# LIMIT-ORDER COMPLETION TIME IN THE LONDON STOCK MARKET

Quan Wen

Submitted for the degree of Doctor of Philosophy

Heriot-Watt University

School of Management and Language

April 2009

# ABSTRACT

This study develops an econometric model of limit-order completion time using survival analysis. Time-to-completion for both buy and sell limit orders is estimated using tick-by-tick UK order data. The study investigates the explanatory power of variables that measure order characteristics and market conditions, such as the limit-order price, limit-order size, best bid-offer spread, and market volatility. The generic results show that limit-order completion time depends on some variables more than on others. This study also provides an investigation of how the dynamics of the market are incorporated into models of limit-order completion. The empirical results show that time-varying variables capture the state of an order book in a better way than static ones. Moreover, this study provides an examination of the prediction accuracy of the proposed models. In addition, this study provides an investigation of the intra-day pattern of order submission and time-of-day effects on limit-order completion time in the UK market. It provides evidence showing that limit orders placed in the afternoon period are expected to have the shortest completion times while orders placed in the mid-day period are expected to have the longest completion times, and the sensitivities of limit-order completion time to the explanatory variables vary over the trading day.

# DEDICATION

To my wife Zhaowei Xu

# ACKNOWLEDGEMENTS

I gratefully acknowledge the advice and support of my supervisor, Dr. Boulis Ibrahim. His teaching and guidance are ultimately responsible for this dissertation. Special thanks go to Professor Paul Draper and Dr. Janusz Brzeszczynski who provided helpful comments and suggestions that further clarified and sharpened the focus of this research. A special note of acknowledgement is due to my family for their love and understanding throughout this journey.

**DECLARATION STATEMENT**

ACADEMIC REGISTRY

**Research Thesis Submission**

| | | | |
|---|---|---|---|
| Name*:* | Quan Wen | | |
| School/PGI: | School of Management and Language | | |
| Version: *(i.e. First, Resubmission, Final)* | Final | Degree Sought (Award **and** Subject area) | PhD in Accounting and Finance |

## Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

| | | | |
|---|---|---|---|
| Signature of Candidate*:* | | Date: | |

## Submission

| | |
|---|---|
| Submitted By *(name in capitals):* | Quan Wen |
| Signature of Individual Submitting: | |
| Date Submitted: | |

## For Completion in Academic Registry

| | | | |
|---|---|---|---|
| Received in the Academic Registry by *(name in capitals):* | | | |
| *Method of Submission* *(Handed in to Academic Registry; posted through internal/external mail):* | | | |
| *E-thesis Submitted (mandatory from January 2009)* | | | |
| Signature: | | Date: | |

# TABLE OF CONTENTS

**LIST OF TABLES** .......... X

**LIST OF FIGURES** .......... XII

**CHAPTER 1 – INTRODUCTION, BACKGROUD, OBJECTIVES AND CONTRIBUTIONS OF THIS STUDY** .......... 1

**1.1 Preliminary Definitions** .......... 1

**1.2 Why Investigate Limit-Order Completion Time?** .......... 3

**1.3 What Has Been Studied?** .......... 5

**1.4 Unique Aspects of this Study** .......... 7

**1.5 Intended Methodology** .......... 7

**1.6 Objectives of this Study** .......... 8
*1.6.1 Organisation of this Study* .......... 9
*1.6.2 Specific Objectives of Chapter 5* .......... 9
*1.6.3 Specific Objectives of Chapter 6* .......... 10
*1.6.4 Specific Objectives of Chapter 7* .......... 10
*1.6.5 Specific Objectives of Chapter 8* .......... 10
*1.6.6 Summary* .......... 11

**CHAPTER 2 – TRADING MECHANISMS OF THE LONDON STOCK EXCHANGE** .......... 12

**2.1 Introduction to Trading Mechanisms** .......... 12
*2.1.1 Quote-Driven Markets* .......... 13
*2.1.2 Order-Driven Markets* .......... 15
*2.1.3 Brokered Markets* .......... 16
*2.1.4 Hybrid Markets* .......... 16
*2.1.5 Summary* .......... 19

**2.2 Trading Services of the London Stock Exchange** .......... 19

**2.3 A SETS Trading Day** .......... 21

**2.4 SETS Orders and Order Books** .......... 29

**2.5 SETS Segments and Sectors** .......... 35

**2.6 SETS Trade and Transaction Reports** .......... 38

**2.7 Summary and Conclusions** .......... 40

**CHAPTER 3 – LITERATURE REVIEW ON MICROSTRUCTURE . 43**

**3.1 Introduction ........ 43**

**3.2 Limit-Order Books ........ 45**
*3.2.1 Information Content of a Limit-Order Book* ........ 45
*3.2.2 Statistical Properties of a Limit-Order Book* ........ 47
*3.2.3 Summary* ........ 49

**3.3 Order Submission Strategies ........ 49**
*3.3.1 Order Submission Strategies* ........ 50
*3.3.2 Empirical Studies on Order Submission Strategies* ........ 52
*3.3.3 Summary* ........ 56

**3.4 Limit-Order Execution ........ 56**
*3.4.1 Theoretical Studies on Limit-Order Execution* ........ 57
*3.4.2 Empirical Studies on Limit-Order Execution* ........ 57
*3.4.3 Summary* ........ 59

**3.5 Summary and Conclusions ........ 59**

**CHAPTER 4 – DATA AND METHODOLOGY ........ 62**

**4.1 Data ........ 62**
*4.1.1 Database* ........ 62
*4.1.2 Sample Selection* ........ 65
*4.1.3 Summary Statistics of the Data Sample* ........ 67
*4.1.4 Summary* ........ 71

**4.2 Methodology ........ 73**
*4.2.1 Basic Concepts of Survival Analysis* ........ 74
*4.2.2 Censoring* ........ 76
*4.2.3 Non-Parametric Estimation of the Survival Function* ........ 77
*4.2.4 Parametric Estimation of the Survival Function* ........ 79
*4.2.5 Incorporating Explanatory Variables* ........ 80
*4.2.6 Assessing the Goodness-of-Fit* ........ 85
*4.2.7 Summary* ........ 86

**CHAPTER 5 – LIMIT-ORDER COMPLETION TIME IN THE UK MARKET: AN INVESTIGATION USING LO, MACKINLAY AND ZHANG MODEL ........ 88**

**5.1 Introduction ........ 88**

**5.2 LMZ Explanatory Variables ........ 89**

**5.3 Empirical Results ........ 102**
*5.3.1 Estimation Results* ........ 103
*5.3.2 Multicollinearity Effects* ........ 111

5.4 **Summary and Conclusions** .......... 120

**CHAPTER 6 – LIMIT-ORDER COMPLETION TIME IN THE UK MARKET: AN INVESTIGATION USING NEWLY CONSTRUCTED VARIABLES** .......... 134

6.1 **Introduction** .......... 134

6.2 **Constructed Variables** .......... 135
*6.2.1 Explanatory Variables* .......... 135
*6.2.2 Summary Statistics of Explanatory Variables* .......... 143

6.3 **Model Specifications and Empirical Estimations** .......... 147
*6.3.1 Empirical Estimations of the GG AFT Model* .......... 147
*6.3.2 Empirical Estimations of the Cox PH Model* .......... 156

6.4 **Summary and Conclusions** .......... 161

**CHAPTER 7 – THE EFFECT OF TIME-VARYING DYNAMICS OF THE ORDER BOOK ON LIMIT-ORDER COMPLETION TIME IN THE UK MARKET** .......... 169

7.1 **Introduction** .......... 169

7.2 **Splitting Survival Times** .......... 170

7.3 **Empirical Estimation of the Dynamic GG AFT Model** .......... 174

7.4 **Empirical Estimation of the Modified Dynamic GG AFT Model** .......... 177

7.5 **Prediction Accuracy of the Modified Dynamic GG AFT Model** .......... 181

7.6 **Summary and Conclusions** .......... 183

**CHAPTER 8 – INTRA-DAILY SEASONALITY IN THE LIMIT-ORDER COMPLETION TIME IN THE UK MARKET** .......... 191

8.1 **Introduction** .......... 191

8.2 **Time-of-Day Patterns in Limit-Order Completion Times** .......... 192

8.3 **Summary and Conclusions** .......... 204

**CHAPTER 9 – MAIN FINDINGS, STRENGTHS, LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH** .......... 212

9.1 **Introduction** .......... 212

**9.2 Main Findings** .......... **212**
*9.2.1 Results of Chapter 5* .......... 212
*9.2.2 Results of Chapter 6* .......... 214
*9.2.3 Results of Chapter 7* .......... 215
*9.2.4 Results of Chapter 8* .......... 216
*9.2.5 Overall Conclusions* .......... 217

**9.3 Strength of this Study** .......... **218**

**9.4 Limitations of this Study** .......... **219**

**9.5 Suggestions for Future Research** .......... **219**

**APPENDIX A** .......... ERROR! BOOKMARK NOT DEFINED.

**APPENDIX B** .......... ERROR! BOOKMARK NOT DEFINED.

**APPENDIX C** .......... ERROR! BOOKMARK NOT DEFINED.

**APPENDIX D** .......... ERROR! BOOKMARK NOT DEFINED.

**APPENDIX E** .......... ERROR! BOOKMARK NOT DEFINED.

**APPENDIX F** .......... ERROR! BOOKMARK NOT DEFINED.

**APPENDIX G** .......... ERROR! BOOKMARK NOT DEFINED.

**APPENDIX H** .......... ERROR! BOOKMARK NOT DEFINED.

**REFERENCES** .......... ERROR! BOOKMARK NOT DEFINED.

# LIST OF TABLES

Table 2.1: Opening Auction, Continuous Trading and Closing Auction........................28

Table 2.2: Summary of Opening Auction, Closing Auction and AESP .........................29

Table 2.3: Suggestions of When to Use Each Type of Order ........................................32

Table 2.4: Summary of Order Execution Options .........................................................32

Table 2.5: Summary of Order Entry Requirements .......................................................33

Table 2.6: Sectors and Segments of SETS.....................................................................37

Table 2.7: Default Period Schedule for SETS Containing Sectors FT10, FS10, FE10, FF10, FT25, FS25 and STOS.....................................................................37

Table 2.8: Default Period Schedule for SETS Containing Sectors SS10, SS25 and STSS ..........................................................................................................38

Table 2.9: Default Period Schedule for SETS Containing Sector SECN .......................38

Table 4.1: Summary Statistics of the Data Sample........................................................68

Table 4.2: Percentage Breakdowns of Completed Orders ..............................................69

Table 4.3: Summary Statistics of Completed and Censored Limit Orders.....................71

Table 4.4: An Example of Kaplan-Meier Estimator of Survival Function.....................79

Table 4.5: A Limit-Order Dataset ..................................................................................83

Table 5.1: Summary Statistics of the LMZ Explanatory Variables..............................123

Table 5.2: Correlation Matrices of the LMZ Explanatory Variables............................125

Table 5.3: Parameter Estimates of the LMZ Models ...................................................126

Table 5.4: Parameter Estimates of the LMZ Models Excluding MKD1X ...................128

Table 5.5: Parameter Estimates of the LMZ Models Excluding SZSD........................130

Table 6.1: Summary Statistics of the Explanatory Variables ......................................163

Table 6.2: Correlation Matrices of the Explanatory Variables .....................................164

Table 6.3: Parameter Estimates of the GG AFT Model...............................................165

Table 6.4: Parameter Estimates of the Cox PH Model ................................................166

Table 7.1: Single-Record Survival Data ........................................................ 172

Table 7.2: Multiple-Record Survival Data ...................................................... 172

Table 7.3: Parameter Estimates of the Dynamic GG AFT Model ................................ 185

Table 7.4: Parameter Estimates of the Modified Dynamic GG AFT Model ................ 186

Table 7.5: Prediction Errors of the Modified Dynamic GG AFT Model...................... 187

Table 8.1: Percentage Breakdown of Limit Orders over a Trading Day ...................... 206

Table 8.2: Summary Statistics of the Explanatory Variables (Three Intra-Daily Periods) ........................................................................................................ 207

Table 8.3: Parameter Estimates of the Modified Dynamic GG AFT Model (Time-of-Day)............................................................................................................ 208

Table 8.4: Parameter Estimates of the Modified Dynamic GG AFT Model (Three Intra-Daily Periods) ......................................................................................... 209

Table 8.5: Statistical Test of the Differences among Estimated Coefficients (Three Intra-Daily Periods) ......................................................................................... 211

# LIST OF FIGURES

Figure 2.1: A SETS Trading Day........................................................................21

Figure 2.2: Opening Auction ............................................................................22

Figure 2.3: Summary of Auction Price Determination for a Crossed Order Book.........23

Figure 2.4: AESP Auction ................................................................................25

Figure 2.5: Closing Auction..............................................................................25

Figure 2.6: The Closing Price Determination for SETS Securities ................................26

Figure 2.7: EDSP Intra-Day Auctions for SETS Securities...........................................27

Figure 2.8: A Typical SETS Order Book.........................................................34

Figure 2.9: Market Segments and Sectors .......................................................36

Figure 5.1: Q-Q Plots of the LMZ Models ......................................................132

Figure 5.2: Q-Q Plots of the LMZ Models Excluding Observations with Cox-Snell Residuals Greater Than 10.........................................................................133

Figure 6.1: Q-Q Plots of the GG AFT Model...................................................167

Figure 6.2: Q-Q Plots of the Cox PH Model....................................................168

Figure 7.1: Q-Q Plots of the Dynamic GG AFT Model ......................................188

Figure 7.2: Q-Q Plots of the Modified Dynamic GG AFT Model...............................189

Figure 7.3: Q-Q Plots of the Modified Dynamic GG AFT Model Excluding Observations with Cox-Snell Residuals Greater Than Five ................................190

# CHAPTER 1 – INTRODUCTION, BACKGROUD, OBJECTIVES AND CONTRIBUTIONS OF THIS STUDY

## 1.1 Preliminary Definitions

This section briefly introduces and defines some terms used in this study. A more detailed coverage is given in subsequent chapters. The remaining sections of this chapter describe the background, objectives and contributions of this study.

Trading is organised differently around world markets and across financial assets. For example traders at the Tokyo Stock Exchange (TSE) submit their orders for execution through a continuous auction process. This mechanism is known as an order-driven market.[1] Usually, an order-driven market features an electronic order book where all unexecuted orders stay for subsequent trading.

The order book is a collection of unexecuted orders arranged in descending price for buy orders and in ascending price for sell orders. These orders are executed with newly arrived orders according to the order precedence rules. The basic rule is price priority. Orders offering the highest (lowest) buying (selling) prices have priority. Another rule is usually time priority. At a given price level, existing orders are executed first-in-first-out with newly arrived orders.

Orders arrive at the order book randomly over time and are compared to existing ones. Executions will only occur when the price is matched. Unexecuted orders will remain on the order book until completion, cancellation or expiration.

The unexecuted buy (sell) order on the order book offering the highest (lowest) price is known as the best bid (offer) order, whose price and size are known as the best bid (offer) price and size. The difference between the best bid and offer prices is known as

[1] A more detailed coverage of an order-driven market is given in Chapter 2.

the best bid-offer spread or market spread. The best bid-offer spread is one important aspect of the state of the order book. A more detailed coverage is given in subsequent chapters.

Since most order-driven markets, such as the TSE, operate continuously, orders arrive at the order book continuously. Hence the state of the order book, such as the best bid-offer spread and the number of orders on the book, also changes continuously. This is the time-varying and dynamic characteristic of an order book.[2] Different types of orders arrive at the dynamic order book for subsequent trading.

The most frequently used types of orders are market and limit orders. The essential difference between a market and a limit order is the price at which the order is submitted. A market order is an order submitted to the trading system with a specified size only but not a price. It immediately fills at the best bid/offer price and may be allowed to execute with the next available liquidity when the liquidity at the best bid/offer price is lacking.[3] A limit order is an order with specified size and price. This usually offers the trader a better execution price than a market order. However, there are some costs inherent in submitting a limit order: it may fail to be completed or may take time to be completed. A limit order can also be adversely picked up if the stock's value moves past the limit-order price before this order can be cancelled. For example, suppose a trader submits a buy limit order at 100 pence. Shortly after the submission, the market price increases to 120 pence. Then this buy limit order is 'picked up' by the seller and is executed at 100 pence. Market and limit orders are widely used by traders to buy or sell stocks in stock markets around the world.

Traders usually need to decide whether to use market or limit orders. They must also decide when to submit market orders and when to submit limit orders before they trade. If they submit limit orders, they must know at what price levels to place their limit-order prices, and when to modify their orders. These decisions are known as order

[2] A more detailed coverage is given in Chapter 7.
[3] This varies across exchanges. For example, Hamao and Hasbrouck (1995) explain that at the TSE if a buy (sell) market order consumes all liquidity at the best offer (bid) price, a 'warning quote' at that price will be temporarily issued to enable participants to offer shares at that price. Biais, Hillion and Spatt (1995) explain that at the Paris Stock Exchange any excess that cannot be executed at the best bid/offer price is converted into a limit order.

submission strategies. A more detailed coverage of these strategies is given in Chapter 3.

This section briefly introduces some terms, such as an order-driven market, a market order and a limit order, which are used in subsequent chapters. The presentation in this chapter proceeds as follows. The next section describes the importance of the limit-order completion time to traders, exchanges and regulators. Section 1.3 briefly reviews existing studies on limit-order execution time and probability. Section 1.4 describes unique aspects of this study. Section 1.5 briefly describes the methodology used in this study. Section 1.6 describes the objectives of this study.

## 1.2 Why Investigate Limit-Order Completion Time?

This section describes the importance of the limit-order completion time to traders, exchanges and regulators, and the reasons to study this topic. First, limit orders are potentially executed at better prices than market orders. Limit orders, however, run the risk of non-execution and take time to be completed. For some traders, the opportunity cost of waiting for a limit order to be completed and the probability of failing to be completed can be significant, as the market may move against them. In this case, these traders may have to accept far worse prices than they had initially hoped for. For example, a trader submits a limit order to buy 1000 BP shares at 600 pence per share and just after the submission the BP share price increases to 610 pence per share. If this trader still wants to buy these 1000 BP shares, he needs to buy them at 610 pence per share, which is far higher than he had initially hoped for. Hence it is important to investigate and understand limit-order completion time in order to minimise this waiting time cost.

Second, many stock exchanges are organised as order-driven markets or at least have some order-driven market functions.[4] These exchanges are facing competitive pressures to make their trading systems more attractive to their traders. Hence they are trying to find out how best to implement limit-order trading in terms of minimising the waiting

[4] Domowitz (1993) documents that approximately 35 financial markets in 16 different countries contain elements of order-driven markets.

time. In addition, limit-order completion time receives much more attention from stock market regulators. For example, the US Securities and Exchange Commission (SEC) proposed in 2001 to consider limit-order completion time as one of the parameters that indicates market quality and to require broker firms to publish statistics on completion times. Thus, investigating and understanding limit-order completion times may result in some benefits for stock exchanges and some policy implications for regulators.

Third, a market is liquid if traders can quickly buy or sell a large number of shares. Four measures are often associated with liquidity in the market microstructure literature: width, depth, immediacy and resilience. According to Amihud and Mendelson (1986) and Harris (1990), width refers to the bid-offer spread for a given number of shares; depth refers to the number of shares that can be traded at given prices; immediacy refers to how quickly trades of a given size can be done at a given cost; and resilience refers to how quickly prices revert to former levels after they change in response to large order flows. Since limit orders provide immediacy, liquidity can be measured by limit-order completion time. Hence, investigating limit-order completion time can increase our understanding of the nature and behaviour of the liquidity of a stock market.

Fourth, from a functional perspective, one of the main economic roles of a stock market is to find prices for the listed stocks. The price-discovery process in an order-driven stock market is dynamic and complex. In such a market, a limit order executes with future orders and competes against existing limit orders. Orders are executed asynchronously so there is no unique market-clearing price. Instead, there is a sequence of transaction prices at which matched orders are executed over time. As a result, investigating limit-order completion time can increase our understanding of the price-discovery process in an order-driven market.

Fifth, Tkatch and Kandel (2006) show that traders base their order submission strategies on the expected order completion times. Whether or not traders can optimise their order submission strategies ultimately depends on the availability of models that provide real-time pre-trade estimates of expected order completion times. The development of such models is still in its infancy. Empirical studies on limit-order completion times might be a step toward developing such models.

Finally, Copeland and Galai (1983) and Stoll (1992) point out that limit orders give options to other traders to trade at the limit-order prices. The trader who places a buy (sell) limit order has written a free put (call) option to market participants. For example, suppose a trader submits a buy limit order at 100 pence. If the stock price falls below 100 pence, this put option will be exercised. For the market participants, the option value of a limit order depends on its expected completion time. Hence understanding limit-order completion time is a step further toward understanding the underlying option value of a limit order.

This section summarises the reasons to investigate limit-order completion time. Investigating and understanding limit-order completion time may lead to reducing the waiting time cost, it may result in some benefits for stock exchanges and some policy implications for regulators, it can also increase our understanding of the nature and behaviour of the liquidity and the price-discovery process of a stock market, and it might also be a step toward developing models which provide real-time pre-trade estimates of expected order completion times and toward understanding the underlying option value of a limit order. The next section will briefly review existing studies on limit-order execution time and probability.

## 1.3 What Has Been Studied?

The recent availability of transaction data from stock markets around the world has stimulated research on limit-order execution time and probability.[5] In the literature there are two main methods of modelling limit-order execution time and probability. One approach is a probit model. For example, Omura, Tanigawa and Jun (2000) reconstruct order flows at the TSE and use a probit model to estimate the probability of a limit order being executed entirely within a day. Another approach is survival analysis, which is a statistical technique for analysing non-negative variables.[6] For example, Cho and Nelling (2000) use survival analysis to investigate the execution time of a limit order at the NYSE. They show that the longer a limit order remains on the order book, the lower the execution probability. They also present evidence showing

[5] A more detailed coverage of market microstructure literature is given in Chapter 3.
[6] An introduction to survival analysis is given in Section 4.2 of Chapter 4.

that the best bid-offer spread, price volatility, limit-order price and size are all related to limit-order execution probability. They estimate, however, a single model for both buy and sell orders and do not provide any goodness-of-fit tests.

In addition to Cho and Nelling (2000), Lo, MacKinlay and Zhang (LMZ) (2002) model limit-order execution time using survival analysis and US data from an institutional broker firm.[7] They find that the execution time is very sensitive to some explanatory variables but not to others. Their data, however, is from a single order routing source and is not comprehensive. For this reason their findings may reflect the characteristics of their data source rather than the NYSE. By comparison, this study uses the entire universe of limit orders at the London Stock Exchange (LSE). As a result, this study is more comprehensive.

In the literature there are some applications of the LMZ models without prior testing of these models to the data or market being studied. Tkatch and Kandel (2006), for example, use the LMZ models to estimate expected order execution times at the Tel Aviv Stock Exchange. Unlike Tkatch and Kandel (2006), this study investigates the LMZ models on UK market data and assesses their applicability.

Recent studies in market microstructure of the UK market have focused on the bid-offer spread. For example, Menyaha and Paudyal (2000) estimate the components of the bid-offer spread at the LSE using intra-day data, Taylor (2002) investigates econometric forecasts of the bid-offer spread, and Levin and Wright (2004) investigate the components of the bid-offer spread.[8] This study is the first attempt to investigate limit-order completion time as one aspect of market microstructure of the UK market. The next section will describe unique aspects of this study.

[7] LMZ refers to Lo, MacKinlay and Zhang in the remainder of this thesis.
[8] A more detailed coverage of the components of the bid-offer spread is given in Section 6.2 of Chapter 6.

## 1.4 Unique Aspects of this Study

This study has several unique aspects. First, investigation of limit-order completion times on data from stock exchanges outside the US is extremely limited. No such attempt has been made on UK data, and to our knowledge this analysis is the only attempt to date that studies limit-order completion time in the UK market. This study provides evidence on several issues related to the interaction between the order book and limit-order completion time, which adds to the existing empirical literature on limit-order execution time and probability. Second, this study investigates the degree to which explanatory variables suggested by LMZ capture the state of the UK order book and affect limit-order completion time. This has not been carried out by previous research. Third, it is believed that this study is the first attempt to construct explanatory variables that capture the state of the UK order book and to investigate their effects on limit-order completion time. Such an investigation may highlight characteristic properties of limit-order completion time in the UK market that might be important to UK traders, the stock exchange and regulators. Fourth, it is believed that this study is the first to incorporate the time-varying dynamics of the order book into models of limit-order completion. Finally, this study examines the effects of the time of order submission (time-of-day) on the completion time. The next section will briefly describe the methodology used in this study.

## 1.5 Intended Methodology

This study uses survival analysis to investigate limit-order completion time in the UK market. Survival analysis, also known as event history analysis, is a statistical technique for analysing non-negative random variables. Typically, the random variable represents the time to an event (known as survival time), e.g. the time to the failure of a physical component. In this study, the survival time refers to the time taken by a limit order to execute following submission, i.e., limit-order completion time. Data containing survival times is known as survival data.

In survival analysis, usually, some observations are censored. Censoring occurs when events are not observed and hence the survival times are unknown. These censored observations, however, still provide some valuable information, since they 'survived' for a certain period of time. In this study, censored observations refer to cancelled or expired limit orders. Accommodating the censored observations is the main advantage of survival analysis. A more detailed coverage of survival analysis is given in Chapter 4. The next section will describe objectives of this study.

## 1.6 Objectives of this Study

This study is motivated by the importance of limit-order completion time for traders, exchanges and regulators. The intention is to investigate limit-order completion time in the UK market. In other words, to take into account the characteristics of limit orders and the UK order book when modelling limit-order completion time. This study is also informative at assessing the impact of these characteristics on limit-order completion time. Finally, in order to check the prediction accuracy of the proposed model, the predicted limit-order completion time is compared to the actual one.

In particular, this study considers the following research questions:

- How do explanatory variables suggested by LMZ capture the state of the UK order book and affect limit-order completion time? This question remains unanswered in previous research.
- What are the determinants of limit-order completion time? Are these determinants rationally justified? Which factors is the completion time more sensitive to? These questions have been given some attention but largely remain unexplored by the literature, especially for the UK market.
- How can the dynamics of an order book be incorporated into models of limit-order completion? These questions have not been investigated by previous research.
- What are the time-of-day effects on limit-order completion time? This question has been given very little attention in the literature, especially for the UK market.

The presentation in this section proceeds as follows. The next subsection describes the organisation of this study. Subsections 1.6.2, 1.6.3, 1.6.4 and 1.6.5 describe the specific objectives of empirical chapters in this study. Subsection 1.6.6 draws a short summary.

### *1.6.1 Organisation of this Study*

Chapter 2 describes trading mechanisms of stock exchanges, with particular emphasis on those of the LSE. Chapter 3 presents a review of related literature. The review of literature is not restricted to Chapter 3 and subsequent empirical chapters contain reviews of specific background and motivational literature. Chapter 4 describes the data sample and methodology used in this study. Chapter 5 is the first empirical chapter; it provides an investigation of the degree to which explanatory variables suggested by LMZ capture the state of the UK order book and affect limit-order completion time. Chapter 6 is the second empirical chapter. It presents the definition of explanatory variables for the UK market. It also provides an investigation into their effects on limit-order completion time. Chapter 7 is the third empirical chapter. It provides an investigation of how to incorporate the dynamics of the order book into models of limit-order completion. Chapter 8 is the fourth empirical chapter. It provides an investigation of time-of-day effects on limit-order completion time in the UK market. Finally, Chapter 9 provides general conclusions, a discussion of strengths and limitations of this study, and suggestions for future research.

### *1.6.2 Specific Objectives of Chapter 5*

As mentioned above, LMZ (2002) model limit-order execution time using survival analysis and data from an institutional broker firm. They find that the execution time is very sensitive to some explanatory variables but not to others. Their models have not been investigated using UK data. Chapter 5 provides an investigation of the degree to which explanatory variables suggested by LMZ capture the state of the UK order book and affect limit-order completion time. It also provides an investigation of a possible multicollinearity effect on parameter estimates of LMZ models.

### 1.6.3 Specific Objectives of Chapter 6

Buy and sell orders may be used with varying degrees of intentions. For example, Keim and Madhavan (1995) show that typically traders are more passive with buy orders to hide their information, and more aggressive with sell orders, perhaps due to greater urgency when selling. Accordingly, in this study separate models are estimated in order to capture possible asymmetries in the effects of the explanatory variables on buy-and-sell-order completion time. Chapter 6 provides an estimation of time-to-completion with explanatory variables - which have lower correlation with each other but still capture the characteristics of the UK order book and limit orders - for the UK market. It also provides an investigation of how these explanatory variables affect limit-order completion time.

### 1.6.4 Specific Objectives of Chapter 7

All proposed models of limit-order execution in the literature mentioned above capture the state of an order book at the time of order submission only. They are ‘snapshot’ or ‘static’ models. However, the state of an order book (e.g. the best bid-offer spread) changes frequently. Thus, ‘static’ models may fail to capture the changing dynamics of an order book. Chapter 7 provides an investigation of how these dynamics are incorporated into models of limit-order completion. This chapter also provides an examination of the prediction accuracy of the proposed models.

### 1.6.5 Specific Objectives of Chapter 8

Researchers have studied intra-day patterns with transaction data from stock exchanges around the world and reveal the persistent u-shape patterns in returns, number of shares traded, volume, bid-offer spreads, and volatility. The intra-day patterns in order placements also draw the attention of research. Biais, Hillion and Spatt (1995) find that the placement of new orders in the Paris Stock Exchange tends to be concentrated in the morning while cancellations and large trades tend to occur late in the trading day. Niemeyer and Sandas (1995) analyse data from the Stockholm Stock Exchange and find a u-shape pattern in the number of limit orders placed, with a surprisingly high frequency of limit orders placed during the last minutes of the trading day. Al-Suhaibani and Kryzanowsky (2000) find that the number of limit orders in the Saudi stock market also exhibits a u-shape pattern. Chapter 8 provides an investigation of the intra-day pattern of order submission in the UK market. It also provides an examination of time-of-day effects on limit-order completion time.

### *1.6.6 Summary*

This study has four empirical chapters. The empirical analysis proceeds by investigating, in Chapter 5, the degree to which explanatory variables suggested by LMZ capture the state of the UK order book and affect limit-order completion time. Particular emphasis is given to investigating the possible multicollinearity effect on parameter estimates of LMZ models. In Chapter 6, explanatory variables, which have lower correlation with each other but still capture the characteristics of the UK order book and limit orders, are constructed. In this chapter the way in which these explanatory variables affect limit-order completion time is also investigated. The investigation of how to incorporate the changing dynamics of an order book into models of limit-order completion is presented in Chapter 7. Finally, Chapter 8 provides an investigation of the intra-day pattern of order submission in the UK market. Particular emphasis is given to investigating the time-of-day effects on limit-order completion time. The next chapter will describe trading mechanisms of stock exchanges, with particular emphasis on those of the LSE.

# CHAPTER 2 – TRADING MECHANISMS OF THE LONDON STOCK EXCHANGE

This chapter provides detailed background information of trading mechanisms of stock exchanges, with particular emphasis on those of the LSE. The next section introduces the main trading mechanisms used in financial markets. The remaining sections of this chapter introduce trading mechanisms of the LSE, with particular emphasis on the Stock Exchange Electronic Trading Service.

## 2.1 Introduction to Trading Mechanisms

Trading is a searching process where buyers are searching for sellers and sellers for buyers. Traders obviously always prefer to trade at a good price. Hence buyers (sellers) prefer sellers (buyers) willing to accept low (high) prices. And buyers (sellers) also must find sellers (buyers) who are going to provide the quantities (sizes) they need. Usually, buyers and sellers need an intermediary to trade with each other.

Trading often involves brokers, dealers or market makers. Brokers act as agents on behalf of traders and arrange trades. Usually brokers will charge fees, also known as commissions. In contrast, dealers or market makers trade on their own behalf and provide liquidity to traders by standing ready to buy from those who are willing to sell and sell to those who are willing to buy.[9] Dealers or market makers are speculators who seek to make profit by buying low and selling high. The prices at which they buy (sell) securities are called bid (offer) prices. The difference between the bid and offer prices is known as the bid-offer spread. Dealers or market makers always keep an inventory (a record of financial securities they have bought long or sold short) that changes as they trade. If the securities bought are markedly different from those sold the inventory is out of balance. Consequently, dealers or market makers may change their bid/offer prices to discourage or induce trading with them.

[9] As Ibrahim (2007) explains, in some markets there is no difference between the functions carried out by a market maker and those by a dealer. However, in other markets, such as the LSE, the main difference is that market makers are obliged to offer the service of liquidity provision, while dealers can choose not to.

Trading can be carried out continuously or periodically. In a continuous market immediate execution of traders' orders is possible. Flexibility is then the main advantage of a continuous market. Continuous markets are particularly common. Almost all major stock markets, such as the LSE, have continuous trading sessions. In contrast, trading in a call market occurs only in periodic auctions. During each auction, buy and sell orders are left to accumulate over a period of time and then matched against each other in a single session. As a result of the accumulation, buyers and sellers can more easily find each other, providing the main advantage of a call market.

Recent developments in stock markets around the world suggest that traders prefer continuous trading mechanisms with opening call market auctions to exclusive call markets. Today all major stock markets, such as the TSE and NYSE, use periodic call auctions to open continuous trading sessions.

Trading is organised differently around world markets and across financial assets. The next three subsections of this introduction describe typical trading mechanisms used in stock markets: a quote-driven market, an order-driven market and a brokered market (c.f., Harris (2003)). Subsection 2.1.1 describes a quote-driven market; Subsection 2.1.2 describes an order-driven market; and Subsection 2.1.3 describes a brokered market. Subsection 2.1.4 describes a hybrid market. Finally, Subsection 2.1.5 draws a summary and introduces the remaining sections of this chapter.

### *2.1.1 Quote-Driven Markets*

A quote-driven market, also known as a dealer market, is where dealers are the market makers and quote prices. In such a market, dealers participate in every trade. Any trader who is willing to trade must trade with a dealer. For example, if Mike wants to buy 1000 BP shares, he must find a dealer who will sell these shares to him. Likewise, if Tom wants to sell 1000 BP shares, he must find a dealer who will buy these shares from him. Although Mike might be willing to buy these shares directly from Tom, in a quote-driven market they generally cannot arrange such trades. Instead, they trade indirectly with each other through the intermediation of dealers.

