groups, period averages, highs and lows, and exposures between financial institutions and exposures to the hidden risks. For all of the sub-indices, we assign a score of 1 if a given bank holding company (BHC) includes the corresponding information in its 10–K, annual, or proxy reports for a given year. Table 4.1 presents the sub-indices used in the analysis.

Our first set of variables is aimed to capture whether a given institution discloses information with regard to its liquidity positions. Institutions reliant on short-term or foreign currency based funding sources are argued as being particularly vulnerable

Table 4.1: Disclosure Index–the Template

Table lists the sub–indices of the disclosure index used in the analysis. For all of the 14 sub-indices, a score of 1 is assigned if disclosure is present in the corresponding 10–K, annual or proxy report of a given company. Otherwise, a score of 0 is assigned.

I. Liquidity Risk

Decomposition of funding sources:

L^1: Liabilities breakdown by term structure: minimum should distinguish between short–term and long–term borrowing

L^2: Liabilities breakdown by currency: minimum should decompose into two currencies

Liquidity resilience:

L^3: Liquidity ratios: any kind of quantitative liquidity ratio that helps investors assess ability to withstand funding stress

L^4: Level or ratios of high–quality unencumbered assets

II. Group Structure

G^1: Balance sheet information of main group subsidiaries, branches or affiliates

G^2: Balance sheet information of sectors, sub–units or segments

G^3: Risk ratios of main group subsidiaries, branches or affiliates (e.g. capital, liquidity, loan loss reserves)

G^4: Risk ratios of sectors, sub–units or segments (e.g. capital, liquidity, loan loss reserves).

III. Intra-annual Information

I^1: Detailed average figures of balance sheet items between reporting dates

I^2: Quarterly information for balance sheet items

I^3: Risk ratios on quarterly basis

IV. Spillover Risk

S^1: Credit exposures to banks/ financial institutions

S^2: Detailed breakdown of off-balance sheet items

S^3: Exposures to off-balance sheet entities (SPEs)

to stresses in financial markets (Fahlenbrach, Prilmeier and Stulz, 2011). Hence, we first collect information on the decomposition of funding sources by maturity and currency. We focus whether a given BHC includes its liabilities breakdown by term structure and whether it is decomposed into different non-local currencies. Second, we focused on the liquidity risk profile of firms' balance sheets and on firms' holdings of liquid asset, i.e., liquidity resilience. We specifically search for the liquidity ratios and level or ratio of high-quality unencumbered assets.

Information on group structure is our second main category. Disclosing information on the profitability of key group affiliates is compulsory for the U.S. bank holding companies. However particularly in the case of large and complex financial groups, detailed information on the riskiness and balance sheets of subsidiaries is non-negligible. Hence, we assign a score of 1 each if the bank holding company disclosures balance-sheet and risk ratios of its subsidiaries. In addition, instead of group subsidiaries, we search the same information regarding the main group segments for example, the derivatives desk, card services, and insurance services. A failure of one segment of a large institution not only increases the risk exposures of the individual bank, but also can trigger a broader systemic failure (Ellul and Yerramilli, 2012).

Another key area we include in our index is the publication of intra-annual information. End-of-year figures can be unrepresentative of banks' behavior either due to intra-period volatility in banks' business activity or window dressing at the period end. Hence, reporting period averages and highs/lows to present a window on the risks that institutions run during reporting periods is helpful (Bank of England, 2009). We look for the detailed annual average figures of balance sheet items, quarterly information on balance sheet items, and risk ratios.

Our final main group is information on the network or spillover risk. First, we look for information on the exposure of assets and liabilities of a given BHC to different types of financial institutions. In his annual conference on Bank Structure and Competition in May 2008, Ben Bernanke underlined the banks' substantial exposures to subprime risk and off-balance sheet vehicles. Similarly, the Senior Supervisors Group (2008) mention the importance of enhanced public disclosures to possibly reduce the uncertainty regarding exposures to off-balance items that the market considers to be high-risk following the crises.[5] Hence, we also check whether the detailed breakdown of the off-balance sheet items and maximum loss exposure to special purpose vehicles (or variable interest entities) are present in a given report.

[5] The Senior Supervisors Group is comprised of bank supervisory executives of France, Germany, Switzerland, the United Kingdom, and the United States.

In order to avoid the subjective judgments regarding the relative importance of disclosure on sub-indices, following Tetlock (2007) and Ellul and Yerramilli (2012) we employ the principal component analysis (PCA) to reach the aggregated disclosure score, DSCORE. DSCORE is obtained as the eigenvector in the decomposition of the correlation matrix of the four main groups with the highest eigenvalue. For each bank b and at a given year t, we have:

$$\mathrm{DSCORE}_{b,t} = \mathrm{PCA}\left(\mathrm{LIQ}_{b,t}, \mathrm{GRP\_STR}_{b,t}, \mathrm{INTRA}_{b,t}, \mathrm{SPIL}_{b,t}\right) \tag{4.1}$$

where $\mathrm{LIQ}_{b,t}$ is the disclosure score on liquidity risk calculated as the first principal component of liquidity related sub-indices. The disclosure scores on group structure (GRP_STR), intra-annual information (INTRA), and finally spillover risk (SPIL) are calculated analogously.

PCA picks the information in group structure as the key constitute of the disclosure index (having the highest eigenvalue). The first principal component explains 43% of the overall variation. The four main groups are positively correlated with each other and with the aggregated score, DSCORE. Each of these correlations, except the correlation between the liquidity risk and intra-annual information, are significant at 1% level.

Assessing the validity of the disclosure index

To quantify a disclosure level is not a straightforward task. Investors can capture information not only through the annual reports or 10–K statements but as well through the reports of financial analysts, rating agencies, intra-annual disclosures of the companies or news channel. Moreover, investors may value not only the quantitative disclosures, but as well quality of a given disclosure. Finally, companies may manipulate their balance sheets around reporting dates–a practice commonly known as “window dressing”. Although we acknowledge all above, in order reaching a metric, we focus only on the information provided via publicly available 10–K, annual or proxy reports. Moreover, we check whether a given characteristic of the bank is disclosed, rather than attempt to measure how well it is disclosed. Keeping these possible limitations in mind, we conduct an analysis that may provide some insights into the reliability of our self-constructed index.

First, we expect the disclosure score to be positively correlated with number of financial analysts following the bank holding company during the sample period, which is another possible proxy for availability of corporate information.[6] The Spearman correlation coefficient between the two measures is 0.5172, which is significant at 1% level.

Second, previous literature identifies variables that have a statistically significant association with the level of disclosure. If our disclosure index measures the disclosure level, it should be significantly correlated with these variables. The positive link between the size of the firm and the level of disclosure is documented by many (see for example Botosan and Plumlee (2002), Bujaki and McConomy (2002), Ho and Taylor (2007), and Francis et al. (2008)). Various studies examine the consequences of voluntary disclosure on capital markets. Firms that make extensive voluntary disclosures benefit from improved liquidity for their stock in the capital market and they face reduced cost of capital (see Healy and Palepu (2001) for a literature review). Hence, we examine the relationship between disclosure, the firm size, liquidity, and finally cost of capital.

We measure the firm size as the natural logarithm of the market value of a given bank holding company at the end of each year. We employ three different proxies to measure liquidity: the bid-ask spread (SPR), Amihud (2002) illiquidity measure (AMD), and stock turnover (TRN). For bank b in a given day d, the *weekly* liquidity

[6]The number of financial analysts following a given financial institution is obtained from Bloomberg.

measures are calculated as follows:

$$\mathrm{SPR}_{b,w} = P^{A}_{b,w} - P^{B}_{b,w} \tag{4.2}$$

$$\mathrm{AMD}_{b,w} = \frac{|R_w|}{TrVol_{b,w} \cdot P^{C}_{b,w}} \tag{4.3}$$

$$\mathrm{TRN}_{b,w} = \frac{TrVol_{b,w}}{Q_{b,w}} \tag{4.4}$$

where P^A is the closing ask price, P^B is the closing bid price, $TrVol$ is the weekly trading volume, R is the weekly return, P^C is the weekly closing price, and Q is the number of shares outstanding. Finally, $\mathrm{SPR}_{b,t}$, $\mathrm{AMD}_{b,t}$, and $\mathrm{TRN}_{b,t}$ are calculated as the annual averages of the weekly measures for each year.

Following Sironi (2003), we proxy the cost of capital as the average of the primary market spread to the benchmark security at the time of the subordinated debt issue. We obtain the subordinated debt issue data from Bloomberg and Dealogic databases. $\mathrm{COSTCAP}_{b,t}$ is the average spread on the subordinated debt issued by bank b following the disclosure in year t. Note that the higher the spread on the subordinated debt, the higher the cost of capital.

Table 4.2 presents the Spearman correlation coefficients between the level of disclosure (DSCORE) and firm size (SIZE), liquidity measures, and finally cost of capital (COSTCAP). $\mathrm{DSCORE}_{b,t}$ is the aggregated disclosure score of the bank b at year t, calculated as in (4.1). In line with the literature, results show that the aggregated disclosure score is significantly and negatively correlated with cost of capital, positively correlated with the size of the firm and liquidity. The ρ is highest in absolute term with the size of the bank. Within the liquidity measures, the highest correlation is with the Amihud (2002) illiquidity measure. Higher disclosure is associated with a lower price impact, i.e. higher liquidity, on average. Finally, note that the small sample size on the analysis on cost of capital is due to missing data points on the subordinated debt spreads.

To sum up, we provide supporting evidence of the validity of our disclosure index

Table 4.2: Verification of the Disclosure Index–Correlation Analysis
Table presents the Spearman correlation coefficients between the level of disclosure and firm size, liquidity measures, and cost of capital. $\text{DSCORE}_{b,t}$ is the aggregated disclosure score of the bank b at year t, calculated as in Equation (4.1). $\text{SIZE}_{b,t}$ is the natural logarithm of the market value of a given BHC at the end of year t. The definitions of the three measures of liquidity, spread (SPR), amihud (AMD), and turnover (TRN), are given in (4.2), (4.3), and (4.4), respectively. $\text{COSTCAP}_{b,t}$ is the cost of capital, calculated as the average of the primary market spread to the benchmark security at the time of the subordinated debt issue. The number of observations and the $p-$values corresponding the null hypothesis that disclosure and the given variable is independent are presented as well.

		$\text{SIZE}_{b,t}$	$\text{SPR}_{b,t}$	$\text{AMD}_{b,t}$	$\text{TRN}_{b,t}$	$\text{COSTCAP}_{b,t}$
$\text{DISC}_{b,t}$	Spearman ρ	0.4883	-0.3444	-0.5169	0.2591	-0.2292
	$p-$value	0.0000	0.0000	0.0000	0.0000	0.0129
	Obs.	1043	1037	1045	1045	117

and the aggregated score DSCORE by two analyses: first the number of financial analysts following the given bank and second, the correlation between DSCORE and various variables identified in prior research to be associated with disclosure level.

Descriptive analysis–disclosure index

Table 4.3 presents the descriptive statistics on the sub–indices of the disclosure index. The first column gives the number of the banks that disclose the particular information in all of the years, whereas the second column reports the number of the banks that never discloses the particular risk category throughout the whole period. The majority of the U.S. bank holding companies (BHCs) disclose the average balance sheet items as well as the risk ratios of the main subsidiaries throughout the whole sample period with an average score very close to 1, which is the maximum attainable score for a given category. On the other hand, we see that disclosures on the currency breakdown of funding sources, risk ratios of sectors or sub-units, and risk ratios on quarterly basis are assigned with relatively lower scores on average. Within the liquidity measures, we see that 31 of BHCs disclose a breakdown of funding sources by term (funding sources categorized by maturity) for the whole sample period. Though, disclosures on high quality collateral as well as the liquidity ratios are relatively poor.

Table 4.3: Descriptive Statistics-Disclosure Sub–indices
Table presents the descriptive statistics on the sub–indices of the disclosure index. Panel includes the largest 85 U.S. BHCs spanning the period 1998–2011. The first column gives the number of the banks that disclose the particular information in all of the years, whereas the second column reports the number of the banks that never discloses the particular category throughout the whole period. The last two columns report the sample average and standard deviation of each disclosure category, respectively. For all of the categories, the minimum attainable score is 0, whereas the maximum attainable score is 1.

	disclosing in all periods	disclosing in no periods	average	stdev
L^1: term breakdown	31	0	0.656	0.475
L^2: currency breakdown	1	2	0.040	0.197
L^3: liquidity ratio	15	0	0.339	0.473
L^4: unencumbered assets	2	4	0.134	0.340
G^1: B/S info of subsidiaries	4	2	0.136	0.343
G^2: B/S info of sectors/sub-units	9	1	0.185	0.389
G^3: risk ratios of subsidiaries	70	1	0.942	0.234
G^4: risk ratios of sectors/sub-units	1	1	0.048	0.215
I^1: average B/S figures	81	0	0.983	0.129
I^2: quarterly B/S figures	8	1	0.150	0.357
I^3: risk ratios on quarterly basis	3	1	0.065	0.247
S^1: credit exposure to financial inst.	4	2	0.147	0.354
S^2: off-balance sheet items	11	3	0.263	0.441
S^3: exposure to SPEs	1	2	0.136	0.343

The average scores attained are far lower than 1 for almost all of the sub-indices.

