HEC MONTRÉAL
École affiliée à l'Université de Montréal

High-Frequency Liquidity, Risk Management and Trading Strategy

Par
Xiaozhou Zhou

Thèse présentée en vue de l'obtention du grade de Ph. D. en administration
(option Finance)

June 2015

HEC MONTRÉAL

École affiliée à l'Université de Montréal

Cette thèse intitulée :

High-Frequency Liquidity, Risk Management and Trading Strategy

Présentée par :

Xiaozhou Zhou

a été évaluée par un jury composé des personnes suivantes :

Pascale Valery
HEC Montréal
Présidente-rapporteuse

Georges Dionne
HEC Montréal
Directeur de recherche

Maria Pacurar
Rowe School of Business
Codirectrice de recherche

Tolga Cenesizoglu
HEC Montréal
Membre du jury

Christian Gourieroux
University of Toronto
Examinateur externe

Patrick Soriano
HEC Montréal
Représentant du directeur de HEC Montréal

Résumé

Cette Thèse propose trois articles à propos de la Microstructure du marché**.** Dans cet objectif, nous avons trois articles différents et complémentaires.

Le premier chapitre développe la mesure du risque de haute-fréquence: la Valeur à Risque Intra-journalière (LIVaR) avec liquidité-ajustée. Notre objectif est de considérer, de façon explicite la dimension de la liquidité endogène, associée avec la grandeur des ordres. En reconstruisant le carnet d'ordre de Deutsche Börse, les changements de rendements sans friction et les changements de rendements actuels sont modélisés conjointement. Ce modèle aide à identifier les mouvements des rendements actuel et sans friction; et aussi pour quantifier le risque relié à la prime de liquidité, qui peut compter jusqu'à 35% du risque total, dépendamment de la grandeur de l'ordre. Notre modèle peut non seulement être utilisé pour identifier l'impact du risque de la liquidité sur le risque total, mais aussi pour avoir une estimation de Valeur à Risque pour le rendement actuel pour un moment précise dans la journée.

Chapitre 2 analyses si les informations intégrées dans le carnet d'ordres à haute fréquence permettent de prédire la dynamique des prix et la direction des transactions. Pour ce faire, nous reconstruisons les 20 premiers niveaux d'un carnet d'ordres (à chaque milliseconde) des actions échangées dans le système électronique Xetra de la Bourse de Francfort. Ensuite, nous construisons différentes variables qui résument les dimensions différentes des informations intégrées dans le carnet d'ordres, tels que la profondeur, la pente, la convexité d'achat et de vente et différents ratios. En suivant le modèle proposé par Hasbrouck (1991), nous estimons un système linéaire d'un vecteur autorégressif (VAR) qui inclut le rendement de milieu de la fourchette, la direction des transactions et une variable d'information de carnet d'ordres. Consistent avec le modèle théorique sur le marché avec un carnet d'ordres, nous constatons que l'état du carnet d'ordres peut prédire, à court terme, le rendement de milieu de la fourchette et la direction des transactions. En revanche, certains variables qui résument les informations des niveaux plus hauts du carnet d'ordres ont également des effets cumulatifs importants à long terme sur le rendement de milieu de la fourchette et la direction des transactions.

Par conséquent, la réalisation complète de ces effets prend du temps. Finalement, par un simple exercice de transactions à haute fréquence, nous démontrons qu'il est difficile d'obtenir des gains économiques de la relation entre les variables du carnet d'ordres et le rendement de milieu de la fourchette à cause des coûts de transaction.

Le troisième chapitre se concentre sur la liquidité ex-ante qui est intégrée dans le carnet d'ordre et ses mouvements. En utilisant les données reconstruites du carnet d'ordre du Système de transactions Xetra, nous utilisons un modèle de décomposition pour chercher l'impact de la durée de la transaction, de la durée de la côte, et autres variables exogènes sur le changement de liquidité ex-ante. Dans ce modèle, le facteur de la durée est présumé être strictement exogène, et ses mouvements sont captés par un processus de Log-ACD. De plus, en prenant en considération les données d'ultra haute fréquence (UHF), notre modèle consiste à décomposer la distribution conjointe de la mesure de la liquidité ex-ante dans des distributions qui sont simples et faciles à interpréter. Nos résultats suggèrent que les durées de négociation et de côte ont une influence sur les variations de liquidité ex-ante, les variables à court terme comme le changement de l'écart de fourchette et le volume peuvent aussi prédire la tendance du changement de la liquidité. Toutefois, contrairement à la littérature précédente, la variable à long terme tel que déséquilibre de négociation est moins informative pour prédire la dynamique de changement de liquidité.

Mots clés : la Valeur à Risque Intrajournalière avec liquidité-ajustée, Carnet d'ordre, Stratégie de négociation, Mesure de liquidité Xetra (XLM).

Abstract

This thesis proposes three articles about the market microstructure. To this end, we have three different and complementary articles.

The first chapter develops a high-frequency risk measure: the Liquidity-adjusted Intraday Value at Risk (LIVaR). Our objective is to explicitly consider the endogenous liquidity dimension associated with order size. By reconstructing the open Limit Order Book of Deutsche Börse, changes in the tick-by-tick (ex-ante) frictionless return and actual return are modeled jointly. This modeling helps to identify the dynamics of frictionless and actual returns, and to quantify the risk related to the liquidity premium. In our sample, liquidity risk can account for up to 35% of total risk depending on order size. Our model can be used not only to identify the impact of ex-ante liquidity risk on total risk, but also to provide an estimation of the VaR for the actual return at a point in time.

The second chapter analyzes whether the state of the limit order book affects future price movements in line with what recent theoretical models predict. We do this in a linear vector autoregressive system which includes midquote return, trade direction and variables that are theoretically motivated and capture different dimensions of the information embedded in the limit order book We find that different measures of depth and slope of bid and ask sides as well as their ratios cause returns to change in the next transaction period in line with the predictions of Goettler, Parlour, and Rajan (2009) and Kalay and Wohl(2009). Limit order book variables also have significant long term cumulative effects on midquote return, which is stranger and takes longer to be fully realized for variables based on higher levels of the book. In a simple high frequency trading exercise, we show that it is possible in some cases to obtain economic gains from the statistical relation between limit order book variables and midquote return.

The third chapter focuses on ex ante liquidity embedded in open Limit Order Book (LOB) and its dynamics. Using the tick-by-tick data and the re-constructed open LOB data from Xetra trading system, we utilize a decomposition model to investigate the impact of trade duration, quote duration and other exogenous variables on ex-ante liquidity embedded in open LOB. More specifically, the duration factor is assumed to be strictly exogenous and its dynamics can be captured by a Log-ACD process. Furthermore, by taking into account of Ultra High Frequency (UHF) data, our modeling involves decomposing consistently the joint distribution of the ex-ante liquidity measure into certain simple and interpretable distributions. In this study, the decomposed factors are Activity, Direction and Size. Our results suggest that trade durations and quote durations do influence the ex-ante liquidity changes. Short-run variables, such as spread change and volume, also predict the tendency of liquidity changes. However, contrary to previous literature, the long-term variable trade imbalance is less informative in predicting the liquidity change dynamics.

Keywords: Liquidity-adjusted Intraday Value at Risk, Limit Order Book, Trading strategy, Xetra Liquidity Measure (XLM)

Table of contents

List of tables

List of figures

List of acronyms

- **VaR** Value at Risk
- **LIVaR** Liquidity-adjusted Intraday Value at Risk
- **IVaR** Intraday Value at Risk
- **AR** Actual return
- **FR** Frictionless return
- **VARMA** Vector autoregressive moving average
- **DCC** Dynamic conditional correlation
- **ACD** Autoregressive conditional duration
- **LogACD** Log-Autoregressive conditional duration
- **GARCH** Generalized Autoregressive Conditional Heteroskedasticity
- **NGARCH** Nonlinear GARCH
- **MGARCH** Multivariate GARCH
- **MC** Monte carlo
- **VAR** Vector autoregression
- **LOB** Limit order book
- **IRF** Impulse response function
- **MA** Moving average
- **MLE** Maximum likelihood estimation
- **XLM** Xetra liquidity measure
- **UHF** Ultra high frequency
- **HFT** High frequency trading

To my family who has always been there to support me

To Eva whose love has made this thesis possible.

Acknowledgements

First I would like to express my sincerest gratitude to my supervisor, Professor Georges Dionne, who has supported me throughout my thesis with his patience and knowledge that allowed me to work in my own way. I also thank him for the detailed comments, the financing for my participation to the conferences and the time invested in the three projects. These elements are essential to the success of all this enterprise.

I also would like to thank my co-director, Professor Maria Pacurar, who patiently provided guidance in the modeling and advise in the writing of the project. Her trust and support helped me move forward in the difficult times. I thank her for all the energy and time invested, especially for the constant encouragement in the first project.

I am extremely grateful to Professor Tolga Cenesizoglu whose research experience in econometrics played a vital role in the second project. I also thank him for always being there to assist, especially for giving me confidence for job search.

I am also thankful to my professors in the PhD program who helped to achieve my goal. To Professor Jean-Guy Simonato for his valuable comments, to professor Benoit Perron for the gift of his time and his helpful suggestions for the first project and to Professor Diego Amaya for his constructive comments and time. .

I acknowledge financial support from the Social Sciences and Humanities Research Council of Canada (SSHRC), Fonds de recherche sur la société et la culture (FRQSC), HEC Montréal, the Canada Research Chair in Risk Management, le Centre Interuniversitaire sur le Risque, les Politiques Économiques et l'Emploi (CIRPEE).

Finally, I thank my parents for supporting and encouraging me throughout all my studies. Without you, I would have never been the person I am today. To my wife, Xiaocen, who has had great patience, while making plenty of sacrifices in order to help me complete my PhD program. Thank you for being there to show your support and love.

Chapter 1

Liquidity-adjusted Intraday Value at Risk modeling and risk management: An application to data from Deutsche Börse

Abstract

This paper develops a high-frequency risk measure: the Liquidity-adjusted Intraday Value at Risk (LIVaR). Our objective is to explicitly consider the endogenous liquidity dimension associated with order size. By reconstructing the open Limit Order Book of Deutsche Börse, changes in the tick-by-tick (ex-ante) frictionless return and actual return are modeled jointly. This modeling helps to identify the dynamics of frictionless and actual returns, and to quantify the risk related to the liquidity premium. In our sample, liquidity risk can account for up to 35% of total risk depending on order size. Our model can be used not only to identify the impact of ex-ante liquidity risk on total risk, but also to provide an estimation of the VaR for the actual return at a point in time.

1.1 Introduction

Following increased computerization, many prominent exchanges around the world such as Euronext, the Tokyo Stock Exchange, the Toronto Stock Exchange, and the Australian Stock Exchange have organized trading activities under a pure automatic order-driven structure: there are no designated market-makers during continuous trading, and liquidity is fully guaranteed by market participants via an open Limit Order Book (LOB hereafter). In other main exchange markets including the NYSE and the Frankfurt Stock Exchange, trading activities are carried out under the hybrid structure, which combines the automatic order-driven structure and the traditional floor-based quote-driven structure. Nevertheless, most trades are executed under the automatic order-driven structure due to its advantages of transparency, efficiency, and immediacy. Consequently, trading frequency is rising and trading activity is easier than ever.

As mentioned above, one of the most important features of open LOB is that liquidity is provided entirely by traders who place limit orders. As a result, the role of the traditional market makers has been replaced by informed and uninformed traders.[1] During trading hours, traders can place limit orders with different prices and quantities according to their trading strategies and preferences. Nowadays, these limit orders are placed, updated or cancelled very frequently. As a result, the liquidity embedded in the open LOB also changes quickly. Another important feature of the open LOB market structure is the existence of active trading[2] whereby traders trade over a very short time horizon by monitoring the market continuously or using high-speed computers during the day and liquidating all open positions before market closing. Leaving aside the question of whether active trading is beneficial or not,[3] this new trading culture requires more and more attention to intraday risk derived from market and open LOB. However, traditional risk management has been challenged by this trend towards high-frequency trading because low-frequency measures of risk such as Value at Risk (VaR), which are usually based on daily data, are ill-suited to capturing the potential liquidity risk hidden in very short horizons.

In this paper, we propose a new high-frequency risk measure — Liquidity-adjusted Intraday Value at Risk (LIVaR) — that accounts for both the market risk and the ex-ante liquidity risk of liquidating a position[4]. Ex-ante liquidity is computed for a given order size before the transaction is executed. The difference between ex-ante liquidity and traditional (ex-post) liquidity is that the latter is the liquidity consumed when transactions are completed.[5] By nature, ex-ante liquidity is forward-looking liquidity and presents important and practical aspects for intraday traders and other participants indirectly involved in active trading, including financial institutions[6] and market regulators. Most importantly, compared with the ex-post liquidity measure, the ex-ante liquidity measure is more related to the original definition of a liquid asset. Typically, the recognized description of a liquid security is its

[1] A large number of studies show that informed traders trade via open LOB (Kumar and Seppi (1994), Handa and Schwartz (1996), Bloomfield, O'Hara and Saar (2005), Foucault, Kadan and Kandel (2005), Kaniel and Liu (2006), and Rosu (2009), among others).

[2] High-frequency trading (HFT) was estimated to make up 51% of equity trades in the U.S. in 2012 and 39% of traded value in the European cash markets (Tabb Group). In this study, we distinguish between active traders and high-frequency traders. As mentioned in Ait-Sahalia and Saglam (2014), HFT refers to trading strategies whose profits strongly rely on their latency advantage. Our active traders include traders using algorithmic strategies and traders working manually at their trading desks.

[3] HFT was strongly criticized after the "flash crash" on May 6, 2010 and "gold halt" on January 6, 2014. However , HFT has been shown to increase market liquidity (Hendershott, Jones, and Menkveld (2011)) and price efficiency (Chaboud, Chiquoine, Hjalmarsson, and Vega (2014), Brogaard, Hendershott and Riordan (2014)).

[4] In this paper we consider the case of a sell transaction, but similar arguments apply to a buy transaction.

[5] The ex-post measure is based on the transaction price and is therefore backward-looking.

[6] As mentioned by Gouriéroux and Jasiak (2010), financial institutions also need intraday risk analysis for internal control of their trading desks.

ability to convert the *desired quantity* of a financial asset into cash *quickly* and with little *impact on the market price* (Demsetz (1968); Black (1971); Kyle (1985); Glosten and Harris (1988)). This definition implicitly covers four dimensions: *volume* (significant quantity), *price impact* (deviation from the best price provided in the market), *time* (speed of completing the transaction), and *resilience* (speed of backfilling). The ex-ante liquidity measure derived from open LOB can provide more information on volume, price impact, and resilience dimensions. In an automated trading system, the speed of completing the transaction varies according to the location and the capacity of the machine and software.

Regarding existing risk measures, the conventional VaR is a risk measure that has been widely used in both academic and industry sectors. However, at both low frequency and high frequency, this risk measure can be interpreted only as an estimation of the potential loss on a predetermined portfolio over a relatively long or short fixed period, namely a measure of market price risk. Thus, VaR is not a liquidation value because it does not take into account the volume dimension; it is solely a 'paper value' for a frozen portfolio. Nevertheless, liquidity risk is always present before the transaction is realized. Further, from a high-frequency market microstructure perspective, the transaction price is an outcome of information shock, trading environment, market imperfections, and the state of the LOB. For very short horizons, all these microstructure effects could cause the transaction price to deviate from the efficient price. Therefore, if we concentrate on the transaction price, the VaR will suffer from a serious omission of liquidity, especially when the liquidation quantity is large. To unwind a large position, a trader might have to move down in the LOB and accept a lower price. In other words, when a trader with a large long position for a given stock wishes to liquidate his position in a very short time horizon, his potential loss will depend on two (related) risks: market risk and liquidity risk. Market risk is assumed by all traders, whereas liquidity risk is a concern only when the trader starts liquidating his position. This is why our proposed LIVaR measure includes the additional dimension of ex-ante liquidity risk. LIVaR measures the potential maximum loss at a given confidence level due to the decrease in both market price and available liquidity in the LOB. The introduction of the ex-ante liquidity dimension can thus offer a more accurate risk measure for active traders and market regulators who aim to closely monitor total risk and compute regulatory capital.

Very few studies have focused on high-frequency risk measures. Dionne, Duchesne and Pacurar (2009) are the first to consider an ultra-high-frequency market risk measure, Intraday

Value at Risk (IVaR), based on *all* transactions. In their study, the informative content of trading frequency is taken into account by modeling the durations between two consecutive transactions. One important practical contribution of their paper is that, instead of being restricted to traditional one- or five-minute horizons, their model allows the IVaR measure to be computed for any time horizon once the model is estimated. The authors found that ignoring the effect of durations can underestimate risk. However, as they noted, similar to other VaR measures, the IVaR ignores the ex-ante liquidity dimension by taking into account only information about transaction prices.

The development of liquidity-adjusted risk measures goes back to Bangia et al. (1999), who first consider the spread between best ask and bid as a measure of liquidity risk in VaR computation. To simplify the modeling, they further assume that the liquidity risk proxied by half spread is perfectly correlated with market risk. Actually, the total risk they attempt to identify is the sum of market risk and trading cost associated with only one share, because their liquidity risk measure does not consider the volume dimension. In line with Bangia et al. (1999), Angelidis and Benos (2006) estimate the liquidity-adjusted VaR by using data from the Athens Stock Exchange. They find that liquidity risk measured by the bid-ask spread accounts for 3.4% of total market risk for high-capitalization stocks and 11% for low-capitalization stocks. Using the same framework as Bangia et al. (1999), Weiß and Supper (2013) address the liquidity risk of a portfolio formed with five NASDAQ stocks by estimating the multivariate distribution of log-return and spread using vine copulas to account for the dependence between the two variables across stocks over a regular interval of 5 minutes. They evidence strong extreme comovements in liquidity and tail dependence between bid-ask spreads and log-returns across the selected stocks.

Our study is related to that of Giot and Grammig (2006), who focus on the ex-ante liquidity risk faced by an impatient trader acting as a liquidity demander by submitting a market order. Using open LOB Xetra data sampled at regular time intervals, the authors construct an actual return that is a potential implicit return for a predetermined volume to trade over a fixed time horizon of 10 or 30 minutes. Ex-ante liquidity risk is quantified by comparing the standard VaR based on frictionless return, i.e. mid-quote return, and the liquidity-adjusted VaR inferred from the actual return.

Our study differs from previous papers in the following ways. First, we examine the ex-ante liquidity risk by focusing on the tick-by-tick frictionless returns (issued from the best bid/ask price) and actual returns (issued from the potential liquidation price) derived from open LOB. Many papers have explored ex-post liquidity (Bacidore, Ross, and Sofianos (2003); Battalio, Hatch and Jennings (2003), Goyenko, Holden and Trzcinka (2009) among others), which is already consumed by the market or marketable orders when the trades are realized. Compared with ex-post liquidity, ex-ante liquidity is more informative and relevant in that it measures the unconsumed liquidity in LOB. Further, the existing literature that analyzes ex-ante measures of liquidity (e.g., Giot and Grammig, 2006) is based on a regular time interval and thus ignores the information within the interval. In our study, we evidence that the durations between two consecutive observations are positively related to volatilities of actual returns and frictionless returns.

Second, our study addresses the questions of the relationship between ex-ante liquidity risk embedded in open LOB and market risk, and how ex-ante liquidity evolves during the trading day. One challenge of directly modeling frictionless returns and actual returns is that they are not time-additive. Therefore, in this article we model the frictionless return changes and actual return changes using an econometric model characterized by the Logarithmic Autoregressive Conditional Duration, Vector Autoregressive Moving Average and Multivariate GARCH processes (denoted by Log-ACD, VARMA and M-GARCH hereafter). The structure not only captures the joint dynamics of both frictionless return changes and actual return changes, but also quantifies the impact of ex-ante liquidity risk on total risk by further defining $IVaR^c$ and $LIVaR^c$ as the VaRs on frictionless return changes and actual return changes, respectively. To make the model more flexible, we allow for the time-varying correlation of volatility of the frictionless return and actual return.

Third, from a practical perspective, our proposed risk measure aims at providing a view of total risk and ex-ante liquidity that can help high-frequency traders develop their timing strategies during a particular trading day. Our model is first estimated on deseasonalized data and then validated on both simulated deseasonalized and re-seasonalized data. The time series of re-seasonalized data is constructed by re-introducing deterministic seasonality factors. One advantage of simulated re-seasonalized data is that risk management can be conducted in calendar time. In addition, because the model is estimated using tick-by-tick observations and takes into account the durations between two consecutive transactions, practitioners can then construct the risk measure for any desired time horizon.[7]

The rest of the paper is organized as follows: Section 1.2 describes the dataset we utilize. Section 1.3 briefly presents the procedure used to test the model and to compute an impact coefficient of ex-ante liquidity risk and an ex-ante liquidity premium. Section 1.4 defines the actual return, frictionless return, IVaR and LIVaR. In Section 1.5, we specify the econometric model used to capture the dynamics of duration, frictionless return changes, actual return changes and their correlations. Section 1.6 applies the econometric model proposed in Section 1.5 to data for three stocks and reports the estimation results. The model performance is assessed by the Unconditional Coverage test of Kupiec (1995), the Independence test proposed by Christoffersen (1998), and the new improved backtests of Ziggel et al. (2014). In Section 1.7, we analyze the ex-ante liquidity risk in $LIVaR^c$ and compare the proposed LIVaR with other high-frequency risk measures. Section 1.8 presents the results on robustness analysis. Section 1.9 concludes the paper and proposes new research directions.

1.2 Xetra dataset

The present study uses data from the automated order-driven trading system Xetra, which is operated by Deutsche Börse at the Frankfurt Stock Exchange and has a similar structure to NASDAQ's Integrated Single Book and NYSE's Super Dot. A more detailed review of Xetra can be found in the supplementary appendix. Our sample period is July 5 to July 16, 2010. We have also considered an additional sample from June 6 to June 17, 2011 and the results are presented in the appendix.

For blue-chip and other highly liquid stocks, during continuous trading, there are no dedicated market makers like the traditional NYSE specialists. Therefore, the liquidity comes from all market participants who submit limit orders in LOB. Our database includes up to 20 levels of LOB information except the hidden part of an iceberg order, which means that by observing the LOB, any trader and registered member can monitor the dynamic of liquidity supply and potential price impact caused by a market or marketable limit order. However, all the trading and order submissions are anonymous; that is, the state and the updates on LOB can be observed but there is no information on the identities of market participants.

[7] If market conditions change significantly, the model can, of course, be re-estimated.

The raw dataset that we analyzed contains all the events that are tracked and sent through the data streams. There are two main types of streams: delta and snapshot. The former tracks all the possible updates in LOB such as entry, revision, cancellation and expiration. Traders can be connected to the delta stream during trading hours to receive the latest information, whereas the snapshot provides an overview of the state of LOB and is sent after a constant time interval for a given stock. Xetra original data with delta and snapshot messages are first processed using the software Xetra Parser developed by Bilodeau (2013) to make Deutsche Börse Xetra raw data usable for academic and professional purposes. Xetra Parser reconstructs the real-time order book sequence including all the information for both auctions and continuous trading by implementing the Xetra trading protocol and Enhanced Broadcast.[8] We further convert the raw LOB information into a readable LOB for each update time and then retrieve useful and accurate information about the state of LOB and the precise timestamp for order modifications and transactions during continuous trading. Inter-trade durations and LOB update durations are irregular. The stocks that we choose for this study— SAP (SAP), RWE AG (RWE), and Merck (MRK) — are blue-chip stocks from the DAX30 index. SAP is a leading multinational software corporation with a market capitalization of 33.84 billion Euros in 2010. RWE generates and distributes electricity to various customers including municipal, industrial, commercial and residential customers. The company produces natural gas and oil, mines coal and delivers and distributes gas. In 2010, its market capitalization was around 15 billion Euros. Merck is the world's oldest operating chemical and pharmaceutical company with a market capitalization of 4 billion Euros in 2010. Plots of the daily and intradaily time evolution of price, volume, returns, and state of the LOB for each of the three stocks during our sample period can be found in the appendix.

1.3 Procedure used for computing the risk measures

To compute the proposed risk measures, the model will first be estimated using deseasonalized data, and then the tests will be carried out on both deseasonalized and seasonalized data. Figure 1.1 presents the flowchart of our methodology:

[Insert Figure 1.1 here]

We now describe the main steps:

[8] See Xetra Release 11.0 – Enhanced Broadcast Solution and Interface Specification for a detailed description.

(a) We first compute the raw tick-by-tick durations defined as the time interval between two consecutive trades and tick-by-tick frictionless returns and actual returns based on the data of open LOB and trades.

(b) We further compute the frictionless return changes and actual return changes by taking the first difference of frictionless returns and actual returns, respectively. This step is required because the frictionless return and actual return are not time-additive and cannot be modeled directly.

(c) We remove seasonality from durations, frictionless return changes and actual return changes to obtain the corresponding deseasonalized data.

(d) The deseasonalized data are modeled by a LogACD-VARMA-MGARCH model.

(e) Once we have estimated the model, we simulate the deseasonalized data based on estimated coefficients and construct VaR measures at different confidence levels for backtesting on out-of-sample deseasonalized data.

(f) The seasonal factors are re-introduced into the deseasonalized data to generate the re-seasonalized data.

(g) As done in (e), we construct different quantiles for backtesting on out-of-sample seasonalized data.

(h) We construct impact coefficients of ex-ante liquidity risk based on the simulated re-seasonalized data.

(i) We further compute the IVaR and LIVaR, which are defined as the VaR for frictionless return and actual return, respectively.

(j) Based on the IVaR and LIVaR, we can finally compute the ex-ante liquidity premium by taking the ratio of the difference between LIVaR and IVaR over LIVaR.

The following sections will explain each step in detail.

1.4 Frictionless return, actual return and the corresponding high-frequency VaRs

We take into account all trading information that is available by modeling tick-by-tick data. The first characteristic in tick-by-tick data modeling is that the durations between two consecutive transactions are irregularly spaced. Consider two consecutive trades that arrive at t_{i-1} and t_i, and define dur_i as the duration from t_{i-1} to t_i. Based on this point process, we can further construct two return processes: frictionless return and actual return. More specifically, the frictionless return is defined as the log ratio of best bid price, $b_i(1)$ at moment i and previous best ask price, $a_{i-1}(1)$. The frictionless return is an ex-ante return indicating the tick-by-tick return for selling only one unit of stock.[9]

$$R_i^F = \ln(\frac{b_i(1)}{a_{i-1}(1)}). \qquad (1.1)$$

The actual return is defined as the log ratio of selling price for a volume v and previous best ask price.[10]

$$R_i^B = \ln(\frac{b_i(v)}{a_{i-1}(1)}),$$

where

$$b_i(v) = \frac{\sum_{k=1}^{K-1} b_{k,i} v_{k,i} + b_{K,i} v_{K,i}}{v} \quad and \quad v_{K,i} = v - \sum_{k=1}^{K-1} v_{k,i}, \qquad (1.2)$$

$b_{k,i}$ *and* $v_{k,i}$ are the *k*th level bid price and volume available, respectively. $v_{K,i}$ is the quantity left after K-1 levels are completely consumed by *v*. The consideration of quantity available in LOB is in line with other ex-ante liquidity measures in the market microstructure literature (Irvine, Benston and Kandel (2000); Domowitz, Hansch and Wang (2005); Coppejans, Domowitz and Madhavan (2004), among others). The choice of volume *v* is motivated by transaction volume and volume available in LOB. Explicitly, for each stock we first compute the cumulative volume available over the 20 levels at each transaction moment for the bid side of LOB, and then we choose the minimum cumulative volume as the maximum volume to construct the actual returns. We thus avoid the situation where the actual price does not

[9] In the market microstructure literature, the mid-quote price is often used as a proxy for the unobserved efficient market price. However, from a practical perspective, traders can rarely obtain the mid-quote price during their transactions. Therefore, the use of the mid-quote price will underestimate the risk faced by active traders. We take a more realistic price, the best bid price, as the frictionless price for traders who want to liquidate their stock position.

[10] The frictionless return and actual return are defined similarly from buyers' viewpoint. This definition implicitly assumes that orders can be executed without any latency, and quote stuffing does not affect order execution.

exist for a given volume. One concern that may arise involves iceberg orders, which keep a portion of the quantity invisible to market participants. In this study, we assume that the liquidity risk is faced by an impatient trader, and the possibility of trading against an iceberg order will not influence his trading behavior. According to Beltran-Lopez, Giot and Grammig (2009), the hidden part of the book does not carry economically significant informational content.[11] The difference between the frictionless return and the actual return is that the actual return takes into account the desired transaction volume, which is essential for the liquidity measure. Intuitively, the actual return measures the ex-ante return when liquidating v units of shares.

One characteristic of our defined frictionless returns and actual returns is that they do not possess the time-additivity property of traditional log-returns. To circumvent this difficulty, we model the frictionless return changes and actual return changes instead of modeling the actual return and frictionless return directly. More specifically, let $r_i^f = R_i^F - R_{i-1}^F$ and $r_i^b = R_i^B - R_{i-1}^B$ be the tick-by-tick frictionless return changes and actual return changes. Following this setup, the L-step forward frictionless return and actual return can be expressed as follows:

$$R_{i+L}^F = \ln(\frac{b_{i+L}(1)}{a_{i+L-1}(1)}) = \ln(\frac{b_i(1)}{a_{i-1}(1)}) + r_{i+1}^f + r_{i+2}^f + ... r_{i+L}^f = R_i^F + \sum_{m=1}^{L} r_{i+m}^f , \tag{1.3}$$

$$R_{i+L}^B = \ln(\frac{b_{i+L}(v)}{a_{i+L-1}(1)}) = \ln(\frac{b_i(v)}{a_{i-1}(1)}) + r_{i+1}^b + r_{i+2}^b + ... r_{i+L}^b = R_i^B + \sum_{m=1}^{L} r_{i+m}^b , \tag{1.4}$$

The terms $\sum_{m=1}^{L} r_{i+m}^f$ and $\sum_{m=1}^{L} r_{i+m}^b$ are the sum of all tick-by-tick changes in return over a predetermined interval and can be considered as the waiting cost related to frictionless returns and actual returns.[12] More specifically, they measure the costs/gains associated with the latter instead of immediate liquidation of one share and v shares of stock, respectively. Moreover, if we define the IVaR[c] and LIVaR[c] as the VaR for $\sum_{m=1}^{L} r_{i+m}^f$ and $\sum_{m=1}^{L} r_{i+m}^b$, respectively, and compute them for a predetermined time interval, then the IVaR[c] and LIVaR[c] will provide the maximal loss over a given interval and at a given confidence level for an investor that trades

[11] In our dataset, the hidden order information is not available. However, from the transaction data and open LOB, we deduced that the proportion of transactions that hit hidden orders is less than 1% for the three stocks.

[12] The waiting cost could be positive or negative, which indicates loss and gain, respectively.

at frictionless or actual return. In other words, the IVaR^c and LIVaR^c will estimate the maximal loss in terms of frictionless return and actual return, which are related to market risk and total risk (market risk and ex-ante liquidity risk), respectively. Mathematically, consider a realization of a sequence of intervals with length *int* and let $y^f_{int,t}$ *and* $y^b_{int,t}$ [13] be the sum of tick-by-tick changes of returns r^f_i *and* r^b_i over the *t*-th interval,

$$y^f_{int,t} = \sum_{j=\tau(t-1)}^{\tau(t)-1} r^f_j \ , \ \ y^b_{int,t} = \sum_{j=\tau(t-1)}^{\tau(t)-1} r^b_j \ , \tag{1.5}$$

where $\tau(t)$ is the index for which the cumulative duration exceeds the *t*-th interval with length *int* for the first time. By definition:

$$int \geq \sum_{j=\tau(t-1)}^{\tau(t)-1} dur_j \ \ \text{and} \ \ int \leq \sum_{j=\tau(t-1)}^{\tau(t)} dur_j . \tag{1.6}$$

The process of duration allows us to aggregate the tick-by-tick data to construct the dynamics of frictionless return and actual return for a predetermined interval that allows consideration of risk in calendar time. Accordingly, the IVaR^c and LIVaR^c [14] for frictionless return changes and actual return changes with confidence level $1-\alpha$ for a predetermined interval *int* are defined as:

$$Pr(y^f_{int,t} < IVaR^c_{int,t}(\alpha) \,|\, I_t) = \alpha \ ;$$

$$Pr(y^b_{int,t} < LIVaR^c_{int,v,t}(\alpha) \,|\, I_t) = \alpha ,$$

I_t is the information set until moment $\tau(t-1)$. Similar to the traditional definition of VaR, $IVaR^c_{int,t}(\alpha)$ and $LIVaR^c_{int,v,t}(\alpha)$ are the conditional α-quantiles for $y^f_{int,t}$ and $y^b_{int,t}$.

We can further define the IVaR and LIVaR as the VaR for the frictionless return and actual return as follows: $IVaR_{int,t} = R^F_{\tau(t-1)} + IVaR^c_{int,t}$,

[13] y^f_{int} *and* y^b_{int} are defined on both seasonalized return changes and deseasonalized return changes.

[14] IVaR^c and LIVaR^c are also defined on both seasonalized return changes and deseasonalized return changes. Moreover, our IVaR^c and LIVaR^c can also be used in the strategy of short selling where IVaR^c and LIVaR^c will be the 1-α quantiles of the distributions.

$$LIVaR_{int,v,t} = R^{B}_{\tau(t-1)} + LIVaR^{c}_{int,v,t} \;\;,$$

where $R^{F}_{\tau(t-1)}$ and $R^{B}_{\tau(t-1)}$ are the frictionless return and actual return at the beginning of the t-th interval. Consequently, IVaR and LIVaR estimate the α-quantiles for frictionless return and actual return at the end of the t-th interval.

1.5 Methodology

In our tick-by-tick modeling, there are three random processes: duration, changes in the frictionless return, and changes in the actual return. The present study assumes that the duration evolution is strongly exogenous but has an impact on the volatility of frictionless and actual return changes. The joint distribution of duration, frictionless return change and actual return change can be decomposed into the marginal distribution of duration and joint distribution of frictionless and actual return change conditional on duration. More specifically, the joint distribution of the three variables is:

$$\begin{aligned} & f^{d,f,b}(dur_i, r_i^f, r_i^b \,|\, dur_{i-1}, r_{i-1}^f, r_{i-1}^b; \Theta^f) \\ &= f^d(dur_i \,|\, dur_{i-1}, r_{i-1}^f, r_{i-1}^b; \theta^d)\, f^{f,b}(r_i^f, r_i^b \,|\, dur_i, dur_{i-1}, r_{i-1}^f, r_{i-1}^b; \theta^d, \theta^{f,b}) \;\;, \end{aligned} \tag{1.7}$$

where $f^{d,f,b}$ is the joint distribution for duration, frictionless return change and actual return change. $f^d(\cdot)$ is the marginal density for duration and $f^{f,b}(\cdot)$ is the joint density for actual and frictionless return changes. Consequently, the corresponding log-likelihood function for each joint distribution can be written as:

$$L(\theta^d, \theta^{f,b}) = \sum_{i=1}^{n} \log f^d(dur_i \,|\, dur_{i-1}, r_{i-1}^f, r_{i-1}^b; \theta^d) + \log f^{f,b}(r_i^f, r_i^b \,|\, dur_{i-1}, r_{i-1}^f, r_{i-1}^b, dur_i; \theta^{f,b}), \tag{1.8}$$

In the next subsections, we specify marginal density for dynamics of duration and joint density for frictionless return and actual return changes. We present the model for deseasonalized duration, frictionless and actual return changes. The deseasonalization procedure is described in detail in Section 6.

1.5.1. Model for duration

The ACD model used to model the duration between two consecutive transactions was introduced by Engle and Russell (1998). The GARCH-style structure is used to capture the duration clustering observed in high-frequency financial data. The basic assumption is that the realized duration is driven by its conditional duration and a positive random variable as an error term. Let $\psi_i = E(dur_i | I_{i-1})$ be the expected duration given all the information up to *i-1*, and ε_i be the positive random variable. The duration can be expressed as: $dur_i = \psi_i \cdot \varepsilon_i$. There are several possible specifications for the expected duration and the independent and identically distributed (i.i.d.) positive random error (see Hautsch (2004) and Pacurar (2008) for surveys). To guarantee the positivity of duration, we adopt the log-ACD model proposed by Bauwens and Giot (2000). The specification for expected duration is

$$\psi_i = \exp\left(\omega + \sum_{j=1}^{p} \alpha_j \varepsilon_{i-j} + \sum_{j=1}^{q} \beta_j \ln \psi_{i-j}\right). \tag{1.9}$$

For positive random errors, we use the generalized gamma distribution, which allows a non-monotonic hazard function and nests the Weibull distribution (Grammig and Maurer (2000); Zhang, Russell and Tsay (2001)):

$$p(\varepsilon_i | \gamma_1, \gamma_2) = \begin{cases} \dfrac{\gamma_1 \varepsilon_i^{\gamma_1\gamma_2 - 1}}{\gamma_3^{\gamma_1\gamma_2} \Gamma(\gamma_2)} \exp[-\left(\dfrac{\varepsilon_i}{\gamma_3}\right)^{\gamma_1}], & \varepsilon_i > 0, \\ 0 \quad , & \text{otherwise.} \end{cases} \tag{1.10}$$

where $\gamma_1, \gamma_2 > 0$, $\Gamma(.)$ is the gamma function, and $\gamma_3 = \Gamma(\gamma_2) / \Gamma(\gamma_2 + \frac{1}{\gamma_1})$.

1.5.2. Model for frictionless return and actual return changes

The high-frequency frictionless return and actual return changes display a high serial correlation. To capture this microstructure effect, we follow Ghysels and Jasiak (1998) and adopt a VARMA (*p,q*) structure:

$$R_i = \left(r_i^b, r_i^f\right)', \; E_\text{i} = \left(e_i^b, e_i^f\right)',$$

$$R_i = \sum_{m=1}^{p} \Phi_m R_{i-m} + E_\text{i} - \sum_{n=1}^{q} \Delta_n E_{i-n} \;, \tag{1.11}$$

where Φ_m and Δ_n are matrices of coefficients for R_{i-m} and E_{i-n}, respectively. As mentioned in Dufour and Pelletier (2011), we cannot directly work with the representation in (11) because of an identification problem. Consequently, we impose the restrictions on Δ_n by supposing that the VARMA representation is in diagonal MA form. More specifically, $\Delta_n = \begin{pmatrix} \delta_{11}^{(n)} & 0 \\ 0 & \delta_{22}^{(n)} \end{pmatrix}$ where $\delta_{11}^{(n)}$ and $\delta_{22}^{(n)}$ are the coefficients for nth-lag error terms of the actual return change and frictionless return change, respectively.

Furthermore, we assume the volatility part follows a multivariate GARCH process:

$$\begin{pmatrix} e_i^b \\ e_i^f \end{pmatrix} = H_i^{1/2},$$

where z_i is the bivariate normal distribution that has the following two moments: $E(z_i) = 0, Var(z_i) = I_2$. The normality assumption for the error term is supported by backtesting. The other distributions we tried, such as Normal Inverse Gaussian (NIG), Student-t and Johnson, overestimated the error distribution in our simulation tests.

H_i is the conditional variance matrix for R_i that should be positive-definite. To model the dynamic of H_i , we use the DCC structure proposed by Engle (2002) in which H_i is decomposed as follows:

$$H_i = D_i C_i D_i,$$

$$\text{where:}$$

$$\begin{cases} D_i = diag(\sqrt{h_{11i}}, \sqrt{h_{22i}}), \quad h_{11i} = \sigma_{11i}^2 dur_i^{\gamma_b}, h_{22i} = \sigma_{22i}^2 dur_i^{\gamma_f}, \\ \\ C_i = (diagQ_i)^{-1/2} Q_i (diagQ_i)^{-1/2}, \end{cases}$$

and

$$Q_i = (1-\theta_1-\theta_2)Q + \theta_1 \varepsilon_{i-1}\varepsilon_{i-1}^{'} + \theta_2 Q_{i-1}, \quad \varepsilon_{k,i-1} = \frac{e_{k,i-1}}{\sqrt{h_{kk,i-1}}}, \ k = 1,2 \ , \tag{1.12}$$

Q_i is the unconditional correlation matrix of $\{\varepsilon_i\}_{i=1}^{N}$, $\sqrt{h_{11i}}$ ($\sqrt{h_{22i}}$) is the conditional variance for actual (frictionless) return change, and γ_b (γ_f) measures the impact of duration on the volatilities of actual and frictionless return changes, respectively.[15]

In the DCC framework, each series has its own conditional variance. For both actual and frictionless return changes, we adopt a NGARCH(*m*,*n*) (Nonlinear GARCH) process as proposed by Engle and Ng (1993) to capture the cluster as well as the asymmetry in volatility. The process can be written as:

$$h_{11,i} = \sigma_{11,i}^2 \cdot dur_i^{\gamma_b},$$
$$\sigma_{11,i}^2 = \omega^b + \sum_{j=1}^{m} \alpha_j^b \sigma_{11,i-m}^2 + \sum_{j=1}^{n} \beta_j^b \sigma_{11,i-j}^2 (\varepsilon_{i-1}^b - \pi^b)^2, \tag{1.13}$$

and

$$h_{22,i} = \sigma_{22,i}^2 \cdot dur_i^{\gamma_f},$$
$$\sigma_{22,i}^2 = \omega^f + \sum_{j=1}^{m} \alpha_j^f \sigma_{22,i-m}^2 + \sum_{j=1}^{n} \beta_j^f \sigma_{22,i-j}^2 (\varepsilon_{i-j}^f - \pi^f)^2, \tag{1.14}$$

where π^b and π^f are used to capture the asymmetry in the conditional volatilities. When $\pi^f \ or \ \pi^b = 0$, the model will become a standard GARCH model, whereas a negative $\pi^f \ or \ \pi^b$ indicates that a negative shock will cause higher conditional volatilities for the next moment.

Our structure also explicitly introduces the duration dimension in conditional volatilities. In his pioneering study of the impact of duration on volatility, Engle (2000) assumes that the impact is linear; that is, $h_{11,i} = \sigma_{11,i}^2 \cdot dur_i$. This modeling for the unit of time might be restrictive for some empirical data for which conditional volatility can depend on duration in a more complicated way. To make the model more general, we follow Dionne, Duchesne and Pacurar (2009) by assuming the exponential form $h_{11,i} = \sigma_{11,i}^2 \cdot dur_i^{\gamma_b}$ and $h_{22,i} = \sigma_{22,i}^2 \cdot dur_i^{\gamma_f}$.

[15]The DCC model can be estimated by a two-step approach. Engle and Sheppard (2001) show that the likelihood of the DCC model can be written as the sum of two parts: a mean and volatility part, and a correlation part. Even though the estimators from the two-step estimation are not fully efficient, the one iteration of a Newton-Raphson algorithm applied to total likelihood provides asymptotically efficient estimators.

When γ_f or $\gamma_b = 0$, the volatility will become a standard NGARCH process, whereas when γ_f or $\gamma_b = 1$, it transforms to the similar model studied in Engle (2000).

As mentioned above, our model has three uncertainties: duration uncertainty, price uncertainty and LOB uncertainty. Deriving a closed form of $LIVaR^c$ would be complicated for multi-period forecasting in the presence of three risks, especially for non-regular time duration. Therefore, once the models are estimated, we follow Christoffersen (2003) and use Monte Carlo simulations to make multi-step forecasting and to test the model's performance.

1.6 Empirical Results

1.6.1. Seasonality adjustment

It is well known that high-frequency data behave very differently from low-frequency data. Table 1.1, 1.2 and 1.3 present the descriptive statistics of raw and deseasonalized duration, frictionless return changes and actual return changes for which various volumes are chosen for the three studied stocks. From Panel A, we can observe that for the sample period (the first two weeks of July 2010), SAP is the most liquid stock: the average duration is the shortest and the number of observations is the largest. MRK is the least liquid one. Moreover, given that the variables are constructed on tick-by-tick frequency, all three stocks have an average of zero and a very small standard deviation for frictionless return changes and actual return changes. All three stocks present high kurtosis due to the fact that most of the observations are concentrated on their average and co-exist with some extreme values. In addition, the raw data are characterized by extremely high autocorrelation for both first and second moments for all of the variables.

High-frequency data are characterized by seasonality, which should be removed before estimating any model. To do so, several approaches have been proposed in the literature: Andersen and Bollerslev (1997) use the Fourier Flexible Functional (FFF) form to take off seasonality, Dufour and Engle (2000) remove seasonality by applying a simple linear regression with a dummy, and Bauwens and Giot (2000) take off seasonality by averaging over a moving window and linear interpolation.

[Insert Table 1.1 to 1.3 here]

However, as found in Anatolyev and Shakin (2007) and Dionne, Duchesne and Pacurar (2009), the high-frequency data could behave differently throughout the day as well as between different trading days. Therefore, to fully account for the deterministic part of the data, we apply a two-step deseasonalization procedure, interday and intraday. Further, an open auction effect in our continuous trading dataset exists that is similar to the one found by Engle and Russell (1998). More precisely, for each trading day, continuous trading follows the open auction in which an open price is set according to certain criteria such as maximization of the volume. Once the open auction is finished, the transactions are recorded. Consequently, the beginning of continuous trading is contaminated by extremely short durations. These short durations could produce negative seasonality factors of duration that are based on previous observations and cubic splines. To address this problem, the data for the first half hour of each trading day are only used to compute the seasonality factor and then discarded.

The interday trend is extracted under a multiplicative form:

$$dur_{i,inter} = \frac{dur_{i,s}}{\overline{dur_s}}, \; r^f_{i,inter} = \frac{r^f_{i,s}}{\sqrt{\overline{(r^f_s)^2}}}, \; r^b_{i,inter} = \frac{r^b_{i,s}}{\sqrt{\overline{(r^b_s)^2}}}, \tag{1.15}$$

where $dur_{i,s}$, $r^f_{i,s}$, $r^b_{i,s}$ are the *i*th duration, frictionless and actual return change for day s, respectively and $\overline{dur_s}$, $\overline{(r^f_s)^2}$, $\overline{(r^b_s)^2}$ are the daily average for day *s* for duration, squared frictionless, and actual return changes, respectively.

Based on interday deseasonalized data, the intra-day seasonality is removed by following Engle and Russell (1998):

$$dur_{i,intra} = \frac{dur_{i,inter}}{\sqrt{E(dur_{i,inter})|I_{i-1}}}, \; r^f_{i,intra} = \frac{r^f_{i,inter}}{\sqrt{E((r^f_{i,inter})^2)|I_{i-1}}}, r^b_{i,intra} = \frac{r^b_{i,inter}}{\sqrt{E((r^b_{i,inter})^2)|I_{i-1}}}, \tag{1.16}$$

where $E(dur_{i,inter})$, $E((r^f_{i,inter}{}^2)$ and $E((r^b_{i,inter})^2)$ are the corresponding deseasonality factors constructed by averaging the variables over 30-minute intervals for each day of the week and then applying cubic splines to smooth these 30-minute averages. The same day of week shares the same intra-deseasonality curve. However, it takes different deseasonality factors according to the moment of transaction. Figure 1.2 illustrates the evolution of the seasonality factors of RWE for duration, frictionless, and actual return changes when *v* = 4000.[16] It is not

[16] Results on other volumes for RWE, SAP, and MRK are available upon request.

surprising to see that the frictionless and actual return changes have similar dynamics because the actual return changes contain the frictionless return changes. However, the magnitude for frictionless and actual return changes differs. Panels B of Tables 1.1, 1.2 and 1.3 report descriptive statistics of deseasonalized durations, frictionless and actual return changes. The raw frictionless and actual return changes have been normalized to bring the mean to zero and standard deviation to one. However, other statistics such as skewness, kurtosis and auto-correlation are not affected by this normalization process. The high kurtosis and auto-correlation will be captured by the proposed models.

[Insert Figure 1.2 here]

1.6.2. Estimation results

We use the model presented in Section 5 to fit SAP, RWE and MRK deseasonalized data. The estimation data cover the first week of July 2010. The data from the second week are used as out-of-sample data to test the model's performance. For robustness, we apply the same procedure to the same stocks for first two weeks of June 2011. As previously mentioned, the estimation is realized jointly for frictionless and actual return changes. The likelihood function is maximized using Matlab v7.6.0 with Optimization toolbox.

Tables 1.4, 1.5 and 1.6 report the estimation results for actual return changes for SAP, RWE and MRK for v = 4000, 4000, and 1800 shares, respectively. It should be noted that for each stock, the frictionless return changes and actual return changes are governed by the same duration process, which is assumed to be strictly exogenous. The high clustering phenomenon is indicated in deseasonalized data by the Ljung-Box statistic (see Table 1.1, 1.2 and 1.3, Panel B). The clustering in duration is confirmed by the Log-ACD model. To better fit the data, we retain a Log-ACD (2,1) specification for SAP durations, a Log-ACD(3,1) model for RWE durations and a Log-ACD(1,1) model for MRK durations. The Ljung-Box statistic on standardized residuals of duration provides evidence that the Log-ACD model is capable of removing the high autocorrelation identified in deseasonalized duration data. The Ljung-Box statistic with 15 lags is dramatically reduced to 25.08 for SAP, 39.7 for RWE and 21.79 for MRK.

Frictionless and actual return changes of the three stocks are also characterized by a high autocorrelation in level and volatility. Moreover, the Ljung-Box statistics with 15 lags on deseasonalized return change and its volatility reject independence at any significance level for the three stocks. Taking the model efficiency and parsimony into consideration, a VARMA(4,2)-MGARCH((1,3),(1,3)) [17] model is retained for SAP, a VARMA(5,1)-MGARCH((1,3),(1,3)) model for RWE, and the specification of VARMA(2,2)-MGARCH((1,3),(1,3)) for MRK. The model adequacy is assessed based on standardized residuals and squared standardized residuals. Taking MRK as an example, the Ljung-Box statistics for standardized residuals and squared standardized residuals of actual return changes, computed with 5, 10, 15, 20 lags, respectively, are not significant at the 5% level. The Ljung-Box statistic with 15 lags has been significantly reduced after modeling to 7.72. Similar results are obtained for stocks RWE and SAP.

Regarding the estimated parameters, the sum of coefficients in each individual GARCH model is close to one, indicating a high persistence in volatility. Further, π^f and π^b are significantly different from zero in both structures, thus evidencing asymmetric effects. In other words, a negative shock generates a higher conditional volatility for the next moment. It also should be noted that γ_b and γ_f are both positive for the three stocks. This means that a longer duration will generate higher volatility for both actual return and frictionless return changes. In addition, due to the fractional exponent, volatility increases slowly as duration lengthens. In our model, volatility is the product of no-duration scaled variance and duration factor, and the γ_b of actual return changes are higher than γ_f of frictionless return changes for the three stocks. This implies that the duration factor has a larger impact on actual returns than on frictionless returns.

The use of dynamic conditional correlation is justified by the fact that θ_1 and θ_2 in equation (12) are both significantly different from zero for the three stocks. As expected, the conditional correlation of actual return and frictionless return changes is time-varying. A sum of the two parameters of around 0.8 confirms the high persistence of conditional correlation.

[Insert Table 1.4 to 1.6 here]

1.6.3. Model performance and backtesting

In this section, we present the simulation procedure and backtesting results on simulated deseasonalized and re-seasonalized frictionless and actual return changes. Once the model is estimated on tick-by-tick frequency, we can test the model performance and compute

[17] NGARCH (1,3) for actual return changes and NGARCH(1,3) for frictionless return changes.

frictionless $IVaR^c$ and $LIVaR^c$ by Monte Carlo simulation. One of the advantages of our method is that once the model is estimated, we can compute the simulated deseasonalized $IVaR^c$ and $LIVaR^c$ for any horizon without re-estimating the model. In addition, we can compute the simulation-based re-seasonalized $IVaR^c$ and $LIVaR^c$ in traditional calendar time using the available seasonal factors.