In a quote-driven market, traders can always obtain firm prices from dealers before submitting their orders and are free to buy or sell financial securities whether or not there exists a willing counterpart in the market. Dealers thus provide liquidity to traders and act as counterparts in transactions.

In a quote-driven market, dealers provide liquidities and facilitate trades. Hence temporary fluctuations in traders' supplies and demands are smoothed over by the dealers. For example, a big buy order arrives at the market at 10:00am. Suppose a big sell order with the same quantity will arrive at the market at 11:00am. Without the dealers, the buyer (seller) will drive up (down) the price temporarily. Since the dealers will provide the liquidity to the buyer and later buy it back from the seller, the market as a whole benefit from this buffering effect, which would lead to less price volatility.

In a quote-driven market, each dealer must quote bid and offer prices at which he is obliged to trade with the traders in quantities within the pre-set market trading limits (market size). The maximum bid-offer spread is fixed and known as the trading spread. Whenever a dealer quotes prices, he does not know whether the traders are willing to buy or sell and therefore has an incentive to quote prices that reflects his true valuation of the financial securities.

In a quote-driven market, a trade typically starts with a trader asking a dealer for price quotes. In response the dealer quotes bid and offer prices, whereupon the trader may buy at the dealer's offer price, sell at the dealer's bid price or do nothing.

The main advantage of a quote-driven market is that immediate execution of traders' orders is guaranteed. The main disadvantage is the low transparency inherent in a quote-driven market. The dealers provide quotes in response to trader inquiries and usually these are not publicly visible. Generally trade price is also not publicly visible.

Quote-driven markets are very common. Some of the largest markets are quote-driven markets, including bond, currency and swap markets.

### 2.1.2 Order-Driven Markets

An order-driven market, such as the TSE, is where trading can be arranged without the intermediation of dealers. Trading is arranged according to order precedence rules. The basic rule is price priority. A buy order priced at 200 pence, for example, will be executed before a buy order priced at 199 pence. Another rule is usually time priority. At a given price level, orders are executed first-in-first-out. For instance, a buy order priced at 100 pence submitted at 10:20am will be executed before an order also priced at 100 pence but submitted at 10:25am.

Order-driven markets vary in how they implement these order precedence rules. In some markets traders negotiate their trades face-to-face on an exchange floor. In most order-driven markets, however, buy and sell orders are matched automatically in electronic systems.

In an order-driven market, traders will not know and cannot choose with whom they trade as trades are arranged according to order precedence rules. All unmatched orders stay on an order book for subsequent trading.[10] Orders arrive at the order book randomly over time and are compared to existing ones. Executions will only occur when the price is matched. For example, suppose a buyer places a buy order to buy 2000 BP shares at 600 pence per share and a seller places a sell order to sell 3000 BP shares at 610 pence per share. There will be no match: a price of 600 pence is not acceptable to the seller; a price of 610 pence is not acceptable to the buyer. A subsequent order to buy 1000 BP shares at 620 pence per share could be matched. However there is an overlap in the acceptable prices. In this case the trade occurs at the price set by the first order: an execution will take place (for 1000 shares) at 610 pence.

Transparency is the main advantage of an order-driven market. The order book clearly shows all of the orders and what prices traders are willing to buy at or sell for. Usually the state of the order book is widely visible to most market participants. The main disadvantage of this market is that there is no guarantee of immediate execution of traders' orders.

[10] A market might have multiple order books, each managed by a different broker or entity.

Order-driven markets are very common. All markets that conduct auctions are order-driven markets. Most stock markets, such as the NYSE, have at least some order-driven market functions.

### *2.1.3 Brokered Markets*

A brokered market, very common for large blocks of stocks trading, is where brokers offer intermediary search services to traders and facilitate large (block) trades. A brokered market (also called the upstairs market) is mainly institutional. When an institutional trader contacts a broker to fill a large order, the broker will locate counterparts for the full amount or execute the order overtime. The brokers' advantage lies in access to knowledge of potential counterparts and expertise in executing large orders over time. The main characteristic of a brokered market is that brokers are finding liquidity and actively search to match buyers and sellers.

Brokered markets are very common in the economy. Real estate markets are examples of broker markets. Some stock markets, such as the NYSE, also have some brokered market functions.

### *2.1.4 Hybrid Markets*

A hybrid market facilitates trading through a blend of a quote-driven market, an order-driven market and a brokered market. This market mixes characteristic of all three types of markets mentioned above. This subsection presents two examples of hybrid markets: NYSE and NASDAQ.

The NYSE, also known as the 'Big Board', is home to the majority of the world's largest and best-known companies, as foreign-based corporations can list their shares on the NYSE if they adhere to certain SEC rules, known as listing standards. The NYSE opens for trading from 9:30am to 4:00pm (ET) every weekday, except public holidays, throughout the year.[11]

The NYSE relied for many years on a traditional call market trading system where buyers and sellers met directly on the trading floor to compete for the best possible price through the interplay of supply and demand. However, from 24 January 2007, all NYSE stocks could be traded via its electronic system known as Hybrid Market (except for a small group of very high-priced stocks).[12] Traders can now send orders for immediate electronic execution or route orders to the floor for trade in the auction market. Hence the NYSE is mainly an order-driven market where buyers and sellers trade with each other directly.

Each stock listed on the NYSE is allocated to a specialist who ensures a fair and orderly market.[13] Specialists act as auctioneers in the specific stocks they are designated to trade at a designated location, called a trading post, and serve as a point of accountability for the smooth functioning of the market. They ensure the correct market price based on supply and demand. When there is a demand-supply imbalance on the order books, the specialists act as dealers and are obliged to commit their own capital and inventories to equalise the market. For example, if there are far more buyers than sellers in the market, the specialists will provide liquidity by selling stocks to equalise the market. If there are far more sellers than buyers in the market, the specialists will commit their own capital to buy stocks as principals to equalise the market. In this case specialists work as dealers and provide additional liquidity. Thus, the NYSE also has some characteristics of a quote-driven market.

At the NYSE, floor brokers, consisting of house brokers and independent brokers, trade stocks on the trading floor.[14] House brokers are employed by broker companies and buy and sell securities as agents for their customers. The majority of independent brokers are institutional brokers who facilitate large trades at low commission rates.

[11] ET refers to the North American Eastern Time Zone.
[12] The NYSE Hybrid Market is the NYSE's new market model currently being phased in, with the best aspects of both the auction market and automated trading.
[13] A specialist is a member of the NYSE who is responsible for maintaining a fair and orderly market in the stocks he is allocated and always represents one of seven NYSE specialist firms (the main facilitators of trade on the exchange).
[14] Floor brokers are the largest single membership group of the NYSE.

Hence the NYSE also has some characteristics of a brokered market.

The NASDAQ is the largest US electronic stock market. Unlike the NYSE, having a physical trading floor that brings together buyers and sellers, all trading on the NASDAQ exchange is done over a network of computers. The NASDAQ also does not employ specialists to buy unfilled orders like the NYSE. The NASDAQ opens for trading from 7:00am to 8:00pm (ET) every weekday, except public holidays, throughout the year.

The NASDAQ has over 600 market makers with an average of 14 market makers for each stock, who are required to give a two-sided quote, meaning they must state a firm bid price and a firm offer price that they are willing to honour. These market makers are large financial companies that will buy and sell stocks through an electronic network. They are openly competing with each other for business and quote competitive prices. These market markers will trade with individual traders and other market makers with their own capital.

Individual traders' orders are sent electronically to the NASDAQ, where market makers list their quotes. Once prices are matched, orders are executed electronically. Hence, the NASDAQ is mainly a quote-driven market. Unexecuted orders can be posted alongside and compete with market makers' quotes to be executed with incoming new orders. Thus, the NASDAQ also has some characteristics of an order-driven market.

Both the NYSE and the NASDAQ accommodate a major portion of all equities trading in the world. The fundamental difference between the NYSE and NASDAQ is the trading mechanisms used by the exchanges. The NASDAQ is mainly a quote-driven market, where traders are buying from and selling to market makers; the NYSE is mainly an order-driven market, where traders are trading with each other directly. At the NYSE, some trading occurs on the trading floor while at the NASDAQ all trading takes place between traders and market makers through an electronic network.

### 2.1.5 *Summary*

This section describes trading mechanisms used in stock markets: a quote-driven market, an order-driven market and a brokered market. The main difference between a quote-driven market and an order-driven market is the involvement of dealers in the trading process. In a quote-driven market, dealers post prices before orders are submitted and traders will trade with dealers directly. Hence traders' order execution is guaranteed. In contrast, in an order-driven market, all traders submit their orders before prices are determined and traders interact with each other directly. Thus immediate execution of traders' orders is not always possible. The main difference between a brokered market and a quote-driven/order-driven market is the involvement of brokers - who are actively matching buyers and sellers to facilitate large trades - in the trading process.

Recently stock markets are converging towards a hybrid market structure, which mixes characteristics of all three types of markets mentioned above. Some major stock markets, such as the NYSE and NASDAQ, all have hybrid market structures.

The presentation in this chapter proceeds as follows. The next section introduces trading services of the LSE with a focus on Stock Exchange Electronic Trading Service (SETS) (c.f., LSE (2007)). Section 2.3 describes a SETS trading day (c.f., LSE (2007)). Section 2.4 describes SETS orders and order books (c.f., LSE (2007)). Section 2.5 describes SETS segments and sectors (c.f., LSE (2007)). Section 2.6 describes trade and transaction reports (c.f., LSE (2007)). Section 2.7 concludes this chapter.

## 2.2 Trading Services of the London Stock Exchange

The LSE, the largest stock exchange in Europe, is a centralised marketplace in London where issuers raise capital (listing of securities) and participants buy and sell securities.[15] Issuers can list their securities on the LSE to have access to a wide pool of

[15] Originated in 1773, the regional exchanges were merged in 1973 to form the Stock Exchange of Great Britain and Ireland, later renamed the LSE.

capital. This is known as a primary market.[16] In addition to the provision of a primary market for securities, The LSE provides mechanisms for securities price formation that results from the interplay of supply and demand and offers investors the facilities to trade securities amongst themselves, either directly or through intermediaries (market makers). This is known as a secondary market.

The LSE provides a secondary market for trading in a wide range of securities that includes UK and international stocks and gilts. It offers three trading services for trading UK stocks as follows.

- SETS (Stock Exchange Electronic Trading Service) is an electronic order book used for trading the constituents of the FTSE All Share Index, Exchange Traded Funds, Exchange Traded Commodities, along with over 180 of the most traded AIM and Irish securities.[17] SETS was introduced on 27 October 1997. It allows participants to submit orders to buy or sell quantities of shares at specific prices. Orders can be submitted either on behalf of clients or for participants' own trading purposes. Participants add orders or execute against existing orders by sending messages electronically to SETS. Executions occur automatically in accordance with strict price then time priority.
- SETSqx (Stock Exchange Electronic Trading Service – quotes and crosses) is a trading platform for securities less liquid than those traded on SETS. It combines a periodic electronic auction with quote-driven market making.[18]
- SEAQ (Stock Exchange Automated Quotation System) is a quote-display system used as a price reference point for execution between market participants and registered market makers. All UK stocks (not on SETS or SETSqx) with two or more registered market makers are traded on SEAQ.

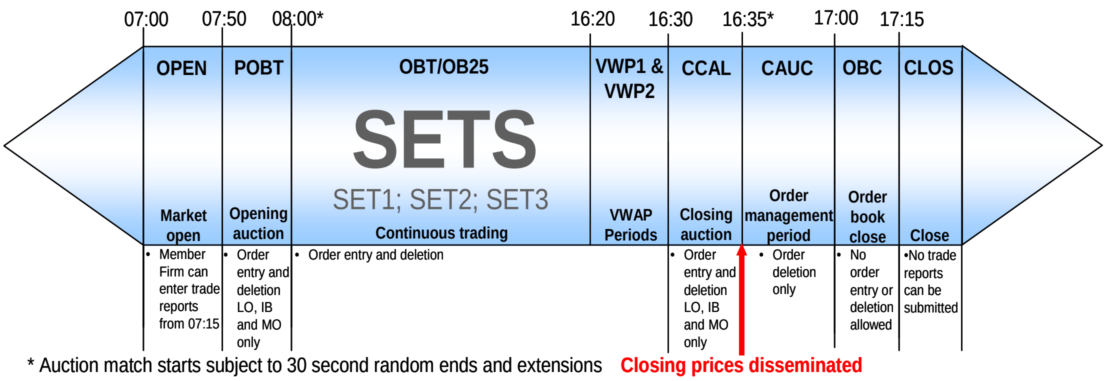
[16] The LSE offers four primary markets: Main Market, Alternative Investment Market (AIM), Professional Securities Market (PSM) and Specialist Fund Market (SFM). These markets are not the focus of this study. The remaining sections of this chapter focus on the secondary market.
[17] Since 8 October 2007, SETS has started to offer market making in all stocks including those deemed to be 'liquid' under the new Markets in Financial Instruments Directive (MiFID). The remaining sections of this chapter contain descriptions of SETS excluding the changes introduced on 8 October 2007, as the data sample period in this study is from October to December 2000.
[18] Since 8 October 2007, all Main Market and AIM equity securities not traded on SETS and with less than two registered market makers are traded on SETSqx.

The LSE provides three trading services for trading UK stocks: SETS, SETSqx and SEAQ. The remaining sections of this chapter will focus on SETS. The next section will describe a typical SETS trading day.

## 2.3 A SETS Trading Day

SETS is open on all working days from Monday to Friday. The trading day is divided into three main trading periods. It starts with an opening auction followed by a period of continuous trading and ends with a closing auction where an official closing price is set (see Figure 2.1). Operations within each period are described as follows.

**Figure 2.1: A SETS Trading Day**




The codes in the top row are 'Period Name' codes given to a particular period that is defined by a specific set of rules: OPEN (the period that defines when the trading service is opened), POBT (Pre Order Book Trading), OBT/ OB25 (Order Book Trading), VWP1 & VWP2 (Order Book Trading Volume Weighted Average Price (VWAP) Calculation), CCAL (Closing Auction Call), CAUC (Closing Auction Execution followed by End of Order Book Trading), OBC (Order Book Closed), CLOS (Closed).

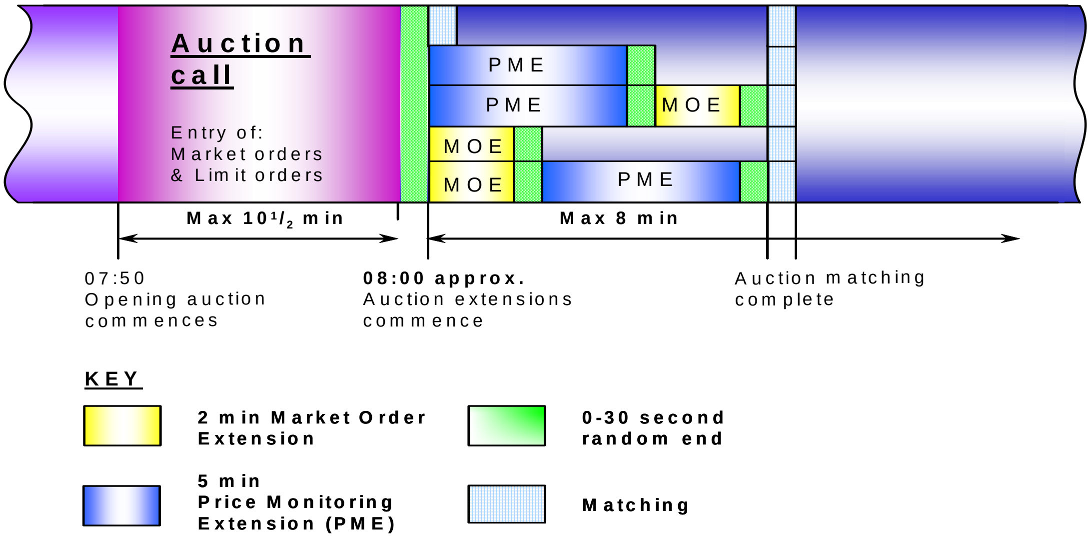
*Source: The LSE Guide to Trading Services. Courtesy of the London Stock Exchange*

*Opening Auction*

The trading day begins with an opening auction. During this period orders can be entered, modified or cancelled. It lasts about 10 minutes from 07:50 to 08:00 subject to extensions and 30-second random end periods.[19] There are five possible scenarios for how the auction extensions can occur (see Figure 2.2).

[19] Auction extension periods prolong the auction call period to provide the market with more time to react to market imbalances and are categorised into market order extension (MOE) and price monitoring extension (PME). MOE occurs if market orders would remain unexecuted following the auction match.

**Figure 2.2: Opening Auction**




*Source: The LSE Guide to Trading Services. Courtesy of the London Stock Exchange*

The auction match is the process by which orders on a crossed order book are executed against each other at the end of the auction call including any auction extension periods.[20] For auction matching to occur each security is temporarily 'frozen' while an algorithm is run and no additional orders may be entered or cancelled until the matching process for that security is completed. After the matching, all unexecuted orders remain on the order book and are carried forward into the subsequent trading period.

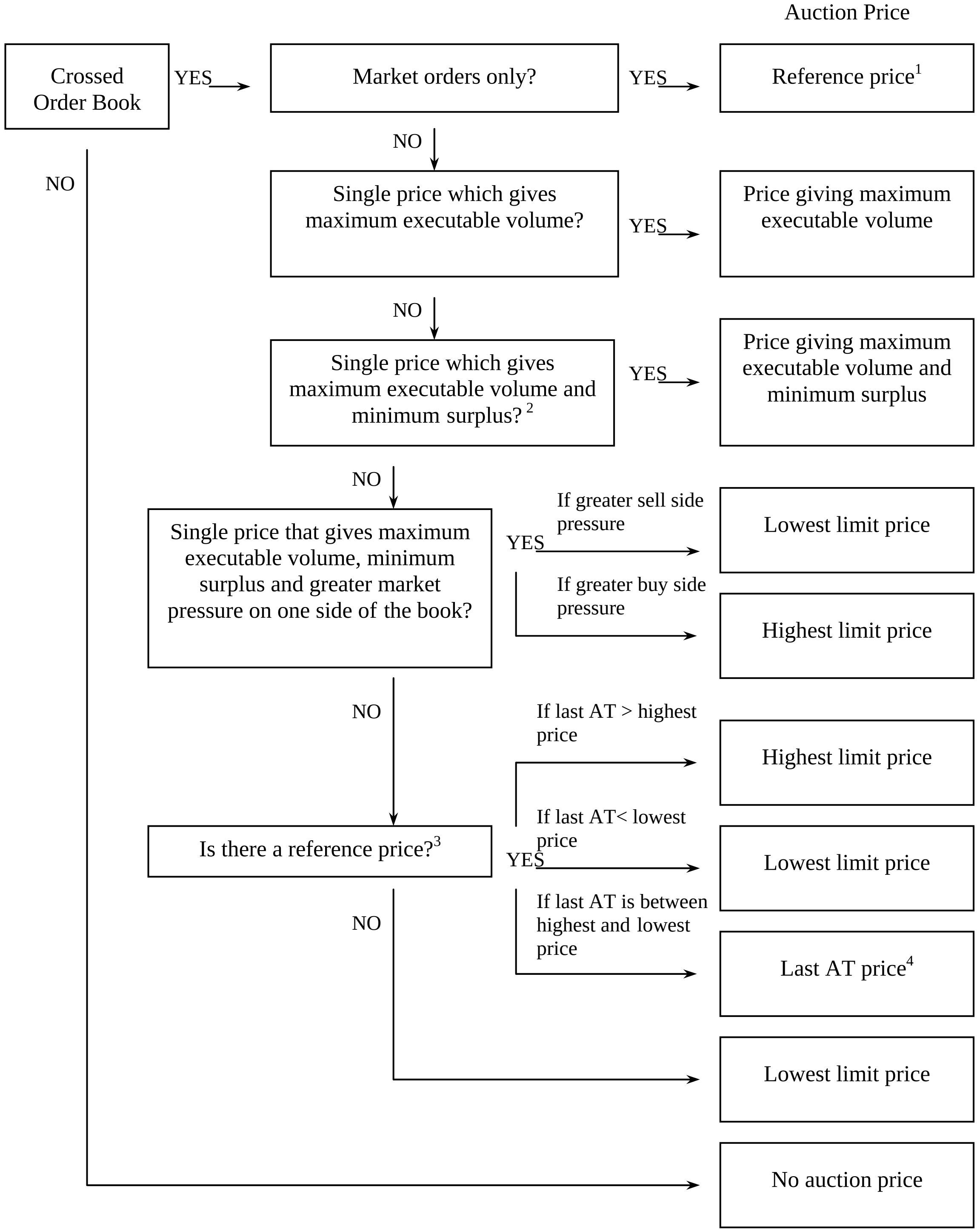
The opening auction will result in an opening price. This opening price is determined by choosing the price at which the largest volume of shares is executed with some additional rules (see Figure 2.3).

---

The purpose of this extension is to minimise the market order surplus following the auction match by giving market participants more time to enter orders that will execute against this surplus. PME occurs when the indicative auction match price breaches the price tolerance level (the percentage up to which the price of a security can increase or decrease from the price of the last automatic or auction match trade). If the conditions described above do not occur, auction matching takes place and there are no auction extensions. A random end to the open auction period of up to 30 seconds ensures that there is high quality price formation.

[20] The order book is crossed where there are buy orders priced at a higher or equal level to sell order prices or market orders are available to execute against limit orders on the other side of the book. If the order book is not crossed, there will be no auction match and trading will move to the subsequent period without an auction price.

**Figure 2.3: Summary of Auction Price Determination for a Crossed Order Book**




1. The reference price is the last automatic trade (AT) price.
2. Maximum executable volume is the highest volume of orders can execute. Minimum surplus is the least order volume unexecuted on the book.
3. If there is no reference price, then the system will prevent market orders from being entered.
4. While 'Last AT' is stated above for simplicity, in fact this could be the last AT or the last auction match trade (UT).

*Source: The LSE Guide to Trading Services. Courtesy of the London Stock Exchange*

*Continuous Trading*

Continuous trading commences at the end of the opening auction, 08:00 (depending on the 30-second random end and auction extensions), and continues until the beginning of the closing auction at 16:30. During this period, orders are automatically executed when order prices match and volume is available.

If the price tolerance levels are reached or breached during the continuous trading period, continuous trading will be suspended and an Automatic Execution Suspension Period (AESP) intra-day auction will occur.[21] This auction provides market participants time to react to the change in market conditions and find a market price. The AESP follows the same process as the opening auction and consists of an auction call period, which lasts for five minutes subject to a market order extension, and random end periods (see Figure 2.4). At the end of AESP this security will return to continuous trading.

*Closing Auction*

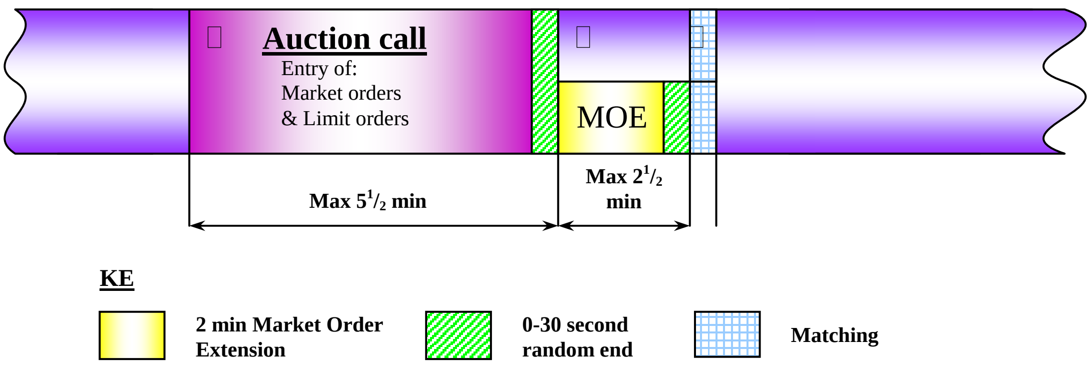
This period follows the same process as the opening auction but with a five-minute call period from 16:30 to 16:35 and probably an additional price monitoring extension to ensure a high quality closing price (see Figure 2.5).[22] During this call period, participants can cancel existing orders or enter new orders. After the auction process end, the closing price will be disseminated.

Usually the price resulting from the closing auction is the closing price of the trading day. If no auction matching occurs, the Volume Weighted Average Price (VWAP) of the last 10 minutes of continuous trading (between 16:20 and 16:30) will be the closing

[21] The price tolerance level is the percentage up to which the price of a security can increase or decrease from the price of the last automatic or auction match trade.
[22] The reference price of a security used in the closing auction will be the VWAP of all automatic trades (AT) occurred between 16:20 and 16:30 or the last AT price prior to 16:20 in the event that there is no VWAP price. If the auction matching process would result in an auction price that is a pre-determined percentage above or below the reference price, then auction matching is temporarily halted. The auction call period is then extended for five minutes (PME1), during which time participants can cancel or enter new orders. If PME1 has occurred and the auction matching process would still result in an auction price that is a pre-determined percentage above or below the reference price, then auction matching is temporarily halted again. The auction call period is then extended for a further five minutes (PME2), during which time participants can cancel or enter new orders. At the end of this extension, the auction matching process is run again and there will be no further price checks.

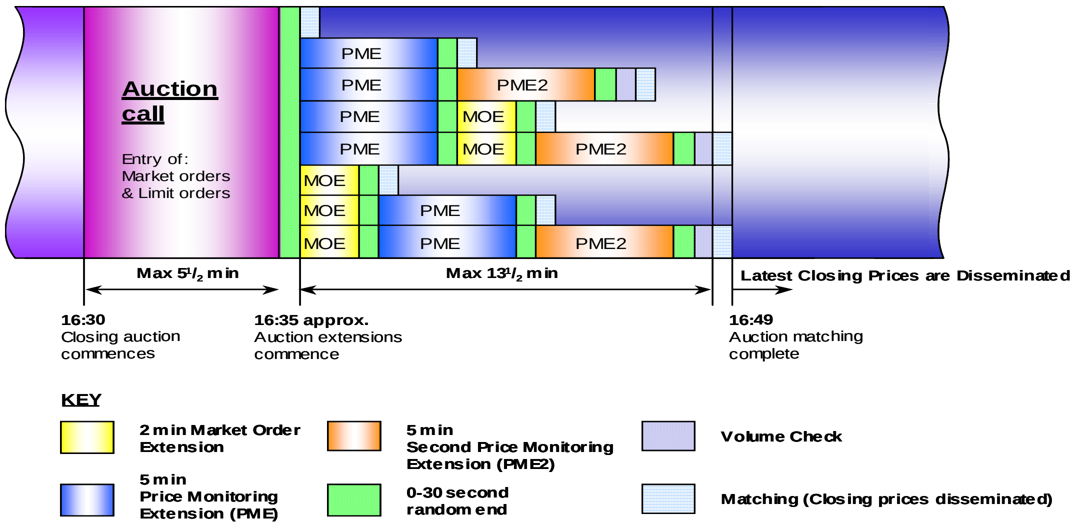
price and if no trades are executed between 16:20 and 16:30, the last automatic trade price will be used as the official closing price (see Figure 2.6).

**Figure 2.4: AESP Auction**



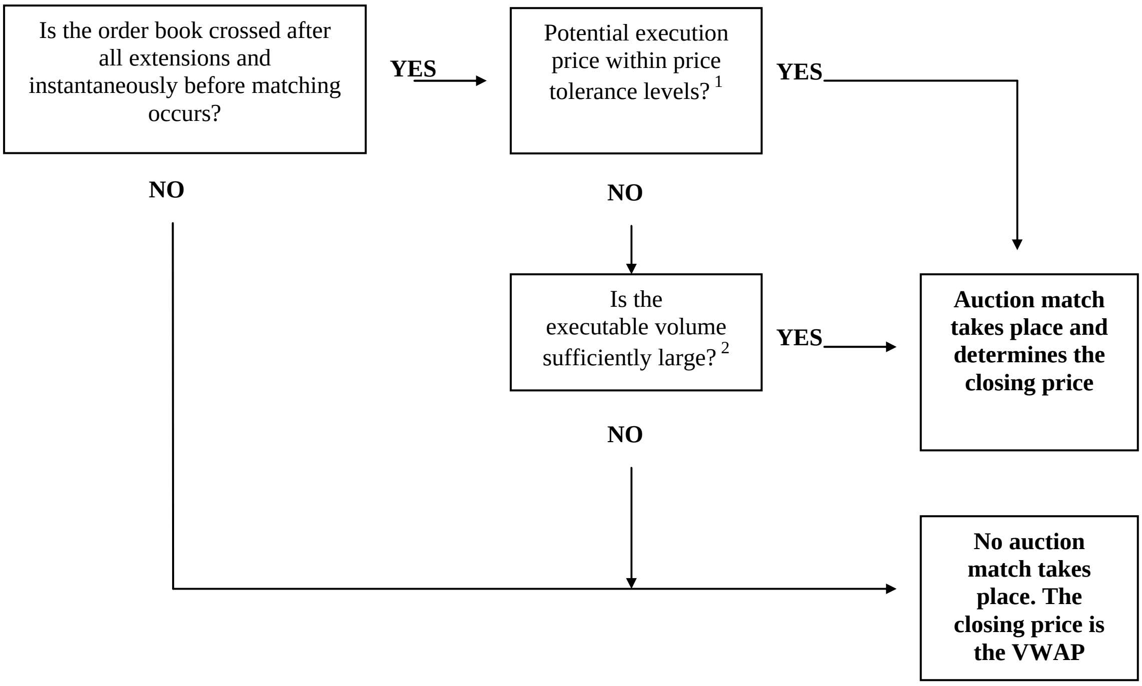

*Source: The LSE Guide to Trading Services. Courtesy of the London Stock Exchange*

**Figure 2.5: Closing Auction**




*Source: The LSE Guide to Trading Services. Courtesy of the London Stock Exchange*

**Figure 2.6: The Closing Price Determination for SETS Securities**




1. The price tolerance level is a pre-determined percentage above or below the reference price.
2. The executable volume is of a large enough size to deem the price to be a representative closing price.

*Source: The LSE Guide to Trading Services. Courtesy of the London Stock Exchange*

*Exchange Delivery Settlement Price Intra-day Auction*

During quarterly and monthly expiry of FTSE 100 index derivatives, market activity is typically very high as derivatives traders trade FTSE 100 securities. This heavy trading could potentially lead to rapid price swings in these securities as baskets of orders enter the system. To prevent increased volatility during these periods, there will be an intra-day auction, known as an Exchange Delivery Settlement Price (EDSP) auction, which takes place on the third Friday of each month for FTSE 100 index constituents.

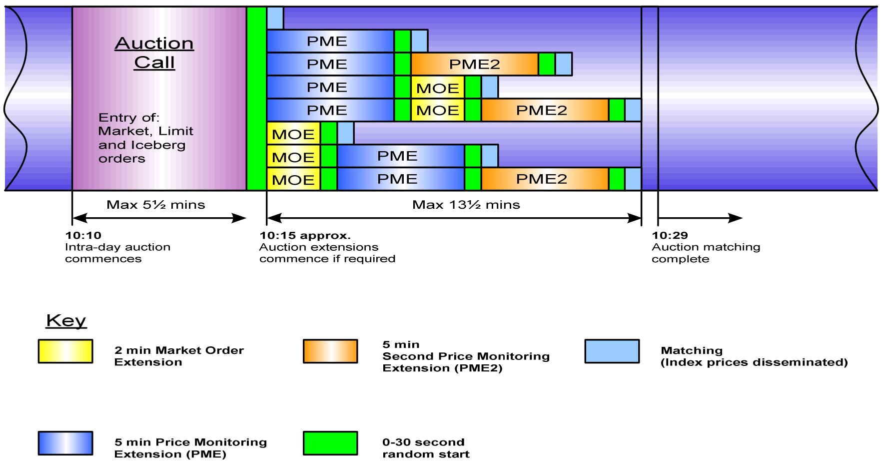
The EDSP auction will be based on a similar process as the closing auction, except with no volume check.[23] The auction will begin at 10:10 on the relevant Friday and may not complete until 10:29. The initial auction call period (from 10:10 to 10:15) enables participants to submit orders.[24] During this period, no automated execution occurs and

[23] There is no minimum required volume in the matching process.
[24] The reference price of a security used in the intra-day auction will be the latter of the last automatic trade price prior to 10:10 or where there is no automatic trade price, the previous day's closing price. When there are market imbalances, auction extension periods prolong the auction call period to provide the market with more time to react. These extensions are categorised into first price monitoring extension (PME), second price monitoring extension (PME2) and market order extension (MOE). If the auction

orders remain on the order book until they are cancelled or expire. During this auction call period an indicative auction match price is displayed if the price of the buy and sell orders are crossed, or market orders are available to execute against orders on the other side of the book. The auction match is the process by which crossed orders are executed against one another at the end of the auction call including any auction extension periods (see Figure 2.7).[25] After the auction process end, the auction price is made available to the market.

FTSE International Limited will calculate and disseminate real time index values during the intra-day auction period. Upon completion of the EDSP auction, it will publish a flagged final EDSP, calculated using the auction price for each of the securities in the index.[26]

**Figure 2.7: EDSP Intra-Day Auctions for SETS Securities**




*Source: The LSE Guide to Trading Services. Courtesy of the London Stock Exchange*

---

matching process would result in an auction price that is a pre-determined percentage above or below the reference price, then auction matching is temporarily halted. The auction call period is then extended for five minutes (PME), during which time participants can cancel or enter new orders. If the PME has occurred and the auction matching process would still result in an auction price, which is a pre-determined percentage above or below the reference price, then auction matching continues to be temporarily halted. Then the auction call period is extended for a further five minutes (PME2), during which time participants can also cancel or enter new orders. If the auction matching process would result in market orders remaining unexecuted on the order book, auction matching is temporarily halted. The auction call period is extended for two minutes (MOE). If a MOE has occurred, the auction match process will run even if further market order imbalances remain. A MOE may occur before or after PME but not after PME2.