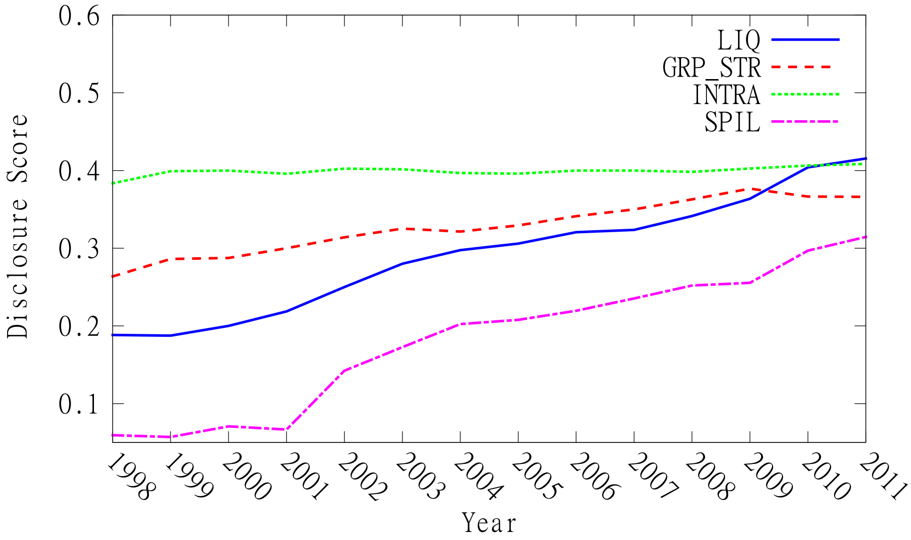
Table 4.4 reports the within year mean and standard deviation of four main categories of the disclosure index, which are obtained by averaging the scores on the corresponding sub-indices. Similarly, Figure 4.1 Panel A and B plots the main categories and composite disclosure index (DSCORE), respectively averaged across the BHCs.

We see an increasing trend for the disclosures throughout the period in study, with a particular improvement in the liquidity risk disclosures as well as the disclosures related to the spillover and hidden risks (information on off balance sheet items or exposure to the special purpose entities). The average highest score, 0.399, is on the disclosures related to intra–annual information, whereas the scores related to the spillover risk are the lowest among the four main categories. Another area where

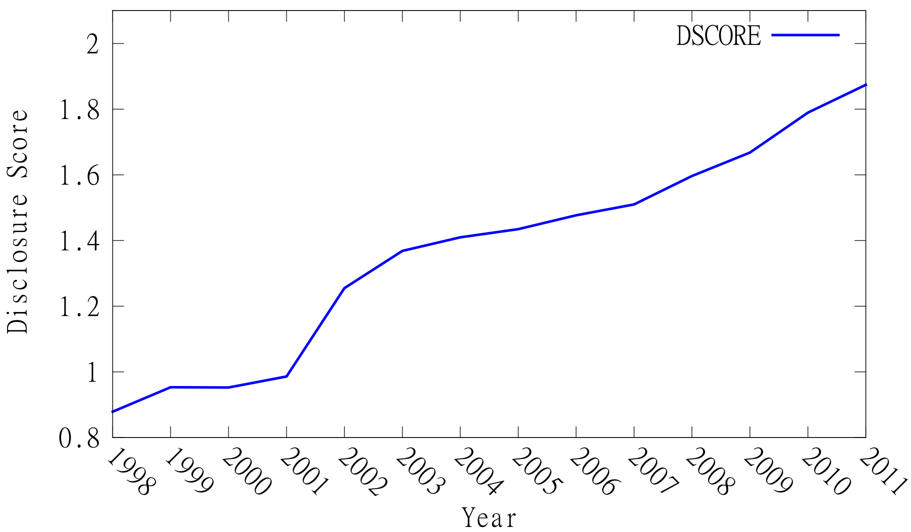
Table 4.4: Descriptive Statistics–Disclosure Index
Table reports the standard deviation of four main categories of the disclosure index along with the aggregated disclosure score throughout the sample period. Panel includes the largest 85 U.S. BHCs spanning the period 1998–2011. Disclosure score on liquidity (LIQ), for example, is obtained as average of the scores on the liquidity-related sub-indices: L^1, L^2, L^3, and L^4 for each bank in a given year. GRP_STR stands for the disclosure on group structure, INTRA for intra-annual information, and finally, SPIL for spillover risk.

	LIQ	GRP_STR	INTRA	SPIL
within year mean				
1998	0.188	0.264	0.384	0.059
1999	0.188	0.286	0.399	0.057
2000	0.200	0.288	0.400	0.071
2001	0.219	0.300	0.396	0.067
2002	0.250	0.314	0.402	0.142
2003	0.280	0.325	0.402	0.173
2004	0.298	0.321	0.397	0.202
2005	0.306	0.329	0.396	0.208
2006	0.321	0.341	0.400	0.220
2007	0.324	0.350	0.400	0.235
2008	0.341	0.363	0.398	0.252
2009	0.364	0.377	0.403	0.255
2010	0.404	0.366	0.406	0.297
2011	0.415	0.366	0.408	0.315
within year standard deviation				
1998	0.185	0.150	0.154	0.151
1999	0.183	0.157	0.172	0.148
2000	0.184	0.144	0.179	0.156
2001	0.192	0.156	0.177	0.154
2002	0.180	0.161	0.187	0.234
2003	0.181	0.169	0.186	0.267
2004	0.192	0.158	0.190	0.269
2005	0.187	0.169	0.189	0.282
2006	0.187	0.172	0.184	0.293
2007	0.188	0.165	0.177	0.299
2008	0.190	0.167	0.177	0.312
2009	0.205	0.170	0.190	0.315
2010	0.234	0.172	0.194	0.317
2011	0.227	0.173	0.197	0.333
uncond. Mean	0.292	0.328	0.399	0.182
uncond. Std	0.206	0.166	0.182	0.273

progress has been slow over the period is the provision of the balance sheet and the risk positions of the main group affiliates and segments. Results as well show that disclosure varies across the BHCs in the sample in a given year. The minimum



(a) Panel A: Disclosure Score–sub-indices



(b) Panel B: Aggregated Disclosure Score

Figure 4.1: Panel A plots the disclosure scores assigned to each of the sub-indices of disclosure score throughout the sample period, averaged across the bank holding companies. Disclosure score on liquidity (LIQ), for example, is obtained as average of the scores on the liquidity-related sub-indices: L^1, L^2, L^3, and L^4 for each bank in a given year. GRP_STR stands for the disclosure on group structure, INTRA for intra-annual information, and finally, SPIL for spillover risk. For all of the categories, the minimum attainable score is 0, whereas the maximum attainable score is 1. Panel B on the other hand presents the cross-sectional average of the aggregate disclosure score (DSCORE) obtained as the first principal component of the four sub-indices.

standard deviation is around 0.144, whereas it increases up to 0.333 in 2011 for the score on spillover risk. We see the highest deviation is on the information provided on spillover risk category.

4.2.3 Measuring the probability of default

The default probability of an institution depends on the unobservable factors such as the value of the company or the firm volatility that needs to be translated from publicly observable data. Several studies use different proxies to estimate the default probability. Nier and Baumann (2006) proxy the default risk by the book leverage, Anginer and Yildizhan (2010) use corporate credit spreads. Using the maximum entropy principle, Jeong (2010) proposes a methodology to estimate the default probability of a firm using binary option prices.

An appealing methodology to estimate the default probability of an institution is proposed by Capuano (2008). The idea is to use the Merton (1974) framework to extract implied probabilities of default from equity option prices. This is quite a flexible framework, the default barrier, i.e., the threshold level under which the firm defaults, is endogenously estimated. Moreover, the employed principle of minimum cross-entropy (Cover and Thomas, 2006) makes it possible to infer the probability distribution of the firm value. Hence, Capuano (2008)'s Option Implied Probability of Default (IPoD) methodology does not require any assumption on neither the recovery rates nor the asset distribution.

One can argue two possible drawbacks of the methodology. First, it estimates the expected level of default in a risk neutral world rather than the actual probability measure. Second, since in case of a default, there is neither stock, nor options trading, we do not have any information regarding the default state. We can only estimate parameters of *entering* to the default state. However, within the alternatives, it is still possibly one of the most flexible methodologies.

The methodology

Merton (1974)'s structural framework suggests that a company goes bankrupt if its value of assets, V, is lower than the face value of its debt, D, in which case the equity holders receive 0. In case of no default, equity holders receive the residual amount. Hence, the value of a stock, S, which is a claim on equity E is $S = E = \max\{V - D, 0\}$. So the payoff of a call option written on a stock can be written as:

$$C_T^K = \max(E_T - K; 0) = \max(V_T - D - K; 0) \tag{4.5}$$

where K is the corresponding strike price. If the default value and the distribution of assets are known, then one can estimate the probability of default for an arbitrary D as follows:

$$PoD(D) = \int_0^D f(V_T)dV_T \tag{4.6}$$

where V is the value of the asset and $f(V)$ is the corresponding probability density function.

Hence, to calculate the probability of default, one needs to estimate the default barrier as well as infer the $f(V_T)$ from the available option contracts. To do so, Capuano (2008) employs the concept of cross entropy. This is a measure of relative distance between the prior and posterior density function.[7] The problem to be solved turns to be:

$$\min_D \left\{ \min_{f(V_T)} \int_0^\infty f(V_T) \log \frac{f(V_T)}{f_0(V_T)} dV_T \right\} \tag{4.7}$$

where $f_0(V)$ is the prior probability density function of the value of asset V and $f(V_T) \log \frac{f(V_T)}{f_0(V_T)}$ is the cross-entropy (or relative entropy) between $f(V)$ and $f_0(V)$. The minimization problem (4.7) is subject to the following constraints:

1. **Option pricing constraint**—The current price of an option is the discounted future cash flows under risk neutral measure:

$$C_0^{K_i} = e^{-rT} \int_{V_T = D + K_i}^{\infty} (V_T - D - K_i) f(V_T) dV_T \tag{4.8}$$

[7] See Capuano (2008) or Vilsmeier (2011) for details of the estimation and for further discussions.

where K_i is the strike price of option i. Note that the current stock price S_0 is included as an option with $K = 0$.

2. **Additivity constraint**—The probability density function must sum up to 1:

$$1 = \int_{V_T=0}^{\infty} f(V_T) dV_T \tag{4.9}$$

Hence, the Lagrangian adds up to:

$$\begin{aligned} L = &\int_0^{\infty} f(V_T) \log \frac{f(V_T)}{f_0(V_T)} dV_T + \lambda_0 \left[1 - \int_{V_T=0}^{\infty} f(V_T) dV_T \right] \\ &+ \sum_{i=1}^{N} \lambda_i \left[C_0^{K_i} - e^{-rT} \int_{V_T=D+K_i}^{\infty} (V_T - D - K_i) f(V_T) dV_T \right] \end{aligned} \tag{4.10}$$

where N is the number of options available, $\lambda_0, ..., \lambda_N$ are the corresponding Lagrange multipliers.

The first step is to determine the optimal values of λs through the first order conditions (Cover and Thomas, 2006). For a given value of D:

$$\begin{aligned} \frac{\partial L(f(V, \lambda), \lambda)}{\partial \lambda} &= e^{-rT} \int_{V_T=0}^{\infty} \mathbb{1}_{V_T > D+K_i} (V_T - D - K_i) f(V_T) dV_T - C_0^{K_i} \\ &= 0, \qquad i = 1, .., N. \end{aligned}$$

The above equation should be solved numerically via a multivariate algorithm, such as the Newton–Paphson algorithm. However, as noted by Vilsmeier (2011), searching for the roots of the system is unfeasible in many applications mainly due to near-singularities of the Jacobian matrix resulting from the first Taylor approximation, unless the initial guess of λs are very “accurate”.[8]

Vilsmeier (2011) suggests technical modifications to the Capuano (2008)'s framework. Following Alhassid, Agmon and Levine (1978), he uses a robust and computationally efficient algorithm to calculate the optimal set of λs. This paper follows Vilsmeier (2011)'s methodology to estimate the optimal λs.[9]

[8] The majority of our optimization trails failed due to the non-singularity of the Jacobian matrix.

[9] We sincerely thank Johannes Vilsmeier for sharing his codes to estimate the probability of default.

Once the optimal λs are obtained, we can get $f^*(V_T, D)$. Given $f^*(V_T, D)$, the default barrier D^* is calculated trough another numerical optimization of:

$$\lim_{\Delta \to 0} \frac{L\left(f^*(V_T, D + \Delta)\right) - L\left(f^*(V_T, D)\right)}{D + \Delta} = 0. \tag{4.11}$$

Finally the IPoD is estimated through (4.6) once we have $f^*(V_T)$ and D^*.

Empirical implementation

At least two option contracts written on the same stock with the same expiry date are needed to solve the problem. The first one is used to shape the density function $f^*(V, D)$, whereas the second one is needed to estimate the threshold level D^*. We apply the framework only to the call options since put options relate by the put-call parity. Moreover, we consider only the options with time to expiry more than 6 months.