We choose different time intervals to test the model performance. The interval lengths are 40, 50, 60, 80, 100, 120, and 140 units of time for the more liquid stocks SAP and RWE and 20, 30, 40, 50, 60, 80, and 100 for the less liquid stock MRK. Given that the model is applied to deseasonalized data, the simulated duration is not in calendar units. However, simulated duration and calendar time intervals are related in a proportional way. Further, depending on the trading intensity, the simulated duration does not correspond to the same calendar time interval. For a more liquid stock, the same simulated interval relates to a shorter calendar time interval. For instance, in the case of MRK, the interval length 50 re-samples the one-week data for 190 intervals and corresponds to 13.42 minutes, and the interval-length relates 100 to 95 intervals and corresponds to 26.84 minutes. However, for a more liquid stock such as SAP, the interval-length 50 corresponds to 5.45 minutes and the interval-length 140 corresponds to 15.27 minutes.

The simulations for frictionless and actual return changes are realized as follows:

1) We generate the duration between two consecutive transactions and assume that the duration process is strongly exogenous.

2) With the simulated duration and estimated coefficients of the VARMA-MGARCH model, we obtain the corresponding return changes.

3) We repeat steps 1 and 2 for 10,000 paths and re-sample the data at each path according to the predetermined interval.

4) For each interval, we compute the corresponding $IVaR^c$ and $LIVaR^c$ at the desired level of confidence. To conduct the backtesting for each given interval, we also need to construct the return changes for original out-of-sample data.

To validate the model, we conduct the Unconditional Coverage and Independence tests by applying the Kupiec test (1995), the Christoffersen test (1998), and the recently proposed tests of Ziggel et al. (2014). The Kupiec test checks whether the empirical failure rate is statistically different from the failure rate we are testing, whereas the Christoffersen test evaluates the independence aspect of the violations. More specifically, it rejects VaR models that generate clustered violations by estimating a first-order Markov chain model on the sequence. The new set of tests proposed by Ziggel et al. (2014) are based on simulation with i.i.d Bernoulli random variables. One advantage is that the new test allow for two-sided testing. Table 1.7 reports the p-values for all tests (two-sided testing for UC and i.i.d) on simulated data for confidence levels of 95%, 97.5%, 99% and 99.5%. The time interval varies from 5 minutes to 15 minutes for SAP, from 5 minutes to 16.67 minutes for RWE and from 5 minutes to 27 minutes for MRK. Most of the p-values are higher than 5%, indicating that the model generally captures the distribution of frictionless and actual return changes well.

[Insert Table 1.7 here]

Because most trading and risk management decisions are based on calendar time and raw data, it might be difficult for practitioners to use simulated deseasonalized data to conduct risk management. To this end, we conduct another Monte Carlo simulation that takes into account the time-varying deterministic seasonality factors. The process is similar to that used for simulating deseasonalized data. However, the difference is that we re-introduce the seasonality factors for duration, actual return changes and frictionless return changes. Given that seasonality factors vary from one day to another, the simulation should take the day of week into account. More precisely, for the first day, simulated durations are converted to a calendar time of that day and the corresponding timestamp identifies seasonality factors for actual and frictionless return changes. The simulation process continues until the corresponding timestamp surpasses the closing time for the underlying day. In the case of a multiple-day simulation, the process continues for another day. Based on the simulated re-seasonalized data, we also compute our $IVaR^c$ and $LIVaR^c$ by repeating the same algorithm.

Table 1.8 presents the backtesting results on the re-seasonalized simulated data. The time interval varies from 5 minutes to 10 minutes for the three stocks, and the confidence levels to test are 95%, 97.5%, 99%, and 99.5%. Similar to the test results for simulated deseasonalized data, the p-values suggest that the simulated re-seasonalized data also provides reliable high-frequency risk measures for all chosen confidence levels over intervals of 5 to 10 minutes.

[Insert Table 1. 8 here]

1.7 Risks for Waiting Cost, Ex-ante Liquidity Risk and Various IVaRs

1.7.1. Risks for waiting cost

As shown in Section 4, the sums of tick-by-tick frictionless return changes and actual return changes over a given interval can be viewed as the waiting costs related to market risk and total risk, which contains market risk and ex-ante liquidity risk. Consequently, the corresponding $IVaR^c$ and $LIVaR^c$ estimate the risk of losses on these waiting costs. Based on simulated re-seasonalized data from the previous section that contain the determinist (seasonal factor) and random (error term) elements, we can further investigate the effect of the ex-ante liquidity risk embedded in the open LOB on the total risk. To this end, we define an impact coefficient of ex-ante liquidity risk[18]:

$$\Gamma_{int,v} = \frac{\widehat{LIVaR^c_{int,v}} - \widehat{IVaR^c_{int}}}{\widehat{LIVaR^c_{int,v}}}. \tag{1.17}$$

As mentioned above, $IVaR^c_{int,t}$ and $LIVaR^c_{int,v,t}$ are the VaRs for frictionless return changes and actual return changes of volume v for the t-th interval. As we simulate the data in the tick-by-tick framework, we can compute the $IVaR^c_{int,t}$ and $LIVaR^c_{int,v,t}$ for any desired interval. Accordingly, $\widehat{IVaR^c_{int}}$ and $\widehat{LIVaR^c_{int,v}}$ are the averages of $IVaR^c_{int,t}$ and $LIVaR^c_{int,v,t}$. As a result, $\Gamma_{int,v}$ assesses, on average, the impact of the ex-ante liquidity risk of volume v on total risk for a given interval. Figure 1.3 shows how the impact coefficients of ex-ante liquidity perform for intervals from 3 minutes to 10 minutes for the three stocks.

[Insert Figure 1.3 here]

There are two interesting points to mention after observing the plots. First, the curve is globally increasing; that is, the impact coefficient of the ex-ante liquidity most often increases when the interval increases and finally converges to its long-run level. The relation of frictionless return changes and actual return changes can be explicitly expressed by $r_i^b = r_i^f + r_i^{LOB}$, where r_i^{LOB} is the volume-dependent LOB return change for the i-th transaction. Econometrically, given that the sum of two ARMA structures is also an ARMA structure (Engel (1984)), the difference of the frictionless return changes and the actual return

[18] The risk measures in equation (17) are computed for confidence level 95%; in practice, other confidence levels can also be used.

changes implicitly follows another ARMA structure. Accordingly, the relationship of $IVaR^{c}_{int,t}$ and $LIVaR^{c}_{int,v,t}$ can be generally written as:

$$LIVaR^{c}_{int,v,t} = IVaR^{c}_{\mathrm{int},t} + LOBIVaR^{c}_{int,v,t} + Dep^{F,LOB}_{int,v,t}. \tag{1.18}$$

$LOBIVaR^{c}_{int,v,t}$ measures the risk associated with the open LOB, and $Dep^{F,LOB}_{int,v,t}$ presents the dependence between the frictionless return changes and the actual return changes, which can stand for various dependence measures. However, in our specific modeling, $Dep^{F,LOB}_{int,v,t}$ is the covariance between r_i^{f} and r_i^{LOB}. Therefore, the numerator of equation (17) is the sum of $LOBIVaR^{c}_{int,v,t}$ and $Dep^{F,LOB}_{int,v,t}$. The fact that the curve is globally increasing in time is due to the higher autocorrelation in LOB return changes over time, which are the changes of magnitude in LOB caused by a given ex-ante volume.

The convergence means that, for the long run, the sum of $LOBIVaR^{c}_{int,v,t}$ and $Dep^{F,LOB}_{int,v,t}$ is proportional to $LIVaR^{c}_{int,v,t}$. Recall that in a general GARCH framework, the forward multi-step volatility converges to its unconditional level. In the present study, once the intervals include sufficient ticks for which the volatilities of both LOB return changes and frictionless return changes reach their unconditional levels, the volatilities of the sum of both return changes increase at the same speed. The impact coefficient of the ex-ante liquidity therefore converges to its asymptotic level.

Second, it is interesting to observe that the impact coefficients of ex-ante liquidity of the RWE stock are negative for a volume of 1,000 shares and become positive when volumes are 2,000, 3,000, and 4,000 shares. A negative impact coefficient of ex-ante liquidity indicates that volatility for the actual return changes is less than that of the frictionless return changes. In other words, the ex-ante liquidity risk embedded in LOB offsets the market risk. This again results from the fact that off-best levels of LOB are more stable than the first level. Based on equation (18), when the volume is small, the negative correlation between frictionless return change and LOB return change plays a more important role in determining the sign of the impact coefficients of ex-ante liquidity. However, for a higher ex-ante volume, the risk of LOB return change also increases but at a faster rate than its interaction with frictionless return change. Consequently, the impact coefficients of ex-ante liquidity become positive.

1.7.2. High-frequency ex-ante liquidity premium

Based on the tick-by-tick simulation, we can also compute the high-frequency IVaR and LIVaR for frictionless return and actual return. Using IVaR and LIVaR of the same stock, we can further define a Relative Liquidity Risk Premium as follows:

$$\Lambda_{int,v,t} = \frac{LIVaR_{int,v,t} - IVaR_{int,t}}{LIVaR_{int,v,t}}. \tag{1.19}$$

Similar to the liquidity ratio proposed in Giot and Grammig (2006), $IVaR_{int,t}$ and $LIVaR_{int,v,t}$ are the VaR measures for frictionless return and actual return at the end of the *t*-th interval and *int* is the predetermined interval such as 5-min, and 10-min. Unlike the frictionless return changes and actual return changes, our defined actual return and frictionless return do not have the time-additivity property. Even though the VaR based on frictionless return changes and actual return changes can be used directly in practice, in some situations practitioners might want to predict their potential loss on frictionless return or actual return instead of frictionless and actual return changes for a precise calendar time point.

To illustrate how our model can be used to provide the ex-ante risk measure for frictionless return and actual return, we first compute the return changes for frictionless returns and actual returns, then calculate the instantaneous frictionless return and actual return at the beginning of the given interval using equations (3) and (4). Once we know the frictionless return and actual return at the beginning of a given interval, we can obtain the frictionless return and actual return for the end of the interval. Figure 1.4 illustrates how the frictionless return and actual return at the end of an interval are computed. Figure 1.5 tracks the evolution of IVaR and LIVaR associated with a large liquidation volume for SAP, RWE, and MRK during one out-of-sample day, July 12, 2010. For the three stocks, the IVaR and LIVaR both present an inverted U shape during the trading day. However, for the more liquid stock SAP, the IVaR and LIVaR are less volatile than those of the less liquid stocks RWE and MRK. It also seems that the total risk is smaller during the middle of the day. Nonetheless, the smaller VaR in absolute terms does not mean we should necessarily trade at that moment. The IVaR and LIVaR only provide the estimates of potential loss for a given probability at a precise point in time.

[Insert Figure 1.4 here]

[Insert Figure 1.5 here]

In addition, the difference between the curves on each graph, which measures the risk associated with ex-ante liquidity, varies with time. This is due to the fact that LOB interacts with trades and changes during the trading days. A smaller (bigger) difference indicates a deeper (shallower) LOB. More specifically, for the least liquid stock, MRK, the ex-ante liquidity risk is more pronounced even for a relatively smaller quantity of 1800 shares. Regarding the more liquid stocks such as SAP and RWE, the ex-ante liquidity risk premiums are much smaller even for the relatively larger quantities of 4,000 shares. This again suggests that the ex-ante liquidity risk heightens when the liquidation quantity is large and the stock is less liquid.

Table 1.9 presents the average relative ex-ante liquidity risk premium given various order sizes at different confidence levels for SAP, RWE, and MRK. The results are based on our one-week simulations. For each stock, we consider order sizes ranging from the mean to the 99th quantile of historical transaction volume. As shown in Table 1.9, the proportion of liquidity risk increases with volume for the three stocks, as expected. At the 95% confidence level, for the average transaction volumes the liquidity risk accounts for only 3.46%, 2.88% and 1.03% of total risk for SAP, RWE, and MRK, respectively, meaning that the LOB seems efficient for providing liquidity for relatively small transaction volumes. However, for large transaction volumes such as the 99th quantile, liquidity risk can account for up to 32%, 25.26%, and 31.88% of total risk at 95% confidence level.

[Insert Table 1.9 here]

1.7.3. Comparison of LIVaR and other intraday VaRs

Our proposed IVaR and LIVaR, which are validated by backtesting, allow us to further analyze ex-ante liquidity and compare the two risk measures with other high-frequency risk measures found in the literature.

The standard IVaR proposed by Dionne, Duchesne and Pacurar (2009) is based on a transaction price that is similar to the closing price in daily VaR computation. However, the resulting IVaR serves as a measure of potential loss of ‘paper value’ for a frozen portfolio and omits the ex-ante liquidity dimension. To some extent, the IVaR accounts for an ex-post liquidity dimension; more specifically, it measures the liquidity already consumed by the market. However, active traders are more concerned with ex-ante liquidity because it is related to their liquidation value. For any trader, the risk related to liquidity is always present and the omission of this liquidity dimension can cause a serious distortion from the observed transaction price, especially when the liquidation volume is large.

Another major difference is that before obtaining LIVaR, we should compute LIVaR^c for actual return changes, which gives the potential loss in terms of waiting costs over a predetermined interval. Accordingly, the resulting LIVaR provides a risk measure for actual return at a given point in time, while the standard IVaR is based on tick-by-tick log-returns, which have a time-additive property. It thus directly gives a risk measure in terms of price for a given interval.

The high-frequency VaR proposed by Giot and Grammig (2006) is constructed on mid-quote price and ex-ante liquidation price over an interval of 10 or 30 minutes. As mentioned in Section 3, because the use of mid-quote might underestimate the risk faced by active traders, we take best bid price as the frictionless price when liquidating a position. Figure 1.6 illustrates the difference in constructing the frictionless returns and actual returns.

[Insert Figure 1.6 here]

Consequently, for active day-traders, our LIVaR can be considered an upper bound of risk measure that provides the maximum p-th quantile in absolute value when liquidating a given volume v. Further, we found that durations impact volatilities of both frictionless and actual returns (with a higher impact on actual returns than on frictionless returns), and therefore should not be ignored.

1.8 Robustness analysis

1.8.1 With data from another period

Regarding to the robustness analysis of our model, we re-applied our model to the data for another period. It should be noted that the model structure used to estimate remain the same, that is, the LogACD-VARMA-MGARCH model; however, the seasonality and the number of lags for three parts of the model can change depending on the characteristics of the trading data and LOB structure. In appendix, the results of average relative ex-ante liquidity risk premium for June 2011 are also presented. As expected, due to different characteristics in trading and LOB structure, the results are slightly different. For example, the LOB of SAP in June 2011 is less deep than that in July 2010, as a result, the average relative ex-ante liquidity risk premium slightly increases.

1.8.2 With simulated data

Another robustness analysis involves using simulation to quantify the effect of volatility change on the liquidity risk premium. There are situations where the volatility for actual return, frictionless return, or both, could be different; using simulation allows us to verify how our model captures these characteristics. It should be noted that our model are the stationary model, the magnitude of volatility could change but the time series should remain stationary. Figure A 1.7 to Figure A 1.9 of appendix illustrate the evolution of liquidity risk premium as a function of the volatility. In order to isolate the effect of volatility, we choose a particular parameter set, time interval, and confidence level equal to 5%. The procedure is as following, first, we take the estimated parameters from MRK for volume equal to 2700 as the initial parameters. Second, we change the volatility for actual return changes on keeping the volatility of frictionless return changes unchanged and simulate 5000 scenarios for this magnitude of volatility. In our exercise, we choose 14 different volatility multipliers. Third, we compute the liquidity risk premium based on the simulated data. We repeat the same procedure to frictionless return changes and both frictionless return changes and actual return changes[19].

For the purpose of comparison, we compute the relative change between initial liquidity risk premium and new simulated-based risk liquidity premium. As shown in the figures, when the total risk (actual return risk) increases and market risk (frictionless return risk) remains unchanged, the liquidity risk premium increases (In our example, an increases of more than 10%). Whereas when the market risk increases and total risk remains unchanged, the liquidity risk premium decreases (In our example, a decreases of more than 40%). If we sum up the two effects by increasing both volatilities, we observe a decreasing curve (In our example, a decrease of around 15%) due to the fact that the effect of market risk is more important than total risk on the liquidity risk premium.

1.9 Conclusion

In this paper, we introduce the ex-ante liquidity dimension in an intraday VaR measure using tick-by-tick data. To take the ex-ante liquidity into account, we first reconstruct the LOB for three blue-chip stocks actively traded in Deutsche Börse (SAP, RWE, and MRK) and define

[19] For the frictionless return changes, we increase the error term variance by keeping actual return changes variance unchanged, whereas, for both frictionless return changes and actual return changes, we increase both error term variance by multiplying the same multiplier.

the tick-by-tick actual return; that is, the log ratio of ex-ante liquidation price computed from a predetermined volume over the previous best ask price. Correspondingly, the proposed $IVaR^c$ and $LIVaR^c$ are based on the frictionless return changes and actual return changes and relate to the ex-ante loss in terms of frictionless return and actual return, respectively. In other words, both risk measures can be considered as the waiting costs associated with market risk and liquidity risk.

To model the dynamic of returns, we use a LogACD-VARMA-MGARCH structure that allows for both the irregularly spaced durations between two consecutive transactions and stylized facts in changes of return. In this setup, the time dimension is supposed to be strongly exogenous. Once the model is estimated, Monte Carlo simulations are used to make multiple-step forecasts. More specifically, the Log-ACD process first generates the tick-by-tick duration while the VARMA-MGARCH simulates the corresponding conditional tick-by-tick frictionless return changes and actual return changes. The model performance is assessed by using the tests of Kupiec (1995), Christoffersen (1998) and the new improved backtests of Ziggel et al. (2014) on both simulated deseasonalized and re-seasonalized data. All tests indicate that our model can capture well the dynamics of frictionless returns and actual returns over various time intervals for confidence levels of 95%, 97.5%, 99%, and 99.5%.

Our LIVaR provides a reliable measure of total risk for short horizons. In addition, the simulated data from our model can be easily converted to data in calendar time. In our sample, we find that the liquidity risk can account for up to 35% of total risk depending on order size. Regarding practical applications, potential users of our measure could be high-frequency traders that need to specify and update their trading strategies within a trading day, or market regulators who aim to track the evolution of market liquidity, along with brokers and clearinghouses that need to update their clients' intraday margins.

Future research could take several directions. Our study is focused on a single stock's ex-ante liquidity risk. A possible alternative is to investigate how IVaR and LIVaR evolve in the case of a portfolio. In particular, the lack of synchronization of the durations between two consecutive transactions for each stock is a challenge. Another direction is to test the role of ex-ante liquidity in different regimes. Our study focuses on the liquidity risk premium in a relatively stable period. It could also be interesting to investigate how the liquidity risk behaves during a crisis period. This study will require a more complicated econometric model to take different regimes into account.

References

Ait-Sahalia, Y., Saglam, M., 2014. High Frequency Traders: Taking Advantage of Speed. Working Paper, Princeton University, 50 pages

Anatolyev, S., Shakin, D., 2007. Trade Intensity in the Russian Stock Market: Dynamics, Distribution and Determinants. Applied Financial Economics 17, 87-104

Andersen, T.G., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance 4, 115-158

Angelidis, T., Benos, A., 2006. Liquidity adjusted value-at-risk based on the components of the bid-ask spread. Applied Financial Economics 16, 835

Bacidore, J., Ross, K., Sofianos, G., 2003. Quantifying market order execution quality at the New York stock exchange. Journal of Financial Markets 6, 281-307

Bangia, A., Diebold, F.X., Schuermann, T., Stroughair, J.D., 1999. Modeling Liquidity Risk With Implications for Traditional Market Risk Measurement and Management. The Wharton Financial Institutions Center WP 90-06

Battalio, R., Hatch, B., Jennings, R., 2003. All else equal?: a multidimensional analysis of retail, market order execution quality. Journal of Financial Markets 6, 143-162

Bauwens, L., Giot, P., 2000. The Logarithmic ACD Model: An Application to the Bid-Ask Quote Process of Three NYSE Stocks. Annals of Economics and Statistics / Annales d'Économie et de Statistique, 117-149

Beltran-Lopez, H., Giot, P., Grammig, J., 2009. Commonalities in the order book. Financial Markets and Portfolio Management 23, 209-242

Bilodeau, Y., 2013. XetraParser [computer software]. HEC Montréal

Black, F., 1971. Toward a fully automated stock exchange. Financial Analysts Journal 27, 28-44

Bloomfield, R., O'Hara, M., 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

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

Chaboud, A., Chiquoine, B., Hjalmarsson, E., Vega, C., 2014. (Forthcoming) Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market. Journal of Finance

Christoffersen, P., 2003. Elements of Financial Risk Management. Academic Press, San Diego, 214

Christoffersen, P.F., 1998. Evaluating Interval Forecasts. International Economic Review 39, 841-862

Coppejans, M., Domowitz, I., Madhavan, A., 2004. Resiliency in an automated auction. Working Paper

Demsetz, H., 1968. The cost of transacting. Quarterly Journal of Economics 82, 33-53

Dionne, G., Duchesne, P., Pacurar, M., 2009. Intraday Value at Risk (IVaR) using tick-by-tick data with application to the Toronto Stock Exchange. Journal of Empirical Finance 16, 777-792

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

Dufour, A., Engle, R.F., 2000. Time and the Price Impact of a Trade. The Journal of Finance 55, 2467-2498

Dufour, J.-M., Pelletier, D., 2011. Practical Methods for Modelling Weak VARMA Processes: Identification, Estimation and Specification with a Macroeconomic Application WORKING PAPER

Engel, E.M.R.A., 1984. A Unified Approach to the Study of Sums, Products, Time-aggregation and Other Functions of ARMA Processes. Journal of Time Series Analysis 5, 159-171

Engle, R., 2002. Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. Journal of Business & Economic Statistics 20, 339-350

Engle, R.F., 2000. The Econometrics of Ultra-High-Frequency Data. Econometrica 68, 1-22

Engle, R.F., Ng, V.K., 1993. Measuring and Testing the Impact of News on Volatility. The Journal of Finance 48, 1749-1778

Engle, R.F., Russell, J.R., 1998. Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. Econometrica 66, 1127-1162

Engle, R.F., Sheppard, K., 2001. Theoretical and Empirical properties of Dynamic Conditional Correlation Multivariate GARCH. NBER Working Papers

Foucault, T., Kadan, O., Kandel, E., 2005. Limit Order Book as a Market for Liquidity. Review of Financial Studies 18, 1171-1217

Ghysels, E., Jasiak, J., 1998. GARCH for Irregularly Spaced Financial Data: The ACD-GARCH Model. Studies in Nonlinear Dynamics & Econometrics 2, 133-149

Giot, P., Grammig, J., 2006. How large is liquidity risk in an automated auction market? Empirical Economics 30, 867-887

Glosten, L.R., Harris, L.E., 1988. Estimating the Components of the Bid/Ask Spread. Journal of Financial Economics 21, 123 - 142

Gourieroux, C., Jasiak, J., 2010. Local Likelihood Density Estimation and Value-at-Risk. Journal of Probability and Statistics 2010, 26Pages

Goyenko, R.Y., Holden, C.W., Trzcinka, C.A., 2009. Do liquidity measures measure liquidity? Journal of Financial Economics 92, 153-181

Grammig, J., Maurer, K.-O., 2000. Non-monotonic hazard functions and the autoregressive conditional duration model. Econometrics Journal 3, 16-38

Handa, P., Schwartz, R.A., 1996. Limit Order Trading. The Journal of Finance 51, 1835-1861

Hautsch, N., 2004. Modelling irregularly spaced financial data : theory and practice of dynamic duration models. doctoral. Springer-Verlag,Universität.

Hendershott, T., Jones, C.M., Menkveld, A.J., 2011. Does Algorithmic Trading Improve Liquidity? The Journal of Finance 66, 1-33

Irvine, P.J., Benston, G.J., Kandel, E., 2000. Liquidity Beyond the Inside Spread: Measuring and Using Information in the Limit Order Book. SSRN eLibrary

Kupiec, P., 1995. Techniques for Verifying the Accuracy of Risk Measurement Models. Journal of Derivatives 2, 73-84

Kyle, A.S., 1985. Continuous Auctions and Insider Trading. Econometrica 53, 1315-1335

Pacurar, M., 2008. Autoregressive Conditional Duration Models in Finance: a Survey of the Theoretical and Empirical Literature. Journal of Economic Surveys 22, 711-751

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

Rosu, I., 2009. A Dynamic Model of the Limit Order Book. The Review of Financial Studies 22, 4601-4641

Seppi, D.J., Kumar, P., 1994. Limit and Market Orders with Optimizing Traders. SSRN eLibrary

Weiß, G.N.F., Supper, H., 2013. Forecasting liquidity-adjusted intraday Value-at-Risk with vine copulas. Journal of Banking & Finance 37, 3334-3350

Zhang, M.Y., Russell, J.R., Tsay, R.S., 2001. A nonlinear autoregressive conditional duration model with applications to financial transaction data. Journal of Econometrics 104, 179-207

Ziggel, D., Berens, T., Weiß, G.N.F., Wied, D., 2014. A new set of improved Value-at-Risk backtests. Journal of Banking & Finance 48, 29-41

Table 1.1: Descriptive Statistics for SAP Raw and Deseasonalized Data

	Mean	Std.Dev	Skew	Kurt	Min	Max	LB(15)	LB2(15)
Panel A : SAP Raw data								
Duration	6.43	13.21	4.58	37.28	1.00E-03	292.75	6909.93	1426.06
FR Change	6.39E-09	1.93E-04	0.10	6.01	-1.98E-03	1.47E-03	6408.48	7082.94
AR(Q=8000)	-7.77E-09	1.66E-04	0.13	8.13	-1.92E-03	2.14E-03	5096.10	6512.58
AR(Q=6000)	-6.25E-09	1.66E-04	0.10	7.43	-1.93E-03	1.72E-03	5252.56	6021.21
AR(Q=4000)	-4.73E-09	1.68E-04	0.09	7.79	-1.96E-03	2.11E-03	5374.51	5185.26
AR(Q=2000)	-2.76E-09	1.72E-04	0.11	7.15	-1.98E-03	1.90E-03	5514.05	5156.26
Panel B : SAP Deseasonalized Data								
Duration	1.00	1.88	4.21	37.82	0.00	52.68	4345.54	1528.45
FR Change	-1.33E-04	1.01E+00	0.07	5.99	-10.96	7.21	6411.40	5923.61
AR(Q=8000)	-4.32E-05	1.04E+00	0.08	9.99	-15.37	11.25	5043.46	6599.80
AR(Q=6000)	-1.05E-05	1.04E+00	0.08	9.45	-13.07	12.60	5184.79	7189.66
AR(Q=4000)	-6.65E-05	1.03E+00	0.07	8.33	-12.78	10.76	5324.84	5661.21
AR(Q=2000)	-1.37E-04	1.02E+00	0.09	7.26	-12.78	9.69	5501.73	4684.47

The table shows the descriptive statistics for raw durations, actual return changes when Q=2,000, 4,000, 6,000, 8,000 and frictionless return changes. The sample period is the first 2 weeks of July 2010 with 44,467 observations.

Table 1.2: Descriptive Statistics for RWE Raw and Deseasonalized Data

	Mean	Std.Dev	Skew	Kurt	Min	Max	LB(15)	LB2(15)
Panel A : RWE Raw data								
Duration	7.64	15.88	4.75	37.53	1.00E-03	296.58	10061.17	2479.01
FR Change	-3.10E-08	2.51E-04	0.12	7.51	-3.07E-03	2.49E-03	5091.89	7971.04
AR(Q=4000)	-4.33E-08	2.23E-04	0.22	10.01	-2.96E-03	2.38E-03	3940.66	10253.80
AR(Q=3000)	-4.05E-08	2.25E-04	0.18	10.00	-3.29E-03	2.37E-03	4022.16	10239.86
AR(Q=2000)	-3.70E-08	2.27E-04	0.18	9.19	-3.22E-03	2.33E-03	4131.99	10306.19
AR(Q=1000)	-3.12E-08	2.32E-04	0.16	8.75	-3.19E-03	2.42E-03	4364.74	11046.85
Panel B : RWE Deseasonalized Data								
Duration	0.97	1.73	3.71	24.51	2.45E-05	28.63	3101.10	577.62
FR Change	-7.80E-05	0.98	0.12	6.00	-8.55	7.80	5123.94	5093.30
AR(Q=4000)	-1.23E-04	0.99	0.20	7.32	-8.61	10.11	3930.82	4781.76
AR(Q=3000)	-1.13E-04	0.98	0.18	7.08	-9.52	9.60	4002.22	4796.94
AR(Q=2000)	-1.01E-04	0.98	0.17	6.66	-9.40	9.14	4101.41	4888.15
AR(Q=1000)	-8.61E-05	0.98	0.17	6.38	-9.27	8.52	4315.98	5519.57

The table shows the descriptive statistics for raw durations, actual return changes when Q=1,000, 2,000, 3,000, 4,000 and frictionless return changes. The sample period is the first 2 weeks of July 2010 with 37,394 observations.

Table 1.3: Descriptive Statistics for MRK Raw and Deseasonalized Data

Panel A : MRK Raw data								
	Mean	Std.Dev	Skew	Kurt	Min	Max	LB(15)	LB2(15)
Duration	16.31	36.40	5.16	47.06	0.001	718.35	2157.42	905.39
FR Change	-3.83E-09	3.11E-04	0.00	15.52	-5.03E-03	4.25E-03	2551.10	4125.30
AR(Q=2700)	-1.33E-08	2.74E-04	-0.02	24.71	-4.71E-03	3.75E-03	1672.50	2303.16
AR(Q=1800)	-9.58E-09	2.79E-04	0.17	24.70	-4.77E-03	4.74E-03	1754.34	2308.44
AR(Q=900)	-6.82E-09	2.84E-04	0.24	21.74	-4.82E-03	4.50E-03	1890.60	3030.64
Panel B : MRK Deseasonalized Data								
Duration	0.97	1.95	4.13	29.42	1.581E-05	26.65	736.41	150.75
FR Change	-2.55E-04	1.01	0.06	13.44	-13.87	11.61	2537.61	4884.19
AR(Q=2700)	-2.78E-04	1.03	0.14	15.67	-14.65	11.65	1670.95	2537.68
AR(Q=1800)	-2.81E-04	1.03	0.20	17.60	-14.15	15.63	1730.01	2201.63
AR(Q=900)	-2.68E-04	1.02	0.33	16.44	-13.79	16.86	1862.13	2289.72

The table shows the descriptive statistics for raw durations, actual return changes when Q=900, 1,800, 2,700 and frictionless return changes. The sample period is the first 2 weeks of July 2010 with 17,472 observations.

Table 1.4: Estimation Results SAP (v = 4000)

	Estimation log-ACD(2,1)-VARMA(4.2)- NGARCH((1.3).(1.3)) (Obs =21089)					
	Parameters	Estimation	Std Error	Statistics		
ACD(2.1)	α_1	0.127	0.009	LB test on Residuals		
	α_2	-0.049	0.009	Lags	Statistic	C_Value
	β_1	0.963	0.004	5	5.364	11.070
	γ_1	0.812	0.022	10	13.045	18.307
	γ_2	0.419	0.016	15	**25.077**	24.996
	ω	-0.081	0.004	20	30.561	31.410
Actual Return Changes VARMA(4.2)-NGARCH((1.3).(1,3))	$\varphi_{11}^{(1)}$	1.043	0.088	LB test on Residuals		
	$\varphi_{12}^{(1)}$	-0.322	0.010			
	$\varphi_{11}^{(2)}$	-0.154	0.061	Lags	Statistic	C_Value
	$\varphi_{12}^{(2)}$	0.089	0.031	5	5.291	11.070
	$\varphi_{11}^{(3)}$	-0.031	0.016	10	13.424	18.307
	$\varphi_{12}^{(3)}$	0.047	0.014	15	16.106	24.996
	$\varphi_{11}^{(4)}$	0.032	0.011	20	18.769	31.410
	$\varphi_{12}^{(4)}$	**0.008**	**0.012**	LB test on Squared Residuals		
	$\delta_{11}^{(1)}$	-1.338	0.087			
	$\delta_{11}^{(2)}$	0.353	0.085	Lags	Statistic	C_Value
	ω^b	0.088	0.004	5	3.395	11.070
	α_1^b	0.412	0.022	10	15.405	18.307
	β_1^b	0.159	0.006	15	18.925	24.996
	π^b	0.404	0.022	20	21.266	31.410
	α_3^b	0.308	0.018			
	γ^b	0.073	0.002			
Frictionless Return Changes ARMA(4,2)-GARCH((1,3),(1,3))	$\varphi_{22}^{(1)}$	0.6055	0.0693	LB test on Residuals		
	$\varphi_{21}^{(1)}$	0.1754	0.0093			
	$\varphi_{22}^{(2)}$	-0.0401	0.0175	Lags	Statistic	C_Value
	$\varphi_{21}^{(2)}$	-0.0593	0.0187	5	**15.811**	11.070
	$\varphi_{22}^{(3)}$	0.0488	0.0133	10	**21.046**	18.307
	$\varphi_{21}^{(3)}$	-0.0199	0.0103	15	23.203	24.996
	$\delta_{22}^{(1)}$	-1.4122	0.0690	20	**32.595**	31.410
	$\delta_{22}^{(2)}$	0.4265	0.0669	LB test on Squared Residuals		
	ω^f	0.0716	0.0037			
	α_1^f	0.4799	0.0305	Lags	Statistic	C_Value
	β_1^f	0.1612	0.0074	5	7.794	11.070
	π^f	0.3466	0.0254	10	17.916	18.307
	α_3^f	0.2472	0.0237	15	21.902	24.996
	γ^f	0.0377	0.0025	20	24.743	31.410
DCC parameter	θ_1	0.0791	0.0032			
	θ_2	0.7537	0.0113			

The Statistics column reports the Ljung-Box statistic on standardized residuals of duration, actual return changes and squared standardized residuals for different lags. The bold entries are the estimation coefficients that are not significantly different from zero and the Ljung-Box statistics that reject the non-correlation in the residuals.

Table 1.5: Estimation Results RWE (v = 4000)

Estimation log-ACD(3,1) VARMA(5.1)- NGARCH((1.3).(1.3)) (Obs =15320)						
	Parameter	Estimation	StdError	Statistics		
ACD(3.1)	α_1	0.108	0.010	LB test on Residuals		
	α_2	-0.020	0.013			
	α_3	-0.024	0.009	Lags	Statistic	C_Value
	β_1	0.970	0.004	5	9.209	11.070
	γ_1	1.145	0.028	10	**20.280**	18.307
	γ_2	0.280	0.010	15	**39.712**	24.996
	ω	-0.068	0.004	20	**44.304**	31.410
Actual Return Changes VARMA(5.1)-NGARCH((1.3).(1,3))	$\varphi_{11}^{(1)}$	0.710	0.012	LB test on Residuals		
	$\varphi_{12}^{(1)}$	-0.333	0.011			
	$\varphi_{11}^{(2)}$	0.047	0.014			
	$\varphi_{12}^{(2)}$	**0.005**	**0.013**	Lags	Statistic	C_Value
	$\varphi_{11}^{(3)}$	**0.006**	**0.014**	5	10.161	11.070
	$\varphi_{12}^{(3)}$	0.024	0.013	10	14.653	18.307
	$\varphi_{11}^{(4)}$	0.006	0.014	15	**27.042**	24.996
	$\varphi_{12}^{(4)}$	**0.009**	**0.013**	20	30.095	31.410
	$\varphi_{11}^{(5)}$	0.041	0.012	LB test on Squared Residuals		
	$\varphi_{12}^{(5)}$	**0.007**	**0.012**			
	$\delta_{11}^{(1)}$	-0.980	0.002			
	ω^b	0.099	0.006	Lags	Statistic	C_Value
	α_1^b	0.437	0.030	5	4.356	11.070
	β_1^b	0.156	0.008	10	8.474	18.307
	π^b	0.461	0.029	15	11.071	24.996
	α_3^b	0.260	0.023	20	14.102	31.410
	γ^b	0.067	0.003			
Frictionless Return Changes VARMA(5.1)-NGARCH((1.3).(1,3))	$\varphi_{22}^{(1)}$	0.2324	0.0123			
	$\varphi_{21}^{(1)}$	0.1447	0.0109	LB test on Residuals		
	$\varphi_{22}^{(2)}$	0.0374	0.0132			
	$\varphi_{21}^{(2)}$	**0.0048**	**0.0131**	Lags	Statistic	C_Value
	$\varphi_{22}^{(3)}$	0.0746	0.0135	5	6.378	11.070
	$\varphi_{21}^{(3)}$	-0.0428	0.0133	10	9.420	18.307
	$\varphi_{22}^{(4)}$	0.0305	0.0138	15	13.824	24.996
	$\varphi_{21}^{(4)}$	**-0.0168**	**0.0132**	20	22.560	31.410
	$\varphi_{22}^{(5)}$	**0.0205**	**0.0128**			
	$\varphi_{21}^{(5)}$	0.0179	0.0114			
	$\delta_{22}^{(1)}$	-0.9805	0.0018	LB test on Squared Residuals		
	ω^f	0.0711	0.0049			
	α_1^f	0.4586	0.0343	Lags	Statistic	C_Value
	β_1^f	0.1436	0.0080	5	7.403	11.070
	π^f	0.4531	0.0359	10	16.284	18.307
	α_3^f	0.2731	0.0276	15	19.750	24.996
	γ^f	0.0299	0.0029	20	22.811	31.410
DCC parameter	θ_1	0.0813	0.0043			
	θ_2	0.7791	0.0135			

The Statistics column reports the Ljung-Box statistic on standardized residuals of duration, actual return changes and squared standardized residuals for different lags. The bold entries are the estimation coefficients that are not significantly different from zero and the Ljung-Box statistics that reject the non-correlation in the residuals.

Table 1.6: Estimation Results MRK (v = 1800)

Estimation log-ACD(1,1)-VARMA(1.1)-NGARCH((1.3).(1.3)) (Obs =7653)						
	Parameter	Estimation	StdError	Statistics LB test on Residuals		
ACD(1.1)	α_1	0.048	0.005	Lags	Statistic	C_Value
	β_1	0.984	0.004	5	8.240	11.070
	γ_1	0.784	0.034	10	17.239	18.307
	γ_2	0.382	0.025	15	21.796	24.996
	ω	-0.049	0.005	20	25.006	31.410
Actual Return Changes VARMA(2.2)-NGARCH((1.3).1)	$\varphi_{11}^{(1)}$	0.9401	0.0599	LB test on Residuals		
	$\varphi_{12}^{(1)}$	-0.3679	0.0149	Lags	Statistic	C_Value
	$\varphi_{11}^{(2)}$	**-0.0316**	**0.0515**	5	11.971	11.070
	$\varphi_{12}^{(2)}$	0.0945	0.0264	10	25.105	18.307
	$\delta_{11}^{(1)}$	-1.1788	0.0584	15	27.396	24.996
	$\delta_{11}^{(2)}$	0.2028	0.0565	20	29.201	31.410
	ω^b	0.1543	0.0109	LB test on Squared Residuals		
	α_1^b	0.3863	0.0341	Lags	Statistic	C_Value
	β_1^b	0.1877	0.0115	5	1.818	11.070
	π^b	0.4853	0.0328	10	3.084	18.307
	α_3^b	0.2238	0.0289	15	7.719	24.996
	γ^b	0.0861	0.0032	20	10.320	31.410
Frictionless Return Changes ARMA(1,3)-GARCH((1,3),1)	$\varphi_{22}^{(1)}$	0.4737	0.0725	LB test on Residuals		
	$\varphi_{21}^{(1)}$	0.1822	0.0129	Lags	Statistic	C_Value
	$\varphi_{22}^{(2)}$	0.0422	0.0228	5	**20.248**	11.070
	$\varphi_{21}^{(2)}$	-0.0652	0.0179	10	**28.844**	18.307
	$\delta_{22}^{(1)}$	-1.2728	0.0712	15	**30.197**	24.996
	$\delta_{22}^{(2)}$	0.3020	0.0677	20	**34.977**	31.410
	ω^f	0.1551	0.0119	LB test on Squared Residuals		
	α_1^f	0.4034	0.0378	Lags	Statistic	C_Value
	β_1^f	0.2373	0.0163	5	4.348	11.070
	α_3^f	0.1670	0.0272	10	11.419	18.307
	γ^f	0.0525	0.0039	15	22.855	24.996
DCC parameter	θ_1	0.1009	0.0027	20	**35.233**	31.410
	θ_2	0.6775	0.0159			

The Statistics column reports the Ljung-Box statistic on standardized residuals of duration, actual return changes and squared standardized residuals for different lags. The bold entries are the estimation coefficients that are not significantly different from zero and the Ljung-Box statistics that reject the non-correlation in the residuals.

Table 1.7: Backtesting on Simulated Re-seasonalized Data

Panel A: SAP Out-of-sample Backtesting on Deseasonalized Actual Return Change (v = 4000)

Nb of Interval	Interval (units)	Time interval (in Min)	Kupiec test				Christoffersen Test				Unconditional Coverage				I.I.D			
			5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%
585	40	4.36	**0.001**	**0.010**	0.717	0.191	0.407	0.724	0.769	0.953	**0.019**	**0.008**	**0.000**	0.241	0.364	0.748	0.133	0.122
468	50	5.45	0.154	0.405	0.160	0.321	0.272	0.575	0.896	0.948	**0.006**	**0.006**	**0.018**	0.099	0.292	0.094	0.205	0.928
390	60	6.54	0.907	0.090	0.286	0.971	0.174	0.747	0.919	0.919	0.188	0.059	0.173	0.776	0.998	0.069	0.050	0.449
292	80	8.73	0.313	0.796	0.267	0.672	0.376	0.530	0.709	0.868	0.101	0.153	**0.024**	0.576	0.097	**0.014**	**0.007**	**0.010**
234	100	10.90	0.078	0.062	0.320	0.872	0.709	0.852	0.926	0.926	0.202	0.203	0.216	0.780	0.116	0.074	0.387	0.384
195	120	13.08	0.342	0.954	0.451	0.980	0.503	0.646	0.919	0.919	0.611	0.911	0.974	0.168	0.191	0.439	0.161	0.150

Panel B: RWE Out-of-sample Backtesting on Deseasonalized Actual Return Change (v = 4000)

Nb of Interval	Interval (units)	Time interval (in Min)	Kupiec test				Christoffersen Test				Unconditional Coverage				I.I.D			
			5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%
535	40	4.77	**0.001**	**0.002**	0.095	0.239	0.477	0.832	0.902	0.951	**0.003**	**0.004**	**0.024**	0.226	0.951	0.876	0.690	0.667
428	50	5.96	0.134	0.589	0.891	0.923	0.313	0.557	0.783	0.891	0.299	0.212	0.322	0.383	0.661	0.427	0.514	0.689
357	60	7.14	0.217	0.292	0.362	0.874	0.341	0.679	0.881	0.881	0.066	0.207	0.051	0.107	0.171	0.442	0.204	0.083
267	80	9.55	0.106	0.258	0.239	0.761	0.510	0.762	0.931	0.931	0.084	0.130	0.790	0.713	0.664	0.528	0.535	0.689
214	100	11.92	0.109	0.877	0.577	0.126	0.591	0.661	0.811	0.811	**0.040**	0.098	0.592	0.855	0.970	**0.021**	**0.020**	**0.012**
178	120	14.33	0.145	0.460	0.522	0.909	0.590	0.748	0.915	0.915	**0.004**	**0.040**	0.235	0.351	0.819	0.625	0.317	0.632
153	140	16.67	0.626	0.299	0.646	0.797	0.538	0.817	0.908	0.908	0.722	0.088	0.253	0.864	0.059	0.884	**0.026**	0.742

Panel C: MRK Out-of-sample Backtesting on Deseasonalized Actual Return Change (v = 1800)

Nb of Interval	Interval (units)	Time interval (in Min)	Kupiec test				Christoffersen Test				Unconditional Coverage				I.I.D			
			5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%
476	20	5.36	**0.006**	0.120	0.384	0.699	0.430	0.647	0.845	0.845	**0.001**	**0.032**	0.496	0.800	0.246	0.229	0.230	0.232
317	30	8.04	**0.026**	0.119	0.152	0.617	0.547	0.782	0.936	0.936	**0.008**	0.052	0.326	0.914	0.646	0.220	0.245	0.236
238	40	10.71	**0.021**	0.058	0.309	0.857	0.642	0.854	0.927	0.927	0.085	0.085	0.602	0.414	0.664	0.219	0.226	0.211
190	50	13.42	0.213	0.384	0.942	0.959	0.530	0.756	0.836	0.918	0.070	0.088	0.577	0.148	0.784	0.978	0.606	0.501
158	60	16.14	0.470	0.613	0.747	0.820	0.490	0.732	0.820	0.910	0.167	0.412	0.888	0.697	0.249	0.240	0.220	0.256
95	100	26.84	0.378	0.800	0.959	0.506	0.656	0.768	0.883	0.883	0.336	0.886	0.592	0.168	0.252	0.215	0.242	0.237

The table contains the p-values for Kupiec, Christoffersen, Unconditional coverage, and I.I.D (Ziggel et al (2013)) tests for the stocks SAP, RWE, and MRK. Interval is the interval length used for computing the LIVaR. Nb of intervals is the number of intervals for out-of sample analysis and Time interval in minutes is the corresponding calendar time. Bold entries indicate the rejections of the model at 95% confidence level. When the number of hits is less than two, the p-values are denoted by #.

Table 1. 8: Backtesting on Simulated Re-seasonalized Data

Panel A: SAP Out-of-sample Backtesting on Raw Actual Return Change (v = 4000)

Interval (in Mins)	Nb of Interval	Kupiec test				Christoffersen Test				Unconditional Coverage				I.I.D			
		5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%
5	485	0.445	0.420	0.091	**0.017**	0.612	0.444	0.130	0.675	0.790	0.978	0.436	0.148	0.851	0.324	0.386	0.650
6	405	0.132	0.483	0.256	0.986	0.334	0.570	0.888	0.888	0.060	0.089	0.236	0.694	0.328	0.564	0.709	0.692
8	300	0.588	0.566	1.000	0.663	0.297	0.651	0.841	#	0.721	0.643	0.408	0.983	0.667	0.237	0.086	**0.016**
9	270	0.670	0.923	0.209	0.220	0.500	0.132	0.697	0.832	0.723	0.780	0.764	0.459	0.562	0.124	0.286	0.689
10	245	0.066	0.726	0.362	0.520	0.081	0.552	0.752	0.898	0.713	0.896	0.925	0.726	0.370	0.416	0.693	**0.020**

Panel B: RWE Out-of-sample Backtesting on Raw Actual Return Change (v = 4000)

Interval (in Mins)	Nb of Interval	Kupiec test				Christoffersen Test				Unconditional Coverage				I.I.D			
		5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%
5	485	0.173	0.341	0.364	0.778	0.266	0.604	0.897	0.927	0.510	0.649	0.922	0.889	0.996	0.682	0.085	0.587
6	405	0.132	0.071	0.583	0.986	0.352	0.777	0.888	0.921	0.358	0.604	0.812	0.325	0.723	0.791	**0.008**	**0.005**
7	345	0.570	0.563	0.772	0.379	0.259	0.618	0.791	0.851	0.536	0.928	0.276	0.455	0.742	0.072	0.170	0.858
8	300	0.070	0.156	0.537	0.697	0.651	0.776	0.908	0.908	**0.002**	0.290	0.400	0.651	0.694	0.677	0.257	0.269
9	270	0.307	0.247	0.857	0.601	0.430	0.795	0.863	0.931	0.273	0.788	0.673	0.386	0.804	0.456	**0.001**	**0.001**
10	245	**0.043**	0.156	0.733	0.175	0.647	0.856	0.856	0.856	0.084	0.357	0.485	0.197	0.639	0.533	0.068	0.069

Panel C: MRK Out-of-sample Backtesting on Raw Actual Return (v = 1800)

Interval (in Mins)	Nb of Interval	Kupiec test				Christoffersen Test				Unconditional Coverage				I.I.D			
		5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%
5	485	0.187	0.201	0.140	0.298	0.344	0.628	0.897	0.949	0.090	0.102	0.086	0.270	0.892	0.831	0.101	0.101
6	405	**0.010**	0.071	0.980	0.220	0.476	0.723	0.777	0.777	0.172	**0.046**	0.179	0.780	0.111	0.590	0.444	0.102
7	345	0.273	0.175	0.394	0.548	0.312	0.701	0.878	0.939	0.431	0.126	0.460	0.836	0.882	0.457	0.204	0.102
9	270	0.098	0.247	0.654	0.752	0.513	0.728	0.863	0.931	**0.013**	0.077	0.504	0.627	0.387	0.512	0.966	0.968
10	245	**0.016**	0.050	0.765	0.520	0.647	0.856	0.856	0.856	**0.002**	0.110	0.366	0.621	0.165	0.577	0.109	0.108

The table contains the p-values for Kupiec and Christoffersen Unconditional coverage, and I.I.D (Ziggel et al (2013)) tests. Intervals are regularly time-spaced from 5 minutes to 10 minutes. Bold entries indicate the rejections of the model at 95% confidence level. When the number of hits is less than two, the p-values are denoted by #.

Table 1.9: The Relative Liquidity Risk Premium

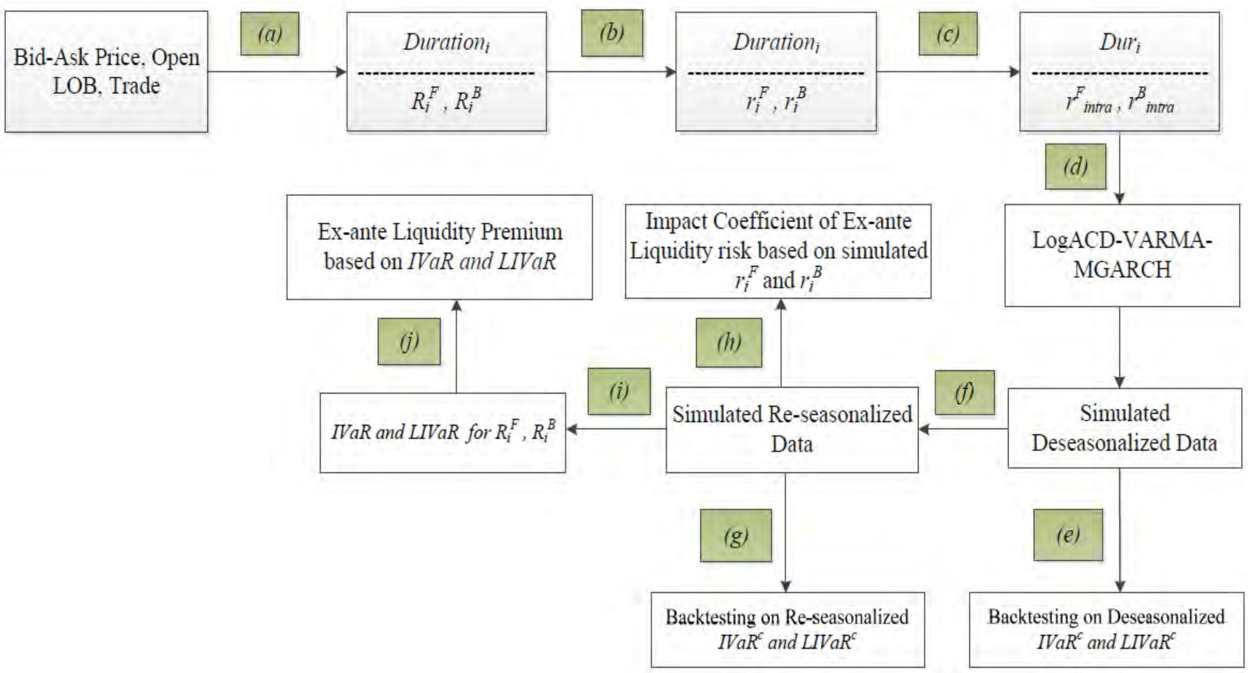
Panel A: SAP					
	Relative Ex-ante Liquidity Risk Premium at the end of 5-min Interval				
Confidence Level	Volume (Shares)/PDHTV				
	550	1500	2000	4000	8000
	mean	>90%	>95%	>97.5%	>99%
10%	4.03%	10.36%	12.86%	22.20%	34.17%
5%	3.46%	9.32%	11.58%	20.55%	32.00%
1%	2.72%	8.22%	10.26%	18.40%	29.24%

Panel B: RWE					
	Relative Ex-ante Liquidity Risk Premium at the end of 5-min Interval				
Confidence Level	Volume (Shares)/PDHTV				
	350	1000	1500	2000	4000
	mean	>90%	>95%	>97.5%	>99%
10%	3.62%	9.78%	14.57%	17.82%	27.80%
5%	2.88%	8.39%	12.87%	15.92%	25.26%
1%	2.09%	6.54%	10.55%	13.45%	21.76%

Panel C: MRK					
	Relative Ex-ante Liquidity Risk Premium at the end of 5-min Interval				
Confidence Level	Volume (Shares)/PDHTV				
	200	500	900	1800	2700
	mean	>90%	>95%	>97.5%	>99%
10%	1.96%	9.83%	16.66%	27.07%	33.40%
5%	1.03%	8.96%	15.81%	25.79%	31.88%
1%	0.39%	8.67%	15.12%	24.65%	30.18%

The panels present the proportion of liquidity risk on total risk given different order sizes at 90%, 95% and 99% confidence level for SAP, RWE, and MRK. The order size for each stock ranges from the average (first column) to the 99th quantile of transaction volume (last column). PDHTV: Percentile in the Distribution of Historical Trade Volume

Figure 1.1: Flowchart for Computing Impact Coefficient of Ex-ante Liquidity Risk and Ex-ante Liquidity Premium



(a): Compute durations, frictionless returns and actual returns based on raw data.