[25] If the order book is not crossed there will be no auction match, and trading will move to the subsequent period without an auction price.

[26] In the case where a security has not produced an auction price due to the security not having a crossed book, the latter of the last automated traded price prior to the EDSP auction or the previous day's closing price will be used to calculate the final EDSP.

A SETS trading day commences with an opening auction followed by a period of continuous trading in which orders are automatically executed against one another and ends with a closing auction where an official closing price is set. Table 2.1 summarises the trading day.

Table 2.2 summarises the opening auction, closing auction and AESP. The auction call periods in the opening auction, closing auction and AESP enable participants to represent firm intentions to trade via entry of orders. No automated execution occurs during these periods, and orders remain on the order book until the auction match, cancellation or expiration. If the auction match volume is less than a pre-determined multiple of the Normal Market Size (NMS) of the security, then auction matching will not occur.[27]

**Table 2.1: Opening Auction, Continuous Trading and Closing Auction**

| | Opening Auction | | Continuous Trading | | Closing Auction | |
|---|---|---|---|---|---|---|
| | **Auction call period** | **Auction matching** | **Continuous trading** | **VWAP** | **Auction call period** | **Auction matching** |
| **Scheduled timings** | 07:50 - 08:00[1] | | 08:00[1] - 16:20 | 16:20 - 16:30 | 16:30 - 16:35[1] | |
| **Trade reporting** | Yes | Yes | Yes | Yes | Yes | Yes |
| **Order entry and deletion** | Yes | No | Yes | Yes | Yes | No |
| **Continuous execution** | No | No | Yes | Yes | No | No |

1. The opening and closing auctions may be longer in duration than those stated here, due to price monitoring or market order extensions, and a random end time.

*Source: The LSE Guide to Trading Services. Courtesy of the London Stock Exchange*

[27] The NMS, a value assigned to a security by the Exchange, indicates the liquidity of that security. The pre-determined multiple for the volume check is: 0.5 x NMS for securities with a NMS of 5,000 or more and 2,500 for securities with a NMS of below 5,000.

**Table 2.2: Summary of Opening Auction, Closing Auction and AESP**

| Type of auction | Auction call period | Random end period | Extensions | | |
|---|---|---|---|---|---|
| | | | Market order extension (MOE) | Price monitoring extension (PME1) | Price monitoring extension (PME2) |
| **Opening auction** | 10 minutes | 30 seconds | 2 minutes | 5 minutes | N/A |
| **Automatic execution suspension period (AESP)** | 5 minutes | 30 seconds | 2 minutes | N/A | N/A |
| **Closing auction** | 5 minutes | 30 seconds | 2 minutes | 5 minutes | 5 minutes |

*Source: The LSE Guide to Trading Services. Courtesy of the London Stock Exchange*

During the auction periods an indicative auction match price is displayed if the price of the buy and sell orders are crossed, or market orders are available to execute against limit orders on the other side of the book. There can only be one auction price per security, and this price is determined according to the status of the orders on the book at the time of matching.

## 2.4 SETS Orders and Order Books

Traders need to submit orders in order to trade. An order represents a firm commitment to trade a security with specified conditions including side indicator (buy or sell), volume, specified price (optional), dealing capacity and specified expiry time and/or date (optional).[28]

Unless an expiry time and date has been entered, SETS will default all expiry dates to submission dates.[29] Orders that expire at midnight on any trading day will be removed from SETS. Orders that are not yet due to expire will remain on the order book and will be available for trading in the next day's opening period.

[28] A dealing capacity is used to indicate whether a participant has acted in an agency or a principal capacity. When the price of an order is specified, it must be an exact multiple of the tick size (minimum price variation) that is 0.01 penny for SETS stocks.
[29] The maximum expiry date allowed by SETS is 90 calendar days from the day of submission.

The most frequently used type of order in SETS is a limit order. A limit order can be partly filled with the 'unfilled' part remaining on the order book with the same time priority pending further fills or cancellation. Thus, a limit order can generate a number of fills. If a limit order executes against more than one order, it is known as a multiple fill order. For example, an order of 10,000 shares may execute against five separate orders of 2,000 shares each.

A limit order can leave the book in the following ways: final execution at the limit-order price, cancellation or expiration. The following examples illustrate three possible paths a buy limit order can follow. A buy limit order is submitted with a limit-order price of 100 pence and after 30 minutes, it is cancelled or expired; a buy limit order is submitted with a limit-order price of 101 pence and after 10 minutes, it is fully executed at the price of 101 pence; or a buy limit order is submitted with a limit-order price of 100 pence and an order size of 1000 shares and, after 15 minutes, it is partly filled, and after another 15 minutes, it is cancelled or expired.

Except limit and market orders, the following types of orders can also be used in SETS: iceberg order, at best order, execute-and-eliminate order and fill-or-kill order. [30] Imbrahim (2007) describes them as follows:

- *Iceberg Order*. An iceberg order is a limit order but with an additional functionality that allows market participants to enter the order with only a certain portion of it (the 'peak') being publicly visible.[31] The non-peak size of the order remains 'invisible' to market participants. Once this 'peak' is executed another 'peak' quantity becomes publicly visible. It allows participants to enter large limit orders onto the order book while revealing only a portion of the orders to the market.
- *At Best Order*. In an At Best Order only the quantity is specified. The order is executed against the best price currently available. If the order is a buy order then it is executed against the prevailing best offer price and if it is a sell order

[30] In SETS, market orders can only be entered during an auction call period. However, if they remain unexecuted during the auction they will be carried forward to continuous trading sessions.
[31] An iceberg order was introduced in September 2003.

then it is executed against the prevailing best bid price. After execution, any unexecuted portion will be rejected.[32]

- *Execute-and-Eliminate Order*. In execute-and-eliminate orders the price and quantity are specified. These orders are executed against immediately available matching orders of any size. The distinguishing feature is that after this first matching round any remaining unmatched quantity in the execute-and-eliminate order is cancelled.[33]

- *Fill-or-Kill Order*. In a fill-or-kill order the quantity is specified while the price does not have to be. If the price is not specified the order is matched at the best available price. The main feature of these types of orders is that they are cancelled if the whole quantity is not immediately matched.

In SETS, market, limit, iceberg, at best, execute-and-eliminate and fill-or-kill orders are anonymous and can be used to trade securities. Table 2.3 presents when to use each type of orders and Table 2.4 summarises order execution options. When entering an order, participants are required to supply SETS with specific information and some orders are only allowed during a certain period of a trading day. Table 2.5 summarises order entry requirements.

Submitted orders are displayed and matched on a computer screen known as the SETS order book (see Figure 2.8). The SETS order book is a queue of all unexecuted orders for subsequent trading. It is maintained for each SETS stock and represents the available liquidity that can be provided immediately to subsequent trading. Orders arrive at the order book randomly over time and are compared to existing ones. Executions will only occur when the price is matched. For example, as Figure 2.8 illustrates, the best bid (offer) price for stock ABC is 524 (525) pence. A limit order to buy 1000 shares at 524.5 pence will not normally be executed immediately. In most circumstances, this price (524.5 pence) will be publicly disseminated as the new prevailing best bid price. A sell order might arrive and execute against this buy order, resulting in a transaction at 524.5 pence. On the other hand, market best bid/offer prices might increase, leaving this buy order unexecuted.

[32] In some exchanges, for example at the NYSE, these orders are called ‘market orders’.
[33] The difference between at best orders and execute-and-eliminate orders is that execute-and-eliminate orders have specified prices.

## Table 2.3: Suggestions of When to Use Each Type of Order

| **Desired action** | **Type of order** |
|---|---|
| **To take part in an auction at a limit-order price.** | Limit/iceberg |
| **To maximise the possibility of execution in an auction.** | Market |
| **To reside on the book at a limit-order price.** | Limit/iceberg |
| **Execute in full or in part against existing order at no worse than a limit-order price, leaving any unexecuted part of the order on the book.** | Limit/iceberg |
| **Enter an order for inclusion in the auction matching process.** | Limit/iceberg/market |
| **Execute in full or in part against existing orders at the best price available, deleting any unexecuted part of the order.** | At best |
| **Execute in full or in part against existing orders at no worse than a limit-order price, deleting any unexecuted part of the order.** | Execute-and-eliminate |
| **Execute in full against existing orders at the best price available or delete the whole order.** | Fill-or-kill without a limit-order price |
| **Execute in full against existing orders at no worse than a specified price or delete the whole order.** | Fill-or-kill with a limit-order price |

*Source: The LSE Guide to Trading Services. Courtesy of the London Stock Exchange*

## Table 2.4: Summary of Order Execution Options

| **Order type** | **Limit-order price required** | **Execution price** | **Partial execution allowed** | **Unexecuted part added to the book?** |
|---|---|---|---|---|
| **Market** | No | Best bid/offer prices | Yes | Yes |
| **Limit** | Yes | Limit-order price[1] | Yes | Yes |
| **Iceberg** | Yes | Limit-order price[2] | Yes | Yes |
| **At Best** | No | Best bid/offer prices | Yes | No |
| **Execute-and-eliminate** | Yes | Limit-order price[3] | Yes | No |
| **Fill-or-kill** | Optional | Best bid/offer prices or limit-order price[4] | No[5] | No |

1. When the limit-order price is better than the best bid/offer prices and volume is available, a limit order will be executed at the best bid/offer prices.
2. When the limit-order price is better than the best bid/offer prices and volume is available, an iceberg order will be executed at the best bid/offer prices.
3. When the limit-order price is better than the best bid/offer prices and volume is available, an execute-and-eliminate order will be executed at the best bid/offer prices.
4. When the limit-order price is better than the best bid/offer prices and volume is available, a fill-or-kill order will be executed at the best bid/offer prices.
5. A fill-or-kill order can result in multiple fills at a time.

*Source: The LSE Guide to Trading Services. Courtesy of the London Stock Exchange*

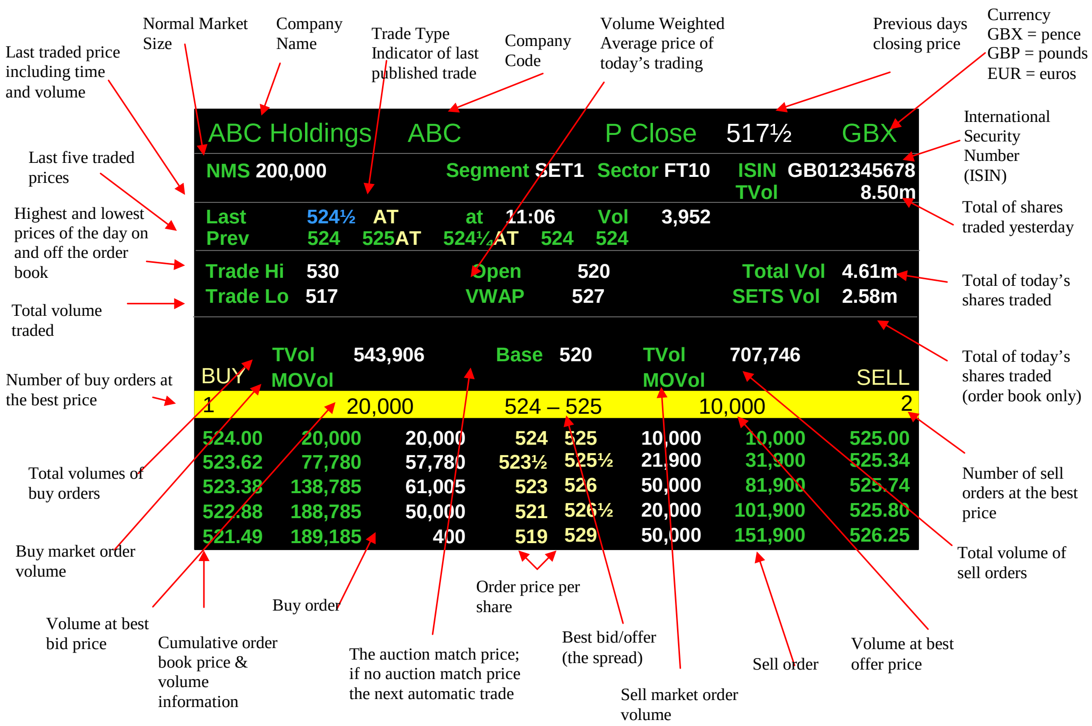
**Table 2.5: Summary of Order Entry Requirements**

| Order type | Information requirement | ENTERED DURING | | |
|---|---|---|---|---|
| | | Auction call | Extension[1] | Continuous trading |
| **Limit** | • Buy or sell<br>• Volume<br>• Limit-order price<br>• Dealing capacity<br>• (Specified expiry time and/or date)[2] | ✓ | ✓ | ✓ |
| **Iceberg** | • Buy or sell<br>• Peak size<br>• Volume<br>• Limit-order price<br>• Dealing capacity<br>• (Specified expiry time and/or date) | ✓ | ✓ | ✓ |
| **Market** | • Buy or sell<br>• Volume<br>• Dealing capacity<br>• (Specified expiry time and/or date) | ✓ | ✓ | × |
| **At best** | • Buy or sell<br>• Volume<br>• Dealing capacity | × | × | ✓ |
| **Fill-or-kill** | • Buy or sell<br>• Volume<br>• (Limit-order price)<br>• Dealing capacity | × | × | ✓ |
| **Execute-and-eliminate** | • Buy or sell<br>• Volume<br>• Limit-order price<br>• Dealing capacity | × | × | ✓ |

1. Market order and price monitoring extensions.
2. Information requirements in brackets are optional.

*Source: The LSE Guide to Trading Services. Courtesy of the London Stock Exchange*

## Figure 2.8: A Typical SETS Order Book




*Source: The LSE Guide to Trading Services. Courtesy of the London Stock Exchange*

Key features of SETS orders and order books are summarised as follows.

- All orders on the book are anonymous.
- Full depth of the order book is displayed.
- Only limit, iceberg and market orders are eligible to be entered during the auction call period.
- Only limit, iceberg and market orders may sit on the order book during continuous trading; other order types are either immediately executed or rejected, in full or in part.
- Market orders take execution priority over limit and iceberg orders as market orders indicate a willingness to trade a specified quantity at any price.
- Market orders appear above limit and iceberg orders on both sides of the order book.
- Market orders can only be entered during auction phases and cannot be entered during continuous trading periods. In the unlikely event that a market order is not fully executed during an auction, it can be cancelled during the continuous trading period.

- When market, limit and iceberg orders sitting on the order book are partially executed, the remaining part of the order keeps the same time priority as the original order.
- The maximum time a market, limit or iceberg order can sit on the order book is 90 calendar days. An order can leave the book in three ways: completion / final execution, cancellation and expiration.

This section describes SETS orders and order books. The next section will introduce SETS segments and sectors.

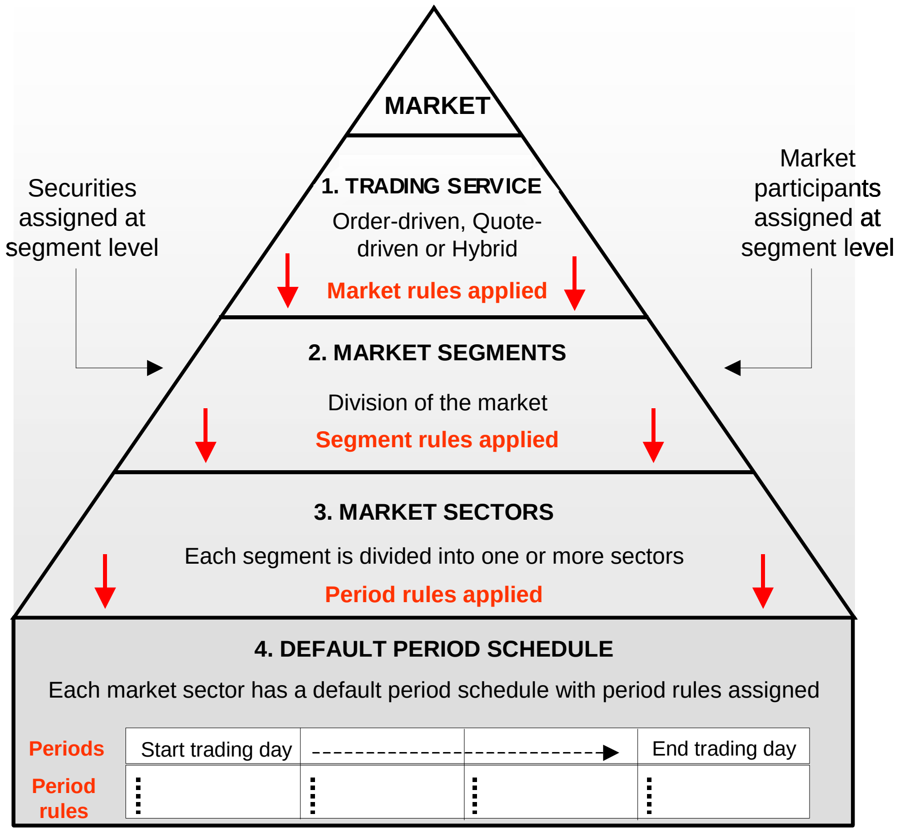
## 2.5 SETS Segments and Sectors

A trading service, such as SETS, characterised by a number of specific rules (known as market rules) is a subdivision of a stock exchange.[34] A market segment is a subdivision of a trading service. Similar tradable instruments and participants are assigned to a particular segment characterised by a number of specific rules, know as segment rules, which govern the trading activities that may take place within that segment.[35] A market segment is divided into one or more market sectors. Each market sector consists of a schedule of periods (default period schedule) characterised by a number of specific rules, known as period rules, each of which is assigned a start time, thus enabling different rules to come into operation at different times of the day.[36]

As Imbrahim (2007) summarises, stocks in market segments are governed by their own standard set of trading rules, which may admit only a few specific types of orders. However, such is the variety of available trading methods that some market segments maybe further divided into market sectors where in any particular trading day a schedule of trading periods is drawn and a different set of trading rules, possibly involving different types of orders, may then apply to each period. Figure 2.9 summarises the classification of market segments and sectors at the LSE.

[34] Each trading service is governed by an independent set of rules.
[35] Each market segment is governed by an independent set of rules.
[36] Period rules are time-dependent rules and govern the functionality relating to trading within a sector. Period rules are grouped within market sectors, as a schedule of periods, each coming into effect at pre-defined times as the trading day progresses.

**Figure 2.9: Market Segments and Sectors**




*Source: The LSE Guide to Trading Services. Courtesy of the London Stock Exchange*

SETS has three segments: SET1, SET2 and SET3, within each of which there are a number of sectors.[37] SET1 includes the FTSE 100 stocks while SET2 includes the FTSE 250 stocks. SET3 includes the remaining SETS stocks. Table 2.6 presents the market segments and sectors of SETS. Tables 2.7, 2.8 and 2.9 present the default schedules associated with each sector. Timings for the SETS sectors are the same (except for SECN) with the following differences. First, OBT (OB25) period price monitoring tolerance level, predefined percentage thresholds either side of the reference price, is set at 3% (25%).[38] Second, VWP1 and VWP2 are VWAP trading rules that are activated for the last 10 minutes of trading and have a 3% price monitoring tolerance. If the price monitoring tolerance levels are breached during VWP1 the stock enters an extended closing auction call. Likewise if the stock enters VWP2 and breaches its 3% tolerance level it enters an extended closing auction call. Finally, VWA1 and VWA2

[37] This was the market segments and sectors arrangement before 8 October 2007.

[38] Price monitoring is based on a comparison between the next potential automatic execution price(s) and a reference price defined as the last order book trade price observed before the aggressive order was entered. Price monitoring functionality tracks the prices at which executions are occurring and will halt execution if certain tolerances would be breached. Price monitoring protects the market against excessive price movements, thereby maintaining an orderly market.

act in exactly the same way as VWP1 and VWP2 with the difference being VWA1 and VWA2 have a 25% price monitoring tolerance.

**Table 2.6: Sectors and Segments of SETS**

| Service | Segment | Sector | Description |
|---|---|---|---|
| SETS | SET 1 | FE10 | FTSE 100 SECURITIES[1] |
| | | FF10 | FTSE 100 SECURITIES[1] |
| | | FS10 | FTSE 100 SECURITIES |
| | | FT10 | FTSE 100 SECURITIES |
| | | SS10 | SETS SECURITIES FTSE 100[2] |
| | SET 2 | FE25 | FTSE 250 SECURITIES |
| | | FT25 | FTSE 250 SECURITIES |
| | | SS25 | SETS SECURITIES FTSE 250[2] |
| | SET 3 | STOS | OTHER SETS SECURITIES |
| | | STSS | SETS SECURITIES[2] |
| | | SECN | SECURITY NOTIFICATIONS[3] |

1. FTSE100 securities traded in SET1 that had an index weighting of 1% or more would be traded in the FE10 and FF10 sectors. During EDSP intra-day auction, those securities would operate on a 1% price monitoring tolerance whilst all the other securities would operate on a 3% or 25% price monitoring tolerance.
2. These securities have a price monitoring tolerance of 25%.
3. The SECN sector consists of dummy securities used for notification purposes of portfolio trades.

*Source: The LSE Guide to Trading Services. Courtesy of the London Stock Exchange*

**Table 2.7: Default Period Schedule for SETS Containing Sectors FT10, FS10, FE10, FF10, FT25, FS25 and STOS**

| Securities | Period name | Start |
|---|---|---|
| **SET 1 (all sectors) SET 2 (sectors FT25 & FS25) SET 3 (sector STOS)** | OPEN | 07:00 |
| | POBT | 07:50 |
| | OBT | 08:00 |
| | VWP1 | 16:20 |
| | VWP2 | 16:30 |
| | CCAL | 16:30 |
| | CAUC | 16:35 |
| | OBC | 17:00 |
| | CLOS | 17:15 |

*Source: The LSE Guide to Trading Services. Courtesy of the London Stock Exchange*

**Table 2.8: Default Period Schedule for SETS Containing Sectors SS10, SS25 and STSS**

| Securities | Period name | Start |
|---|---|---|
| **SET 1 (sector SS10)**<br>**SET 2 (sector SS25)**<br>**SET 3 (sector STSS)** | OPEN | 07:00 |
| | POBT | 07:50 |
| | OT25 | 08:00 |
| | VWA1 | 16:20 |
| | VWA2 | 16:25 |
| | CCAL | 16:30 |
| | CAUC | 16:35 |
| | OBC | 17:00 |
| | CLOS | 17:15 |

*Source: The LSE Guide to Trading Services. Courtesy of the London Stock Exchange*

**Table 2.9: Default Period Schedule for SETS Containing Sector SECN**

| Securities | Period name | Start |
|---|---|---|
| **SET 3 (Sector SECN)** | OPEN | 07:00 |
| | OBC | 17:00 |
| | CLOS | 17:15 |

*Source: The LSE Guide to Trading Services. Courtesy of the London Stock Exchange*

This section described SETS segments and sectors. The next section will describe SETS trade and transaction reports.

## 2.6 SETS Trade and Transaction Reports

SETS participants are obligated to report the detail and confirmation of a transaction to the LSE. These reports are known as trade and transaction reports.

*Trade Reports*

A trade report is defined as a report of the detail of a transaction effected at the LSE. The LSE will publish this information subject to certain criteria.

Trades executed automatically on SETS order books will generate automatic trade reports.[39] All order book trades will publish immediately, irrespective of size. A SETS automatic trade report should have the following information (See Table A.4 in Appendix A).

- Identity of the reporting member firm using the Bank Identification Code (BIC).
- Date and time of the transaction.
- Whether transaction is a buy or a sell.
- Appropriate trade type indicator (e.g. 'AT' for an automatic trade and 'UT' for an auction match trade - see Table A.13 in Appendix A).
- Tradable instrument code of security.
- Security country of register.
- Dealing capacity indicator, either 'A' for agent or 'P' for principal.
- The number of shares traded.
- The transaction price in the currency that the security is quoted or if not quoted the denominated currency of the security.
- The converted currency indicator.
- Settlement due date.
- The unique transaction identifier entered in either the buy or sell client reference field depending on whether the transaction is a purchase or a sell.
- The BIC for the counter party to the transaction when dealing with another member firm, or the text 'NONMEMBER' if dealing with a non-member

*Transaction Reports*

Transaction reports are confirmation reports entered by both parties to a trade and must be submitted to the LSE by SETS participants before the end of the day on which the

[39] Trades conducted away from the order book by member firms are required to report manually. Trades should generally be reported to the Exchange immediately after execution within 3 minutes of trading for business executed during the market opening hours. Trades executed after the market closes and before it opens again should be reported before the opening auction.

transaction was effected.[40] A SETS transaction report should have the following information (See Table A.5 in Appendix A).

- Identity of the reporting member firm using its recognised BIC, participant code or settlement agent code.
- Date and time of the transaction.
- Whether transaction is a buy or a sell.
- The International Securities Identification Number (ISIN).
- The dealing capacity.
- The number of shares traded.
- Settlement due date.
- Identity of counter party.
- Price and currency of the transaction.

This section described SETS trade and transaction reports. The next section will draw a short summary of this chapter.

## 2.7 Summary and Conclusions

Trading often involves brokers, dealers or market makers. Brokers act as agents on behalf of traders and arrange trades. By contrast, dealers or market makers trade on their own behalf and provide liquidity to the market.

Trading can be carried out continuously or periodically. Recent developments in stock markets around the world suggest that traders prefer continuous trading mechanisms with opening call market auctions to exclusive call markets. Today all major stock markets, such as the TSE and NYSE, use periodic call auctions to open continuous trading sessions.

[40] The LSE requires transaction reports for the purpose of maintaining an audit trail of ‘on Exchange’ business conducted by member firms. This is necessary for the LSE to meet its regulatory obligations in accordance with the ‘Recognition Requirements Regulations’, paragraph 4(1) of the ‘Financial Services and Markets Act 2000’. Transaction reports are also used by the LSE for the production of market statistics for publication, and monitoring the quality of the markets that it regulates.

Trading is organised differently around world markets and across financial assets. Three trading mechanisms used in financial markets are introduced in this chapter: a quote-driven market, an order-driven markets and a brokered market. A quote-driven market, also known as a dealer market, is where dealers are the market makers and quote prices. An order-driven market, such as the TSE, is where trading can be arranged without the intermediation of dealers. A brokered market, very common for large blocks of stocks trading, is where brokers offer intermediary search services to traders and facilitate large trades.

A hybrid market mixes characteristics of all three types of markets mentioned above. Recently stock markets are converging towards a hybrid market structure. Some major stock markets, such as the NYSE and NASDAQ, all have hybrid market structures.

The LSE provides secondary markets for a wide range of stocks. It offers three trading services, including SETS, for UK stocks. SETS is the central price formation and trading mechanism for stocks including FTSE 100 index constituents. It allows participants to submit orders to buy or sell quantities of shares at specific prices. Executions occur automatically in accordance with strict price then time priority.

A SETS trading day starts with an opening auction followed by a period of continuous trading and ends with a closing auction where an official closing price is set. If the price tolerance levels are reached or breached during continuous trading, continuous trading will be suspended and an AESP intra-day auction will occur. An EDSP intra-day auction takes place on the third Friday of each month for FTSE 100 index constituents to prevent increased volatility when quarterly and monthly FTSE 100 index derivatives expire.

SETS allows participants to use the following types of anonymous order: market, limit, iceberg, at best, execute-and-eliminate and fill-or-kill. Only limit, iceberg and market orders are eligible to be entered during the auction period. Only limit, iceberg and market orders may sit on the order book during the continuous trading period. Other order types are either immediately executed or rejected.

The SETS order book is a queue of all unexecuted orders for subsequent trading. Orders arrive at the order book randomly over time. A new order is compared to existing ones. Executions will only occur when the price is matched.

SETS has three segments: SET1, SET2 and SET3, within each of which there are a number of sectors. SET1 includes the FTSE 100 stocks while SET2 includes the FTSE 250 stocks. SET3 includes the remaining SETS stocks.

SETS participants are obligated to supply trade and transaction reports. Trade reports are reports of the detail of transactions effected at the LSE. Transaction reports are confirmation reports entered by both parties to a trade.

Key features of SETS are summarised as follows.

- All orders are firm.
- Orders are always displayed and executed according to price then time priority.
- When a pending limit order sitting on the order book is partially executed, its remaining part keeps the same time priority as the original one.
- Amendments to orders can be achieved by order cancellation or replacement. A new order is given a new time stamp and appended to the back of the queue of orders that have the same price.
- All trades that result from order execution on the order book generate automatic trade reports that are published immediately.
- The market as a whole, member firms, trading participants and institutions can view the entire order book. There are no specialist participants or privileged information flows.

This chapter describes trading mechanisms of stock exchanges, with particular emphasis on those of the LSE. The next chapter will review related market microstructure literature.

# CHAPTER 3 – LITERATURE REVIEW ON MICROSTRUCTURE

## 3.1 Introduction

Over the past decade, there has been a significant growth in market microstructure studies. Academics define market microstructure in different ways. Madhavan (2000) defines it as an area of finance that studies the process by which investors' latent demands are ultimately translated into prices and volumes. O'Hara (1995) defines market microstructure as a study of the process and outcomes of exchanging assets under a specific set of rules. Stoll (2003) sees market microstructure as an area of study that deals primarily with market price as reflected in the best bid-offer spread. In general, market microstructure studies focus on the mechanisms used for trading financial securities and the process of price formation.

Several leading researchers have provided excellent review of market microstructure literature. O'Hara (1995) is the first overview. Madhavan (2000) provides a comprehensive review including the theoretical, empirical and experimental literature relating to price formation, market structure and design, and transparency. Biais, Glosten and Spatt (2005) provide a coverage of subsequent advances in market microstructure theory. Hasbrouck (2007) shows tests and methods used in empirical market microstructure studies. Harris (2003) reviews market microstructure studies for practitioners.

Some research also reviews subtopics of market microstructure literature. For example, Stoll (2003) focuses on the transaction cost aspect of market microstructure; Calamia (1999) focuses on the impacts that a trading mechanism might have on price behaviour.

One of the main research interests in market microstructure is trading costs. They play an important role in determining the returns for investors and traders. In general, trading involves two types of costs: explicit and implicit costs. Explicit costs, such as brokerage fees and transaction taxes, are easily identified and measured. In contrast implicit costs, such as market impact and timing costs, are difficult to identify and measure. The timing costs depend on the time of execution of orders. One of the aims of this thesis is to investigate the characteristics of limit-order completion time in the UK market. This investigation will be carried out in subsequent empirical chapters.

Harris and Hasbrouck (1996) point out that one of the outstanding questions in market microstructure is how to optimise order submission strategies in an order-driven market. Developing these strategies is still in its infancy. This thesis develops models of limit-order completion that are the foundation of many order submission strategies (see Section 3.4). If limit-order completion time and completion probability can be modelled, then it may become possible for traders to predict them, at least partially, and hence reduce the risk of non-execution and the probability of the order adversely executing before it is cancelled if the stock's value moves past the limit-order price. This would enable traders to optimise their order submission strategies.

Many stock markets are organised as order-driven markets. Thus exchanges are also interested in reducing limit-order completion time and in facilitating limit-order execution in order to attract businesses. This thesis also attempts to identify the determinants of limit-order completion time, which would certainly be of interest to exchanges. Thus, this study provides some implications, which will be discussed in Chapter 6, for exchanges to facilitate limit-order completion.

Some market microstructure studies focus on trading mechanisms. Madhavan (1992), Brennan and Cao (1996), Biais, Martimortand and Rochet (2000), Maslov (2000), Slanina (2001), Viswanathana and Wang (2002), Parlour and Seppi (2003), Daniels *et al.* (2003), and Li Calzi and Pellizzari (2003) model and compare different trading mechanisms. Amihud and Mendelson (1986), Jain (2002), Kalay, Wei and Wohl (2002) and Bottazzi, Dosi and Rebesco (2002) have empirically investigated different trading mechanisms. Reviewing the literature on trading mechanisms, however, is not the aim of this chapter. This chapter, instead, concentrates on a portion of market microstructure literature that investigates the limit-order book, order submission strategies, limit-order execution time and probability. The review of literature is not restricted to this chapter with subsequent empirical chapters containing reviews of more specific background and motivational literature.

This chapter follows a 'top-down' approach and the presentation proceeds as follows: Section 3.2 reviews studies on the limit-order book. Section 3.3 reviews studies on order submission strategies. Section 3.4 reviews studies on limit-order execution time and probability. Section 3.5 draws a conclusion.

## 3.2 Limit-Order Books

An order-driven market usually features a central limit-order book that contains essential information. This section reviews studies on a limit-order book. The next subsection reviews literature on the information content of a limit-order book. Subsection 3.2.2 reviews both theoretical and empirical studies on the statistical properties of a limit-order book. Subsection 3.2.3 draws a short conclusion.

### *3.2.1 Information Content of a Limit-Order Book*

A limit-order book is a central mechanism in an order-driven market. It provides traders with valuable market information, such as the best bid-offer spread and the number of limit orders stored on the order book. This market information could potentially affect traders' order submission strategies and consequently limit-order completion time.

The NYSE opened its limit-order book to off-exchange traders during trading hours in January 2002. This has motivated recent studies on the information content of a limit-order book. Baruch (2005) models a market similar to the auction with which the NYSE starts the trading day. He considers two different scenarios: only the specialists can see the limit-order book and the information on the book is available to all traders. He shows that the price set in the opening auction reveals more information (implying lower post-open volatility) when the book is open to all traders. Boehmer, Saar and Yu

(2004) further investigate some issues raised by Baruch (2005). They study pre-trade transparency by looking at the move to opening the limit-order book using 1,332 stocks data from the NYSE. They find an increase in displayed liquidity on the book and a decline in the price impact of market orders following the introduction of an open limit-order book.

The TSE opened its order book in April 1990. Madhavan, David and Weaver (2000) examine market transparency when the TSE publicly disseminated its limit-order book on both its traditional floor and automated trading system. The data sample used in their study is drawn from the TSE equity history files for the months of February through June 1990. These files contain every trade and quote with associated prices, volumes, and best bid/offer sizes. In contrast to Boehmer, Saar and Yu (2004) they find that the increase in transparency reduces liquidity and that volatility increases after the limit-order book is publicly displayed.