Option prices are estimated as the average of the best bid and ask prices. Finally, in order to capture the liquidity differences, we weight the option contracts by using the open interest of each option.[10] That is, the weight for option i trading in day τ expires in day T is:

$$w_i(\tau, T) = \frac{\text{OpenInt}_i(\tau, T)}{\sum_{i=1}^{N} \text{OpenInt}_i(\tau, T)} \tag{4.12}$$

where N is the total number of options traded at date τ and expires at T. Table C.2 in Appendix C presents the information we use to estimate the IPoD of J.P. Morgan (JPM) for 12^{th} of December 2010 as an example.

Descriptive analysis–IPoD

Table 4.5 reports the descriptive statistics of the IPoD estimates. For each bank holding company (BHC), each year, we first calculate the average of the IPoD esti-

[10] Capuano (2008) uses the trading volume as weight, whereas Vilsmeier (2011) uses the open interests. We estimate the IPoD using both trading volume and open interests and results are qualitatively similar. However, the IPoD estimated through open interests are more stable. Hence, we report only the results, where open interest is used to weight the liquidity of an option.

Table 4.5: Descriptive Statistics–IPoD, annual averages
Table reports the mean, median, and the 25th and the 75th percentile of the annual averages of the estimated option implied probability of default (IPoD) throughout the sample period, averaged across BHCs. The annual figure is obtained as the average value for a given calendar year. The sample includes the largest 80 U.S. bank holding companies for a period of 1998 to 2011. All of the figures are scaled by 100.

year	mean	p25	median	p75	std. dev.
1998	0.815	0.239	0.472	1.180	0.951
1999	0.790	0.306	0.496	1.040	0.721
2000	1.537	0.564	1.015	1.540	2.510
2001	0.605	0.227	0.458	0.813	0.648
2002	0.626	0.139	0.283	0.620	1.062
2003	0.269	0.062	0.100	0.248	0.516
2004	0.138	0.014	0.031	0.257	0.189
2005	0.148	0.015	0.049	0.164	0.255
2006	0.148	0.015	0.035	0.185	0.227
2007	0.741	0.164	0.389	0.921	1.137
2008	4.653	1.783	3.439	7.110	3.833
2009	10.296	3.078	5.459	11.904	12.575
2010	4.196	0.659	1.818	4.032	7.319
2011	2.812	0.706	1.435	3.233	3.904

mates corresponding to the trading days within that year. The mean, standard deviation, 25th, 50th and 75th percentiles of the estimated default probabilities (scaled by 100) averaged across the BHCs are presented. Results reveal relatively low market based default probabilities for the 2003–2006 period, where the average expected default is only 0.17%. On the other hand, we see a significant increase in 2007 with a peak in 2009. From 2006 to 2009, the average value increased from 0.15% to over 10%. In order to proxy the overall uncertainty in the stock market, we get the Chicago Board Options Exchange Market Volatility Index (VIX) daily data from the Exchange's website. As expected, the Spearman correlation coefficient between the VIX index and IPoD is 0.2566 and significant with a $p-$value of 0.0000. This suggests that market's assessment on the riskiness of a stock increases as the market–wide uncertainty increases.

Similarly, Figure 4.2 plots the estimated values throughout the sample period for the whole sample, for defaulted BHCs only, and finally for non–defaulted ones. The

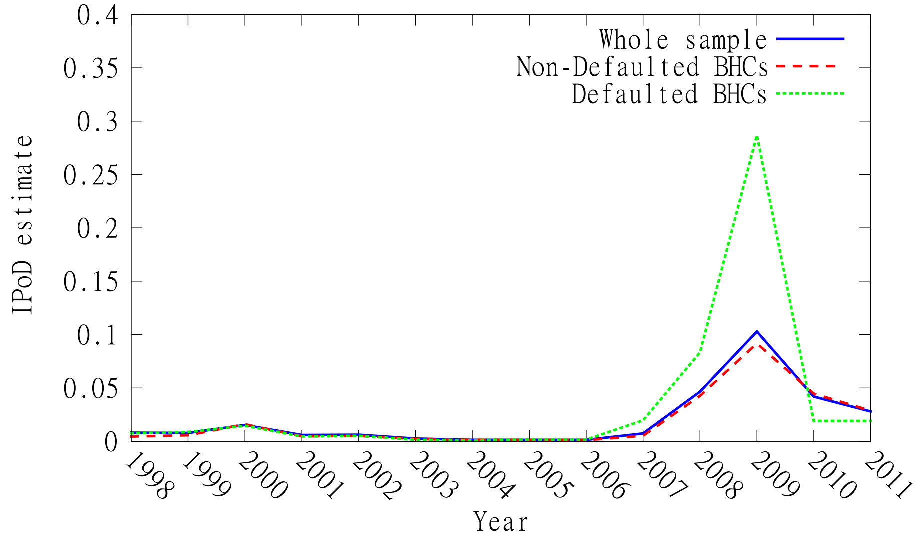


Figure 4.2: Figure plots the market implied probability of default (IPoD) estimates throughout the sample period, averaged across the bank holding companies (BHCs). A BHC is identified as defaulted if it is delisted in a given year.

average IPoD estimates are almost always higher for the defaulted BHCs compared to non–defaulted ones. The difference is significant particularly for the 2006–2010 period. Indeed, sample mean comparison test fails to reject the null hypothesis that the mean value of IPoD for the defaulted companies are significantly higher than the non–defaulted companies with a $p-$value of 0.1465. Both the augmented Dickey-Fuller and Phillips-Perron tests reject the unit–root in the IPoD with $p-$value of 0.0000.

4.3 Empirical Methodology and Preliminary Analysis

This section introduces the empirical methodology we employ to investigate the impact of total disclosure level on the market implied default probability, followed by a preliminary analysis conducted on key variables.

4.3.1 Empirical Methodology

We test our main hypothesis by employing the following panel regression:

$$\text{IPoD}_{b,t+1} = \gamma_0 + \gamma_1 \text{DSCORE}_{b,t} + \kappa * X_{b,t} + \text{Year FE} + \text{BHC FE} + \varepsilon_{b,t}. \quad (4.13)$$

We estimate the above regression on a panel that has one observation for each bank holding company–year combination, where subscript b denotes the bank holding company (BHC) and t denotes the year. Year and BHC fixed effects are included in the regression to capture for any time-invariant unobserved BHC characteristics.

First, we measure our dependent variable, $\text{IPoD}_{b,t+1}$, as the natural logarithm of the average implied probability of default for bank b between two annual 10–K statements disclosure dates.[11] For example, if a bank's 2008 10–K report became public on the SEC–Edgar database on the 26$^{\text{th}}$ of February 2009, $\text{IPoD}_{b,t+1}$ is calculated as the natural logarithm of the average IPoD estimates from 27$^{\text{th}}$ of February 2009 until 16$^{\text{th}}$ of February 2010, the disclosure date of 10–K statement for the year 2009. Thus we allow the revealed information has an all-year-long effect. Second, instead of annual averages, we use the log–transformed three-months averages of the implied default probability estimates ($\text{IPoD}_{b,t+3M}$) following the announcement of a 10–K statement as the dependent variable. Since filling quarterly 10–Q reports to SEC is compulsory for the U.S. bank holding companies, it is likely that investors update their information set with the releases of 10–Q statements. In other words, the informativeness of an annual report decreases with the public 10–Q statements, which is roughly after three months of the release of 10–K report.

The main independent variable, $\text{DSCORE}_{b,t}$, is the aggregated disclosure score of the bank b at year t, calculated as in Equation (4.1). In line with our main hypothesis, the coefficient of interest γ_1 is expected to be negative; investors assess high disclosed

[11] Given the high skewness of the distribution, we use the logarithm of the IPoD estimates instead of levels in our analysis (see for instance Laeven and Levine (2009)). For brevity, we use the "IPoD" in referring to the natural logarithm of IPoD in the rest of the paper.

banks as less likely to default. We include the size of the given bank holding company (SIZE) as a control variable because extant research presents a positive correlation between the disclosure score, the size of the institution, and the bank's risk taking behavior (e.g., Barrell, Davis, Fic and Karim (2010), Hermalin and Weisbach (2012)). It is measured as the natural logarithm of the year-end total market capitalization.

We then control for the volatility of the firm value. The volatility of the firm value is unobservable due to the lack of data on the market value of a firm's debt. However, under Merton (1974)'s assumptions, the volatility of the firm value can be estimated using the volatility of the equity, which is observable. Following Lewis (2011) we proxy the annual volatility as follows:[12]

$$\mathrm{VOLA}_{b,t} = \sqrt{\frac{1}{W_{b,t}} \sum_{w=1}^{W_{b,t}} \left(\frac{P^H_{b,w} - P^L_{b,w}}{P^C_{b,w}} \right)^2} \tag{4.14}$$

where $W_{b,t}$ is the number of weeks available in year t for stock b, $P^H_{b,w}$ and $P^L_{b,w}$ are the average weekly highest and the lowest prices for equity of bank b in week w, respectively. It is represented as a percentage of closing price P^C to adjust for stocks trading at different prices. We expect volatility to be positively associated with IPoD.

$\mathrm{BETA}_{b,t}$ is the estimated beta of bank b for year t and it is included in the analysis to account for the systematic risk. It is calculated for each bank and each year from regressions of bank weekly equity returns on the weekly returns of CRSP value-weighted index.

Finally, we include other bank holding company financial characteristics. Baumann and Nier (2003) proxy the (inverse) default probability as individual banks' capital buffers and document a positive relationship between disclosure and the capital buffer. In a cross-country analysis, Beltratti and Stulz (2011) show that large banks with more Tier 1 capital perform significantly better during the crises. Following the definition of Baumann and Nier (2003) and Nier and Baumann (2006),

[12] As a robustness, we measure volatility as the standard deviation of weekly returns for a given period. The results are presented in Section 4.4.5.

we define capital buffers, CAPBUF, a bank's equity capital divided by its total liabilities. We expect banks with higher capital buffers to default less. To capture other accounting risks, we consider non-performing loans, return on equity ratio, and finally the level of deposits in log terms.[13] For all of the variables, the value reported in the FR Y–9C reports of the Federal Reserve Bank of Chicago for the end of year t is used. The non-performing loans ratio, NPL, is calculated as the ratio of the sum of loans past due 90 days or more and non-accrual loans to total assets. The return on equity ratio, ROE, is the ratio of the income before extraordinary items to total book equity. Finally, DEPO is the natural logarithm of the total deposits. We expect the risk taking behavior of the BHC to be positively associated with the IPoD of the firm. The definitions of the variables are presented in detail in Appendix C.

4.3.2 Preliminary analysis

This section presents a descriptive analysis of the key variables used throughout the analysis. Panels A and B of Table 4.6 report the summary statistics and the pair-wise correlations, respectively. In Panel C, we present a univariate comparison analysis for bank holding companies with high versus low level of disclosure. The superscripts * and ** are used to denote the 10% and 5% levels of statistical significances, respectively.

Panel A presents the mean, median, minimum, maximum and standard deviation of the key variables. Disclosure score has an annual mean of 1.54. Its value ranges from 0.06 to 7.69 with standard deviation of 1.38. The minimum annual IPoD value is -18.47 (corresponding to a 0 probability of default) and increases up to -0.76 (equivalently 47% of implied default probability). The mean market value of common equity (SIZE) is \$2.65 billion and the median is \$1.80 billion. The highest deviation is in

[13] In unreported results, we control for other accounting risk variables such as loan to asset ratio, asset growth, and loan growth. We conclude that our results are robust to the inclusion of various different accounting variables.

the return on equity. The value ranges from almost -6.2 to 0.3 within the sample. Similarly, the ratio of the bank's equity to its liabilities, CAPBUF has a significant variation across the bank holding companies.

Table 4.6: Descriptive Analysis–Key Variables

Table presents the descriptive analysis for the key variables used throughout the paper. Panel A and B report the summary statistics and the pair-wise correlations. In Panel C, we present a univariate comparison analysis for banks with high versus low level of disclosure. A bank is identified as high-disclosed (low-disclosed) if its disclosure score is higher (lower) than the median score in a given year. The superscript * (**) denotes the 10% (5%) level statistical significance. $\text{IPoD}_{b,t+1}$ is the natural logarithm of the average implied probability of default estimates, calculated between two annual report disclosure dates, whereas $\text{IPoD}_{b,t+3M}$ is the average estimates over three months following the 10–K report disclosure. DSCORE is the aggregated disclosure score defined in Equation (4.1). All of the variables are introduced in Section 4.3.1 and as well defined in Appendix C. The sample contains the largest 80 U.S. bank holding companies for a period of 1998 to 2011.