(b): Construct the frictionless return changes and actual return changes.

(c): Remove seasonality from durations, frictionless return changes and actual return changes.

(d): Estimate the model.

(e): Backtest deseasonalized $IVaR^c$ and $LIVaR^c$.

(f): Compute the re-seasonalized data by re-introducing the seasonal factors.

(g): Backtest re-seasonalized $IVaR^c$ and $LIVaR^c$.

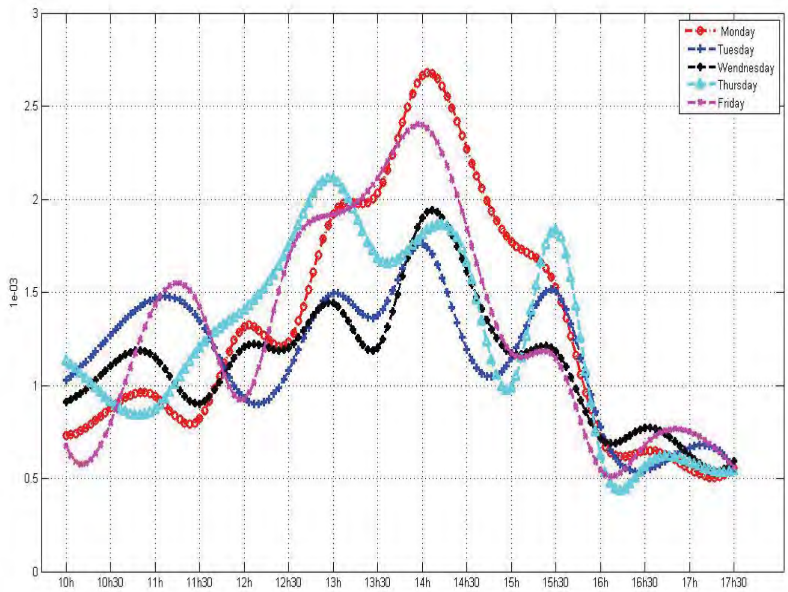
(h): Compute the impact coefficient of ex-ante liquidity risk.

(i): Compute the IVaR and LIVaR for frictionless return and actual return.

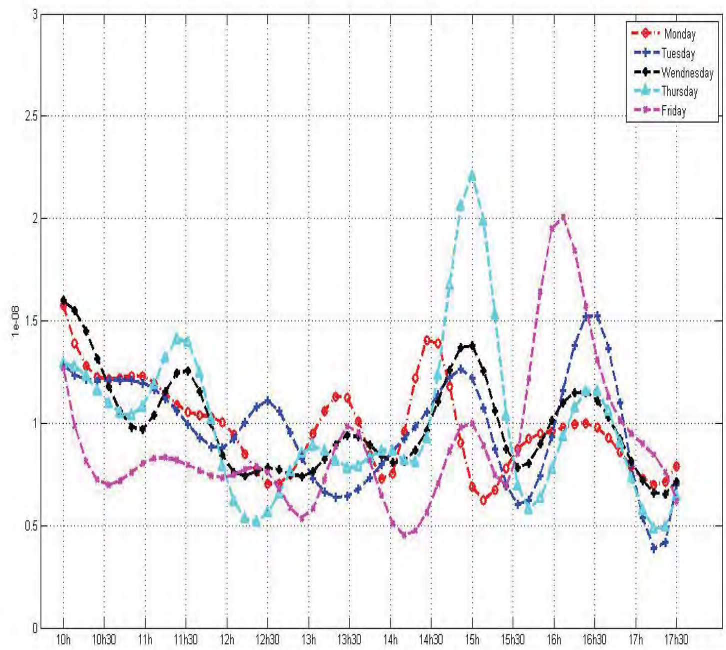
(j): Derive the ex-ante liquidity risk premium.

Figure 1.2: Seasonality Factor For RWE

Panel A: Seasonality Factor for Duration



Panel B: Seasonality Factor for Actual Return Changes (v = 4000)



Panel C: Seasonality Factor for Frictionless Return Changes

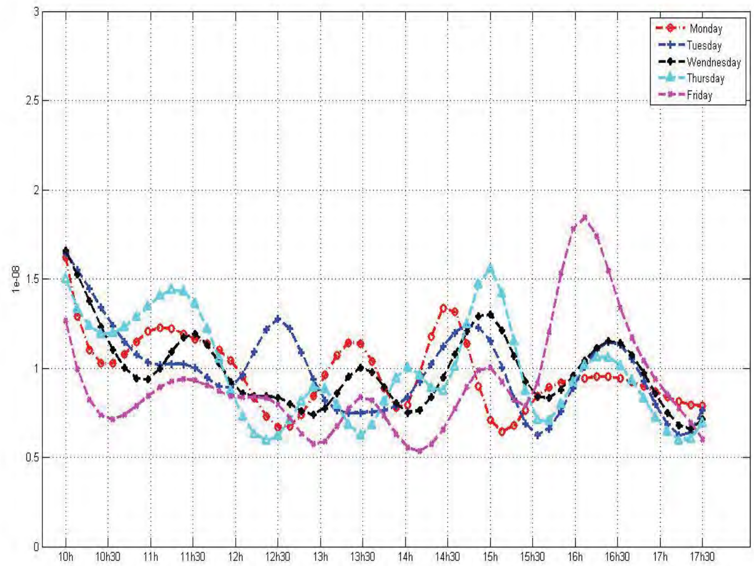
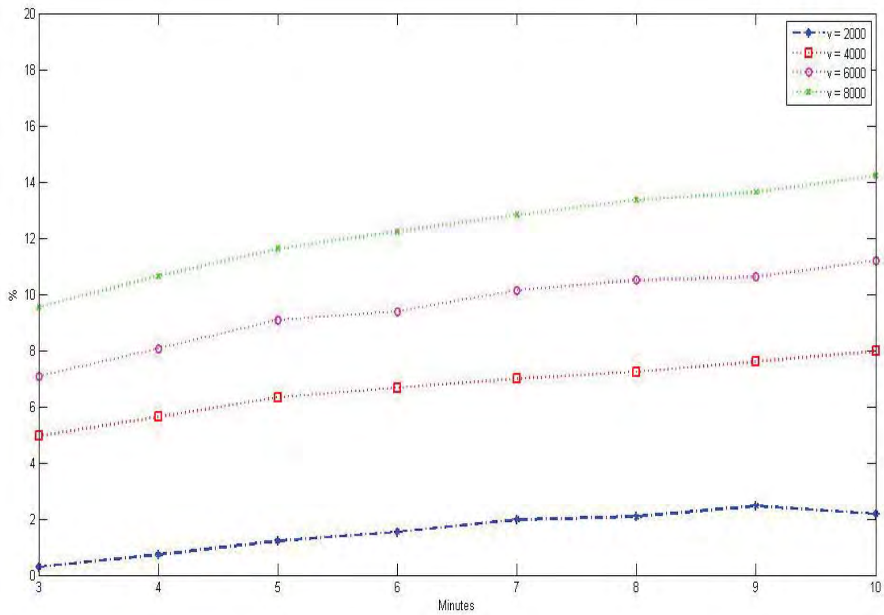
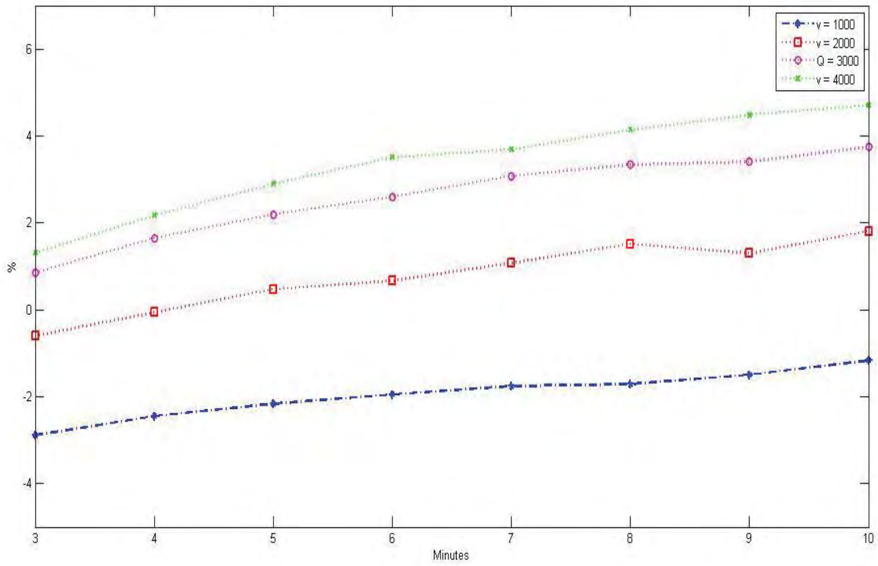


Figure 1.3: Impact Coefficients of Ex-ante Liquidity for Different Time Intervals

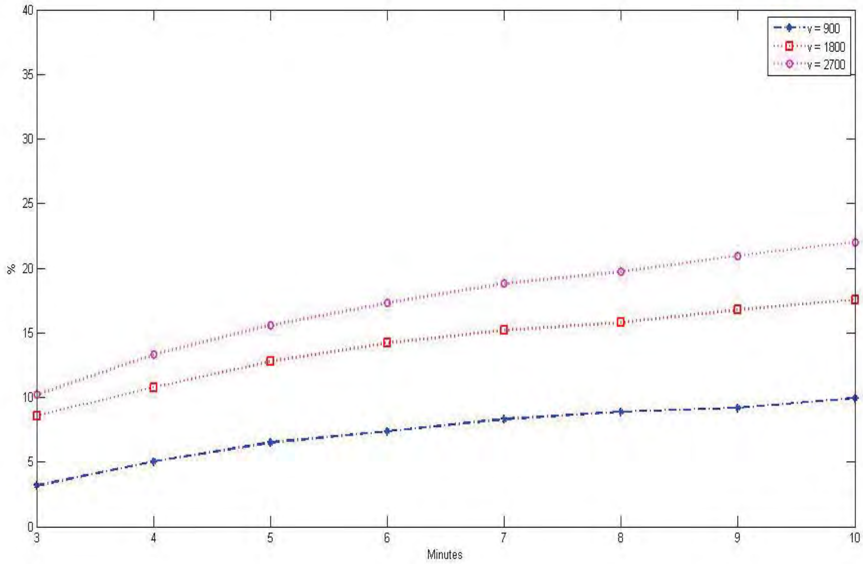
Panel A: SAP



Panel B: RWE



Panel C: MRK



Panels A, B and C illustrate how the impact coefficients of ex-ante liquidity evolve for intervals from 3 minutes to 10 minutes for stocks SAP, RWE, and MRK. The selected volumes for the actual return changes are 2,000, 4,000, 6,000, and 8,000 shares for SAP, 1,000, 2,000, 3,000, and 4,000 shares for RWE, and 900, 1,800 and 2,700 shares for MRK.

Figure 1.4: Computation of the Frictionless Return and the Actual Return for the End of an Interval

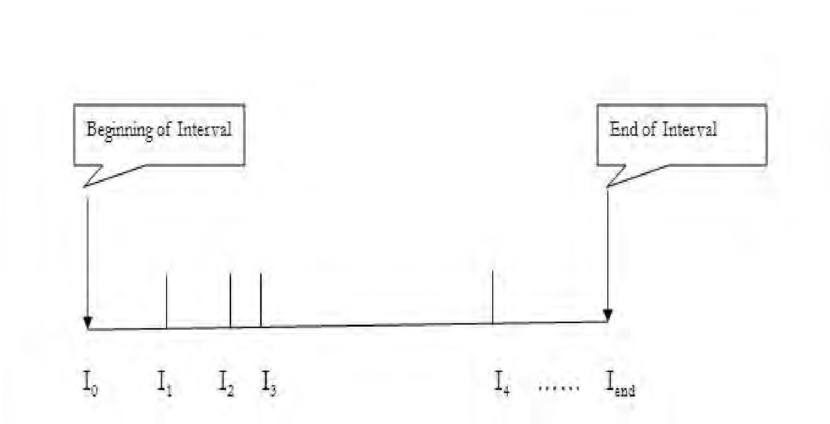
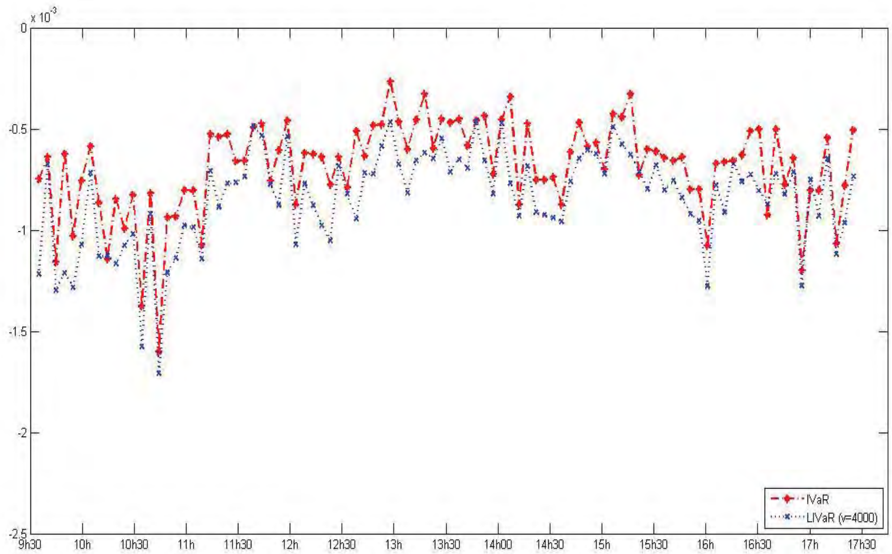


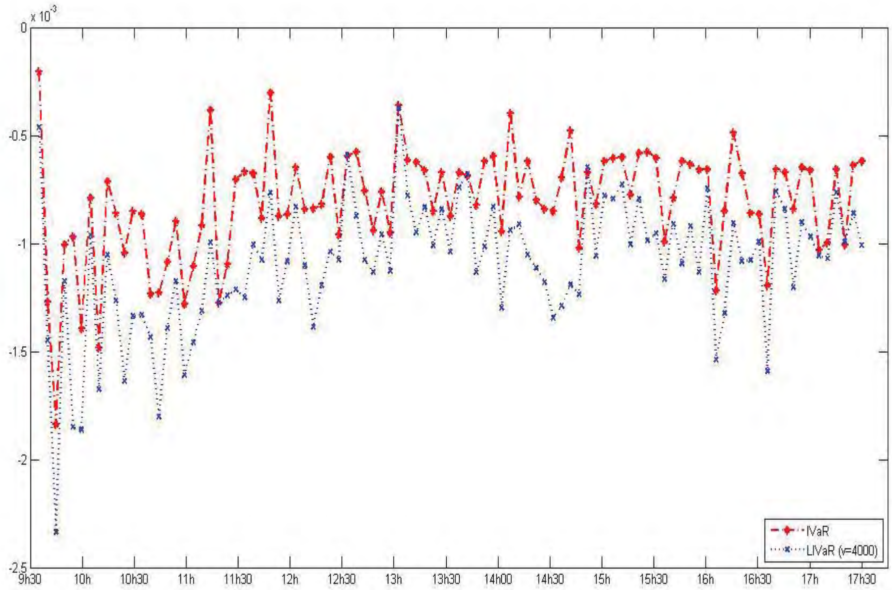
Figure 1.4 illustrates the computation of the frictionless return and the actual return for the end of an interval. I indicates the transaction. At the beginning of the interval, we compute the frictionless return and actual return using real market data. Each I corresponds to a frictionless (actual) change that comes from the simulations. Consequently, the frictionless return (actual return) at the end of the interval is the sum of the initial frictionless return (actual return) and all the corresponding changes in the interval.

Figure 1.5: IVaR and LIVaR of 5-minute for July 12, 2010

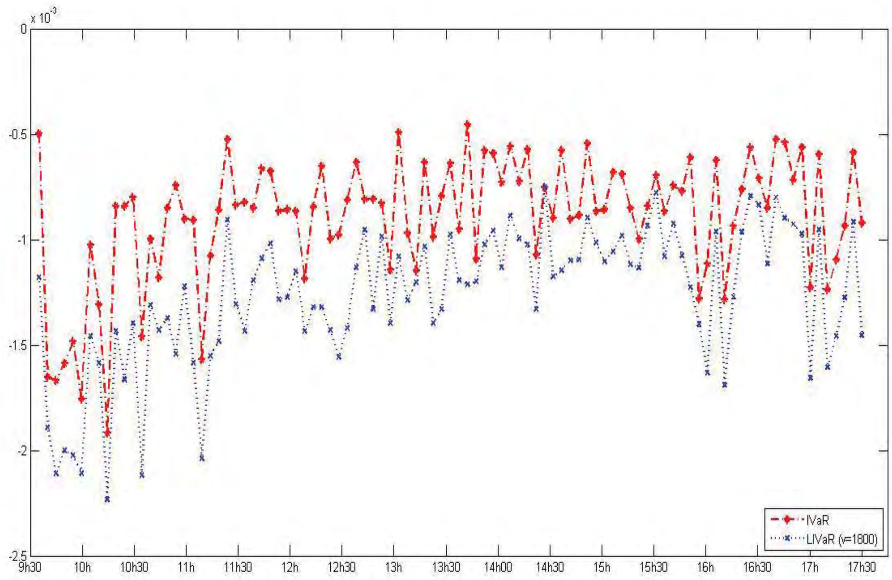
Panel A: SAP with LIVaR (*v* =4000)



Panel B: RWE with LIVaR (v =4000)

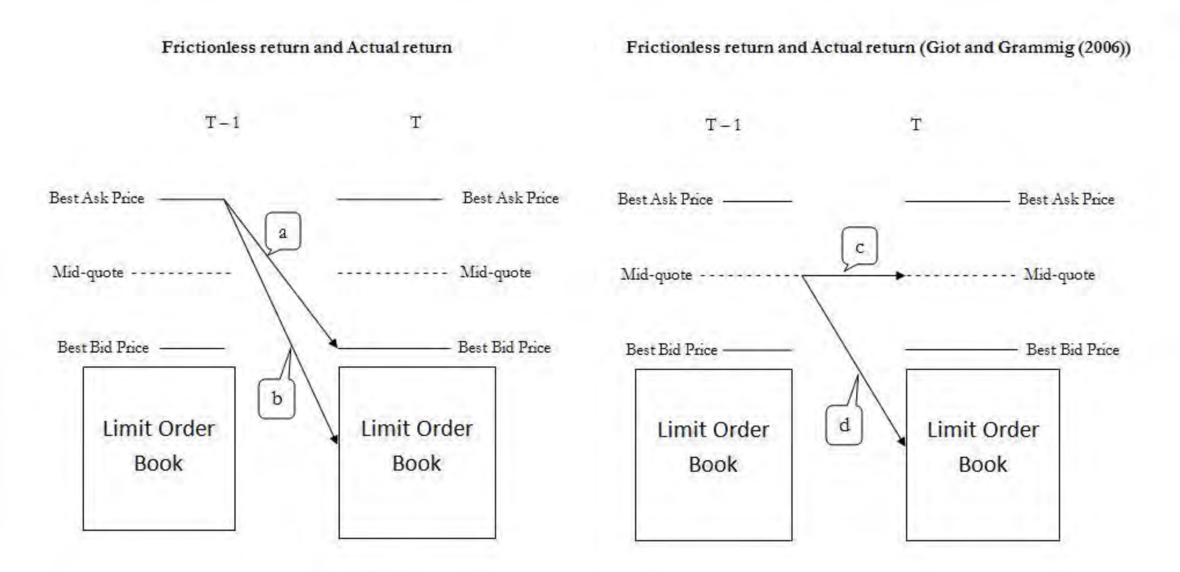


Panel C: MRK with LIVaR (v =1800)



Panels A, B and C present the VaRs for frictionless returns and actual returns at the end of each 5-minute interval on July 12, 2010, for the three stocks of SAP, RWE, and MRK, respectively. The selected volumes for the actual returns are 4,000 shares for SAP, 4,000 shares for RWE and 1,800 shares for MRK.

Figure 1.6: Frictionless Returns and Actual Returns



This figure presents the difference between our frictionless (actual) return and the frictionless (actual) return proposed by Giot and Grammig (2006). Arrow *a* presents the starting price and end price in constructing our frictionless return, while arrow *b* shows the starting price and end price for the actual return given a liquidation quantity *v*. Both frictionless returns and actual returns take the previous best ask price as starting price. Arrows *c* and *d* give the starting price and end price for computing the frictionless return and the actual return (for quantity *v)* proposed by Giot and Grammig (2006). Their frictionless return takes the previous mid-quote as the starting price and the following mid-quote as the end price. Their actual return takes the previous mid-quote as the starting price and the actual price as the end price.

Chapter 2
Effects of the Limit Order Book on Price Dynamics

Abstract

In this paper, we analyze whether the state of the limit order book affects future price movements in line with what recent theoretical models predict. We do this in a linear vector autoregressive system which includes midquote return, trade direction and variables that are theoretically motivated and capture different dimensions of the information embedded in the limit order book. We find that different measures of depth and slope of bid and ask sides as well as their ratios cause returns to change in the next transaction period in line with the predictions of Goettler, Parlour and Rajan (2009) and Kalay and Wohl (2009). Limit order book variables also have significant long term cumulative effects on midquote return, which is stronger and takes longer to be fully realized for variables based on higher levels of the book. In a simple high frequency trading exercise, we show that it is possible in some cases to obtain economic gains from the statistical relation between limit order book variables and midquote return

2.1 Introduction

Regardless of their original trading mechanism, almost all of the world's major exchanges now feature electronic limit order books. Some of them such as Euronext Paris have completely abandoned any form of floor trading and operate as pure electronic limit order markets without any designated market makers. Others such as NASDAQ also had to adapt their trading mechanisms to reflect the growing importance of electronic limit order books originating from alternative trading systems such as Electronic Communication Networks (ECN).

As the importance of electronic limit order books in financial markets increases, so does the demand for information embedded in them. Most exchanges such as those operated by NYSE Euronext now offer investors access to historical and real-time data on their limit order books for a fee while others such as Frankfurt Börse make their electronic limit order book data available with a minor delay on their websites. More importantly, historical and real-time data on limit order books are available at ever increasing frequencies, thanks to recent technological advancements in electronic trading systems. For example, Frankfurt Börse offers historical data on its electronic limit order book including trades and quotes up to 20 levels with millisecond time stamps. Thus, there is an immense wealth of historical and real-time information embedded in high frequency limit order books available to investors.

Whether information embedded in the limit order book should have any effect on future price movements is a theoretical question. Earlier microstructure models such as Kyle (1985), Glosten and Milgrom (1985), Rock (1996) and Glosten (1994) treated limit orders as free options provided by uninformed investors to the market and susceptible to being picked off by later better informed investors. To put differently, most earlier microstructure models implicitly assumed that the limit order book cannot possibly be informative for future price movements. However, recent theoretical models allow informed investors to choose between limit and market orders and show that they indeed use not only market, as assumed in the previous literature, but also limit orders in rational expectations equilibria.[1] Regardless of the channel through which information gets embedded in the limit order book, the common prediction of these models is that limit orders should contain relevant information for the true value of the underlying asset and, thus, affect future price movements.

In this paper, we analyze whether the state of the limit order book affects future price movements in line with what the theory predicts. To this end, we reconstruct the first 20 levels of the historical limit order book every millisecond for several companies traded at Frankfurt Stock Exchange in July 2010 and June 2011 based on the data from the

[1] For example, informed investors might use limit orders to avoid detection (as in Kumar and Seppi (1994)), to insure themselves against the price they might obtain for their market orders (as in Chakravarty and Holden (1995)), to take advantage of their sufficiently persistent private information (as in Kaniel and Liu (2006) and Kalay and Wohl (2009)). There is also a more recent literature on dynamic limit order markets with strategic traders, such as Foucault, Kadan and Kandel (2005), Rosu (2009) and Goettler, Parlour, and Rajan. Foucault, Kadan and Kandel (2005) show that patient trades tend to submit limit orders while impatient ones submit market orders in equilibrium. Rosu (2009) shows that fully strategic, symmetrically informed liquidity traders can choose between market and limit orders based on their trade off between execution price and waiting costs. Goettler, Parlour, and Rajan find that limit orders tend to be submitted mostly by speculators and competition among them results in their private information being reflected in the limit order book.

Xetra electronic trading system. The information embedded in high frequency limit order book is quite rich and is not easy to summarize with a single variable. Instead, we consider variables such as measures of depth and slope that are theoretically motivated and capture different dimensions of the information embedded in the limit order book. Based on existing theoretical models, we then develop our hypotheses on how depth and slope of bid and ask sides should affect future returns. Specifically, we argue based on Goettler, Parlour and Rajan (2009) that an increase in depth at lower levels of the ask (bid) side results in lower (higher) future prices, while an increase in the depth at higher levels of the ask (bid) side results in higher (lower) future prices. Similarly, we expect based on Kalay and Wohl (2009) that an increase in the slope of the ask (bid) side results in higher (lower) future prices, regardless of the levels considered to measure it.

To test these hypotheses, we consider data in transaction, rather than calendar period, and calculate midquote returns and limit order book variables right after a trade, as in Hasbrouck (1991). We then estimate a linear vector autoregressive system (VAR) that includes midquote return, trade direction and each limit order book variable one at a time, while controlling for the contemporaneous effect of trade direction on returns and limit order book variables. This empirical specification allows us to analyze the effect of limit order book variables on return while still controlling for the effect of trade direction and autocorrelation in returns.

In this framework, we first focus on the coefficient estimates on the first lags of limit order book variables in the return equation, which reveal the initial effect of limit order book variables on return. Most limit order book variables considered in this paper have significant parameter estimates on their first lags in the return equation, even after controlling for lagged values of returns, trade directions and the limit order book variable itself. This in turn suggests that limit orders contain relevant information about future price movements in line with the predictions of recent theoretical models. More importantly, the coefficient estimates on the first lag of most limit order book variables have signs predicted by our hypotheses based on Goettler, Parlour and Rajan (2009) and Kalay and Wohl (2009). To be more precise, depth, especially of the ask side, has coefficient estimates of different signs on its first lag depending on the levels used to measure it, in line with our hypothesis based on Goettler, Parlour and Rajan (2009). On the other hand, slopes of both sides have coefficient estimates of the same sign regardless of levels used to measure them, confirming our hypothesis based on Kalay and Wohl (2009).

The coefficient estimates on the first lags of limit order book variables in the return equation provide information about the initial effects of these variables on return. However, they do not immediately reveal the long term cumulative effects of these variables on returns given that return, trade direction and limit order book variables have significant dynamics of their own. To take this into account, we calculate the impulse response functions of return to each limit order book variable. The impulse response functions suggest that the long term cumulative effects of most limit order book variables are generally in the same direction as their initial effects with few exceptions. For example, the long term cumulative effect of bid side depth depends closely on the levels considered. Bid side depth based on the first two, five and twenty levels have positive long term cumulative effects on return. On the other hand, bid side depth excluding the first level have negative long term cumulative effects on return. More importantly, our results on the long term cumulative effects of limit order book variables on return are mostly in line with our hypotheses and, thus, provide further empirical evidence from a different angle in their support.

We then analyze whether this relation is a causal one. To this end, we analyze causality between limit order book variables and return in the sense of Granger (1969). Most measures of ask and bid side depth as well as their ratio causes return to change at the next transaction period. The only exception is the bid side depth between the second and fifth levels, for which we fail to reject the null hypothesis that it does not cause return based on the F-statistic. Similarly, most measures of ask and bid side slope as well as their ratio causes return to change at the next transaction period. Once again, the exception to this is the slope measure between the second and fifth levels of ask and bid sides as well as their ratio. Furthermore, the test statistics tend to be much higher for limit order book variables measured based on lower levels (up to and including the fifth level) of the limit order book, suggesting that there is stronger empirical evidence for the relevance of information embedded in the lower levels of the limit order book.

Having analyzed the statistical relation between return and the information embedded in the limit order book, we then ask whether an investor can use this statistical relation to obtain economic gains. To this end, we consider an in-sample trading exercise similar to those considered in Kozhan and Salmon (2012). Specifically, having observed a transaction (and, thus, the return, trade direction and limit order book variables), we calculate our forecast of the midquote return at the next transaction period based on the estimated coefficients of the VAR model. We then calculate our trading signal for the next transaction period based on whether our forecast is greater or less than a threshold. In particular, we consider a forecast greater than a positive threshold to be a buy opportunity or signal and a forecast less than a symmetric negative threshold to be a sell opportunity or signal while a forecast between these two thresholds is not considered to be a strong enough signal. Tick-by-tick transaction prices tend to be quite noise. Hence, similar to a technical analysis trading strategy, we only consider trading signals when they are further confirmed by the relative movements of short- and long-run moving averages of recent transaction prices. To be more precise, we take a long position based on our trading signal, only when the short-run moving average crosses from below the long-run moving average by more than a specified amount. Similarly, we take a short position based on our trading signal, only when the short-run moving average crosses from above the long-run moving average by more than 0.06 euros. Otherwise, we do not consider a trading signal to be strong enough and, thus do not trade, if it is not confirmed by the relative movements of short- and long-run moving averages.

We start our trading exercise at the beginning of July 2010. We do not take any position until we observe a signal strong enough. When we receive such a signal, we then take a long or short position of one share depending on the signal. At every transaction period, we calculate our trading signal and re-evaluate our position. Specifically, if we have a long (short) position and receive a strong enough buy (sell) signal or a signal that is not strong enough, we continue to keep our long (short) position of one share. On the other hand, if we have a long (short) position and receive a strong enough sell (buy) signal, we then close our long (short) position and hold a short (long) position of one share. We continue in this fashion until the last transaction in July 2010 and keep the position in the last transaction till trading terminates in July 2010. We evaluate the performance of trading strategies based on their cumulative returns when we sell at bid and buy at ask prices.

Trading strategies based on most limit order book variables outperform a benchmark model that does not utilize the information embedded in the limit order book. To be more specific, trading strategies based on 22 out of 30 limit order book variables outperform the benchmark model. Of this 22 trading strategies, 21 provide positive returns. Most trading strategies based on ask side or ratio variables outperform the benchmark strategy while only half of the trading strategies based on bid side variables do so. Furthermore, trading strategies based on ask side variables tend to outperform those based on the corresponding bid side variable. Also, trading strategies based on variables that capture the information embedded in the lower levels of the limit order book tend to outperform those based on variables that do not include this information. Finally, these results hold when we consider a latency of 0, 500 or 1000 milliseconds, suggesting the robustness of our results to a more realistic assumption of 500 or 1000 milliseconds latency. As one would expect, profits based on a latency of 500 or 1000 milliseconds are lower than those based on a latency of 0 milliseconds for the same threshold parameter. Furthermore, profits tend to decrease monotonically as we consider higher latencies. However, we should note that the performance of trading strategies depends on a set of chosen parameters. There are some sets of parameters for which trading strategies using limit order book information outperform the benchmark and there are others for which the opposite holds. In other words, it is feasible to obtain economic gains using information embedded in the limit order book for some sets of parameters but this is not always the case.

Our paper is related to a growing body of papers which focus on the relation between high frequency returns and the information embedded in the limit order book. Biais, Hillion, and Spatt (1995) are among the first to analyze the dynamics of limit order markets and document many interesting facts. Specifically, they find that price revisions tend to move in the direction of previous limit order flows, suggesting that the limit order book contains some relevant information for the future path of prices. In contrast, Griffiths, Smith, Turnbull, and White (2000) find that limit orders tend to have a negative impact on prices in the Toronto Stock Exchange due to the possibility that limit orders can be ``picked off'' by better informed investors. This finding in turn suggests that limit orders are placed by less informed investors and, thus, do not convey much relevant information about prices. On the other hand, using data from the Australian Stock Exchange, Cao, Hansch, and Wang (2009) find that the limit order book is somewhat informative with a contribution of approximately 22% to price discovery. They also find that order imbalances between the demand and supply schedules along the book are significantly related to future short-term returns, even after controlling for the autocorrelations in return, the inside spread, and the trade imbalance. Similarly, using data from NYSE's TORQ (Trades, Orders, Reports, and Quotes), Kaniel and Liu (2006) find that the informed traders prefer limit orders to market orders and, thus, limit orders are more informative than market orders. More recently, Beltran-Lopez, Giot, and Grammig (2009) also finds that factors extracted from the limit order book have non-negligible information relevant for the long run evolution of prices in the German Stock Exchange. Specifically, they find that shifts and rotations of the order book can explain between 5% to 10% of the long run evolution of prices depending on the liquidity of the asset.

Our paper differs from the previous literature in many dimensions. First of all, most papers consider a single variable that is supposed to summarize all the information embedded in the limit order book. However, as we demonstrate in this paper, the dynamic relation between return, trade direction and the state of the limit order book is more complex than what can be captured by a single variable. Instead, we consider a wide range of variables that capture different dimensions of the information embedded in the limit order book. More importantly, we show that different variables, and even sometimes the same variable measured based on different levels of the limit order book, can have different effects on the short and long term dynamics of returns. Second, most of our variables are closely related to and motivated by the recent dynamic models of limit order markets and have not been previously considered in a similar context. Furthermore, we provide some preliminary empirical evidence in support of the causal relation between limit order book variables and return as implied by these theoretical models. Third, to the best of our knowledge, we are among the first to report the impulse response function of return to limit order book variables, which allows us to analyze the long term cumulative effect of limit order book variables on return after several transactions. Fourth, most papers focus on the statistical relation between return, trade direction and the state of the limit order book and do not provide much evidence whether this statistical relation is economically important for investors. In this paper, we do this by providing some preliminary evidence from a simple trading strategy based on the statistical relation between return, trade direction and the state of the limit order book. Finally, we should note that our data set is relatively unique and allows us to capture the state of the limit order book at a higher frequency than the ones that have been used in the previous empirical literature, providing a much finer analysis. Hence, part of our analysis can be considered as new out-of-sample tests of existing theoretical models and our findings provide empirical support for these models while the rest of our analysis is novel empirical evidence that might provide some guidance to new generation of models.

The rest of the paper is organized as follows. Section 2.2 presents the details of our data set. Section 2.3 discusses the variable definitions. Section 2.4 develops the empirical hypotheses based on theoretical models of limit order markets. Section 2.6 presents the coefficient estimates of limit order book variables in the return equation, Section 2.7 presents impulse response functions of return to limit order book variables and Section 2.8 presents results on whether limit order book variables Granger cause return to change. Section 2.9 analyzes the economic value of information embedded in the limit order book based on a trading strategy. Section 2.10 discusses out-of-sample results based on another company and time period. Section 2.11 discusses some further robustness checks. Section 2.12 concludes.

2.2 Data

Our data comes from the automated order-driven trading system Xetra operated by the Deutsche Börse Group at Frankfurt Stock Exchange (FSE). It is the main German trading platform accounting for more than 90% of total transactions at all German exchanges. The trading and order processing (entry, revision, execution and cancellation) of Xetra system is highly computerized. Since September 20, 1999, the normal trading hours are from 9h00 to 17h30 CET (Central European Time).

The raw dataset contains all events that are tracked and sent through the data streams. We first process the raw dataset using a software called XetraParser developed by Bilodeau (2013).[2] We then reconstruct the first twenty levels of the limit order book at millisecond time intervals. The limit order book can change when either a trade is executed or a limit order is placed, modified or canceled. In the unlikely event that these two types of events have the same millisecond time stamp, we need to make an assumption on the sequence of events given that we do not observe which one of them arrived earlier. We assume that a trade is always executed before any other change to the limit order book with the same millisecond time stamp. Thus, we first modify the limit order book to reflect the trade execution before taking its snapshot. In other words, if a trade is executed at a given millisecond, then the snapshot of the limit order book for that millisecond already reflects the executed trade. To avoid any problems due to this assumption, we ignore the state of the limit order book when a trade is executed and use its snapshots either one millisecond before or after a trade. Based on these snapshots, we measure variables summarizing the state of the limit order book every millisecond. We also eliminate all data corresponding to the three call auctions during a trading day since the price during these auctions is determined based on a certain set of rules and not by trading activity.

We consider data for two blue chip stocks in the DAX30 index, namely Merck (MRK) and SAP (SAP). The two stocks are in completely different industries: Merck is world's oldest operating chemical and pharmaceutical company while SAP specializes in enterprise software and related services to manage business operations and customer relations. Merck is a relatively small company in the DAX30 index with a market capitalization of approximately 4 billion Euros at the end of June 2010 corresponding to 0.75% of total market capitalization of stocks in the DAX30 index. On the other hand, SAP had market capitalization of approximately 33 billion Euros at the end of June 2010, almost 8 times that of Merck, representing 6.25% of total market capitalization of stocks in the DAX30 index. Panel (a) of Table 1 shows that SAP is much more actively traded than Merck with an average daily volume of approximately 8 times that of Merck in

[2] We thank Yann Bilodeau for his help in constructing the dataset and comments.

July 2010. Specifically, SAP and Merck had, respectively, average daily volumes of approximately 4 million and 0.5 million shares in July 2010, corresponding to around 15% and 2% of total trading volume of the stocks in the DAX30 index. However, Merck with an average share turnover of 0.8% in July 2010 seems to be more liquid than SAP which has an average share turnover of 0.3%. Fourth, Panel (b) of Table 1 shows that the characteristics of their daily returns are also quite different in July 2010. For example, Merck had a positive return while SAP had a negative return with relatively lower volatility than that of Merck. This is also reflected in the range of returns for both companies.

[Insert Table 1 here]

Similarly, we consider two relatively different time periods, July 2010 and June 2011. Both companies are much bigger in June 2011 with market capitalizations of almost 5 and 50 million for Merck and SAP, respectively. Trading volume for both companies are lower in June 2011 as presented in Panel (a) of Table 1. This is despite of an increase in the total trading volume for all the stocks in the DAX30 index, suggesting that these stocks are relatively less traded in June 2011 compared to July 2010. Both of the companies had negative returns in June 2011 with Merck having a relatively better performance and lower volatility than SAP.

Although we analyze both companies in both periods, we present empirical results for Merck in July 2010 as our main results due to space limitations. However, given the differences between companies and time periods discussed above, we consider other results as an out-of-sample test of our main results. To this end, we discuss these additional results, which are available in the appendix, in detail in Section 2.10.

2.3 Variable Definitions

The information embedded in the limit order book is quite rich and is not easy to summarize with a single variable. Hence, we consider several variables based on different levels of the limit order book to capture different dimensions of the information embedded in the limit order book. These variables can be categorized into two groups depending on whether they summarize information embedded in one or both sides of the limit order book.

We start with the variables that use information embedded in only one side of the limit order book. The first variable we consider is the depth between levels l_1 and l_2 of bid side , $D^B_{l_1,l_2,t}$, or the ask side, $D^A_{l_1,l_2,t}$, at period t. It is simply defined as the cumulative quantity available between levels l_1 and l_2 at period t:

$$D^B_{l_1,l_2,t} = \sum_{i=l_1}^{l_2} Q^B_{i,t}, \tag{2.1}$$

$$D^A_{l_1,l_2,t} = \sum_{i=l_1}^{l_2} Q^A_{i,t}, \tag{2.2}$$

for $l_1 = 1,\ldots,20$ and $l_2 \geq l_1$. $Q^B_{i,t}$ and $Q^A_{i,t}$ are the quantities available at the ith level of the bid and ask side in period t, respectively. The depth measures the cumulative demand and supply for the stock at different levels of demand and supply schedules, respectively. In other words, the higher is the bid side depth, the higher is the overall demand for the stock and the higher is the ask side depth, the higher is the overall supply of the stock.

The second variable is the slope of bid, $S^B_{l_1,l_2,t}$, or ask, $S^A_{l_1,l_2,t}$ sides between levels l_1 and l_2 at t. It is defined as the change in the price relative to the cumulative quantity available between levels l_1 and l_2 in period t:

$$S^B_{l_1,l_2,t} = \frac{P^B_{l_2,t} - P^B_{l_1,t}}{D^B_{l_1+1,l_2,t}} \tag{2.3}$$

$$S^A_{l_1,l_2,t} = \frac{P^A_{l_2,t} - P^A_{l_1,t}}{D^A_{l_1+1,l_2,t}} \tag{2.4}$$

for $l_1 = 1,\ldots,19$ and $l_2 > l_1$. $P^B_{l,t}$ and $P^A_{l,t}$ are the lth best bid and ask prices, respectively in period t. The slope of the bid side is a measure of price sensitivity to changes in quantity demanded and is always negative. A high (in absolute value) bid-side slope coefficient implies that the bid price will decrease more for a given change in quantity demanded. Ceteris paribus, this in turn suggests that the investors are willing to buy at lower prices for the same total quantity demanded. Similarly, the slope of the ask side is a measure of price sensitivity to changes in quantity supplied and is always positive. A high ask-side slope coefficient implies that the ask price will increase more for a given change in quantity supplied, which in turn suggests that the investors are willing to sell at higher prices for the same total quantity supplied.

To combine information embedded in both sides of the limit order book, we consider the ratios of the variables introduced above. Specifically, the depth ratio is simply defined as the (normalized) difference in the cumulative quantity available at different levels of bid and ask sides:

$$DR_{l_1,l_2,t} = \frac{D^A_{l_1,l_2,t} - D^B_{l_1,l_2,t}}{D^A_{l_1,l_2,t} + D^B_{l_1,l_2,t}} \quad (2.5)$$

for $l_1 = 1,\ldots,20$ and $l_2 > l_1$. This variable is bounded above by one and below by minus one. Positive values of depth ratio indicate that the total supply of the stock between two given levels of the ask side is greater than the total demand between the same two levels of the bid side. Furthermore, as the quantity supplied increases relative to the quantity demanded between the same two levels of the bid and ask sides, the depth ratio increases. Similarly, we define the slope ratio as the (normalized) difference in the slopes of bid and ask sides between different levels:

$$SR_{l_1,l_2,t} = \frac{S^A_{l_1,l_2,t} - |S^B_{l_1,l_2,t}|}{S^A_{l_1,l_2,t} + |S^B_{l_1,l_2,t}|} \quad (2.6)$$

for $l_1 = 1,\ldots,19$ and $l_2 > l_1$. We take the absolute value of the bid slope since it is always negative by definition. This variable is also bounded between one and minus one. Positive values of slope ratio indicate that the supply schedule is steeper than the demand schedule. Furthermore, as the supply schedule becomes steeper relative to the demand schedule the same two levels of the bid and ask sides, the slope ratio increases.

[Insert Table 2.2 here]

Figure 3.1 presents two snapshots of the limit order book for Merck in July 2010. As it can easily be seen from Figure 3.1, both bid and ask sides of the limit order book can take on different shapes at different points in time. The limit order book variables are designed to summarize these different shapes that the limit order book can take on. To see what information different limit order book variables capture, Table 2.2 presents the values of the limit order book variables that correspond to these two snapshots. For example, it is easy to see from Figure 3.1 that there is more total demand and less total supply in the first snapshot compared to the second one. These facts are captured by $D^{B}_{1,20}$ and $D^{A}_{1,20}$. Furthermore, there seems to be more total demand and less total supply in the higher levels of the limit order book in the first snapshot compared to the second one. This can indeed be verified by comparing $D^{B}_{5,20}$ and $D^{A}_{5,20}$ of the first snapshot to those of the second one. As another example, compare the slopes of the bid side in these two snapshots. The bid side in the first snapshot appears to be overall steeper than that in the second snapshot, which is confirmed by comparing $D^{B}_{1,20}$'s. On the other hand, the bid side of the first snapshot seems to be flatter compared to the second snapshot when we focus on the first two and five levels. This turns out to be the case when we compare $S^{B}_{1,2}$ and $S^{B}_{1,5}$ of the two snapshots. Similarly, the ratio variables allow us to infer about the shapes of bid and ask sides with respect to each other. For example, the bid side appears to be overall steeper than the ask side in the first snapshot, which is confirmed by the fact that $SR_{1,20,t}$ is negative.

[Insert Figure 2.1 here]

Table 2.3 presents the transformations applied to the limit order book variables and summary statistics for the transformed limit order book variables as well as log returns. Although not presented in Table 2.3, we also analyze whether any of the limit order book variables have unit root based on augmented Dickey-Fuller test. We reject the existence of a unit root for all limit order book variables at 1% statistical significance level, suggesting the stationarity of these variables.

[Insert Table 3.3 here]

2.4 Hypotheses Development

In this section, based on recent theoretical models of limit order markets, we develop our hypotheses on the effect of limit order book variables on return. We first consider the effect of depth on return. Goettler, Parlour, and Rajan (2009) develop a theoretical model where traders optimally choose the type of order to submit and whether to acquire information about the asset. They solve for the equilibrium of this model and show that depth at different levels of the limit order book should not only be informative about future prices but also have different effects on them. Specifically, they show that there are only a few stale orders in the book since traders submitting limit orders revisit the market and resubmit orders, on average, twice as often as the true value of the asset changes. Thus, orders submitted in the higher levels of ask (bid) side suggest that the current best ask (bid) is too low (high) and hence lead to an upward (downward) revision in expectations about the true value of the asset. On the other hand, given that the transactions prices and traders' prices are, on average, equal to the true value of the asset, depth at lower levels of ask (bid) side lead to lower (higher) prices. Let us denote our hypothesis for the ask side with *a* and that for the bid side with *b*. Then, our first hypotheses on the effect of depth on return can be summarized as follows:

Hypothesis 1a (Goettler, Parlour, and Rajan): An increase in depth at lower levels of the ask side results in lower future prices, while an increase in the depth at higher levels of the ask side results in higher future prices.

Hypothesis 1b (Goettler, Parlour, and Rajan): An increase in depth at lower levels of the bid side results in higher future prices, while an increase in the depth at higher levels of the bid side results in lower future prices.

We then consider the effect of slope on return. Kalay and Wohl (2009) analyze the relation between slope and future returns in the Noisy Rational Expectations Equilibrium (NREE) models of Hellwig (1980), Kyle (1989), Admati (1985), and Easley and O'Hara (2004). Specifically, they consider the framework of Hellwig (1980) and solve for the equilibrium. Given their assumption that limit orders can only be submitted by informed traders, as the number of informed traders on the bid side increases, the bid side becomes flatter, or equivalently, the slope of the bid side decreases. An increase in the number of informed traders on the bid side also implies a decrease in the relative importance of liquidity traders on the bid side. This in turn results in a decrease in the future price of the stock. The opposite model holds for the ask side. Thus, our hypotheses can be summarized as follows:

Hypothesis 2a (Kalay and Wohl (2009)): An increase in the slope of the ask side results in higher future prices.

Hypothesis 2b (Kalay and Wohl (2009)): An increase in the slope of the bid side results in lower future prices.

We should note that the predictions of Kalay and Wohl (2009) on the effect of slope on return do not depend on the levels of the limit order book used to measure it, unlike the predictions of Goettler, Parlour, and Rajan for the effect of depth on return. However, we still consider slope measured based on different levels of the limit order book to analyze whether this prediction of Kalay and Wohl (2009) holds.

We now consider the effect of depth and slope ratios on return. Recall that the depth ratio is defined as the normalized difference between ask and bid side depths. Given our hypotheses 1a and 1b, one would expect to observe an effect of depth ratio on return similar to but stronger than those proposed in Hypothesis 1a and 1b. Similarly, recall that the slope ratio is defined as the normalized difference between ask and bid side slopes. Given our hypotheses 2a and 2b, one would also expect to observe an effect of slope ratio on return similar to but stronger than those proposed in Hypothesis 2a and 2b. Let *c* denote the combined versions of hypotheses a and b. Then our hypotheses on the effect of depth and slope ratios on return can then be summarized as follows:

Hypothesis 1c (Goettler, Parlour, and Rajan): An increase in depth ratio based on lower levels of the limit order book results in lower future prices while an increase in the depth ratio based on higher levels results in higher future prices.

Hypothesis 2c (Kalay and Wohl (2009)): An increase in the slope ratio results in higher future prices.

2.5 The Empirical Model

In this section, we discuss the empirical model that we use to test our hypotheses. To this end, we follow Hasbrouck (1991) who shows that a vector autoregressive system for the interactions between return and trade direction is consistent with stylized market microstructure models such as Glosten and Milgrom (1985). Specifically, Hasbrouck (1991) suggests using the following vector autoregressive model to analyze the effects of information embedded in trades on prices:

$$r_t = \sum_{\tau=1}^{\infty} \alpha_{r,\tau} r_{t-\tau} + \sum_{\tau=1}^{\infty} \alpha_{x,\tau} x_{t-\tau} + \varepsilon_{r,t} \quad (2.7a)$$

$$x_t = \sum_{\tau=1}^{\infty} \beta_{r,\tau} r_{t-\tau} + \sum_{\tau=1}^{\infty} \beta_{x,\tau} x_{t-\tau} + \varepsilon_{x,t} \quad (2.7b)$$

where *t* indexes trades, x_t is the sign of the trade in period *t* (+1 for a trade initiated by a buyer and -1 for a trade initiated by a seller), r_t is the midquote return defined as the change in the average of best bid and ask quotes between period *t-1* and *t*, i.e. $r_t = \Delta q_t = q_t - q_{t-1}$ and q_t is the simple average of best bid and ask quotes in period *t*. This is a very general and flexible model that nests many of the standard microstructure models as special cases. The disturbances in this framework, $\varepsilon_{r,t}$ and $\varepsilon_{x,t}$, are generally modeled as white noise processes and can be interpreted as public information embedded in unexpected returns and private information embedded in unexpected trades, respectively.