The NYSE is a hybrid market (order-driven/quote-driven) where specialists are obliged to quote prices when it is necessary, but the TSE is a pure order-driven market. This difference between the two markets could potentially be the reason behind the contradictory findings. Nevertheless, these studies show that opening an order book could potentially change traders' behaviour, which then affects limit-order completion.

The content of a limit-order book could also explain price fluctuations. Farmer *et al*. (2004) investigate the causes of large price fluctuations at the LSE using data on 16 stocks over a four-year period (1999-2002). They show that large price fluctuations are unrelated to large transactions or to placements of large orders. Instead, these large price fluctuations are driven by changes in the number of limit orders at a particular price. This finding is not specific to the LSE. Weber and Rosenow (2006) study large price fluctuations of the NYSE stocks. They find that a low number of limit orders stored on the order book causes extreme price fluctuations. Bouchaud *et al*. (2004) investigate price fluctuations using transaction data from the French stock market and provide evidence suggesting that price fluctuations are to a large extent induced by trading activities. Any trading activities would have changed the number of limit orders on the order book. Thus, their finding is consistent with Farmer *et al*. (2004) and Weber and Rosenow (2006). Accordingly, the number (size) of limit orders on the order book could potentially affect traders' order submission strategies and hence limit-order completion.

Studies on the information content of a limit-order book have also concentrated on future trading activities in the market. Harris, Lawrence and Panchapagesan (1999) examine whether the limit-order book is informative about future price changes. Using limit-order data in the TORQ database, they find that the limit-order book is informative, especially about short-term price changes.[41] Corwin and Lipson (2000) study order flows and liquidity around the NYSE trading halts and examine the role of the limit-order book in determining post-halt prices. They find that a price set at the reopening of the limit-order book is a good estimate of future price. Irvine, Benston and Kandel (2000) use data from the TSE to show that the content of a limit-order book can be used to predict subsequent trading activity (number of trades). Thus, a limit-order book impounds forward-looking information about future price volatility and order flows. This information can be used to predict completion times of future limit orders. In Chapter 6 the effect on limit-order completion time of explanatory variables that are designed to capture the state of an order book, such as the best bid-offer spread, are investigated.

### *3.2.2 Statistical Properties of a Limit-Order Book*

A limit-order book offers an extensive source of detailed data on the collective behaviour of interacting traders. Thus order book statistics are often considered as ultimate descriptions of stock markets. Traders, then, are interested in a complete description of the limit-order book, as they base their order submission strategies on the state of the limit-order book.

Some researchers have focused on the relationship between order submission strategies and the state of the limit-order book. For example, Biais, Hillion and Spatt (1995) find that in the French stock market order submissions are contingent on the state of the

[41] The TORQ database contains transactions, quotes and order processing data for a sample of 144 NYSE stocks for a three-month period (from November 1990 to January 1991).

order book. They find that traders intend to place orders aggressively when the best bid/offer sizes are large or when the best bid-offer spread is large in order to gain execution priority. Therefore, the state of a limit-order book would affect traders' order submission strategies and hence limit-order completion.

Some research has focused on the statistical properties of a limit-order book. Bouchaud, Mezard and Potters (2002) investigate several statistical properties of order books of three liquid stocks on the French stock market. They find that the price at which new buy (sell) limit orders are placed is very broadly distributed around the current best bid (offer) price, and that the best bid (offer) size follows a Gamma distribution. Zovko and Farmer (2002) demonstrate a striking regularity in the way traders place limit orders. They use a dataset that consists of roughly seven million orders from the LSE and find a fat-tailed distribution of buy (sell) limit-order price around the best bid (offer) price. Challet and Stinchcombe (2001) study the statistical properties of the NASDAQ order book and find that the order execution time has a fat-tailed distribution. Potters and Bouchaud (2003) use data from the NASDAQ to further investigate statistics of incoming limit-order prices and a typical lifetime of a buy (sell) limit order as a function of the distance from the best bid (offer) price. They show that the lifetime of a given buy (sell) order increases as it moves away from the best bid (offer) price. Thus, the position of a buy (sell) limit order from the best bid (offer) price could potentially affect its completion time. This point is taken into account in the empirical investigation on limit-order completion time in Chapters 5 and 6.

Lillo and Farmer (2004) study signs (buy or sell) of orders submitted at the LSE. They show that the signs of future orders are quite predictable given the signs of past orders. Order splitting and correlated trading on private information could be the reasons behind this finding. Biais, Hillion and Spatt (1995) put forward further explanations of these 'correlated' orders. First, if different traders are imitating each other, the cause of the correlation of order sign is the order submission itself. Second, traders could react similarly to the same events that are related to a particular stock or the economy as a whole.

Some literature also reports findings that orders with a particular level of aggressiveness (the distance between the buy (sell) limit-order price and the best bid (offer) price) tend to be followed by similar orders. For example, Griffiths *et al*. (2000) study determinants of order aggressiveness at the TSE and find that aggressive buy (sell) orders tend to follow other aggressive buy (sell) orders when the best bid-offer spreads are narrow and the depth on the same (opposite) side of an order book is large (small). Thus the best bid-offer spread and depth of the order book could potentially affect order submission strategies and hence limit-order completion time. In Chapter 6 this point is taken into account in the empirical investigation on limit order completion time.

### *3.2.3 Summary*

A limit-order book provides information to traders. It is informative about future trading activities. Thus, it is possible to predict limit-order completion time and probability with the information provided by the order book.

The statistics of the order books are the ultimate descriptions of stock markets and have been investigated. The statistics indicate potential variables, such as the best bid-offer spread, which could affect limit-order completion time. In Chapter 6 the effects on limit-order completion time of these variables are investigated.

## 3.3 Order Submission Strategies

Traders can choose different orders with different prices and sizes to optimise submission strategies.[42] They must also decide when to submit market orders and when to submit limit orders. When they submit limit orders, they must also know what price levels to place their limit orders at, and when to modify their orders.

[42] Kissell and Glantz (2003) have described a comprehensive top-down approach to evaluate trading costs and optimise order submission strategies. Researchers also develop order submission strategies in some special cases. For example, Coggins, Blazejewski and Aitken (2003) develop an approach for optimising trade execution in a limit-order market when the trade is a significant proportion of the day's turnover; Almgren and Chriss (2002) construct optimal strategies of trading through scheduled news events such as earnings announcement.

The decision whether to submit a limit order or a market order involves a trade-off between the expected profit and the option value of a limit order that is given away to market participants. The expected profit depends on the probability of a buy (sell) limit order being executed and the difference between the order price and the prevailing best offer (bid) price assuming that liquidity is available. The option value of a limit order depends on the probability of arrival of adverse information, which moves market price against the submitted limit order, and the expected order completion time, which also depends on the probability of order execution.

When developing order submission strategies, some traders are more concerned about execution price while others are more concerned about execution time. Traders who face deadlines (impatient traders) tend to develop strategies to minimise the waiting time to complete orders. These traders are more concerned about time rather than price. In contrast, traders who are more sensitive to price levels are more willing to wait until favourable prices arise (patient traders); they tend to develop strategies to optimise the completion price rather than time. For both types of traders, the decision to place and modify orders depends on the expected limit-order completion times. Tkatch and Kandel (2006), for example, provide evidence showing that investors care about the expected execution times when trading equities on the Tel Aviv stock market. One of the main aims of this thesis is to investigate limit-order completion time and identify the potential variables that could affect the expected waiting time. This investigation will be carried out in subsequent empirical chapters.

This section reviews literature on order submission strategies and related empirical studies. Subsection 3.3.1 reviews theoretical studies on order submission strategies. Subsection 3.3.2 reviews empirical studies on order submission strategies. Subsection 3.3.3 draws a short conclusion.

### *3.3.1 Order Submission Strategies*

Some research focuses on order submission strategies in a market where orders are aggregated across multiple investors and executed simultaneously in one round of trade. Mendelson (1982), for example, studies order placement strategies in this type of market. He points out the trade-off between the price of a limit order and the execution time: the higher (lower) the price of a sell (buy) order, the lower the probability of execution and the longer the execution time. This thesis attempts to investigate this trade-off in a continuous order-driven market. This investigation will be carried out in Chapters 5 and 6.

Other research focuses on order submission strategies in a market where traders arrive and submit orders one at a time. Kumar and Seppi (1993) find that traders in this type of market can use market orders and schedules of limit orders to optimise their trading strategies. Chakravarty and Holden (1995) develop an integrated model of order submission strategies in which a risk-neutral informed trader optimally chooses any combination of a buy market order, a sell market order, a buy limit order, and a sell limit order, for this type of market. They find that combining a buy market order and a sell limit order is more profitable than a buy market order only. Seppi (1997) shows that traders' order choices in this type of market depend on their prediction about the arrival of a market order, since existing limit orders will be executed against incoming market orders.

Some research also attempts to study order submission strategies in a continuous market where traders can arrive at random times and submit orders at anytime. Research into such strategies is still in its early stages and further studies are needed. Harris (1998) develops a model of order submission strategies for this type of market and shows that a limit order should be placed close to the market price to maximise its execution probability. Bertsimas and Lo (1997) show that in this type of market the common approach of breaking up a large trade into a number of small trades of equal size is optimal only in a special case, and that is when price impact is linear in the trade size and permanent in its effects on future prices, and prices follow an arithmetic random walk. Foucault (1999) develops a theoretical model of order submission strategies for this type of market. He assumes that traders can choose to place either limit or market orders and shows that limit orders result in better execution prices but face a risk of non-execution and a winner's curse problem. Risk of non-execution arises when a market moves against a limit order, and the winner's curse problem occurs when a limit order is adversely picked up before it can be cancelled when the stock's value moves past the limit-order price.

### *3.3.2 Empirical Studies on Order Submission Strategies*

Extensive empirical research has been carried out on order submission and execution strategies. Some examine the interaction between order flows and the state of an order book; others study the relation between price movements and the state of an order book; some focus on the dynamic relation between volatility and the order flows; and others investigate order submission strategies of more specific types of traders, such as informed traders.

Most empirical studies on order submission strategies focus on the interaction between order flows and the state of the NYSE order book. Harris and Hasbrouck (1996) study performance measures for market and limit orders of a sample of day orders covering 144 randomly chosen stocks traded on the NYSE during the period from November 1990 through January 1991.[43] They suggest that buy (sell) limit orders placed at or better than the prevailing best bid (offer) price perform better than market orders. Bae, Jang and Park (2003) also examine traders' choice between market and limit orders using a sample of orders submitted to the NYSE. They find that traders place more limit orders relative to market orders when the best bid-offer spread or the order size is large. Wald and Horrigan (2005) use a probit model and 100 stocks data from the S&P 500 in the 1990 to study order submission strategies from the perspective of a risk-averse trader. They provide a solution to the trader's decision of whether or not to place a limit order, and at what price, given individual beliefs and stock characteristics. They show how the limit-order price and the best bid-offer spread affect whether or not a limit order executes. Thus the best bid-offer spread, the order size and price may affect order submission strategies and hence limit-order completion. In Chapter 6 this point is taken into account in the empirical investigation on limit-order completion time.

The determinants of order choice also attract the attention of research. Ellul *et al.* (2003) examine determinants of order choices using a sample of 148 stocks traded on the NYSE during 20 trading days prior to 29 January 2001. They estimate a multinomial logit model of order choice using the existing theoretical literature to

[43] Harris and Hasbrouck (1996) employ two performance measures. The first measure compares the execution price of a buy (sell) order to the prevailing best offer (bid) price when the order is submitted, and the second measure compares the execution price of a buy (sell) order to the prevailing best bid (offer) price after the execution.

suggest variables influencing traders' order submission strategies.[44] They also estimate the likelihood of observing the arrival of different types of orders, that of cancelling of previous orders and that of no order activity. They find that wider (narrower) best bid-offer spreads increase the probabilities of limit (market) orders arrival. Since existing limit orders will be executed against incoming market orders, this finding implies that best bid-offer spreads may also affect limit-order completion. Caglio and Beber (2005) examine order submission strategies using a three-month (from November 1990 through January 1991) sample of ten high transacted NYSE stocks and find that some microstructure-related and market-related variables representing a trader's information set significantly affect a trader's decision to submit an aggressive versus a passive order. They show that the depth on the same side of the book is one of the most important determinants of an order choice. Knez and Ready (1996) examine price improvement, which is the difference between execution price of a buy (sell) order and best offer (bid) price when an order is submitted. Using a dataset covering a period of three month for a sample of 144 NYSE stocks and nonparametric regression techniques they show that conditional expected price improvement is strongly and nonlinearly related to the difference between best bid (offer) size and sell (buy) order size, which indicate the difference between liquidity available and liquidity demanded. LMZ (2002) suggest some explanatory variables to capture this difference, an aspect that is dealt with in more detail in Chapters 5 and 6.

There is also empirical evidence on order submission strategies from stock markets other than the NYSE. Biais, Hillion and Spatt (1995) study order flows of the Paris Bourse, which is a computerised pure order-driven market, using a dataset containing the history of order books for the 40 stocks in the CAC 40 index for 19 trading days between 29 October and 26 November 1991.[45] They find that order submission strategies depend on market conditions, such as the best bid-offer spread. This is consistent with Bae, Jang and Park (2003). Cao, Hansch and Wang (2004) use a probit model to investigate the connection between the state of an order book and order submission strategies using a dataset from the Australian Stock Exchange (ASX) covering 21 constituent stocks of the ASX 20 index for the month of March 2000. They find that traders use available information on the order book when developing their

[44] This study also uses the existing theoretical and empirical literature to suggest variables influencing limit-order completion time. See detail in Chapter 6.

[45] The Paris Bourse (or 'Bourse de Paris' in French) is the historical Paris stock exchange, known as Euronext Paris from 2000 onwards.

order submission strategies. Verhoeven, Ching and Ng (2004) use a logit model to examine the determinants of order submission strategies using data on two stocks from the ASX (from 1 January to 30 June 1997). Their results indicate that best bid-offer spread, best bid/offer size, price changes in the last five minutes, and order imbalance are major determinants of a trader's decision whether to place market or limit orders. These variables can also potentially affect limit-order completion time and hence will be investigated in Chapter 6.

Some empirical studies focus on the dynamic relation between price movement, volatility and the state of an order book. Chan (2005) investigates order submission strategies in the Hong Kong Stock Exchange by examining the relation between the state of the limit-order book and previous price movement. He shows that traders become more aggressive in buying and more patient in selling following positive stock returns. Ahn, Bae and Chan (2000) focus on the dynamic relation between short-term price volatility and order flow. Using a data sample covering 33 component stocks of the Hang Seng Index between July 1996 and June 1997 they examine the role of limit orders in liquidity provision. They find that traders submit more sell (buy) limit orders than sell (buy) market orders when volatility has recently arisen from the selling (buying) side of the market.

Some empirical studies focus on order submission strategies used by a certain type of trader. Keim and Madhavan (1995) use a probit model and a dataset including orders submitted by 21 institutional investors from January 1991 to March 1993 to examine the selection of order type (limit or market order) by institutional traders. They find significant differences in the choice of order type across institutional styles. Their results suggest that institutional traders have a surprisingly strong demand for immediacy and tend to spread buy orders over longer periods than equivalent sell orders. Handa and Schwartz (1996) test trading strategies using 30 Dow Jones Industrial stocks for the year 1988 and find that some patient traders intend to submit limit orders while some impatient traders intend to submit market orders.

Further studies focus on order submission strategies of informed traders. Kaniel and Liu (2006) use a simple, Glosten-Milgrom type equilibrium model to study the decision of informed traders of whether to use limit or market orders. They show that informed traders prefer limit orders to market orders and that the horizon of private information is critical for this choice. More specifically, this horizon is positively related to the probability of choosing limit orders. Bloomfield, O'Hara and Saar (2005) use simulated markets to investigate the evolution of liquidity in an electronic order-driven market. They also find that informed traders use more limit orders than do uninformed traders. Anand, Chakravarty and Martell (2005) empirically examine the relative use of market versus limit orders by informed and uninformed traders using data from the NYSE. They find that informed traders tend to take liquidity earlier on in the trading day and act as liquidity suppliers later on in the day. Since traders behave differently over the trading day, the time of order submission may affect order completion time. This time-of-day effect will be investigated in Chapter 8.

Some research focuses on only marketable limit orders.[46] Peterson and Sirri (2002), for example, investigate marketable limit orders using two-week NYSE data (from 30 June to 11 July 1997). They show that marketable limit orders are used more frequently for large orders and perform worse than market orders on a trading cost basis.[47]

The change of tick size could also affect order submission strategies. Bacidore, Battalio and Jennings (2003), for example, examine changes in order submission strategies around the switch to decimal pricing for the NYSE pilot stocks and find no evidence that traders switch from traditional limit orders to market orders. They find, however, that traders cancel limit orders more frequently than before and decrease limit-order size, and that these changes in order submission strategies appear to result in less displayed liquidity on the limit-order book.

[46] A marketable buy (sell) limit order is a limit order with a limit price at or better than the prevailing best offer (bid) price. Since marketable limit and market orders are similar, they usually have the same (highest) priority for execution.

[47] The trading cost is measured by the difference between the order execution price and the midpoint of the best bid-offer spread.

### *3.3.3 Summary*

Order submission strategies have attracted the attention of researchers. Some study order submission strategies in a market where orders are aggregated across multiple investors and executed simultaneously in one round of trade; some study order submission strategies in a market where traders arrive and submit orders one at a time; and others attempt to study order submission strategies in a continuous market where traders can arrive at random times and submit orders at anytime. Although extensive research has been carried out on order submission strategies, more studies are still needed.

Extensive empirical research has been carried out on order submission strategies. Most examine the interaction between order flows and the state of an order book; some study the relation between price movements and the state of an order book; some focus on the dynamic relation between volatility and the order flows; and others investigate order submission strategies of more specific type of traders, such as informed traders. These studies suggest potential variables, such as best bid/offer sizes, which may influence limit-order completion time and, hence, will be investigated in Chapter 6.

One of the major difficulties in optimising order submission strategies is to estimate the waiting time for a limit order to complete execution and the probability of this completion. This will affect a trader's decision on when to submit a limit order and when to submit a market order. The main aim of this thesis is to model and investigate limit-order completion time, which is a step forward toward optimising order submission strategies. The next section will review literature on limit-order execution.

## 3.4 Limit-Order Execution

Understanding limit-order execution is essential to optimise order submission strategies. Predicting limit-order execution time is the foundation of optimising order submission strategies. This section reviews theoretical and empirical studies on limit-order execution. Subsection 3.4.1 reviews theoretical studies. Subsection 3.4.2 reviews empirical studies. Subsection 3.4.3 draws a short conclusion.

### *3.4.1 Theoretical Studies on Limit-Order Execution*

Some research focuses on limit-order execution. Cohen *et al*. (1981) argue that with transaction costs and a finite number of investors the market price would generally not behave as a Wiener process and hence order execution probability does not approach one as the limit-order price approaches the market price. Cohen, Conroy and Maier (1985) model the number of executed and cancelled orders as Poisson processes. They assume buy limit orders submitted at the best bid price and sell limit orders submitted at the best offer price, and use standard results from queuing theory to compute properties such as expected number of stored limit orders, expected execution time, and relative probability of execution versus cancellation. They find that there is a lower chance that a limit order will be executed when the bid-offer spread is wider. Angel (1994) investigates the conditional probability of limit-order execution on a trader's information, such as the bid-offer spread. He shows that the probability is lower when the bid-offer spread is wider. This finding is consistent with Cohen, Conroy and Maier (1985). His studies, however, are only applied to a special case: batch trading with informed traders knowing conditions of an entire limit-order book. Domowitz and Wang (1994) assume that order placement and cancellation processes are fixed in time and do not respond to changes at the best bid/offer price, and predict that buy (sell) order execution time will decrease as a function of sell (buy) order arrival rate

### *3.4.2 Empirical Studies on Limit-Order Execution*

Some research empirically investigates limit-order execution using high frequency transaction data. Some studies compare limit-order execution across different stock exchanges. For example, SEC (1997) studies limit-order fill rates and execution times, and shows that fill rates are higher on regional stock exchanges; Battalio *et al*. (2002) use an event methodology to investigate limit-order fill rates and execution times across trading venues. Their analysis shows that overall differences in limit-order fill rates and execution times between regional stock exchanges and the NYSE are small. None of these studies attempts to model limit-order execution times.

Further studies focus on the execution time of a limit order and the probability of its execution. Ranaldo (2003) uses a probit model to examine the information content of a limit-order book in a pure order-driven market with a data sample containing the transaction history of 15 stocks quoted on the Swiss stock market over a two-month period from March to April 1997. He finds that the depth of a limit-order book is a proxy for the execution probability of an incoming limit order. He also finds that the depth on the same (opposite) side of an incoming order increases (decreases) its probability of execution. Accordingly the depth of an order book could potentially affect limit-order completion time, which is an aspect that will be investigated in Chapter 6. Al-Suhaibani and Kryzanowsky (2000) study the execution time of a limit order and the probability of its execution using a sample of 267,517 orders on 56 stocks listed on the Saudi stock market for 65 trading days (from 31 October 1996 to 14 January 1997). They find that an aggressive buy (sell) limit order (above the prevailing best bid (offer) price) has a short expected execution time and a high execution probability. Thus the position of a buy (sell) limit order from the best bid (offer) price could potentially affect the completion time, which is an aspect that will be investigated in Chapter 6.

Some research further investigates the execution time of a limit order and its probability of execution. Cho and Nelling (2000) investigate the execution time of limit orders on the NYSE data. They show that the longer a limit order remains on the order book, the lower the execution probability. They also present evidence that the best bid-offer spread, price volatility, limit-order price and size are all related to limit-order execution probability. Thus these variables may also potentially affect limit-order completion time and similar effects of some of these variables will be investigated in Chapter 6. Cho and Nelling (2000), however, do not provide goodness-of-fit tests in their study. Omura, Tanigawa and Jun (2000) use transaction data to reconstruct order flows at the TSE. They use a probit model to estimate the probability of a limit order executing fully within a single trading day. They find that the execution probability of a limit order is low when the depth of the same side of the limit-order book is thick, high when the depth of the opposite side of the book is thick, low when there are two or more ticks between the best bid-offer spread, and higher for limit orders submitted during the earlier hours of the day. A limitation of Omura, Tanigawa and Jun (2000), however, is that their data sample excludes cancelled orders, which have remained on the order book for a period of time and provide valuable information about order execution probabilities. Some of the variables suggested by Omura, Tanigawa and Jun (2000) would be used in Chapters 6 and 7.[48] LMZ (2002) model limit-order execution time using survival analysis (a statistical technique analysing non-negative variables) and data from Investment Technology Group (ITG).[49] They find that the execution time is very sensitive to some explanatory variables, but not to others. Chapter 5 of this thesis will present an investigation of the degree to which explanatory variables suggested by LMZ capture the state of the UK order book and affect limit-order completion time.

### *3.4.3 Summary*

Understanding limit-order execution is important for the optimisation of order submission strategies. Some empirical studies on limit-order execution time and probability have been carried out. Among these studies, some use a probit model to investigate limit-order execution probability; others use survival analysis to study limit-order execution time.

Most empirical studies reviewed in this section have been carried out on the NYSE data. As far as we know, there has been no attempt to study limit-order completion time and the probability of its completion in the UK market. This thesis is the first attempt to study limit-order completion time in the UK market.

## 3.5 Summary and Conclusions

An order-driven market features a central limit-order book that contains forward-looking information. This information can be used to predict the behaviour of future

[48] Given the computing facilities and technical support available at the time of this study, only some of the variables suggested by Omura, Tanigawa and Jun (2000) can be extracted from the data sample used in this thesis.
[49] ITG is an institutional broker firm.

orders. The statistics of order books, which summarize this information, are the ultimate descriptions of stock markets. These statistics suggest potential variables, such as the best bid-offer spread, which would influence limit-order completion time. The effects of these variables on limit-order completion time will be investigated in Chapters 5 and 6.

Traders can choose different orders with different prices and sizes to obtain optimal submission strategies. Empirical studies on order submission strategies also indicate potential variables, such as best bid/offer sizes, which would influence limit-order completion time. The effects of these variables on limit-order completion time will be investigated in Chapter 6.

Predicting the execution time of a limit order and its probability of execution is the key challenge faced by attempts to optimise order submission and execution strategies. The execution time is also a key component of trading cost. Both traders and exchanges dislike this cost. Thus it is important to identify determinants of the execution time. Two main methodologies of modelling the execution time of a limit order and the probability of execution are reviewed in this chapter: probit models and survival analysis.

Omura, Tanigawa and Jun (2000) use a probit model to estimate the probability of a limit order executing fully within a single day. The main disadvantage of the probit model is that it cannot accommodate censored (unexecuted) orders. Omura, Tanigawa and Jun (2000) argue that traders can cancel limit orders without paying any extra fee. Hence, cancelled orders are not necessary to their study and are excluded. This ignores taking the timing cost of cancelling limit orders into account. This cost is especially high for retail traders (household traders), since it usually takes time for them to cancel an order through their brokers. Another disadvantage of the probit model is that it treats executed orders equally without accommodating the differences of their lifetimes.

Survival analysis can overcome the disadvantages of the probit model. Cho and Nelling (2000) initiate the use of survival analysis to model the probability of limit-order execution with the NYSE data. They propose a single model for both buy and sell orders and do not provide goodness-of-fit tests. LMZ (2002) also model limit-order execution time with survival analysis. Their study is more comprehensive, in that they propose separate models for buy and sell orders. Their limit-order dataset, however, comes from a single routing source, ITG.

Modelling limit-order execution time with data from stock exchanges other than the NYSE is limited. It is believed that no such attempt has been made for the UK market, and, to our knowledge, this thesis is the first one. This thesis uses survival analysis as the modelling methodology.

# CHAPTER 4 – DATA AND METHODOLOGY

This chapter contains descriptions of the database, selected data sample and methodology used in this thesis. Section 4.1 introduces the database and data sample. Section 4.2 introduces the methodology.

## 4.1 Data

The database used in this thesis is obtained from the LSE. The next subsection contains a description of this database. Subsection 4.1.2 presents a description of the sample selection process. Subsection 4.1.3 presents descriptive statistics of the data sample. Subsection 4.1.4 contains a short summary.

### *4.1.1 Database*

The database contains transaction data covering a period of three years from September 1998 to August 2001. It is a collection of files containing all trades, orders and quotes for the LSE traded by its members. It is published by the LSE on CD-ROMs and the data for each month are on two CD-ROMs. All data files are in delimited text format. Overall the database has a size of about 40 GB. Appendix A presents more detail of these files with associated field descriptions and key elements.

The data for each month are converted into Microsoft Access 2003 format and duplicate observations are deleted. Several attempts are made to analyse and extract data from the database with Microsoft Access 2003. None of them were successful, as Microsoft Access 2003 imposes 2 GB limit. Hence Microsoft SQL Server 2000 is used to analyse the database and extract the data.[50]

[50] It took about 40 hours to load the data into Microsoft SQL Server 2000.

This analysis uses four data files from this database, which are described in detail in Table A.1 to Table A.4 in Appendix A. The following is a summary of the data files:

- *BP (Best Price) file:* this contains information about the best price for SETS stocks. It contains the following variables: TimeStamp (date/time at which best bid/offer prices change); BestBidPrice (the highest price at which a stock may be bought); BestOfferPrice (the lowest price at which a stock may be sold); MidPrice (market price based on (BestBidPrice + BestOfferPrice)/2); CountryOfRegister (a code that specifies the country of register for a stock) and TradableInstrumentCode (a code that uniquely identifies a stock with the CountryOfRegister). SETS can manage around ten 'order events' (e.g. submit, execute, cancel) per second, and the system updates the best bid/offer prices after each event. Accordingly the best bid/offer price could theoretically change more than once per second. As a result, it is possible for a number of best bid/offer prices to exist at the same time stamp. This file does not provide the best bid/offer size. The best bid/offer size, however, can be extracted from Order History file described below.[51]

- *SNITS (New Order Entry) file:* this contains detail of new orders entered including the price, size, time and side indicators (buy or sell). It includes TradableInstrumentDisplayMnemonic (four character code allocated by the Exchange to identify a stock); TimeStamp (date/time of order entry); OrderSize; BuySellIndicator (buy or sell); LimitPrice; OrderType ('LO' for limit order); and OrderCode (a unique code allocated to each order by the Exchange that can be linked to OrderCode in Order History file described below). This file also contains TradableInstrumentCode and CountryOfRegister.

- *OHe (Order History) file:* this file documents every event that happens to an order from when it is submitted to the time it is removed from the book either through completion, cancellation or expiration. It includes TimeStamp (date/time of order event), OrderType and BuySellIndicator (buy or sell). It also includes HistType (a code that indicates order events: 'N' for new orders; 'M' for fully executed orders; 'P' for partially executed orders; 'D' for cancelled orders and 'E' for expired orders); OrderCode (as described above); TradeCode (a code assigned by the Exchange to uniquely identify executed trades that can

[51] A code written in Structured Query Language (SQL), which is a standard interactive and programming language designed for the retrieval and management of data in a database, is used to extract the best bid/offer size from Order History file.

be linked to TradeCode on Trade Report file described below); TradeSize (the number of shares that are traded as part of an automatically executed trade, or in the case of a limit order the number of shares cancelled or expired); and AggSize (the number of shares for a new order and the number of shares that are on the order book after each subsequent event).

- *TRe (Trade Report) file:* this contains information about the trades that occur during a trading session. It includes TimeStamp (date/time of trade); TradePrice; TradeQuantity; TradeCode (as described above); TradableInstrumentCode; and CountryOfRegister.

These data files can be combined to extract information on market conditions and the specification of each limit-order. All limit orders are divided into two categories: censored and completed orders. Censored orders are those that are cancelled or expired. Censored observations include censoring time, which is defined as the difference between the time at which cancellation or expiration occurs and the time of submission.[52] Completed orders include all fully completed limit orders. Associated observations include completion time, which is defined as the difference between the time at which the final execution occurs and the time of submission.

The database contains some undesirable peculiarities that are due to either the characteristics of the data (e.g. time stamps are in seconds) or the retrospective compilation of data. For example, there are multiple best bid/offer prices at some time stamps,[53] and for some limit orders documented submission times are slightly different in the SNITS file than in the OHe file.[54]

It is impossible to use the whole database in this study, due to its size and limited resources.[55] Instead, a sample has been selected. The next subsection will explain the process to select this sample.

[52] Cancelled orders include partially executed and eventually cancelled limit orders. In this section, 'censored' means 'cancelled' or 'expired'.
[53] In these cases, the best prices (i.e., the highest bid price and the lowest offer price) are selected.
[54] In these cases, the submission time shown in the SNITS file is selected.
[55] To analyse such a large database requires professional technical support, which is not available at the time of study.

### *4.1.2 Sample Selection*

Initially this analysis intends to use three-year FTSE 100 stocks data. The available computing facility, however, can only process three-month FTSE 100 stocks data.[56] The criteria to select three consecutive months are as follows.

- The three-month data need to be as recent as possible. Thus they are chosen from the latest 12-month data in the database (from September 2000 to August 2001).
- Possible low volumes of thin market effects during summer months are avoided by excluding July, August and September.
- The UK stock market plunged significantly in 2001. In order to avoid this 'bear' market effect, the year of 2001 is excluded.

Accordingly, October, November and December 2000, which are the only remaining possibilities, are chosen as the sample period. Out of the roughly 100 stocks that constitute the FTSE 100 index, 58 survived the period from September 1999 to August 2001. Only 43 of these did not experience any major corporate event, such as a share split and change in market sector, during the sample period. Only 38 of these 43 stocks have data available on DataStream for daily closing price, daily share volume and daily number of shares outstanding.[57] The sample is then confined to data on these 38 stocks.

The steps to select a data sample from the database are described as follows. First, the data files for the three chosen months are loaded on Microsoft SQL Server 2000. Second, an SQL code is written to trace each limit order from submission to completion, cancellation or expiration, and extract information required for the constitution of certain variables that will be explained in Chapters 5 and 6.[58] Third, a final data file is exported and saved in a text format.[59] Finally, STATATRANSFER 8.0 is used to load

---

[56] The computer used in this study has the following specifications: Intel Pentium 4 processor (2.8 GHz), 3 GB memory and 300 GB Hard Drive.

[57] These variables are used in Chapter 5.

[58] Appendix B illustrates a sample of the main programs used, whose running time is about 500 hours.

[59] The data file includes the following information: limit-order code; time stamp of submission; time stamp of completion/censoring; order completion (censoring) time (in seconds); limit-order price; limit-order size; best bid/offer price; best bid/offer size; the number of trades in the last 30 minutes; and the number of trades in the last one hour.

the data file into STATA 8.0 (statistical software). The data sample used in this thesis is extracted from this data file by applying the following filters.

- Only buy and sell limit orders submitted between 09:00 and 16:30 are included as some explanatory variables used later in Chapter 5 are based on the number of trades that occur in the hour preceding the time of order submission. Orders are concentrated around the open of the market (c.f., Jain and Joh (1988), McInish and Wood (1990, 1991), Gerety and Mulherin (1992) and Foster and Viswanathan (1993)). Since a large number of small orders are executed at the market opening as Al-Suhaibani and Kryzanowsky (2000) show, the exclusion of orders submitted in the first hour of trading could potentially affect the empirical results of this study. For example, the low sensitivity of limit-order completion time to the limit-order size discussed later could be due to the exclusion of small orders. In addition, the exclusion of order submitted before 09:00am could also have an impact on intra-day patterns discussed in Chapter 8. For example, more orders are submitted in the afternoon than in the morning as Chapter 8 reveals. This could be due to the exclusion of these orders.

- The sample is restricted to day limit orders that are completed, cancelled or expired during the same day of submission.[60] Only 2% of limit orders in this data file are completed or censored immediately after submission and hence have zero completion or censoring times. In most cases, these orders have prices equal to or better than the best bid/offer prices. Accordingly, they behave as market orders. For this reason they are excluded from the data sample.

- When FTSE 100 index derivatives expire, there is an intra-day auction on stocks in SETS. In order to avoid the effects of this intra-day auction, this study excludes limit orders submitted, completed, cancelled or expired between 10:00 and 10:30 on the third Friday of every month.