PANEL A: Summary statistics

	$\text{IPOD}_{b,t+1}$	$\text{IPOD}_{b,t+3M}$	$\text{DSCORE}_{b,t}$	$\text{SIZE}_{b,t}$	$\text{VOLA}_{b,t}$	$\text{BETA}_{b,t}$	$\text{CAPBUF}_{b,t}$	$\text{NPL}_{b,t}$	$\text{ROE}_{b,t}$	$\text{DEPO}_{b,t}$
mean	-5.256	-5.825	1.538	14.789	0.036	1.021	0.130	0.011	0.079	16.259
median	-5.032	-5.671	1.092	14.404	0.029	0.967	0.102	0.006	0.119	15.884
min	-18.470	-31.135	0.058	10.263	0.011	-0.364	0.010	0.000	-6.215	12.612
max	-0.758	-0.340	7.686	19.428	0.258	3.593	3.766	0.243	0.266	20.844
std. dev.	2.107	2.563	1.380	1.621	0.025	0.558	0.253	0.016	0.305	1.511
Obs.	671	638	1055	1060	1072	1055	1032	958	1032	1032

PANEL B: Pair-wise correlations among key variables

	$\text{IPOD}_{b,t+1}$	$\text{IPOD}_{b,t+3M}$	$\text{DSCORE}_{b,t}$	$\text{SIZE}_{b,t}$	$\text{VOLA}_{b,t}$	$\text{BETA}_{b,t}$	$\text{CAPBUF}_{b,t}$	$\text{NPL}_{b,t}$	$\text{ROE}_{b,t}$	$\text{DEPO}_{b,t}$
$\text{IPOD}_{b,t+3M}$	0.848**	1								
$\text{DSCORE}_{b,t}$	-0.050	-0.076*	1							
$\text{SIZE}_{b,t}$	-0.267**	-0.286**	0.547**	1						
$\text{VOLA}_{b,t}$	0.587**	0.693**	0.021	-0.247**	1					
$\text{BETA}_{b,t}$	0.471**	0.518**	0.197**	0.084**	0.611**	1				
$\text{CAPBUF}_{b,t}$	0.012	0.038	-0.044	0.121**	-0.009	0.086**	1			
$\text{NPL}_{b,t}$	0.403**	0.387**	0.087**	-0.186**	0.636**	0.455**	-0.062*	1		
$\text{ROE}_{b,t}$	-0.306**	-0.338**	0.026	0.265**	-0.526**	-0.337**	0.042	-0.508**	1	
$\text{DEPO}_{b,t}$	-0.086**	-0.111**	0.630**	0.868**	0.019	0.255**	-0.186**	0.122**	0.033	1

PANEL C: Comparison of high and low disclosed banks

	$\text{IPOD}_{b,t+1}$	$\text{IPOD}_{b,t+3M}$	$\text{SIZE}_{b,t}$	$\text{VOLA}_{b,t}$	$\text{BETA}_{b,t}$	$\text{CAPBUF}_{b,t}$	$\text{NPL}_{b,t}$	$\text{ROE}_{b,t}$	$\text{DEPO}_{b,t}$
High-disclosed	-5.205	-6.049	15.471	0.034	1.049	0.106	0.010	0.102	16.985
Low-disclosed	-4.791	-5.446	14.052	0.037	0.986	0.136	0.011	0.054	15.562
Difference	-0.414	-0.603	1.419	-0.003	0.062	-0.029	-0.001	0.048	1.423
$p-$value	0.034	0.006	0.000	0.042	0.059	0.031	0.436	0.015	0.000

The figures of Panel B reveal that the correlation between the annual and three month IPoD is 0.85 and significant at a 5% level. The correlation between the level of disclosure and implied default probability is negative as expected, however disclosure is significantly correlated with the three-months-ahead IPoD only at a 10% level with a $p-$value of 0.0576. Size is negatively related to IPoD and positively related to disclosure, both being highly significant. In other words, investors assess bigger banks as less likely to default and bigger banks disclosure more. We as well find a strong statistical relationship between the size of a BHC and BHC characteristics: bigger banks hold higher capital buffers, have better accounting performance, lower ratio of non-performing loans, and have higher level of deposits. Both volatility and beta are significantly and positively correlated with IPoD, suggesting that higher market risk increases the investors' risk-neutral expectations of default probability. Within the accounting variables, we see that the level of deposits and the ratio of non-performing assets are significantly related to both disclosure and IPoD irrespective of the chosen period.

Finally, we identify a bank as high-disclosed one if its disclosure score is higher than the median disclosure score in a given year. Similarly, low-disclosed BHCs are the ones whose DSCORE values are less than the median score of the corresponding year. In order to understand the differences in characteristics between high-disclosed and low-disclosed bank holding companies, we then employ a univariate mean comparison test between these two samples. Panel C reports the mean values, differences, and $p-$values corresponding to a null hypothesis that both samples have the same mean for a given characteristics. We see that BHCs disclosing more information related to their risk profiles have significantly lower average implied default probabilities in the following three months and year, which is consistent with our main hypothesis. Not surprisingly, BHCs with higher levels of disclosure are larger in size. Larger BHCs are more likely to be complex in structure, involved in riskier non-banking activities and have higher incentives to mitigate the informational frictions by disclosing more information. We as well see that high-disclosed BHCs have higher levels of deposits and benefits from higher operating performance compared to their low-disclosed pairs. Obviously, the causality can go both ways: higher disclosure can lead an increase on those variables, or higher values of the aforementioned accounting characteristics can encourage the management to disclose more information.

4.4 Results

In Section 4.4.1 below, we test our main hypothesis that the previous level of disclosure is associated with lower levels of current market implied default probability by employing the baseline panel regression (4.13). Then we allow IPoD and DSCORE being dynamically endogenous and replicate our panel regression by employing the Arellano and Bond (1991) dynamic panel generalized method of moments (GMM) estimator. The methodology, discussion, and results are presented in Section 4.4.2. A natural question arises whether there is an association between disclosure and other enterprise risks. In Section 4.4.3 we test this question. Section 4.4.4 examines the determinants of disclosure in a cross–sectional setting. Finally in Section 4.4.5, we present further robustness tests.

The panel includes the largest 80 U.S. BHCs and spans the time period 1998–2011. In all of the specifications year and bank fixed effects are included. For the sake of brevity, we do not report the estimated coefficients of fixed effects. To improve the ease of interpretation of the estimated coefficients, all of the explanatory variables are standardized to have a zero mean–unit variance. The t-statistics are reported in parentheses.

4.4.1 Relationship between disclosure and implied default probability

Estimated coefficients of (4.13) are presented in Table 4.7. Columns I through V present the results where the dependent variable is $\text{IPoD}_{b,t+1}$; the natural logarithm of the average IPoD estimates between two 10–K reports dates throughout the year $t+1$. On the other hand, in columns VI through X, we report the results where the dependent variable is $\text{IPoD}_{b,t+3M}$; the natural logarithm of the average IPoD for bank b between the annual 10–K and quarterly 10–Q reports disclosure dates.

The results confirm our hypothesis. The coefficient on DSCORE is negative and statistically significant at a 5% level irrespective of the chosen period. This suggests that BHCs with higher disclosure compared to their peers in this year, ceteris paribus, have lower market implied default probabilities in the following period. The IPoD–disclosure relationship is stronger for three-months ahead; both the statistical and economic significance of the estimated coefficient is higher for the specifications presented in columns VI to X compared to the ones presented in I to V. In terms of economic significance, one standard deviation increase in the level of disclosure is associated with a decrease of a 19% and 27% decrease in IPoD in the following year and three months, respectively.

In columns II and VII we include the size of the bank holding company. Results reveal that the bigger banks are assessed as less likely to default, which could be a result of implicit too-big-to-fail guarantees. The third and eighth columns present the regression of the current level of IPoD on previous year's disclosure, size, and realized volatility. As expected, higher levels of stock volatility is significantly and positively associated with higher levels of market implied default probability. In columns IV and IX, we control for the systematic risk, proxied by BETA. The estimated coefficient is positive and significant. In other words, stocks with higher beta are assessed as more likely to default by investors.

Table 4.7: Relationship Between Disclosure Score and IPoD

Table provides the results of panel regressions that examine the impact of the level of disclosure on a bank holding company's market implied probability of default (IPoD). The panel includes the largest 80 U.S. BHCs and spans the time period 1998–2011. In columns I through V, the dependent variable is $\text{IPoD}_{b,t+1}$, is the natural logarithm of the average IPoD estimates for bank b between two annual 10–K report disclosure dates. On the other hand, columns VI through X report the results where the dependent variable $\text{IPoD}_{b,t+3M}$, is the natural logarithm of the average IPoD for bank b between the annual 10–K and quarterly 10–Q report disclosure dates. DSCORE is the aggregated disclosure score and VOLA is the realized volatility calculated from the bid and ask prices. BETA is estimated from regressions of bank weekly equity returns on the weekly returns of CRSP value weighted index. SIZE is measured as the natural logarithm of the yearend total market capitalization. CAPBUF is the ratio of bank's equity capital to total liabilities, NPL is the non-performing loans ratio, ROE is the return on equity, and finally DEPO is the natural logarithm of the total deposits. In all of the specifications year and bank fixed effects are included. The explanatory variables are standardized to have a zero mean–unit variance. The $t-$statistics are reported in parentheses. For each of the specification, the sample size and the adjusted R^2s are also reported.

	$\text{IPoD}_{b,t+1} = \gamma_0 + \gamma_1\text{DSCORE}_{b,t} + \text{controls}_{b,t} + \varepsilon_{b,t}$					$\text{IPoD}_{b,t+3M} = \gamma_0 + \gamma_1\text{DSCORE}_{b,t} + \text{controls}_{b,t} + \varepsilon_{b,t}$				
	I	II	III	IV	V	VI	VII	VIII	IX	X
$\text{DSCORE}_{b,t}$	-0.188	-0.232	-0.214	-0.177	-0.183	-0.267	-0.329	-0.303	-0.247	-0.298
	(-2.17)	(-2.74)	(-2.60)	(-2.18)	(-2.11)	(-2.33)	(-2.97)	(-2.80)	(-2.34)	(-2.59)
$\text{SIZE}_{b,t}$		-0.959	-0.406	-0.367	-0.593		-1.412	-0.705	-0.740	-1.007
		(-5.53)	(-2.15)	(-1.97)	(-2.38)		(-6.26)	(-2.74)	(-2.96)	(-3.05)
$\text{VOLA}_{b,t}$			0.671	0.429	0.215			0.735	0.316	0.143
			(6.38)	(3.59)	(1.61)			(5.32)	(2.07)	(0.82)
$\text{BETA}_{b,t}$				0.293	0.284				0.524	0.571
				(4.16)	(3.87)				(5.81)	(5.76)
$\text{CAPBUF}_{b,t}$					-0.195					-0.229
					(-1.76)					(-1.64)
$\text{NPL}_{b,t}$					0.381					0.143
					(3.63)					(1.52)
$\text{ROE}_{b,t}$					0.125					0.108
					(1.72)					(1.05)
$\text{DEPO}_{b,t}$					0.62					0.543
					(1.88)					(1.23)
cons	-4.956	-4.789	-4.874	-5.001	-4.438	-5.463	-5.128	-5.273	-5.496	-4.256
	(-30.90)	(-30.03)	(-31.49)	(-32.10)	(-26.63)	(-25.84)	(-24.28)	(-25.37)	(-26.76)	(-19.33)
Obs.	668	667	667	666	613	635	635	635	635	584
adjR^2	0.679	0.695	0.714	0.721	0.740	0.604	0.630	0.648	0.668	0.681
Year & Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Finally, we include the bank holding company accounting characteristics in the analysis. In general, the association between the accounting variables and annual IPoD is stronger than the three-months IPoD. The estimated sign of CAPBUF is negative and significant at a 10% level; banks with higher capital buffers are perceived as less likely to default. The ratio of non-performing loans, deposits, and return on equity are positively associated with next period's IPoD. The signs are as expected; higher non-performing loans indicate higher expected losses and associated with higher default risk. Higher profitability may signal greater efficiency and lower default risk. However, a higher value might also indicate higher risk-taking activities. Our results suggest that above–sample–average ROE is assessed as increased risk. Similarly, higher deposits could be a signal of maturity gap since deposits are more likely to have short term maturity.

We conclude that, even controlled by the market and the accounting risk, an increase in the disclosure score reduces the market based implied probability of default. When the level of disclosure increases, this may be perceived as an increased transparency by the investors which in turn affect the agents perceptions regarding the default risk of the given bank holding company. The result is robust to all of the specifications considered. We would like to underline that our regressions include bank holding company fixed effects, so the documented association between disclosure and IPoD cannot be explained by differences in management quality across BHCs. Besides the level of disclosure, size of the company, stock beta, and ratio of non-performing loans are significantly associated with the next year's market default probabilities and robust to different specifications. The riskiness of the company is associated with higher default probabilities, whereas bigger banks are expected to default less.

4.4.2 Adjusting for dynamic endogeneity

The underlying assumption under our regression model noted in (4.13) is that level of disclosure, DSCORE, is exogenous after controlling for the market risk, bank holding company characteristics, year, and bank fixed effects. However, causality may run in both directions–from management's decision on the level of disclosure to default probability and vice versa. A bank holding company's past level of market implied default risk can affect both the current level of default risk and the decision on the level of disclosure. In other words, IPoD and DSCORE can be dynamically endogenous. For instance, a bank holding company that is exposed to higher risk may choose to disclose more information to reduce the uncertainty and change investors' assessment on its risk. Otherwise, some unobserved time-invariant bank characteristics may jointly affect the implied default probability and the level of disclosure, even though controlled by year and bank fixed effects. In these cases, the regressors will be correlated with the error term, which produces biased coefficients.