In this paper, we assume that the dynamics of a limit order market can also be approximated by a linear vector autoregressive system similar to that proposed by Hasbrouck (1991). Specifically, we consider each variable summarizing different limit order book-related information separately as a third state variable in the VAR in addition to return and trade direction:

$$r_t = \sum_{\tau=1}^{\infty} \alpha_{r,\tau} r_{t-\tau} + \sum_{\tau=1}^{\infty} \alpha_{x,\tau} x_{t-\tau} + \sum_{\tau=1}^{\infty} \alpha_{z,\tau} z_{t-\tau} + \varepsilon_{r,t} \quad (2.8a)$$

$$x_t = \sum_{\tau=1}^{\infty} \beta_{r,\tau} r_{t-\tau} + \sum_{\tau=1}^{\infty} \beta_{x,\tau} x_{t-\tau} + \sum_{\tau=1}^{\infty} \beta_{z,\tau} z_{t-\tau} + \varepsilon_{x,t} \quad (2.8b)$$

$$z_t = \sum_{\tau=1}^{\infty} \gamma_{r,\tau} r_{t-\tau} + \sum_{\tau=1}^{\infty} \gamma_{x,\tau} x_{t-\tau} + \sum_{\tau=1}^{\infty} \gamma_{z,\tau} z_{t-\tau} + \varepsilon_{z,t} \quad (2.8c)$$

where z_t is a variable that summarizes a certain dimension of the information embedded in the limit order book.

This specification can be considered as a reduced form linear approximation that is designed to capture dynamics of limit order market models discussed in the introduction. Secondly, it is flexible and allows us to analyze the effect of limit order book-related information on prices while still controlling for trade-related information. For example, the immediate effect of limit order book-related information on prices is captured by $\alpha_{z,1}$. Finally, Goettler, Parlour and Rajan (2009) argue that competition among speculators results in their private information being partially revealed in the limit order book. Hence, $\varepsilon_{z,t}$ in this framework can be interpreted as an unexpected private information shock embedded in the limit order book variable of interest.

As mentioned above, we can measure limit order book variables including best bid and ask prices every millisecond. However, a trade can only be matched to a millisecond interval and thus one needs to decide whether to take a snapshot of the limit order book right before or right after a trade. The theory does not provide much guidance on this issue. In this paper, we follow the previous literature, e.g. Hasbrouck (1991) and Dufour and Engle (2000), and measure the limit order book variables right after a trade occurs. [3]This sampling approach implies that the midquote return and limit order book variables in period *t* are observed right after (less than a millisecond after) the trade, and thus trade direction. Thus, one can include trade direction in period *t* to control for its contemporaneous effect on return and limit order book variables in the estimated version of (2.8). Our results are similar regardless of whether or not we control for this contemporaneous effect. We choose to present results on the statistical relation between return and limit order book variables based on the specification that includes this contemporaneous effect.

However, we should note that any contemporaneous effects of trade direction on return and limit order book variables cannot be interpreted as a causal relation even if trade direction is observed right before these variables. This is mainly due to the fact that market participants (human or non-human) cannot be possibly reacting to and trading based on any information embedded in contemporaneous return and limit order book

[3] We also considered the alternative sampling approach of measuring the limit order book variables right before a trade occurs. Our results on the effects of limit order book variables on return do not change significantly when consider this alternative sampling approach.

variables which are only observed less than a millisecond before the trade occurs. Xetra reports that the average time required for an order to travel from the trading participant's system across the network to its back-end and for confirmation of receipt to be sent back to the participant is about 13 milliseconds.[4] Assuming that the two legs of this round trip are equally fast, it takes about 6.5 milliseconds for an order to travel from the trading participant's system across to the Xetra back-end. Hence, even if we make the unrealistic assumption that any necessary computations of an algorithmic trading strategy or the reaction of a human trader take less than a millisecond, it is physically impossible for their orders to arrive at the market within a millisecond. Hence, we choose to present our results on the trading strategy based on the empirical specification that does not include the contemporaneous effect of trade direction on return and limit order book variables.

Following Hasbrouck (1991) and Dufour and Engle (2000), we also assume that the infinite sums in the model in Equation (2.8) can be truncated at J lags. Furthermore, the timing convention discussed above is reflected in the starting points of the summations in the estimated version of (2.8). pecifically, the summations for trade direction in the equations for return and limit order book variables start at zero instead of one. Then, the estimated version of the model in Equation (2.8) can be expressed as follows:

$$r_t = \sum_{\tau=1}^{J} \alpha_{r,\tau} r_{t-\tau} + \sum_{\tau=0}^{J} \alpha_{x,\tau} x_{t-\tau} + \sum_{\tau=1}^{J} \alpha_{z,\tau} z_{t-\tau} + \varepsilon_{r,t} \tag{2.9a}$$

$$x_t = \sum_{\tau=1}^{J} \beta_{r,\tau} r_{t-\tau} + \sum_{\tau=1}^{J} \beta_{x,\tau} x_{t-\tau} + \sum_{\tau=1}^{J} \beta_{z,\tau} z_{t-\tau} + \varepsilon_{x,t} \tag{2.9b}$$

$$z_t = \sum_{\tau=1}^{J} \gamma_{r,\tau} r_{t-\tau} + \sum_{\tau=0}^{J} \gamma_{x,\tau} x_{t-\tau} + \sum_{\tau=1}^{J} \gamma_{z,\tau} z_{t-\tau} + \varepsilon_{z,t} \tag{2.9c}$$

Hasbrouck (1991) and Dufour and Engle (2000) consider $J=5$ assuming that five lags are sufficient to capture the dynamics of the variables of interest. In this paper, we

[4] http://deutscheboerse.com/dbg/dispatch/en/binary/gdb_content_pool/imported_files/public_files/10_downloads/31_trading_member/10_Products_and_Functionalities/20_Stocks/BR_Xetra_Speed.pdf

consider different lag structures up to a maximum of five lags. To be consistent with the previous literature, we present results based on a lag structure of five lags.[5]

We estimate the empirical model via ordinary least squares (OLS), which provides consistent parameter estimates for return and limit order book variables given that they are stationary. OLS also yields consistent estimates for the parameters in the trade equation even though we estimate a linear specification for a limited dependent variable. As discussed in detail in Dufour and Engle (2000), this is mainly due to the fact that the conditional mean of the trade sign is generally correctly specified given that the probability of a buy or sell is never far from 1/2. A similar argument holds for depth and slope ratios. Specifically, the OLS yields consistent estimates for the parameters in the equations for depth and slope ratios even though they are limited between -1 and +1. This is again due to the fact that their conditional mean is correctly specified since their means are not far from zero. However, these estimates are inefficient and standard errors are biased. Hence, we present heteroskedasticity and autocorrelation consistent standard errors (Newey and West (1987)). The results are presented in Table 2.4.

2.6 Coefficient Estimates

We start with the effect of ask side variables on returns. Ask side depth measured based on the first two, five and twenty levels have significantly negative coefficient estimates on their first lags. On the other hand, ask side depth between the second and fifth levels has a significantly positive coefficient estimates on its first lag. These results provide some empirical evidence in support of Hypothesis 1a that an increase in depth at lower levels of the ask side results in lower future prices at the next transaction period while an increase in the depth at higher levels results in higher future prices. Furthermore, ask side depth between the fifth and twentieth levels has a statistically insignificant coefficient estimate on its first lag. This in turn suggests that any information, that is relevant for future prices, embedded in the ask side depth is mostly available at the lower levels of the ask side. These results are not only statistically but also economically significant. To understand the economic magnitude of these coefficient estimates, one

[5] Results based on the model estimated based on different numbers of lags are similar to those presented in the paper and available from the authors upon request.

needs to take into account the fact that we consider the logarithm of the ask side depth in the VAR system. For example, assuming that everything else remains constant, the log stock price decreases by 0.18 basis points ($(-0.180 b.p. = -0.260 \times \log(2))$) at the next trade following a two-fold increase in the depth of the first two levels of the ask side. Although this might look like an economically insignificant change at first sight, it is indeed more than 18 times the mean log return between two trades, which is around 0.01 basis points.

[Insert Table 2.4 here]

Different measures of ask side slope have positive and significant coefficient estimates on their first lags with the exception of the slope between second and fifth levels of the ask side. This in turn confirms Hypothesis 2a that an increase in the slope of the ask side results in higher future prices. The economic magnitude of this effect is a 0.09 ($0.132 \times \log(2)$), 0.13 ($0.189 \times \log(2)$) and 0.40 ($0.575 \times \log(2)$) basis points increase in the price at the next trade following a two-fold increase in the slope of the first two, five and twenty levels of the ask side, respectively. Once again, considering that the mean log return between two trades is around 0.01 basis points, these effects are economically important. Furthermore, these results suggest that there is relevant information for future prices not only in lower but also in higher levels of the ask side in terms of its slope, as suggested by the significant coefficient estimate on the first lag of the ask side slope between the fifth and twentieth levels.

We now consider the bid side variables. The coefficient estimates on the first lag of most bid side variables have the opposite signs of those on the first lag of the corresponding ask side variable. This is in line with our hypotheses in Section 2.4. Furthermore, the economic interpretations of the coefficient estimates on the bid side variables are similar to those on the corresponding ask side variables. Thus, we only briefly discuss our results on the bid side variables.

First, the coefficient estimates on the first lag of bid side depth in the first two and five levels are both significantly positive. This in turn suggests that an increase in depth at lower levels of the bid side results in higher future prices at the next transaction period, in line with the first part of our Hypothesis 1b. However, the coefficient estimates on the first lag of depth in the higher levels of the bid side are either significantly positive for $D^{B}_{1,20}$ and $D^{B}_{5,20}$ or negative but insignificant for $D^{B}_{2,5}$. In other words, these coefficient estimates do not provide any supporting evidence for the second part of our Hypothesis 1b. Second, different measures of bid side slope have negative and significant coefficient estimates on their first lags with the exception of the slope between second and fifth levels of the ask side. These provide empirical evidence in support of Hypothesis 2b that an increase in the slope of the bid side results in lower future prices. Finally, the information embedded in both lower and higher levels of the bid side is relevant for future prices as suggested by the significant coefficient estimates on the first lag of bid side variables based on both lower and higher levels.

We now turn our attention to the variables that use information embedded in both sides of the limit order book. The coefficient estimates on the first lag of depth ratio are all negative regardless of the levels considered. These estimates are statistically significant with the exception of the depth ratio between second and fifth levels of the limit order book. These results suggest that an increase in depth ratio results in a lower future price regardless of the levels considered, providing evidence for the first part of Hypothesis 1c but against its second part. On the other hand, the coefficient estimates on the first lag of slope ratio are all significantly positive, with the exception of the slope ratio between the second and fifth levels of the limit order book. These results in turn suggest that an increase in the slope ratio results in higher future prices, providing evidence in support of Hypothesis 2c. Given that these variables are not transformed like variables based on one side of the limit order book, their economic interpretation is straightforward. For example, the log stock price decreases by 14 basis points at the next trade following a 10% increase in the depth ratio and increases by 19 basis points following a 10% increase in the slope ratio.

2. 7 Impulse Response Functions

So far, we have focused on the coefficient estimates on the first lags of limit order book variables in the return equation. These coefficient estimates reveal the initial effects of each limit order book variable on return but not their long term cumulative effects. Given that most of the variables in the VAR system have also significant dynamics of their own, one needs to calculate the impulse response functions of return to limit order book variables to infer about long term cumulative effects of limit order book variables on returns. To do this, we first simulate the estimated VAR system for a long enough period setting all residual terms to zero, i.e. $\varepsilon_{z,t} = \varepsilon_{r,t} = \varepsilon_{x,t} = 0$, to obtain its steady state. Second, starting with the steady state, we simulate the VAR system once again but this time assuming that the initial residual of the limit order book variable of interest is +1 while all other residuals (initial or future) remain zero. The difference between these two simulations of the VAR system is the impulse response function of return to a unit shock to the limit order book variable of interest.

Figure 2.2 presents the impulse response functions of return to limit order book variables. The cumulative effect of an increase in depth between any levels of the ask side is negative. This is in line with the initial effects of ask side depth on return discussed above based on the coefficient estimates on the first lags. The only exception is the ask side depth between second and fifth levels, which has a significantly positive initial effect but a negative long term cumulative effect. These results provide further supporting empirical evidence for the first part of Hypothesis 1a. However, they also suggest that the second part of Hypothesis 1a might not hold in the data when one considers the long term cumulative effect of ask side depth rather than its initial effect.

[Insert Figure 2.2 here]

In contrast to ask side depth, the long term cumulative effect of bid side depth depends on the levels considered. On one hand, bid side depth based on the first two, five and twenty levels have positive long term cumulative effects on return, although their cumulative effects in the short and medium term tend to be higher than those in the long term. These findings are in line with the initial effects of these variables on return and provide further empirical support for the first part of Hypothesis 1b. On the other hand, measures of bid side depth that exclude the first level have negative long term cumulative effects on return, although they have positive initial effects as well as positive cumulative effects in the short and medium term. These findings provide empirical evidence in support of the second part of Hypothesis 1b, unlike the coefficient estimates discussed above. Hence, these impulse response functions of return to bid side depth based on different levels suggest that Hypothesis 1b holds in the data when one considers long term cumulative effects rather than initial or short to medium run effects.

We now turn our attention to the long term cumulative effects of slope measures. Ask side slope has a positive long term cumulative effect regardless of the levels considered to measure it. These results are in line with their initial effects discussed above based on the coefficient estimates on their first lags in the return equation and provides further empirical evidence in support of Hypothesis 2a. On the other hand, the long term cumulative effects of bid side slope measures are negative with the exception of bid side slope between the second and fifth levels. This is mostly in line with Hypothesis 2b but also suggests that bid side slope might have different long term cumulative effects on returns depending on the levels considered to measure it, in contrast to its initial effect. Finally, we consider the long term cumulative effects of depth and slope ratios on return. Regardless of the levels considered, depth ratio has a negative long term cumulative effects on return. This provides further empirical evidence for the first part of Hypothesis 1c. However, it also provides some evidence contrary to the second part of Hypothesis 1c. This in turn suggests that Hypothesis 1c might not hold when we consider the long term cumulative effects rather than initial effects. Measures of slope ratio based on different levels of the limit order have positive long term cumulative effects on return, providing further empirical support for Hypothesis 2c.

The impulse response functions can be economically interpreted as the change in the log price of the stock in basis points as a function of transaction periods in response to a one unit positive shock to the limit order book variable of interest while taking into account its transformation. For example, log stock price decreases by almost 0.1 basis points in the long term following almost a three-fold increase[6] in between the first and fifth levels of the ask side. Furthermore, the long term cumulative effects of limit order book

[6] Note that a one unit positive shock to a limit order book variable with a log transformation implies e($\approx$ 2.718) -fold increase in the limit order book variable of interest.

variables based on higher levels of the limit order book tend to be stronger and take longer to be fully realized. This suggests that there might be more information (relevant for returns) embedded in the higher levels of the limit order book but it takes longer for this information to be fully incorporated in prices.

2.8 Granger Causality

We have discussed whether limit order book variables have any significant effect on returns based on coefficient estimates and impulse response functions. However, we have not yet answered whether the state of the limit order book causes returns to change at the next transaction period. In this section, we address this question by analyzing the causal effect of limit order book variables on returns.

To do this, we consider the statistical test of causality in the sense of Granger (1969). Specifically, consider that we are interested in whether a given limit order book variable Granger causes return. We run a regression of return on its own five lags and five lags of the limit order book variable. We then conduct an F-test of the null hypothesis that the coefficients on the lags of the limit order book variable are jointly zero. Rejecting this null hypothesis suggests that the limit order book variable causes the return. To conduct the test, we consider the asymptotic F-statistic that has a χ^2 distribution with 5 degrees of freedom asymptotically. Last column of Table 2.4 presents statistics testing whether limit order book variables Granger cause return.

Our results can be summarized as follows: Most measures of ask and bid side depth as well as their ratio causes return to change at next transaction period. The only exception is the bid side depth between the second and fifth levels, for which we fail to reject the null hypothesis that it does not cause return based on the F-statistic. Similarly, most measures of ask and bid side slope as well as their ratio causes return to change at the next transaction period. Once again, the exception to this is the slope measure between the second and fifth levels of ask and bid sides as well as their ratio. Furthermore, the test statistics tend to be much higher for limit order book variables measured based on lower levels of the limit order book, suggesting that there is stronger empirical evidence for the relevance of information embedded in the lower levels of the limit order book.

To sum up, these results provide further empirical evidence that the state of the limit order book affects returns at the next transaction period, although it is hard to interpret them in terms of individual hypotheses discussed in Section 2.4 as they do not provide any information on the directional effect of limit order book variables on return.

2.9 Economic Value of the Information in the Limit Order Book

So far, we have discussed the statistical relation between returns and the information embedded in the limit order book. In this section, we analyze whether an investor can use this significant statistical relation to obtain economic gains. To this end, we consider simple trading strategies similar to those discussed in Kozhan and Salmon (2012).

Specifically, at each transaction period *t*, we first obtain forecast of the midquote return at the next transaction period *t+1*, $\hat{r}_{t+1}^{(j)}$, based on a restricted version of the VAR model in Equation (9a) that excludes the contemporaneous effect of trade direction on return.[7] We then calculate our trading signal for the transaction period based on whether our forecast is greater or less than a threshold. In particular, we consider a forecast greater than a positive threshold, i.e. $\hat{r}_{t+1}^{(j)} > \kappa$ where $\kappa > 0$, to be a buy opportunity or signal and a forecast less than a symmetric negative threshold, i.e. $\hat{r}_{t+1}^{(j)} < -\kappa$ where $\kappa > 0$, to be a sell opportunity or signal while we do not trade on any forecast between these two thresholds, i.e. $|\hat{r}_{t+1}^{(j)}| \leq \kappa$. Similar to Kozhan and Salmon (2012), κ can be considered as a parameter to filter out potentially weak signals. In this paper, we consider a κ of 1 basis points. As we consider higher values for κ, we trade less frequently.

In addition, we remove further noise from the transaction data and, thus from our trading strategy, using a moving average filter. Specifically, for each transaction period *t*, we calculate a short-run moving average based on the last three and a long-run moving average based on the last forty midquote prices including the midquote price at

[7] We also consider trading strategies based on a forecasting model that includes the contemporaneous effect of trade direction on returns. As discussed above, any contemporaneous effect of trade direction on return is not a causal relation even if trade direction is observed right before the return is calculated. Our results based on this VAR model are similar to those based on the restricted VAR model and, thus, are not presented.

transaction period *t*. Similar to a technical analysis trading strategy, we consider the relative movement of the short and long-run moving averages to confirm the strength of our trading signal discussed above. To be more precise, we take a long position based on our trading signal discussed above, only when the short-run moving average crosses from below the long-run moving average by more than 0.06 euros, which correspond to approximately 10 basis points for MRK in July 2010. Similarly, we take a short position based on our trading signal discussed above, only when the short-run moving average crosses from above the long-run moving average by more than 0.06 euros. Otherwise, we do not consider a trading signal to be strong enough and, thus do not trade, if it is not confirmed by the relative movements of short- and long-run moving averages.

We start our trading exercise at the beginning of July 2010. We do not take any position until we observe a signal strong enough based on both of the filters discussed above. When we receive such a signal, we then take a long or short position of one share depending on the signal. At every transaction period, we calculate our trading signal and reevaluate our position. Specifically, if we have a long (short) position and receive a strong enough buy (sell) signal or a signal that is not strong enough, we continue to keep our long (short) position of one share. On the other hand, if we have a long (short) position and receive a strong enough sell (buy) signal, we then close our long (short) position and hold a short (long) position of one share. We continue in this fashion until the last transaction in July 2010 and keep the position in the last transaction till trading terminates in July 2010.

Several remarks are in order about our trading strategy. First of all, we implicitly assume that our trading does not alter the dynamics of the relation between return, trade direction and limit order book. We believe that this is a reasonable assumption since we only consider trading one share at a time which should be negligible given the trading volume of Merck in July 2010. Secondly, it is an in-sample trading exercise since we use the coefficient estimates based on the whole sample to obtain our return forecasts at each transaction period. An out-of-sample trading exercise would require the estimation of the VAR model at each transaction period based on the information available only up to that transaction period, which is computationally quite intensive. Finally, in our trading exercise, we calculate our signal based on the snapshot of the limit order book right after (less than a millisecond after) a transaction and assume that we can trade at the prices in that snapshot of the limit order book. However, in reality, the computation of our trade signal after a transaction and the processing of our order at the exchange are not instantaneous. Hence, we might end up trading at prices different than those in the snapshot of the limit order book right after a transaction. Hence, we perform the same exercise when we assume a latency of 500 and 1000 milliseconds and trade at prices observed 500 and 1000 milliseconds after observing a transaction. This exercise allows us not only to check the robustness of our results to a more realistic assumption about latency but also to analyze the effect of speed on the profitability of our trading strategies. Given that our forecasting model is designed to predict returns with a zero delay, we expect profits based on a latency of zero millisecond to be higher than those based on a latency of 500 and 1000 milliseconds.

We consider different trading strategies based on different limit order book variables as separate predictors in the forecasting model in Equation (9). We evaluate the performance of these trading strategies based on their cumulative returns over the month. We also distinguish between returns based on trading at midquote and bid and ask prices. Returns based on midquote prices might not necessarily be attainable by traders. Returns based on bid and ask prices, on the other hand, provide a more realistic performance measure for the trading strategy since we consider actual prices at which the traders can buy and sell the stock. Thus, we present results based on trading at bid and ask rather than midquote prices.

As a benchmark, we consider a trading strategy based on a forecasting model that ignores information embedded in limit order book variables. Specifically, we obtain our trade signals in the benchmark trading strategy based on the estimation of the empirical model in Equation (9) with the restriction that $\alpha_{z,\tau} = 0 \;\; for \;\; \tau = 1,\ldots,5$. This benchmark model allows us to analyze whether limit order book variables provide additional information useful to the trader above and beyond what is embedded in past returns and trade direction. One can think of this benchmark strategy as the equilibrium strategy of a trader who faces barriers to entry. For example, a trader might want to use the benchmark strategy if she cannot process or have access to the limit order book due to technological or financial reasons. Furthermore, we also consider this trading strategy with and without applying the moving average filter. This allows us to have an idea about the effect of the moving average filter on the profits from this trading strategy.

[Insert Table 2.5 here]

Table 2.5 presents returns from trading strategies when we sell at bid and buy at ask prices as well as the number of trades required to implement these trading strategies. We start by comparing results based on the benchmark strategies with and without the moving average filter. This allows us to provide some intuition on why we apply the additional moving average filter. First of all, as expected, the number of trades decreases significantly from 270 to 96 when we apply the moving average filter. Furthermore, although still negative, returns based on the benchmark strategy increases significantly, regardless of the assumed latency, when we apply the moving average filter. The negative returns on the benchmark strategy without the moving average filter is due to the large number of trades required to implement this strategy and the loss associated with each round trip transaction due to the bid-ask spread. Kozhan and Salmon (2012) also report large negative returns from a trading strategy based on a linear model, with a daily cumulative return of as low as -92% (corresponding to a compound monthly return of -100% assuming that there are 22 trading days in a month) based on actual bid and ask prices. As discussed above, these findings are due to the fact that transaction data and, thus, trading returns are noisy and the additional filter based on the difference between short- and long-run moving averages allows to us to decrease this noise further.

More importantly, most trading strategies based on limit order book information outperform the benchmark model. To be more specific, trading strategies based on 22 out of 30 limit order book variables outperform the benchmark model. Of this 22 trading strategies, 21 provide positive returns. Most trading strategies based on ask side or ratio variables outperform the benchmark strategy while only half of the trading strategies based on bid side variables do so. Furthermore, trading strategies based on ask side variables tend to outperform those based on the corresponding bid side variable. Also, trading strategies based on variables that capture the information embedded in the lower levels of the limit order book tend to outperform those based on variables that do not include this information. Finally, these results hold when we consider a latency of 0, 500 or 1000 milliseconds, suggesting the robustness of our results to a more realistic assumption of 500 or 1000 milliseconds latency. As one would expect, profits based on a latency of 500 or 1000 milliseconds are lower than those based on a latency of 0 milliseconds for the same threshold parameter. Furthermore, profits tend to decrease monotonically as we consider higher latencies. On the other hand, the relative performance of strategies based on limit order book variables with respect to the benchmark strategy generally increases as the latency increases from 0 millisecond to 1000 milliseconds, suggesting that the information embedded in the limit order book can still be economically important for traders with a relatively high latency.

Several remarks are in order concerning these results. First of all, there are four parameters that we had to choose to operationalize our trading strategy: the time period for short- and long-run moving averages, the threshold parameters for the moving average filter and our trading signal. We chose parameters which we considered to be reasonable. However, we should note that they were still chosen in an ad hoc fashion as there is no theoretical model that provides any guidance on how to choose these parameters. We considered other sets of parameters and results undoubtedly change. There are some sets of parameters for which trading strategies using limit order book information continue to outperform the benchmark and there are others for which this is no longer the case. In other words, our results suggest that it is feasible to obtain economic gains using information embedded in the limit order book for some sets of parameters but this is not always the case.

2.10 Robustness Checks based on Another Company and Time Period

In this paper, we have focused on a single company, i.e. Merck, in a single month, i.e. July 2010. However, our findings are not due to our choice of company and time period. To demonstrate this, we have also analyzed the effect of limit order book variables on return for Merck in June 2011 and SAP in July 2010 and June 2011. In this section, we briefly summarize these additional findings which are available in the appendix.

First of all, regardless of the company and time period, most limit order book variables have significant coefficient estimates on their first lags in the return equation and these coefficient estimates have the same signs and similar magnitudes as our benchmark results. Furthermore, similar to our benchmark results, the coefficient estimates on higher order lags of most limit order book variables are not significantly different than zero. Second, the long term cumulative effects of most limit order book variables on returns are similar to those presented in Figure 2.2. Third, similar to our benchmark results, most limit order book variables significantly cause return to change at the next transaction period. Finally, the results from the trading exercise are also similar to our benchmark results. To sum up, these additional results suggest that the effect of limit order book variables do not change significantly with the stock and time period and are very similar to those in our benchmark analysis.

2.11 Further Robustness Checks

In this section, we discuss several checks we have implemented to test the robustness of our results. The results are available in the appendix.

2.11.1 Daily Results

Our results are based on the estimation of the VAR system in Equation (9) over the whole sample period between 1st July 2010 and 31th July 2010. We have enough observations in a given day to estimate the VAR system for each day of our sample period separately. This allows us to analyze whether results based on the whole sample period continue to hold when we focus on a single day. Furthermore, it also provides some information about the stability of our results. Not surprisingly, daily coefficient estimates on the first lag of limit order book variables in the return equation tend to change from one day to another but most of them tend to be stable and move around the corresponding estimates based on the whole sample.

2.11.2 Controlling for Other Factors

In this section, we analyze the robustness of our results by controlling for other factors that might potentially affect the relation between return, trade direction and the state of the limit order book. One such variable is the duration measured as the waiting time between consecutive transactions. Dufour and Engle (2000) analyze the effect of duration on the relation between return and trade direction by considering the interaction between duration and trade direction as an additional variable in their empirical specification. In this framework, they find that as duration decreases, the price impact of trades, the speed of price adjustment to trade-related information, and the positive autocorrelation of signed trades all increase. They argue that markets have reduced liquidity when they are most active and there is an increased presence of informed trades. Our empirical approach differs from theirs in the sense that we include (log) duration itself in the empirical specification in Equation (9) as an additional state variable. Other variables that might also capture market activity are transaction volume and return volatility. We consider log number of shares traded at each transaction and log absolute midquote return (right before each transaction) to proxy for transaction volume and return volatility, respectively, and include these variables separately as additional state variables in the empirical specification in Equations (9). Our results available in an online appendix suggest that the relation between return, trade direction and the state of the limit order book do not change significantly when we control for duration, transaction volume and volatility.

2.12 Conclusion

In this paper, we analyze whether the state of the limit order book affects future price movements in line with what the theory predicts. To this end, we reconstruct the first 20 levels of the historical limit order book every millisecond for several companies traded at Frankfurt Stock Exchange in July 2010 and June 2011 based on the data from the Xetra electronic trading system. We consider several variables that summarize different dimensions of the information embedded in the limit order book, such as depth and slope of both ask and bid sides separately as well as their ratios. Following Hasbrouck (1991), we estimate a linear vector autoregressive system (VAR) that includes midquote return, trade direction and a limit order book variable one at a time. In line with theoretical models of dynamic limit order markets, we find that the state of the limit order book help predict short run midquote return. Limit order book variables also have significant long term cumulative effects on midquote return and trade directions. This long term cumulative effect is stronger and takes longer to be fully realized for variables based on higher levels of the book. In a simple high frequency trading exercise, we show that it is possible to obtain economic gains from the relation between limit order book variables and midquote return.

References

Admati, A.R., 1985. A noisy rational expectations equilibrium for multi-asset securities markets. Econometrica 53, 629-657

Beltran-Lopez, H., Giot, P., Grammig, J., 2009. Commonalities in the order book. Financial Markets and Portfolio Management 23, 209-242

Biais, B., Hillion, P., Spatt, C., 1995. An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse. The Journal of Finance 50, 1655-1689

Bilodeau, Y., 2013. Xetraparser [computer software]. HEC Montréal

Cao, C., Hansch, O., Wang, X., 2009. The information content of an open limit-order book. Journal of Futures Markets 29, 16-41

Chakravarty, S., Holden, C.W., 1995. An Integrated Model of Market and Limit Orders. Journal of Financial Intermediation 4, 213-241

Dufour, A., Engle, R.F., 2000. Time and the Price Impact of a Trade. The Journal of Finance 55, 2467-2498

Easley, D., O'Hara, M., 2004. Information and the Cost of Capital. The Journal of Finance 59, 1553-1583

Foucault, T., Kadan, O., Kandel, E., 2005. Limit Order Book as a Market for Liquidity. Review of Financial Studies 18, 1171-1217

Glosten, L.R., 1994. Is the Electronic Open Limit Order Book Inevitable? The Journal of Finance 49, 1127-1161

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

Goettler, R.L., Parlour, C.A., Rajan, U., 2009. Informed traders and limit order markets. Journal of Financial Economics 93, 67-87

Granger, C.W.J., 1969. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica 37, 424-438

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

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

Hellwig, M.F., 1980. On the aggregation of information in competitive markets. Journal of Economic Theory 22, 477-498

Kalay, A., Wohl, A., 2009. Detecting Liquidity Traders. Journal of Financial and Quantitative Analysis 44, 29-54

Kozhan, R., Salmon, M., 2012. The information content of a limit order book: The case of an FX market. Journal of Financial Markets 15, 1-28

Kyle, A.S., 1985. Continuous Auctions and Insider Trading. Econometrica 53, 1315-1335

Kyle, A.S., 1989. Informed Speculation with Imperfect Competition. The Review of Economic Studies 56, 317-355

Newey, W.K., West, K.D., 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica 55, 703-708

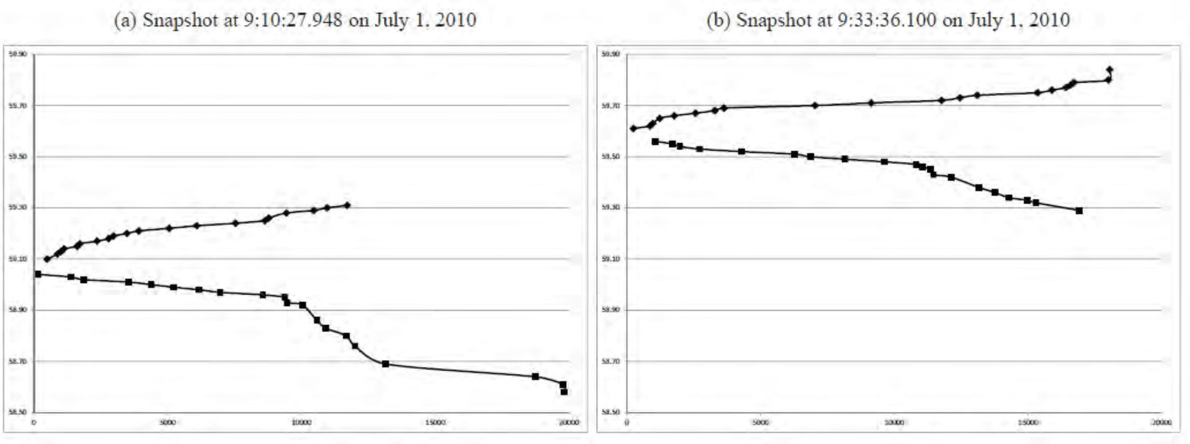
Rock, K., 1996. The Specialist’s Order Book and Price Anomalies. Working Paper,Harvard University

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

Rosu, I., 2009. A Dynamic Model of the Limit Order Book. The Review of Financial Studies 22, 4601-4641

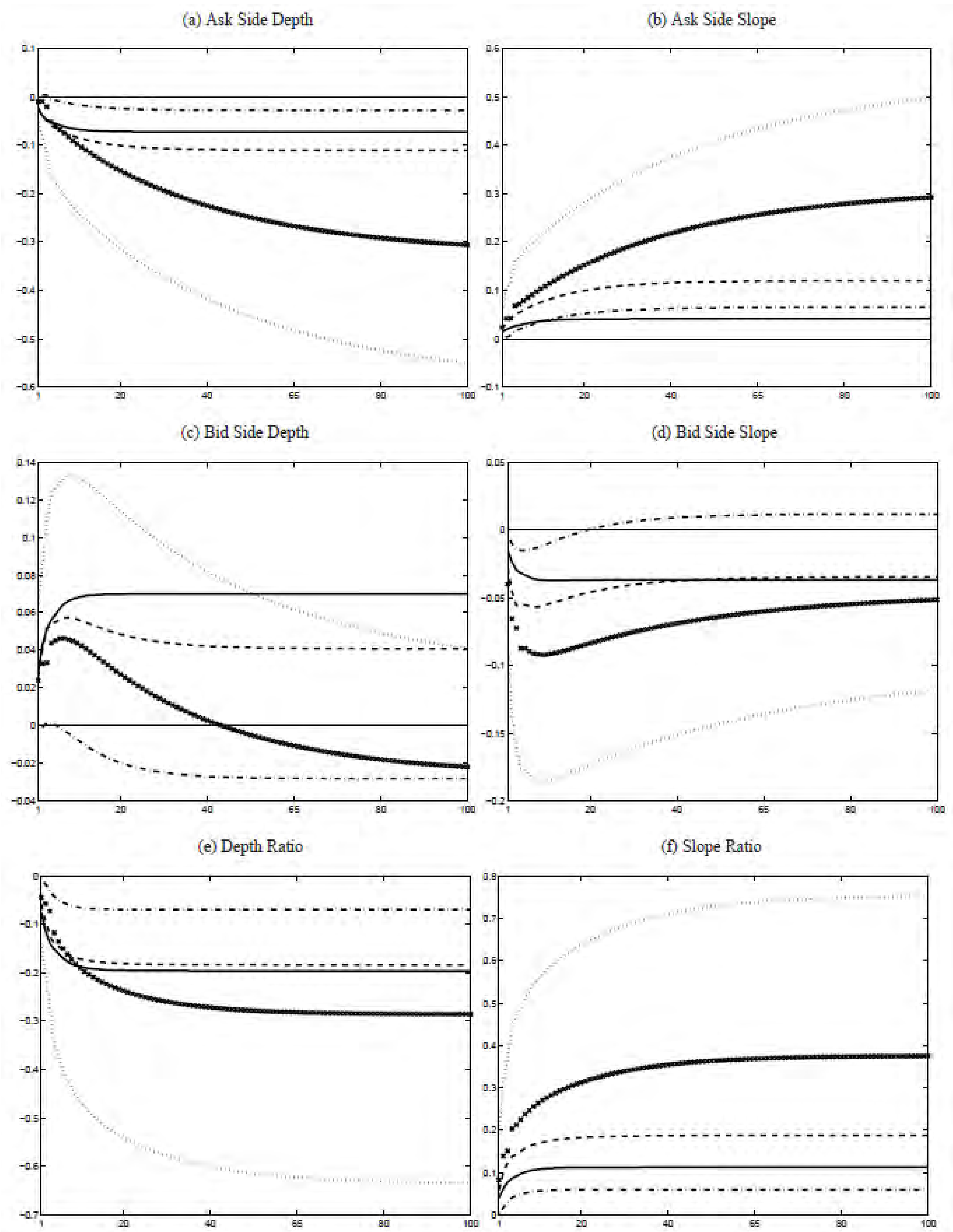
Seppi, D.J., Kumar, P., 1994. Limit and Market Orders with Optimizing Traders. SSRN eLibrary

Figure 2.1 Two Snapshots of the Limit Order Book



Panel (a) and (b) present the price and depth of the first 20 levels of the limit order book for Merck on July 1, 2010 at 9:10:27.948 and 9:33:36.100, respectively. The diamonds and squares represent different levels of the ask and bid sides, respectively.

Figure 2.2 Impulse Response Function of Return to Limit Order Book Variables



This figure presents the impulse response functions of midquote return to a one unit positive shock to limit order book variables. The horizontal axis is transaction periods, i.e. the number of transactions since the initial shock, and the vertical axis is the response of midquote return in basis points. The solid, dashed, small dotted and dashed-dotted line correspond to the limit order book variable measured based on the first two, five, ten and twenty levels, respectively. The big dotted line corresponds to the limit order book variable measured between second and fifth levels.

Table 2.1 Summary Statistics for Daily Trading Volume and Returns

(a) Daily Trading Volume

	MRK		SAP	
	July-2010	June-2011	July-2010	June-2011
Mean	532,321	356,218	4,198,806	3,308,930
Median	478,219	316,050	4,185,447	3,070,861
Std. Deviation	294,149	121,066	1,547,879	1,491,676
Min	202,535	210,017	2,286,050	1,296,397
Max	1,629,225	681,365	8,676,879	8,618,285

(b) Daily Returns

	MRK		SAP	
	July-2010	June-2011	July-2010	June-2011
Mean	0.590%	-0.087%	-0.208%	-0.152%
Median	0.336%	-0.019%	-0.331%	-0.018%
Std. Deviation	1.662%	1.073%	1.140%	1.177%
Min	-2.867%	-2.849%	-2.266%	-2.608%
Max	4.507%	2.039%	2.626%	2.114%

This table presents summary statistics for daily trading volume (Panel (a)) and returns (Panel (b)) for MRK and SAP in July 2010 and June 2011

Table 2.2 Variables for the Snapshots of the Limit Order Book in Figure 2.1

LOB Variable	Snapshot 1	Snapshot 2
Ask Side Variables		
Depth (Levels 1-2)	890	864
Depth (Levels 1-5)	1624	1774
Depth (Levels 1-20)	11723	18056
Depth (Levels 2-5)	1124	1528
Depth (Levels 6-20)	10599	16826
Slope (Levels 1-2)	0.000051	0.000016
Slope (Levels 1-5)	0.000044	0.000033
Slope (Levels 1-20)	0.000019	0.000013
Slope (Levels 2-5)	0.000041	0.000044
Slope (Levels 6-20)	0.000016	0.000011
Bid Side Variables		
Depth (Levels 1-2)	1378	1705
Depth (Levels 1-5)	4390	4282
Depth (Levels 1-20)	19840	16923
Depth (Levels 2-5)	4228	3218
Depth (Levels 6-20)	16290	14200
Slope (Levels 1-2)	-0.000008	-0.000016
Slope (Levels 1-5)	-0.000009	-0.000012
Slope (Levels 1-20)	-0.000023	-0.000017
Slope (Levels 2-5)	-0.000010	-0.000012
Slope (Levels 6-20)	-0.000027	-0.000018
Ratio Variables		
Depth Ratio (Levels 1-2)	-0.215168	-0.327365
Depth Ratio (Levels 1-5)	-0.459927	-0.414135
Depth Ratio (Levels 1-20)	-0.257168	0.032391
Depth Ratio (Levels 2-5)	-0.579970	-0.356089
Depth Ratio (Levels 6-20)	-0.211648	0.084639
Slope Ratio (Levels 1-2)	0.723600	0.018268
Slope Ratio (Levels 1-5)	0.649243	0.449419
Slope Ratio (Levels 1-20)	-0.110834	-0.137310
Slope Ratio (Levels 2-5)	0.608115	0.581224
Slope Ratio (Levels 6-20)	-0.263582	-0.244090

This table presents the values of the limit order book variables based on the snapshots of the limit order book for Merck on July 1, 2010 at 9:10:27.948 (Snapshot 1) and 9:33:36.100 (Snapshot 2) presented in Figure 2.1

Table 2.3 Summary Statistics

	Transformation	Mean	Median	Std. Dev.	Min	Max
Log Return	$\times$ 10000					
Ask Side Variables						
Depth (Levels 1-2)	$\ln(x)$	6.599	6.640	0.852	0.693	10.017
Depth (Levels 1-5)	$\ln(x)$	7.973	8.008	0.531	5.521	10.188
Depth (Levels 1-20)	$\ln(x)$	9.616	9.628	0.310	8.231	10.667
Depth (Levels 2-5)	$\ln(x)$	7.555	7.609	0.626	4.466	10.140
Depth (Levels 6-20)	$\ln(x)$	9.367	9.378	0.350	7.742	10.579
Slope (Levels 1-2)	$\ln(x)$	-10.156	-10.233	1.170	-14.601	-3.219
Slope (Levels 1-5)	$\ln(x)$	-10.856	-10.932	0.684	-13.393	-7.367
Slope (Levels 1-20)	$\ln(x)$	-10.882	-10.911	0.462	-12.262	-8.865
Slope (Levels 2-5)	$\ln(x)$	-10.956	-11.054	0.737	-13.647	-6.359
Slope (Levels 6-20)	$\ln(x)$	-10.865	-10.879	0.513	-12.272	-8.658
Bid Side Variables						
Depth (Levels 1-2)	$\ln(x)$	6.570	6.627	0.792	2.398	9.264
Depth (Levels 1-5)	$\ln(x)$	7.883	7.912	0.532	4.754	9.778
Depth (Levels 1-20)	$\ln(x)$	9.558	9.585	0.332	8.001	10.869
Depth (Levels 2-5)	$\ln(x)$	7.459	7.506	0.632	4.220	9.379
Depth (Levels 6-20)	$\ln(x)$	9.322	9.353	0.360	7.745	10.745
Slope (Levels 1-2)	$\ln(\|x\|)$	-10.167	-10.272	1.116	-13.403	-2.813
Slope (Levels 1-5)	$\ln(\|x\|)$	-10.768	-10.831	0.685	-12.654	-7.371
Slope (Levels 1-20)	$\ln(\|x\|)$	-10.801	-10.870	0.525	-12.242	-8.879
Slope (Levels 2-5)	$\ln(\|x\|)$	-10.856	-10.942	0.744	-12.886	-7.212
Slope (Levels 6-20)	$\ln(\|x\|)$	-10.792	-10.857	0.572	-12.306	-8.615
Ratio Variables						
Depth Ratio (Levels 1-2)		0.012	0.012	0.444	-0.995	0.991
Depth Ratio (Levels 1-5)		0.040	0.047	0.320	-0.912	0.922
Depth Ratio (Levels 1-20)		0.028	0.022	0.154	-0.566	0.694
Depth Ratio (Levels 2-5)		0.041	0.050	0.339	-0.911	0.942
Depth Ratio (Levels 6-20)		0.021	0.016	0.177	-0.691	0.713
Slope Ratio (Levels 1-2)		0.005	0.004	0.551	-0.999	0.999
Slope Ratio (Levels 1-5)		-0.039	-0.048	0.364	-0.954	0.928
Slope Ratio (Levels 1-20)		-0.038	-0.039	0.219	-0.775	0.696
Slope Ratio (Levels 2-5)		-0.042	-0.052	0.386	-0.984	0.966
Slope Ratio (Levels 6-20)		-0.034	-0.034	0.273	-0.779	0.795

This table presents the transformation applied to limit order book variables and the summary statistics for these transformed variables

Table 2.4 Coefficient Estimates on Lagged Values of Limit Order Book Variables (Z_{i-j}) and F-statistics for Granger Causality Tests

	$\alpha_{z,1}$	$\alpha_{z,2}$	$\alpha_{z,3}$	$\alpha_{z,4}$	$\alpha_{z,5}$	F-stat
Ask Side Variables						
Depth (Levels 1-2)	-0.260***	0.003	0.008	0.028	0.019	408.518***
Depth (Levels 1-5)	-0.224***	0.022	0.033	-0.005	0.062*	75.322***
Depth (Levels 1-20)	-0.532***	0.228	0.003	-0.226	0.398***	51.619***
Depth (Levels 2-5)	0.058*	-0.021	0.021	-0.074**	0.026	11.236**
Depth (Levels 6-20)	-0.100	0.130	-0.091	-0.215*	0.194*	16.472***
Slope (Levels 1-2)	0.132***	-0.009	0.001	-0.009	-0.018	193.198***
Slope (Levels 1-5)	0.189***	-0.028	-0.038	0.013	-0.074**	67.759***
Slope (Levels 1-20)	0.575***	-0.064	-0.325**	0.268	-0.381***	93.289***
Slope (Levels 2-5)	-0.019	0.030	-0.021	0.059*	-0.033	6.714
Slope (Levels 6-20)	0.238**	-0.009	-0.171	0.197*	-0.200**	32.749***
Bid Side Variables						
Depth (Levels 1-2)	0.286***	0.000	-0.002	-0.029	-0.020	443.729***
Depth (Levels 1-5)	0.260***	-0.040	-0.010	-0.064	-0.027	93.188***
Depth (Levels 1-20)	0.627***	-0.304	-0.047	-0.058	-0.129	52.157***
Depth (Levels 2-5)	-0.008	-0.005	0.016	-0.022	0.022	1.177
Depth (Levels 6-20)	0.240**	-0.123	-0.084	0.068	-0.055	11.778**
Slope (Levels 1-2)	-0.151***	-0.022	-0.011	0.022	0.007	303.520***
Slope (Levels 1-5)	-0.277***	0.014	0.045	0.094**	0.031	154.872***
Slope (Levels 1-20)	-0.840***	0.186	0.291**	0.152	0.130	172.180***
Slope (Levels 2-5)	-0.050	-0.006	0.003	0.022	0.000	9.005
Slope (Levels 6-20)	-0.400***	0.081	0.181*	-0.037	0.117	62.047***
Ratio Variables						
Depth Ratio (Levels 1-2)	-0.751***	0.003	0.009	0.092***	0.047	942.974***
Depth Ratio (Levels 1-5)	-0.579***	0.073	0.018	0.080	0.096*	182.466***
Depth Ratio (Levels 1-20)	-1.386***	0.571**	0.039	-0.170	0.460**	129.497***
Depth Ratio (Levels 2-5)	-0.060	0.006	-0.058	0.004	-0.014	11.557**
Depth Ratio (Levels 6-20)	-0.436***	0.264	-0.004	-0.280	0.188	31.213***
Slope Ratio (Levels 1-2)	0.404***	0.024	0.033	-0.035	-0.038	475.614***
Slope Ratio (Levels 1-5)	0.582***	-0.042	-0.080	-0.091	-0.096*	225.576***
Slope Ratio (Levels 1-20)	1.897***	-0.276	-0.724***	0.100	-0.561***	340.444***
Slope Ratio (Levels 2-5)	0.041	0.035	0.008	0.022	-0.032	7.520
Slope Ratio (Levels 6-20)	0.825***	-0.103	-0.407**	0.284*	-0.351***	115.302***

This table presents the parameter estimates of lagged limit order book variables in the return equation $(\alpha_{z,\tau})$ and F-statistics for Granger Causality tests. ***, **, * represent statistical significance at 1%, 5%, 10%, respectively.

Table 2.5 Bid-Ask Return on Trading Strategies

	Number of Trades	0 ms	Latency 500 ms	1000 ms
Ask Side Variables				
Depth (Levels 1-2)	119	3.676%	3.351%	2.619%
Depth (Levels 1-5)	98	6.607%	6.330%	5.591%
Depth (Levels 1-20)	98	5.384%	4.954%	4.400%
Depth (Levels 2-5)	96	0.045%	-0.466%	-0.815%
Depth (Levels 6-20)	102	-3.666%	-4.108%	-4.582%
Slope (Levels 1-2)	120	2.229%	2.016%	1.478%
Slope (Levels 1-5)	99	2.119%	1.791%	1.289%
Slope (Levels 1-20)	100	0.284%	-0.176%	-0.469%
Slope (Levels 2-5)	88	1.557%	1.044%	0.718%
Slope (Levels 6-20)	100	-3.445%	-4.198%	-4.335%
Bid Side Variables				
Depth (Levels 1-2)	118	2.223%	0.910%	0.377%
Depth (Levels 1-5)	111	-4.928%	-5.478%	-5.857%
Depth (Levels 1-20)	93	-1.580%	-2.452%	-3.344%
Depth (Levels 2-5)	94	0.972%	0.311%	-0.157%
Depth (Levels 6-20)	93	0.599%	0.306%	-0.299%
Slope (Levels 1-2)	127	-6.685%	-7.805%	-8.461%
Slope (Levels 1-5)	120	-6.846%	-7.798%	-8.328%
Slope (Levels 1-20)	103	1.404%	0.632%	0.215%
Slope (Levels 2-5)	92	0.013%	-0.937%	-1.607%
Slope (Levels 6-20)	85	-2.272%	-3.057%	-3.388%
Ratio Variables				
Depth Ratio (Levels 1-2)	141	0.726%	-0.708%	-1.181%
Depth Ratio (Levels 1-5)	109	-0.578%	-2.034%	-2.571%
Depth Ratio (Levels 1-20)	97	5.393%	5.056%	4.414%
Depth Ratio (Levels 2-5)	94	1.491%	0.726%	0.134%
Depth Ratio (Levels 6-20)	89	0.582%	0.187%	-0.118%
Slope Ratio (Levels 1-2)	121	1.169%	0.673%	0.048%
Slope Ratio (Levels 1-5)	115	-4.944%	-5.900%	-6.420%
Slope Ratio (Levels 1-20)	112	2.937%	2.270%	1.944%
Slope Ratio (Levels 2-5)	90	1.471%	0.740%	0.150%
Slope Ratio (Levels 6-20)	104	4.778%	4.089%	3.848%
Benchmark with MA filter	96	-1.429%	-1.964%	-2.478%
Benchmark without MA filter	270	-3.305%	-6.425%	-8.061%

This table presents the cumulative bid-ask returns on trading strategies over the whole trading period of July 2010. The trading strategy is based on the empirical model in equation 9 that excludes the contemporaneous effect of trade direction on return and described in detail in Section 2.9. The parameter of the filter for the return forecasts, κ, is set to 1 basis points. The long and short-run moving average filters are calculated based on the last three and forty transaction prices, including the most recent one, respectively. The parameter of the moving average filter is set to 0.06 euros. Benchmark with and without moving average (MA) filter present results from trading strategies based on a forecasting model that ignores information embedded in limit order book variables.

Chapter 3
The Dynamics of Ex ante High-Frequency Liquidity: An Empirical Analysis

Abstract

The ex ante liquidity embedded in open Limit Order Book (LOB) and its dynamics have been one of the most important issues in financial research and evolves with development of financial infrastructure. Using the tick-by-tick data and the re-constructed open LOB data from Xetra trading system, we utilize a decomposition model to investigate the impact of trade duration, quote duration and other exogenous variables on ex-ante liquidity embedded in open LOB. More specifically, the duration factor is assumed to be strictly exogenous and its dynamics can be captured by a Log-ACD process. Furthermore, by taking into account of Ultra High Frequency (UHF) data, our modeling involves decomposing consistently the joint distribution of the ex-ante liquidity measure into certain simple and interpretable distributions. In this study, the decomposed factors are Activity, Direction and Size. Our results suggest that trade durations and quote durations do influence the ex-ante liquidity changes. Short-run variables, such as spread change and volume, also predict the tendency of liquidity changes. However, contrary to previous literature, the more long-term variable trade imbalance is less informative in predicting the liquidity change dynamics.

3.1 Introduction

Liquidity has been one of the most important issues in financial research for a long time and evolves with development of financial infrastructure. However, the term of liquidity might be interpreted differently by various market participants: market regulators see liquidity as a capacity of buying and selling a large quantity of financial securities or the total turnover of assets in a given time interval. For the individual traders, the level of liquidity might only relate to the quantity available when changing their positions at buying or selling side. In this paper, we focus on the dynamic of ex-ante liquidity, which is offered by the Limit Order Book (LOB), from a general view through the analysis of available electronic data in Xetra open LOB system.