- This study also excludes executed orders during AESP auction.[61]

- The median distance between the mid-quote price $P_q$ and the limit-order price $P_l$ of all limit orders in this data file is 0.5%, while the 99th percentile is 2.5%.[62]

[60] About 98% of limit orders in this data file are day limit orders.

[61] If the price tolerance levels are reached or breached during continuous trading period, continuous trading will be suspended and an Automatic Execution Suspension Period (AESP) intra-day auction will occur. See Section 2.3 of Chapter 2.

[62] The distance between the mid-quote price and the limit-order price is: $|(P_q - P_l)/P_q|$.

Accordingly, limit orders that are more than 2.5% away from the mid-quote prices are treated as outliers and excluded from this data sample.

- This study excludes any limit order of a size greater than 250,000 shares. The median of the limit-order size in this data file is 1,000 shares, while the 99th percentile is 250,000 shares. Accordingly, limit orders that have a size greater than 250,000 shares are treated as outliers and excluded from the data sample.
- Some best bid/offer sizes and best bid/offer prices are recorded as zero.[63] The sample excludes associated orders.

1,392,565 out of a total of 1,510,789 limit orders pass the filtering criteria. They constitute the sample used in this thesis. The next subsection contains a description of this data sample.

### *4.1.3 Summary Statistics of the Data Sample*

Table 4.1 reports summary statistics of the limit-order data sample, which is almost evenly split between buy orders (51.31%) and sell orders (48.69%) for the pooled sample of 1,392,565 limit orders.[64] 41.39% of the orders have at least one fill. 23.91% are completely filled on the first execution (first fill) while 12.01% are completed with multiple fills. Overall, only about 36% of limit orders are completed while the rest are censored. The low completion rate is not specific to the LSE. Hasbrouck and Saar (2004) report that more than 80% of all limit orders submitted on Island ECN in the last quarter of 1999 did not execute and were ultimately cancelled.[65] In contrast, LMZ (2002) find over 80% of orders in their data sample experience at least one fill.[66] The above studies only use data from a small subset of the universe of transactions in listed American securities (Island ECN and ITG). Thus the low/high execution rate could very well reflect the characteristics of their data sources rather than the markets.

[63] An SQL program is used to extract the best bid/offer sizes. After extractions, some limit orders have zero best bid/offer sizes, zero best bid/offer prices or both.
[64] Traders submit more buy than sell orders. This evidence could be due to the bull market that characterises the sample period. This finding is not specific to the LSE. Biais, Hillion and Spatt (1995) and Griffiths *et al.* (2000) also find that there are more buy orders in their data samples. Table C.1 in Appendix C reports the summary statistics for each individual stock.
[65] Their sample includes 300 largest NASDAQ stocks based on market capitalisation of 30 September 1999. Island ECN is a trading platform for NASDAQ stocks.
[66] Their sample is provided by a single institutional broker (ITG).

Another reason behind the difference in order completion rates across these samples could be the difference in transparency across these markets. Boehmer, Saar and Yu (2004) find that when the NYSE began to open its order book to off-exchange traders, the cancellation rate of limit orders increased by 17%. This suggests that the more transparent a market is, the higher the order cancellation rate and the lower the order completion rate.

**Table 4.1: Summary Statistics of the Data Sample**

This table contains summary statistics of the limit-order data (October 2000-December 2000) for a pooled sample of 38 stocks (Pool). The 'partially filled' category includes orders that experience at least one fill. Around 42% of submitted orders experience at least one fill. Around 24% are completed with one fill and 12% are completed with multiple fills. Hence around 6% (42%-24%-12%) experience at least one fill but are cancelled or expired eventually.

| | **No. of orders** | **% buy** | **% sell** | **% partially filled** | **% completed with one fill** | **% completed with multiple fills** |
|---|---|---|---|---|---|---|
| **Pool** | 1392565 | 51.31% | 48.69% | 41.39% | 23.91% | 12.01% |

Table 4.2 reports the percentage of executed limit orders that are completed with a given number of fills.[67] 66.56% of completed orders are completed with one fill, 20.40% are completed with two fills while 0.95% of completed orders require seven or more fills. Thus around 87% of completed orders require no more than two fills. LMZ (2002) find that over 92% of completed orders, which include around 81% that are completed with the first fill, require no more than two fills and 1% require seven or more fills. Their data sample is, however, more representative of a specific trading clientele than the NYSE. In contrast to LMZ (2002), this analysis finds that at the LSE, as presented by Table 4.2, over 33% of completed limit orders require more than one fills. Large size of limit orders for FTSE 100 stocks could be the reason behind the high number of fills required for completion.

[67] Table C.2 in Appendix C reports the percentage breakdowns of completed orders for each individual stock.

**Table 4.2: Percentage Breakdowns of Completed Orders**

This table contains the percentage breakdowns of completed orders by the number of fills required for completion, for a pooled sample of 38 stocks (Pool). The sample period is from October 2000 to December 2000.

| | No. of fills | | | | | | |
|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **>=7** |
| **Pool** | 66.56% | 20.40% | 7.28% | 2.97% | 1.36% | 0.65% | 0.95% |

Table 4.3 reports the mean and standard deviation of completion and censoring times in minutes.[68] The mean of completion time is 9.16 minutes for buy orders and 9.47 minutes for sell orders. The 'no fills' columns contain limit orders that are not executed and the 'partial fills' columns contain orders that are partially but not completely filled. The average censoring time for 'no fills' limit orders is 7.59 minutes for buy orders and 7.53 minutes for sell orders.[69] The mean of censoring time for 'partial fills' is 16.64 minutes for buy orders and 16.18 minutes for sell orders.[70] One consistent trend is that the average censoring time for 'partial fills' is longer than the average completion time. It could be due to the fact that the partially filled orders are usually large orders.[71] LMZ (2002) find that the mean of completion time in their data sample is 29.07 minutes for buy orders and 12.29 minutes for sell orders. The long buy-order completion time could indicate that their market is a 'buy' market, as buy orders are competing with each other and taking longer to be completed. In addition, the average limit-order completion time in their data sample is considerably longer than that in the data sample used in this study. This could reflect the institutional feature of the limit orders in their data sample. They also find that the average censoring time for 'no fills' and 'partial fills' limit orders is 46.92 minutes and 41.14 minutes for buy orders, and 34.15 minutes and 49.16 minutes for sell orders, respectively. The 'no fills' and 'partial fills' censoring times in the data sample used in this study are shorter than those in their data sample. This could, at least partially, be due to the fact that the data sample used in this study is restricted to day limit orders only and excludes orders that remain on the order book for longer and are eventually cancelled or expired.

[68] Table C.3 in Appendix C reports additional statistics of limit-order completion and censoring times. Table C.4 in Appendix C reports the summary statistics of limit-order completion and censoring times for each individual stock.

[69] The censoring time is the difference between the time at which cancellation or expiration occurs and the time of submission.

[70] The censoring times for 'partial fills' orders are the difference between the time at which final cancellation or expiration occurs and the time of submission.

[71] The average size of completed orders is 83% of that of partially filled orders.

As Table 4.3 shows, the average order completion time in the data sample used in this study is over nine minutes. The opportunity cost of waiting for these nine minutes could be significant for traders as market prices could move against them.[72] In this case, these traders may have to accept far worse prices than they had initially hoped for. For example, assume that the best offer price of BT is 401 pence per share. A trader submits a limit order to buy 2000 BT shares at 400 pence per share and just after the submission the best offer price of BT increases to 410 pence per share. If this trader still wants to buy these 2000 BT shares, he needs to buy them at 410 pence per share, which is far higher than the previous best offer price (401 pence per share). This trader can choose to wait and face the probability that the share price could keep on increasing. As a result, these traders could have an incentive to use market orders rather than limit orders in order to avoid this opportunity cost. Impatient traders who try to meet deadlines may find this waiting time more significant. As Handa and Schwartz (1996) argue, impatient traders should therefore intend to submit market orders other than limit orders. In addition, some traders (e.g. hedge funds) use computer programmes to trade and the subsequent trading decision depends on the completion of previous orders. This waiting time means that these traders should mainly use market orders rather than limit orders. This study will show how to reduce this opportunity cost. This issue will be discussed in Chapter 6.

As Table 4.3 shows, the average PLMQ value is higher for censored ‘no fills’ orders than completed and ‘partial fills’ orders. Since orders are executed according to price priority rule as discussed in Chapter 2 and PLMQ is a measure of the relative distance between the limit-order price and the mid-quote price, this higher PLMQ value indicates that traders could cancel their orders due to longer expected execution time. The average order size is larger for ‘partial fills’ orders than for completed orders. This is expected since large orders are difficult to be filled. The average order size is smaller for ‘‘no fills’ orders than for completed orders. This could be due to the strategic behaviours of professional traders. Some professional traders could use small limit

[72] Opportunity cost is the value of the next best alternative foregone as the result of making a decision. Traders can always use market orders to execute their orders at the best quotes. For example, assume that the best offer price of BT is 450 pence per share and traders can only choose between market and limit orders. If the traders choose buy market orders, they can always execute their orders at 450 pence per share assuming that the liquidity is available. However assume that the traders choose buy limit orders and these orders remain unexecuted on the order book. The subsequent best offer price of BT increases to 460 pence per share. Hence the opportunity cost of waiting the limit orders to be executed is 10 pence per share. This opportunity cost depends on the expected waiting time of limit-order completion.

orders to detect the potential counterparts on the opposite side of the market.[73] For example, a trader could submit a small order (e.g. one-share order) to find out whether there is a willing counterpart on the opposite side of the market to trade at this price. After this order remains unexecuted on the order book for a short period of time, this trader could cancel this order and submit another small order to detect the potential counterparts on the opposite side of the market at this new price level.

**Table 4.3: Summary Statistics of Completed and Censored Limit Orders**

This table contains summary statistics (Mean and Standard Deviation) of limit-order completion and censoring times for a pooled sample of 38 stocks. Columns labelled 'Completions' report statistics for completed limit orders. Columns labelled 'No fills' report statistics for the censored limit orders without any fill. Columns labelled 'Partial fills' report statistics for the censored limit orders partially but not completely filled. Standard deviation is in parentheses. PLMQ is a measure of the relative distance between the limit-order price and the mid-quote price. This variable is given by: (limit-order price – mid-quote price)/ mid-quote price. S refers to shares.

| | Units | Completions | | No fills | | Partial fills | |
|---|---|---|---|---|---|---|---|
| | | Buy orders | Sell orders | Buy orders | Sell orders | Buy orders | Sell orders |
| Average Time | Minutes | 9.16 | 9.47 | 7.59 | 7.53 | 16.64 | 16.18 |
| Std.Dev | Minutes | (28.38) | (28.94) | (25.87) | (25.96) | (39.08) | (38.78) |
| Average PLMQ | % | -0.16 | 0.16 | -0.43 | 0.40 | -0.16 | 0.15 |
| Std.Dev | % | (0.18) | (0.18) | (0.49) | (0.47) | (0.17) | (0.16) |
| Average Order Size | 1000S | 22.13 | 23.38 | 14.48 | 13.59 | 27.23 | 27.62 |
| Std.Dev | 1000S | (36.64) | (38.10) | (23.14) | (22.63) | (38.78) | (40.19) |

### *4.1.4 Summary*

The original database contains three-year transaction data (from September 1998 to August 2001). It is impossible to analyse this database with the available computing facilities and technical support. Thus a sample is selected, which consists of all limit orders that are submitted through SETS during the period from October to December 2000 for 38 FTSE 100 stocks and pass filtering criteria. An SQL code is written to trace each limit order from submission to completion, cancellation or expiration, and extract

[73] Some one-share limit orders are found in the data sample used in this study. This could be due to the behaviours of professional traders.

information required for the constitution of certain variables that will be explained in Chapters 5 and 6.

The preceding subsection shows the summary statistics of the data sample. As Table 4.1 shows, 42% of limit orders experienced at least one fill. Hasbrouck and Saar (2004) report 56% of limit orders in a four-month sample of TORQ limit orders have experienced at least one fill. Compared to their data sample, the data sample used in this study has a lower percentage of execution. One possible reason is the difference of trading mechanisms used by the LSE (SETS) and the NYSE. The LSE (SETS) is a pure order-driven market and the NYSE is a hybrid market (order-driven/quote-driven) where specialists are obliged to quote prices when it is necessary. In addition to the limit-order book, these specialists also provide liquidity and hence improve order execution rate. For this reason a hybrid market (order-driven/quote-driven) could offer a better limit-order execution rate than a pure order-driven market.

Table 4.2 reports 66.56% of completed orders are completed with the first fill. LMZ (2002) report over 80% of completed orders are completed with the first fill. Their sample is drawn from limit orders handled by a single institutional broker. The high first-fill completion rate could be more representative of their data source other than the NYSE. The hybrid structure of the NYSE, however, could be the reason behind this high first-fill completion rate, as the specialists provide additional liquidity. If this is the case, in terms of improving first-fill completion rate, a hybrid market (order-driven/quote-driven) is superior to a pure order-driven market.

Table 4.3 reports that the average completion time of limit orders is over nine minutes. In an order-driven market, all traders submit their orders before prices are determined and traders interact with each other directly, meaning immediate execution of traders' orders is not always possible. Limit-order completion time is, then, an implicit trading cost faced by these traders. In contrast, in a quote-driven market, a trader can trade with dealers immediately, so he would not incur this waiting cost. In terms of reducing this waiting cost, a hybrid market (order-driven/quote-driven), such as the NYSE, is superior to a pure order-driven market, such as the LSE, since specialists could provide additional liquidity, which could facilitate limit-order completion.

## 4.2 Methodology

This section introduces the methodology used in this study: survival analysis. Survival analysis, also known as event history analysis, is a statistical technique for analysing non-negative random variables. Typically, the random variable represents the time to an event known as survival time, for example the time to the failure of a physical component or the time it takes a patient to recover and leave a hospital after admission. In this study, it is the time taken by a limit order to execute following submission, i.e., limit-order completion time. Data containing survival time is also known as survival data.

In survival analysis, usually some observations are censored, which occurs when events are not observed and hence the survival times are unknown. These censored observations, however, still provide some valuable information, since they 'survived' for a certain period of time. The main advantage of survival analysis is that it can accommodate these censored observations.

The data sample used in this study includes censored limit orders whose completion times are unknown. Some commonly used regression models cannot be applied to this data sample. For example, Ordinary Least Squares (OLS) linear regressions cannot handle one aspect of the data - censoring. A binary dependent regression model can get round the censoring issue by simply modelling whether or not a limit order completes. Completed orders would have a '1' for the dependent variable while censored orders would have a '0'. However it takes no account of the differences in survival time of each order and consequently loses a large amount of information about when a limit order completes. Therefore, this study will use survival analysis.

The next subsection introduces basic concepts of survival analysis (c.f., Kalbfleisch and Prentice (1980)). Subsection 4.2.2 introduces censoring mechanisms in survival analysis (c.f., Jenkins (2004)). Subsections 4.2.3 and 4.2.4 introduce non-parametric and parametric estimation methods used in survival analysis (c.f., Smith (2002)). Subsection 4.2.5 explains the method of incorporating explanatory variables in survival analysis (c.f., Lee (2003)). Subsection 4.2.6 explains the method to assess the goodness-of-fit of estimated survival models (c.f., Smith (2002)). Subsection 4.2.7 contains a short summary.

### *4.2.1 Basic Concepts of Survival Analysis*

Survival times measure the times to certain events such as failure, death, response, relapse, etc and hence cannot be negative. These survival times are subject to random variations and, like any random variable, form a distribution. The distribution of survival times is usually characterised by one of three mathematically equivalent functions: *survival function*, *probability density function* and *hazard function*. These will now be explained.

Let T denote the survival time. The distribution of T can be characterised by the following three equivalent functions.

*Survival Function*

The survival function, denoted by $S(t)$, is defined as the probability that an observation survives longer than t:

$$S(t) = P(T > t). \tag{4.1}$$

From the definition of the cumulative distribution function $F(t)$ of T,

$$S(t) = 1 - F(t) \tag{4.2}$$

$S(t)$ is a non-increasing function of time t with the properties $S(t) = 1$ for $\mathrm{t} = 0$ and $S(t) = 0$ for $\mathrm{t} = \infty$. That is, the probability of surviving at least at time zero is one and that of surviving an infinite time is zero. The function $S(t)$ is also known as the *survival rate* and the function $F(t)$ is also known as the *failure function.*

*Probability Density Function*

Like any other continuous random variable, the survival time T has a probability density function defined as the limit of the probability that an observation 'fails' (e.g. a patient dies) in the short interval $t$ to $t+\Delta t$, or simply the probability of failure in a small time interval. It can be expressed as

$$f(t) = \lim_{\Delta t \to 0} \frac{P(\mathrm{t} \le \mathrm{T} \le \mathrm{t} + \Delta\mathrm{t})}{\Delta\mathrm{t}} . \tag{4.3}$$

*Hazard Function*

The hazard function $h(t)$ of survival time is defined as the probability of failure during a very small time interval, assuming that the observation has survived to the beginning of the interval:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(\mathrm{t} \le \mathrm{T} \le \mathrm{t} + \Delta\mathrm{t} \mid \mathrm{T} \ge \mathrm{t})}{\Delta\mathrm{t}} . \tag{4.4}$$

The hazard function is also known as the *instantaneous failure rate*, *force of mortality*, *conditional mortality rate* and *age-specific failure rate* in insurance and medical applications. It is a measure of the proneness to failure as a function of the age of the 'individual' in the sense that the quantity $h(t)\Delta t$ is the expected proportion of age t 'individual' that will fail in the short time interval $t$ to $t+\Delta t$. The hazard function, therefore, gives the risk of failure per unit time during the aging process.

The hazard function can also be defined in terms of the cumulative distribution function $F(t)$:

$$h(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1-F(t)} . \tag{4.5}$$

It then follows that

$$h(t) = -\frac{d}{dt}\{\log S(t)\}, \tag{4.6}$$

and hence

$$S(t) = \exp\{-H(t)\} , \tag{4.7}$$

where

$$H(t) = \int_0^t h(u)d(u) . \tag{4.8}$$

The function $H(t)$ is called the *integrated or cumulative hazard function* and widely used in survival analysis. From Equation (4.7), the cumulative hazard function can be defined in terms of the survival function as

$$H(t) = -\log S(t)\,. \tag{4.9}$$

The survival function, the probability density function, the hazard function and the cumulative hazard function are mathematically equivalent. If one function is known, the rest can be derived mathematically.

### ***4.2.2 Censoring***

Censoring occurs when events (e.g. the failure of a physical component) cannot be observed. Thus the survival time (from entry time, which is the time when the object entered observation, until failure) is not known exactly. For example, in a clinical trial that monitors the survival times of cancer patients, some patients may still be alive at the end of the trial so that the survival times of these patients are longer than the duration of the trial. In the data sample used in this study, some limit orders are cancelled or expired before completion. For these orders, the survival times (completion times) are longer than their cancellation or expiration times. Censoring can be categorised into the following types.

- *Right censoring:* at the time of observation, the relevant failure event has not yet occurred and so the survival time is unknown. Given entry at time 0 and observation at time t, it is only known that the survival time is greater than the observed time spell (T > t). For example, if five physical components are tested and only three have failed by the end of the test, the two remaining components are the right-censored observations. If the two remaining physical components were to keep on operating, the failure would occur at some point after the end of the test. This type of censoring features in this study of limit-order completion time.

- *Left censoring:* when the entry time is not observed, the exact survival time is not known. For example, in a medical study of monitoring the survival time of cancer patients from ill state to death state, patients are already ill for a while when an examination reveals their illness. Then the time of onset is unknown. This type of censoring, however, does not feature in this study of limit-order completion time.

- *Interval censoring:* interval censoring occurs when the failure event occurs between time intervals. Then the exact survival time is unknown. For example, interval censoring occurs if a mortality investigation records only the calendar year of death rather than the exact date. Hence the exact survival time is unknown. This type of censoring, however, does not feature in this study of limit-order completion time.

Censoring can be both informative and non-informative. Non-informative censoring occurs if the censoring has no information about the survival time. Thus, if the survival time and the censoring mechanism are independent, censoring is non-informative. An example from medical studies would be a patient who moves away from the study area for a reason unconnected with the progress of his illness (for example, due to a heating facility break down). However, if the patient moves away from the study area in order to receive a different treatment, then it is informative censoring. In this study, censoring is assumed to be non-informative.

In the data sample used in this study, some traders may cancel their orders due to long expected completion times. In this case, it would be considered as informative censoring. Without knowing why traders cancel orders from the dataset, it is impossible to know whether the censoring is informative or non-informative, which makes this a limitation of the dataset used in this thesis.

### *4.2.3 Non-Parametric Estimation of the Survival Function*

This subsection introduces a non-parametric estimator of survival functions since no prior assumptions are usually made about the shapes of the survival functions.

*Kaplan-Meier Estimator*

The main non-parametric estimator of survival function used is known as the Kaplan-Meier estimator or Product Limit estimator.[74] This allows for right censoring and provides an estimate of the survival function from survival data.

Let the observed survival times until events be

$$t_1 \le t_2 \le t_3 .......... \le t_m . \quad (4.10)$$

Corresponding to each $t_i$ is $n_i$, the number of observations 'at risk' just prior to time $t_i$, and $f_i$, the number of failures at time $t_i$. When there is no censoring, $n_i$ is just the number of survivors just prior to time $t_i$. With censoring, $n_i$ is the number of survivors less the number of censored observations. The Kaplan-Meier estimator of $S(t)$ is given by

$$\hat{S}(t) = \prod_{t_i \le t} \frac{n_i - f_i}{n_i} \ . \quad (4.11)$$

From Equations (4.2) and (4.9) estimates of the failure function $\hat{F}(t)$ and the integrated hazard function $\hat{H}(t)$ can be derived as

$$\hat{F}(t) = 1 - \hat{S}(t) \text{ and} \quad (4.12)$$

$$\hat{H}(t) = -\log \hat{S}(t) . \quad (4.13)$$

The following is an example of Kaplan-Meier estimator. Suppose that 500 patients are tracked over a period of time to determine how many survive for one year, two years, three years, and so forth. Without right censoring, the estimation of year-by-year survival rate for patients in general would be an easy matter. The survival of 475 patients at the end of the first year would give a one-year survival rate estimate of 475/500=0.95; the survival of 450 patients at the end of the second year would yield a two-year estimate of 450/500=0.90; and so forth.

[74] Kaplan and Meier (1958) introduce this estimator.

With right censoring (e.g. patients can no longer be tracked), the estimation of survival rate would be more complicated. Of the 500 patients at the beginning of the study, four are censored and become unavailable during the first year and 24 are known to have died by the end of the first year. Another six are censored during the second year and another 15 are known to have died by the end of the second year. And so on for the other years. Table 4.4 shows how to compute the Kaplan-Meier estimators of survival functions.

The Kaplan-Meier estimator can be easily derived and the survival function is estimated non-parametrically. In contrast, in the parametric estimation of the survival function described below a specific parametric family is assumed for the distribution of survival times (e.g. the generalised gamma distribution). Given the distributional assumption, maximum likelihood estimation can then be performed.

**Table 4.4: An Example of Kaplan-Meier Estimator of Survival Function**

| Time Period | No. of patients at the start of each year | No. of censored patients | No. of dead patients | No. of survived patients | Kaplan-Meier estimator of survival function |
|---|---|---|---|---|---|
| **Year 1** | 500 | 4 | 24 | 472 | (472/(500-4))=0.95 |
| **Year 2** | 472 | 6 | 15 | 451 | (472/(500-4)) x (451/(472-6))=0.92 |
| **Year 3** | 451 | 10 | 21 | 420 | (472/(500-4)) x (451/(472-6)) x (420/(451-10))=0.88 |
| **Year 4** | 420 | 7 | 20 | 393 | (472/(500-5) x (451/(472-6)) x (420/(451-10) x (393/(420-7))=0.83 |

### *4.2.4 Parametric Estimation of the Survival Function*

Parametric estimation of the survival function begins with the specification of the distribution of survival time T, from which the likelihood function is obtained.

*Likelihood Function*

Let $(t_1,...,t_n)$ denote a sequence of n realizations of T. In the case where there is no censoring, the likelihood function may be written as

$$L(\theta) = \prod_{i=1}^{n} f(t_i;\theta) = \prod_{i=1}^{n} S(t_i;\theta)h(t_i;\theta), \tag{4.14}$$

where $\theta$ is a vector of the parameters of interest.

In the case of non-informative right-censored survival data, if $m$ observations are still alive at $t_i$, the only thing known about these observations is that their survival times exceed $t_i$. In this case, the contribution of $m$ censored observations to the likelihood function is

$$L(\theta) = \prod_{i=1}^{m} S(t_i;\theta). \tag{4.15}$$

Suppose that $(T_1,...,T_n)$ are right-censored by $(t_1,...,t_n)$ and $(\delta_1,...,\delta_n)$ are used to denote as the censoring indicators:

$$\delta_i = \begin{cases} 1 & \text{if observation i is not censored } (\mathrm{T_i} \leq \mathrm{t_i}) \\ 0 & \text{if observation i is censored } (\mathrm{T_i} > \mathrm{t_i}) \end{cases}. \tag{4.16}$$

The likelihood function of non-informative right-censored survival data may then be written as

$$L(\theta) = \prod_{i=1}^{n} h(t_i;\theta)^{\delta} S(t_i;\theta) = \prod_{i=1}^{n} f(t_i;\theta)^{\delta} S(t_i;\theta)^{1-\delta}. \tag{4.17}$$

Given a specific mathematical form for the distribution function $f(t;\theta)$, the vector $\theta$ is estimated by the method of maximum likelihood. An estimate of the survival function can then be obtained.

### *4.2.5 Incorporating Explanatory Variables*

In this subsection, the relationship between the survival time T and the value of an explanatory variable x are examined. For example, if T denotes limit-order completion time, x may denote limit-order size or some other variables that might affect the limit-order completion time. The aim is to investigate whether x has explanatory power over T and acts as a determinant. Generally, survival time T is expected to depend on the values of several explanatory variables, denoted by a vector of X.

Two frequently used models for adjusting survival functions for effects of explanatory variables are the Accelerated Failure Time (AFT) and Proportional Hazard (PH) models. In an AFT model, the dependence of survival times on explanatory variables can be addressed by assuming that the effects of explanatory variables can be captured by rescaling time. In a PH model, it assumes the hazard function is rescaled. These models will now be described.

*The Accelerated Failure Time Model*

In an AFT model, an exponential factor is used to rescale time. Specifically, an AFT model takes the following form:

$$T = e^{X'\beta} T_0 \ , \tag{4.18}$$

where T is the survival time, X is a vector of explanatory variables, $\beta$ is a vector of parameters, and $T_0$ is called the baseline survival time and its distribution is the baseline distribution. According to Equation (4.18), T is a scaled transformation of the baseline time $T_0$ and the explanatory variables and parameter coefficients determine the scaling.

If the baseline distribution is estimated to take a specific parametric form, such as that of the generalised gamma distribution, it is not difficult to incorporate explanatory variables into the likelihood function of Equation (4.17). The likelihood function can then be maximised to obtain an estimate of the parameter vector $\beta$.

*The Proportional Hazard Models*

In a PH model, explanatory variables have a multiplicative effect on the hazard function:

$$h(t) = e^{X'\beta} h_0(t) \ , \tag{4.19}$$

where $h(t)$ is the hazard function, X is a vector of explanatory variables, $\beta$ is a vector of parameters, and $h_0(t)$ is called the baseline hazard function.

If the baseline hazard function $h_0(t)$ is estimated to take a specific parametric form, then the explanatory variables can easily be incorporated into the likelihood function of Equation (4.17), which can then be maximised to give an estimate of the parameter vector $\beta$.

In the PH model the baseline function $h_0(t)$ may be left unspecified, yielding the Cox Proportional Hazard (PH) model which is frequently employed in survival analysis.

*The Cox Proportional Hazard Model*

A unique feature of the Cox PH model is that one can estimate the relationship between the hazard function and the explanatory variables without any assumptions about the shape of the baseline hazard function. The fact that $h_0(t)$ can be completely unspecified can be seen as a great advantage that avoids potential problems of specifying the wrong shape. However it can also be a disadvantage because the shape of the baseline hazard function, which might need to be investigated, is unknown. The Cox PH model is described next.

Assume that there are no ties in the n observed survival times.[75] Suppose that k observations failed (uncensored) and the remaining n-k are censored. Let the ordered survival times as

$$t_1 < t_2 < t_3 .......... < t_k . \tag{4.20}$$

Let $R(t_i)$ denote the risk set at time $t_i$, which is the set of subjects alive and under observation at time $t_i^-$, immediately prior to $t_i$, $i$ =1,2,3…k. An example of constructing risk sets is described next.

[75] No observations have the same survival times.

The following are five limit-orders. Two of these limit orders are right-censored (cancelled or expired before completion) and three are completed.

**Table 4.5: A Limit-Order Dataset**

| Individual limit order | Survival time (minutes) | Censoring |
|---|---|---|
| **A** | 4 | No |
| **B** | 12 | No |
| **C** | 25 | Yes |
| **D** | 68 | No |
| **E** | 108 | Yes |

For this dataset, n=5 and k=3. Accordingly $t_1$=4, $t_2$=12 and $t_3$=68, with associated risk sets: $R(t_1) = \{ \text{A, B, C, D, E} \}$; $R(t_2) = \{ \text{B, C, D, E} \}$; $R(t_3) = \{ \text{D, E} \}$.

For the Cox PH model $\beta$ can be estimated without directly estimating $h_0(t)$. Parameter estimates in this model are obtained by maximising the partial likelihood (PL). The partial likelihood is given by

$$L(\beta) = \prod_{i=1}^{k} \frac{e^{X_i^{'}\beta}}{\sum_{j \in R(t_i)} e^{X_j^{'}\beta}} . \tag{4.21}$$

The partial likelihood behaves like an ordinary likelihood and can be used to estimate $\beta$.

In the above example no tied survival times exist.[76] However, in the sample data used in this thesis, due to the way that times are recorded (in seconds), tied survival times do occur. In such a case, approximation needs to be used. Breslow (1974) suggests an approximation method. Suppose that there are $d_i$ repetitions of an observed survival time $t_i$ and these observations with tied survival times record different values of explanatory variables X as

$$X_{(i)1}, X_{(i)2}, ..., X_{(i)d_i} . \tag{4.22}$$

[76] Equal survival times are known as tied survival times.

These appear in the numerator of Equation (4.21) as

$$e^{X_{(i)1}'\beta} \times e^{X_{(i)2}'\beta} \times, \ldots, e^{X_{(i)d_i}'\beta} = e^{\sum_{l=1}^{d_i} X_{(i)l}'\beta} = e^{S_i'\beta}, \tag{4.23}$$

where

$$S_i = \sum_{l=1}^{d_i} X_{(i)l}, \tag{4.24}$$

$S_i$ is the sum of the explanatory variable vectors for tied survival times. Accordingly the partial likelihood of Equation (4.21) is transformed to

$$L(\beta) = \prod_{i=1}^{k} \frac{e^{S_i'\beta}}{\left[\sum_{j \in R(t_i)} e^{X_j'\beta}\right]^{d_i}}. \tag{4.25}$$

In this study, the Breslow method is used to accommodate tied survival times.

*Hazard Ratio*

The PH model presented above indicates that absolute differences in the vector of explanatory variables X mean differences in the hazard function at time t. Consider two limit orders A and B with associated vectors of explanatory variables $X_A$ and $X_B$. The ratio of the hazard rates is

$$\frac{h_A(t)}{h_B(t)} = \frac{h_0(t)e^{X_A'\beta}}{h_0(t)e^{X'_B\beta}} = \frac{e^{X_A'\beta}}{e^{X'_B\beta}} = e^{(X'_A - X_B')\beta} \tag{4.26}$$

If limit orders A and B have identical explanatory variables except the kth, then Equation (4.26) can be simplified to

$$\frac{h_A(t)}{h_B(t)} = e^{(X_{Ak} - X_{Bk})\beta_k}. \tag{4.27}$$

Furthermore if there is only a one-unit change in $X_k$ (i.e., $X_{Ak} - X_{Bk}$ =1), then

$$\frac{h_A(t)}{h_B(t)} = e^{\beta_k}. \tag{4.28}$$

$e^{\beta_k}$ is known as the hazard ratio. It shows the proportionate change in the hazard function given a one-unit change in an explanatory variable when all remaining explanatory variables are held constant.

### *4.2.6 Assessing the Goodness-of-Fit*

To check the goodness-of-fit of estimated survival models, this analysis uses a graphical diagnostic measure.[77] This measure is motivated as follows.

Let $S(t)$ be the true survival function of a continuous random variable T, and that the cumulative hazard function is $H(t) = -\log S(t)$. Thus $S(t) = e^{(-H(t))}$. Let $Y = H(T)$ be a transformation of T based on the cumulative hazard function. Then the survival function for Y is as follows:

$$S(y) = P(H(T) > y) = P(T > H^{-1}(y)) = S(H^{-1}(y)) = e^{(-H(H^{-1}(y)))} = e^{-y} \ . \qquad (4.29)$$

Thus, regardless of the distribution of T, the new measure $Y = H(T)$ has an exponential distribution with $\lambda = 1$.

Let $\hat{S}(t)$ be an estimated survival function and $\hat{H}(t)$ an estimated cumulative hazard function. The 'Cox-Snell residual' is defined as $z_i = \hat{H}(t_{i,} X_i) = -\log \hat{S}(t_i, X_i)$, where $X_i$ is a vector of explanatory variables. According to Equation (4.29), if $\hat{S}(t)$ is close to the true survival function $S(t)$, then the Cox-Snell residual Z should have a standard exponential distribution with $\lambda = 1$.

In this study the Cox-Snell residuals are used to check the goodness-of-fit of estimated survival models. The steps to check whether the Cox-Snell residual Z has a standard exponential distribution with $\lambda = 1$ are described as follows. First, the Kaplan-Meier

[77] The likelihood ratio test cannot be used in this study, since for such a large sample tiny departures from null hypothesis are always significant. This analysis attempts to use the likelihood ratio test and finds that the null hypothesis is always rejected. The likelihood ratio test for survival analysis is described in more detail by Lee (2003).

estimators on the Cox-Snell residuals $\hat{S}(z_i)$ are computed. Second, from these estimates the cumulative hazard function of the Cox-Snell residuals $\hat{H}(z_i)$ can be estimated as Equation (4.13) illustrates. Finally, this $\hat{H}(z_i)$ is plotted against the Cox-Snell residual $z_i$. In this study this plot is known as Q-Q plot. If the estimated model is the true model, then the plot should be a straight line through the origin with slope of 1. Any departure from this line indicates lack of fit.