One solution is to run fixed-effects instrumental variables estimation (two-stage least squares). However, finding strong instruments is a challenge and with weak instruments, the fixed-effects IV estimators are likely to be biased as well. Therefore, to adjust for the dynamic endogeneity, we employ the Arellano and Bond (1991) dynamic panel GMM estimator, which enables us to use the lags of the endogenous variables to provide instruments for identifying the relationship between the level of disclosure and IPoD.[14] Recall our main panel regression model:

$$\mathrm{IPoD}_{b,t+1} = \gamma_0 + \gamma_1 \mathrm{DSCORE}_{b,t} + \kappa X_{b,t} + \varepsilon_{b,t} \tag{4.15}$$

where X stands for the control variables. The error term can be composed of unob-

[14] Another advantage of the Arellano and Bond (1991) GMM is that it is designed for situations with small T, large N panels, as in the case of our sample.

served bank-specific effects and observation specific errors:

$$\varepsilon_{b,t} = \eta_b + e_{b,t}$$

The Arellano and Bond (1991) GMM first uses first-differences to transform (4.15) into:

$$\Delta\text{IPoD}_{b,t+1} = \gamma_1 \Delta\text{DSCORE}_{b,t} + \kappa \Delta X_{b,t} + \Delta\varepsilon_{b,t}$$

Note that the first-differencing eliminates the unobserved η_b:

$$\Delta\varepsilon_{b,t} = (\eta_b - \eta_b) + (e_{b,t} - e_{b,t-1}) = \Delta e_{b,t}$$

The final stage in the estimation is to estimate the first-differenced equation via GMM by using lagged values of the endogenous variables. We include three lags of IPoD as regressors and estimate the following dynamic panel regression:

$$\begin{aligned}\text{IPoD}_{b,t+1} = \gamma_0 + \gamma_1\text{DSCORE}_{b,t} + \gamma_2\text{IPoD}_{b,t} + \gamma_3\text{IPoD}_{b,t-1} \\ + \gamma_3\text{IPoD}_{b,t-2} + \kappa X_{b,t} + \eta_b + e_{b,t}\end{aligned} \tag{4.16}$$

Table 4.8 report the results where we use both the $\text{IPoD}_{b,t+1}$; the natural logarithm of the average IPoD for bank b between the two annual report disclosure dates and $\text{IPoD}_{b,t+3M}$; is the natural logarithm of the average IPoD for bank b between the annual 10–K and quarterly 10–Q report disclosure dates as dependent variables. The full set of control variables introduced in Section 4.4.1 are included in the analysis, but for the sake of brevity we do not report the estimated coefficients.

We try two different specifications. The first and the third columns present the estimated coefficients and robust t-statistics for the specification where we treat only IPOD and DSCORE as endogenous. The first two lags of the endogenous variables and controls are used as instruments. The coefficient on DSCORE is negative and significant. However, a crucial assumption for the validity of the GMM estimates is that the instruments are exogenous. The Hansen test statistic for over-identifying

Table 4.8: Dynamic Panel GMM Estimaton

Table provides the results of the GMM panel regression in Equation (4.16) that examine the impact of the level of disclosure on a bank holding company's (BHC) market implied probability of default (IPoD). The panel includes the largest 80 U.S. BHCs and spans the time period 1998–2011. In columns I and II, the dependent variable $\text{IPoD}_{b,t+1}$, is the natural logarithm of the average IPoD for bank b between the two annual report disclosure dates. In columns III and IV, we report the results where the dependent variable $\text{IPoD}_{b,t+3M}$, is the natural logarithm of the average IPoD for bank b between the annual 10–K and quarterly 10–Q report disclosure dates. DSCORE is the aggregated disclosure score. All of the control variables introduced in Table 4.7; size, beta of the company, realized variance, capital buffers, non-performing loans, return on equity, and finally level of deposits, are included in all of the specifications. For the sake of brevity, the estimated coefficients are not reported. Column I and III present the results in which the IPoD and DSCORE is allowed to be dynamically endogenous, whereas all of the other variables are assumed to be exogenous. In columns II and IV on the other hand, in addition to disclosure and IPoD, we allow the accounting risk variables (CAPBUF, NPL, ROE, and DEPO) to be endogenous as well. The $t-$statistics are reported in parentheses. For each of the specification, the sample size, the Hansen test statistics for over-identifying restrictions with the corresponding $p-$values are also reported.

	$Y_{b,t} = \text{IPoD}_{b,t+1}$		$Y_{b,t} = \text{IPoD}_{b,t+3M}$	
	I	II	III	IV
$\text{Y}_{b,t-1}$	1.222	0.971	-0.097	-0.109
	(12.59)	(10.13)	(-0.88)	(-1.00)
$\text{Y}_{b,t-2}$	-0.162	-0.235	-0.419	-0.281
	(-1.29)	(-2.62)	(-3.20)	(-2.54)
$\text{Y}_{b,t-3}$	-0.605	-0.555	0.398	0.237
	(-5.25)	(-4.80)	(4.10)	(2.68)
$\text{DSCORE}_{b,t}$	-1.603	-0.792	-2.296	-1.428
	(-2.24)	(-2.22)	(-3.46)	(-2.92)
Obs.	337	337	317	317
Hansen χ^2	30.90	45.16	36.24	43.49
Hansen $p-$value	0.010	0.507	0.001	0.578
BHC controls	Yes	Yes	Yes	Yes

restrictions rejects the null of joint validity of our instruments. Hence, in column II and IV we report the estimated coefficients when IPoD, DSCORE, and all of the accounting variables; capital buffers, non-performing loans, return on equity, and deposits are assumed to be endogenous. In other words, we allow a dynamic relationship between the market assessment on bank's risk, the decision on the level of disclosure, and some of the key accounting identities. The coefficient on DSCORE continues to be negative and significant and the joint validity of our instruments cannot be rejected with a p–value of over 0.5 in both of the specifications. We conclude that bank holding companies with a higher level of disclosure in the previous year have lower market implied default probabilities in the current year, even after controlling for dynamic endogeneity between IPoD and disclosure.

4.4.3 Relationship between disclosure and other enterprise risks

This section examines whether disclosure has any impact on other enterprise risks. We focus on four different risk measures. First, we use the standard deviation of a bank's weekly equity returns as a proxy for aggregate risk (AGG) (Demsetz, Saidenberg and Strahan (1997) and Nier and Baumann (2006)). Second, we measure downside risk (DOWN) as the mean of implied volatility estimates from the option prices written on the bank's stock (Cremers and Weinbaum (2010) and Xing, Zhang and Zhao (2010)). Third, we measure the systematic risk (BETA) by the beta of the firm, estimated from the CAPM model, and finally we include the idiosyncratic risk (IDIO) calculated as the standard deviation of the weekly residuals of the CAPM model.

Table 4.9 present the panel regression estimates of (4.13), where the dependent variable is one of the above defined risk measures. Again, we report two different specifications where the dependent variable is calculated as the average value of the risk over a year and over three-months following an announcement of an annual report.

Table 4.9: Relationship Between Disclosure Score and BHC Risk

Table provides the results of panel regressions that examine the impact of the level of disclosure on a bank holding company's risk. The panel includes the largest 80 U.S. BHCs and spans the time period 1998–2011. AGG is the aggregate risk, calculated as the standard deviation of weekly stock returns, DOWN is the downside risk, measured as the natural logarithm of option implied volatility written on bank's stock. BETA captures the systematic risk and calculated as the estimated beta of the bank from the CAPM model, and finally IDIO is the idiosyncratic risk calculated as the standard deviation of the weekly residuals of the CAPM model. In columns I through IV, the dependent variables are calculated as the annual averages of the corresponding risk, whereas in columns V through VIII we have three-month averages of the risks as dependent variables. The definitions of explanatory variables are presented in Table 4.7. In all of the specifications year and bank fixed effects are included and the explanatory variables are standardized. The $t-$statistics, the sample size and the adjusted R^2s are also reported.

	$\text{AGG}_{b,t+1}$	$\text{DOWN}_{b,t+1}$	$\text{BETA}_{b,t+1}$	$\text{IDIO}_{b,t+1}$	$\text{AGG}_{b,t+3M}$	$\text{DOWN}_{b,t+3M}$	$\text{BETA}_{b,t+3M}$	$\text{IDIO}_{b,t+3M}$
	I	II	III	IV	V	VI	VII	VIII
$\text{DSCORE}_{b,t}$	-0.435	-0.039	-0.092	-0.284	-0.289	-0.045	-0.133	-0.317
	(-3.46)	(-2.12)	(-2.15)	(-2.50)	(-1.53)	(-2.73)	(-2.95)	(-1.75)
$\text{SIZE}_{b,t}$	-4.099	-0.170	-0.298	-3.913	-5.119	-0.280	-1.004	-4.749
	(-7.29)	(-3.23)	(-2.09)	(-7.51)	(-4.00)	(-3.49)	(-3.72)	(-4.03)
$\text{VOLA}_{b,t}$		0.063	0.125	0.005		0.0597	0.305	0.035
		(1.92)	(2.26)	(0.03)		(1.51)	(2.75)	(0.10)
$\text{BETA}_{b,t}$	0.294	0.039		0.320	0.438	0.064		0.187
	(2.44)	(2.52)		(3.29)	(2.80)	(3.55)		(1.80)
$\text{CAPBUF}_{b,t}$	-0.035	-0.015	-0.024	-0.077	-0.562	-0.037	-0.088	-0.117
	(-0.31)	(-0.85)	(-0.87)	(-0.84)	(-3.58)	(-2.93)	(-2.11)	(-1.08)
$\text{NPL}_{b,t}$	0.391	0.049	0.017	0.464	0.364	0.035	-0.356	0.619
	(2.22)	(2.35)	(0.40)	(2.20)	(0.82)	(1.60)	(-2.64)	(1.20)
$\text{ROE}_{b,t}$	-0.020	0.040	0.071	-0.129	-0.347	0.013	0.292	-0.372
	(-0.12)	(2.77)	(1.93)	(-0.86)	(-1.42)	(0.49)	(2.05)	(-1.52)
$\text{DEPO}_{b,t}$	3.749	0.182	0.580	3.412	4.487	0.214	1.247	3.746
	(5.64)	(2.35)	(3.92)	(5.84)	(3.50)	(2.25)	(4.83)	(3.56)
cons	5.914	-0.846	0.543	5.904	7.117	-0.753	0.338	6.367
	(18.83)	(-25.04)	(8.77)	(21.24)	(15.96)	(-19.48)	(3.56)	(16.64)
Obs.	851	614	845	844	868	584	871	867
adjR^2	0.783	0.873	0.562	0.751	0.765	0.883	0.546	0.688
Year & Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

In columns I and V, the dependent variable is the aggregate risk, AGG. The expected sign of disclosure on volatility is ambiguous. On one hand, there is a literature providing evidence that disclosure moves the stock prices and increases volatility (See for example, Ross (1989) and Leuz and Verrecchia (2000)). On the other hand, a number of studies argue that the level of disclosure is negatively associated with the stock return volatility (See for example, Bushee and Noe (2000), Baumann and Nier (2004), Kothari et al. (2009), and So (2011)). Our findings provide supporting evidence for the latter group; banks with higher level of disclosure benefit from lower stock volatility. However, the relationship is significant at a 5% level only for longer horizons. One standard deviation increase in the level of disclosure in the previous year decreases the stock return volatility by 0.435% in the following year. This could be explained by two: first, enhanced disclosure may reduce the impact of news about a firm's performance since it reduces the uncertainty. Second, as market microstructure theory suggests, in a market where some investors have access to better information than others, by reducing the information asymmetries, disclosure diminishes the advantage to be better-informed. To the extent that this is true, enhanced disclosure reduces the price impact of a trade initiated by informed agents (Diamond and Verrecchia, 1991).

Columns II and VI reveal a negative and significant relationship between the disclosure and the downside risk irrespective of the period chosen. In line with the findings of Ederington and Lee (1996) and Rogers et al. (2009) implied volatility; the market's participants risk-neutral expectations of volatility, declines following information releases in long-window changes. Finally, results presented in columns III, IV, VII, and VIII show that enhanced disclosure in the previous year significantly decreases both the systematic and the idiosyncratic risk of a bank holding company, whereas the association is stronger for the latter.

Besides, DSCORE, in all of the specifications, size is significantly and negatively associated with the enterprise risks we consider. Bigger bank holding companies benefit from lower enterprise risk, all else being equal.

4.4.4 Determinants of disclosure

In this section, we examine the determinants of disclosure in a cross-sectional setting. First, we include the one period-lagged value of disclosure in our analysis. We expect the firms' past disclosure behavior affecting their prosperity to provide voluntary disclosures in the future. In their multiperiod analysis, Einhorn and Ziv (2008) argue that by providing higher levels of voluntary disclosure, the managers make an implicit commitment to provide similar disclosures in the future.

There is a rich empirical literature providing evidence that the bigger firms disclosure more information.[15] Due to their more risk taking behavior and complex activities, the bigger firms are more likely to disclose more if disclosure is a tool to reduce information asymmetries. Following this, we include the size of the bank holding company, measured as the natural logarithm of the market capitalization at the end of each year.