There already exist a large number of theoretical microstructure models emphasizing on state of LOB because of its importance in discovering price formation and explaining the trading mechanism. However, the empirical research on ex-ante liquidity in LOB itself is still limited. There is a huge gap between theoretical and empirical research for the analysis of high-frequency trading mechanism and liquidity. This is due to the fact that market order traders and LOB traders behave in a complicated way, and the theoretical models are reluctant to capture all these features. For example, the LOB traders in Parlour (1998) cannot choose multiple limit-order price strategy; in Foucault (1999), the lifetime of a limit order could not last more than one period; and in Foucault, Kadan and Kadel (2005), the limit orders cannot be canceled once they are placed in the open LOB. All these restrictions have impact on the ex-ante liquidity provision.

In market microstructure framework, beside the spontaneously demand and supply in the market, the liquidity provision also inherently depends on various exogenous factors such as trading mechanism, information disclosure process and regulatory issues. For instance, Viswanathan and Wang (2002) show that the slope of the equilibrium bid price is flatter than that in dealership since the discriminatory pricing rule intensifies the liquidity provision. Another important concern is the availability of data. In the previous studies, liquidity measures such as quoted spread and effective spread are obtained from trade related database. One problem with these liquidity measures is that they only reflect the liquidity for one share trading. With the introduction of open LOB trading mechanism, the access to more complete high-frequency data becomes possible. As a result, we can use the information in LOB to compute two different measures of the liquidity: ex-ante liquidity and ex-post liquidity. Up to now, a large number of liquidity (illiquidity) measures[1] or proxies have been proposed. There exists little consensus on the most appropriate liquidity measure in terms of efficiency and accuracy due to its multi-facets complexity in both cross-section (depth, width, resiliency etc) and frequency (high-frequency or low-frequency based) dimensions. In practice, researchers

[1] Liquidity measure designs liquidity or illiquidity measure.

propose and use various liquidity measures, consistent with the available data, to examine one or several liquidity dimensions they focus on.

This empirical paper concentrates on ex-ante liquidity embedded in open LOB (depth and price impact). Recent studies related to ex-ante liquidity include Irvine (2000), Coppejans, Domowitz and Madhavan (2004), Domowitz, Hansch and Wang (2005), Giot and Grammig (2006), Beltran-Lopez, Giot and Grammig (2009), Beltran-Lopez, Grammig and Menkveld (2011). According to Aitken and Comerton-Forde (2003), ex-post liquidity measures involve trade-based measures and ex-ante liquidity measures are related to order-based. The former measures involve most used liquidity measures and indicate what the traders have obtained in the realized transaction. The second group captures the cost related to potential immediate trading. For traders with small quantity to trade, the liquidity cost is quasi-fixed and is equal to the half of bid-ask spread. In this case we do not need any information about the LOB. However, for the larger traders, the spread will surely underestimate the associated cost when the quantity to trade is larger than the quantity available at first level. This requires a liquidity measure that can allow for volume-related price impact.

Another characteristic of the ex-ante liquidity measure is that it can be computed even where there is no trade. As noted by Beltran-Lopez, Grammig and Menkveld (2011), for the average stock in DAX30 of Xetra trading system, the average daily number of limit orders is 12,785, i.e. 25 limit orders per minute, whereas the average number of trades is four per minute. As the trend of high frequency trading continues to increase rapidly, market regulators and institutional investors need more continuous measure of liquidity to update their information set. The traditional ex-post liquidity measure cannot meet this requirement as it can be computed only when transactions occur. The temporal relation between updates of LOB and transactions can be shown in Figure (3.1) where transactions and updates of LOB are denoted by circles and squares respectively.

[Insert Figure 3.1 here]

The objective of this study is to capture the evolution of ex-ante liquidity measure by a decomposition model which allows for various factors in a flexible way. To our knowledge, we are the first to consider modeling the ex-ante liquidity by a decomposition model and to include a large set of factors. Given the particularity of UHF data and the motivation of microstructure analysis, one of possible modeling frameworks involves decomposing consistently the joint distribution of a target variable into certain simple and interpretable distributions.

The idea of decomposition is pioneered by Rogers and Zane (1998) and aims at constructing observation-driven models in the sense of Cox et al. (1981). The decomposition model has been first used in analyzing transaction price dynamics. Hausman, Lo, and MacKinly (1992) and Russell and Engle (2005) propose an Autoregressive Conditional Multinomial (ACM) and ordered Probit model respectively. Rydberg and Shephard (2003) manage to achieve the same goals by decomposing the joint distribution of tick-by-tick transaction price changes into three sequential components. The first component is designed active and indicates whether the price change will occur. The second called direction component relates to the direction of price change, and the third one measures the absolute size of the price change. Finally, the decomposition model is used to provide the prediction for price movements or price level with the help of simulations. McCulloch and Tsay (2001) aim at modeling the transaction price change process with a decomposition model. In their framework, they initialize a price change and duration (PCD) model that decomposes the price changes into four factors and introduces time and liquidity dimension in modeling price change dynamics. The duration between two consecutive transactions and the number of trades during this duration are considered as implicit factors for the price changes. In total, they use six conditional models to capture the dynamic of price changes. Manganelli (2005) applies the decomposition methodology in investigating the simultaneous interaction between duration, volume and return. Two subgroups of different trade intensity perform different dynamics. The decomposition framework remains flexible for more complicated modeling and according to different modeling assumptions, addition or deletion of certain factors is possible.

Our paper differs from the existing literature in several dimensions. First of all, instead of aggregating the time for a fix interval, our analysis contains a time dimension.

Specifically, our paper investigates the role of trade and quote durations in explaining the dynamic of ex-ante liquidity provision. It's widely known that trades may contain private information that will be further incorporated in the quote updates. Consequently, trade duration and quote duration are closely related to the speed of information flow and quotes revision, respectively. Our model takes them as two factors for the liquidity changes and assumes that the trade duration factor is strictly exogenous. The joint modeling of trade duration and corresponding quote duration is challenging due to the fact that they are not synchronic by nature. To circumvent this problem, Engle and Lunde (2003) propose a bivariate point process. As a general conclusion, they found that information flow variables, such as trade duration, large volume of trade and spread, predict more rapid quote revision. We first apply their bivariate model to our Xetra dataset and then extend the analysis to the impact of both trade and quote durations on ex-ante liquidity changes.

Second, by applying a decomposition model, we make a much finer analysis of ex-ante liquidity changes and take advantage of econometric modeling by attempting to capture a more general and realistic LOB trading pattern which is much more complicated than that characterized by structural models. Specifically, following Engle and Lunde (2003) and Rydberg and Shephard (2003), we use different factors to model the dynamics of liquidity changes, including trade duration, quote duration, activity, direction and size. Our objective is to identify the possible determinants of each factor from a wide range of variables. Our empirical findings will not only provide support to existing theoretical models but will offer a guidance for new theoretical models in market microstructure.

Third, regarding the explanatory variables, in addition to the lagged dependent variables, we also put various exogenous variables in the different factor equations. In existing literature, most papers take one variable as the explanatory variable and suppose that this variable can summarize all the trade information. Instead, our variables are volume-related, duration-related and trade imbalance related. Among these variables, we also distinguish the short-run and long-run variables[2] to reflect the time dimension of the variables.

The rest of the paper is organized as follows: Section 3.2 describes the Xetra trading system and the ex-ante liquidity measure Xetra Liquidity Measure (XLM) we model in this study. Section 3.3 briefly presents the decomposition model and the exogenous variables we use for explaining the dynamics of each factor. Section 3.4 applies the econometric model proposed in Section 3.3 to our data for the selected stocks and reports the estimation results. Section 3.5 concludes and provides possible new research directions.

3.2 Xetra trading system and Xetra Liquidity Measure (XLM)

3.2.1 Xetra trading system

Electronic trading systems have been adopted by many stock exchanges during last two decades. The data used in this study is from the trading system Xetra, which is operated by Deutsche Börse at Frankfurt Stock Exchange (FSE) and has a similar structure to Integrated Single Book of NASDAQ and Super Dot of NYSE. Xetra trading system realizes more than 90% of total transactions at German exchanges. Since September 20, 1999, trading hours have been from 9h00 to 17h30 CET (Central European Time). However, during the pre- and post-trading hours, entry, revision and cancellation are still permitted.

There are two types of trading mechanism during normal trading hours: the call auction and the continuous auction. A call auction could be organized one or several times during the trading day in which the clearance price is determined by the state of LOB and remains as the open price for the following continuous auction. During each call auction, market participants can submit both round-lot and odd-lot orders. Both start and end time for a call auction are randomly chosen by a computer to avoid scheduled trading. Between the call auctions, the market is organized as a continuous auction where traders can only submit round-lot-sized limit orders or market orders.

---

[2] Short-run variables are variables at a given timepoint, whereas the long-run variables are variables that summarize the information over an interval.

For the highly liquid stocks, there are no dedicated market makers during the continuous trading. As a result, all liquidity comes from limit orders in LOB. Xetra trading system imposes the Price Time Priority condition where the electronic trading system places the incoming order after checking the price and timestamps of all available limit orders in LOB. Our database includes 20 levels of LOB information[3] meaning that, by monitoring the LOB, any registered member can evaluate the liquidity supply dynamic and potential price impact caused by a market order. However, there is no information on the identities of market participants.

The reconstruction of LOB is mainly based on two main types of data streams: delta and snapshot. The delta tracks all the possible updates in LOB such as entry, revision, cancellation and expiration, whereas the snapshot gives an overview of the state of LOB and is sent after a constant time interval for a given stock. Xetra original data with delta and snapshot messages are first processed using the software XetraParser developed by Bilodeau (2013) in order to make Deutsche Börse Xetra raw data usable for academic and professional purposes. XetraParser reconstructs the real-time order book sequence including all the information for both auctions and continuous trading by implementing the Xetra trading protocol and Enhanced Broadcast. We then put the raw LOB information in order under a readable format for each update time and retrieve useful and accurate information about the state of LOB and the precise timestamp for order modifications and transactions during the continuous trading. The stocks Metro AG (MEO), Merck (MRK), RWE AG (RWE) and ThyssenKrupp AG (TKA) that we choose for this study are blue chip stocks from the DAX30 index. The selected stocks have different levels of market capitalization and are in different markets. Metro AG, with market capitalization of 9.8 billion Euros, operates retail stores, supermarkets and hypermarkets on-line and off-line. Merck is the world's oldest operating chemical and pharmaceutical company with a market capitalization of 4 billion Euros in 2010. RWE generates and distributes electricity to various customers including municipal, industrial, commercial and residential customers. The company produces natural gas and oil, mines coal and delivers and distributes gas. In 2010, its market capitalization was around 15

[3] Our dataset does not include the hidden part of an iceberg order.

billion Euros. ThyssenKrupp AG manufactures iron and steel industrial components with market capitalization of 11.9 billion Euros.

3.2.2 (Il)liquidity measure in Xetra Trading System

As mentioned above, liquidity is the central quality criterion for the efficiency of marketplaces in electronic securities trading. We use the following definition of liquidity: the ability to convert the desired quantity of a financial asset into cash quickly and with little impact on the market price (Demsetz (1968); Black (1971); Glosten and Harris(1988)).The Xetra trading system defines its own ex-ante (il)liquidity measure Xetra Liquidity Measure (XLM) as follows[4]:

$$XLM^{q} = \frac{P^{q}_{net,buy} - P^{q}_{net,sell}}{P_{mid}} \times 10000 \tag{3.1}$$

$$P^{q}_{net,buy} = \frac{\sum_{k=1}^{K-1} P_{k,i} \cdot v_{k,i} + P_{K,i} \cdot v_{K,i}}{v} \text{ and } v_{K,i} = v - \sum_{k=1}^{K-1} v_{k,i}$$

where q is the potential size in Euro, the conventional value for q in Xetra trading system is 25 000 euros. $P^{q}_{net,buy}$ is the average price when a buy market order of q euros arrives and $P^{q}_{net,sell}$ relates to the average price for a sell market order of q euros. v is the total volume bought by the market order of q euros. $P^{q}_{net,sell}$ and $v_{k,i}$ are the kth level ask price and volume available, respectively. $v_{k,i}$ is the quantity left after K-1 levels are completely consumed by the market order of q euros. $P^{q}_{net,sell}$ is computed in the similar way. P_{mid} is the mid-quote of bid-ask spread. XLM measures the relative potential round-trip impact when buying and liquidating a volume of q euros at the same time. Intuitively, it is also the cost in basis point for an immediate demand for liquidity from buy and sell market orders. For example, a XLM of 10 and a market order volume of 25 000 means that the market impact of buying and selling is 25 euros. The previous market microstructure literature that considers the quantity available in LOB includes

[4] More information is available in the official website of Xetra.

Irvine, Benston and Kandel (2000), Domowitz, Hansch and Wang (2005), Coppejans, Domowitz and Madhavan (2004), among others.

XLM is based on P_{mid} and the difference between $P^{q}_{net,buy}$ and $P^{q}_{net,sell}$. Theoretically, there are infinite combinations of $P^{q}_{net,buy}$ and $P^{q}_{net,sell}$ for the same difference. That is, the illiquidity may come from either side or both sides of LOB. However, our study focuses on stocks' global liquidity and considers the lack of depth in either side as an illiquid situation. For the buy side or sell side investors, the corresponding one-side XLM can be defined and computed in a similar fashion.

The XLM measure can be used for several ends: first, XLM measure can help decision making in security selection when constructing a portfolio. Among the stocks with same correlation with market portfolio, a small XLM stock will decrease the trading cost and at last provide a higher net return. Second, XLM can also be used for comparison purpose. For instance, a cross listing stock may perform different liquidity features in different markets. By using XLM, one can quantify this difference by choosing a given volume. Third, by choosing different volume q, we can identify the global liquidity or one-side liquidity pattern.

3.3 Methodology

3.3.1 Model

In our dataset, there are three variables to model: trade, quote and ex-ante liquidity changes. Following Engle and Lunde (2003), we consider trade and quote as a bivariate point process. Based on the timestamps of these point processes, we can define two types of duration: trade duration and quote duration which constitute a bivariate duration process. However, due to the no-synchronization problem, we further assume that the trade times are the initiators for both the following trade and the next quote update. Consequently, trade durations and quote durations with same index share the same original timestamp. The economic intuition behind the assumption is that the limit order traders in open LOB update their quotes by observing the transactions. After each transaction, we compute the quote duration based on the very last transaction.

As mentioned by Engle and Lunde (2003), by taking the transaction times as the origin of each pair of durations, two possible situations may occur for quote duration: an uncensored observation or a censored observation. The uncensored duration occurs when the update of quote is before the next trade arrival and the censored duration happens when the following trade arrives before the update of quote. We denote x_i and y_i as the trade duration and quote duration, respectively, and further define the observed quote duration $\overline{y_i} = (1-d_i)\cdot y_i + d_i x_i$,where $d_i = I_{\{y_i > x_i\}}$.

Apart from the time dimension, there is also a liquidity dimension to model. We use the above mentioned XLM as the liquidity measure. In the no-censored situation, we take the average of the measure within the first quote update timestamp and the following trade timestamp. We suppose its evolution could be written as:

$$XLM_i^q = XLM_0^q + \sum_{k=1}^{i} Z_k \tag{3.2}$$

where Z_k is defined as the kth rounded signed change for XLM.

We define p as the joint density for trade duration, quote duration and XLM changes. One possible decomposition for this joint density of kth mark can be written as:

$$\begin{aligned} & p(x_k, \overline{y_k}, z_k^q \mid F_{k-1};\omega) \\ & = g(x_k \mid F_{k-1};\omega_1)\cdot f^{Dur}(\overline{y_k} \mid x_k, F_{k-1};\omega_2)\cdot f^{A}(A_k \mid x_k, \overline{y_k}, F_{k-1};\omega_3)\cdot \\ & f^{D}(D_k \mid x_k, \overline{y_k}, A_k, F_{k-1};\omega_4)\cdot f^{S}(S_k \mid x_k, \overline{y_k}, A_k, D_k, F_{k-1};\omega_5) \end{aligned}$$

We define x_k and $\overline{y_k}$ be the exogenous factor, which relates to the trade duration and observed quote duration, for XLM changes. Conditional on information set F_{k-1} and two durations, A_k takes on value 0 or 1 indicating if there is a change on kth XLM. Conditional on A_k = 1, D_k relates to the direction of the XLM change by taking on the value -1 and +1. Finally, given the information set F_{k-1} with A_k = 1 and Dir_k, S_k takes on positive integers and indicates the size of the change. ω relates to the parameter set including ω_1, ω_2, ω_3, ω_4 and ω_5 which are the parameters for factors of trade duration, observed quote duration, activity, direction and size.

More precisely, we adopt Log-ACD model originally introduced by Bauwens and Giot (2000) in modeling the irregularly spaced trade durations which is considered as a main characteristic in high frequency data:

$$\frac{x_k}{\psi_k} = \varepsilon_k , \quad \psi_k = \exp\left(\omega + \sum_{j=1}^{p} \alpha_j \varepsilon_{k-j} + \sum_{j=1}^{l} \beta_j \ln \psi_{k-j} + \Psi' W_{k-1} \right) \tag{3.3}$$

where $\psi_k = E(x_k | F_{k-1})$ and ε_k is a i.i.d random variable following the generalized gamma distribution with unit expectation. The process of duration is composed of a sequence of deseasonalized durations. W_{k-1} is a vector of exogenous variables available at k-1 which include trade-related variables, quote-related variables and dummy variables to capture the intraday seasonality (hereafter, all exogenous vector will include dummy variables to capture the intraday seasonality). We use the Log-ACD to model durations as it is more flexible and the non-negativity constraint on the coefficients can be relaxed[5].

We adopt a similar structure for observed quote duration, that is

$$\frac{\overline{y_k}}{\phi_k} = \xi_k , \quad \phi_k = \exp\left(\mu + \sum_{j=1}^{p} \rho_j \xi_{k-j} + \rho_{j+1} \xi_{k-1} d_{k-1} + \sum_{j=1}^{l} \delta_j \ln \phi_{k-j} + \Phi' V_{k-1} \right) \tag{3.4}$$

Wher $\phi_k = E(\overline{y_k} | F_{k-1})$ and ξ_k is supposed to be i.i.d exponential distributed and the error distribution is supported by the estimation convergence. In this equation, we add the term with censored dummy variable to capture the impact of censored observation and the vector of exogenous variables V_{k-1} which may have some common variables as in W_{k-1}. We show in the following section that as quote durations are conditional on transactions; one possible exogenous variable could be the expected trade duration available at tick k.

Regarding the liquidity dimension, we decompose the XLM change into three factors: Activity, Direction and Size. The advantage of the decomposition approach is that we

[5] The ACD (Autoregression Conditional Duration) model is initialized by Engle and Russell (1998) and widely used for duration modeling.

can use simple and interpretable factors to model the complicated variables. To this end, the first factor Activity is a bivariate variable that takes values of 0 or 1 to indicate whether there is a change in XLM measure. To model the activity factor, we use the auto-logistic model (Cox et al.(1981)). As the log-likelihood function of the auto-logistic is concave, the numerical optimization can be easily and reliably realized. However, the high-frequency data often performs a slow decay for longer lags in autoregressive structure. Thus there is a trade-off between bias and variance, i.e., inference with too few parameters may be biased, while that with too many parameters may cause precision and identification problems. To solve this, we adopt another structure called the GLARMA (Generalized Linear Autoregressive Moving Average) binary model which is a generalized structure of auto-logistic structure by allowing moving average-type behavior (Shephard (1995)). The auto-logistic model for activity is defined as:

$$f(A_k = 1 | F_{k-1}, x_k, \overline{y_k}) = p(\theta_k^A), \quad \text{where } p(\theta_k^A) = \frac{\exp(\theta_k^A)}{1+\exp(\theta_k^A)} \tag{3.5}$$

$$\text{and } \theta_k^A = \Pi'_A M_{k-1}^A + g_k^A, \quad g_k^A = \sum_{j=1}^{p} \gamma_j^A g_{k-j}^A + \sum_{j=1}^{l} \lambda_j^A A_{k-j}$$

$$\text{Consequently, } f(A_k = 0 | F_{k-1}, x_k, \overline{y_k}) = \frac{1}{1+\exp(\theta_k^A)}$$

M_{k-1}^A is the vector of exogenous variables for activity factor known at *k-1*. In this logistic modeling, the parameter θ_k^A is time-varying and depends on both its own lag variables, such as lags of g_k and A_k, and some exogenous variables. The model will be validated by applying the Ljung-Box test on the standardized errors defined by:

$$u_k^A = \frac{A_k - p(\theta_k^A)}{\sqrt{p(\theta_k^A)(1-p(\theta_k^A))}} \tag{3.6}$$

which should be uncorrelated with zero mean and unit variance.

In a similar way, the Direction of change of the liquidity measure conditional on the activity factor is specified by another binary process on 1 or -1 (positive indicates that more liquidity cost should be paid when trading volume q and negative is related to less liquidity cost situation) and is estimated by another auto-logistic model:

$$f(D_k = 1 \mid F_{k-1}, x_k, \overline{y_k}, A_k = 1) = p(\theta_k^D), \text{ where } p(\theta_k^D) = \frac{\exp(\theta_k^D)}{1+\exp(\theta_k^D)} \tag{3.7}$$

$$\text{and } \theta_k^D = \Pi_D^{'} M_{k-1}^D + g_k^D, \; g_k^D = \sum_{j=1}^{p} \gamma_j^D g_{k-j}^D + \sum_{j=1}^{l} \lambda_j^D D_{k-j}$$

$$\text{Consequently, } f(D_k = -1 \mid F_{k-1}, x_k, \overline{y_k}, A_k = 1) = \frac{1}{1+\exp(\theta_k^D)}$$

M_{k-1}^D is a vector including exogenous variables of subset F_{k-1} and Π_D is a parameter vector. It should be noted that the vectors and M^A and M^D might have some identical exogenous variables.

Once the model is estimated, we use the Ljung-Box test to validate its ability of capturing the main features of data. The test will be applied to standardized residuals:

$$u_k^D = \frac{D_k - (2p(\theta_k^D) - 1)}{2\sqrt{p(\theta_k^D)(1 - p(\theta_k^D))}} \tag{3.8}$$

Finally, the last factor is Size which captures the magnitude of the change of XLM. Less than half of observations in the sample have no size change, that is, they stay at their previous level. Thus we will adopt a geometric process for size changes. The choice of geometric distribution is motivated by its simplicity and generality:

$$S_k \mid F_{k-1}, x_k, \overline{y_k}, A_k = 1) \sim 1 + \mathrm{g}(\lambda_k) \tag{3.9}$$

$$\lambda_k = \frac{\exp(\theta_k^{Siz})}{1+\exp(\theta_k^{Siz})},$$

$$\theta_k^{Siz} = \Pi_{siz}^{'} M_{k-1}^{Siz} + g_k^{Siz} \text{ and } g_k^{Siz} = \sum_{j=1}^{p} \gamma_j^{Siz} g_{k-j}^{Siz} + \sum_{j=1}^{l} \lambda_j^{Siz} S_{k-j}$$

where M_{k-1}^{Siz} is a vector of exogenous variables and Π_{siz} is the corresponding parameter vector. $g(\lambda_k)$ indicates the geometric distribution with parameter λ_k [6]. In order to capture the asymmetry between up-move size and down-move size, we add a direction variable in the vector of exogenous variables. In Equation (3.9), we add one to the geometric distribution since the minimum change is one. We also apply the Ljung-Box statistics to standardized residuals to evaluate the model. Given the conditional distribution of Size, we have

$$E(S_k - 1 \mid F_{k-1}, x_k, \overline{y_k}, A_k = 1)) = \frac{1-\lambda_k}{\lambda_k} \tag{3.10}$$

$$Var(S_k - 1 \mid F_{k-1}, x_k, \overline{y_k}, A_k = 1)) = \frac{1-\lambda_k}{(\lambda_k)^2}. \tag{3.11}$$

Standardized residuals is computed as

$$u_k^{Siz} = \frac{S_k - 1 - E(S_k - 1 \mid F_{k-1}, x_k, \overline{y_k}, A_k = 1)}{\sqrt{Var(S_k - 1 \mid F_{k-1}, x_k, \overline{y_k}, A_k = 1)}} \tag{3.12}$$

and an adequate modeling requires u_k^{Siz} being uncorrelated with zero mean and unit variance.

In summary, for the estimation process, we can separately estimate each factor by using the Maximum Likelihood approach and the BIC criteria will be applied for the model selection, especially for the choice of number of lags. Moreover, in order to test their ability to capture the main feature of time series data, we perform a Portmanteau test in which residuals will be served to calculate the Ljung-Box statistic as a measure of residual dependence.

[6] The general probability distribution function is $f(x = m) = \lambda(1-\lambda)^m, m = 0,1,2,\ldots$.

Then, with the previous specification, all the observations can be classified into one of the three following categories:

1) There is no change in XLM, that is, activity factor $A_k = 0$ and no direction and size factors.

2) Liquidity decreases and the change size is at least one unit. The corresponding factors are : $A_k = 1$, $D_k = 1$, and $S_k = s_k$

3) Liquidity increases and change size is at least one unit. The corresponding factors are : $A_k = 1$, $D_k = -1$, and $S_k = s_k$

The overall log likelihood function is:

$$L(\omega_1, \omega_2, \omega_3, \omega_4, \omega_5; x_k, \overline{y_k}, z_k^q) =$$

$$\sum_{k=1}^{n} \begin{bmatrix} \log[g(x_k \mid F_{k-1}; \omega_1)] + \log[f^{Dur}(\overline{y_k} \mid x_k, F_{k-1}; \omega_2)] \\ +I_k(1) \cdot [\log(f^A(A_k \mid F_{k-1}; \omega_3))] \\ +I_k(2) \cdot [\log(1 - f^A(A_k \mid F_{k-1}; \omega_3)) + \log(f^D(D_k \mid F_{k-1}; \omega_4)) + \log(f^S(S_k \mid F_{k-1}; \omega_5))] \\ +I_k(3) \cdot [\log(1 - f^A(A_k \mid F_{k-1}; \omega_3)) + \log(1 - f^D(D_k \mid F_{k-1}; \omega_4)) + \log(f^S(S_k \mid F_{k-1}; \omega_5)) \end{bmatrix} \quad (3.13)$$

where $\omega_1, \omega_2, \omega_3, \omega_4, \omega_5$ are parameter sets for factors trade duration, quote duration, activity, direction and size respectively. In addition, $I_k(1)$, $I_k(2)$, $I_k(3)$, correspond to the indicator function relating to the three categories mentioned above.

The advantage of this modeling is that this partition enables us to simplify the modeling and computation task by specifying the suitable econometric models for the marginal densities of trade duration and conditional densities for quote duration and factors such as Activity, Direction and Size. In addition, for different purposes, the model could also be extended to a more or less complicated context by including other factors. In these decomposition models, one of the crucial tasks is to identify the exogenous variables. Apart from the irregularly spaced duration, it is reasonable to test whether the rest of the factors contain also seasonality effects.

3.3.2 Exogenous Variables Set

Given the model defined above, we need to identify the possible exogenous variables, apart from own lags, for each component. Our goal is to find some variables that have economic interpretation. In previous literature, the most widely used variables are spread, trading volume and price (Hasbrouck (1996), Goodhart and O'Hara (1997), Coughenour and Shastri (1999) and Madhavan (2000)). The intuition is that, trading activities and LOB trader behavior are related. For instance, in a volatile trading period, trading volume will increase and trade duration and quote duration will decrease, consequently, these will generate a volatile open LOB.

The first exogenous variable is relative spread change which is computed with the following formula:

$$RelativeSpread_k = 100 \cdot (ln(ask_k) - ln(bid_k))$$

Its variation is measured by:

$$DeltaSpread_k = RelativeSpread_k - RelativeSpread_{k-1}$$

where *ask* and *bid* are the best sell price and buy price available in open LOB. The advantage of relative spread is that it is dimensionless and can be used to directly compare different stocks.

As the relative spread captures quasi-instantaneous information and might be noisy, another spread-related variable is the average relative spread over most recent ten observations:

$$AveSpread_k = \frac{1}{10}\sum_{i=1}^{10} RelativeSpread_{k-i}$$

Regarding to the volume dimension, the first variable we take as the exogenous variable is the square root of the volume, *SquareRoot(vol)*, that initiates the current trade. There are two reasons for the use of square root, one is to down weight the large trade volume, and the second is that the price impact proves to be a concave function of market order size (Hasbrouck (1991)). If the volume that initiates the current trade is large, we expect a volatile situation and, by consequence, the trade duration and quote duration are likely to be short.

The second volume-related variable is to capture the imbalance of the signed trade. For this end, we adopt the depth measure proposed by Engle and Lange (2001) which is defined as follows:

$$Abs(sign.vol)_k = \left| \sum_{i=1}^{10} sign_{k-i} volume_{k-i} \right|$$

where the $sign_{k-i}$ and $volume_{k-i}$ are the trade sign and trade volume for the $(k\text{-}i)$th trade. The trades are classified into buy-initiated and sell-initiated according to the rule of Lee and Ready (1991). Intuitively, when the depth measure increases, it indicates that the trades are imbalanced and the market is dominated by one-side pressure.

The third dimension is the duration. Different from trade duration and quote duration, we also define two sorts of duration: back-quote duration and quote-quote duration. The back-quote duration is used to consider the duration between the first update of LOB after the previous trade and the following trade which contains quote information. It should be noted that the back-quote duration could be zero due to the fact that the quote duration might be censored when the trade occurs before the update of open LOB. The way by which the data is sampled ignores some quotes when there are more than one updates between two trades. It might not be a concern when 75% of the quotes are preserved as in Engle and Lunde (2003). However, in a market where the open LOB is more active such as the Xetra trading system, ignoring the quote activity may be a concern. In fact, only around 20% of the quotes are preserved in our dataset.

Another exogenous variable is quote-quote duration which considers the duration for which there is no change of XLM. By consequence, this variable will be used only in explaining the components such as direction and size which have observations when the liquidity measure changes.

The above variables will be all (or partially) included in the exogenous variable vectors for different components. In addition, we also put time-of-day dummy variables in the vectors to remove the seasonality which is a stylized fact in high-frequency data. There exist several techniques for this end. In our study, we use eight time-of-day dummy variables to capture the intraday seasonality for each component for which we desire modeling. Among the dummy variables, one for the first half hour after the market opening and then one for each hour of the trading day until the market closing.

Table 3.1 presents the summary statistics for durations and exogenous variables. The sample period is the first week of July 2010. The number of trades for all stocks ranges from 8 256 to 19 488. The trade frequency is also confirmed by the corresponding average trade durations. That is, stocks with larger number of trades correspond to shorter trade durations. Regarding to the average trade volumes, there is a big difference varying from 193 to 746 shares meaning that the selected stocks have different levels of liquidity. The average quote durations are relatively small and evidence that the dynamics in open LOB is more active than that of trades. Considering other exogenous variables, it is natural to see that the averages of *DeltaSpread* and ΔXLM are close to zero. As *AveSpread* is dimensionless, we can consider it as an indicator of the transaction cost. All stocks have a average of *AveSpread* around 0.05% meaning that the average of spread remains stable across the stocks. The average of trade imbalance variable *Abs(sign.vol)* varies from 944 to 3616 indicating the existence of different trading patterns across the stocks.

[Insert Table 3. 1 here]

3.4 Estimation and results

3.4.1 Trade Duration Equation

The log-likelihood is maximized by using the quadratic hill climbing method and the maximization program is run with Matlab v7.6.0 with Optimization toolbox. The estimate results of trade durations are presented in Table 3.2. In the modeling, we decompose the explanatory variables of trade durations into two groups: lagged dependent variables and exogenous variables. More specifically, the lagged dependent variables are used to capture the high degree of persistence in the trade durations. One part of exogenous variables will test the effect of mark variables on trade durations, and the other part with dummy variables helps to remove the seasonality.

[Insert Table 3.2 here]

Panel A of Table 3.2 presents the estimation results of lagged dependent variables for all the stocks in the sample. The ACD-related models are based on the ARMA structure meaning that the determinant part of the dynamics of durations are specified by the lagged duration and lagged error term[7]. As we expected, the trade durations of all stocks are highly autocorrelated. Taking the model efficiency and parsimony into consideration, the number of lagged trade durations varies from one to three. The sum of the coefficients of lagged trade durations for all stocks is around 0.9 which confirms a very high degree of persistence in autocorrelation. TKA has the most persistent trade durations among all stocks and has been specified by a Log-ACD(3,2) dynamics. Regarding to the error terms for all stocks, the parameters *Gamma1* (ranges from 0.512 to 0.974) and *Gamma2* (ranges from 0.350 to 0.780) are small indicating that the distribution of the random part is abrupt and near to zero, that is, expected trade duration shocks are very small.

Considering the exogenous variables mentioned above, we put six such variables in the trade duration equations: *DeltaSpread*, *AveSpread*, *SquareRoot(vol)*, *Abs(sign.vol)*, *BackQuote duration* and ΔXLM. Panel B of Table 3.2 reports the estimation results. Consistent with Engle and Lunde (2003), for all the stocks, an increase in spread, that is, a positive *DeltaSpread*, increases the trade duration. The parameters are all positive and significant at 5% level. Intuitively, when the stocks become less liquid, traders will slow their trading intensities, by consequence, trading durations will become longer. Different from Engle and Lunde (2003), we also find a positive sign for *AveSpread* coefficient. That is, the higher the *AveSpread* is, the longer the trade durations are. The difference between *DeltaSpread* and *AveSpread* is that the former captures the short-run spread change and the latter relates to the long-term spread change. It reveals that trade

[7] Regarding to the model selection, the first criteria is the Ljung-Box statistics, and the second one is the BIC criteria for the selection of number of lags.

intensity is sensitive for both short-run and long-run spread change. When the stocks become less liquid, traders slow down their activities and then wait for next moment when the liquidity increases.

Other important exogenous variables are volume-related. For all stocks, the coefficients of *SquareRoot(vol)* are all negative and significant at 5% level. This means that large trades will generate higher trading intensity and shorter trade durations. According to Easley and O'Hara(1987), large trades are likely to be related to information trade since the informed traders try to exploit their information advantage by increasing both the trading intensity and trading quantity. Another volume-related variable is *Abs(sign.vol)* which measures the volume imbalance (in Euros) of the last ten trades. As we take the absolute value, a higher value means a higher level of imbalance and a zero value means a complete balanced trading activity. All stocks have small positive coefficients, and only two of them are significant at 5% level. The measure of imbalance also relates to the information trading. Our results suggest that when the trades become imbalanced, the trade intensity will decrease. Presumably, when the non-informed traders observe an imbalanced trading history, they will then slow down their trading activity to protect themselves.

Regarding the *BackQuote duration*, we find that the effect of duration between the first quote and the following trade on the trade duration is not clear. Two stocks have positive sign and two others have negative sign. Moreover, three of them are significant at 5% level. Intuitively, the trade activity should be positively correlated with quote activity. As a result, the coefficient should have positive sign. Our results do not seem to be completely consistent with this intuition. One explanation might be that as the trade duration consists of quote duration and *BackQuote duration* depends on both trade duration and quote duration. Therefore, the explanatory power of *BackQuote duration* is also influenced by quote duration which is censored. For instance, a longer *BackQuote duration* could be either a long trade duration or a shorter quote duration. The longer trade duration has a positive impact on trade duration, whereas a shorter quote duration has a negative one. Therefore, the sign of *BackQuote duration* on trade duration could be positive or negative.

The last exogenous variable is the change of liquidity measure XLM. Like *DeltaSpread*, a high XLM means a less liquid situation. If we follow the same argument as *DeltaSpread*, higher XLM will decrease the trade intensity and then generate longer trade duration. Our estimate results suggest that three stocks have positive estimated coefficients and one has negative estimated coefficient. However, only one is significant at 5% level. Generally speaking, the XLM is not likely to have effect on the trade duration. However, if we combine the XLM result with that of *DeltaSpread*, it suggests that the trade activity is more sensitive to the first level of LOB and the quote activities beyond first level are not likely to explain the trade duration dynamics.

The last part consists of dummy variables. Panel C of Table 3.2 reports the estimated results for these dummy variables. To make the results more visible, we piecewisely plot the coefficient of dummy variables in Figure 3.2 In line with previous literature, the trade durations do have a diurnal pattern. For all stocks, there is an inverse U-shaped pattern for trade durations.

[Insert in Figure 3.2 here]

The model is validated by the Ljung-Box statistic. Table 3.3 presents the value of Ljung-Box statistic at different lags for trade durations and the standardized residuals of trade durations. The left side of the table shows that there is a high persistence in autocorrelation of the trade duration dynamics. The right side of the table presents evidence that the Log-ACD model is capable of removing this autocorrelation feature in the trade durations since the Ljung-Box statistics have been reduced dramatically and, in most of cases, the hypothesis of no autocorrelation cannot be rejected.

[Insert Table 3.3 here]

The overall results on trade durations are stable across stocks and provide some new empirical evidence about the trade duration dynamics. First, trade durations are highly persistent. Second, volume-related variables and best level quote have more explanatory power in predicting the expected trade duration. However, the quote duration and XLM have less influence on trade durations. Third, trading durations do perform an intraday pattern.

3.4.2 Quote Duration Equation

The estimated results for quote duration are presented in Table 3.4. Similar to trade duration, we also decompose explanatory variables of quote duration into two parts: lagged dependent variables and exogenous variables.

[Insert Table 3.4 here]

Panel A of Table 3.4 reports the estimate results of lagged dependent variables for all stocks. Similar to trade durations, quote durations are also highly persistent. The number of lagged value varies from two to three and the number of lagged error terms varies from one to four. TKA performs the most persistent dynamics and takes a Log-ACD(3,4) model to fit its dynamics. Comparing the Log-ACD model to the traditional ARMA structure, the sum of the coefficient of "AR" part in quote duration equation varies from 0.239 to 0.505. The relatively small coefficients can be explained by the fact that the start point of our quote duration is the trade timestamp, when there are more than one update between two trades, only the first one is used to calculate the quote duration. To some extent, this sampling "deletes" some autocorrelation in quote durations. Regarding to the coefficients of error term, the estimated coefficients vary from 0.042 to 0.072 confirming the sampling effect in autocorrelation.

We include more exogenous variables for quote duration than in trade duration since we assume that the trade durations are exogenous and can explain the quote duration dynamics. More specifically, the exogenous variables we use in explaining the dynamics of quote durations are: trade-duration-related variables, censored effect variable, *DeltaSpread*, *AveSpread*, *SquareRoot(vol)*, Abs(sign.vol), *BackQuote duration* and ΔXLM. Panel B of Table 3.4 presents the corresponding estimate results.

As in Engle and Lunde (2003), the censored effect is captured by the product of censored dummy variable and lagged error term. However, our censored effect is positive and significant at 5% level for all stocks. The implication is that if the last quote duration is censored, the next quote duration is likely to be longer and censored again. Trade duration related variables consist of current error term, lagged error term and lagged expected trade duration. The estimate results show that the current error term and expected trade duration have positive effect on the quote durations, the coefficients are positive and significant. It suggests that when the (current and lagged) trade duration innovations and expected trade durations are high, the quote duration will also become longer. Intuitively, as quote activity adjusts to trade activity, trade duration and quote duration are correlated positively.

The spread-related variables, *DeltaSpread* and *AveSpread*, have negative impact on the quote durations, which is opposite to the trade duration. The coefficients are negative and significant at 5% level for all stocks. Spread variables are very important in explaining the trade activity and quote activity. It suggests that when the spread is large, the market order traders and LOB traders react in different fashions. Market order traders slow down their trading speed when observing an increasing spread, whereas LOB traders speed up to update their price or quantity of limit orders. As an intraday LOB trader, a main risk to face is the adverse selection risk. By monitoring attentively the change of spread, LOB traders attempt to avoid this risk by updating their quote rapidly.

The effect of volume-related variables, *SquareRoot(vol)* and *Abs(sign.vol)*, on quote duration is similar to that of trade duration. More specifically, the large trades predict shorter quote durations. As mentioned before, large trades relate to informed trades. Informed traders exploit short-lived information by increasing trade intensity and trade quantity. When trades become more intensive, so do quote revisions. As a result, the quote durations become shorter. Regarding to imbalanced measure *Abs(sign.vol),* its coefficients are all positive but only two of them are significant. Compared to *SquareRoot(vol), Abs(sign.vol)* is a long-run variable as it consists of information over last ten trades. In a high-frequency trading framework, trades and information are short-lived. The estimate results suggest that the short-run measure is more predictive than the long-run measure. The positive coefficient of *Abs(sign.vol)* is not very intuitive if we consider the trade imbalance as the result of information asymmetry. One explanation might be that the positive effect is the evidence of the high-frequency algorithm trading. The algorithm (probably not human) traders "enjoy" providing the liquidity by creating this trade imbalance. For example, if an algorithm trader desires buying a given quantity of stocks, at ask side, he can place limit orders in different price levels with different quantities. As a result, the pressure on sell side increases and then more trades occur on the bid side. In this case, trade imbalance is not a fear and quote revision can be postponed.

The *BackQuote duration* has a positive and significant impact on the following quote duration. A longer *BackQuote duration* can be either a low trade intensity or a high quote intensity. That is, a long trade duration or a short quote duration. Moreover, a long trade duration implies a long quote duration. The estimated results suggest that the long trade duration effect dominates the short quote duration effect. Concerning the effect of liquidity measure XLM, empirically, the effect is negative and significant at 5% level for three stocks. One stock has a positive significant sign. Theoretically, there are arguments to support either sign. A high XLM reveals a more risky situation, especially a high adverse selection risk. LOB traders speed up their quote revisions to reduce this kind of risk. This is more pronounced for less liquid stocks. On the other side, the increase of XLM can be interpreted as the evidence of algorithm trading. The objective is to provide liquidity at different price levels without updating the quote rapidly. Panel C of Table 3.4 reports the results of estimation for dummy variables. We also find a seasonality effect for quote durations. However, as the quote durations are censored and the trade durations enter in the model of quote durations, the pattern of the seasonality for quote durations is not like that of trade durations. Figure 3. 2 illustrates the intraday seasonality pattern for different stocks.

[Insert Figure 3. 2 here]

To test the model, we present the Ljung-Box statistics on quote duration and the deduced standardized residuals. The results are reported in Table 3.5. The left side of the table shows that there is a huge autocorrelation in the quote durations, similar to the trade durations. Comparing the two sides of the table, we find that the Ljung-Box statistics have been largely reduced and, in most of cases, the hypothesis of no autocorrelation cannot be rejected. It should be noted that, despite of the large number of the lagged variables, TKA still has a relatively high autocorrelation in the standardized residuals.

[Insert Table 3.5 here]

3.4.3 Activity Factor

Up to now, we have analyzed the dynamics of trade durations and quote durations. Conditional on the time dimension, we can further make the analysis of liquidity dimension. As we mentioned above, we decompose the change of the liquidity measure XLM into three components: Activity, Direction and Size. Similar to the time dimension, we also put two groups of explanatory variables into the activity equation: lagged dependent variables and exogenous variables. The M^A vector includes expected trade duration, expected quote duration, *DeltaSpread*, *AveSpread*, *SquareRoot(vol)*, *abs(sign.vol)*, *BackQuote duration*, ΔXLM and dummy variables.

[Insert Table 3.6 here]

Table 3.6 reports the estimated results for activity factor. The activity process is a binary process in which $A_k = 1$ means the change of liquidity measure XLM. To model the activity factor, we adopt the GLARMA structure introduced by Rydberg and Shephard (2003). In addition to GLARMA structure, we also include exogenous variables. Panel A of Table 3.6 presents the estimated results of GLARMA part. For all stocks, we use a lag set of (1,2) to capture the autocorrelation of activity factor. Consistent with previous literature, the coefficients of "GLAR" part are positive and significant, ranging from 0.66 to 0.95. The result suggests a high persistence in autocorrelation for activity factor. More specifically, there is a cluster effect in activity, namely, the change of liquidity is more likely to be followed by another change.

In this study, we are also interested in the effect of exogenous variables on the dynamics of liquidity. Panel B of Table 3.6 shows the estimated results for these exogenous variables. For the time dimension variables, expected trade duration and expected quote duration do not have the same effect on the probability of liquidity change. In particular, a longer expected trade duration increases the probability of liquidity change, whereas a longer quote duration decreases this probability. In the tick-by-tick trading framework, as found by Dionne, Duchesne and Pacurar (2009), a longer trade duration will have a positive impact on price volatility. As a result, a longer trade duration will increase the probability of liquidity change. On the other hand, quote duration measures the quote intensity. A longer quote duration means a less active open LOB. By consequence, the quote is likely to be unchanged.

Regarding to the spread-related variables, both *DeltaSpread* and *AveSpread* have positive effect on the probability of liquidity change. These two measures are related to the liquidity itself. When the liquidity decreases, LOB traders are more prudential in their quotes and then likely to update their quotes, therefore, the probability of the liquidity change increases. Recall that, in trade duration, *DeltaSpread* and *AveSpread* also have positive effect on trade duration. We find that different types of traders use different ways to protect themselves, more concretely, when market is less liquid, market order traders become less active and LOB traders become more active and prudential. It should be noted that the activity factor only tells us whether the liquidity changes or not, there is no information on the direction and magnitude of change.

Regarding to volume-related variables, *SquareRoot(vol)* and *Abs(sign.vol)* affect the probability of liquidity change in different ways. As expected, the coefficient of *SquareRoot(vol)* is positive meaning that the large trades will increase the probability of liquidity change. As mentioned before, large trades are likely to be informative and increase the volatility. Under this circumstance; the LOB traders are more likely to review their quotes and then the liquidity changes. As for trade duration and quote duration, the effect of *Abs(sign.vol)* is again not clear. Only one stock has significant coefficient. The results suggest that the probability of liquidity change hardly depends on the long-run trade imbalance. Possible explanations are that: First, *Abs(sign.vol)* captures last ten trades information and the activity component has very short memory; Second, algorithm traders might provide the liquidity by “creating” this trade imbalance.

Another time dimension variable, *BackQuote duration*, also has a positive impact on the probability of liquidity change. This is in line with the estimate results for trade duration. As we can see from Panel B of Table 3.6, the effect of expected trade duration is higher than that of expected quote duration. Therefore, the longer *BackQuote duration* implies a more volatile market and liquidity is likely to be updated. For the liquidity measure *XLM*, as expected, the coefficient is positive and significant at 5% level. This means that when the market is less liquid, there is more chance that LOB traders review their quote and then the liquidity of this stock changes. Concerning the dummy variables, we find that there is a seasonality pattern only for some stocks in certain time periods, most of periods do not perform seasonality pattern. The results are shown in the Panel C of Table 3.6.

Similar to trade durations and quote durations, we use the Ljung-Box statistics to validate the model. The statistics for activity factor and the corresponding standardized residuals are reported in left side and right side of Table 3.7, respectively. On one hand, the activity factor is also highly autocorrelated. At five-lag level, the Ljung-Box statistics ranges from 358 to 1127 for all stocks and the hypothesis of no autocorrelation is rejected at any confidence level. On the other hand, we find that the model which includes the GLARMA part and exogenous variables can capture this autocorrelation feature very well. On the right side of the table, all statistics have been reduced less than the critical values except one for MRK.

[Insert Table 3.7 here]

3.4.4 Direction Factor

Another component of liquidity measure is direction which is also a binary process: value 1 means the decrease of liquidity and -1 means increase of liquidity. Conditional on the activity, the direction factor gives more information about the change of XLM. To capture the dynamics of direction, we use the similar variables set as for the other factors. Three stocks take a GLARMA(2,1) structure and one stock takes GLARMA(3,1). Panel A of Table 3.8 illustrates the estimated results on GLARMA structure. Interestingly, the sums of the "MA" part are all negative and smaller than -0.5

indicating a mean-reverting feature. That is, the increase of liquidity is likely to be followed by a decrease of liquidity, vice versa.

[Insert Table 3.8 here]

The exogenous variables we include in the direction factor are: *QuoteQuote duration*, *DeltaSpread*, *AveSpread*, *SquareRoot(vol)*, *Abs(sign.vol)*, *BackQuote duration* and ΔXLM. Panel B of Table 3.8 presents the estimate results of these exogenous variables. To explain the direction factor, the new variable *QuoteQuote duration* is defined as the duration between two liquidity measure changes. As the direction component is observed only when the activity factor equal to one, it is more reasonable to put a temporal variable to capture this time interval. Based on Panel B of Table 3.8, we find that three stocks have positive coefficient for *QuoteQuote duration* and only one is significant. It appears that *QuoteQuote duration* is not a relevant variable in predicting the direction factor.

Consider the spread related variables, a high *DeltaSpread* increase the probability of liquidity decrease if there is a liquidity change. Intuitively, when the spread increase, this means the LOB traders keep off from mid-quote and then XLM is likely to increase. Compared to *DeltaSpread*, the *AveSpread* has the opposite effect on the direction factor. It suggests that the permanent increase of the spread will indeed increase the probability of liquidity increase. It seems evident that when the traders have to pay a higher liquidity premium, the LOB traders (i.e, liquidity providers) are willing to provide liquidity.

Regarding to the volume-related variables, coefficient of *SquareRoot(vol)* is positive meaning that the current large trades predict a less liquid situation, which is consistent with previous literature. Intuitively, the large trade is likely to be informed trade which will create volatility in the market. Accordingly, LOB traders are likely to keep away from mid-quote and the liquidity will decrease. However, the trade imbalance variable *Abs(sign.vol)* has no significant effect on the direction factor. This confirms the results in activity component. As there is no clear impact of *Abs(sign.vol)* on liquidity change, by consequence, the effect of *Abs(sign.vol)* on direction is not clear. As we focus on the dynamics of liquidity measure XLM, the reason to the lack of information content of trade imbalance is out of our scope. However, it might be interesting for trading strategy analysis or algorithm trading to continue exploring this aspect.

Another temporal variable *BackQuote duration* has a positive impact on the probability of liquidity increase. This suggests that when the trades become less active, the LOB traders are likely to review their quotes and incite the trades by providing more liquidity. Regarding to liquidity measure XLM, an increase of XLM will increase the probability of being liquid. This is the evidence of the mean-reverting in XLM dynamics. As other components, we also use dummy variables to capture the seasonality in direction factor. The estimated results show that the coefficients are not significant meaning that there do not exist a clear seasonality effect in direction factor. The results for coefficients of dummy variables are presented in Panel C of Table 3.8.

The direction factor is observed when the liquidity measure XLM changes. The Ljung-Box statistics for direction factor and corresponding standardized residuals are reported in left side and right side of Table 3.9, respectively. Similar to the activity factor, the direction factor is highly autocorrelated. The hypothesis of no autocorrelation is rejected at any confidence level for all stocks. After modeling the Ljung-Box statistics for standardized residuals from 5 to 200 lags are not significant at the 5% level.

[Insert Table 3.9 here]

3.4.5 Size Factor

The last factor is the size of liquidity change. Table 3.10 reports the estimated results. Panel A of Table 3.10 shows the results of GLARMA part. We find that the number of lags of GLARMA part ranges from (2,1) to (2,2). Depending on stocks, the effect of lagged value is either positive or negative. As shown in equation (3.10), a higher λ_k indicates a smaller expectation of size factor. Combined with the estimated results of direction component, the implication is that the XLM does have a mean-reverting feature, however, the magnitude of reverting varies from one stock to another.

[Insert Table 3.10 here]

The exogenous variables are the same as in the direction factor. In addition, we put current direction and lagged direction to explain the size factor. In total, there are nine exogenous variables for the size factor. Panel B of Table Table 3.10 reports the estimated results of exogenous variables. In line with estimated results for direction factor, *QuoteQuote duration* does not have significant effect for all stocks. Recall that *QuoteQuote duration* is the duration between two XLM changes. Again, it evidences that the LOB structure only has short memory and does not have predict power on the dynamics of size change.