### *4.2.7 Summary*

The methodology used in this thesis is survival analysis, which is a statistical technique for analysing non-negative random variables. Usually, the random variable represents the time to an event known as survival time. In this study, it is the limit-order completion time. The distribution of survival times is usually characterised by either of three mathematically equivalent functions: survival function, probability density function and hazard function.

In survival analysis, censoring occurs when events are not observed and hence their survival times are unknown. Censoring can be categorised into right, left and interval censoring. It can also be categorised into informative and non-informative censoring. In this study, censoring is assumed to be non-informative.

In survival analysis, estimation of survival function can be obtained non-parametrically and parametrically. In the non-parametric approach, no prior assumptions are made about the shapes of the survival function. The main non-parametric estimator of survival function used is known as the Kaplan-Meier estimator. In the parametric approach, the distribution of survival time is specified, from which the likelihood function is obtained. An estimate of survival function is then obtained by the method of maximum likelihood.

In survival analysis explanatory variables can be incorporated. Two frequently used models are the AFT and PH models. In the PH (AFT) model, explanatory variables are used to rescale the baseline hazard function (survival time). In the PH model, the baseline hazard function may be left unspecified, yielding the Cox PH model, which is frequently employed in survival analysis.

Recently survival analysis has been employed in finance studies. Lunde, Timmermann and Blake (1999) use survival analysis to study the lifetimes of mutual funds. Maggiolini and Mistrulli (2005) use survival analysis to study survival rate of de novo Co-operative Credit Banks established in Italy during the 1990s. Survival analysis is also used to study high frequency transaction data. Cho and Nelling (2000) use survival analysis to study the execution probability of a limit order. LMZ (2002) use survival analysis to model limit-order execution time.

The next chapter is the first empirical chapter in this study, which will investigate the degree to which explanatory variables suggested by LMZ capture the state of the UK order book and affect limit-order completion time.

# CHAPTER 5 – LIMIT-ORDER COMPLETION TIME IN THE UK MARKET: AN INVESTIGATION USING LO, MACKINLAY AND ZHANG MODEL

## 5.1 Introduction

Cho and Neiling (2000) was the first to model limit-order execution probability using survival analysis. They use a Proportional Hazard (PH) model to incorporate explanatory variables.[78] They propose a single model to describe execution probabilities for both buy and sell limit orders. However, they do not provide any goodness-of-fit tests. LMZ (2002) investigate limit-order execution time also using survival analysis. They use an Accelerated Failure Time (AFT) model to incorporate explanatory variables and assume the generalised gamma distribution as the baseline survival time distribution.[79] In their study, censoring is assumed to be non-informative.[80] They model time-to-first-fill and time-to-completion for buy and sell limit orders separately and provide goodness-of-fit tests. The data sample in their study is unique, containing all limit orders submitted through an institutional brokerage firm (ITG) during one year for the 100 largest stocks in the S&P500 index. Chakrabarty *et al.* (CHTZ) (2006) investigate limit-order execution and cancellation time using competing risk analysis (an extension of survival analysis) and modified versions of some of the explanatory variables suggested by LMZ. They model time-to-execution and time-to-cancellation for buy and sell limit orders separately. The data sample in their study contains all limit orders submitted through a specific trading platform (INET) during the period from July 2005 to December 2005 for five stocks traded on INET.[81]

[78] See Subsection 4.2.5 of Chapter 4 for details of a PH model.
[79] See Subsection 4.2.5 of Chapter 4 for details of an AFT model.
[80] See Subsection 4.2.2 of Chapter 4 for details of censoring mechanisms in survival analysis. This study assumes that censoring is non-informative.
[81] INET is an automated limit-order platform for trading equities in the US. It is open when the US equity markets are open, and generally accepts orders between 7 am and 8 pm EST. Only broker/dealers can submit orders to INET, and the only type of order allowed is limit order, which can be either open for display or hidden. The data sample consists of four randomly selected stocks and Intel Corp, which is one of the most liquid equities, traded on US markets. Orders in the sample are priced in cents.

Empirical Investigations in the studies mentioned above focus on the US market. As far as we know, research on limit-order completion time has not been carried out on the UK data. The main aim of this chapter is to investigate the degree to which explanatory variables suggested by LMZ capture the state of the UK order book and affect limit-order completion time at the LSE.

The motivations behind replicating the LMZ models are as follows. Firstly, LMZ (2002) is the leading research in limit-order completion. However the LMZ models have not been tested on market data outside the US. Hence it is important and necessary to test the universal applicability of the LMZ models before these models can be used.[82] This study replicates the LMZ models on UK market data, identifies the potential problems in these models and suggests that these models should be used cautiously in future research. Secondly, a replication of the LMZ models on UK market data is useful to compare limit-order completion between UK and US markets. Finally, the replication of the LMZ models on UK market data will show whether the LMZ variables are significant in explaining the limit-order completion time in the UK market.

The remainder of this chapter is organised as follows. Section 5.2 introduces explanatory variables suggested by LMZ and offers a discussion on their construction. Section 5.3 presents empirical results and investigates the possible multicollinearity effect on the estimation of parameters involved the LMZ models. Section 5.4 offers a summary, some conclusions, and provides an introduction to the work in subsequent chapters.

## 5.2 LMZ Explanatory Variables

Before investigating LMZ explanatory variables, differences between the data sample used in this study and the data sample used by LMZ are listed as follows.

- *Different markets:* this study uses LSE (SETS) stocks data and LMZ use NYSE stocks data. During the period covered by the sample used in this study, October

[82] Tkatch and Kandel (2006) use the LMZ models without prior testing of these models to the market data being studied.

2000 to December 2000, SETS operated as a pure order-driven market and disclosed its limit-order books to traders. Thus, it could be described as a highly pre-trade transparent market. This high transparency could be the reason behind the low order completion rate observed in the data sample used in this study, as discussed in Subsection 4.1.4. Assuming no private information, in a pre-trade transparent market such as the LSE, traders will submit orders depending mainly on the state of the order book. Hence, limit-order completion time is more sensitive to the state of the order book (e.g. the best bid-offer spread) than that in a less transparent market. Also all SETS orders were executed continuously and electronically. In contrast, during the period covered by the sample used in the LMZ study, August 1994 to August 1995, the NYSE only displayed a snapshot of all price steps and associated depth on the limit-order books at ten seconds intervals. Hence, the NYSE is not as transparent as the LSE (SETS). Unlike the LSE (SETS), a pure order-driven market, the NYSE had hybrid trading mechanisms. This potentially could affect the sensitivity of limit-order completion time to explanatory variables constructed next. This will be discussed later.

- *Different data sample periods*: this study uses a data sample that includes all limit orders submitted to the order books for 38 FTSE 100 stocks from October 2000 to December 2000. In contrast, LMZ use a data sample provided by Investment Technology Group (ITG) only. ITG is an institutional brokerage firm that provides technology-based equity trading services such as POSIT (an electronic crossing system), QuantEX (a decision support and routing system) and a full-service trading desk. Their data sample consists of all limit orders submitted through the ITG trading desk from 1 August 1994 to 31 August 1995 for the 100 largest stocks (in terms of market capitalization at the end of September 1995) in the S&P 500.[83]

- *Different data sources:* the data sample used in this study includes all 38 FTSE 100 stocks orders submitted to the LSE and reflects some characteristics of the UK stock market as discussed in Chapter 2, since trading in these 38 FTSE 100 stocks represents a large percentage of trades in the UK market. In contrast, LMZ use data from a single source (ITG) and thus their data sample reflects the

[83] In the data sample used by LMZ orders are priced in dollars and in the data sample used in this study orders are priced in pence. This difference in pricing could cause the difference in explanatory variables. This will be discussed later.

characteristics of their data source as follows. Their limit orders are almost exclusively submitted by institutional investors and other brokers/dealers and hence reflect an institutional feature. Since the ITG trading desk handles unfilled trades from some liquidity sources, such as SuperDOT and Instinet, some limit orders in their data sample have arisen from absence of liquidity.[84]

- *Different tick size:* the minimum tick of the LMZ sample is 0.125 dollars, while that of the data sample used in this study is 0.01 pence. The tick size imposes price discreteness and forms a lower bound on the spread.[85] If the minimum tick size is a binding constraint for liquidity providers in posting quotes, then it is likely that a reduction in the tick size will encourage competition among traders and decrease the spread. Hasbrouck (1992) examines the cross-sectional relative frequency of limit order executions using the TORQ data sample and finds that the number of limit orders increases with tick sizes. He argues that with a smaller tick size a trader can strategically jump ahead of standing limit orders by marginally improving on the limit-order price. Hence, the smaller tick size of the data sample used in this study could decrease the protection time priority offered to limit orders on the order book. This may result in a high cancellation rate, as discussed in Subsection 4.1.4, because limit orders that lose their time priority are likely to be cancelled. The smaller tick size of the sample used in this thesis could also be a possible reason for the observed difference between the variables constructed in this chapter and those proposed by LMZ. This will be discussed below in more detail.

LMZ use an AFT model, which assumes that rescaling time can capture the effect of explanatory variables on survival times, to incorporate a vector of explanatory variables (see Subsection 4.2.5).[86] In this section, the vector of explanatory variables used in the LMZ study is constructed. Unlike LMZ, who model both time-to-first-fill and time-to-

[84] SuperDOT is an electronic system used to place orders for listed stocks, which usually refers to those trading on the NYSE. Instinet is a professional stock trading system owned by Reuters. Institutions use it to trade large blocks of shares with each other without using the exchanges.

[85] Harris (1994) argues that the bid-offer spread should be positively related to the minimum tick size. He also argues that liquidity providers cannot improve on a spread that is just one tick, since the tick size sets the lower bound for the quoted bid-offer spread.

[86] LMZ (1999) also use a Cox PH model to incorporate these explanatory variables. LMZ find that the AFT model performs better than the Cox PH model basing on the Q-Q plots, which will be discussed later in this chapter. For this reason in this chapter this study will use only the AFT model to incorporate these variables as LMZ (2002) suggest. In the next chapter this study will use both the AFT and Cox PH models to incorporate newly constructed variables.

completion, this study model time-to-completion only, which is of primary interest to traders as their aims normally are to complete their orders. Separate models are constructed, however, for buy and sell orders. These are henceforth referred to as 'buy model' and 'sell model'. The construction of explanatory variables is described next.[87]

Let $P_l$ denote the limit-order price, $S_l$ the limit-order size, $P_b$ the best bid price, $S_b$ the best bid size, $P_o$ the best offer price, $S_o$ the best offer size, $P_q$ the mid-quote price and $P$ the market price of the most recent transaction.

The following are the explanatory variables suggested by LMZ in their model of time-to-completion of buy limit orders (all variables are measured at the submission time of a limit order).

MQLP is a measure of the distance between the limit-order price and current midpoint. This variable is given by Equation (5.1) below:

$$MQLP = P_q - P_l \,. \tag{5.1}$$

A large value of MQLP indicates that this limit order does not improve the current best bid price given the market spread. For example, as Figure 2.8 of Chapter 2 illustrates, the best bid price is 524 pence and the best offer price is 525 pence. Hence, the mid-quote price is 524.5 pence. For a limit order, a MQLP of 5 pence means that the limit-order price is 519.5 pence, which is far below the best bid price, and hence this limit order does not improve on the current best bid price (524 pence). Large values of MQLP may indicate low execution probability and long execution time, since orders are executed according to price priority as discussed in Chapter 2. However, MQLP does not capture the market spread, which is an important microstructure aspect of a limit-order book, and transaction costs faced by traders. A wider market spread increases transaction costs and hence discourages trading. Thus this would obviously reduce order execution probability and extend execution time.

[87] LMZ (2002) do not present any rationales behind the construction of their explanatory variables.

BSID is an indicator of whether the prior trade was buyer-initiated (BSID equals to one) or seller-initiated (BSID equals to negative one).[88] This variable is given by Equation (5.2) below:

$$BSID = \begin{cases} 1 & \text{if prior trade occured above } P_q \\ 0 & \text{if prior trade occured at } P_q \\ -1 & \text{if prior trade occured below } P_q \end{cases} . \qquad (5.2)$$

As Figure 2.8 of Chapter 2 illustrates, the best bid price is 524 pence and the best offer price is 525 pence. Hence, the mid-quote price is 524.5 pence. If the price of the prior-trade price is 524.75 pence, for example, then the value BSID takes is one as the price of the prior-trade price is above the mid-quote price and hence closer to the best offer price. Two scenarios could have happened. The best offer price just before the last trade was 524.75 pence and a buy market order was executed at this price. This would have moved the best offer price to 525 pence. Or the best bid price just before the last trade was 524.75 pence and a sell market order was executed at this price, which moved the best bid price to 524 pence. The first scenario is more probable as the prior-trade price (524.75 pence) was closer to the current best offer price (525 pence). Hence, it is more probable that the prior trade was buyer-initiated as buyers submitted market orders. If the prior trade was buyer-initiated, then the current market is considered as a 'buy' market and buyers are submitting more buy orders, since trades are usually clustered as Biais, Hillion and Spatt (1995) show.[89] Assuming that sellers fail to react by submitting more sell orders, buy orders will compete with each other for execution against a limited number of sell orders. As a result, buy limit orders are expected to have lower execution probabilities and longer execution times.

MKD1 measures the minimum number of shares that have higher priority for execution scaled by the distance between the limit-order price and the best bid price. This variable is given by Equation (5.3) below:

---

[88] LMZ interpret this variable in the opposite way. An email was sent to the LMZ for classification. However no reply has been received.

[89] Buy orders tend to follow buy trades and sell orders tend to follow sell trades. The 'buy' and 'sell' markets could just be the result of a random arrival of orders, which is triggered by economic events. For example, on 14 September 2007, Northern Rock share price plunged 32% after BBC revealed that this bank had to ask the Bank of England for emergency funding. The plunge of the share price is the result of the inflow of a large number of sell orders, which is triggered by the news. During the sample period, the 'buy'/'sell' market is the result of the arrival of buy/sell orders triggered by economic events such as companies reporting corporate earnings.

$$MKD1 = \begin{cases} (1 + P_b - P_l)\log S_b & \text{if } P_l \le P_b \\ 0 & \text{if } P_l > P_b \end{cases}. \tag{5.3}$$

It is different from zero only for buy limit orders that do not improve the best bid price. This variable measures an interaction between two terms. The first term is the distance between the limit-order price and the best bid price, and the second is the best bid size. Even if the limit-order price equals to the best bid price, the existing best bid order still has priority for execution, as orders are executed according to price and then time priority as discussed in Chapter 2. Thus for this limit order the best bid size is the number of shares that have higher priority for execution. MKD1 has high values if either the best bid size or the distance between the best bid price and the limit-order price is large. The larger the value MKD1 takes (the further away a limit order is priced from the best bid price or the larger the best bid size), the larger the number of shares that have higher priority for execution over this limit order, and hence the lower the execution probability and the longer the expected time-to-execution.

MKD1X is an interactive term to capture non-linearity between MKD1 and market price relative to the limit-order price. This variable is given by Equation (5.4) below:

$$MKD1X = \begin{cases} (P - P_l)MKD1 & \text{if } P \ge P_l \\ 0 & \text{if } P < P_l \end{cases}. \tag{5.4}$$

MKD1X is different from zero only for buy limit orders submitted below the last transaction price. This variable measures the interaction between two terms. The first term is the distance between the last transaction price and the limit-order price, and the second term is the variable MKD1. MKD1X can be interpreted as a version of MKD1. Since MKD1X is derived from MKD1, the correlation between them could be high and could cause multicollinearity in the model. This multicollinearity issue will be discussed later in Subsection 5.3.2.

MKD2 is a measure of liquidity available from the sell side of the market and the interaction between liquidity on the best offer price and the distance between the best offer price and the limit-order price. This variable is given by Equation (5.5) below:

$$MKD2 = \begin{cases} \log S_o / (1 + P_o - P_l) & \text{if } P_o \ge P_l \\ \log S_o & \text{if } P_o < P_l \end{cases}. \tag{5.5}$$

When the limit-order price is less than or equal to the best offer price, it is an interaction of two terms. The first term, $\log(S_o)$, measures the best offer size, while the second, $1+P_o-P_l$, measures the distance between the limit-order price and the best offer price. MKD2 is large if the best offer size is large. High values of this variable may indicate that traders price their limit orders aggressively (close to the best offer price). When the limit-order price is higher than or equal to the best offer price, two scenarios are possible. If liquidity is available, the limit order will behave as a market order and will be completed immediately. If liquidity is lacking, the limit order will be partly filled with the available liquidity and the unexecuted part will remain on the order book for subsequent trading, as discussed in Chapter 2. The more liquidity is on the sell side of the order book, the higher the limit-order execution probability and the shorter the expected time-to-execution.

SZSD is a measure of liquidity demanded by the limit order scaled by the distance between the limit-order price and the best offer price. This variable is given by Equation (5.6) below:

$$SZSD = \begin{cases} \log(S_l)(1+P_o-P_l) & \text{if } P_o > P_l \\ \log(S_l-S_o) & \text{if } P_o = P_l \text{ and } S_l > S_o \\ 0 & \text{otherwise} \end{cases} . \tag{5.6}$$

When the best offer price is lower than or equal to the limit-order price ($P_o \le P_l$) and liquidity on the sell side is available, the variable takes the value of zero. When the best offer price is equal to the limit-order price ($P_o = P_l$) and liquidity is lacking ($S_o < S_l$) the limit order executes immediately but partly and hence the liquidity demanded depends on the unexecuted part of the limit order ($S_l - S_o$). When the limit-order price is lower than the best offer price, the liquidity demanded is defined by an interaction of two terms. The first term, $\log(S_l)$, measures the limit-order size, while the second, $1+P_o-P_l$, measures the distance between the limit-order price and the best offer price. The higher the liquidity demand, the lower the limit-order execution probability and the longer the expected time-to-execution.

STKV is used to capture recent shifts in trading activity. This variable is given by Equation (5.7) below:

$$STKV = \text{trades last half hour / trades last one hour}. \tag{5.7}$$

This variable assumes that the number of trades during the hour prior to order submission cannot zero.[90] It is high if there is an increase in the level of activity in the market during the last half hour, which means that more orders have arrived at the market. The more active a market becomes prior to order submission, the higher the limit-order execution probability and the shorter the expected time-to-execution.

TURN is a trading activity measure that provides an absolute measure of volatility. This variable is given by Equation (5.8) below:

$$TURN = \log(\text{trades last one hour}). \tag{5.8}$$

This variable assumes that the number of trades during the hour prior to order submission is not zero.[91] It is high if the volatility in the market in the last one-hour has been high. The more active a market becomes the higher the limit-order execution probability and the shorter the expected time-to-execution.

LSO is the logarithm of the number of shares outstanding and is updated monthly.[92] This variable is given by Equation (5.9) below:

$$LSO = \log(\text{previous month - end shares outstanding, in thousands}). \tag{5.9}$$

It is a measure of market size in terms of the number of shares outstanding. Trading activity is usually correlated with the number of shares available for trading. The higher the number of shares available for trading, the more active the market could become. Therefore this variable can be potentially correlated with TURN and could cause multicollinearity in the model. This multicollinearity issue will be discussed later in Subsection 5.3.2. The higher the value this variable takes, the higher the limit-order execution probability and the shorter the expected time-to-execution.

LPR is the logarithm of historical share price.[93] This variable is given by Equation (5.10) below:

[90] If no trade occurs in the last one hour, STKV will be invalid. Limit orders with invalid STKV would be excluded from the data sample in estimation.
[91] If no trade occurs in the last hour, TURN will be invalid. Limit orders with invalid TURN would be excluded from the data sample in estimation.
[92] This variable is extracted from DataStream.

$$LPR = \log(\text{previous month's average daily closing price}). \tag{5.10}$$

Usually high-price stocks tend to be liquid.[94] Hence, trading activity can be correlated with the share price. The higher the share price, the more liquid the stock becomes and hence the more active trading could become. Therefore this variable can be potentially correlated with TURN and LSO and cause multicollinearity in the model. This multicollinearity issue will be discussed later in Subsection 5.3.2. The higher the value this variable takes, the higher the limit-order execution probability and the shorter the expected time-to-execution.

LVO is a measure of the historical trading activity in the stock.[95] This variable is given by Equation (5.11) below:

$$LVO = \log(\text{previous month's average daily share volume}). \tag{5.11}$$

This variable can be potentially correlated with TURN, LSO and LPR, and cause multicollinearity in the model. This multicollinearity issue will be discussed later in Subsection 5.3.2. The higher the value this variable takes, the higher the limit-order execution probability and the shorter the expected time-to-execution.

The last three variables are ‘primitive’ variables designed to capture differences across stocks and are updated monthly. In contrast, the remaining variables capture the characteristics of a limit order and market conditions at the submission time.

The following are redefined variables for the sell model, which retain the underlying economic interpretation:

$$MKD1 = \begin{cases} (1 + P_l - P_o)\log S_o & \text{if } P_l \geq P_o \\ 0 & \text{if } P_l < P_o \end{cases}; \tag{5.12}$$

$$MKD1X = \begin{cases} (P - P_l)MKD1 & \text{if } P \leq P_l \\ 0 & \text{if } P > P_l \end{cases}; \tag{5.13}$$

[93] This variable is extracted from DataStream.
[94] Baker and Stein (2004) argue that the more liquid stock would have a somewhat higher price.
[95] This variable is extracted from DataStream.

$$MKD2 = \begin{cases} \log S_b /(1 + P_l - P_o) & \text{if } P_o \le P_l \\ \log S_b & \text{if } P_o > P_l \end{cases} \text{; and} \tag{5.14}$$

$$SZSD = \begin{cases} \log(S_l)(1 + P_l - P_b) & \text{if } P_l > P_b \\ \log(S_l - S_b) & \text{if } P_l = P_b \text{ and } S_l > S_b \\ 0 & \text{otherwise} \end{cases} . \tag{5.15}$$

Table 5.1 reports summary statistics of these explanatory variables constructed using the data sample described in Section 4.1.4.[96] There is not much difference across buy and sell orders. Summary statistics of these explanatory variables are to be compared to those reported in the LMZ and CHTZ studies.

The mean of MQLP shows that the limit-order price of buy orders is almost two pence below the mid-quote price while that of sell orders is almost two pence above the mid-quote price. Both buyers and sellers price their orders equally in terms of the distance from the mid-quote price and hence the sample period covers a market that is balanced (i.e., neither a 'buy' market nor a 'sell' market) in terms of this distance. In the LMZ study, the mean of MQLP is 0.25 dollars for buy limit orders and -0.04 dollars for sell orders. Hence, on average, the limit-order price of buy orders is 25 cents below the mid-quote price and that of sell orders is four cents above the mid-quote price. Thus their market is a 'sell' market, as sellers price their orders much closer to the mid-quote price than do buyers. In the CHTZ study, the buy (sell) limit-order price is three cents below (above) the mid-quote price. The orders in both this and the CHTZ studies are priced more aggressively than those in the LMZ study. Hence, traders in this and the CHTZ markets are more willing to trade, and higher execution probability and shorter execution time are expected.

The mean of BSID is -0.17 for buy orders and 0.12 for sell orders. This indicates that, on average, for buy limit orders it is sellers who initiate the prior trade and for sell limit orders it is buyers. In the LMZ study, the mean of BSID is -0.008 for buy orders and -0.003 for sell orders. Hence in their market, on average, for both buy and sell limit orders, it is sellers who initiate the prior trade. Accordingly their market is a 'sell' market as discussed above.

[96] LMZ reports summary statistics of their explanatory variables in LMZ (1999), but not in (2002).

The mean of MKD1 is 17.95 for buy orders and 17.03 for sell orders. This indicates that the number of shares that have higher priority for execution over buy orders is larger than that over sell orders. In the LMZ study, the mean of MKD1 is 2.14 for limit orders and 1.33 for sell orders. Hence, the number of shares that have higher priority for execution over buy orders is much larger than that over sell orders. Compared to those reported in the LMZ study, the higher values of MKD1 in this study could be due to larger best bid/offer sizes or a smaller tick size of the data sample used in this study as a smaller tick size could inflate the terms of $(1+P_b-P_l)$ in Equation (5.3) for buy orders and $(1+P_l-P_o)$ in Equation (5.12) for sell orders. In the CHTZ study, the mean of MKD1 is 6.17 for buy orders and 6.08 for sell orders. Compared to those reported in the CHTZ study, the higher values of MKD1 in this study could probably be due to larger best bid/offer sizes.

The mean of MKD1X is 88.35 for buy orders and -82.63 for sell orders. Since MKD1X is derived from MKD1, these values of MKD1X indicate that, on average, buy (sell) orders are priced below (above) the previous transaction price. This indicates that both buyers and sellers are ‘patient’ and price their orders away from the market price of the previous transaction, and would wait for better execution opportunities. In the LMZ study, the mean of MKD1X is 0.89 for buy orders and -0.11 for sell orders. Hence, buy (sell) orders are also priced below (above) the previous transaction price. However, in their study buy orders are priced much further away from previous transaction prices than sell orders. Compared to those reported in the LMZ study, the higher values of MKD1X in this study could be due to the higher values of MKD1 discussed above or a smaller tick size of the data sample used in this study as a smaller tick size could inflate the terms of $(P-P_l)$ in Equations (5.4) and (5.13).

The mean of MKD2 is 3.66 for buy orders and 7.90 for sell orders. This indicates that more liquidity is available on the buy side of the market. In the LMZ study, the mean of MKD2 is 1.75 for buy orders and 2.03 for sell orders. Hence, more liquidity is also available on the buy side of their market. The higher values of MKD2 reported in this study could be due to larger best bid/offer sizes. The higher value of MKD2 for sell orders could be due to the incorrect construction of this variable in the sell model.[97]

The mean of SZSD is 36.58 for buy orders and 35.43 for sell orders. This indicates that buy orders demand more liquidity than sell orders. In the LMZ study, the mean of SZSD is 2.25 for buy orders and 2.33 for sell orders. In their market, buy orders demand less liquidity than sell orders. Compared to those reported in the LMZ study, the higher values of SZSD reported in this study could be due to larger order sizes, or a smaller tick size of the data sample used in this study as a smaller tick size could inflate the terms of $(1+P_o-P_l)$ in Equation (5.5) for buy orders and $(1+P_l-P_b)$ in Equation (5.14) for sell orders. In the CHTZ study, the mean of SZSD is 5.58 for buy orders and 5.54 for sell orders. Compared to those reported in the CHTZ study, the higher values of SZSD reported in this study could be due to larger order sizes or the fact that buy (sell) orders are priced further away from the best offer (bid) price.

The mean of STKV is 0.54 for buy orders and 0.55 for sell orders. The number of trades during the last 30 minutes for buy (sell) orders is 54% (55%) of the number of trades in the last one hour. This indicates that the market becomes slightly more active in the 30 minutes prior to order submission in terms of the number of trades. Similarly, CHTZ report a mean of 0.53 for both buy and sell orders. LMZ report a mean 0.52 for buy orders and 0.50 for sell orders. Hence, for buy orders the LMZ market becomes slightly more active in the 30 minutes prior to order submission, but for sell orders it does not.

The mean of TURN is 3.88 for buy orders and 3.87 for sell orders. This indicates that there is roughly the same number of trades in the last one hour prior to order submission for both buy and sell orders. This also indicate that, in the data sample used in this study, the number of trades in the last 30 minutes prior to order submission is about 48 ($e^{3.875}$). In the LMZ study, the mean of TURN is 4.23 for buy orders and 4.22 for sell orders. Hence, in their data sample, the number of trades in the last 30 minutes prior to order submission is about 69 ($e^{4.235}$). Thus their market is slightly more active than this

[97] An email was sent to LMZ for clarification of the construction of MKD2 in the sell model. However no reply was received. CHTZ (2006) also point out the incorrect construction of MKD2 in the sell model.

market. In the CHTZ study, the mean of TURN is 11.48 for buy orders and 11.46 for sell orders. Hence, in their data sample, the number of trades in the last 30 minutes prior to order submission is about 9579 ($e^{11.47}$).[98] Thus their market is much more active than this market.

The mean of LSO is 14.91 for buy orders and 14.87 for sell orders. This indicates that the average number of shares outstanding at the end of the previous month is about 2,928,497,000 ($e^{14.89} \times 1000$). The mean of LPR is 6.46 for buy orders and 6.44 for sell orders. This indicates that, on average, the average daily closing price for the previous month is 632 ($e^{6.45}$) pence. The mean of LVO is 16.16 for buy orders and 16.15 for sell orders. This indicates that, on average, the previous month's average daily share volume is 10,375,938 ($e^{9.155}$).

Table 5.2 reports correlation matrices of the explanatory variables.[99] The cross-correlations of the explanatory variables are generally high, with most being greater than 30% in magnitude. This is a sign of multicollinearity, which will be investigated later in this chapter. For buy orders, the highest correlation between STKV that captures changing volatility and the variables related to the limit order, MQLP, BSID, MKD1, MKD1X, MKD2 and SZSD, is -3.68%. Similarly, the highest correlation between TURN and these variables is 25.85%. However, those variables related to the limit order are highly correlated with each other. For instance, the highest correlation is between MQLP and SZSD (94.37%), and the second highest is between MKD1 and MKD1X (86.70%). These high correlations are expected given the way in which the variables are defined by LMZ. For example, SZSD for buy orders includes the distance between the best offer price and the limit-order price ($P_o - P_l$) and MQLP captures the distance between the limit-order price and the mid-quote price ($P_q - P_l$). Therefore, SZSD already includes a part of MQLP. A similar observation holds true for sell orders. In contrast, the correlations across the explanatory variables reported in the

[98] It is suspected that in the CHTZ study TURN captures the volume rather than the number of trades in the last 30 minutes. An email was sent to the author for classification. However, no reply has been received.
[99] LMZ reports correlation matrices of their explanatory variables in LMZ (1999), and LMZ (2002) only mentions that the cross-correlations of their explanatory variables are generally low, with most being less than 30% in magnitude.

LMZ study are much lower than those reported in this study. The low correlations reported in the LMZ study are mainly due to differences between the data samples used.

This study identifies the following problems in the LMZ models. Firstly, MQLP is an absolute measure, which cannot capture the position of a limit order in the market well. This issue will be discussed later in Chapter 6. Secondly, there is a lack of rationales behind the construction of some variables (e.g. MKD1X).[100] Thirdly, the construction of MKD2 in the sell model could be incorrect as discussed early. Finally, the cross-correlations of the explanatory variables are generally high as Table 5.2 shows. This is a sign of multicollinearity, which will be investigated later in this chapter. It is necessary to interpretate the emprical findings cautiously when multicollinearity exists.

## 5.3 Empirical Results

This section presents empirical results of time-to-completion models that involve the explanatory variables suggested by LMZ. This section also provides coverage of goodness-of-fit tests and an investigation into multicollinearity in the models.

As LMZ (2002) suggest, an Accelerated Failure Time (AFT) model, which assumes that rescaling time can capture the effect of explanatory variables on survival times, is used to incorporate a vector of explanatory variables (see Subsection 4.2.5). Also as LMZ (2002) suggest, the generalised gamma distribution, which nests a number of other distributions as special cases, is chosen as the distribution of baseline survival times. It has two ancillary parameters, kappa and sigma. A Weibull distribution results when kappa is one, an exponential distribution when both kappa and sigma are one, and a lognormal distribution when kappa is zero. An AFT model with a generalised gamma distribution of baseline survival times is known as a Generalised Gamma (GG) AFT model.

[100] An email was sent to LMZ for clarification of the rationales behind the construction of some variables. However no reply was received. CHTZ (2006) also point out the lack of rationales and drop some of the LMZ variables.

### *5.3.1 Estimation Results*

Table 5.3 shows parameter estimates along with corresponding standard errors of the model in Equation (4.18) of Chapter 4. The estimates of the parameters associated with the conditioning variables generally are statistically significant for both models. These coefficients reflect the sensitivities of limit-order completion time to the corresponding explanatory variables.

*The Coefficients of MQLP*

The coefficient of MQLP is negative (-0.31) in the buy model and positive (0.22) in the sell model. The negative (positive) sign in the buy (sell) model means that the lower the limit-order price below the mid-quote price, the shorter (longer) the expected completion times for buy (sell) orders. These signs are unexpected and are against the basic price priority in an order-driven market as discussed in Chapter 2. This finding, together with the observed high correlations among the variables, as Table 5.2 presents, raises the question of multicollinearity in the models which will be discussed in Subsection 5.3.2. The Q-Q plots discussed later imply that the LMZ models could be wrongly specified. Hence it is necessary to interpretate these emprical findings cautiously. In the LMZ study, the reported coefficients of MQLP have the expected signs: positive for buy orders and negative for sell orders.

Assume that the estimated coefficients of MQLP are correct. These estimates indicate that traders could reduce the waiting time of limit-order completion by pricing their orders further away from the mid-quote price and hence reduce the associated opportunity cost. In this case, if the limit order were completed eventually, these traders would receive a better price than the market price since the order is priced further away from the mid-quote price. Hence pricing a limit order further away from the mid-quote price can not only reduce its completion time but also improve its execution price. Intuitively the market could not function and hence this could not be true.[101]

[101] If both buyers and sellers price their orders further away from the mid-quote price, the market spread will increase to a degree that no one will submit a market order. As a result, no trade will occur and the market will collapse.

*The Coefficients of BSID*

The coefficient of BSID is positive (0.47) in the buy model and negative (-0.51) in the sell model. For buy orders, the positive sign indicates that if the previous transaction has been seller-initiated (BSID equals to negative one), then a shorter completion time is expected. For sell orders, the negative sign indicates that if the previous transaction has been buyer-initiated (BSID equals to one), then a shorter completion time is expected. Therefore the reported signs are as expected.

Traders who have private information will use market orders. For example, assume that some traders know the actual loss figure of RBS just before this figure is made public.[102] These traders will use market orders to cash in this private information.