Third, we consider the profitability of the bank holding company, proxied by return on equity ratio and the average weekly returns in the current year. Both theoretical and empirical evidence on the relationship between profitability and the level of disclosure is mixed. Several papers argue that the managers of more profitable firms have incentives to disclose more to signal the quality (e.g., Verrecchia (1983), Verrecchia (1990), and Dye (1985)). On the other hand Lang and Lundholm (1993) and Skinner (1994) argue that the management is more likely to disclose bad news than good news. In an empirical analysis conducted on the largest 50 U.S. and Japanese companies, Ho and Taylor (2007) show that firms with lower profitability disclose more information. Bujaki and McConomy (2002) find that the extent of

[15] See for example Botosan and Plumlee (2002), Bujaki and McConomy (2002), Ho and Taylor (2007), and Francis et al. (2008).

Table 4.10: Determinants of Disclosure

Table presents the results for the cross sectional regressions that examine the determinants of disclosure. Sample includes the largest 80 U.S. BHCs and spans the time period 1998–2011. The dependent variable, DSCORE, is the aggregated disclosure score. All of the control variables are listed in Appendix C. All of the specifications include the year fixed effects and t-values reported in the parentheses are based on the robust standard errors that are clustered at the bank holding company level. For each of the specification, the sample size and the adjusted R^2s are also reported.

	$\text{DSCORE}_{b,t+1} = \gamma_0 + \gamma_1\text{DSCORE}_{b,t} + \gamma_2\text{DSCORE}_{b,t-1} + \text{controls}_{b,t} + \varepsilon_{b,t}$			
	I	II	III	IV
$\text{DSCORE}_{b,t}$	1.335	1.293	1.283	1.287
	(11.60)	(11.49)	(11.29)	(11.94)
$\text{DSCORE}_{b,t-1}$	0.045	0.035	0.045	0.037
	(0.39)	(0.30)	(0.38)	(0.34)
$\text{SIZE}_{b,t}$		0.094	0.108	0.102
		(4.68)	(4.71)	(3.75)
$\text{ROE}_{b,t}$			-0.033	-0.042
			(-2.24)	(-2.36)
$\text{RET}_{b,t}$			-0.045	-0.040
			(-2.32)	(-1.70)
$\text{IPoD}_{b,t}$				-0.011
				(-0.76)
$\text{BETA}_{b,t}$				0.072
				(2.16)
$\text{VOLA}_{b,t}$				-0.055
				(-1.83)
Obs.	895	889	861	612
adjR^2	0.927	0.932	0.933	0.93
Year FE	Yes	Yes	Yes	Yes

disclosure is affected by revenue growth, with higher growth firms disclosing less than their peers. On the other hand, Cheung et al. (2010) show that more profitable companies tend to disclose more.

Finally, we include the annual IPoD, volatility of the weekly equity returns and estimated beta of the company in our analysis. One can expect a significant relationship between the disclosure and risk taking behavior of a bank if there is an effective market discipline; increased public information helps investors to assess the riskiness of a bank, punish accordingly, and changes in market dynamics influence the decisions made by bank management in turn. Table 4.10 present the results.

All of the specifications include the year fixed effects and t-values reported in the parentheses are based on the robust standard errors that are clustered at the bank holding company level. In line with the findings of Einhorn and Ziv (2008), we find that the current year's level of disclosure is positively and significantly associated with next year's disclosure, whereas we find the two-period lagged value of disclosure does not significantly associated with the current level. Second column reports the results where the size of the company is included in the analysis. In harmony with the extant literature, we find that the bigger companies disclose more information on their risk structure. In column III we include the two proxies of the past performance; return on equity and average of the weekly equity returns. Results reveal the negative association of the variables with the next year's level of disclosure. Finally column IV reports the results where IPoD, volatility and beta of the company are included in the analysis. Only the relationship between current year's BETA and next year's DSCORE is statistically significant at a 5% level. Higher levels of observed systematic risk this year encourage the managers to disclose more information. This is consistent with the market discipline hypothesis mentioned earlier.

4.4.5 Further robustness checks

We perform four sets of robustness tests to confirm the validity of our results. First, we examine whether the documented relationship between disclosure and the annual and three-months-level of implied default probability (IPoD) holds for other time intervals. To examine this, we re-estimate our baseline panel regression (4.13) with dependent variable equal to the bimonthly and semi-annually averages of IPoD estimates following the disclosure date. Second, we assess the sensitivity of our results to model specification by using the logit-transformed market implied default probability instead of a log-transformed IPoD. Third, we allow the disclosure measure to enter the model as a dummy variable. We define a variable DISCDECILE, which takes the value one if the corresponding bank's level of disclosure in a given year belongs to the top deciles compared to its peers. Finally, we change the definition of our volatility and instead of calculating volatility from bid and ask prices of the equity, we use the standard deviation of weekly equity returns (VOLAret).

Table 4.11 presents the results. The first two columns repeat the estimated coefficients for the baseline regressions, where the main dependent variables are the annual and three-month averages of IPoD estimates. In columns III and IV, the dependent variables are the bimonthly and semi-annually averages of IPoD estimates following the announcement of an annual report. Column V reports the estimated coefficients when we use logit transformation instead of a log-transformation of IPoD. In column VI we change the definition of the disclosure index and finally the last column reports the robustness when volatility is calculated from the weekly returns.

Results confirm the robustness of the documented relationship between market implied default probability and disclosure. The economic significance of the association is almost monotonically decreasing with the period; it is highest for the three-months ahead and lowest for the annual. Irrespective of the horizon, disclosure is negatively and significantly associated with the next period IPoD. Logit-transformed dependent variable produces almost identical results. When we allow disclosure to enter the specification as a dummy variable, we see that the economic relationship increases significantly.

Table 4.11: Robustness Analysis

Table presents the results for the robustness analysis. The panel includes the largest 80 U.S. BHCs and spans the time period 1998–2011. Columns I and II repeat the estimated coefficients for the baseline regressions, where the main dependent variables are the annual and three-month averages of IPoD estimates, respectively. In columns III and IV, the dependent variables are the bimonthly and semi-annually averages of IPoD following the announcement of an annual report. Column V reports the estimated coefficients when we use logit transformation instead of a log-transformation. In column VI we measure disclosure with a dummy variable (DISCDECILE), which takes 1 of the bank is in the top decile, and finally the last column reports the robustness when volatility is calculated as the standard deviation of weekly equity returns. The definitions of the other variables are presented in Table 4.7. In all of the specifications year and bank fixed effects are included and the explanatory variables are standardized. The $t-$statistics, the sample size and the adjusted R^2s are also reported.

	$\text{IPoD}_{b,t+1}$	$\text{IPoD}_{b,t+3M}$	$\text{IPoD}_{b,t+2M}$	$\text{IPoD}_{b,t+6M}$	$\text{logitIPoD}_{b,t+1}$	$\text{IPoD}_{b,t+1}$	$\text{IPoD}_{b,t+1}$
	I	II	III	IV	V	VI	VII
$\text{DSCORE}_{b,t}$	-0.183	-0.298	-0.279	-0.199	-0.188		-0.179
	(-2.11)	(-2.59)	(-2.38)	(-2.06)	(-2.13)		(-2.06)
DISCDECILE						-0.631	
						(-2.70)	
$\text{SIZE}_{b,t}$	-0.593	-1.007	-1.015	-0.851	-0.673	-0.577	-0.649
	(-2.38)	(-3.05)	(-2.87)	(-3.07)	(-2.66)	(-2.33)	(-2.61)
$\text{VOLA}_{b,t}$	0.215	0.143	0.226	0.202	0.206	0.214	
	(1.61)	(0.82)	(1.26)	(1.39)	(1.52)	(1.61)	
$\text{VOLAret}_{b,t}$							0.015
							(0.13)
$\text{BETA}_{b,t}$	0.284	0.571	0.550	0.411	0.289	0.283	0.334
	(3.87)	(5.76)	(5.36)	(4.95)	(3.88)	(3.87)	(4.00)
$\text{CAPBUF}_{b,t}$	-0.195	-0.229	-0.183	-0.259	-0.193	-0.181	-0.195
	(-1.76)	(-1.64)	(-1.30)	(-2.21)	(-1.71)	(-1.64)	(-1.75)
$\text{NPL}_{b,t}$	0.381	0.143	0.0451	0.142	0.397	0.384	0.442
	(3.63)	(1.52)	(0.44)	(1.79)	(3.72)	(3.67)	(4.38)
$\text{ROE}_{b,t}$	0.125	0.108	-0.171	0.127	0.142	0.123	0.114
	(1.72)	(1.05)	(-1.04)	(1.47)	(1.92)	(1.70)	(1.56)
$\text{DEPO}_{b,t}$	0.62	0.543	0.660	0.410	0.692	0.726	0.607
	(1.88)	(1.23)	(1.43)	(1.10)	(2.07)	(2.18)	(1.83)
cons	-4.438	-4.256	-4.220	-4.349	-4.407	-4.262	-4.435
	(-26.63)	(-19.33)	(-18.77)	(-23.42)	(-26.05)	(-25.47)	(-26.43)
Obs.	613	584	568	593	613	613	613
adjR^2	0.740	0.681	0.693	0.703	0.740	0.742	0.739
Year & Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

4.5 Conclusion

Increased uncertainty is argued as one of the main reasons of the breakdown of trading and the associated withdrawal of liquidity in many markets during the crisis. In periods of stress, there is a flight to quality and safe-heaven. Hence, investors with imperfect information over the quality of assets reduce their holdings, while holders of "safe" assets are unwilling to sell at prevailing market prices, leading to a collapse of market functioning. The evidence presented in this paper suggests that disclosure may help to mitigate some of these informational frictions. In particular, we show that an increased disclosure affects the investors' believes on the riskiness of a bank and is followed by reduced market implied default probabilities. This result is robust to the inclusion of a number of other bank characteristics and as well the adjustment of dynamic endogeneity.

Our sample includes the largest 85 U.S. bank holding companies in terms of asset size as of December 2007, for the period 1998–2011. We measure the level of disclosure by a self-constructed disclosure index and we estimate the market implied default probabilities from traded options of a given company. By construction, option implied default probabilities reflect investors' expectations of the bank holding companies' average default probability over option time to expiration. Hence, one can argue that the communication processes increase transparency and eliminate disparities between what investors understand and expect and what management intends to deliver. Moreover, we provide evidence that voluntary disclosure has a sizeable and beneficial effect on other enterprise risks; return volatility, downside risk, market risk, and idiosyncratic risk of a bank holding company. Finally, our results show that bigger companies, companies with lower performance, and higher market risk are more likely to provide higher level of disclosures compared to their peers, all else being equal.

Our analysis provides possible policy implications. High disclosure is a necessity condition for the market discipline and it seems to provide incentives for investors to reward the high disclosed banks. This could be beneficial for the bank as well if reduced risk is translated into a reduced cost of capital as documented in the literature. Our analysis shows that there is a number of areas in which the banks are fail to provide sufficient information, where more granular quantitative disclosures are required on the liquidity risk, especially on the information on unencumbered funding and liquid asset holdings. Moreover there an evidence of lack of disclosure on information related to the credit exposure to other financial institutions and exposure to the special purpose entities.

Chapter 5

Conclusions

Chapters 2 and 3 of this thesis have explored a number of different issues in pure order driven exchanges. Generally, the results presented in these two chapters suggest that the state of a limit order book contains non-negligible information about the short-term aggregate price formation process and order choice strategies of the agents. Specifically, "Global Depth and Future Volatility" investigates the predictive power of a limit order book distribution on intraday market and individual stock volatilities. We summarize the relative accumulation of quoted depth in a given limit order book and provide empirical evidence that our summary measure, global depth, has a leading in-sample and out-of-sample predictive power over market volatility.

Our results presented in Chapter 2 points out several directions for further research. First, we left the relationship between aggregate liquidity and subsequent returns as a further research. One can easily argue that the relative price position of the quoted depth can signal and can change the expectations of the direction of the price level as well. However, when we test the predictability of the intraday returns, we fail to obtain any significant explanatory power of either global depth or other variables, such as trading activity measures or bid-ask spread. Next, we focus on the daily predictability regressions. We find that global depth is a significant predictor of daily returns at a 10% level, which could be results of a small daily sample size. Hence, as a next step, one could increase the sample size, or adopt the measure on another exchange with longer time-series data availability and test the relationship between global depth and future returns.

Whether traders can design execution strategies that allow them to reduce execution costs using the informativeness of the state of a limit order book is another possible interesting question. There is a vast theoretical literature on order choice concluding that the agents submit limit orders rather than the market orders under high volatility states to reduce the picking-off risk (see for instance Foucault (1999)). Moreover, due to option-like features of the limit orders, the execution probability of an order increases with volatility. Putting differently, as option values depend on volatility, volatility information can be used by traders to price their orders. For instance, in an anticipation of increased volatility, a trader can reduce their exposure to the risk of being picked off by submitting less aggressive orders ((Foucault et al., 2007)). Similarly, given access to relevant data set, volatility information can be used to price index options written on the ISE-30 index.