To explain the size factor, current direction is used to investigate if there exists a leverage effect, that is, the negative change (less liquid) is higher than positive change (more liquid). Consistent with Rydberg and Shephard (2003), the current direction variable has negative and significant effect on the λ_k, which implies a higher expected liquidity change when the liquidity decreases and confirms the leverage effect. However, the lagged direction is less significant than the current one. The spread-related variable, *DeltaSpread*, has positive impact on λ_k and only two of them are significant. Combined with the results of direction factor, it shows that when the lagged spread increase, the probability of direction equal to one increases and stock is likely to be more illiquid, however, the magnitude of change for this illiquidity is likely to be smaller. Only for TKA, *AveSpread* has a negative significant effect on λ_k. For the rest of stocks, the *AveSpread* has no significant effect on λ_k. The implication is that as the dynamics of liquidity only has short memory, the long-run variable is less significant. The negative effect of *AveSpread* on λ_k, combined with the results of direction factor, means that the permanent increase of spread will increase the liquidity, moreover, the magnitude of this increase is likely to be higher, that is, the stock will become more liquid.

Regarding to the volume-related variable, *SquareRoot(vol)* is not significant for three stocks but significant for TKA. Recall that the value of size component is defined as the size change minus one. Therefore, the nonsignificance means that the size component changes less than one unit. Up to now, we can quantify more about the effect of trade quantity on the liquidity change. The large trades do increase the probability of illiquidity, however, the effect of large trade on the magnitude of liquidity change seems limited for three stocks. Only for TKA, the effect of large trade on size factor is significantly more than one unit. The effect of trade imbalance variable *Abs(sign.vol)* is not significant for all stocks. Taking into consideration of the nonsignificance of *Abs(sign.vol)* in direction, this again confirms the no informativeness of trade imbalance.

The temporal variable *BackQuote duration* has some positive and significant effect on λ_k indicating that even though the liquidity provider try to incite the traders to trade by increasing the liquidity provision (conclusion from results of direction factor estimation), the magnitude of liquidity increase is limited. The last variable ΔXLM itself has a negative and significant effect on λ_k. As mentioned before, liquidity measure XLM has the mean-reverting feature, the negative effect implies that when there is a mean-reverting in liquidity measure, the magnitude has tendency to stay the same. More specifically, it is likely to observe the similar or symmetry magnitude in the mean-reverting of liquidity measure. We also attempt to test if there is a seasonality in size factor by using dummy variables. As shown in Panel C of Table 3.10, similar to activity and direction factors, there is little seasonality feature in size component since most of the coefficients are not significant from zero at 5% level.

Same as the direction factor, the size factor is also observed when the liquidity measure XLM changes. Table 3.11 reports the Ljung-Box statistics of size factor and the corresponding standardized residuals. Interestingly, the statistics for size factor vary significantly across the stocks. For instance, the size factors of RWE and TKA, with Ljung-Box statistics of 1406 and 889 for 5 lags, are relatively high autocorrelated. Given the high Ljung-Box statistics of standardized residuals for stocks RWE and TKA at small lags (from 5 lags to 20), the model might have the misspecification problem on θ_k^{Siz} or a mild distributional failure. It should be noted that θ_k^{Siz} can be specified in many different ways and the distribution for size factor can be others than geometric distribution. However, a parsimonious and interpretable model is always preferred.

[Insert Table 3.11 here]

3.4.6 Effect of exogenous on XLM changes

After estimating each factor, we are able to obtain a more general understanding on the dynamics of ex-ante liquidity measure XLM. The results are shown in

Table 3.12. Specifically, all exogenous variables will increase the probability of having a change in XLM measure except the expected quoted duration and *Abs(sign.vol)*. Longer expected quoted duration relates to a less active LOB, by consequence, the probability of change on XLM decreases. The *Abs(sign.vol)* is less informative on dynamics of XLM, this is further confirmed by the insignificance of coefficient for direction and size factors.

[Insert Table 3.12 here]

Conditional on the activity factor equal to one, exogenous variables will have different impact on XLM changes. Particularly, if the non-change time is longer, the XLM will be likely to increase, that is, the stock will become less liquid but the magnitude is limited to one unit (in our case, 0.25 basic point). The previous direction has a negative effect on the next direction factor when activity factor being one, moreover, the size of change will decrease for both directions. Previous short-run delta spread will cause a less liquid situation for the next change in which the size of change is like to increase or limited to one unit depending on the stock characteristics. Interestingly, XLM will change in the opposite direction of long-run liquidity measure *AveSpread*. It suggests that the future short term liquidity has trend of reverting to its average level, however, the speed of this reverting related to size factor varies according to the stock characteristics.

Higher *SquareRoot(vol)* will be likely to cause a less liquid situation, however, the effect of volume is limited corresponding to a non-significant or negative parameter. As we have showed in previous estimation, the imbalance trade has little effect on next liquidity change, it suggest that the use of imbalance trade as a predictor of liquidity might be more accurate for the low-frequency framework. Recall that the *BackQuote duration* is the time between first update of LOB after a trade and the next trade. From a perspective of liquidity traders, it measures the time of liquidity resilience. A shorter *BackQuote duration* means the liquidity goes back to its original level quickly. Therefore, a longer *BackQuote duration* has negative impact on liquidity direction, however, the magnitude of this effect varies according to stock characteristics. XLM itself is mean-reverting as previous XLM has a negative effect on the next XLM changes, again the speed of mean-reverting changes across stocks.

3.5 Conclusion

After the introduction of the open LOB trading mechanism, trading frequency has become shorter than ever before. By consequence, the liquidity has become an important issue for active traders, investors and financial institutions. This paper make an analysis of the dynamics of ex-ante liquidity changes. The ex-ante liquidity measure used in this study is XLM proposed by Xetra trading system. Different from ex-post liquidity measure, the XLM is an ex-ante volume dependent measure. The computation of the measure requires the information such as the prices and the corresponding quantity available in the open LOB.

To model the dynamics of the liquidity changes, we adopt the decomposition approach proposed by Rogers and Zane (1998). The liquidity changes have been decomposed into five factors: trade durations, quote durations, activity, direction and size. Trade durations and quote durations are the temporal variables and are not synchronized. To solve this problem, we follow Engle and Lunde (2003) by defining the last trade time as the initial quote time in quote duration computation. The other three factors are directly related to the change of XLM itself. Both activity factor and direction factor are binary process taking value 0 and 1, or -1 and 1, respectively. The size factor captures the magnitude of XLM change. To investigate the dynamics of each component, we apply the relevant econometric models to each factor and put a wide range of trade-related exogenous variables into the models which include volume-related variables, trade-balance-related variables and temporal variables. The models are validated by the t-test of the coefficients and the Ljung-Box test on the deduced standardized residuals.

The essential findings are that first, most trade-related variables can influence the dynamic of trade durations, however, the impact of ΔXLM on trade duration is not clear. Moreover, the quote durations are influenced by the dynamics of trade durations, trade-related variables and ΔXLM. Second, expected trade durations and quote durations are likely to affect the probability of liquidity change. Moreover, most of trade-related variables have an impact on the activity factor except the long-run trade imbalance variable, *Abs(sign.vol)*. Third, temporal variable *QuoteQuote duration* seems not have impact on direction factor. Similar to the activity factor, most of trade-related variables can affect the direction factor except *Abs(sign.vol)*. Fourth, there is a leverage effect in size factor, that is, the magnitude of liquidity decrease is higher than that of increase. Among the trade-related variables, only spread change has significant effect on size factor for most stocks. The effects of the other trade-related variables are not significant. The variable ΔXLM has negative and significant effect on size factor meaning that the magnitude of liquidity change is likely to stay the same despite the mean-reverting feature. Fifth, the trade durations and quote durations have obvious seasonality pattern, whereas the seasonality pattern for other factors is not clear.

Future research can continue in several directions. Our study focuses on the impact of trade-related variables on the liquidity changes. A possible alternative is to investigate how the liquidity change co-moves with trades. Another direction is to decompose the liquidity changes in a different order or into different factors to answer other microstructure questions. It could also be interesting to generalize the model from one particular stock to a portfolio. Again, the no-synchronization of the trade durations and quote durations between different stocks is a challenge. It will require a more complicated econometric model and reasonable assumptions.

References

Aitken, M., Comerton-Forde, C., 2003. How should liquidity be measured? Pacific-Basin Finance Journal 11, 45-59

Bauwens, L., Giot, P., 2000. The Logarithmic ACD Model: An Application to the Bid-Ask Quote Process of Three NYSE Stocks. Annals of Economics and Statistics / Annales d'Économie et de Statistique, 117-149

Beltran-Lopez, H., Giot, P., Grammig, J., 2009. Commonalities in the order book. Financial Markets and Portfolio Management 23, 209-242

Beltran-Lopez, H., Grammig, J., Menkveld, A.J., 2011. Limit order books and trade informativeness. The European Journal of Finance, 1-23

Bilodeau, Y., 2013. Xetraparser [computer software]. HEC Montréal

Black, F., 1971. Toward a fully automated stock exchange. Financial Analysts Journal 27, 28-44

Coppejans, M., Domowitz, I., Madhavan, A., 2004. Resiliency in an automated auction. Working Paper

Coughenour, J., Shastri, K., 1999. Symposium on Market Microstructure: A Review of Empirical Research Financial Review 34, 1-27

Cox, D.R., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., Juselius, K., Lauritzen, S.L., 1981. Statistical Analysis of Time Series: Some Recent Developments [with Discussion and Reply]. Scandinavian Journal of Statistics 8, 93-115

Demsetz, H., 1968. The cost of transacting. Quarterly Journal of Economics 82, 33-53

Dionne, G., Duchesne, P., Pacurar, M., 2009. Intraday Value at Risk (IVaR) using tick-by-tick data with application to the Toronto Stock Exchange. Journal of Empirical Finance 16, 777-792

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

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

Engle, R.F., Lange, J., 2001. Predicting VNET: A model of the dynamics of market depth. Journal of Financial Markets 4, 113-142

Engle, R.F., Lunde, A., 2003. Trades and Quotes: A Bivariate Point Process. Journal of Financial Econometrics 1, 159-188

Engle, R.F., Russell, J.R., 1998. Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. Econometrica 66, 1127-1162

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., Kandel, E., 2005. Limit Order Book as a Market for Liquidity. Review of Financial Studies 18, 1171-1217

Giot, P., Grammig, J., 2006. How large is liquidity risk in an automated auction market? Empirical Economics 30, 867-887

Glosten, L.R., Harris, L.E., 1988. Estimating the Components of the Bid/Ask Spread. Journal of Financial Economics 21, 123 - 142

Goodhart, C.A.E., O'Hara, M., 1997. High frequency data in financial markets: Issues and applications. Journal of Empirical Finance 4, 73-114

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

Hasbrouck, J., 1996. Order characteristics and stock price evolution An application to program trading. Journal of Financial Economics 41, 129-149

Hausman, J.A., Lo, A.W., MacKinlay, A.C., 1992. An ordered probit analysis of transaction stock prices. Journal of Financial Economics 31, 319-379

Irvine, P.J., Benston, G.J., Kandel, E., 2000. Liquidity Beyond the Inside Spread: Measuring and Using Information in the Limit Order Book. SSRN eLibrary

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

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

Manganelli, S., 2005. Duration, volume and volatility impact of trades. Journal of Financial Markets 8, 377-399

McCulloch, R.E., Tsay, R.S., 2001. Nonlinearity in High-Frequency Financial Data and Hierarchical Models. Studies in Nonlinear Dynamics & Econometrics 5, 1-17

Parlour, C., A. , 1998. Price Dynamics in Limit Order Markets. The Review of Financial Studies (1986-1998) 11, 789

Rogers, L.C.G., Zane, O., 1998. Designing and estimating models of high-frequency data Unpublished paper

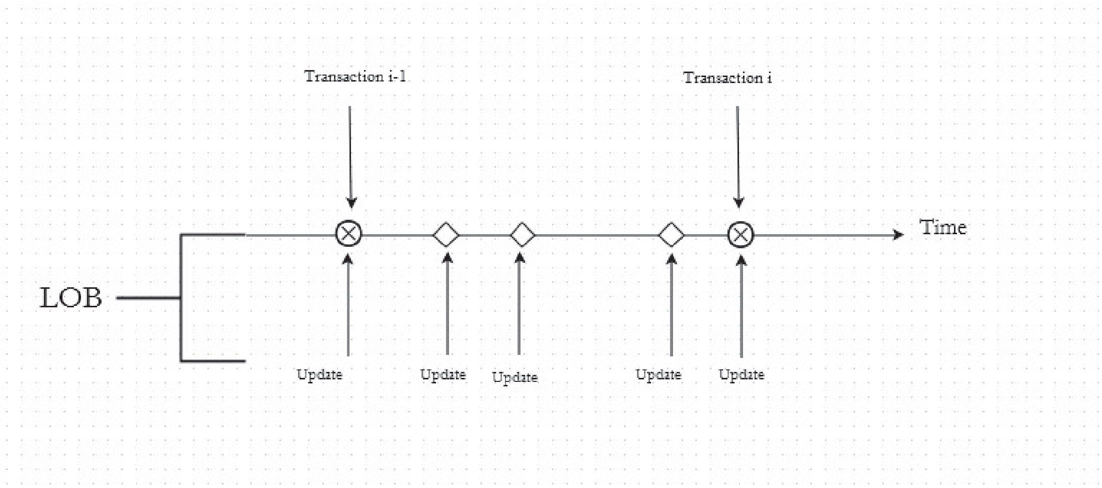
Russell, J., Engle, R., 2005. A Discrete-State Continuous-Time Model of Financial Transactions Prices and Times: The Autoregressive Conditional Multinomial-Autoregressive Conditional Duration Model. Journal of Business & Economic Statistics 23, 166

Shephard, N., 1995. Generalized linear autoregressions. Economics Group, Nuffield College, University of Oxford

Tina Hviid, R., Neil, S., 2003. Dynamics of Trade-by-Trade Price Movements: Decomposition and Models. Journal of Financial Econometrics 1, 2

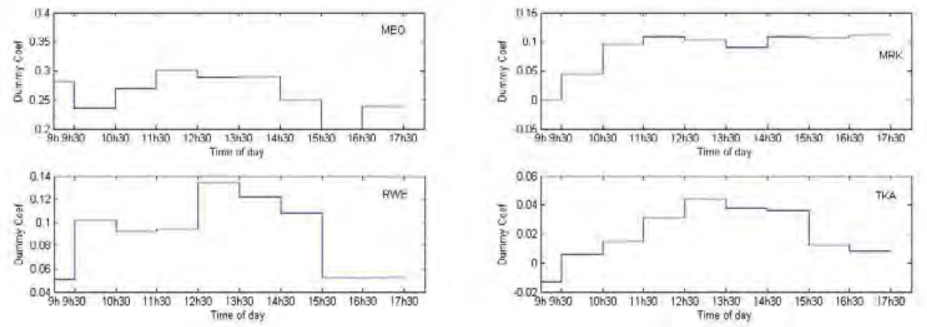
Viswanathan, S., Wang, J.J.D., 2002. Market architecture: limit-order books versus dealership markets. Journal of Financial Markets 5, 127-167

Figure 3. 1 Timestamps for trades and quote update in open LOB



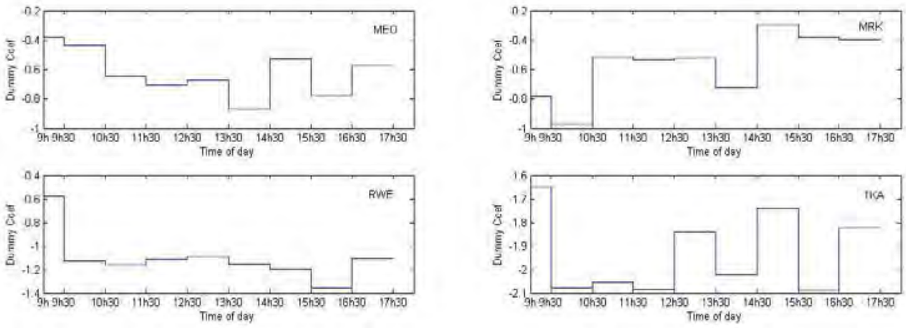
This figure presents the temporal relation between trades and quote updates. The trades and the quote updates are presented by circles and squares, respectively.

Figure 3. 2 Intraday Seasonality Pattern For Trade Durations



This figure presents the temporal relation between trades and quote updates. The trades and the quote updates are presented by circles and squares, respectively.

Figure 3. 3 Intraday Seasonality Pattern For Trade Durations



This figure presents the intraday patterns of seasonality variables for trade durations for different stocks.

Table 3. 1 Summary Statistics of Trade Related Variables and Exogenous Variables

	Trade Related and Duration Variables					
	Nb.Obs	Average Price (euros)	Average Volume (shares)	Ave.Trade Duration (seconds)	Ave.Quote Duration (seconds)	
MEO	10742	43.19	262.00	14.12	0.12	
MRK	8256	59.37	193.74	18.38	0.14	
RWE	16487	54.37	390.84	9.20	0.16	
TKA	19488	21.00	746.05	7.79	0.14	
	Average of Exogenous Variables					
	DeltaSpread %	AveSpread %	SquareRoot(Vol) (shares)	Abs(sign.vol) (euros)	BackQuote Dur (seconds)	ΔXLM (1/4 bps)
MEO	-1.52E-04	0.049	14.344	1322.493	14.008	-4.75E-03
MRK	-4.77E-05	0.049	12.309	944.487	18.231	-3.75E-03
RWE	-4.01E-05	0.041	17.021	1924.649	9.043	-1.03E-03
TKA	-4.01E-06	0.048	23.538	3616.967	7.653	-4.62E-04

The table reports the summary statistics for trade related variables, durations and exogenous variables. The Nb.Obs is the number of observations in the period of first week of July 2010. The lower part of the table presents the average of the exogenous variables we defined in Section 3.2. Instead of using basic point, we use1/4 basic point as one unit of liquidity change.

Table 3.2 Estimated Results For Trade Durations

		MEO			MRK			RWE			TKA		
		Est	STD	T-Stat	Est	STD	T-Stat	Est	STD	T-Stat	Est	STD	T-Stat
Panel A	ln(dur(lag1))	0.875	0.040	**21.904**	0.804	0.040	**19.958**	0.975	0.026	**37.655**	1.194	0.037	**32.168**
	ln(dur(lag2))	-0.021	0.034	-0.627	0.126	0.056	**2.239**	-0.130	0.037	**-3.481**	-0.306	0.062	**-4.969**
	ln(dur(lag3))	0.063	0.021	**2.958**	0.024	0.038	0.629	0.110	0.027	**4.045**	0.094	0.033	**2.848**
	x/phi (lag1)	0.081	0.008	**10.057**	0.060	0.006	**9.960**	0.119	0.007	**17.467**	0.130	0.009	**14.863**
	x/phi (lag2)										-0.059	0.009	**-6.810**
		0.512	0.021	**24.364**	0.515	0.024	**21.552**	0.974	0.026	**36.793**	0.798	0.019	**42.379**
		0.780	0.052	**15.077**	0.711	0.054	**13.170**	0.350	0.014	**25.582**	0.449	0.016	**27.562**
	cons	0.239	0.033	**7.147**	0.111	0.027	**4.128**	0.052	0.016	**3.282**	0.008	0.007	1.087
Panel B	delta spread	14.848	0.276	**53.892**	26.521	0.937	**28.303**	18.192	0.600	**30.336**	14.341	0.556	**25.782**
	lev.spread	1.146	0.187	**6.119**	1.232	0.226	**5.444**	0.807	0.176	**4.597**	0.488	0.084	**5.777**
	square root(vol)	-0.013	0.001	**-9.499**	-0.008	0.002	**-5.075**	-0.008	0.001	**-12.042**	-0.003	0.000	**-9.700**
	abs(sign.vol)	4.51E-06	4.94E-06	0.913	3.45E-06	7.05E-06	0.489	4.54E-06	1.64E-06	**2.768**	1.24E-06	5.30E-07	**2.340**
	Back_quote_dur	3.90E-04	5.53E-04	0.705	8.17E-04	3.25E-04	**2.512**	-0.002	0.001	**-3.337**	-0.002	3.92E-04	**-3.971**
	ΔXLM	0.019	0.014	1.404	-0.012	0.013	-0.902	0.019	0.007	**2.488**	0.001	0.013	0.109
Panel C	Dummy1	0.043	0.024	1.773	-0.112	0.020	**-5.492**	-0.002	0.011	-0.156	-0.021	0.006	**-3.690**
	Dummy2	-0.003	0.016	-0.174	-0.066	0.013	**-5.191**	0.049	0.008	**5.925**	-0.002	0.004	-0.574
	Dummy3	0.030	0.015	**1.991**	-0.015	0.012	-1.294	0.040	0.009	**4.549**	0.007	0.004	1.900
	Dummy4	0.063	0.018	**3.542**	-0.002	0.012	-0.144	0.041	0.009	**4.513**	0.023	0.005	**4.757**
	Dummy5	0.050	0.016	**3.165**	-0.008	0.014	-0.593	0.082	0.011	**7.743**	0.037	0.006	**6.512**
	Dummy6	0.051	0.017	**2.939**	-0.021	0.012	-1.713	0.069	0.010	**7.228**	0.030	0.006	**4.683**
	Dummy7	0.011	0.013	0.885	-0.002	0.012	-0.189	0.056	0.009	**6.502**	0.028	0.005	**5.315**
	Dummy8	-0.039	0.011	**-3.404**	-0.004	0.011	-0.365	-0.001	0.007	-0.088	0.005	0.004	1.202

The table reports the estimated results for trade durations. Panel A reports the results for Log-ACD part, Panel B includes the results for the exogenous variables and Panel C consists of the results for dummy variables. Bold entries indicate the coefficients are significant at 5% level.

Table 3.3 Ljung-Box Statistics For Trade Durations and Deduced Standardized Residuals

Ljung-Box Statistics on Trade Durations						Ljung-Box Statistics on Corresponding Standardized Residuals					
Lags	MEO	MRK	RWE	TKA	**C_Value**	**Lags**	MEO	MRK	RWE	TKA	**C_Value**
5	557.20	412.30	172.20	172.20	11.07	5	3.55	4.29	3.80	1.35	11.07
10	975.90	762.70	267.10	290.60	18.31	10	8.26	7.53	13.16	10.17	18.31
15	1192.90	1088.00	343.50	384.90	25.00	15	9.97	9.12	17.60	19.12	25.00
20	1430.60	1404.30	411.70	450.10	31.41	20	11.27	12.38	24.02	29.27	31.41
50	3167.20	3137.30	723.50	774.40	67.50	50	45.86	44.97	90.65	51.34	67.50
100	5164.20	5240.20	1123.40	1313.10	124.34	100	102.40	70.91	178.10	95.96	124.34
150	6445.60	6632.60	1353.90	1752.80	179.58	150	168.24	106.01	246.26	162.45	179.58
200	7350.40	7601.20	1564.20	2102.80	233.99	200	234.11	131.46	316.12	228.84	233.99

The table reports the Ljung-Box statistic on trade durations (left side) and standardized residuals of trade duration (right side) for different stocks at different lags.The column of Lags is the number of lags we use to compute the statistic and the C_Value is the critical value for the corresponding lags.

Table 3.4 Estimated Results For Quote Durations

		MEO			MRK			RWE			TKA		
		Est	Std	T-Stat	Est	Std	T-Stat	Est	Std	T-Stat	Est	Std	T-Stat
Panel A	ln(dur(lag1))	0.299	0.004	**70.792**	0.309	0.006	**52.986**	0.301	0.004	**75.553**	0.199	0.005	**39.587**
	ln(dur(lag2))	0.041	0.004	**9.511**	0.131	0.004	**29.611**	0.127	0.004	**35.714**	0.023	0.004	**5.949**
	ln(dur(lag3))	0.165	0.003	**48.008**							0.017	0.004	**4.831**
	x/phi (lag1)	0.072	0.001	**57.179**	0.042	0.001	**31.149**	0.063	0.001	**51.643**	0.047	0.001	**42.350**
	x/phi (lag2)										0.021	0.001	**22.672**
	x/phi (lag3)										0.017	0.001	**17.146**
	x/phi (lag4)										0.016	0.001	**23.487**
	cons	-0.575	0.016	**-36.725**	-0.394	0.012	**-31.852**	-1.110	0.013	**-83.628**	-1.821	0.017	**-106.145**
Panel B	x/phi (lag1)*dummy	0.070	0.004	**19.370**	0.069	0.005	**13.471**	0.075	0.004	**18.651**	0.071	0.003	**25.239**
	trade_dur/expected_trade_dur	0.154	0.002	**70.405**	0.269	0.003	**107.065**	0.267	0.002	**108.278**	0.254	0.002	**105.179**
	lag_trade_dur/expected_trade_dur	0.037	0.003	**12.023**	-0.001	0.002	-0.584	0.026	0.003	**8.439**	0.091	0.003	**34.055**
	lag_expected_trade_dur	0.108	0.005	**23.294**	-0.064	0.003	**-18.381**	0.202	0.004	**46.566**	0.226	0.004	**51.182**
	delta spread	-10.671	0.176	**-60.510**	-12.345	0.187	**-66.007**	-13.301	0.116	**-114.931**	-8.662	0.090	**-95.923**
	lev.spread	-4.573	0.098	**-46.586**	-1.560	0.109	**-14.254**	-9.825	0.116	**-84.423**	-5.026	0.106	**-47.236**
	square root(vol)	-0.072	3.72E-04	**-193.950**	-0.065	0.001	**-125.819**	-0.040	1.82E-04	**-219.646**	-0.029	1.69E-04	**-170.569**
	abs(sign.vol)	9.73E-05	1.69E-06	**57.550**	-3.51E-05	2.80E-06	**-12.525**	2.82E-05	1.10E-06	**25.673**	2.96E-05	7.10E-07	**41.690**
	Back_quote_dur	6.68E-04	1.94E-04	**3.444**	0.004	1.14E-04	**35.804**	0.010	2.83E-04	**35.118**	0.001	3.08E-04	**2.637**
	ΔXLM	-0.074	0.002	**-36.090**	-0.033	0.002	**-17.847**	-0.023	0.002	**-14.136**	0.012	0.003	**4.697**
Panel C	Dummy1	0.193	0.011	**17.960**	-0.390	0.012	**-33.898**	0.531	0.010	**55.626**	0.173	0.010	**17.074**
	Dummy2	0.139	0.007	**19.935**	-0.577	0.009	**-65.518**	-0.018	0.006	**-3.088**	-0.258	0.007	**-36.187**
	Dummy3	-0.067	0.006	**-11.290**	-0.115	0.008	**-15.127**	-0.055	0.006	**-8.858**	-0.231	0.007	**-31.943**
	Dummy4	-0.128	0.006	**-20.111**	-0.141	0.009	**-15.891**	-0.003	0.007	-0.506	-0.266	0.010	**-27.010**
	Dummy5	-0.097	0.006	**-15.359**	-0.131	0.009	**-15.115**	0.021	0.007	**2.790**	-0.020	0.009	**-2.186**
	Dummy6	-0.298	0.007	**-43.491**	-0.329	0.007	**-45.212**	-0.047	0.008	**-6.227**	-0.201	0.011	**-17.957**
	Dummy7	0.049	0.005	**9.788**	0.101	0.008	**12.343**	-0.089	0.006	**-13.938**	0.081	0.008	**9.593**
	Dummy8	-0.203	0.005	**-41.558**	0.015	0.008	1.795	-0.245	0.006	**-44.364**	-0.268	0.007	**-36.236**

The table reports the estimated results for quote durations. Panel A reports the results for Log-ACD part, Panel B includes the results for the exogenous variables and Panel C consists of the results for dummy variables. Bold entries indicate the coefficients are significant at 5% level.

Table 3.5 Ljung-Box Statistics For Quote Durations and Deduced Standardized Residuals

Ljung-Box Statistics on Quote Durations						Ljung-Box Statistics on Corresponding Standardized Residuals					
Lags	MEO	MRK	RWE	TKA	**C_Value**	**Lags**	MEO	MRK	RWE	TKA	**C_Value**
5	120.09	34.11	297.80	492.90	11.07	5	10.43	5.92	8.10	**8.12**	11.07
10	134.80	42.80	369.80	611.70	18.31	10	15.54	15.96	9.87	**49.33**	18.31
15	164.76	52.64	392.20	687.90	25.00	15	24.77	23.10	11.25	**72.68**	25.00
20	181.16	63.90	424.90	727.10	31.41	20	45.75	32.93	12.38	**116.32**	31.41
50	278.87	117.20	650.20	952.40	67.50	50	74.21	47.55	53.54	**251.10**	67.50
100	408.87	169.94	1019.40	1373.70	124.34	100	151.98	91.85	131.14	**345.85**	124.34
150	559.91	212.97	1200.20	1623.90	179.58	150	220.29	128.82	178.94	**441.45**	179.58
200	664.12	257.83	1373.40	2121.50	233.99	200	278.68	172.90	217.33	**518.50**	233.99

The table reports the Ljung-Box statistic on quote durations (left side) and standardized residuals of quote duration (right side) for different stocks at different lags.The column of Lags is the number of lags we use to compute the statistic and the C_Value is the critical value for the corresponding lags.

Table 3.6 Estimated Results For Activity Factor

		MEO			MRK			RWE			TKA		
		Est	Std	T_stat	Est	Std	T_stat	Est	Std	T_stat	Est	Std	T_stat
Panel A	g(t-1)	0.661	0.084	**7.837**	0.848	0.053	**15.933**	0.762	0.041	**18.459**	0.954	0.016	**59.194**
	activity(t-1)	0.807	0.042	**19.351**	0.892	0.049	**18.153**	0.776	0.034	**22.695**	0.399	0.029	**13.564**
	activity(t-2)	-0.359	0.102	**-3.534**	-0.657	0.087	**-7.514**	-0.410	0.062	**-6.651**	-0.345	0.033	**-10.517**
	cons	-2.023	0.107	**-18.962**	-2.031	0.143	**-14.196**	-1.479	0.102	**-14.492**	-1.384	0.101	**-13.745**
Panel B	Expect Trade Dur	0.269	0.035	**7.735**	0.278	0.033	**8.529**	0.220	0.029	**7.471**	0.143	0.025	**5.852**
	Expect Quote Dur	-0.097	0.020	**-4.934**	-0.086	0.021	**-4.117**	-0.124	0.022	**-5.611**	-0.016	0.013	-1.245
	delta spread	3.633	0.649	**5.601**	6.324	1.003	**6.305**	6.341	0.859	**7.378**	1.433	0.565	**2.535**
	lev.spread	2.375	1.038	**2.287**	2.831	1.477	1.918	6.284	1.327	**4.737**	6.126	0.854	**7.176**
	square root(vol)	0.026	0.003	**9.192**	0.035	0.004	**8.876**	0.004	0.002	**2.132**	0.008	0.001	**7.517**
	abs(sign.vol)	1.52E-05	1.82E-05	0.834	4.05E-05	3.13E-05	1.293	-7.21E-06	0.000	-0.755	-1.54E-05	4.34E-06	**-3.541**
	Back_quote_dur	0.006	7.36E-04	**8.027**	0.004	6.40E-04	**7.010**	0.006	1.19E-03	**4.959**	0.005	9.99E-04	**4.526**
	ΔXLM	0.037	0.013	**2.945**	0.062	0.012	**5.162**	0.071	0.009	**7.852**	0.076	0.015	**5.129**
Panel C	Dummy1	0.190	0.134	1.418	0.028	0.134	0.211	-0.051	0.089	-0.574	0.174	0.071	**2.464**
	Dummy2	-0.084	0.087	-0.966	-0.205	0.090	**-2.272**	-0.097	0.065	-1.500	0.038	0.052	0.729
	Dummy3	-0.096	0.084	-1.152	-0.042	0.093	-0.452	-0.075	0.066	-1.141	0.104	0.056	1.862
	Dummy4	-0.193	0.093	**-2.067**	-0.224	0.097	**-2.310**	-0.140	0.066	**-2.132**	-0.065	0.064	-1.013
	Dummy5	-0.133	0.089	-1.503	-0.192	0.107	-1.799	-0.243	0.082	**-2.981**	-0.020	0.073	-0.273
	Dummy6	-0.096	0.093	-1.036	-0.103	0.098	-1.053	-0.219	0.077	**-2.844**	0.011	0.074	0.153
	Dummy7	-0.122	0.074	-1.659	-0.315	0.095	**-3.322**	-0.136	0.067	**-2.026**	-0.083	0.064	-1.298
	Dummy8	-0.034	0.064	-0.534	-0.112	0.088	-1.270	-0.033	0.055	-0.605	0.084	0.051	1.650

The table reports the estimated results for Activity factor. Panel A reports the results for GLARMA part, Panel B includes the results for the exogenous variables and Panel C consists of the results for dummy variables.Bold entries indicate the coefficients are significant at 5% level.

Table 3.7 Ljung-Box Statistics For Activity and Deduced Standardized Residuals

Ljung-Box Statistics on Activity						Ljung-Box Statistics on Corresponding Standardized Residuals					
Lags	MEO	MRK	RWE	TKA	**C_Value**	**Lags**	MEO	MRK	RWE	TKA	**C_Value**
5	887.04	754.79	1127.06	358.71	11.07	5	3.08	4.51	3.68	8.32	11.07
10	979.11	850.97	1272.35	441.53	18.31	10	5.36	6.64	6.72	12.80	18.31
15	1007.67	923.88	1322.27	512.89	25.00	15	8.12	16.48	11.28	15.51	25.00
20	1022.06	957.06	1329.06	540.47	31.41	20	11.45	23.48	16.43	21.91	31.41
50	1233.96	1056.15	1417.73	662.32	67.50	50	54.39	61.57	50.13	54.50	67.50
100	1453.23	1163.59	1502.42	813.11	124.34	100	99.63	116.12	108.88	111.38	124.34
150	1555.10	1274.25	1538.54	917.70	179.58	150	147.95	177.37	145.24	151.97	179.58
200	1665.43	1332.34	1586.24	994.88	233.99	200	203.86	234.75	190.94	189.69	233.99

The table reports the Ljung-Box statistic on activity factor (left side) and standardized residuals of activity factor (right side) for different stocks at different lags.The column of Lags is the number of lags we use to compute the statistic and the C_Value is the critical value for the corresponding lags.

Table 3.8 Estimated Results For Direction

		MEO			MRK			RWE			TKA		
		Est	Std	T-stat	Est	Std	T-stat	Est	Std	T-stat	Est	Std	T-stat
Panel A	g(t-1)	1.06	0.07	**14.77**	0.90	0.10	**9.28**	0.94	0.07	**13.97**	1.04	0.04	**23.25**
	g(t-2)	-0.30	0.06	**-4.72**	-0.14	0.08	-1.65	-0.30	0.09	**-3.45**	-0.35	0.04	**-8.82**
	g(t-3)							0.08	0.04	1.73			
	direction(t-1)	-0.57	0.05	**-10.77**	-0.55	0.06	**-9.11**	-0.67	0.04	**-14.95**	-0.85	0.06	**-14.95**
	cons	0.04	0.11	0.38	-0.13	0.13	-0.98	-0.06	0.09	-0.71	0.23	0.09	**2.54**
Panel B	Expect QuoteQuote Dur	-0.01	0.01	-1.25	0.02	0.01	1.51	0.01	0.01	1.32	0.02	0.01	**2.60**
	delta spread	4.84	1.03	**4.70**	3.05	1.12	**2.73**	2.40	0.95	**2.54**	6.83	0.75	**9.06**
	lev.spread	-5.37	1.37	**-3.91**	-1.14	1.91	-0.60	-3.00	1.64	-1.83	-9.54	0.99	**-9.62**
	square root(vol)	1.15E-02	4.07E-03	**2.82**	7.88E-04	0.01	0.15	4.18E-03	2.41E-03	1.73	9.72E-03	1.90E-03	**5.12**
	abs(sign.vol)	-3.51E-05	2.52E-05	-1.39	6.55E-05	4.01E-05	1.64	-4.53E-06	1.34E-05	-0.34	-1.82E-05	7.48E-06	**-2.44**
	Back_quote_dur	-2.97E-03	8.40E-04	**-3.54**	-1.15E-03	6.75E-04	-1.70	-0.01	1.25E-03	**-4.61**	-5.77E-03	1.48E-03	**-3.89**
	ΔXLM	-0.33	0.02	**-14.10**	-0.22	0.02	**-9.52**	-0.29	0.02	**-14.65**	-0.74	0.03	**-22.72**
Panel C	Dummy1	0.11	0.17	0.62	-0.12	0.17	-0.71	0.23	0.12	1.88	0.21	0.11	1.94
	Dummy2	0.01	0.13	0.05	0.20	0.12	1.64	0.12	0.09	1.33	-0.03	0.09	-0.30
	Dummy3	0.07	0.12	0.60	0.07	0.12	0.56	0.22	0.09	**2.47**	0.11	0.10	1.13
	Dummy4	0.16	0.13	1.18	0.03	0.13	0.20	0.14	0.09	1.54	-0.06	0.11	-0.57
	Dummy5	0.05	0.13	0.36	0.10	0.15	0.69	0.10	0.11	0.94	0.03	0.12	0.26
	Dummy6	0.02	0.13	0.17	-0.08	0.13	-0.63	0.20	0.10	1.91	-0.10	0.12	-0.79
	Dummy7	0.10	0.11	0.92	0.09	0.13	0.71	0.15	0.09	1.65	-0.17	0.11	-1.45
	Dummy8	0.09	0.10	0.90	-0.04	0.12	-0.34	0.09	0.08	1.17	-0.04	0.09	-0.42

The table reports the estimated results for Direction factor. Panel A reports the results for GLARMA part, Panel B includes the results for the exogenous variables and Panel C consists of the results for dummy variables. Bold entries indicate the coefficients are significant at 5% level.

Table 3.9 Ljung-Box Statistics For Direction and Deduced Standardized Residuals

Ljung-Box Statistics on Direction						Ljung-Box Statistics on Corresponding Standardized Residuals					
Lags	MEO	MRK	RWE	TKA	**C_Value**	**Lags**	MEO	MRK	RWE	TKA	**C_Value**
5	869.93	721.00	1963.59	3051.37	11.07	5	6.14	2.79	11.05	0.00	11.07
10	871.34	722.20	1977.68	3059.01	18.31	10	11.95	4.73	20.31	0.03	18.31
15	888.31	729.80	1987.41	3060.47	25.00	15	20.70	12.29	22.22	0.03	25.00
20	897.14	733.60	1996.95	3066.62	31.41	20	24.56	17.98	29.49	0.03	31.41
50	950.96	781.00	2020.14	3114.22	67.50	50	50.02	51.02	53.49	0.03	67.50
100	1004.69	871.70	2073.82	3238.26	124.34	100	84.32	109.29	90.04	0.04	124.34
150	1051.74	967.00	2148.45	3272.82	179.58	150	138.29	169.60	139.51	0.05	179.58
200	1114.54	1045.60	2194.39	3392.66	233.99	200	205.41	223.39	193.70	0.06	233.99

The table reports the Ljung-Box statistic on direction factor (left side) and standardized residuals of direction factor (right side) for different stocks at different lags.The column of Lags is the number of lags we use to compute the statistic and the C_Value is the critical value for the corresponding lags.

Table 3.10 Estimated Results For Size

		MEO			MRK			RWE			TKA		
		Est	Std	T-stat	Est	Std	T-stat	Est	Std	T-stat	Est	Std	T-stat
Panel A	g(t-1)	0.700	0.161	**4.3591**	0.793	0.083	**9.598**	0.595	0.150657	**3.9487**	-0.008	0.361	-0.023
	g(t-2)	-0.162	0.022	**-7.447**	-0.133	0.017	**-7.9**	0.138	0.0855645	1.6149	0.182	0.073	**2.4763**
	Size (t-1)	0.075	0.042	1.7716	0.070	0.026	**2.6568**	-0.220	0.0130852	**-16.83**	-0.493	0.035	**-13.92**
	Size (t-2)							0.105	0.0311978	**3.3676**	-0.002	0.185	-0.009
	cons	0.390	0.097	**4.0358**	0.278	0.114	**2.4404**	0.420	0.077398	**5.4252**	1.901	0.082	**23.289**
Panel B	Expect QuoteQuote Dur	2.72E-04	0.000	0.559	-8.05E-04	0.000	-1.871	-6.14E-04	0.001	-0.965	-1.53E-03	0.001	-1.501
	Direction(t)	-0.069	0.028	**-2.505**	-0.061	0.026	**-2.312**	-0.103	0.020	**-5.269**	-0.145	0.036	**-4.021**
	Direction(t-1)	-0.006	0.048	-0.121	0.056	0.044	1.2584	0.085	0.031	**2.7106**	0.147	0.057	**2.5808**
	delta spread	1.030	0.534	1.9306	1.158	0.869	1.3324	1.974	0.715	**2.7618**	0.254	0.375	0.6775
	lev.spread	0.959	1.047	0.9152	0.225	1.364	0.1648	0.388	1.243	0.3124	-1.758	0.746	**-2.358**
	square root(vol)	-2.23E-03	0.003	-0.669	-2.75E-03	0.004	-0.711	2.38E-03	0.002	1.3135	-5.31E-03	0.002	**-3.162**
	abs(sign.vol)	-1.78E-06	0.000	-0.082	-5.57E-06	0.000	-0.19	-7.41E-06	0.000	-0.744	-1.05E-05	0.000	-1.409
	Back_quote_dur	0.001	0.001	1.5792	0.001	0.001	1.6047	0.005	0.001	**5.347**	0.008	0.002	**4.3816**
	ΔXLM	-0.006	0.022	-0.276	-0.033	0.017	-1.951	-0.068	0.013	**-5.285**	-0.145	0.034	**-4.304**
Panel C	Dummy1	-0.134	0.131	-1.018	-0.167	0.126	-1.327	-0.161	0.088	-1.831	-0.543	0.103	**-5.256**
	Dummy2	-0.072	0.099	-0.724	-0.096	0.087	-1.096	-0.046	0.064	-0.712	-0.262	0.096	**-2.723**
	Dummy3	-0.121	0.101	-1.205	-0.037	0.088	-0.426	-0.048	0.063	-0.759	-0.194	0.100	-1.938
	Dummy4	-0.165	0.104	-1.591	-0.033	0.101	-0.322	-0.091	0.064	-1.414	-0.289	0.115	**-2.508**
	Dummy5	0.059	0.103	0.5749	-0.097	0.111	-0.874	-0.017	0.078	-0.214	-0.309	0.123	**-2.51**
	Dummy6	-0.096	0.101	-0.954	-0.095	0.096	-0.992	0.027	0.075	0.3611	-0.198	0.131	-1.505
	Dummy7	-0.029	0.088	-0.331	-0.017	0.096	-0.175	-0.040	0.069	-0.579	-0.125	0.121	-1.032
	Dummy8	-0.058	0.080	-0.722	-0.038	0.087	-0.432	-0.013	0.055	-0.235	-0.137	0.097	-1.411

The table reports the estimated results for Size factor. Panel A reports the results for GLARMA part, Panel B includes the results for the exogenous variables and Panel C consists of the results for dummy variables.Bold entries indicate the coefficients are significant at 5% level.

Table 3.11 Ljung-Box Statistics For Size and Deduced Standardized Residuals

Ljung-Box Statistics on Direction						Ljung-Box Statistics on Corresponding Standardized Residuals					
Lags	MEO	MRK	RWE	TKA	**C_Value**	**Lags**	MEO	MRK	RWE	TKA	**C_Value**
5	247.48	297.58	1406.50	889.70	11.07	5	2.73	4.59	18.64	21.23	11.07
10	252.01	330.53	1592.70	1054.60	18.31	10	2.96	11.11	23.22	27.89	18.31
15	263.01	343.90	1624.90	1107.20	25.00	15	12.67	13.86	24.03	36.50	25.00
20	264.68	350.70	1655.50	1125.00	31.41	20	14.05	19.96	32.73	39.11	31.41
50	301.70	383.77	1701.20	1218.50	67.50	50	49.69	45.29	73.52	68.95	67.50
100	335.13	429.84	1774.40	1266.30	124.34	100	80.15	88.45	116.21	114.54	124.34
150	389.70	497.10	1814.40	1301.90	179.58	150	141.83	132.88	163.70	155.42	179.58
200	458.14	550.77	1909.40	1358.20	233.99	200	205.49	177.29	219.94	199.67	233.99

The table reports the Ljung-Box statistic on size factor (left side) and standardized residuals of size factor (right side) for different stocks at different lags.The column of Lags is the number of lags we use to compute the statistic and the C_Value is the critical value for the corresponding lags.

Table 3.12 Effect of Exogenous Variables on XLM

	Activity	Direction	Size	Total Effect on XLM
Quote_Quote Dur	#	+/#	#	little
Direction(t-1)	+	-	-	negative
delta spread	+	+	+/#	positive / little
lev.spread	+	-/#	-/#	negative / little
square root(vol)	+	+/#	-/#	negative /little
abs(sign.vol)	#	-/#	#	little
Back_quote_dur	+	-	+/#	positive /little
XLM	+	-	-/#	negative / little

The table summarizes the effect of various exogenous variables on the XLM changes, # relates to a non-significant estimated coefficient.

Appendix

A1.1 Xetra is the main trading platform in Germany and realizes more than 90% of total transactions at German exchanges. Trading and order processing (entry, revision, execution and cancellation) in the Xetra system is highly computerized and maintained by the German Stock Exchange. Since September 20, 1999, trading hours have been from 9:00 a.m. to 5:30 p.m. CET (Central European Time). However, during the pre- and post-trading hours, operations such as entry, revision and cancellation are permitted.

To ensure trading efficiency, Xetra operates with different market models that define order matching, price determination, transparency, etc. Most of the market models impose the Price-Time-Priority condition where the electronic trading system places the incoming order after checking the price and timestamps of all available limit orders in LOB. One of the important parameters of a market model is the trading model that determines whether the trading is organized continuously, discretely or both ways. During normal trading hours, there are two types of trading mechanisms: call auction and continuous auction. A call auction, in which the clearance price is determined by the state of LOB and remains as the open price for the following continuous auction, can occur one or several times during the trading day. At each call auction, market participants can submit both round-lot and odd-lot orders, and start and end times for a call auction are randomly chosen by a computer to avoid scheduled trading. For stocks in the Deutscher Aktien-Index 30 (German stock index, DAX30[1]), there are three auctions during a trading day: the open, mid-day, and closing auctions. The mid-day auction starts at 1:00 p.m. and lasts around two minutes. Between the call auctions, the market is organized as a continuous auction where traders can submit round-lot-sized limit orders or market orders only.

[1] The DAX30 includes 30 major German companies trading on the Frankfurt Stock Exchange. The DAX30 includes 30 major German companies trading on the Frankfurt Stock Exchange.

Table A 1.1: Descriptive Statistics for SAP Raw and Deseasonalized Data

	Mean	Std.Dev	Skew	Kurt	Min	Max	LB(15)	LB2(15)
Panel A: SAP Raw data								
Duration	5.81	12.89	5.03	43.03	1.00E-03	239.8	12748.87	3409.66
FR Change	4.83E-09	1.99E-04	0.2	20.78	-3.93E-03	4.76E-03	7228.65	12678.9
AR Change (Q=6000)	-3.33E-09	1.81E-04	0	38.23	-5.15E-03	4.99E-03	6482.09	11491.39
AR Change (Q=4000)	-1.10E-09	1.82E-04	0.11	35.28	-4.76E-03	5.09E-03	6564.74	11995.11
AR Change (Q=2000)	1.22E-09	1.85E-04	0.24	30.43	-4.05E-03	5.18E-03	6656.12	12334.21
Panel B: SAP Deseasonalized Data								
Duration	0.97	1.9	3.94	26.79	0	29.83	6466.15	2249.41
FR Change	-4.00E-05	1.01	0.11	21.13	-22.91	22.18	6321.16	8731.99
AR Change (Q=6000)	-4.00E-05	1.01	0.11	21.13	-22.91	22.18	6321.16	8731.99
AR Change (Q=4000)	-5.24E-05	1	0.15	19.86	-21.1	22.59	6405.89	9233.8
AR Change (Q=2000)	-2.92E-05	1	0.23	17.88	-17.88	22.89	6483.35	9895.25

The table shows the descriptive statistics for raw durations, actual return changes when Q=2,000, 4,000, 6,000 and frictionless return changes. The sample period is the first 2 weeks of June 2011 with 49,250 observations.

Table A 1.2 : Descriptive Statistics for RWE Raw and Deseasonalized Data

	Mean	Std.Dev	Skew	Kurt	Min	Max	LB(15)	LB2(15)
Panel A : RWE Raw data								
Duration	6.1	12.38	5.08	48.98	1.00E-03	308.7	12478.3	2998.71
FR Change	8.39E-09	2.18E-04	0.15	9.19	-3.29E-03	3.04E-03	6453.84	6426.85
AR Change (Q=4000)	1.88E-08	1.99E-04	-0.02	15.33	-4.32E-03	3.04E-03	5232.02	4274.08
AR Change (Q=2000)	1.58E-08	2.00E-04	0.06	12.98	-4.07E-03	3.04E-03	5398.41	4279.5
AR Change (Q=1000)	1.05E-08	2.03E-04	0.13	10.9	-3.67E-03	3.04E-03	5596.1	4495.29
Panel B : RWE Deseasonalized Data								
Duration	0.97	1.76	4.17	37.09	4.59E-05	44.4	5721.78	2844.72
FR Change	-5.97E-05	0.98	0.02	14.47	-20.23	15.83	5155.43	4565.87
AR Change (Q=4000)	-5.97E-05	0.98	0.02	14.47	-20.23	15.83	5155.43	4565.87
AR Change (Q=2000)	-7.25E-05	0.98	0.09	12.24	-18.81	15.88	5324.74	4390.58
AR Change (Q=1000)	-6.20E-05	0.98	0.16	10.37	-16.72	15.58	5529.7	4344.41

The table shows the descriptive statistics for raw durations, actual return changes when Q=1,000, 2,000, 4,000 and frictionless return changes. The sample period is the first 2 weeks of June 2011 with 46,893 observations.

Table A 1.3: Descriptive Statistics for MRK Raw and Deseasonalized Data

	Mean	Std.Dev	Skew	Kurt	Min	Max	LB(15)	LB2(15)
Panel A : MRK Raw data								
Duration	6.1	12.38	5.08	48.98	1.00E-03	308.7	12478.3	2998.71
FR Change	8.39E-09	2.18E-04	0.15	9.19	-3.29E-03	3.04E-03	6453.84	6426.85
AR Change(Q=2700)	1.88E-08	1.99E-04	-0.02	15.33	-4.32E-03	3.04E-03	5232.02	4274.08
AR Change (Q=1800)	1.58E-08	2.00E-04	0.06	12.98	-4.07E-03	3.04E-03	5398.41	4279.5
AR Change (Q=900)	1.05E-08	2.03E-04	0.13	10.9	-3.67E-03	3.04E-03	5596.1	4495.29
Panel B : MRK Deseasonalized Data								
Duration	0.97	1.76	4.17	37.09	4.59E-05	44.4	5721.78	2844.72
FR Change	-5.97E-05	0.98	0.02	14.47	-20.23	15.83	5155.43	4565.87
AR Change (Q=2700)	-5.97E-05	0.98	0.02	14.47	-20.23	15.83	5155.43	4565.87
AR Change (Q=1800)	-7.25E-05	0.98	0.09	12.24	-18.81	15.88	5324.74	4390.58
AR Change (Q=900)	-6.20E-05	0.98	0.16	10.37	-16.72	15.58	5529.7	4344.41

The table shows the descriptive statistics for raw durations, actual return changes when Q= 900, 1,800, 2,700 and frictionless return changes. The sample period is the first 2 weeks of June 2011 with 18,631 observations.