Assume that a seller use a market order to initiate the previous transaction. Perhaps this seller might have some private information, which suggests that this stock is overvalued and the share price might start to fall. Assume that more and more sellers have access to this information. Hence more sellers would enter the market and submit more market orders. As a result, the market becomes a 'sell' market since sellers are more willing to trade. Since more sellers are submitting market orders and these market orders will be executed with buy limit orders, the expected time-to-completion for these buy orders will be shorter.

This empirical finding indicates that traders should 'watch out' their counterparts on the opposite side of the market. If these traders notice an inflow of market orders from the opposite side of the market, they need to think carefully before jumping in to trade any shares as their orders may be 'picked up' quickly by informed counterparts. However this may also present an opportunity, since it would be easier to fill limit orders and the waiting time could also be reduced.

In the LMZ study, the coefficient of BSID is negative (-5.52) in the buy model and positive (6.77) in the sell model. For buy orders, the negative sign indicates that if the previous transaction has been buyer-initiated (BSID equals to one), then a shorter

[102] Royal Bank of Scotland (RBS) reported a record-breaking loss on 19 January 2009.

completion time is expected. In the data sample used by LMZ, institutional traders submit the majority of orders, and thus these orders are large orders**.** If buyers submit market orders to initiate the previous transactions, then buyers are more willing to trade than sellers. It could be argued that the market is a 'buy' market and so an additional large buy order can significantly increase the order imbalance (more buy orders than sell orders) in the market. Since the NYSE specialists are obliged to provide additional liquidity when the market is imbalanced, the submission of a large buy order could cause the specialists to provide the necessary liquidity and the order will consequently have a shorter expected completion time. For sell orders the positive sign indicates that if the previous transaction has been seller-initiated (BSID equals to negative one), then a shorter completion time is expected. Similarly, if sellers submit market orders to initiate the previous transactions, then sellers are more willing to trade than buyers. It could be argued that the market is already a 'sell' market and an additional large sell order can significantly increase the order imbalance (more sell orders than buy orders) in the market. The submission of a large sell order could consequently cause the specialists to provide additional liquidity giving this order a shorter expected completion time.

*The Coefficients of MKD1*

The coefficient of MKD1 is 0.10 in the buy model and 0.05 in the sell model. The positive sign suggests that the expected completion time increases with the minimum number of shares that have higher priority for execution. This is expected and consistent with the prediction of Parlour (1998).[103] Since unexecuted orders stay on the order book for subsequent execution, the higher the minimum number of shares that have higher priority for execution, the longer the expected completion time. If traders can access the full information of the order book, they should monitor the length of 'queue' (the minimum number of shares that have higher priority for execution) on the order book and submit limit orders when the 'queue' is shorter in order to reduce the waiting time.

Since both MKD1 and MQLP capture the position of a submitted order in the 'queue' on the order book, the estimated coefficients of these two variables should have the

[103] Parlour (1998) presents a model of an order-driven market, in which traders choose to submit a limit order or a market order depending on the state of the limit order book.

same interpretation. However the estimated coefficients of MKD1 have the opposite interpretation of those of MQLP discussed above. This indicates that the LMZ models may be wrongly specified. This issue will be discussed later.

In the LMZ study, the coefficient of MKD1 is 0.62 in the buy model and 0.46 in the sell model. Hence, in their study, limit-order completion time is more sensitive to MKD1. In the NYSE, specialists are required to provide liquidity when the market is imbalanced. For example, assume that the number of shares that have higher priority for execution on the buy side of the order book is 1000 shares for stock A at 10:00am and a buy order of 1000 shares is submitted. At this point, the market becomes so imbalanced that specialists are required to provide the additional liquidity to complete this buy order. The higher the number of shares that have higher priority for execution, the more likely it is that a market will become imbalanced and the more likely it is that specialists will provide additional liquidity and hence the shorter the expected completion time. Thus in the LMZ study MKD1 also captures the probability of specialists providing additional liquidity and hence LMZ limit-order completion time is more sensitive to this variable.

*The Coefficients of MKD1X*

The coefficient of MKD1X is -0.003 in the buy model and 0.002 in the sell model. Since MKD1X captures the distance between the last transaction price and the limit-order price, these estimates indicate that the further away the limit-order price from the last transaction price, the shorter the expected completion time. These signs are unexpected since the last transaction price could be a good proxy of subsequent transaction prices assuming that no new information arrives at the market. These unexpected signs indicate that the LMZ models could be wrongly specified. This issue will be discussed later.

Assume that the estimated coefficients of MKD1X are correct. These estimates indicate that traders could reduce the waiting time of limit-order completion by pricing their orders further away from the last transaction price and hence reduce the associated opportunity cost. If this were the case, both buyers and sellers would price their orders further away from the last transaction price. As a result, the market spread would increase to a degree that no one would submit market orders, no trade would occur and hence the market could not function.

In the LMZ study, the coefficient of MKD1X is -0.90 in the buy model and 0.94 in the sell model. Hence, their limit-order completion time is more sensitive to this variable. This is expected as MKD1X is derived from MKD1 and LMZ limit-order completion time is more sensitive to MKD1.

*The Coefficients of MKD2*

The coefficient of MKD2 is -0.35 in the buy model and -0.23 in the sell model. The negative signs of the estimates indicate that the greater the depth of the opposite side of the order book, the shorter the expected completion time.

The depth of the order book is a proxy of the liquidity on the order book. Hence these signs are expected since the more liquidity is available on the opposite side of the order book, the higher the limit-order execution probability and consequently the shorter the expected time-to-completion.

The depth of the order book could also be a proxy of traders' willingness to trade. For example, assume that at 10:00am the order book is empty. From 10:00am to 10:30am, 20 buy limit orders and five sell limit orders enter the market. As a result, more liquidity is available on the buy side of the market. It can be argued that during this period of time buyers are more willing to trade than sellers. It is possible that some buyers may become impatient, cancel their existing orders and submit market orders instead. Hence the expected completion times of existing sell limit orders will be shorter. This finding indicates that in order to reduce the waiting time traders should submit limit orders when more liquidity is available on the opposite side of the order book.

In the LMZ study, the coefficient of MKD2 is -0.33 in the buy model and -0.15 in the sell model. Accordingly in their study completion time of buy orders is more sensitive to this variable than that of sell orders. In the LMZ data sample there are more buy orders than sell orders. Thus, the LMZ market is a ‘buy’ market in which more buy orders are competing for the liquidity available from the sell side. Hence buy-order completion time in their study is more sensitive to this variable.

*The Coefficients of SZSD*

The coefficient of SZSD is 0.03 in the buy model and 0.04 in the sell model. The positive signs of the coefficient estimates indicate that the greater the liquidity demanded by the limit order, the longer the expected completion time. Intuitively it is more difficult and takes a longer period of time to fill a large order. Since SZSD captures the limit-order size scaled by the distance between the limit-order price and the best offer price, these signs are expected. This finding suggests that traders could reduce the waiting times to complete their orders by dividing large orders into small orders. However since SZSD does not just capture the limit-order size, the pure effect of limit-order size on limit-order completion time is still unknown. This issue will be addressed later in Chapter 6.

In the LMZ study, the coefficient of SZSD is 0.07 in the buy model and 0.19 in the sell model. Accordingly, in their study, completion time of buy orders is less sensitive to this variable than that of sell orders. Similar to the discussion above, the LMZ market is a ‘buy’ market and more buy orders are competing for execution. Hence, buy-order completion time is less sensitive to this variable.

*The Coefficients of STKV and TURN*

The coefficient of STKV is -1.17 in the buy model and -1.24 in the sell model. The coefficient of TURN is -0.75 in the buy model and -0.83 in the sell model. The negative signs on these coefficients imply that a shorter completion time is expected when the market is more active and volatile.

Higher volatility will generally increase the probability of limit-order completion and hence reduce limit-order completion time (benefit), since a limit order placed away from the current market price will fill when the limit-order price is first met. However the higher volatility may also increase the probability of being ‘picked up’ by a counterpart with better information from the opposite side of the market (cost). Due to the potential cost of being 'picked up' traders could switch from limit orders to market orders. In addition, if the market moves against submitted limit orders, traders may have to cancel these limit order and use market orders to chase the price. As a result, the completion probability of existing limit orders will be improved since these orders will be executed with market orders from the opposite side of the market. From what have been discussed, the negative signs on the coefficients of STKV and TURN are expected.

In general the probability and time of limit-order completion depend on the overall volatility and the cost of being 'picked up' depends on the 'information-driven' component of the volatility since the probability of being 'picked up' depends on the arrival of informed traders from the opposite side of the market. It can be argued that 'extreme' volatility will discourage the use of limit orders since the cost of being 'picked up' exceeds the benefit. If the volatility becomes extremely high, the market could fail to function since no one would submit limit orders and the order book would be empty. Although higher volatility will generally reduce the expected waiting time of limit-order completion, traders should be always cautious and careful when submitting a limit order in a volatile market since the cost could be far higher than the benefit.

In the LMZ study, the coefficient of STKV is -0.39 in the buy model and -0.57 in the sell model, and the coefficient of TURN is -0.26 in the buy model and -0.33 in the sell model. Accordingly in their study limit-order completion time is less sensitive to these variables. This probably is because the orders in their data sample are large, and large orders in the NYSE could be less sensitive to market volatility in terms of the execution probability and time. For example, we assume that a large buy order of 10000 shares is submitted at 10:00am. At this point, the market becomes so imbalanced that specialists are required to provide the additional liquidity to complete this buy order. Thus the execution of a large order in the NYSE depends more on the market imbalance and the involvement of specialists.

*The Coefficients of LSO, LPR and LVO*

The estimates of the coefficients on the three variables that capture the cross-stock differences are significant in both models. The coefficient of LSO is 0.61 in the buy model and 0.40 in the sell model. The positive signs indicate that the higher the number of shares outstanding, the longer the expected completion time. This is unexpected since the greater the number of shares, the more liquid the stock should be. These unexpected signs indicate that the LMZ models could be wrongly specified. This issue will be discussed later. The coefficient of LPR is -0.66 in the buy model and -0.36 in the sell model. The negative sign on the estimates of these coefficients indicates that the higher the price, the shorter the expected completion time. This is expected as higher priced stocks tend to be more liquid as discussed above.[104] The coefficient of LVO is -0.53 in the buy model and -0.29 in the sell model. The negative sign on the estimates of these coefficients indicates that the higher the share volume, the shorter the expected completion time. This is expected since stocks with higher share volume traded tend to be more liquid.

Simplifications of the generalised gamma to the Weibull (kappa is one) or the exponential (both kappa and sigma are one) distribution are strongly rejected. As Table 5.3 reports, the coefficient of Kappa is 0.00006 with a standard error of 0.01 in the buy model and 0.012 with a standard error of 0.01 in the sell model. Accordingly the estimates of the shape parameter (kappa) for the buy and sell models are about 100 standard errors from one. The coefficient of sigma is 2.26 with a standard error of less than 0.01 in the buy model and 2.22 with a standard error of less than 0.01 in the sell model. Hence the estimated scale parameter (sigma) for buy and sell models is at least more than 120 standard errors from one. Simplifications of the generalised gamma to the lognormal distribution (kappa is zero) cannot be strongly rejected, since the z-statistic of the estimated shape parameter (kappa) is 2.03 for sell orders and -0.01 for buy orders.

To check the goodness-of-fit of the Generalised Gamma AFT model estimated above, the graphical diagnostic (Q-Q plot) discussed in Section 4.2.6 is used. Since the empirical survival function is subject to sampling variation, an exact straight line is not

[104] In the UK market, the most illiquid stocks are those ‘penny’ shares. Hence the share price could be a proxy of liquidity.

expected. However if the model is correctly specified, the plot should show points closely clustered about the 45-degree line. In general Q-Q plots that deviate significantly from the 45-degree line are an indication of model misspecification. Figure 5.1 shows Q-Q plots for both models and these plots obviously deviate from the 45-degree line. These indicate that the LMZ models could be wrongly specified. There are only about 20 limit orders with Cox-Snell residuals greater than 10 in each model and hence are suspected to be the main cause of the deviation from the 45-degree line. Figure 5.2 shows Q-Q plots after excluding these orders. Although the fitness is greatly improved, the models still do not fit the data as well as they have been reported in the LMZ study. [105] The next subsection will therefore investigate further the multicollinearity problem that was reported earlier and suspected to be the main culprit.

### *5.3.2 Multicollinearity Effects*

Multicollinearity is a statistical phenomenon in which two or more explanatory variables in a multiple regression model are highly correlated. In this situation the coefficient estimates may change erratically in response to small changes in the model or the data. Perfect multicollinearity occurs if the correlation between two explanatory variables is equal to one or negative one. In practice, perfect multicollinearity rarely occurs in a data sample. More commonly, the issue of multicollinearity arises when there is a high degree of correlation (either positive or negative) between two or more explanatory variables.

*Multicollinearity in a Bivariate Model*

This subsection starts the presentation through an example of a bivariate linear regression model:

$$y = \beta_0 + \beta_1 X_1 + B_2 X_2 + e \ , \qquad (5.16)$$

where y is a dependent variable, $X_1$ and $X_2$ are two explanatory variables and e is a random error term. $\beta_0$, $\beta_1$ and $\beta_2$ are coefficients.

[105] In the LMZ study, Q-Q plots shows points closely clustered about the 45-degree line.

Consider a simple analysis where $X_2$ is a direct linear function of $X_1$:

$$X_2 = c + dX_1, \tag{5.17}$$

where c and d are constants.

Substituting Equation (5.17) in Equation (5.16), one gets

$$y = \beta_0 + \beta_1 X_1 + \beta_2 (c + dX_1) + e = \alpha_1 + \alpha_2 X_1 + e, \tag{5.18}$$

where $\alpha_1$ is the intercept and $\alpha_2$ is the slope.

The intercept $\alpha_1$ and the slope $\alpha_2$ can be estimated. The relationship between the coefficients in Equations 5.16 and 5.18 is

$$\alpha_1 = \beta_0 + \beta_2 c \text{, and} \tag{5.19}$$

$$\alpha_2 = \beta_1 + \beta_2 d\ . \tag{5.20}$$

The coefficients c and d in the exact relation between $X_1$ and $X_2$ can be easily computed. However from the estimates of $\alpha_1$ and $\alpha_2$, it is impossible to solve for $\beta_0$, $\beta_1$ and $\beta_2$ since there are two equations and three unknowns. This means that there is no unique estimator for $\beta_0$, $\beta_1$ and $\beta_2$ that minimises the sum of squared errors (or maximises the likelihood function). Therefore it is impossible to distinguish between the effects of two variables that have an exact linear relation.

Let $\hat{\beta}_1$ be the estimator of $\beta_1$ and $\hat{\beta}_2$ be the estimator of $\beta_2$. The variances of $\hat{\beta}_1$ and $\hat{\beta}_2$ are as follows:

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{S_1(1-r^2)}; \tag{5.21}$$

$$Var(\hat{\beta}_2) = \frac{\sigma^2}{S_2(1-r^2)}; \tag{5.22}$$

where $\sigma^2$ is the variance of the error term e, $S_1$ and $S_2$ are the sample variances of $X_1$ and $X_2$ and r is the correlation coefficient between these variables, respectively.

If the correlation coefficient is close to one or negative one (perfect multicollinearity), then the variance of $\hat{\beta}_1$ and $\hat{\beta}_2$ will become very large.

*Multicollinearity in a Multivariate Model*

In practice, usually a multivariate regression model, which has more than two explanatory variables, is used. Griffiths, Hill and Judge (1993), Kennedy (2003) and McIntosh (2007) explain the multicollinearity in a multivariate regression model as follows. A standard multivariate linear regression model is presented as

$$y = X\beta + e \ , \tag{5.23}$$

where e is a random error term, $\beta$ is a vector of coefficients,

$$X = \begin{bmatrix} 1, X_{11} & \dots & X_{1k} \\ ., \quad . & \dots & . \\ 1, X_{n1} & \dots & X_{nk} \end{bmatrix} , \tag{5.24}$$

n is the number of observations and k represents the number of explanatory variables.

The classical regression model’s OLS solution for computing the vector of coefficients $\beta$ necessitates the following step:

$$X'X\hat{\beta} = X'y ; \tag{5.25}$$

$$\hat{\beta} = (X'X)^{-1} X'y ; \tag{5.26}$$

where $X'$ is the transpose of X and $\hat{\beta}$ is the estimator of $\beta$.

Going from Equation (5.25) to Equation (5.26) obviously requires $X'X$ to be invertible and the inverse $(X'X)^{-1}$ actually to exist. In other words, $X'X$ has to be square and ‘non-singular’.

X can be rewritten as:

$$X = (1, x_1, x_2, \dots x_k) , \tag{5.27}$$

where $x_i$ is the i+1th column of the X matrix.

Perfect multicollinearity is said to exist when the X matrix is not of full rank; that is, it has rank less than k+1. This occurs when one or more exact relations exist among the columns of X. That is, there are one or more relations of the form

$$Xc = c_0 + x_1c_1 + x_2c_2 + \cdots + x_kc_k = 0, \tag{5.28}$$

where $c = (c_0,\ldots,c_k)'$ is a vector of constants not all of which are zero. Alternatively, assuming $c_1$ is not zero:

$$x_1 = (-c_0/c_1) + x_2(-c_2/c_1) + \cdots + x_k(-c_k/c_1). \tag{5.29}$$

So that the first variable can be written as an exact linear combination of the rest, and thus the columns of X are linearly dependent. From the linear algebra, it is known that when one column of X is a linear combination of the other columns of X, $X'X$ will be singular and the rank of X is less than k+1 (X is said to not be of 'full' rank) and consequently the OLS estimators do not exist since the normal equations cannot be solved uniquely for $\hat{\beta}$ because $X'X$ cannot be inverted. This means that there is no unique estimator for the vector $\beta$ that minimises the sum of squared errors (or maximises the likelihood function).

Let us now consider the case where the linear relationship between the columns of X is not an exact one, so that there exists one or more relations of the form

$$Xc = c_0 + x_1c_1 + x_2c_2 + \cdots + x_kc_k \approx 0. \tag{5.30}$$

where $\approx$ means 'almost equal to'. Suppose that $c_1 \neq 0$ so that

$$x_1 \approx (-c_0/c_1) + x_2(-c_2/c_1) + \cdots + x_k(-c_k/c_1), \tag{5.31}$$

or

$$x_1 = d_0 + x_2d_2 + \cdots + x_kd_k + u_1, \tag{5.32}$$

where $d_i = -c_i/c_1$ and $u_1$ is just the difference between the right and left hand sides of Equation (5.31). Relations where one explanatory variable is written as a linear function of the other explanatory variables plus the difference are called 'auxiliary regressions', since they do have the form of a regression equation. It is intuitively clear that the better the fit of this auxiliary regression of the $x_1$ on the other (k-1) explanatory variables, the more nearly exact the relationship and the more 'severe' the multicollinearity. In fact if it fits exactly then it would hold exactly and X would not be of full rank.

Let us partition X = (x1X2) where $x_1$ is the second column of X and X2 contains the other k columns. Define $N = I - X_2(X_2'X_2)^{-1}X_2'$, then

$$(X'X)^{-1} = \frac{1}{x_1'Nx_1} \times \left[ \begin{array}{c:c} 1 & -x_1'X_2(X_2'X_2)^{-1} \\ \hdashline -(X_2'X_2)^{-1}X_2'x_1 & x_1'Nx_1(X_2'X_2)^{-1} + (X_2'X_2)^{-1}X_2'x_1x_1'X_2(X_2'X_2)^{-1} \end{array} \right]. \quad (5.33)$$

If $\hat{\beta} = (X'X)^{-1}X'y$ is the OLS estimator then the variance-covariance matrix for the estimator $\hat{\beta}$ is $\text{cov}(\hat{\beta}) = \sigma^2(X'X)^{-1}$. The variances of the parameter estimates $\hat{\beta}$ are the diagonal vector from the variance-covariance matrix. Using Equation (5.33), the variance of $\hat{\beta}_1$ is

$$\text{var}(\hat{\beta}_1) = \frac{\sigma^2}{x_1'Nx_1} = \frac{\sigma^2}{\hat{u}_1'\hat{u}_1}, \quad (5.34)$$

where $\sigma^2$ is the variance of the error term e in Equation (5.23) and $\hat{u}_1$ is the vector of residuals from estimating of the auxiliary regression equation. Consequently, the variance of $\hat{\beta}_1$ depends on $\sigma^2$ and the variation in $x_1$ not explained by the linear influence of the other explanatory variables. Thus the better the fit of the auxiliary regression the larger $\text{var}(\hat{\beta}_1)$ since the better the fit, the smaller is $\hat{u}_1'\hat{u}_1$.

The consequences of multicollinearity in the above sample are:

- Perfect multicollinearity resulting from an exact linear relationship between two columns of the X matrix causes $X'X$ to be singular, and the OLS estimator $\hat{\beta} = (X'X)^{-1}X'y$ is not defined. This means that there is no unique estimator for the vector $\beta$ that minimises sum of squared errors (or maximises the likelihood function).
- It becomes very difficult to identify the separate effects of the variables involved precisely. In fact, since the regression coefficients are interpreted as reflecting the effects of changes in their corresponding variables, all other things held constant, the ability to interpret the coefficients declines the more persistent and severer the multicollinearity is.

- The estimated standard deviations of estimators become large when the explanatory variables in the regression model are highly correlated with each other. Even though a definite statistical relation exists between the dependent variable and the vector of explanatory variables, the estimates of unknown parameters may not appear significantly different from zero, because the sample is inadequate to isolate the effect precisely. And adding or deleting an explanatory variable could change the estimated coefficients of correlated explanatory variables.

In general, when high multicollinearity is present, confidence intervals for estimators tend to be very wide and z-statistics tend to be very small. Estimators will have to be larger in order to be statistically significant and it will be harder to reject the null model when multicollinearity is present.

Although there have been many suggestions about how to detect multicollinearity, unfortunately most are ineffective rules of thumb. Usually high sample correlations between explanatory variables are the sign of multicollinearity. Coefficients change is also a sign of multicollinearity when a variable is dropped. Low z-statistics on the individual estimators can also indicate the multicollinearity.

A problem that arises in this study is the existence of multicollinearity in the LMZ explanatory variables as discussed in Section 5.2. The multicollinearity could be due to various reasons. For instance, a variable included in the model is computed from other variables, such as the variable MKD1X in the LMZ models is constructed from MKD1. As discussed above, multicollinearity could make these estimators unreliable and inflate standard errors (c.f., Leeflang *et al*. (2000) and Morrow-Howell (1994)). Economic variables will always have some correlations with one another. However the key question is whether the multicollinearity is so severe as to alter the estimation of the coefficients. The remainder of this subsection will investigate the severity of multicollinearity in the LMZ models.

Before investigating the multicollinearity in the LMZ model, some literature on multicollinearity in survival model is reviewed. Some research has proposed procedures to identify multicollinearity and its effects on survival model estimates. Lariviere and Poel (2005) introduce a new variable 'rec_corr' that has a correlation of 97% with the variable 'recency' in order to understand better the impact of higher inter-correlations on parameter estimates produced by their survival model. They reveal that the introduction of the highly correlated 'rec_corr' variable dramatically changes the parameter estimate of the variable 'recency'. Their findings indicate that survival models also suffer from the effects of multicollinearity (like traditional statistical techniques) and suggest that researchers should be cautious when interpreting the estimates of highly inter-correlated variables in survival models. Mitra and Golder (2002) sequentially add variables to their survival model to assess the stability of parameters and hence to ensure that multicollinearity has no harmful impact on their results. They find that multicollinearity does not pose a problem in their results when the correlation between the variables is below 40%. Lariviere and Poel (2003) handle different correlation cut-offs starting from 20%, and then sequentially introduce new variables into the model based on a less restrictive (and hence higher) cut-off point. They show that 80% is too high to be a correlation cut-off value for their study, whereas lower cut-off values tend to result in more stable parameter estimates, and a 40% correlation cut-off seems to be an appropriate benchmark. All the above studies are carried out with the PH model discussed in Subsection 4.2.5.

Section 5.3.1 shows that the estimates of the coefficients on MQLP have an unexpected sign and Table 5.3 shows that the explanatory variables are highly correlated with each other. These are signs of multicollinearity. In order to investigate the extent of the multicollinearity effect on parameters estimated in the LMZ models this study excludes, one at a time, the two most inter-correlated variables (SZSD and MKD1X), from the models.

Table 5.4 presents estimation results when MKD1X is excluded. Compared to those reported in Table 5.3, all coefficient estimates have the same sign and, except ones of MQLP, MKD1 and SZSD, have similar magnitude.

MQLP has a correlation of 78.16% with MKD1X for buy orders and 78.84% for sell orders as Table 5.2 presents. The coefficient estimate of MQLP is still negative in the buy model and positive in the sell model. It is lower by about 59% (from -0.31 to -0.49) in the buy model and higher by about 78% (from 0.22 to 0.39) in the sell model.

MKD1 has a correlation of 86.70% with MKD1X for buy orders and -87.83% for sell orders as Table 5.2 presents. The coefficient estimate of MKD1 is still positive. It is lower by about 27% (from 0.10 to 0.08) in the buy model and about 40% (from 0.05 to 0.03) in the sell model.

SZSD has a correlation of 62.34% with MKD1X for buy orders and -63.23% for sell orders as Table 5.2 presents. The coefficient estimate of SZSD is still positive. It is higher by about 33% (from 0.03 to 0.04) in the buy model and by about 25% (from 0.04 to 0.05) in the sell model.

In the buy model, the estimate of kappa becomes significant and is higher by about 116567% (from -0.00006 to 0.07). In the sell model, it is higher by about 500% (from 0.01 to 0.06). This shows that multicollinearity could also affect estimated shape of the baseline survival time distribution.

Table 5.5 presents estimation results when SZSD is excluded. Compared to those reported in Table 5.3, all estimates have the same sign and, except ones of MQLP and MKD1X, have similar magnitude.

MQLP has a correlation of 94.37% with SZSD for buy orders and -94.33% for sell orders as Table 5.2 presents. The coefficient estimate of MQLP becomes positive for buy orders and negative for sell orders. It is higher by about 68% (from -0.31 to 0.10) in the buy model and lower by about 286% (from 0.22 to -0.41) in the sell model.

MKD1X has a correlation of 62.34% with SZSD for buy orders and -63.23% for sell orders as Table 5.2 presents. The coefficient estimate of MKD1X is still negative for buy orders and positive for sell orders. It is lower by about 15% (from -0.0026 to 0.0030) in the buy model and higher by about 40% (from 0.002 to 0.0028) in the sell model.

In the buy model, the estimate of kappa becomes significant and is lower by about 83233% (from -0.00006 to -0.05). In the sell model, it is lower by about 600% (from 0.01 to -0.05). This also shows that multicollinearity could also affect estimated shape of the baseline survival time distribution.

These results discussed above show that the LMZ models suffer from the effects of multicollinearity even with the large dataset used here, as some coefficient estimates change significantly. Thus applying the variables defined by LMZ to UK market data has caused problems of multicollinearity. Accordingly, in pursuing the objects of properly modelling order completion time in the UK market, the next chapter presents a set of survival models that use constructed explanatory variables to solve the multicollinearity problem inherent in the LMZ variables.

This study shows that the LMZ models suffer from the multicollinearity effect and the LMZ variables do not capture the state of the order book well. Hence new explanatory variables will be constructed in the next chapter. In addition, the LMZ variables are fixed at the order submission time and hence the LMZ variables do not capture the dynamics of the order book. This could be one of the reasons behind the poor goodness-of-fit revealed in this chapter. This issue will be addressed in Chapter 7.

In this chapter, heteroscedasticity has not been tested on the pooled data for the following reasons.[106] Firstly, any traditional heteroscedasticity tests cannot handle one aspect of the survival data: censoring.[107] As far as we know, no formal heteroscedasticity tests have been put forward in the literature on survival analysis reviewed in this study. Secondly, this study provides a new approach to test

[106] Heteroscedasticity means a situation in which the variance of the dependent variable varies across the data.

[107] Since traditional heteroscedasticity tests cannot handle censoring aspect of the survival data, in this chapter this study tests heteroscedasticity on data excluding censored orders. Without incorporating the censored data, the AFT model can be transformed into a loglinear model. Hence Equation (4.18) may be re-written as: $\ln(T) = \beta X^{'} + \varepsilon$. Then the Breusch-Pagan test is used to test for heteroscedasticity. This test is based on testing whether the estimated variance of the residuals from a linear regression depends on the values of the independent variables. In other words, it tests whether the variance of the residuals is homogenous. The Breusch-Pagan test is a chi-square test under the null hypothesis that the variance of the residuals is homogenous. The result of the Breuch-Pagan test on data including completed orders only suggests that heteroskedasticity does exist since the chi-square distribution is statistically significant (it is 222.70 for buy orders and 271.48 for sell orders). Hence the null hypothesis is rejected. This study will also test heteroskedasticity on the pooled data in Chapter 7.

heteroscedasticity on survival data in Chapter 7. Accordingly this issue will be discussed in Chapter 7.

Also since the LMZ models have the problems mentioned above, it is expected that addressing the heteroscedasticity issue alone would not improve the goodness-of-fit to a degree that the Q-Q plots could show. Hence the heteroscedasticity issue will be addressed in Chapter 7, as new explanatory variables need to be constructed first. The next chapter will focus on the construction of new explanatory variables.

## 5.4 Summary and Conclusions

The main aim of this chapter is to investigate the degree to which explanatory variables suggested by LMZ capture the state of the UK order book and affect limit-order completion time. This chapter shows that the data sample used in this thesis is significantly different from the LMZ data sample due to the differences of markets, data sources and data sample periods. In this chapter, explanatory variables used in the LMZ study are constructed. Unlike LMZ who model time-to-first-fill and time-to-completion, this study models time-to-completion only, which is of primary interest to traders as their aims normally are to complete their orders. However, separate models are constructed for buy and sell orders.

LMZ variables have the interpretation as follows. MQLP is a measure of the distance between the limit-order price and current midpoint. BSID is an indicator of whether the prior trade was buyer-initiated or seller-initiated. MKD1 measures the number of shares that have higher priority for execution. MKD1X is an interactive term to capture non-linearity between market depth and market price relative to the limit-order price. MKD2 is a measure of liquidity available from the opposite side of the market. SZSD is a measure of liquidity demanded by the limit order. STKV is used to capture recent shifts in trading activity. TURN is an absolute measure of volatility. In the monthly updated variables, LSO is the logarithm of the number of shares outstanding, LPR is the logarithm of share price and LVO is the logarithm of average daily volume. They are ‘primitive’ variables included to capture differences across stocks. As Table 5.2 shows, the cross-correlations of these explanatory variables are generally high, with most being greater than 30% in magnitude.

In this chapter, as LMZ (2002) suggest, an AFT model is used to incorporate a vector of explanatory variables and the generalised gamma distribution, which nests a number of other distributions as special cases, is chosen as the distribution of baseline survival times. It has two ancillary parameters, kappa and sigma. A Weibull distribution results when kappa is one, an exponential distribution when both kappa and sigma are one and a lognormal distribution when kappa is zero. Table 5.3 shows parameter estimates along with corresponding standard errors. The estimates of the parameters associated with the conditioning variables generally are statistically significant and have expected sign, except ones of MQLP, for both models.

Simplifications of the generalised gamma to the Weibull (kappa is one) or the exponential (both kappa and sigma are one) distribution are strongly rejected. As Table 5.3 reports, the coefficient of Kappa is 0.00006 with a standard error of 0.01 in the buy model and 0.012 with a standard error of 0.01 in the sell model. Accordingly the estimates of the shape parameter (kappa) for the buy and sell models are about 100 standard errors from one. The coefficient of sigma is 2.26 with a standard error of less than 0.01 in the buy model and 2.22 with a standard error of less than 0.01 in the sell model. Hence, the estimated scale parameter (sigma) for buy and sell models is at least more than 120 standard errors from one. Simplifications of the generalised gamma to the lognormal distribution (kappa is zero) cannot be rejected strongly, since the z-statistic of the estimated shape parameter (kappa) is 2.03 for sell orders and -0.01 for buy orders.

To check the goodness-of-fit of the Generalised Gamma AFT model estimated above, the graphical diagnostic (Q-Q plot) discussed in Section 4.2.6 is used. Since the empirical survival function is subject to sampling variation, an exact straight line is not expected. However, if the model is correctly specified then the plot should show points closely clustered about the 45-degree line. In general, Q-Q plots that deviate significantly from the 45-degree line are an indication of model misspecification. Figure 5.1 shows Q-Q plots for both models and these plots obviously deviate from the

45-degree line. There are only about 20 limit orders with Cox-Snell residuals greater than 10 in each model and hence these are suspected to be the main cause of the deviation from the 45-degree line. Figure 5.2 shows Q-Q plots after excluding these orders. Although the fitness is greatly improved, the models still do not fit the data as well as they have been reported in the LMZ study.

In order to investigate the extent of the multicollinearity effect on estimated parameters in the LMZ models, this study excludes, one at a time, the two most inter-correlated variables, SZSD and MKD1X, from the models. The results show that the LMZ models suffer from the effects of multicollinearity even with the large dataset used here, as some coefficient estimates change significantly. Accordingly, in pursuing the objects of properly modelling order completion time in the UK market the next chapter presents a set of survival models that use explanatory variables constructed to solve the multicollinearity problem inherent in the LMZ variables. In the next chapter, as LMZ (1999) suggest, both the Generalised Gamma AFT and Cox PH models are used to incorporate these constructed variables, since they are the most frequently used models in survival analysis.