Finally, in this thesis, we mostly focus on the "static" properties of a limit order book. In other words, we focus on the distribution of non-executed limit orders or we consider the state of a book both at a given time. Many other "dynamic" properties can also be analyzed, such as the response of the order flow (i.e. order choice) after following a high volatility state. These can be commonly obtained by fitting the volatility and limit order book distribution data to a Vector Auto-Regression model.

Chapter 3, "Competition, Signaling and Non-walking Through the Book: Effects on Order Choice" investigates the following questions: (1) Do agents consider the competition and signaling effects during their order submission?; (2) How does the feature of non-walking through the book affect order decision of an impatient trader?; and (3) What is the difference in trading behavior between institutional and individual investors? We find that the competition effect is stronger than the signaling effect for both sides of the market. For a limit order trader, the competition effect is the strongest for the volume at the second best quotes. On the other hand, we show that under non-walking through the book rule, only volatility and price trend affect the order choice decision of a market order trader. 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.

The novel feature of the chapter is an emphasis on the effect of competition/signaling in a pure order driven market when walking through the order book is not allowed. The research questions we addressed have not been thoroughly treated in the current literature. Chapter 3 focuses the effects a particular market mechanism; non-walking through the book 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, "mechanically" the immediate price impact of such an order is higher under walking through the book. Whether this creates any permanent price impact is important information for stock exchanges while designing their market mechanism rules.

Chapter 4 of this thesis on the other hand, addresses a question in the area of corporate governance. It specifically asks whether enhanced disclosure can alleviate the informational frictions between the management and investors and in turn can affect the investor's assessment on the riskiness of a firm. By proposing a template to measure the level of voluntary disclosure, I conclude that increased disclosure is associated with lower levels of market implied default probability.

Increased uncertainty is argued as one of the main reasons of the breakdown of trading and the associated withdrawal of liquidity in many markets during the crisis. The evidence presented in this chapter suggests that disclosure may help to mitigate some of these informational frictions and is beneficial for various enterprise risks. This finding is important for policy makers. High disclosure is a necessity condition for market discipline and it seems to provide incentives for investors to reward the high disclosed banks. If reduced risk is translated into a reduced cost of capital as documented in the current literature (see for instance Botosan (1997), Leuz and Verrecchia (2000), Botosan and Plumlee (2002), and Barth et al. (2013)), disclosure can be argued as beneficial for the bank management as well.

The key methodological issue in this chapter is the endogeneity of the disclosure variable, making it difficult to interpret the findings. A bank holding company's past level of market implied default risk can affect both the current level of default risk and the decision on the past and current levels of disclosure. In other words, the main explanatory variable and the response variable of the baseline panel regression model can be dynamically endogenous. In this case, the dynamic endogeneity can be solved through employing a panel GMM estimation, as discussed in Section 4.4.2. However, a bank holding company that is exposed to higher risk may choose to disclose more information to reduce the uncertainty and change investor's assessment on its risk. In other words, the bank holding company's risk culture can jointly determine both the choice of risk and the level of disclosure. Otherwise, an institution can optimally choose the level of disclosure given the level of risk it undertakes, i.e. the causality can be reversed.

One possible solution to assess whether the relation between high levels of disclosure and lower market implied default probabilities for these firms is actually a result of a change in the level of disclosures, rather than the changes that the firm is experiencing, is adopting a diff-in-diff estimation methodology. The other solution is to find a relevant instrument for the level of disclosure and employ a fixed-effects instrumental variables estimation (two-stage least squares).

However, finding strong instruments is a challenge and with weak instruments, the fixed-effects IV estimators are likely to be biased as well. Moreover, to employ a diff-in-diff methodology, I need a natural experiment or policy change that "treats" only a particular subsample of the bank holding companies in my sample. In other words, adopting both of the aforementioned solutions are not straightforward and hence, left as further work.

Another possible route to take as a further analysis of this chapter is to employ a cross country analysis to examine the differences in the level of disclosure, and whether there is a link between the transparency and the soundness of the overall banking system. The question has obvious policy implications and could be interesting for regulators.

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			$GD(\lambda=0)_\tau^{buy}$		
$DEPTH0_\tau$	0.15			0.14	0.17
	[100/100]			[93/100]	[100/100]
$DEPTH03_\tau$		0.05		0.01	
		[87/87]		[57/53]	
$DEPTH05_\tau$			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			$GD(\lambda=0.5)_\tau^{buy}$		
$DEPTH0_\tau$	0.70			1.10	1.13
	[100/100]			[100/100]	[100/100]
$DEPTH03_\tau$		0.14		-0.16	
		[67/80]		[87/0]	
$DEPTH05_\tau$			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			$GD(\lambda=1)_\tau^{buy}$		
$DEPTH0_\tau$	0.78			1.38	1.36
	[100/100]			[100/100]	[100/100]
$DEPTH03_\tau$		0.24		-0.31	
		[57/63]		[100/0]	
$DEPTH05_\tau$			-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)
Dopp^{1_2}/ 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

In this appendix, we first list and define the key variables used in Chapter 3, Disclosure Practices and Option Implied Probability of Default. The expressions within the parentheses denote the corresponding variable names in the FR Y-9C reports from Federal Reserve Bank of Chicago. Then in Table C.1 we present the sample of bank holding companies (BHCs) with the corresponding identifiers. In the first three columns, the name of the bank holding company, its total asset size as of December 2007 in billion U.S. dollars, the state where the company is registered, RSSID, and PERMNO are presented. The last column presents time span of the BHC in panel data. Finally, Table C.2 presents a sample of option data used to estimate implied probability of default (IPoD) J.P. Morgan for one of the stocks in our sample, for a given trading day. The first two columns give the option trading date (Date) and the expiry date of the option (Exdate). The trading price (Price) calculated as the average of closing bid and ask prices, K is the strike price (note that the first line with the strike price 0 corresponds to the stock price), R_f is the risk free rate corresponding the trading date, T is time to expiry in days, Volume is the trading volume, and Open Int. is the open interest of the option. Finally, w_1 and w_2 correspond to weights calculated using the trading volumes and open interests of the option, respectively as introduced in (4.12).

Variable Descriptions

- **DSCORE**: Total disclosure score. It is calculated as the first principal component of the four main groups: liquidity risk, group structure, intra-annual information and spillover risk.

- **IPoD**: Option implied probability of default. Implied probability of default of a given bank is extracted from equity option prices using the methodology proposed by Capuano (2008) and introduced in Section 4.2.3.

Risk Measures:

- **AGG**: Aggregate risk, calculated as the standard deviation of a bank's weekly equity returns.

- **DOWN**: Downside risk. It is average implied volatility estimated from options written on a bank's stock.

- **BETA**: The estimated beta of a bank from regressions of bank weekly equity returns on the weekly returns of CRSP value-weighted index.

- **IDIO**: Idiosyncratic risk, calculated as the standard deviation of the weekly residuals of the CAPM model.

Bank holding company characteristics:

- **SIZE**: Natural logarithm of the BHC's total market value at the end of the year.

- **VOLA**: Volatility calculated from the weekly bid and ask prices of the bank's stock as defined in Equation (4.14).

- **CAPBUF**: Capital buffer of a BHC at the end of the year. Calculated as the bank's equity capital as a proportion of its total liabilities (BHCK3210/BHCK2948).

- **NPL**: The non-performing loans ratio. It is calculated as the ratio of the sum of loans past due 90 days or more (BHCK5525) and non-accrual loans (BHCK5526) to total assets (BHCK2170).

- **ROE**: Return on equity, calculated as the ratio of the income before extraordinary items (BHCK4300) to total book equity (BHCK3210).

- **DEPO**: The natural logarithm of total deposits (BHDM6631 + BHDM6636 + BHFN6631 + BHFN6636).

Table C.1: List of Bank Holding Companies
This table lists the sample of bank holding companies (BHCs) we included in the analysis with the corresponding identifiers.

NAME	**2007 TA ($bn)**	**STATE**	**RSSID**	**PERMNO**	**SAMPLE**
AMCORE FNCL	5.20	IL	1208661	10304	1998–2009
ASSOCIATED BANC CORP	21.59	WI	1199563	15318	1998–2011
BANCORPSOUTH	13.20	MS	1097614	85789	1998–2011
BANK OF AMER CORP	1720.69	NC	1073757	58827/ 59408	1998–2011
BANK OF HI CORP	10.47	HI	1025309	16548	1998–2011
BANK OF NY MELLON CORP	197.84	NY	3587146	49656	2002–2011
BB&T CORP	132.62	NC	1074156	71563	1998–2011
BOK FC	20.90	OK	1883693	76892	1998–2011
BOSTON PRIVATE FNCL HOLD	6.83	MA	1248078	80223	1998–2011
CAPITAL ONE FC	150.59	VA	2277860	81055	1998–2011
CATHAY GEN BC	10.40	CA	1843080	76504	1998–2011
CENTRAL PACIFIC FC	5.68	HI	1022764	11628	1998–2011
CITIGROUP	2187.63	NY	1951350	70519	1998–2011
CITIZENS REPUBLIC BC	13.52	MI	1205688	86685	1998–2011
CITY NAT CORP	15.89	CA	1027518	23916	1998–2011
COLONIAL BANCGROUP	25.97	AL	1080465	24628	1998–2008
COMERICA	62.76	TX	1199844	25081	1998–2011
COMMERCE BC LLC	49.37	NJ	1117679	86845	1998–2007
COMMERCE BSHRS	16.21	MO	1049341	25129	1998–2011
CORUS BSHRS	8.93	IL	1200393	67046	1998–2008
CULLEN/FROST BKR	13.65	TX	1102367	27888	1998–2011
CVB FC	6.29	CA	1029222	20395	1998–2011
EAST W BC	11.85	CA	2734233	86719	1998–2011
FIFTH THIRD BC	110.96	OH	1070345	34746	1998–2011
FIRST BC	17.19	PR	2744894	11018	1998–2011
FIRST CITIZENS BSHRS	16.23	NC	1075612	10777	1998–2011
FIRST COMMONWEALTH FNCL	5.89	PA	1071306	77643	1998–2011
FIRST HORIZON NAT CORP	37.02	TN	1094640	36397	1998–2011
FIRST MIDWEST BC	8.10	IL	1208184	35917	1998–2011
FIRSTMERIT CORP	10.41	OH	1070804	35167	1998–2011
FNB CORP	6.09	PA	3005332	10629	1998–2011
FRANKLIN RESOURCES	9.63	CA	1246216	37584	1998–2011
FULTON FNCL CORP	15.92	PA	1117129	88197	1998–2011
HANCOCK HC	6.10	MS	1086533	76684	1998–2011
HUNTINGTON BSHRS	54.63	OH	1068191	42906	1998–2011
INTERNATIONAL BSHRS CORP	11.17	TX	1104231	85875	1998–2011
IRWIN FC	6.17	IN	1199732	89237	1998–2008
JPMORGAN CHASE & CO	1562.15	NY	1039502	47896	2000–2011
KEYCORP	99.57	OH	1068025	64995	1998–2011
M&T BK CORP	64.88	NY	1037003	35554	1998–2011

Table C.1: List of BHCs in the sample (cont.)

NAME	**2007 TA ($bn)**	**STATE**	**RSSID**	**PERMNO**	**SAMPLE**
MB FNCL	7.83	IL	1090987	81541	1998–2011
NATIONAL CITY CORP	150.38	OH	1069125	56232	1998–2007
NATIONAL PENN BSHRS	5.82	PA	1117026	56611	1998–2011
NBT BC	5.20	NY	1139279	77415	1998–2011
NEW YORK CMNTY BC	30.60	NY	2132932	79859	1998–2011
NEWALLIANCE BANCSHARES	8.23	CT	3214095	90132	2003–2011
NORTHERN TR CORP	67.61	IL	1199611	58246	1998–2011
OLD NAT BC	7.85	IN	1098303	12068	1998–2011
PACIFIC CAP BC	7.39	CA	1029884	83551	1998–2011
PACWEST BC	5.19	CA	2875332	88343	2000–2011
PARK NAT CORP	6.50	OH	1142336	76266	1998–2011
PNC FNCL SVC GROUP	138.98	PA	1069778	60442	1998–2011
POPULAR	44.41	PR	1129382	16505	1998–2011
PROSPERITY BSHRS	6.38	TX	1109599	86432	1998–2011
PROVIDENT BSHRS CORP	6.47	MD	1247633	11823	1998–2008
PROVIDENT FNCL SVC	6.36	NJ	3133637	89653	2002–2011
REGIONS FC	141.04	AL	3242838	35044	2004–2011
SANTANDER BC	9.15	PR	2847115	86398	2000–2009
SOUTH FNCL GROUP	13.87	SC	1141599	10825	1998–2009
STATE STREET CORP	142.94	MA	1111435	72726	1998–2011
STERLING FC	12.15	WA	3152245	11056	1998–2011
SUNTRUST BK	179.57	GA	1131787	68144	1998–2011
SUSQUEHANNA BSHRS	13.08	PA	1117156	73809	1998–2011
SVB FNCL GRP	6.45	CA	1031449	11786	1998–2011
SYNOVUS FC	33.02	GA	1078846	20053	1998–2011
TCF FC	16.07	MN	2389941	10375	1998–2011
TRUSTMARK CORP	8.97	MS	1079562	35263	1998–2011
U S BC	237.62	MN	1119794	66157	1998–2011
UCBH HOLD	11.80	CA	2694814	86437	1998–2008
UMB FC	9.34	MO	1049828	78829	1998–2011
UMPQUA HC	8.35	OR	2747644	86004	1999–2011
UNIONBANCAL CORP	55.73	CA	1378434	20694	1998–2011
UNITED BSHRS	7.99	WV	1076217	11369	1998–2011
UNITED CMNTY BK	8.21	GA	1249347	89323	1998–2011
VALLEY NAT BC	12.75	NJ	1048773	80072	1998–2011
W HOLD CO	17.93	PR	2801546	93105	1999–2008
WACHOVIA CORP	782.90	NC	1073551	36469	1998–2007
WEBSTER FNCL CORP	17.21	CT	1145476	10932	1998–2011
WELLS FARGO & CO	575.44	CA	1120754	38703	1998–2011
WESBANCO	5.38	WV	1070448	11293	1998–2011
WESTERN ALLI BC	5.02	AZ	2349815	90776	2005–2011
WHITNEY HC	11.03	LA	1079740	77053	1998–2011
WILMINGTON TR CORP	11.62	DE	1888193	83030	1998–2011
WINTRUST FC	9.37	IL	2260406	84636	1998–2011
ZIONS BC	52.95	UT	1027004	84129	1998–2011