Table A 1.4: Estimation Results SAP (v = 4000, First Week of June 2011)

Estimation Actual return and Frictionless return changes Q=4000						
Estimation VARMA(3.1)- NGARCH((1.3).1) (Obs =20799)						
	Parameter	Estimation	StdError	Ljung_box test on Residual		
ACD(1.1)	alpha1_dur	0.145	0.009			
	alpha2_dur	**-0.013**	0.012			
	alpha3_dur	-0.039	0.009	Lags	Statistic	C_Value
	beta1_dur	0.968	0.003	5	6.288	11.070
	gam1_dur	0.802	0.021	10	23.202	18.307
	gam2_dur	0.384	0.015	15	31.564	24.996
	Constant	-0.097	0.005	20	35.334	31.410
Actual Return Changes VARMA(3.2)-NGARCH((1.3).1)	phi11_1	0.6097	0.0107			
	phi12_1	-0.2420	0.0095			
	phi11_2	0.1427	0.0132	Lags	Statistic	C_Value
	phi12_2	-0.0433	0.0123	5	20.828	11.070
	phi11_3	0.0359	0.0126	10	25.616	18.307
	phi12_3	**0.0127**	0.0120	15	39.492	24.996
	phi11_4	0.0431	0.0108	20	46.121	31.410
	phi12_4	**-0.0033**	0.0107			
	theta11_1	-0.9844	0.0011			
	Constant	0.0005	0.0001	Ljung_box test on Residual squared		
	Beta0	0.1119	0.0041			
	Beta1_1	0.3868	0.0187	Lags	Statistic	C_Value
	Beta2_1	0.2067	0.0069	5	17.214	11.070
	theta	0.4511	0.0202	10	32.292	18.307
	Beta1_3	0.2456	0.0134	15	37.733	24.996
	gam	0.0989	0.0020	20	39.815	31.410
Frictionless Return Changes ARMA(1,3)-GARCH((1,3),1)	phi22_1	0.2352	0.0107			
	phi21_1	0.1153	0.0096			
	phi22_2	0.0534	0.0126	Lags	Statistic	C_Value
	phi21_2	0.0411	0.0121	5	25.026	11.070
	phi22_3	0.0708	0.0116	10	34.263	18.307
	phi21_3	**-0.0176**	0.0106	15	44.319	24.996
	theta22_1	-0.9763	0.0014	20	55.459	31.410
	Constant	0.0007	0.0001	Ljung_box test on Residual squared		
	Beta0	0.0879	0.0040			
	Beta1_1	0.4048	0.0245	Lags	Statistic	C_Value
	Beta2_1	0.1688	0.0070	5	25.026	11.070
	theta	0.4326	0.0240	10	34.263	18.307
	Beta1_3	0.2888	0.0182	15	44.319	24.996
	gam	0.0889	0.0023	20	55.459	31.410
DCC parameter	$\xi 1$	0.0890	0.0017			
	$\pi 1$	0.7489	0.0076			

The Statistics column reports the Ljung-Box statistic on standardized residuals of duration, actual return changes and squared standardized residuals for different lags. The bold entries are the estimation coefficients that are not significantly different from zero.

Table A 1.5: Estimation Results RWE (v = 4000, First Week of June 2011)

Estimation Actual return and Frictionless return changes Q=4000						
Estimation VARMA(5.1)- NGARCH((1.3).1) (Obs =22427)						
	Parameter	Estimation	StdError	Ljung_box test on Residual		
ACD(3.1)	alpha1_dur	0.107	0.008			
	alpha2_dur	**-0.004**	0.011	Lags	Statistic	C_Value
	alpha3_dur	-0.026	0.008	5	3.819	11.070
	beta1_dur	0.962	0.003	10	8.974	18.307
	gam1_dur	1.093	0.022	15	12.131	24.996
	gam2_dur	0.297	0.009	20	16.275	31.410
	Constant	-0.082	0.004			
Actual Return Changes VARMA(5.1)-NGARCH((1.3).1)	phi11_1	0.7212	0.0101	Lags	Statistic	C_Value
	phi12_1	-0.2930	0.0089	5	12.720	11.070
	phi11_2	0.1109	0.0120	10	24.254	18.307
	phi12_2	**-0.0104**	0.0105	15	37.274	24.996
	phi11_3	**0.0020**	0.0122	20	49.151	31.410
	phi12_3	0.0335	0.0114			
	phi11_4	**0.0062**	0.0121			
	phi12_4	0.0219	0.0115			
	phi11_5	**0.0153**	0.0103			
	phi12_5	0.0218	0.0105	Ljung_box test on Residual squared		
	theta11_1	-0.9877	0.0009			
	Constant	0.0002	0.0001	Lags	Statistic	C_Value
	Beta0	0.1537	0.0049	5	7.271	11.070
	Beta1_1	0.3617	0.0168	10	10.894	18.307
	Beta2_1	0.1955	0.0067	15	11.615	24.996
	theta	0.5689	0.0221	20	18.830	31.410
	Beta1_3	0.2031	0.0123			
	gam	0.0778	0.0017			
Frictionless Return Changes VARMA(5.1)-NGARCH((1.3).1)	phi11_1	0.2325	0.0100	Lags	Statistic	C_Value
	phi12_1	0.1643	0.0091	5	14.827	11.070
	phi11_2	0.0888	0.0116	10	27.592	18.307
	phi12_2	**-0.0029**	0.0111	15	43.001	24.996
	phi11_3	0.0820	0.0117	20	48.085	31.410
	phi12_3	-0.0349	0.0114			
	phi11_4	0.0288	0.0116			
	phi12_4	**0.0005**	0.0109			
	phi11_5	0.0578	0.0108	Ljung_box test on Residual squared		
	phi12_5	-0.0288	0.0097			
	theta11_1	-0.9846	0.0011	Lags	Statistic	C_Value
	Constant	0.0004	0.0001	5	22.193	11.070
	Beta0	0.1098	0.0045	10	24.741	18.307
	Beta1_1	0.3882	0.0206	15	28.203	24.996
	Beta2_1	0.2223	0.0078	20	33.346	31.410
	theta	0.4218	0.0208			
	Beta1_3	0.2123	0.0146			
	gam	0.0637	0.0022			
DCC parameter	ξ1	0.1295	0.0044			
	π1	0.5276	0.0156			

The Statistics column reports the Ljung-Box statistic on standardized residuals of duration, actual return changes and squared standardized residuals for different lags. The bold entries are the estimation coefficients that are not significantly different from zero.

Table A 1.6: Estimation Results MRK (v = 1800, First Week of June 2011)

Estimation Actual return changes (Q=1800)						
Estimation VARMA(2.2)- NGARCH((1.3).1) (Obs =9428)						
	Parameter	Estimation	StdError	Ljung_box test on Residual		
ACD(1.1)	alpha1_dur	0.091	0.010	Lags	Statistic	C_Value
	beta1_dur	0.900	0.017	5	11.319	11.070
	gam1_dur	0.591	0.028	10	15.434	18.307
	gam2_dur	0.525	0.039	15	20.436	24.996
	Constant	-0.092	0.009	20	27.879	31.410
Actual Return Changes VARMA(2.2)-NGARCH((1.3).1)	phi11_1	1.2982	0.0547	Lags	Statistic	C_Value
	phi12_1	-0.2352	0.0165	5	8.931	11.070
	phi11_2	-0.3454	0.0456	10	17.498	18.307
	phi12_2	0.1307	0.0227	15	28.640	24.996
	theta11_1	-1.5935	0.0482	20	29.687	31.410
	theta11_2	0.6063	0.0459			
	Constant	0.0002	0.0001	Ljung_box test on Residual squared		
	Beta0	0.0914	0.0035			
	Beta1_1	0.3646	0.0298	Lags	Statistic	C_Value
	Beta2_1	0.1550	0.0068	5	7.903	11.070
	theta	0.5161	0.0363	10	13.586	18.307
	Beta1_3	0.3443	0.0246	15	17.091	24.996
	gam	0.0579	0.0028	20	45.001	31.410
Frictionless Return Changes ARMA(1,3)-GARCH((1,3),1)	phi22_1	0.5550	0.1480	Lags	Statistic	C_Value
	phi21_1	0.1853	0.0137	5	7.600	11.070
	phi22_2	**-0.0200**	0.0395	10	12.122	18.307
	phi21_2	-0.0635	0.0316	15	15.541	24.996
	theta22_1	-1.3188	0.1460	20	19.757	31.410
	theta22_2	0.3399	0.1406	Ljung_box test on Residual squared		
	Constant	0.0008	0.0003			
	Beta0	0.0561	0.0039	Lags	Statistic	C_Value
	Beta1_1	0.3940	0.0423	5	11.012	11.070
	Beta2_1	0.1689	0.0099	10	23.809	18.307
	Beta1_3	0.3698	0.0358	15	25.171	24.996
	gam	0.0320	0.0032	20	34.358	31.410
DCC parameter	ξ1	0.0598	0.0015			
	π1	0.8526	0.0037			

The Statistics column reports the Ljung-Box statistic on standardized residuals of duration, actual return changes and squared standardized residuals for different lags. The bold entries are the estimation coefficients that are not significantly different from zero.

Table A 1.7: Backtesting on Simulated Deseasonalized Data

Panel A: SAP Out-of-sample Backtesting on Deseasonalized Actual Return Change (v = 4000)

Interval (in units)	Nb of Interval	Time interval in Minutes	Kupiec test				Christoffersen Test				UC (Ziggel et al (2013))				I.I.D (Ziggel et al (2013))			
			5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%
40	689	3.70	0.659	0.282	**0.005**	0.119	**0.040**	0.479	0.957	0.957	0.648	0.367	0.016	0.200	0.360	0.491	0.082	0.088
50	551	4.63	0.914	0.621	0.497	0.222	0.095	0.464	0.809	0.952	0.897	0.582	0.444	0.219	0.417	0.344	0.194	0.666
60	459	5.56	0.519	0.150	0.777	0.334	0.176	0.641	0.791	0.947	0.514	0.188	0.924	0.345	0.774	0.417	0.994	0.549
80	344	7.41	0.414	0.343	0.807	0.550	0.275	0.644	0.818	0.939	0.413	0.450	0.943	0.498	0.874	0.314	0.984	0.657
100	275	9.27	0.276	0.229	0.633	0.617	0.384	0.731	0.864	0.864	0.286	0.254	0.754	0.660	0.641	0.366	0.226	0.222
120	229	11.14	0.891	0.441	**0.032**	0.130	0.291	0.705	#	#	0.877	0.476	0.034	0.192	0.160	0.655	0.640	0.649
140	196	13.01	0.485	0.964	0.977	0.984	0.209	0.608	0.839	0.919	0.467	0.916	0.923	0.780	0.410	0.613	0.868	0.540

Panel B: RWE Out-of-sample Backtesting on Deseasonalized Actual Return Change (v = 4000)

Interval (in units)	Nb of Interval	Time interval in Minutes	Kupiec test				Christoffersen Test				UC (Ziggel et al (2013))				I.I.D (Ziggel et al (2013))			
			5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%
40	593	4.30	0.086	0.099	0.060	0.184	0.214	0.598	0.907	0.954	0.069	0.104	0.129	0.249	0.946	0.355	0.490	0.629
50	475	5.37	**0.006**	0.057	0.909	0.336	0.430	0.695	0.744	0.794	**0.004**	**0.048**	0.982	0.206	0.158	0.964	0.716	0.263
60	395	6.46	0.054	0.532	0.276	**0.047**	0.385	0.565	0.886	#	0.070	0.450	0.420	0.234	0.714	0.977	0.691	0.757
80	296	8.61	0.289	0.166	**0.015**	0.085	0.356	0.740	#	#	0.364	0.178	**0.017**	**0.046**	0.752	0.812	0.579	0.079
100	237	10.76	0.055	0.059	0.312	0.861	0.576	0.853	0.926	0.926	0.075	0.118	0.228	0.914	0.535	0.518	0.112	0.108
120	197	12.94	0.327	0.973	0.443	0.160	0.471	0.609	0.919	#	0.312	0.914	0.706	**0.016**	0.296	0.984	0.376	0.596
140	169	15.09	0.189	0.057	0.065	0.193	0.621	0.913	#	#	0.243	0.111	**0.016**	0.066	0.420	0.671	0.597	0.053

Panel C: MRK Out-of-sample Backtesting on Deseasonalized Actual Return Change (v = 1800)

Nb of Interval	Interval (in units)	Time interval in Minutes	Kupiec test				Christoffersen Test				UC(Ziggel et al (2013))				I.I.D (Ziggel et al (2013))			
			5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%
40	220	11.59	**0.038**	0.238	0.362	0.923	0.629	0.773	0.924	0.924	0.065	0.270	0.616	0.698	0.190	0.058	0.054	0.062
50	176	14.49	**0.005**	0.195	0.859	0.305	0.830	0.830	0.830	0.830	**0.008**	0.309	0.858	0.285	0.723	0.704	0.715	0.706
60	147	17.35	0.053	**0.006**	0.086	0.225	0.907	#	#	#	**0.006**	**0.002**	0.172	0.072	0.284	0.097	0.934	0.312
80	110	23.18	**0.016**	**0.018**	0.137	0.294	0.892	#	#	#	0.052	**0.026**	0.372	0.219	0.624	0.469	0.060	0.602
100	88	28.98	0.056	0.359	0.900	0.468	0.879	0.879	0.879	0.879	**0.030**	0.293	0.828	0.587	0.511	0.484	0.485	0.505
140	63	40.48	0.477	0.619	0.260	0.427	0.715	0.856	#	#	0.575	0.899	0.620	0.297	0.407	0.513	0.756	0.272

The table contains the p-values for Kupiec, Christoffersen, Unconditional coverage, and I.I.D (Ziggel et al (2013)) tests for the stocks SAP, RWE, and MRK. Interval is the interval length used for computing the LIVaR. Nb of intervals is the number of intervals for out-of sample analysis and Time interval in minutes is the corresponding calendar time. Bold entries indicate the rejections of the model at 95% confidence level. When the number of hits is less than two, the p-values are denoted by #.

Table A 1.8: Backtesting on Simulated Re-seasonalized Data

Panel A: SAP Out-of-sample Backtesting on Raw Actual Return Change (v = 4000)

Interval (in Mins)	Nb Interval	Kupiec test				Christoffersen Test				UC (Ziggel et al (2013))				I.I.D (Ziggel et al (2013))			
		5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%
5	485	**0.039**	0.341	0.613	0.147	0.327	0.559	0.698	0.747	0.138	0.144	0.612	0.672	0.392	0.053	0.224	0.903
6	405	0.210	0.293	0.583	0.986	0.299	0.646	0.832	0.888	0.054	**0.025**	0.761	0.121	0.930	0.788	0.347	0.058
7	345	0.099	0.175	0.803	0.548	0.417	0.731	0.818	0.939	0.654	0.404	0.487	0.588	0.323	0.252	**0.009**	0.162
8	300	**0.043**	0.326	0.537	0.663	0.507	0.680	0.870	0.935	**0.004**	0.070	**0.025**	0.408	0.670	0.368	0.422	0.721
9	270	0.098	0.247	0.857	0.601	0.484	0.728	0.795	0.863	**0.020**	0.064	0.051	0.377	0.928	0.650	0.814	0.267
10	245	0.185	0.959	0.765	0.833	0.461	0.582	0.856	0.928	0.340	**0.002**	0.055	0.078	0.579	0.201	0.436	0.449

Panel B: RWE Out-of-sample Backtesting on Raw Actual Return Change (v = 4000)

Interval (in Mins)	Nb Interval	Kupiec test				Christoffersen Test				UC (Ziggel et al (2013))				I.I.D (Ziggel et al (2013))			
		5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%
5	485	0.173	**0.049**	0.140	0.298	0.238	0.698	0.897	0.949	0.207	**0.039**	0.154	0.326	**0.006**	0.874	0.906	0.912
6	405	0.132	0.071	0.068	0.423	0.316	0.723	0.944	0.944	0.154	0.081	0.128	0.328	0.109	0.064	0.885	0.894
7	345	0.099	**0.006**	**0.008**	0.063	0.394	0.878	#	#	0.090	**0.012**	**0.045**	0.171	0.909	0.308	**0.050**	0.717
8	300	0.160	0.326	0.178	0.663	0.405	0.680	0.935	0.935	0.200	0.389	0.213	0.934	0.940	0.525	0.422	0.443
9	270	0.307	**0.030**	0.233	0.100	0.379	0.863	0.931	#	0.278	0.055	0.194	0.370	0.479	0.616	0.936	0.454
10	245	0.318	0.354	0.765	0.520	0.406	0.715	0.856	0.856	0.286	0.449	0.993	0.408	0.402	0.625	0.196	0.200

Panel C: MRK Out-of-sample Backtesting on Raw Actual Return (v = 1800)

Interval (in Mins)	Nb Interval	Kupiec test				Christoffersen Test				UC (Ziggel et al (2013))				I.I.D (Ziggel et al (2013))			
		5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%	5%	2.50%	1%	0.50%
5	485	**0.039**	0.106	0.613	0.354	0.479	0.650	0.698	0.796	**0.019**	0.076	0.671	0.869	0.723	0.061	**0.027**	0.660
6	405	0.078	0.715	0.363	0.220	0.352	0.522	0.671	0.777	0.051	0.410	0.878	0.921	0.168	0.226	**0.011**	0.665
7	345	0.171	0.898	0.394	0.838	0.373	0.512	0.878	0.878	0.082	0.542	0.194	0.627	0.938	0.804	0.270	0.275
8	300	0.160	0.156	0.178	0.663	0.405	0.742	0.935	0.935	0.105	0.1078	0.116	0.2796	0.885	0.596	0.707	0.7198
9	270	0.471	0.766	0.458	0.601	0.356	0.633	0.728	0.863	0.228	0.4716	0.778	0.6942	0.523	0.3178	0.164	0.4926
10	245	0.318	0.463	0.151	0.175	0.434	0.491	0.647	0.785	0.179	0.7636	0.91	0.9398	0.577	0.3998	0.227	0.5956

The table contains the p-values for Kupiec and Christoffersen Unconditional coverage, and I.I.D (Ziggel et al (2013)) tests. Intervals are regularly time-spaced from 5 minutes to 10 minutes. Bold entries indicate the rejections of the model at 95% confidence level. When the number of hits is less than two, the p-values are denoted by #.

Table A 1.9: The Relative Liquidity Risk Premium

Panel A: SAP					
	Relative Ex-ante Liquidity Risk Premium at the end of 5-min Interval				
Confidence Level	Volume (Shares)/PDTV				
	550	1500	2000	4000	6000
10%	6.33%	12.56%	15.56%	23.81%	29.88%
5%	5.76%	11.19%	13.96%	21.60%	27.29%
1%	6.15%	10.51%	12.84%	20.17%	25.38%

Panel A: RWE					
	Relative Ex-ante Liquidity Risk Premium at the end of 5-min Interval				
Confidence Level	Volume (Shares)/PDTV				
	350	1000	1500	2000	4000
10%	3.76%	11.17%	15.61%	17.51%	26.41%
5%	3.30%	10.37%	14.48%	16.34%	24.58%
1%	2.52%	9.45%	13.02%	14.75%	21.94%

Panel C: MRK					
	Relative Ex-ante Liquidity Risk Premium at the end of 5-min Interval				
Confidence Level	Volume (Shares)				
	200	500	900	1800	2700
10%	2.03%	7.64%	12.88%	22.19%	30.14%
5%	0.77%	5.55%	10.41%	19.25%	27.42%
1%	0.67%	3.02%	7.17%	15.34%	24.42%

The panels present the proportion of liquidity risk on total risk given different order sizes at 90%, 95% and 99% confidence level for SAP, RWE, and MRK. For the end of comparison, the order sizes remain the same as used in July 2010.

Figure A 1.1: Daily price, volume, return, and LOB depth for SAP

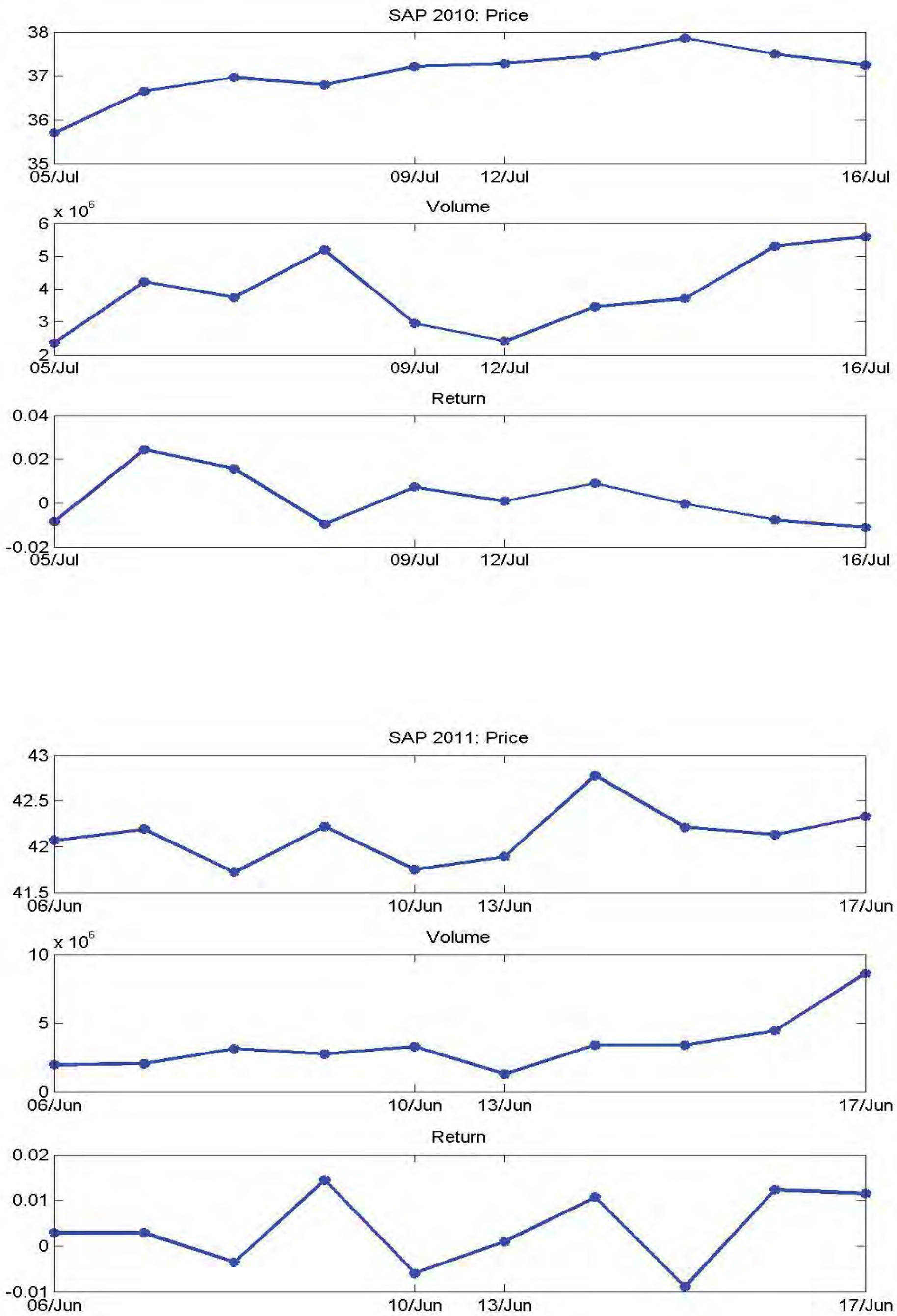
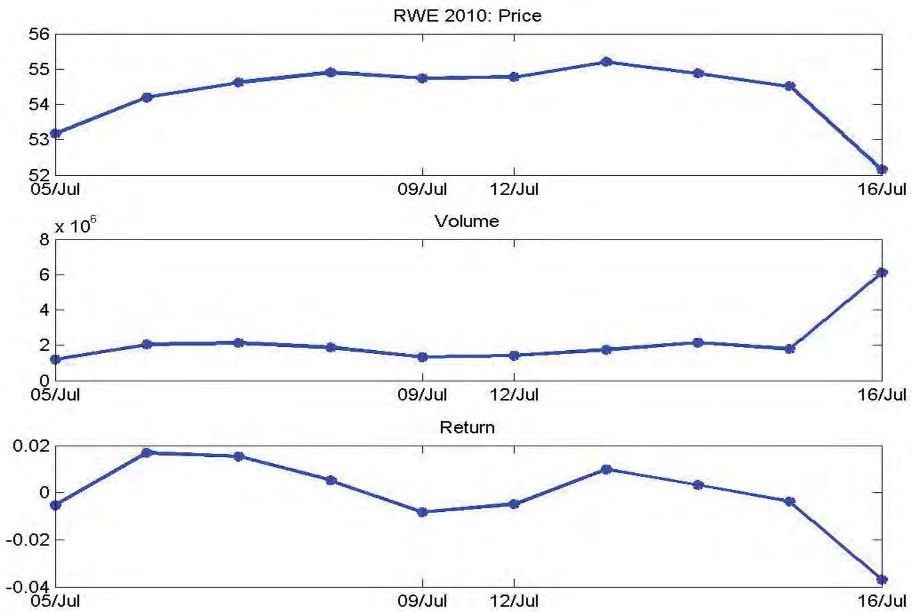


Figure A 1.2: Daily price, volume, return, and LOB depth for RWE



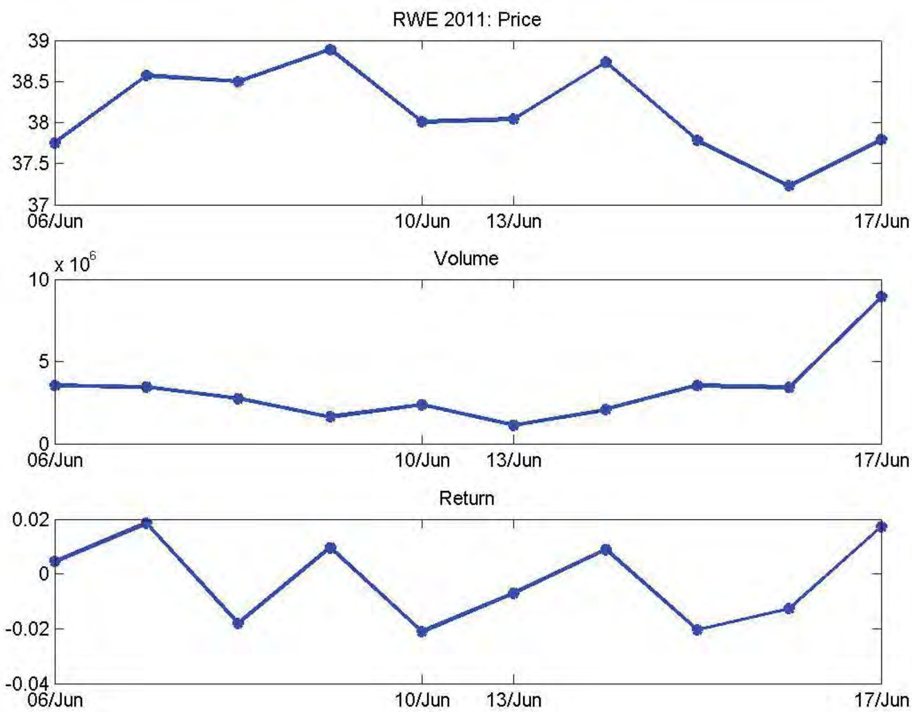
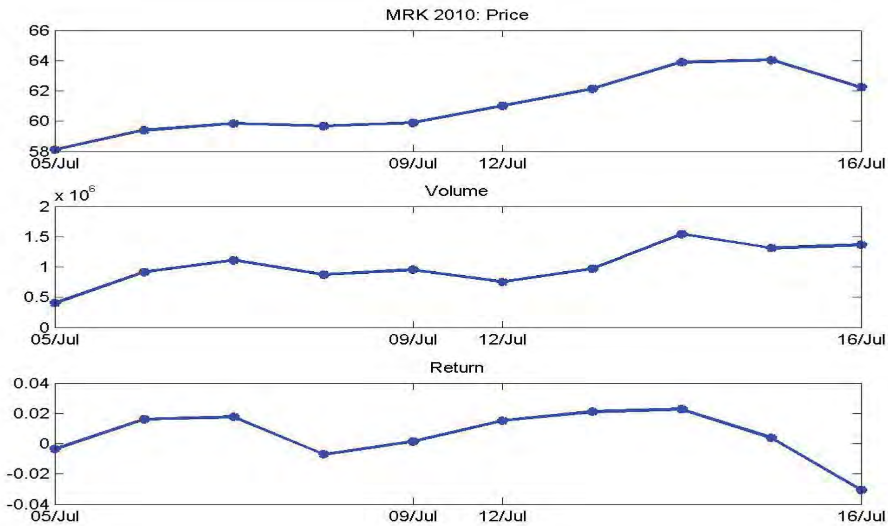


Figure A 1.3: Daily price, volume, return, and LOB depth for MRK



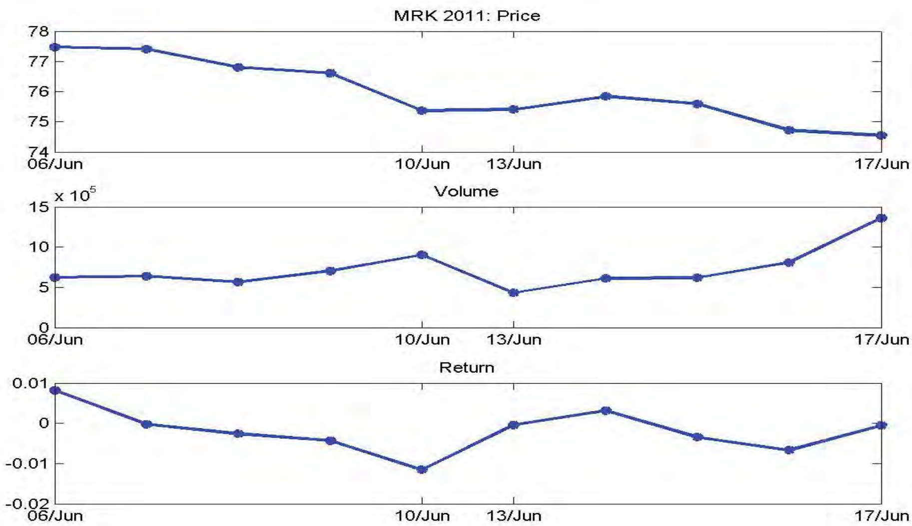


Figure A 1.1-1.3 plots daily price, trading volume, and return for the sample periods of July 2010 and June 2011 for stock SAP, RWE, and MRK.

Figure A 1.4: Intraday price, return, trading volume, and price impact (v=2000) for SAP

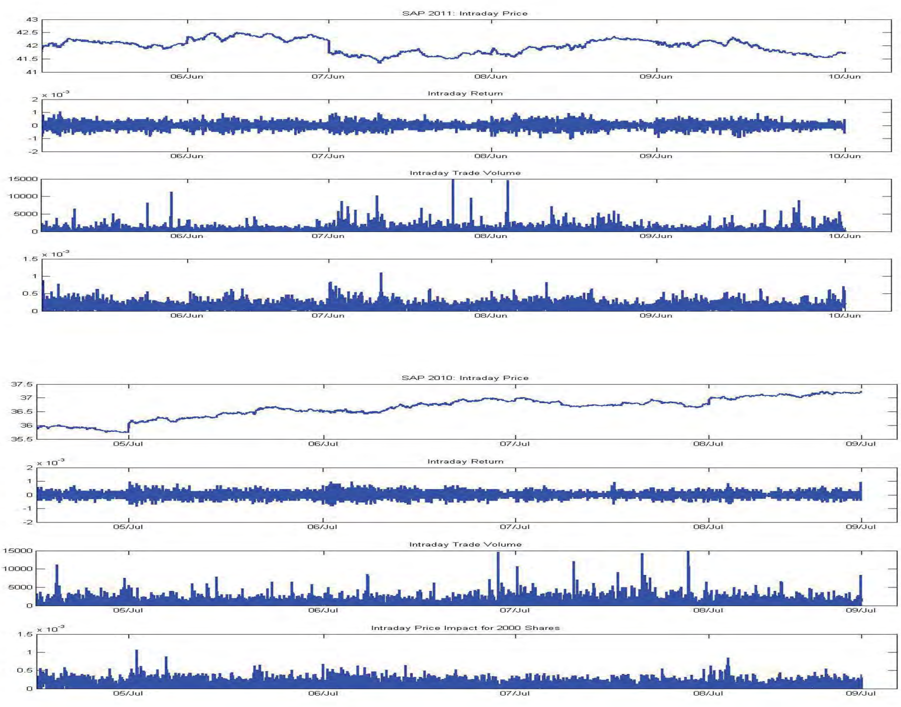


Figure A 1.5: Intraday price, return, trading volume, and price impact (v=1500) for RWE

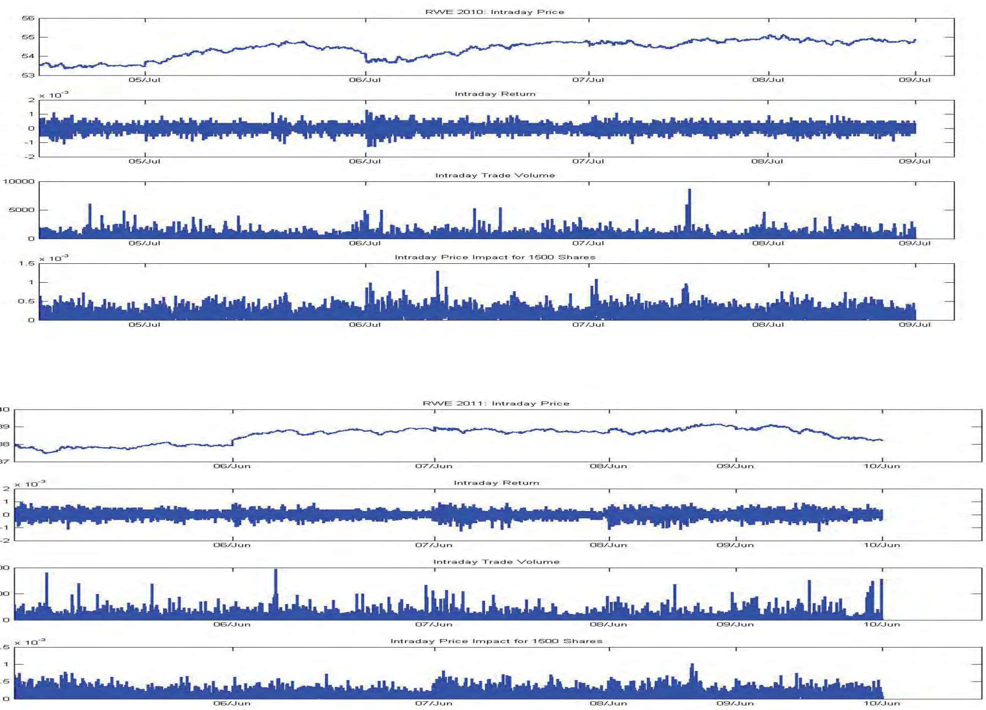


Figure A 1.6 : Intraday price, return, trading volume, and price impact (v=900) for MRK

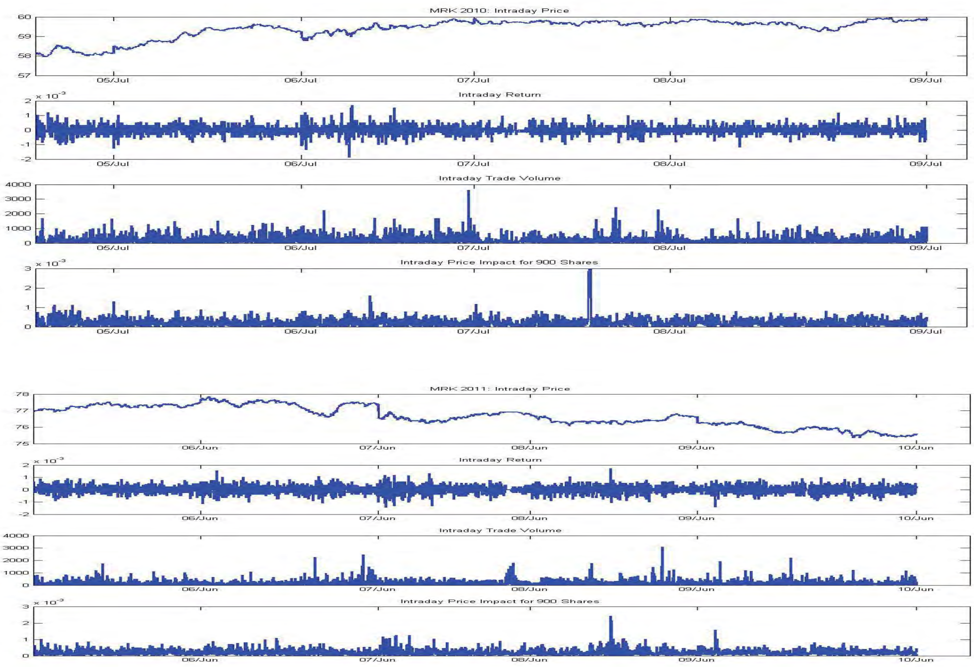


Figure A 1.4-1.6 present intraday trade price, trading volume, tick-by-tick return, and price impact for a given volume for the sample periods of July 2010 and June 2011 for stock SAP, RWE, and MRK.

Figure A 1.7: The effect of Volatility of Actual Return Changes on Liquidity Risk Premium

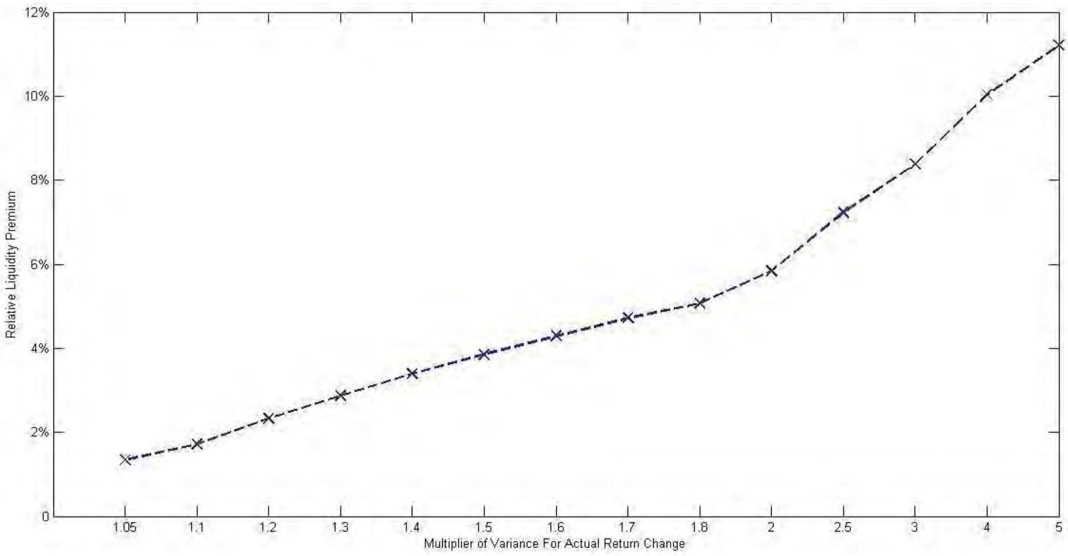


Figure A 1. 8: The effect of Volatility of Frictionless Return Changes on Liquidity Risk Premium

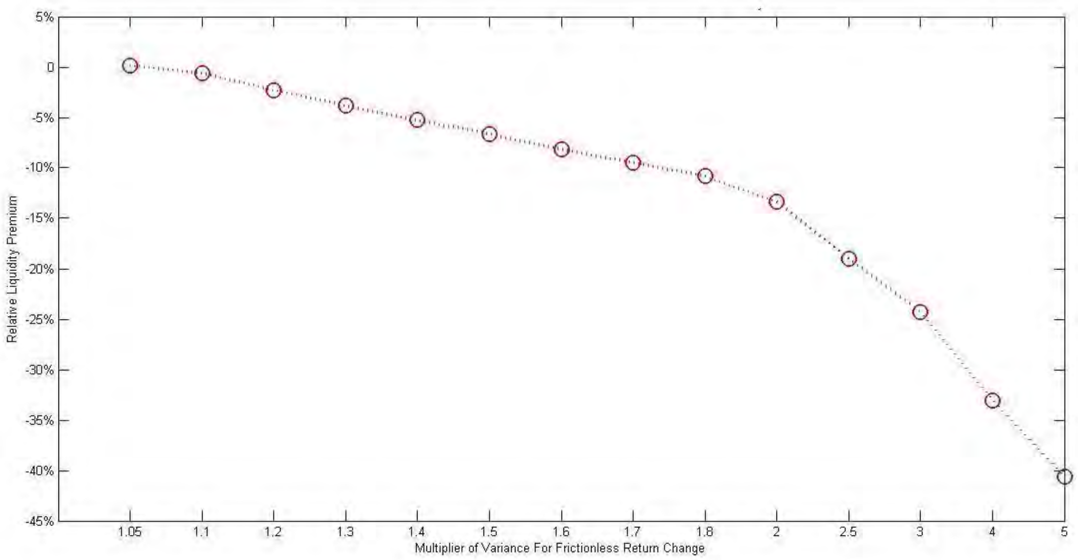


Figure A 1. 9: The effect of Volatility of Both Actual Return Changes and Frictionless Return Changes on Liquidity Risk Premium

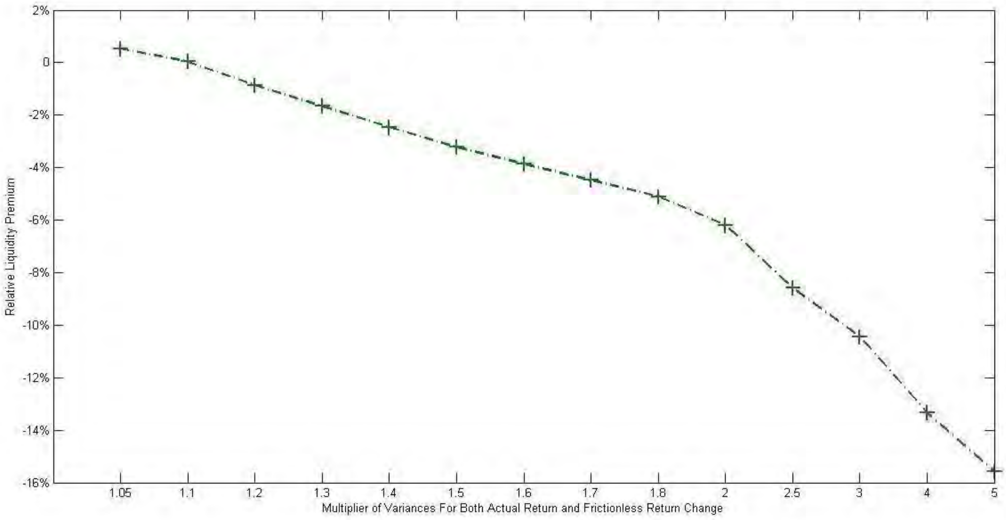
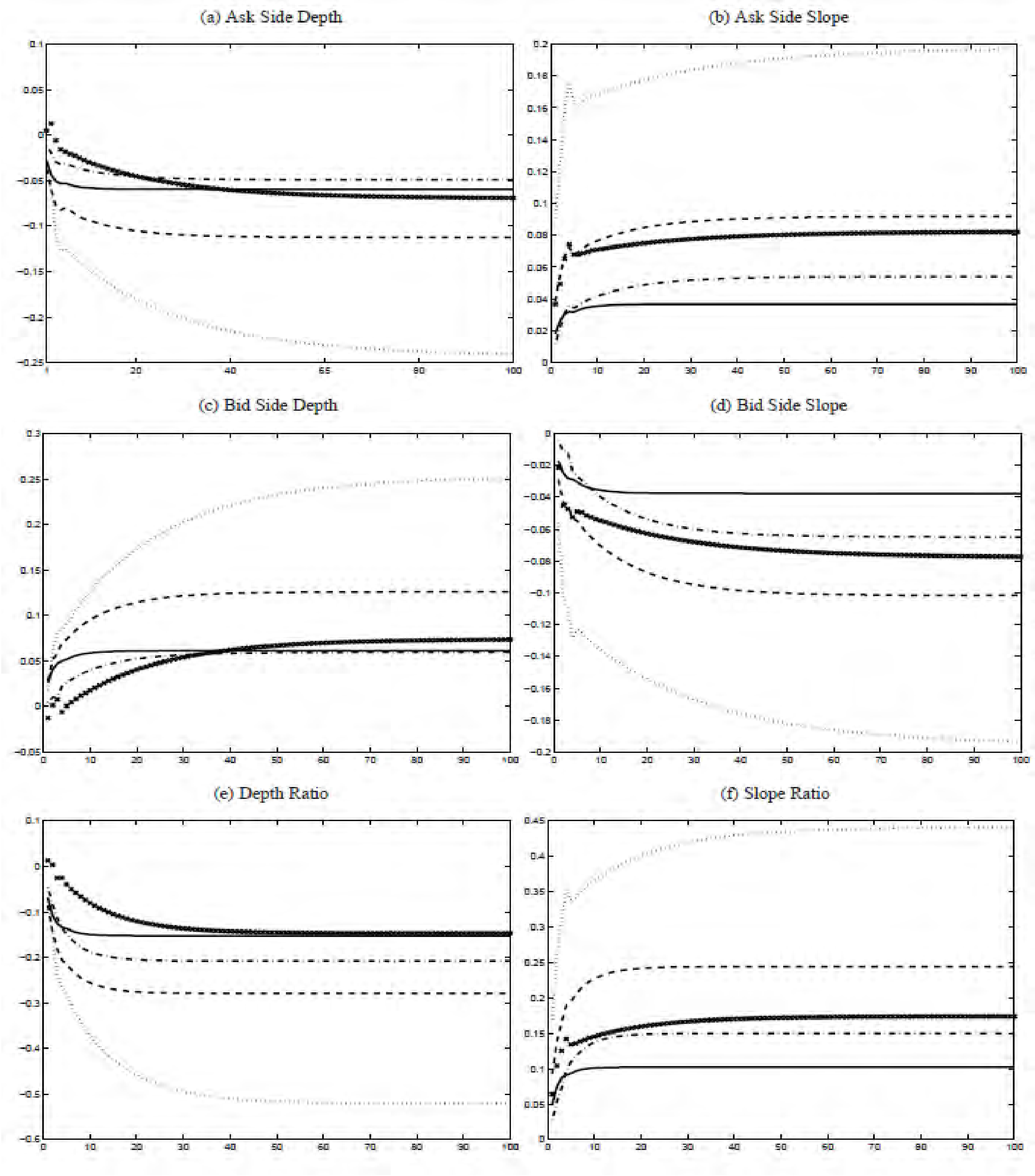


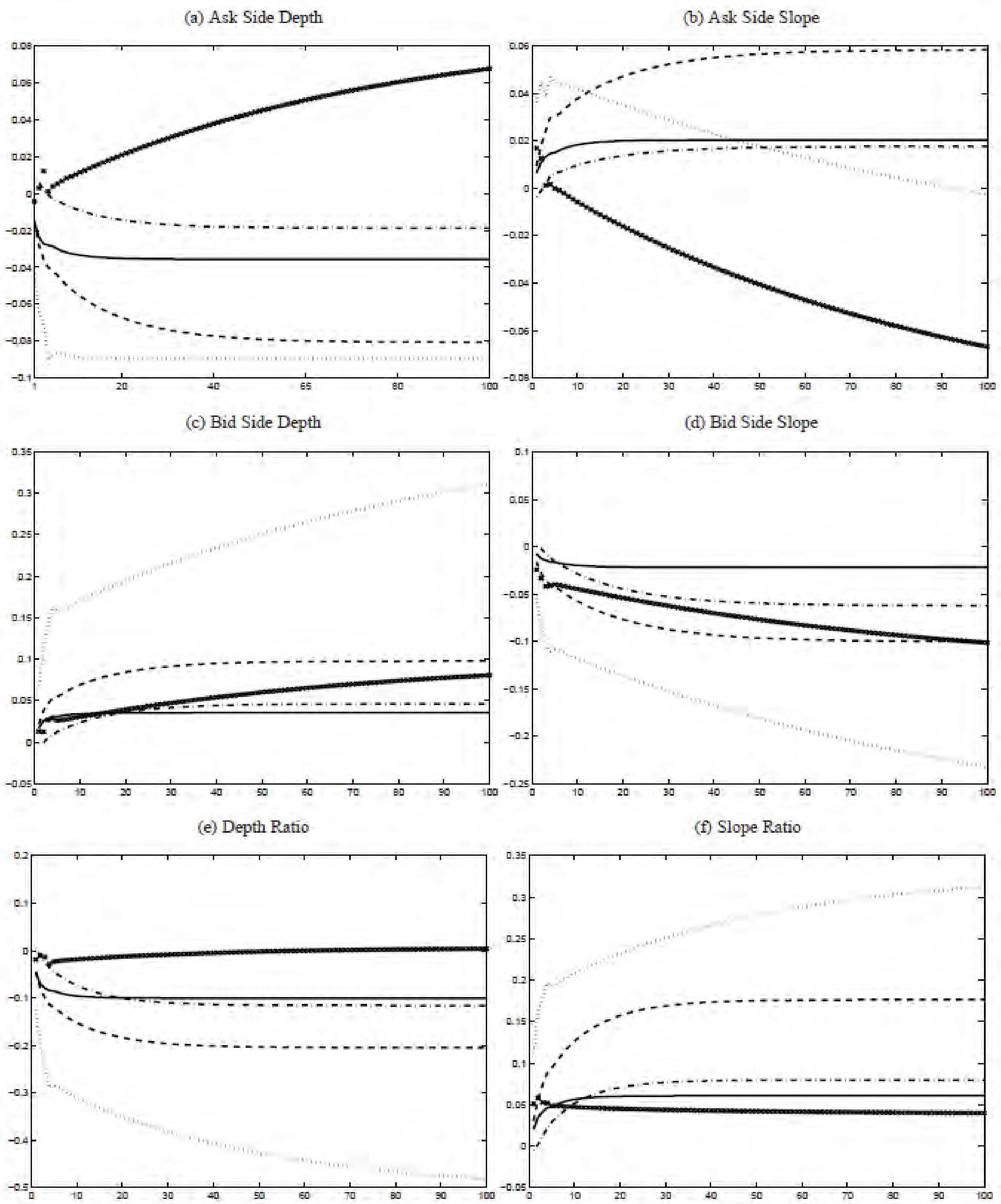
Figure A 1.7-1.9 present the evolution of liquidity risk premium, in relative term, as a function of the volatility of actual return changes, frictionless return changes, and both actual return changes and frictionless return changes, respectively. Relative liquidity premium indicate the difference between the new liquidity risk premium and the initial liquidity risk premium. For the illustration purpose, we use the estimated parameters from MRK when volume = 2700. The liquidity risk premium is for confidence level equal to 5%.