Table C.2: Sample of Option Data Used to Estimate IPoD
Table presents a sample of option data used to estimate implied probability of default (IPoD) for one of the stocks in our sample, J.P. Morgan. Date is the option trading date, Exdate is the expiry date of the option, Price is the trading price calculated as the average of closing bid and ask prices, K is the strike price (note that the first line with the strike price 0 corresponds to the stock price), R_f is the risk free rate corresponding the trading date, T is time to expiry in days, Volume is the trading volume, and Open Int. is the open interest of the option. Finally, w_1 and w_2 correspond to weights calculated using the trading volumes and open interests of the option, respectively as introduced in Equation (4.12).

Date	Exdate	Price	K	R_f	T	Volume	Open Int.	w_1	w_2
20101215	20110618	40.21	0	0.001	185			1.000	1.000
20101215	20110618	21.325	20	0.001	185	0	0	0.000	0.000
20101215	20110618	18.925	22.5	0.001	185	0	1	0.000	0.000
20101215	20110618	15.45	25	0.001	185	50	61	0.049	0.001
20101215	20110618	10.85	30	0.001	185	0	701	0.000	0.008
20101215	20110618	8.725	32.5	0.001	185	20	274	0.020	0.003
20101215	20110618	7.55	34	0.001	185	69	137	0.067	0.002
20101215	20110618	6.75	35	0.001	185	15	279	0.015	0.003
20101215	20110618	6.05	36	0.001	185	4	224	0.004	0.002
20101215	20110618	5.375	37	0.001	185	12	884	0.012	0.010
20101215	20110618	4.7	38	0.001	185	21	5420	0.021	0.060
20101215	20110618	4.1	39	0.001	185	2	1239	0.002	0.014
20101215	20110618	3.55	40	0.001	185	249	12211	0.243	0.135
20101215	20110618	3.05	41	0.001	185	54	1256	0.053	0.014
20101215	20110618	2.595	42	0.001	185	2	6441	0.002	0.071
20101215	20110618	2.185	43	0.001	185	29	894	0.028	0.010
20101215	20110618	1.825	44	0.001	185	61	1137	0.060	0.013
20101215	20110618	1.505	45	0.001	185	160	24056	0.156	0.267
20101215	20110618	1.24	46	0.001	185	183	148	0.179	0.002
20101215	20110618	0.52	50	0.001	185	22	24111	0.022	0.267
20101215	20110618	0.16	55	0.001	185	70	10772	0.068	0.119

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.

Akerlof, G., A. (1970). The market for "lemons": Quality uncertanity and the market mechanism, *The Quarterly Journal of Economics* **84**: 488–500.

Alhassid, Y., Agmon, N. and Levine, R. D. (1978). An upper bound for the entropy and its applications to the maximal entropy problem, *Chemical Physics Letters* **653**: 22–26.

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.

Anginer, D. and Yildizhan, C. (2010). Is there a distress risk anomaly? corporate bond spread as a proxy for default risk. World Bank Policy Research Working Paper, 5319.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations., *Review of Economic Studies* **58**: 277–297.

Balakrishnan, K., Billings, M., B., Kelly, B., T. and Ljungqvist, A. (2013). Shaping liquidity: On the causal effects of voluntary disclosure. Working paper.

Bank of England (2009). Financial stability report. December.
URL: *http://www.bankofengland.co.uk/publications/Documents/fsr/2009/fsrfull0912.pdf*

Barrell, R., Davis, E., P., Fic, T. and Karim, D. (2010). Is there a link from bank size to risk taking? NIESR Discussion Paper No. 367.

Barth, M., Konchitchki, Y. and Landsman, W. (2013). Cost of capital and earnings transparency, *Journal of Accounting and Economics* **55**: 206–224.

Baumann, U. and Nier, E. (2003). Market discipline and financial stability: some empirical evidence, *Technical report*, Bank of England.
URL: *http://www.bankofengland.co.uk/publications/fsr/2003/fsr14art7.pdf*

Baumann, U. and Nier, E. (2004). Disclosure, volatility and transparency: an empirical investigation into the value of bank disclosure, *Technical report*, Federal Reserve Bank of New York.
URL: *http://www.ny.frb.org/research/epr/04v10n2/0409Baum.pdf*

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.

Beltratti, A. and Stulz, R., M. (2011). The credit crisis around the globe: Why did some banks perform better? Fisher College of Business Working Paper Series.

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.

Botosan, A., C. (1997). Disclosure level and the cost of equity capital, *The Accounting Review* **72**: 323–349.

Botosan, A., C. and Plumlee, M., A. (2002). A re-examination of disclosure level and the expected cost of equity capital, *Journal of Accounting Research* **40**: 21–40.

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

Bujaki, M. and McConomy, B., J. (2002). Corporate governance: Factors influencing voluntary disclosure by publicly traded canadian firms, *Canadian Accounting Perspectives* **1**: 105–139.

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

Bushee, B., J. and Noe, C., F. (2000). Corporate disclosure practices, institutional investors, and stock return volatility, *Journal of Accounting Research* **38**: 171–202.

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

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.

Capuano, C. (2008). The option-ipod. the probability of default implied by option prices based on entropy, *Technical report*, IMF, WP/08/194.

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

Cheung, Y., Jiang, P. and Tan, W. (2010). A transparency disclosure index measuring disclosures: Chinese listed companies, *Journal of Accounting Public Policy* **29**: 259–280.

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.

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.

Cover, T. and Thomas, J. (2006). *Elements of Information Theory*, Wiley, New York.

Cremers, M. and Weinbaum, D. (2010). Deviations from put-call parity and stock return predictability, *Journal of Financial and Quantitative Analysis* **45**: 335–367.

Crockett, A. (2002). Market discipline and financial stability, *Journal of Banking and Finance* **26**: 977–987.

Demsetz, R., S., Saidenberg, M., R. and Strahan, P., E. (1997). Agency problems and risk taking at banks, *Technical report*, Federal Reserve Bank of New York.

Diamond, D. and Verrecchia, R. (1991). Disclosure, liquidity, and the cost of capital, *Journal of Finance* **66**: 1325–1355.

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.

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

Dye, R. (1985). Disclosure of nonproprietary information, *Journal of Accounting Research* **23**: 123–145.

Ederington, L., H. and Lee, J. H. (1996). The creation and resolution of market uncertainty: the impact of information releases on implied volatility, *Journal of Financial and Quantitative Analysis* **31**: 513–539.

Einhorn, E. and Ziv, A. (2008). Intertemporal dynamics of corporate voluntary disclosures, *Journal of Accounting Research* **46**: 567–589.

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.

Ellul, A. and Yerramilli, V. (2012). Stronger risk controls, lower risk: Evidence from u.s. bank holding companies, *Journal of Finance* **Forthcoming**.

Fahlenbrach, R., Prilmeier, R. and Stulz, R., M. (2011). This time is the same: Using bank performance in 1998 to explain bank performance during the recent financial crisis, *Journal of Finance* **Forthcoming**.

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.

Francis, J., Nanda, D. and Olsson, P. (2008). Voluntary disclosure, earnings quality, and cost of capital, *Journal of Accounting Research* **46**: 53–99.

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.

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

Grossman, S., J. (1981). The informational role of warranties and private disclosure about product quality, *Journal of Law and Economics* **24**: 461–483.

Grossman, S. J. and Miller, O. D. (1980). Disclosure laws and takeover bids, *Journal of Finance* **35**: 323–334.

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.

Healy, P., Hutton, A. and Palepu, K. (1999). Stock performance and intermediation changes surrounding sustained increased in disclosure, *Contemporary Accounting Research* **16**: 485–520.

Healy, P. and Palepu, K. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature, *Journal of Accounting and Economics* **31**: 405–440.

Heflin, F., L., Shaw, K., W. and Wild, J., J. (2005). Disclosure policy and market liquidity: Impact of depth quotes and order sizes, *Contemporary Accounting Research* **22**: 829–865.

Hermalin, B., E. and Weisbach, M., S. (2012). Information disclosure and corporate governance, *Journal of Finance* **67**: 195–233.

Ho, L., J. and Taylor, M., E. (2007). An empirical analysis of triple bottom-line reporting and its determinants: Evidence from the united states and japan, *Journal of International Financial Management and Accounting* **18**: 123–150.

Jeong, M., G. (2010). The probability of default implied by binary option prices, *International Journal of Contemporary Mathematical Sciences* **5, no 5**: 201–207.

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.

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

Kothari, S., P., Li, X. and Short, J., E. (2009). The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis, *The Accounting Review* **84**: 1639–1670.

Laeven, L. and Levine, R. D. (2009). Bank governance, regulation and risk taking, *Journal of Financial Economics* **93**: 259–275.

Lambert, R., Leuz, C. and Verrecchia, R. (2007). Accounting information, disclosure, and the cost of capital, *Journal of Accounting Research* **45**: 385–420.

Lang, M. and Lundholm, R. (1993). Cross-sectional determinants of analyst ratings of corporate disclosures, *Journal of Accounting Research* **31**: 246–271.

Lang, M. and Lundholm, R. (2000). Voluntary disclosure and equity offerings: Reducing information asymmetry or hyping the stock?, *Contemporary Accounting Research* **17**: 623–662.

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

Leuz, C. and Verrecchia, R. (2000). The economic consequences of increased disclosure, *Journal of Accounting Research* **38**: 91–124.

Lewis, A., N. (2011). Equity volatility across industries from ipo day. Princeton University Department of Economics Working paper.

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.

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

Merton, R. (1974). On the pricing of corporate debt: The risk structure of interest rates, *Journal of Finance* **29**: 449–470.

Milgrom, P., R. (1981). Good news and bad news: Representation theorems and applications, *The Bell Journal of Economics* **12**: 380–391.

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.

Nier, E., W. and Baumann, U. (2006). Market discipline, disclosure and moral hazard in banking, *Journal of Financial Intermediation* **15**: 332–361.

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.

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.

Rogers, J., L., Skinner, D., J. and Van Buskirk, A. (2009). Earnings guidance and market uncertainty, *Journal of Accounting and Economics* **48**: 90–109.

Ross, S. A. (1989). Information and volatility: The no–arbitrage martingale approach to timing and resolution irrelevancy, *Journal of Finance* **1**: 1–17.

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.

Senior Supervisors Group (2008). Leading-practice disclosures for selected exposures. **URL:** *http://www.newyorkfed.org/newsevents/news/banking/2008/SSG_Leading_Practice_Disclosures.pdf*

Sironi, A. (2003). Testing for market discipline in the european banking industry: Evidence from subordinated debt issues, *Journal Money Credit and Banking* **35**: 332–361.

Skinner, D., J. (1994). Why firms voluntarily disclose bad news, *Journal of accounting research* **32**: 38–60.

So, E. (2011). Disclosure and the option value of equity.

Tadesse, S. (2006). The economic value of regulated disclosure: Evidence from the banking sector, *Journal of Accounting and Public Policy* **25**: 32–70.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market, *Journal of Finance* **62**: 1139–1168.

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

Verrecchia, R. (1983). Discretionary disclosure, *Journal of Accounting and Economics* **5**: 179–194.

Verrecchia, R. (1990). Information quality and discretionary disclosure, *Journal of Accounting and Economics* **12**: 365–380.

Verrecchia, R., E. (2001). Essays on disclosure, *Journal of Accounting and Economics* **32**: 97–180.

Vilsmeier, J. (2011). Updating the option implied probability of default methodology. BGPE Discussion Paper, No. 107.

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

Xing, Y., Zhang, X. and Zhao, R. (2010). What does individual option volatility smirk tell us about future equity returns?, *Journal of Financial and Quantitative Analysis* **45**: 641–662.

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.