Figure A 2.1 (MRK-June 2011): Impulse Response Function of Return to Limit Order Book Variables



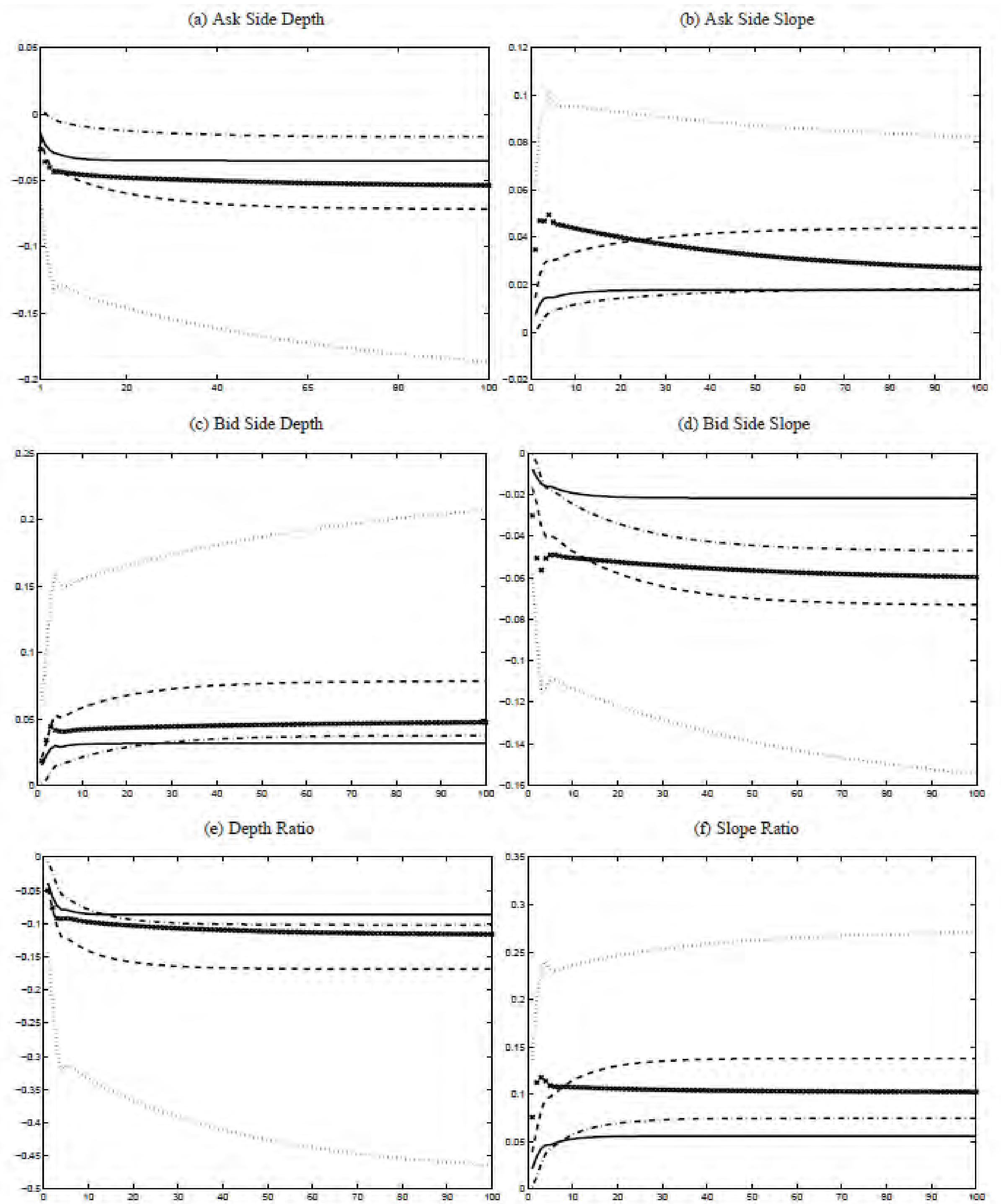
This figure presents the impulse response functions of midquote return to a one unit positive shock to limit order book variables. The horizontal axis is transaction periods, i.e. the number of transactions since the initial shock, and the vertical axis is the response of midquote return in basis points. The solid, dashed, small dotted and dashed-dotted line correspond to the limit order book variable measured based on the first two, five, ten and twenty levels, respectively. The big dotted line corresponds to the limit order book variable measured between second and fifth levels.

Figure A 2.2 (SAP-July 2010): Impulse Response Function of Return to Limit Order Book Variables



This figure presents the impulse response functions of midquote return to a one unit positive shock to limit order book variables. The horizontal axis is transaction periods, i.e. the number of transactions since the initial shock, and the vertical axis is the response of midquote return in basis points. The solid, dashed, small dotted and dashed-dotted line correspond to the limit order book variable measured based on the first two, five, ten and twenty levels, respectively. The big dotted line corresponds to the limit order book variable measured between second and fifth levels.

Figure A 2.3 (SAP-June 2011): Impulse Response Function of Return to Limit Order Book Variables



This figure presents the impulse response functions of midquote return to a one unit positive shock to limit order book variables. The horizontal axis is transaction periods, i.e. the number of transactions since the initial shock, and the vertical axis is the response of midquote return in basis points. The solid, dashed, small dotted and dashed-dotted line correspond to the limit order book variable measured based on the first two, five, ten and twenty levels, respectively. The big dotted line corresponds to the limit order book variable measured between second and fifth levels.

Table A 2.1 (MRK-June 2011): Coefficient Estimates on Lagged Values of Limit Order Book Variables (Z_{i-j}) and F-statistics for Granger Causality Tests

	$\alpha_{z,1}$	$\alpha_{z,2}$	$\alpha_{z,3}$	$\alpha_{z,4}$	$\alpha_{z,5}$	F-stat
Ask Side Variables						
Depth (Levels 1-2)	-0.280***	-0.021	0.013	0.039*	0.029	484.462***
Depth (Levels 1-5)	-0.387***	0.031	-0.023	0.119***	0.086**	225.580***
Depth (Levels 1-20)	-0.460***	0.194	-0.368***	0.307**	0.163	73.831***
Depth (Levels 2-5)	-0.132***	0.019	-0.038	0.052	0.025	39.603***
Depth (Levels 6-20)	0.046	0.052	-0.253**	0.041	0.050	14.659**
Slope (Levels 1-2)	0.183***	-0.004	-0.008	-0.020	-0.016	347.733***
Slope (Levels 1-5)	0.398***	-0.100*	-0.017	-0.063	-0.100***	265.603***
Slope (Levels 1-20)	0.940***	-0.395***	-0.004	-0.172	-0.258**	241.749***
Slope (Levels 2-5)	0.134***	-0.003	0.020	-0.027	-0.050	51.356***
Slope (Levels 6-20)	0.377***	-0.149	0.032	-0.066	-0.132*	64.606***
Bid Side Variables						
Depth (Levels 1-2)	0.258***	0.034*	-0.013	-0.044**	-0.003	436.874***
Depth (Levels 1-5)	0.286***	0.071	-0.104**	-0.001	-0.042	160.880***
Depth (Levels 1-20)	0.231*	0.214	-0.170	-0.133	0.036	39.615***
Depth (Levels 2-5)	0.045	0.041	-0.055	0.080*	-0.012	28.334***
Depth (Levels 6-20)	-0.089	0.133	-0.017	-0.128	0.153*	10.034*
Slope (Levels 1-2)	-0.168***	-0.007	-0.001	0.032**	-0.004	330.725***
Slope (Levels 1-5)	-0.285***	0.022	0.097**	-0.009	0.040	159.143***
Slope (Levels 1-20)	-0.605***	0.039	0.305**	-0.033	0.203*	119.103***
Slope (Levels 2-5)	-0.050	-0.045	0.066	-0.089*	0.033	35.210***
Slope (Levels 6-20)	-0.249***	-0.001	0.140	0.003	0.074	32.606***
Ratio Variables						
Depth Ratio (Levels 1-2)	-0.682***	-0.077**	0.032	0.091**	0.036	958.045***
Depth Ratio (Levels 1-5)	-0.872***	-0.107	0.042	0.103	0.075	492.646***
Depth Ratio (Levels 1-20)	-0.864***	-0.031	-0.238	0.438**	0.043	129.374***
Depth Ratio (Levels 2-5)	-0.491***	-0.087	-0.011	0.013	-0.001	209.102***
Depth Ratio (Levels 6-20)	0.090	-0.076	-0.260*	0.168	-0.124	16.998***
Slope Ratio (Levels 1-2)	0.500***	0.028	0.007	-0.073**	-0.011	699.074***
Slope Ratio (Levels 1-5)	0.937***	-0.060	-0.045	-0.045	-0.090	556.052***
Slope Ratio (Levels 1-20)	1.780***	-0.479***	-0.330*	-0.138	-0.461***	381.203***
Slope Ratio (Levels 2-5)	0.302***	0.089	-0.008	0.101	-0.029	126.460***
Slope Ratio (Levels 6-20)	0.690***	-0.167	-0.104	-0.064	-0.201*	92.848***

This table presents the parameter estimates of lagged limit order book variables in the return equation ($\alpha_{z,\tau}$) and F-statistics for Granger Causality tests. ***, **, * represent statistical significance at 1%, 5%, 10%, respectively.

Table A 2.2 (SAP-July 2010): Coefficient Estimates on Lagged Values of Limit Order Book Variables (Zi-j) and F-statistics for Granger Causality Tests

	$\alpha_{z,1}$	$\alpha_{z,2}$	$\alpha_{z,3}$	$\alpha_{z,4}$	$\alpha_{z,5}$	F-stat
Ask Side Variables						
Depth (Levels 1-2)	-0.149***	-0.005	0.018**	0.028***	0.010	725.592***
Depth (Levels 1-5)	-0.170***	0.004	0.017	0.030	0.032	200.519***
Depth (Levels 1-20)	-0.435***	0.134	0.139	-0.070	0.213**	81.138***
Depth (Levels 2-5)	0.036**	-0.005	-0.035*	-0.017	0.004	20.061***
Depth (Levels 6-20)	-0.014	0.050	0.051	-0.172**	0.100	11.644**
Slope (Levels 1-2)	0.066***	0.010	-0.002	-0.005	-0.011*	292.209***
Slope (Levels 1-5)	0.119***	-0.019	-0.001	0.003	-0.051***	115.713***
Slope (Levels 1-20)	0.366***	-0.234***	-0.126	0.087	-0.096	72.913***
Slope (Levels 2-5)	-0.010	0.008	0.017	0.021	-0.021	9.634*
Slope (Levels 6-20)	0.138***	-0.126*	-0.093	0.074	-0.006	24.225***
Bid Side Variables						
Depth (Levels 1-2)	0.164***	0.006	-0.032***	0.001	-0.031***	866.889***
Depth (Levels 1-5)	0.233***	-0.004	-0.044*	-0.006	-0.066***	348.940***
Depth (Levels 1-20)	0.720***	-0.167	-0.086	-0.113	-0.287***	221.762***
Depth (Levels 2-5)	0.003	-0.001	0.046**	-0.008	0.010	34.827***
Depth (Levels 6-20)	0.123*	-0.084	0.099	-0.071	-0.044	14.388**
Slope (Levels 1-2)	-0.075***	-0.014**	0.010	0.003	0.011*	370.853***
Slope (Levels 1-5)	-0.174***	0.003	0.036	0.011	0.038*	248.443***
Slope (Levels 1-20)	-0.561***	0.162	0.153	0.104	0.103	210.339***
Slope (Levels 2-5)	-0.017	0.000	-0.029	-0.006	-0.003	49.262***
Slope (Levels 6-20)	-0.238***	0.094	0.030	0.080	0.013	54.028***
Ratio Variables						
Depth Ratio (Levels 1-2)	-0.445***	-0.013	0.077***	0.036*	0.053***	1,792.649***
Depth Ratio (Levels 1-5)	-0.484***	0.013	0.071	0.039	0.099***	598.490***
Depth Ratio (Levels 1-20)	-1.304***	0.326**	0.250*	0.037	0.523***	318.873***
Depth Ratio (Levels 2-5)	-0.028	-0.074*	-0.033	-0.026	0.013	90.826***
Depth Ratio (Levels 6-20)	-0.144	0.137	-0.031	-0.131	0.158	9.335*
Slope Ratio (Levels 1-2)	0.208***	0.045***	-0.020	-0.007	-0.032***	700.599***
Slope Ratio (Levels 1-5)	0.354***	-0.023	-0.021	-0.010	-0.077**	398.015***
Slope Ratio (Levels 1-20)	1.158***	-0.434***	-0.353**	-0.028	-0.242**	305.225***
Slope Ratio (Levels 2-5)	0.007	0.018	0.060	0.028	-0.006	55.836***
Slope Ratio (Levels 6-20)	0.463***	-0.250**	-0.168	-0.002	-0.021	62.026***

This table presents the parameter estimates of lagged limit order book variables in the return equation ($\alpha_{z,\tau}$) and F-statistics for Granger Causality tests. ***, **, * represent statistical significance at 1%, 5%, 10%, respectively.

Table A 2.3 (SAP-June 2011): Coefficient Estimates on Lagged Values of Limit Order Book Variables (Zi-j) and F-statistics for Granger Causality Tests

	$\alpha_{z,1}$	$\alpha_{z,2}$	$\alpha_{z,3}$	$\alpha_{z,4}$	$\alpha_{z,5}$	F-stat
Ask Side Variables						
Depth (Levels 1-2)	-0.145***	-0.011	0.010	0.011	0.017**	759.731***
Depth (Levels 1-5)	-0.181***	0.001	0.027	0.026	0.048**	227.392***
Depth (Levels 1-20)	-0.676***	0.262***	0.124	0.042	0.181**	250.467***
Depth (Levels 2-5)	0.022	-0.028	-0.014	0.012	-0.013	11.329**
Depth (Levels 6-20)	-0.276***	0.147**	0.049	0.012	0.040	55.613***
Slope (Levels 1-2)	0.075***	0.004	-0.004	-0.017**	-0.012*	298.485***
Slope (Levels 1-5)	0.150***	-0.020	-0.031	-0.034	-0.030	146.611***
Slope (Levels 1-20)	0.635***	-0.279***	-0.199**	-0.001	-0.131**	342.564***
Slope (Levels 2-5)	0.004	0.021	0.004	-0.005	-0.004	9.628*
Slope (Levels 6-20)	0.350***	-0.161***	-0.112	-0.005	-0.059	154.401***
Bid Side Variables						
Depth (Levels 1-2)	0.147***	0.007	0.005	-0.019**	-0.033***	784.009***
Depth (Levels 1-5)	0.212***	-0.027	0.030	-0.042	-0.093***	323.400***
Depth (Levels 1-20)	0.672***	-0.229**	0.045	-0.186*	-0.257***	271.469***
Depth (Levels 2-5)	0.019	0.009	0.035	-0.011	-0.026	22.445***
Depth (Levels 6-20)	0.210***	-0.058	-0.037	-0.094	-0.013	36.769***
Slope (Levels 1-2)	-0.077***	-0.001	-0.005	0.009	0.019***	336.439***
Slope (Levels 1-5)	-0.171***	0.044*	-0.035	0.046	0.064***	222.854***
Slope (Levels 1-20)	-0.608***	0.200**	0.094	0.218***	0.071	333.443***
Slope (Levels 2-5)	-0.016	-0.020	-0.036	0.016	0.027	32.190***
Slope (Levels 6-20)	-0.315***	0.068	0.129	0.116*	-0.007	142.415***
Ratio Variables						
Depth Ratio (Levels 1-2)	-0.384***	-0.020	-0.001	0.037*	0.059***	1,613.299***
Depth Ratio (Levels 1-5)	-0.476***	0.022	-0.024	0.065	0.138***	646.022***
Depth Ratio (Levels 1-20)	-1.503***	0.536***	0.086	0.230	0.453***	570.258***
Depth Ratio (Levels 2-5)	-0.093**	-0.058	-0.092	0.031	0.057*	121.941***
Depth Ratio (Levels 6-20)	-0.541***	0.242*	0.085	0.109	0.048	96.736***
Slope Ratio (Levels 1-2)	0.218***	0.013	0.013	-0.032**	-0.037***	666.664***
Slope Ratio (Levels 1-5)	0.405***	-0.066	0.035	-0.079	-0.096***	436.497***
Slope Ratio (Levels 1-20)	1.401***	-0.498***	-0.336*	-0.241**	-0.219**	716.291***
Slope Ratio (Levels 2-5)	0.041	0.056	0.054	0.001	-0.033	62.194***
Slope Ratio (Levels 6-20)	0.778***	-0.281**	-0.281	-0.129	-0.047	317.660***

This table presents the parameter estimates of lagged limit order book variables in the return equation ($\alpha_{z,\tau}$) and F-statistics for Granger Causality tests. ***, **, * represent statistical significance at 1%, 5%, 10%, respectively.

Table A 2.4 (MRK-June 2011): Bid-Ask Return on Trading Strategies

	Number of Trades	Latency 0 ms	500 ms	1000 ms
Ask Side Variables				
Depth (Levels 1-2)	25	2.633%	1.941%	1.996%
Depth (Levels 1-5)	19	2.266%	1.744%	1.796%
Depth (Levels 1-20)	21	-2.133%	-2.784%	-2.708%
Depth (Levels 2-5)	19	-7.106%	-8.311%	-8.251%
Depth (Levels 6-20)	11	2.148%	1.842%	1.816%
Slope (Levels 1-2)	24	1.633%	0.957%	0.435%
Slope (Levels 1-5)	25	1.312%	0.565%	0.073%
Slope (Levels 1-20)	22	1.545%	0.650%	0.131%
Slope (Levels 2-5)	15	-1.594%	-2.098%	-2.035%
Slope (Levels 6-20)	11	0.190%	-0.020%	-0.046%
Bid Side Variables				
Depth (Levels 1-2)	18	1.762%	1.167%	1.010%
Depth (Levels 1-5)	18	4.569%	3.709%	3.709%
Depth (Levels 1-20)	13	-5.807%	-5.795%	-5.598%
Depth (Levels 2-5)	13	3.497%	3.267%	3.267%
Depth (Levels 6-20)	13	-2.260%	-2.452%	-2.452%
Slope (Levels 1-2)	20	-0.763%	-1.760%	-1.833%
Slope (Levels 1-5)	19	3.063%	1.954%	1.796%
Slope (Levels 1-20)	18	-2.432%	-3.259%	-3.409%
Slope (Levels 2-5)	10	3.075%	2.764%	2.764%
Slope (Levels 6-20)	9	2.503%	2.116%	2.169%
Ratio Variables				
Depth Ratio (Levels 1-2)	35	1.262%	0.136%	0.032%
Depth Ratio (Levels 1-5)	35	1.379%	-0.107%	0.002%
Depth Ratio (Levels 1-20)	15	1.137%	0.500%	0.710%
Depth Ratio (Levels 2-5)	24	-2.799%	-2.929%	-2.903%
Depth Ratio (Levels 6-20)	13	-1.399%	-1.464%	-1.258%
Slope Ratio (Levels 1-2)	35	2.346%	1.097%	0.497%
Slope Ratio (Levels 1-5)	40	1.625%	-0.278%	-0.496%
Slope Ratio (Levels 1-20)	26	3.483%	3.103%	3.183%
Slope Ratio (Levels 2-5)	22	0.381%	-0.006%	-0.006%
Slope Ratio (Levels 6-20)	8	2.598%	2.344%	2.344%
Benchmark with MA filter	9	-0.233%	-0.402%	-0.377%
Benchmark without MA filter	86	-10.645%	-8.105%	-9.061%

This table presents the cumulative bid-ask returns on trading strategies for MRK over the whole trading period of June 2011. The trading strategy is based on the empirical model in Equation 2.9 that excludes the contemporaneous effect of trade direction on return and described in detail in Section 2.9. The parameter of the filter for the return forecasts, κ, is set to 1 basis points. The long and short-run moving average filters are calculated based on the last transaction price and last forty transaction prices, including the most recent one, respectively. The parameter of the moving average filter is set to 0.12 euros. Benchmark with and without moving average (MA) filter present results from trading strategies based on a forecasting model that ignores information embedded in limit order book variables.

Table A 2.5 (SAP-July 2010): Bid-Ask Return on Trading Strategies

	Number of Trades	Latency 0 ms	500 ms	1000 ms
Ask Side Variables				
Depth (Levels 1-2)	31	8.102%	7.336%	6.717%
Depth (Levels 1-5)	31	-0.634%	-0.672%	-0.896%
Depth (Levels 1-20)	33	0.296%	0.081%	-0.253%
Depth (Levels 2-5)	29	2.757%	2.634%	2.404%
Depth (Levels 6-20)	27	1.210%	1.421%	1.329%
Slope (Levels 1-2)	31	6.881%	6.752%	6.267%
Slope (Levels 1-5)	32	4.988%	4.451%	4.004%
Slope (Levels 1-20)	27	-0.955%	-1.082%	-1.384%
Slope (Levels 2-5)	31	-3.292%	-3.482%	-3.880%
Slope (Levels 6-20)	27	2.902%	2.560%	2.246%
Bid Side Variables				
Depth (Levels 1-2)	37	1.456%	1.269%	1.232%
Depth (Levels 1-5)	39	1.417%	0.872%	0.237%
Depth (Levels 1-20)	29	1.979%	1.817%	1.760%
Depth (Levels 2-5)	31	0.574%	0.251%	-0.190%
Depth (Levels 6-20)	29	-0.467%	-0.288%	-0.404%
Slope (Levels 1-2)	35	-3.190%	-4.397%	-5.165%
Slope (Levels 1-5)	37	1.629%	1.093%	0.392%
Slope (Levels 1-20)	25	2.196%	1.829%	1.815%
Slope (Levels 2-5)	27	1.161%	1.358%	1.076%
Slope (Levels 6-20)	27	3.472%	3.366%	3.078%
Ratio Variables				
Depth Ratio (Levels 1-2)	45	3.822%	3.529%	3.377%
Depth Ratio (Levels 1-5)	39	5.662%	5.304%	5.038%
Depth Ratio (Levels 1-20)	30	7.201%	7.258%	6.824%
Depth Ratio (Levels 2-5)	29	-0.355%	-0.430%	-0.734%
Depth Ratio (Levels 6-20)	29	4.221%	4.322%	4.228%
Slope Ratio (Levels 1-2)	39	8.819%	9.001%	8.664%
Slope Ratio (Levels 1-5)	37	3.951%	3.130%	2.339%
Slope Ratio (Levels 1-20)	29	4.067%	4.087%	3.754%
Slope Ratio (Levels 2-5)	29	2.328%	2.237%	1.925%
Slope Ratio (Levels 6-20)	27	1.082%	0.729%	0.585%
Benchmark with MA filter	29	-2.231%	-2.001%	-2.115%
Benchmark without MA filter	536	-23.693%	-26.451%	-27.623%

This table presents the cumulative bid-ask returns on trading strategies for SAP over the whole trading period of July 2010. The trading strategy is based on the empirical model in Equation 2.9 that excludes the contemporaneous effect of trade direction on return and described in detail in Section 2.9. The parameter of the filter for the return forecasts, κ, is set to 0.5 basis points. The long and short-run moving average filters are calculated based on the last transaction price and last forty transaction prices, including the most recent one, respectively. The parameter of the moving average filter is set to 0.05 euros. Benchmark with and without moving average (MA) filter present results from trading strategies based on a forecasting model that ignores information embedded in limit order book variables.

Table A 2.6 (SAP-June 2011): Bid-Ask Return on Trading Strategies

	Number of Trades	Latency 0 ms	500 ms	1000 ms
Ask Side Variables				
Depth (Levels 1-2)	22	2.546%	2.475%	2.388%
Depth (Levels 1-5)	24	3.391%	3.841%	3.383%
Depth (Levels 1-20)	24	0.297%	0.323%	0.264%
Depth (Levels 2-5)	24	3.292%	3.841%	3.383%
Depth (Levels 6-20)	26	2.753%	3.323%	2.844%
Slope (Levels 1-2)	25	0.373%	0.257%	0.222%
Slope (Levels 1-5)	24	3.490%	3.940%	3.482%
Slope (Levels 1-20)	24	0.777%	0.947%	0.864%
Slope (Levels 2-5)	24	3.292%	3.841%	3.383%
Slope (Levels 6-20)	26	2.802%	3.175%	2.696%
Bid Side Variables				
Depth (Levels 1-2)	27	4.693%	4.539%	4.783%
Depth (Levels 1-5)	27	4.618%	4.370%	4.384%
Depth (Levels 1-20)	26	5.597%	5.535%	5.460%
Depth (Levels 2-5)	25	3.102%	3.526%	3.070%
Depth (Levels 6-20)	25	4.137%	4.565%	4.105%
Slope (Levels 1-2)	25	1.597%	1.990%	1.516%
Slope (Levels 1-5)	27	3.938%	3.668%	3.631%
Slope (Levels 1-20)	25	5.268%	5.010%	4.910%
Slope (Levels 2-5)	25	3.102%	3.526%	3.070%
Slope (Levels 6-20)	24	2.877%	2.843%	2.794%
Ratio Variables				
Depth Ratio (Levels 1-2)	25	4.543%	4.264%	4.533%
Depth Ratio (Levels 1-5)	23	5.095%	5.044%	5.059%
Depth Ratio (Levels 1-20)	24	5.707%	5.811%	5.620%
Depth Ratio (Levels 2-5)	25	3.005%	3.576%	3.120%
Depth Ratio (Levels 6-20)	27	3.520%	4.094%	3.611%
Slope Ratio (Levels 1-2)	25	0.166%	-0.265%	-0.129%
Slope Ratio (Levels 1-5)	25	5.193%	5.172%	5.135%
Slope Ratio (Levels 1-20)	23	3.028%	2.913%	3.061%
Slope Ratio (Levels 2-5)	25	3.005%	3.576%	3.120%
Slope Ratio (Levels 6-20)	25	2.601%	2.960%	2.469%
Benchmark with MA filter	25	3.005%	3.428%	2.972%
Benchmark without MA filter	24536	-99.877%	-99.961%	-99.969%

This table presents the cumulative bid-ask returns on trading strategies for SAP over the whole trading period of June 2011. The trading strategy is based on the empirical model in Equation 2.9 that excludes the contemporaneous effect of trade direction on return and described in detail in Section 2.9. The parameter of the filter for the return forecasts, κ, is set to 0.3 basis points. The long and short-run moving average filters are calculated based on the last transaction price and last forty transaction prices, including the most recent one, respectively. The parameter of the moving average filter is set to 0.06 euros. Benchmark with and without moving average (MA) filter present results from trading strategies based on a forecasting model that ignores information embedded in limit order book variables.

Table A 2.7 Coefficient Estimates on Lagged Values of Limit Order Book Variables and Lagged Values of Volume and F-statistics for Granger Causality Tests for Limit Order Book Variables

	$\alpha_{z,1}$	$\alpha_{z,2}$	$\alpha_{z,3}$	$\alpha_{z,4}$	$\alpha_{z,5}$	$\theta_{z,1}$	$\theta_{z,2}$	$\theta_{z,3}$	$\theta_{z,4}$	$\theta_{z,5}$	F-stat
Ask Side Variables											
Depth (Levels 1-2)	-0.260***	-0.001	0.006	0.028	0.018	0.021**	0.010	-0.006	0.007	-0.001	415.482***
Depth (Levels 1-5)	-0.225***	0.019	0.033	-0.005	0.063*	0.012	0.007	-0.007	0.005	-0.004	76.81***
Depth (Levels 1-20)	-0.533***	0.225	0.004	-0.228	0.401***	0.010	0.006	-0.006	0.006	-0.004	52.36***
Depth (Levels 2-5)	0.057*	-0.021	0.022	-0.074**	0.026	0.007	0.006	-0.006	0.006	-0.004	11.16**
Depth (Levels 6-20)	-0.101	0.131	-0.091	-0.216*	0.194*	0.008	0.006	-0.006	0.006	-0.003	16.71***
Slope (Levels 1-2)	0.132***	-0.009	0.002	-0.009	-0.018	0.011	0.007	-0.007	0.005	-0.004	194.29***
Slope (Levels 1-5)	0.190***	-0.027	-0.039	0.014	-0.074**	0.010	0.005	-0.008	0.005	-0.005	68.35***
Slope (Levels 1-20)	0.575***	-0.066	-0.324**	0.267	-0.380***	0.007	0.005	-0.007	0.005	-0.004	92.95***
Slope (Levels 2-5)	-0.019	0.030	-0.022	0.059*	-0.033	0.008	0.006	-0.006	0.006	-0.004	6.79
Slope (Levels 6-20)	0.237**	-0.010	-0.169	0.196*	-0.199**	0.007	0.006	-0.006	0.005	-0.003	32.37***
Bid Side Variables											
Depth (Levels 1-2)	0.286***	0.001	-0.001	-0.028	-0.019	-0.006	0.000	-0.010	0.003	-0.008	444.45***
Depth (Levels 1-5)	0.260***	-0.042	-0.009	-0.065	-0.027	0.002	0.005	-0.007	0.006	-0.004	92.20***
Depth (Levels 1-20)	0.627***	-0.308	-0.046	-0.059	-0.128	0.006	0.006	-0.006	0.006	-0.004	51.74***
Depth (Levels 2-5)	-0.008	-0.006	0.017	-0.022	0.023	0.008	0.006	-0.006	0.006	-0.004	1.24
Depth (Levels 6-20)	0.240**	-0.123	-0.084	0.068	-0.056	0.008	0.006	-0.006	0.005	-0.004	11.67**
Slope (Levels 1-2)	-0.151***	-0.022	-0.011	0.022	0.007	0.003	0.005	-0.007	0.005	-0.004	302.62***
Slope (Levels 1-5)	-0.278***	0.015	0.045	0.095**	0.031	0.004	0.006	-0.005	0.008	-0.002	154.25***
Slope (Levels 1-20)	-0.841***	0.186	0.292**	0.152	0.131	0.008	0.007	-0.005	0.007	-0.003	172.45***
Slope (Levels 2-5)	-0.049	-0.006	0.002	0.022	0.000	0.007	0.006	-0.006	0.006	-0.003	8.76
Slope (Levels 6-20)	-0.401***	0.080	0.182*	-0.038	0.118	0.009	0.006	-0.006	0.006	-0.004	62.40***
Ratio Variables											
Depth Ratio (Levels 1-2)	-0.751***	0.003	0.008	0.092***	0.047	0.008	0.005	-0.008	0.005	-0.005	943.24***
Depth Ratio (Levels 1-5)	-0.579***	0.073	0.018	0.081	0.096*	0.007	0.005	-0.008	0.005	-0.004	182.23***
Depth Ratio (Levels 1-20)	-1.386***	0.573**	0.039	-0.170	0.460**	0.007	0.005	-0.007	0.006	-0.004	129.26***
Depth Ratio (Levels 2-5)	-0.060	0.007	-0.058	0.005	-0.014	0.008	0.006	-0.007	0.005	-0.004	11.46**
Depth Ratio (Levels 6-20)	-0.436***	0.265	-0.004	-0.280	0.188	0.008	0.006	-0.007	0.006	-0.004	31.14***
Slope Ratio (Levels 1-2)	0.404***	0.024	0.033	-0.035	-0.038	0.008	0.006	-0.007	0.005	-0.005	475.73***
Slope Ratio (Levels 1-5)	0.582***	-0.043	-0.080	-0.091	-0.096*	0.007	0.005	-0.008	0.005	-0.004	225.27***
Slope Ratio (Levels 1-20)	1.897***	-0.278	-0.724***	0.099	-0.561***	0.006	0.005	-0.006	0.006	-0.004	339.93***
Slope Ratio (Levels 2-5)	0.040	0.035	0.008	0.022	-0.032	0.008	0.006	-0.007	0.005	-0.004	7.41
Slope Ratio (Levels 6-20)	0.825***	-0.104	-0.407**	0.283*	-0.351***	0.007	0.006	-0.006	0.006	-0.004	115.06***

This table presents the parameter estimates of lagged limit order book variables ($\alpha_{z,\tau}$) and lagged volume ($\theta_{z,\tau}$) in the return equation and F-statistics for Granger Causality tests for limit order book variables. ***, **, * represent statistical significance at 1%, 5%, 10%, respectively.

Table A 2.8 Coefficient Estimates on Lagged Values of Limit Order Book Variables and Lagged Values of Volatility and F-statistics for Granger Causality Tests for Limit Order Book Variables

	$\alpha_{z,1}$	$\alpha_{z,2}$	$\alpha_{z,3}$	$\alpha_{z,4}$	$\alpha_{z,5}$	$\theta_{z,1}$	$\theta_{z,2}$	$\theta_{z,3}$	$\theta_{z,4}$	$\theta_{z,5}$	F-stat
Ask Side Variables											
Depth (Levels 1-2)	-0.261***	0.004	0.007	0.028	0.019	-0.005	-0.013	0.015	-0.001	0.007	408.561***
Depth (Levels 1-5)	-0.224***	0.024	0.032	-0.006	0.061*	-0.006	-0.013	0.015	0.000	0.006	74.764***
Depth (Levels 1-20)	-0.532***	0.236	-0.001	-0.227	0.395***	-0.006	-0.013	0.015	-0.001	0.006	51.014***
Depth (Levels 2-5)	0.058**	-0.019	0.022	-0.074**	0.025	-0.006	-0.012	0.016	0.001	0.007	11.359**
Depth (Levels 6-20)	-0.101	0.131	-0.089	-0.215*	0.192*	-0.006	-0.013	0.015	0.000	0.007	16.162***
Slope (Levels 1-2)	0.132***	-0.010	0.001	-0.008	-0.018	-0.006	-0.014	0.014	-0.001	0.006	192.846***
Slope (Levels 1-5)	0.190***	-0.028	-0.041	0.015	-0.073**	-0.007	-0.014	0.014	-0.001	0.006	67.657***
Slope (Levels 1-20)	0.581***	-0.058	-0.332**	0.267	-0.383***	-0.010	-0.014	0.014	-0.001	0.007	95.005***
Slope (Levels 2-5)	-0.020	0.030	-0.022	0.059*	-0.031	-0.006	-0.013	0.015	0.000	0.007	6.517
Slope (Levels 6-20)	0.241**	-0.005	-0.171	0.193*	-0.202**	-0.008	-0.014	0.014	0.000	0.007	33.211***
Bid Side Variables											
Depth (Levels 1-2)	0.286***	0.001	-0.003	-0.030	-0.020	-0.006	-0.012	0.016	0.002	0.008	445.273***
Depth (Levels 1-5)	0.262***	-0.037	-0.012	-0.067	-0.027	-0.007	-0.013	0.016	0.002	0.008	95.265***
Depth (Levels 1-20)	0.627***	-0.298	-0.048	-0.062	-0.129	-0.006	-0.012	0.016	0.001	0.008	52.519***
Depth (Levels 2-5)	-0.007	-0.004	0.016	-0.023	0.022	-0.006	-0.013	0.015	0.000	0.007	1.123
Depth (Levels 6-20)	0.237**	-0.122	-0.081	0.068	-0.056	-0.005	-0.012	0.015	0.000	0.007	11.546**
Slope (Levels 1-2)	-0.151***	-0.022	-0.012	0.022	0.007	-0.004	-0.010	0.017	0.003	0.010	306.058***
Slope (Levels 1-5)	-0.278***	0.013	0.043	0.095**	0.030	-0.004	-0.011	0.017	0.002	0.009	156.625***
Slope (Levels 1-20)	-0.837***	0.188	0.287**	0.145	0.131	-0.001	-0.010	0.016	0.000	0.007	170.746***
Slope (Levels 2-5)	-0.050*	-0.008	0.003	0.023	0.001	-0.005	-0.012	0.016	0.001	0.008	9.553*
Slope (Levels 6-20)	-0.397***	0.083	0.179*	-0.044	0.117	-0.003	-0.011	0.016	0.001	0.007	61.430***
Ratio Variables											
Depth Ratio (Levels 1-2)	-0.752***	0.004	0.008	0.093***	0.047	-0.006	-0.014	0.015	0.000	0.008	944.711***
Depth Ratio (Levels 1-5)	-0.581***	0.074	0.019	0.083	0.095*	-0.007	-0.014	0.015	0.000	0.007	183.617***
Depth Ratio (Levels 1-20)	-1.385***	0.571**	0.039	-0.167	0.457**	-0.006	-0.013	0.014	-0.001	0.007	129.066***
Depth Ratio (Levels 2-5)	-0.062	0.007	-0.056	0.005	-0.014	-0.006	-0.013	0.015	0.000	0.007	11.469**
Depth Ratio (Levels 6-20)	-0.432***	0.264	-0.005	-0.281	0.186	-0.006	-0.013	0.015	0.000	0.007	30.852***
Slope Ratio (Levels 1-2)	0.404***	0.023	0.032	-0.034	-0.038	-0.005	-0.012	0.015	0.001	0.008	475.647***
Slope Ratio (Levels 1-5)	0.582***	-0.041	-0.082	-0.090	-0.096*	-0.006	-0.013	0.015	0.001	0.007	225.613***
Slope Ratio (Levels 1-20)	1.893***	-0.275	-0.725***	0.105	-0.562***	-0.005	-0.012	0.015	0.000	0.007	339.446***
Slope Ratio (Levels 2-5)	0.040	0.037	0.007	0.021	-0.031	-0.006	-0.013	0.015	0.000	0.007	7.473
Slope Ratio (Levels 6-20)	0.823***	-0.104	-0.406**	0.287*	-0.352***	-0.006	-0.012	0.015	0.000	0.007	114.885***

This table presents the parameter estimates of lagged limit order book variables ($\alpha_{z,\tau}$) and lagged volatility ($\theta_{z,\tau}$) in the return equation and F-statistics for Granger Causality tests for limit order book variables. ***, **, * represent statistical significance at 1%, 5%, 10%, respectively.

Table A 2.9 Coefficient Estimates on Lagged Values of Limit Order Book Variables and Lagged Values of Duration and F-statistics for Granger Causality Tests for Limit Order Book Variables

	$\alpha_{z,1}$	$\alpha_{z,2}$	$\alpha_{z,3}$	$\alpha_{z,4}$	$\alpha_{z,5}$	$\theta_{z,1}$	$\theta_{z,2}$	$\theta_{z,3}$	$\theta_{z,4}$	$\theta_{z,5}$	F-stat
Ask Side Variables											
Depth (Levels 1-2)	-0.261***	0.001	0.008	0.027	0.022	0.008**	0.000	0.007**	-0.003	0.000	411.239***
Depth (Levels 1-5)	-0.225***	0.017	0.035	-0.010	0.064*	0.006*	0.000	0.009**	-0.003	0.001	78.238***
Depth (Levels 1-20)	-0.536***	0.224	-0.001	-0.226	0.402***	0.006*	0.000	0.008**	-0.003	0.001	53.656***
Depth (Levels 2-5)	0.057*	-0.024	0.023	-0.076**	0.025	0.005*	-0.001	0.008**	-0.003	0.001	11.375**
Depth (Levels 6-20)	-0.105	0.134	-0.100	-0.208*	0.192*	0.006*	-0.001	0.008**	-0.002	0.001	17.265***
Slope (Levels 1-2)	0.133***	-0.006	0.002	-0.007	-0.018	0.008**	0.001	0.009**	-0.002	0.001	200.595***
Slope (Levels 1-5)	0.190***	-0.021	-0.039	0.021	-0.076**	0.007**	0.001	0.009***	-0.003	0.002	73.024***
Slope (Levels 1-20)	0.585***	-0.062	-0.316**	0.264	-0.385***	0.008**	0.001	0.008**	-0.002	0.001	98.034***
Slope (Levels 2-5)	-0.017	0.033	-0.022	0.062*	-0.032	0.006*	0.000	0.008**	-0.003	0.001	8.238
Slope (Levels 6-20)	0.248**	-0.015	-0.160	0.187	-0.197**	0.007**	0.000	0.008**	-0.002	0.001	34.776***
Bid Side Variables											
Depth (Levels 1-2)	0.286***	0.000	-0.002	-0.031	-0.019	0.002	-0.003	0.007**	-0.003	0.000	439.588***
Depth (Levels 1-5)	0.259***	-0.044	-0.008	-0.070	-0.024	0.004	-0.002	0.007**	-0.004	0.000	90.746***
Depth (Levels 1-20)	0.623***	-0.309	-0.046	-0.062	-0.125	0.005	-0.001	0.007**	-0.003	0.000	50.207***
Depth (Levels 2-5)	-0.009	-0.008	0.017	-0.025	0.022	0.006*	-0.001	0.008**	-0.003	0.001	1.348
Depth (Levels 6-20)	0.235**	-0.122	-0.088	0.071	-0.058	0.005	-0.001	0.008**	-0.003	0.000	10.840*
Slope (Levels 1-2)	-0.151***	-0.022	-0.012	0.023*	0.006	0.001	-0.004	0.006*	-0.004	-0.001	299.381***
Slope (Levels 1-5)	-0.277***	0.018	0.042	0.101**	0.027	0.003	-0.002	0.007**	-0.004	0.000	151.560***
Slope (Levels 1-20)	-0.837***	0.185	0.296**	0.151	0.128	0.002	-0.002	0.007**	-0.003	0.001	167.886***
Slope (Levels 2-5)	-0.048	-0.004	0.002	0.025	0.000	0.005	-0.001	0.007**	-0.003	0.000	7.551
Slope (Levels 6-20)	-0.394***	0.076	0.189*	-0.046	0.120	0.004	-0.002	0.007**	-0.004	0.000	59.089***
Ratio Variables											
Depth Ratio (Levels 1-2)	-0.751***	0.004	0.009	0.092***	0.047	0.004	-0.002	0.007**	-0.004	0.000	939.840***
Depth Ratio (Levels 1-5)	-0.578***	0.074	0.019	0.082	0.095*	0.005	-0.001	0.008**	-0.003	0.001	181.279***
Depth Ratio (Levels 1-20)	-1.383***	0.575**	0.036	-0.165	0.459**	0.005	-0.001	0.007**	-0.003	0.000	127.238***
Depth Ratio (Levels 2-5)	-0.060	0.006	-0.057	0.006	-0.014	0.005	-0.001	0.008**	-0.003	0.001	11.056*
Depth Ratio (Levels 6-20)	-0.432***	0.267	-0.008	-0.276	0.189	0.005	-0.001	0.007**	-0.003	0.000	29.960***
Slope Ratio (Levels 1-2)	0.404***	0.024	0.033	-0.035	-0.038	0.005	-0.001	0.007**	-0.003	0.001	474.658***
Slope Ratio (Levels 1-5)	0.582***	-0.042	-0.081	-0.092	-0.096*	0.005	-0.001	0.008**	-0.003	0.001	225.031***
Slope Ratio (Levels 1-20)	1.893***	-0.278	-0.722***	0.098	-0.560***	0.004	-0.002	0.006*	-0.004	0.000	337.104***
Slope Ratio (Levels 2-5)	0.041	0.035	0.007	0.021	-0.032	0.005	-0.001	0.008**	-0.003	0.001	7.387
Slope Ratio (Levels 6-20)	0.821***	-0.105	-0.404**	0.282*	-0.351***	0.005	-0.001	0.007**	-0.003	0.000	113.145***

This table presents the parameter estimates of lagged limit order book variables ($\alpha_{z,\tau}$) and lagged duration ($\theta_{z,\tau}$) in the return equation and F-statistics for Granger Causality tests for limit order book variables. ***, **, * represent statistical significance at 1%, 5%, 10%, respectively.

Table A 2.10 Daily Coefficient Estimates on the First Lag of Limit Order Book Variables in the Return Equation

	Trading Days in July 2010																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Ask Side Variables																						
Depth (Levels 1-2)	-0.307***	-0.146	-0.492***	-0.307***	-0.339***	-0.202***	-0.284***	-0.173**	-0.286***	-0.177***	-0.206***	-0.227***	-0.100	-0.284***	-0.358***	-0.192*	-0.346***	-0.493***	-0.165***	-0.431***	-0.303***	-0.196***
Depth (Levels 1-5)	-0.522**	-0.031	-0.357*	-0.192	-0.218	-0.013	-0.092	-0.151	-0.023	0.028	-0.201*	-0.128	-0.097	-0.264	-0.404***	0.186	-0.313*	-0.441**	-0.257**	-0.990***	-0.396**	-0.202
Depth (Levels 1-20)	-1.704***	-0.018	-1.122	0.005	-2.135***	-0.179	-0.742	0.078	0.048	-0.687*	0.093	-0.430	-0.567	0.214	-1.789*	0.136	-1.295**	-2.101***	-0.459	-1.533***	-0.103	-0.385
Depth (Levels 2-5)	-0.113	0.132	0.020	0.067	0.066	0.142	0.174*	0.034	0.239**	0.163**	0.008	0.082	-0.097	0.100	-0.158	0.364*	-0.067	0.005	-0.025	-0.370**	0.078	0.046
Depth (Levels 6-20)	-0.965	0.166	-0.013	0.270	-1.095**	-0.268	-0.405	0.264	0.063	-0.669*	0.450	-0.093	0.202	0.647	-0.661	-0.202	-0.716	-1.523**	0.050	-0.215	0.423	-0.189
Slope (Levels 1-2)	0.267***	0.080	0.315***	0.088*	0.129**	0.134***	0.124***	0.060	0.119**	0.128***	0.077*	0.086	-0.027	0.196***	0.251***	0.075	0.241***	0.314***	0.054**	0.219***	0.158***	0.084**
Slope (Levels 1-5)	0.511***	0.224	0.305	0.086	0.140	0.053	-0.032	0.105	0.002	-0.039	0.169	0.129	0.158	0.278	0.382***	-0.168	0.424***	0.282	0.140	0.709***	0.249**	0.158
Slope (Levels 1-20)	-0.081	0.829**	0.406	-0.228	0.224	0.512	-0.631	0.098	1.174**	0.724*	-0.489	0.956**	0.838*	0.593	1.249*	0.413	1.973***	1.130**	0.365	1.409***	0.672*	0.968***
Slope (Levels 2-5)	0.094	-0.038	-0.048	-0.056	-0.020	-0.068	-0.185**	0.003	-0.210*	-0.153**	0.001	0.019	0.193	0.046	0.115	-0.283	0.120	-0.135	0.072	0.322*	-0.061	0.017
Slope (Levels 6-20)	-0.357	0.367	-0.225	-0.249	-0.008	0.311	-0.348	-0.064	0.708*	0.592*	-0.622*	0.462	0.357	0.179	0.364	0.608**	1.000*	0.897**	0.096	0.328	0.184	0.750**
Bid Side Variables																						
Depth (Levels 1-2)	0.279***	0.169*	0.240**	0.386***	0.337***	0.309***	0.281***	0.248***	0.272***	0.156***	0.268***	0.200***	0.371***	0.376***	0.308***	0.334***	0.345***	0.268***	0.343***	0.136	0.388***	0.356***
Depth (Levels 1-5)	0.449**	-0.094	0.118	0.455***	-0.020	0.191	0.177	0.327**	0.335**	0.153	0.315***	-0.075	0.213	0.465**	0.061	0.467**	0.097	0.110	0.508***	0.256	0.534***	0.349***
Depth (Levels 1-20)	-0.344	0.366	1.087	0.791	-0.302	0.046	0.917**	1.048**	0.746	0.609	1.017***	1.211**	1.631*	1.874***	1.029	1.260*	1.198	1.279**	0.860**	-0.532	0.426	0.298
Depth (Levels 2-5)	0.116	-0.193	-0.057	0.119	-0.293	-0.163	-0.009	0.109	0.163	-0.024	-0.047	-0.297***	-0.060	0.096	-0.165	0.156	-0.228	-0.150	0.220***	0.058	0.038	0.057
Depth (Levels 6-20)	-1.057	0.327	0.692	-0.347	-0.230	-0.343	0.375	0.338	0.110	0.200	0.644**	1.059***	1.238*	1.219**	1.261***	0.205	0.904	0.928*	0.033	-0.659	0.094	0.059
Slope (Levels 1-2)	-0.112*	-0.119**	-0.174**	-0.226***	-0.199***	-0.144***	-0.166***	-0.112**	-0.185***	-0.068*	-0.043	-0.050	-0.132	-0.106*	-0.099**	-0.129*	-0.132**	-0.277***	-0.158***	-0.103*	-0.258***	-0.192***
Slope (Levels 1-5)	-0.384**	-0.063	-0.080	-0.376***	0.045	-0.183	-0.162	-0.286**	-0.327**	-0.201*	-0.069	0.189*	-0.255	-0.278	-0.298	-0.261	-0.095	-0.445**	-0.395***	-0.538***	-0.474***	-0.386***
Slope (Levels 1-20)	-0.906	-0.739	-0.624	0.023	-0.986*	-1.117**	-0.667**	-0.559	-0.435	-0.436	-0.942***	-1.473***	-1.189*	-1.164*	-1.753***	-0.989***	-1.057	-0.728*	-0.627**	-0.355	-0.751**	-0.865***
Slope (Levels 2-5)	-0.083	0.048	0.033	-0.060	0.281	0.128	0.004	-0.103	-0.165	0.020	-0.055	0.320***	-0.040	-0.169	-0.051	-0.081	0.254*	0.098	-0.204***	-0.301**	-0.123	-0.113
Slope (Levels 6-20)	-0.349	-0.478	-0.449	0.384	-0.616	-0.445	-0.270	-0.026	-0.031	0.020	-0.721***	-1.191***	-0.423	-0.913*	-1.148***	-0.414	-0.797	-0.122	-0.193	0.211	-0.200	-0.455*
Ratio Variables																						
Depth Ratio (Levels 1-2)	-0.806***	-0.451***	-0.937***	-0.917***	-0.880***	-0.765***	-0.798***	-0.513***	-0.738***	-0.487***	-0.637***	-0.603***	-0.546***	-0.869***	-0.910***	-0.780***	-0.942***	-1.127***	-0.645***	-0.822***	-0.939***	-0.699***
Depth Ratio (Levels 1-5)	-1.034***	0.070	-0.672*	-0.762***	-0.219	-0.194	-0.336	-0.554**	-0.412	-0.146	-0.576***	-0.035	-0.268	-0.778***	-0.617**	-0.500	-0.503*	-0.773**	-1.061***	-1.396***	-1.060***	-0.650***
Depth Ratio (Levels 1-20)	-1.335	-0.457	-2.732**	-0.701	-2.093**	-0.402	-1.880**	-1.094	-0.787	-1.518**	-1.285**	-2.080***	-2.398**	-2.100**	-3.484**	-1.077	-2.576**	-3.529***	-1.502***	-0.744	-0.780	-0.752
Depth Ratio (Levels 2-5)	-0.483*	0.195	-0.329	-0.151	0.313	0.198	0.100	-0.167	-0.051	0.155	0.054	0.489**	0.239	-0.191	-0.092	0.281	0.005	-0.215	-0.413***	-0.640**	-0.183	-0.134
Depth Ratio (Levels 6-20)	0.440	-0.211	-0.717	0.633	-1.027	-0.067	-0.859	-0.054	-0.029	-1.079**	-0.345	-1.562***	-0.852	-0.533	-2.121*	-0.423	-1.698*	-2.152**	0.013	0.612	0.100	-0.243
Slope Ratio (Levels 1-2)	0.562***	0.312**	0.688***	0.439***	0.392***	0.475***	0.421***	0.224**	0.433***	0.280***	0.152*	0.185	0.190	0.395***	0.446***	0.277**	0.509***	0.783***	0.263***	0.503***	0.565***	0.399***
Slope Ratio (Levels 1-5)	0.948***	0.372	0.646*	0.533**	0.101	0.278	0.133	0.459**	0.348	0.150	0.311	-0.204	0.510	0.732***	0.876***	0.166	0.652**	1.080***	0.756***	1.539***	0.894***	0.652***
Slope Ratio (Levels 1-20)	1.039	2.073***	1.571	-0.118	1.460	2.415***	0.090	0.767	2.143***	1.426**	1.025*	2.990***	2.275**	2.138***	3.483***	2.016***	3.656***	1.962***	1.180**	1.852**	1.610***	2.023***
Slope Ratio (Levels 2-5)	0.144	-0.102	0.016	0.014	-0.316	-0.254	-0.245	0.118	-0.103	-0.236*	-0.018	-0.455**	0.241	0.157	0.223	-0.209	-0.106	-0.089	0.381***	0.800***	0.135	0.110
Slope Ratio (Levels 6-20)	0.141	0.990	0.104	-0.650	0.688	1.077**	-0.054	-0.035	1.080*	0.729	0.383	2.086***	0.924	1.211*	1.613	1.341**	2.064**	0.911	0.317	0.130	0.451	1.324***

This table presents the parameter estimates on the first lag of limit order book variables in the return equation ($\alpha_{z,\tau}$) in each trading day in July 2010. ***, **, * represent statistical significance at 1%, 5%, 10%, respectively.