**Table 5.1: Summary Statistics of the LMZ Explanatory Variables**

This table contains summary statistics of the explanatory variables for limit orders of a pooled sample of 38 FTSE 100 stocks, throughout a sample period from October 2000 to December 2000, and the LMZ and CHTZ samples. The variables have the interpretation as follows. MQLP is a measure of the distance between the limit-order price and current midpoint. BSID is an indicator of whether the prior trade was buyer-initiated or seller-initiated. MKD1 measures the number of shares that have higher priority for execution. MKD1X is an interactive term to capture non-linearity between market depth and market price relative to the limit-order price. MKD2 is a measure of liquidity available from the opposite side of the market. SZSD is a measure of liquidity demanded by the limit order. STKV is used to capture recent shifts in trading activity. TURN is an absolute measure of volatility. In the monthly updated variables, LSO is the logarithm of the number of shares outstanding, LPR is the logarithm of share price and LVO is the logarithm of average daily volume. P refers to pence, D refers to dollars, C refers to cents, T refers to trades and S refers to shares.

| | **UK** | | | **US (LMZ)** | | | **US (CHTZ)** | | |
|---|---|---|---|---|---|---|---|---|---|
| **Variable** | **Units** | **Mean** | **Std. Dev.** | **Units** | **Mean** | **Std. Dev.** | **Units** | **Mean** | **Std. Dev.** |
| | | | | | | | | | |
| **Buy Orders** | | | | | | | | | |
| **MQLP** | P | 2.05 | 2.78 | D | 0.25 | 0.30 | C | 3.05 | 3.09 |
| **BSID** | N/A | -0.17 | 0.94 | N/A | -0.01[1] | 0.10 | N/A | -0.02 | 0.97 |
| **MKD1** | P*logS | 17.95 | 23.54 | D*logS | 2.14 | 2.65 | C*logS | 6.17 | 1.05 |
| **MKD1X** | $P^2$*logS | 88.35 | 340.28 | $D^2$*logS | 0.89 | 1.83 | | | |
| **MKD2** | LogS/P or LogS | 3.66 | 2.30 | LogS/D or LogS | 1.75 | 1.60 | LogS/C or LogS | 5.73 | 0.96 |
| **SZSD** | P*logS or LogS | 36.58 | 32.37 | D*logS or LogS | 2.25 | 1.49 | P*logS or LogS | 5.58 | 0.87 |
| **STKV** | N/A | 0.54 | 0.15 | N/A | 0.52 | 0.12 | N/A | 0.53 | 0.13 |
| **TURN** | logT | 3.88 | 0.88 | logT | 4.29 | 0.87 | logT | 11.48 | 0.55 |
| **LSO** | Log 1000S | 14.91 | 1.15 | Log 1000S | 12.79 | 0.81 | | | |
| **LPR** | LogP | 6.46 | 0.67 | LogD | 3.94 | 0.38 | | | |
| **LVO** | LogS | 16.16 | 1.09 | LogS | 13.54 | 0.78 | | | |

## Table 5.1 continued …

| Sell Orders | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **MQLP** | P | -1.94 | 2.70 | D | -0.04 | 0.09 | C | -2.95 | 3.00 |
| **BSID** | N/A | 0.12 | 0.95 | N/A | 0.00[2] | 0.08 | N/A | 0.08 | 0.96 |
| **MKD1** | P*logS | 17.03 | 22.46 | D*logS | 1.33 | 2.00 | C*logS | 6.08 | 1.15 |
| **MKD1X** | $P^2$*logS | -82.63 | 336.78 | $D^2$*logS | -0.11 | 0.25 | | | |
| **MKD2** | LogS/P or LogS | 7.90 | 3.05 | LogS/D or LogS | 2.03 | 1.78 | LogS/C or LogS | 5.80 | 0.93 |
| **SZSD** | P*logS or LogS | 35.43 | 31.24 | D*logS or LogS | 2.33 | 1.48 | P*logS or LogS | 5.54 | 0.86 |
| **STKV** | N/A | 0.55 | 0.15 | N/A | 0.50 | 0.12 | N/A | 0.53 | 0.13 |
| **TURN** | logT | 3.87 | 0.88 | logT | 4.20 | 0.87 | logT | 11.46 | 0.55 |
| **LSO** | Log 1000S | 14.87 | 1.14 | Log 1000S | 12.75 | 0.78 | | | |
| **LPR** | LogP | 6.44 | 0.66 | LogD | 3.94 | 0.40 | | | |
| **LVO** | LogS | 16.15 | 1.09 | LogS | 13.53 | 0.80 | | | |

[1] -0.008
[2] -0.003

**Table 5.2: Correlation Matrices of the LMZ Explanatory Variables**

This table contains correlation matrices of the explanatory variables for limit orders of a pooled sample of 38 stocks, throughout a sample period from October 2000 to December 2000. The variables have the interpretation as follows. MQLP is a measure of the distance between the limit-order price and current midpoint. BSID is an indicator of whether the prior trade was buyer-initiated or seller-initiated. MKD1 measures the number of shares that have higher priority for execution. MKD1X is an interactive term to capture non-linearity between market depth and market price relative to the limit-order price. MKD2 is a measure of liquidity available from the opposite side of the market. SZSD is a measure of liquidity demanded by the limit order. STKV is used to capture recent shifts in trading activity. TURN is an absolute measure of volatility. In the monthly updated variables, LSO is the logarithm of the number of shares outstanding, LPR is the logarithm of share price and LVO is the logarithm of average daily volume.

| | | **Sell Orders** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **MQLP** | **BSID** | **MKD1** | **MKD1X** | **MKD2** | **SZSD** | **STKV** | **TURN** | **LSO** | **LPR** | **LVO** |
| | **MQLP** | 100.00% | 7.89% | -81.87% | 78.84% | 64.35% | -94.33% | -1.54% | 6.46% | 14.79% | -35.64% | 21.74% |
| | **BSID** | 4.71% | 100.00% | -6.44% | 8.69% | 11.56% | -6.88% | 0.37% | 5.45% | 5.20% | -2.64% | 5.78% |
| | **MKD1** | 81.50% | 2.61% | 100.00% | -87.83% | -69.03% | 65.06% | -2.13% | 7.02% | 9.43% | 27.01% | -5.68% |
| | **MKD1X** | 78.16% | 6.00% | 86.70% | 100.00% | 47.91% | -63.23% | 1.12% | -2.01% | -1.44% | -20.17% | -5.68% |
| **Buy Orders** | **MKD2** | -62.05% | -8.25% | -36.72% | -29.32% | 100.00% | -52.28% | 1.03% | 7.03% | 15.45% | -32.98% | 24.20% |
| | **SZSD** | 94.37% | 4.55% | 64.60% | 62.34% | -65.03% | 100.00% | 3.27% | -7.87% | -18.32% | 35.34% | -22.96% |
| | **STKV** | 1.66% | -0.78% | -1.95% | -0.79% | -3.68% | 3.35% | 100.00% | -8.36% | -6.20% | -2.47% | -4.78% |
| | **TURN** | -4.88% | -2.68% | 9.20% | 4.38% | 25.85% | -6.85% | -7.44% | 100.00% | 62.05% | 1.54% | 59.28% |
| | **LSO** | -12.85% | -1.38% | 11.81% | 2.72% | 46.58% | -16.88% | -6.37% | 61.22% | 100.00% | -16.90% | 82.16% |
| | **LPR** | 38.14% | 1.66% | 30.30% | 21.69% | -45.26% | 37.17% | -2.62% | 0.20% | -15.13% | 100.00% | -50.28% |
| | **LVO** | -22.67% | -2.00% | -6.75% | -7.51% | 46.23% | -23.94% | -4.62% | 59.19% | 81.14% | -50.87% | 100.00% |

**Table 5.3: Parameter Estimates of the LMZ Models**

The table below shows parameter estimates of the accelerated failure time specification of limit-order completion under the generalised gamma distribution for limit orders of a pooled sample of 38 stocks, throughout a sample period from October 2000 to December 2000. The variable '_cons' denotes the intercept and the definitions of the remaining explanatory variables are given as follows. MQLP is a measure of the distance between the limit-order price and current midpoint. BSID is an indicator of whether the prior trade was buyer-initiated or seller-initiated. MKD1 measures the number of shares that have higher priority for execution. MKD1X is an interactive term to capture non-linearity between market depth and market price relative to the limit-order price. MKD2 is a measure of liquidity available from the opposite side of the market. SZSD is a measure of liquidity demanded by the limit order. STKV is used to capture recent shifts in trading activity. TURN is an absolute measure of volatility. In the monthly updated variables, LSO is the logarithm of the number of shares outstanding, LPR is the logarithm of share price and LVO is the logarithm of average daily volume. Z-statistics are asymptotically standard normal under the null hypothesis that the corresponding coefficient is zero. All estimated parameters including estimated kappa and sigma are multiple by $\times 10^{-1}$. The explanatory variables are measured at the time of order submission.

| **Variable** | **Coef. $\times 10^{-1}$** | **Std. Err.** | **z** | **P>z** | **[95% Conf.** | **Interval]** |
|---|---|---|---|---|---|---|
| | | | | | | |
| **Buy Orders** | | | | | | |
| **MQLP** | -3.08 | 0.01 | -35.22 | 0.00 | -0.33 | -0.29 |
| **BSID** | 4.70 | 0.00 | 121.39 | 0.00 | 0.46 | 0.48 |
| **MKD1** | 1.05 | 0.00 | 147.50 | 0.00 | 0.10 | 0.11 |
| **MKD1X** | -0.03 | 0.00 | -91.55 | 0.00 | 0.00 | 0.00 |
| **MKD2** | -3.53 | 0.00 | -141.58 | 0.00 | -0.36 | -0.35 |
| **SZSD** | 0.27 | 0.00 | 50.54 | 0.00 | 0.03 | 0.03 |
| **STKV** | -11.74 | 0.02 | -48.81 | 0.00 | -1.22 | -1.13 |
| **TURN** | -7.50 | 0.01 | -127.82 | 0.00 | -0.76 | -0.74 |
| **LSO** | 6.09 | 0.01 | 81.14 | 0.00 | 0.59 | 0.62 |
| **LPR** | -6.59 | 0.01 | -77.88 | 0.00 | -0.68 | -0.64 |
| **LVO** | -5.28 | 0.01 | -62.55 | 0.00 | -0.54 | -0.51 |
| **_cons** | 136.23 | 0.11 | 124.28 | 0.00 | 13.41 | 13.84 |
| | | | | | | |
| **/ln_sig** | 8.17 | 0.00 | 437.55 | 0.00 | 0.81 | 0.82 |
| **/kappa** | 0.00[1] | 0.01 | -0.01 | 0.99 | -0.01 | 0.01 |
| | | | | | | |
| **sigma** | 22.64 | 0.00 | | | 2.26 | 2.27 |

**Table 5.3 continued …**

| Sell Orders | | | | | | |
|---|---|---|---|---|---|---|
| **MQLP** | 2.24 | 0.01 | 23.67 | 0.00 | 0.21 | 0.24 |
| **BSID** | -5.07 | 0.00 | -130.22 | 0.00 | -0.51 | -0.50 |
| **MKD1** | 0.50 | 0.00 | 62.16 | 0.00 | 0.05 | 0.05 |
| **MKD1X** | 0.02 | 0.00 | 60.48 | 0.00 | 0.00 | 0.00 |
| **MKD2** | -2.27 | 0.00 | -102.06 | 0.00 | -0.23 | -0.22 |
| **SZSD** | 0.40 | 0.00 | 69.89 | 0.00 | 0.04 | 0.04 |
| **STKV** | -12.41 | 0.02 | -51.38 | 0.00 | -1.29 | -1.19 |
| **TURN** | -8.27 | 0.01 | -139.81 | 0.00 | -0.84 | -0.82 |
| **LSO** | 3.96 | 0.01 | 54.77 | 0.00 | 0.38 | 0.41 |
| **LPR** | -3.62 | 0.01 | -44.39 | 0.00 | -0.38 | -0.35 |
| **LVO** | -2.87 | 0.01 | -34.21 | 0.00 | -0.30 | -0.27 |
| **_cons** | 119.56 | 0.11 | 109.01 | 0.00 | 11.74 | 12.17 |
| | | | | | | |
| **/ln_sig** | 7.97 | 0.00 | 409.56 | 0.00 | 0.79 | 0.80 |
| **/kappa** | 0.12 | 0.01 | 2.03 | 0.04 | 0.00 | 0.02 |
| | | | | | | |
| **sigma** | 22.19 | 0.00 | | | 2.21 | 2.23 |

[1] 0.0006

**Table 5.4: Parameter Estimates of the LMZ Models Excluding MKD1X**

The table below shows parameter estimates of the accelerated failure time specification of limit-order completion under the generalised gamma distribution for limit orders of a pooled sample of 38 stocks, throughout a sample period from October 2000 to December 2000. The variable '_cons' denotes the intercept and the definitions of the remaining explanatory variables are given as follows. MQLP is a measure of the distance between the limit-order price and current midpoint. BSID is an indicator of whether the prior trade was buyer-initiated or seller-initiated. MKD1 measures the number of shares that have higher priority for execution. MKD1X is an interactive term to capture non-linearity between market depth and market price relative to the limit-order price. MKD2 is a measure of liquidity available from the opposite side of the market. SZSD is a measure of liquidity demanded by the limit order. STKV is used to capture recent shifts in trading activity. TURN is an absolute measure of volatility. In the monthly updated variables, LSO is the logarithm of the number of shares outstanding, LPR is the logarithm of share price and LVO is the logarithm of average daily volume. Z-statistics are asymptotically standard normal under the null hypothesis that the corresponding coefficient is zero. All estimated parameters including estimated kappa and sigma are multiple by $\times 10^{-1}$. The explanatory variables are measured at the time of order submission and MKD1X is excluded.

| **Variable** | **Coef.** $\times 10^{-1}$ | **Std. Err.** | **z** | **P>z** | **[95% Conf.** | **Interval]** |
|---|---|---|---|---|---|---|
| | | | | | | |
| **Buyer Orders** | | | | | | |
| **MQLP** | -4.89 | 0.01 | -58.05 | 0.00 | -0.51 | -0.48 |
| **BSID** | 4.42 | 0.00 | 114.42 | 0.00 | 0.44 | 0.45 |
| **MKD1** | 0.78 | 0.00 | 125.48 | 0.00 | 0.08 | 0.08 |
| **MKD2** | -3.91 | 0.00 | -156.11 | 0.00 | -0.39 | -0.38 |
| **SZSD** | 0.38 | 0.00 | 65.14 | 0.00 | 0.03 | 0.04 |
| **STKV** | -11.73 | 0.02 | -48.16 | 0.00 | -1.21 | -1.12 |
| **TURN** | -7.47 | 0.01 | -126.53 | 0.00 | -0.76 | -0.74 |
| **LSO** | 6.53 | 0.01 | 86.78 | 0.00 | 0.64 | 0.67 |
| **LPR** | -6.77 | 0.01 | -79.58 | 0.00 | -0.69 | -0.66 |
| **LVO** | -5.46 | 0.01 | -64.59 | 0.00 | -0.56 | -0.53 |
| **_cons** | 138.12 | 0.11 | 125.36 | 0.00 | 13.60 | 14.03 |
| | | | | | | |
| **/ln_sig** | 8.09 | 0.00 | 425.65 | 0.00 | 0.80 | 0.81 |
| **/kappa** | 0.72 | 0.01 | 12.09 | 0.00 | 0.06 | 0.08 |
| | | | | | | |
| **sigma** | 22.38 | 0.00 | | | 2.24 | 2.25 |

**Table 5.4 continued …**

| | | | | | | |
|---|---|---|---|---|---|---|
| **Sell Orders** | | | | | | |
| **MQLP** | 3.87 | 0.01 | 42.48 | 0.00 | 0.37 | 0.41 |
| **BSID** | -4.78 | 0.00 | -120.48 | 0.00 | -0.49 | -0.47 |
| **MKD1** | 0.29 | 0.00 | 44.68 | 0.00 | 0.03 | 0.03 |
| **MKD2** | -2.71 | 0.00 | -125.56 | 0.00 | -0.27 | -0.27 |
| **SZSD** | 0.48 | 0.00 | 85.92 | 0.00 | 0.05 | 0.05 |
| **STKV** | -12.22 | 0.02 | -48.99 | 0.00 | -1.27 | -1.17 |
| **TURN** | -8.11 | 0.01 | -133.39 | 0.00 | -0.82 | -0.80 |
| **LSO** | 4.44 | 0.01 | 58.75 | 0.00 | 0.42 | 0.45 |
| **LPR** | -3.63 | 0.01 | -43.21 | 0.00 | -0.38 | -0.35 |
| **LVO** | -3.01 | 0.01 | -35.26 | 0.00 | -0.32 | -0.29 |
| **_cons** | 121.09 | 0.11 | 108.19 | 0.00 | 11.97 | 12.41 |
| | | | | | | |
| **/ln_sig** | 8.21 | 0.00 | 410.31 | 0.00 | 0.81 | 0.82 |
| **/kappa** | 0.59 | 0.01 | 10.43 | 0.00 | 0.05 | 0.07 |
| | | | | | | |
| **sigma** | 22.61 | 0.00 | | | 2.25 | 2.27 |

**Table 5.5: Parameter Estimates of the LMZ Models Excluding SZSD**

The table below shows parameter estimates of the accelerated failure time specification of limit-order completion under the generalised gamma distribution for limit orders of a pooled sample of 38 stocks, throughout a sample period from October 2000 to December 2000. The variable '_cons' denotes the intercept and the definitions of the remaining explanatory variables are given as follows. MQLP is a measure of the distance between the limit-order price and current midpoint. BSID is an indicator of whether the prior trade was buyer-initiated or seller-initiated. MKD1 measures the number of shares that have higher priority for execution. MKD1X is an interactive term to capture non-linearity between market depth and market price relative to the limit-order price. MKD2 is a measure of liquidity available from the opposite side of the market. SZSD is a measure of liquidity demanded by the limit order. STKV is used to capture recent shifts in trading activity. TURN is an absolute measure of volatility. In the monthly updated variables, LSO is the logarithm of the number of shares outstanding, LPR is the logarithm of share price and LVO is the logarithm of average daily volume. Z-statistics are asymptotically standard normal under the null hypothesis that the corresponding coefficient is zero. All estimated parameters including estimated kappa and sigma are multiple by $\times 10^{-1}$. The explanatory variables are measured at the time of order submission and SZSD is excluded.

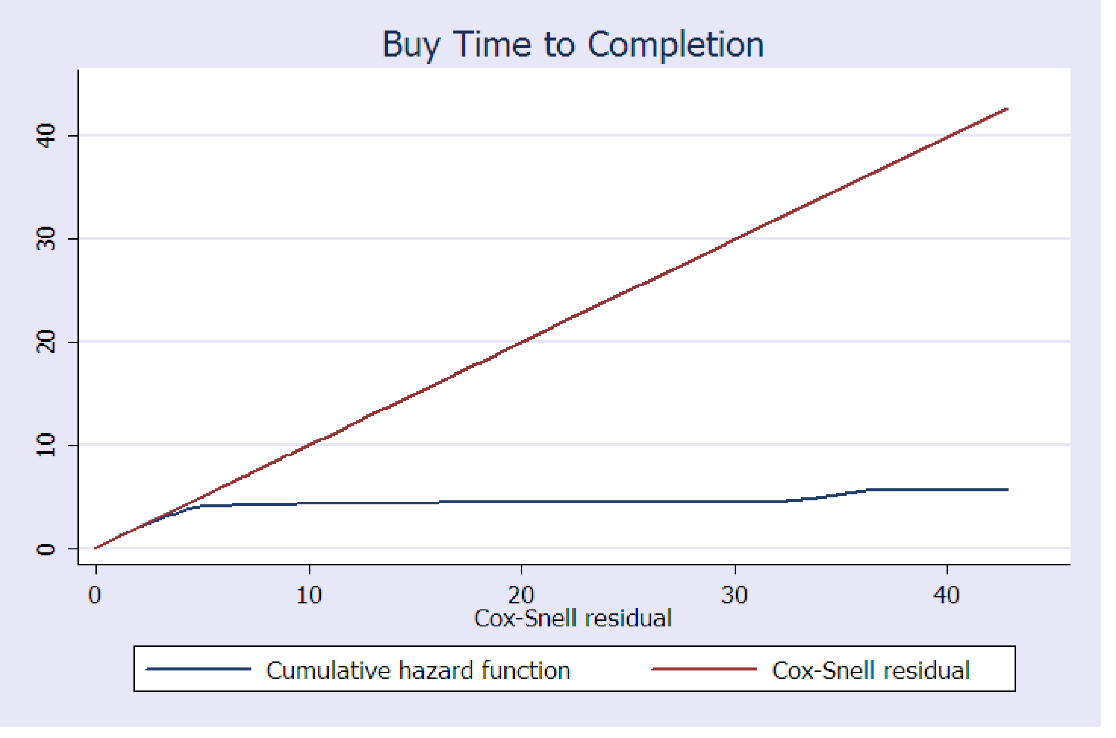
| Variable | Coef. $\times 10^{-1}$ | Std. Err. | z | P>z | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| | | | | | | |
| **Buy Orders** | | | | | | |
| **MQLP** | 1.01 | 0.00 | 27.90 | 0.00 | 0.09 | 0.11 |
| **BSID** | 4.70 | 0.00 | 120.45 | 0.00 | 0.46 | 0.48 |
| **MKD1** | 0.92 | 0.00 | 139.05 | 0.00 | 0.09 | 0.09 |
| **MKD1X** | -0.03 | 0.00 | -101.26 | 0.00 | 0.00 | 0.00 |
| **MKD2** | -3.68 | 0.00 | -147.68 | 0.00 | -0.37 | -0.36 |
| **STKV** | -11.38 | 0.02 | -46.99 | 0.00 | -1.19 | -1.09 |
| **TURN** | -7.31 | 0.01 | -123.97 | 0.00 | -0.74 | -0.72 |
| **LSO** | 6.60 | 0.01 | 87.94 | 0.00 | 0.65 | 0.67 |
| **LPR** | -6.88 | 0.01 | -80.90 | 0.00 | -0.70 | -0.67 |
| **LVO** | -5.59 | 0.01 | -65.86 | 0.00 | -0.58 | -0.54 |
| **_cons** | 139.03 | 0.11 | 125.88 | 0.00 | 13.69 | 14.12 |
| | | | | | | |
| **/ln_sig** | 8.34 | 0.00 | 461.75 | 0.00 | 0.83 | 0.84 |
| **/kappa** | -0.53 | 0.01 | -9.05 | 0.00 | -0.06 | -0.04 |
| | | | | | | |
| **sigma** | 23.03 | 0.00 | | | 2.29 | 2.31 |
| | | | | | | |

**Table 5.5 continued …**

| Sell Orders | | | | | | |
|---|---|---|---|---|---|---|
| **MQLP** | -4.05 | 0.00 | -109.58 | 0.00 | -0.41 | -0.40 |
| **BSID** | -5.08 | 0.00 | -126.23 | 0.00 | -0.52 | -0.50 |
| **MKD1** | 0.45 | 0.00 | 54.76 | 0.00 | 0.04 | 0.05 |
| **MKD1X** | 0.03 | 0.00 | 86.86 | 0.00 | 0.00 | 0.00 |
| **MKD2** | -1.90 | 0.00 | -83.46 | 0.00 | -0.19 | -0.19 |
| **STKV** | -11.72 | 0.03 | -46.86 | 0.00 | -1.22 | -1.12 |
| **TURN** | -7.96 | 0.01 | -130.37 | 0.00 | -0.81 | -0.78 |
| **LSO** | 4.10 | 0.01 | 54.63 | 0.00 | 0.40 | 0.42 |
| **LPR** | -3.53 | 0.01 | -41.75 | 0.00 | -0.37 | -0.34 |
| **LVO** | -2.97 | 0.01 | -34.09 | 0.00 | -0.31 | -0.28 |
| **_cons** | 118.23 | 0.11 | 103.92 | 0.00 | 11.60 | 12.05 |
| | | | | | | |
| **/ln_sig** | 8.46 | 0.00 | 448.69 | 0.00 | 0.84 | 0.85 |
| **/kappa** | -0.46 | 0.01 | -7.65 | 0.00 | -0.06 | -0.03 |
| | | | | | | |
| **sigma** | 23.29 | 0.00 | | | 2.32 | 2.34 |

**Figure 5.1: Q-Q Plots of the LMZ Models**

The following graph displays Q-Q plots of empirical estimates of the cumulative hazard functions of Cox-Snell residuals for limit orders of a pooled sample of 38 stocks, throughout a period from October 2000 to December 2000.



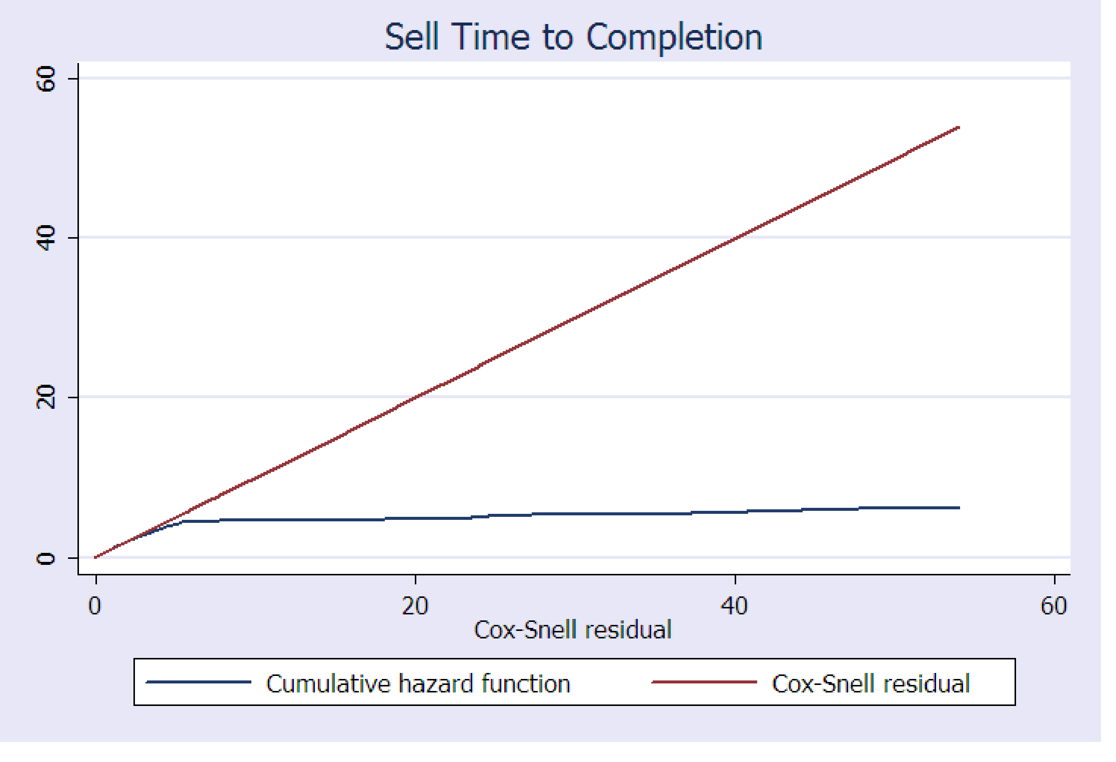
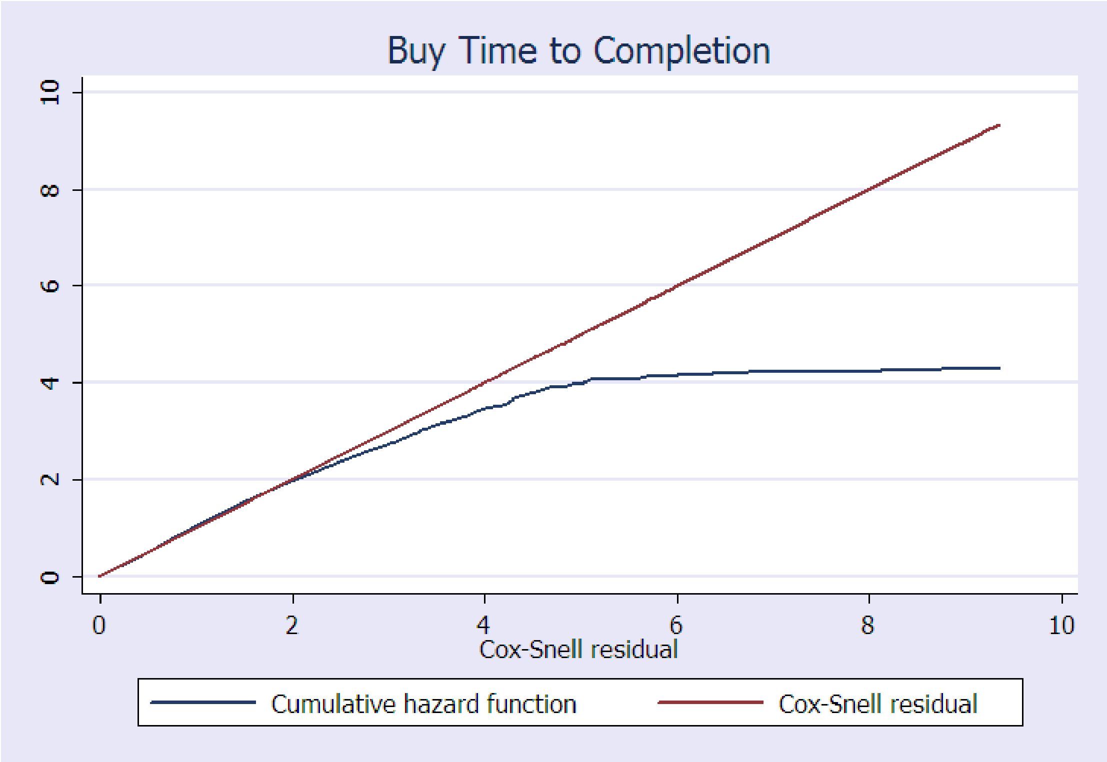
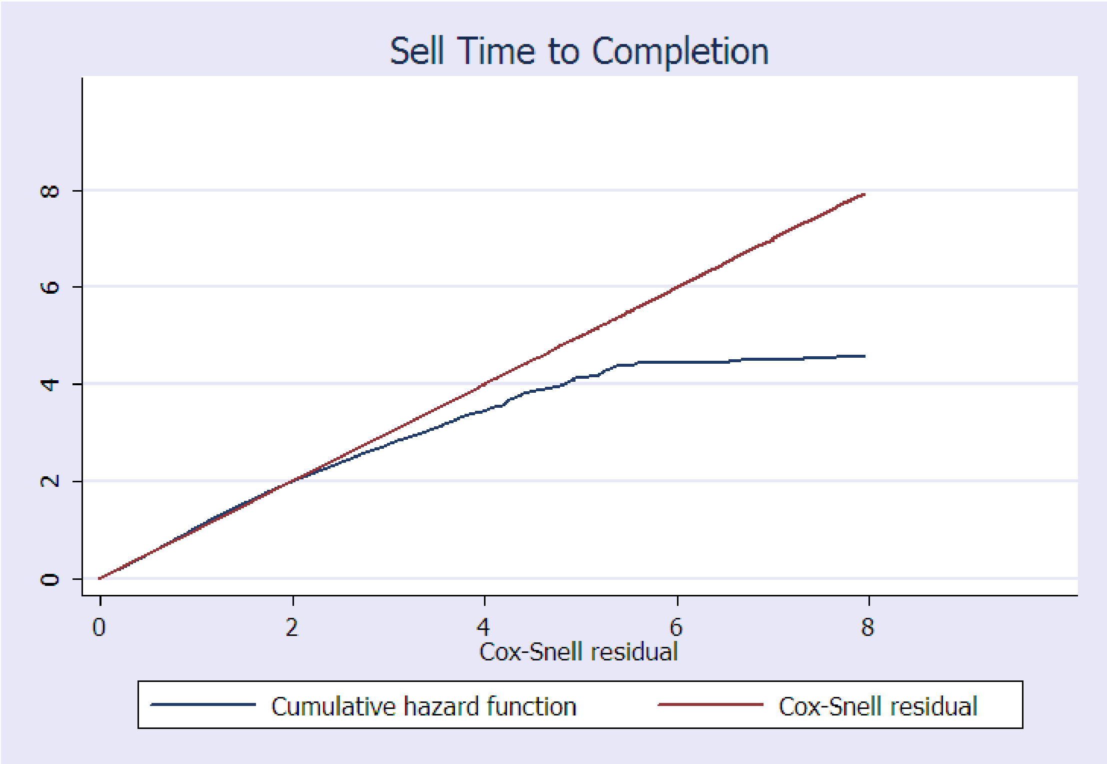



**Figure 5.2: Q-Q Plots of the LMZ Models Excluding Observations with Cox-Snell Residuals Greater Than 10**

The following graph displays Q-Q plots of empirical estimates of the cumulative hazard functions of Cox-Snell residuals for limit orders of a pooled sample of 38 stocks, throughout a period from October 2000 to December 2000, after excluding these observations with Cox-Snell residuals greater than 10.



# CHAPTER 6 – LIMIT-ORDER COMPLETION TIME IN THE UK MARKET: AN INVESTIGATION USING NEWLY CONSTRUCTED VARIABLES

## 6.1 Introduction

In Chapter 5, the empirical results show the LMZ models suffer from the effects of multicollinearity even with the large dataset used here. Thus in pursuing the objects of properly modelling order completion time in the UK market this chapter presents a set of newly developed survival models that use explanatory variables constructed to solve the multicollinearity problem inherent in the LMZ variables. These models are developed by the author to investigate the limit-order completion time in the UK market. Also unlike LMZ who model both time-to-first-fill and time-to-completion, in this chapter this study models time-to-completion only, which is of primary interest to traders and exchanges as discussed in Chapter 1, and provides empirical results and goodness-of-fit tests.

Traders use buy and sell orders with varying degrees of assertion. Keim and Madhavan (1995) provide evidence showing that typically traders are more passive when submitting buy orders, and more aggressive when submitting sell orders, perhaps due to greater urgency when selling. In order to capture possible asymmetries in the effects of the explanatory variables on the completion time of buy and sell orders, in this chapter separate models of limit-order completion time are constructed for buy and sell orders.

The remainder of this chapter is organised as follows. Section 6.2 provides a presentation of the constructed explanatory variables and offers a discussion on their construction. In Section 6.3 both the Generalised Gamma (GG) AFT and the Cox PH models, which are the most frequently used models in survival analysis, are used to incorporate the explanatory variables.[108] This section presents empirical results and

[108] The Generalised Gamma AFT model is henceforth referred to as the GG AFT model in the remainder of this thesis.

goodness-of-fit tests of these models. Section 6.4 offers a summary and some conclusions, and provides an introduction to the work in subsequent chapters.

## 6.2 Constructed Variables

The existing literature on limit-order execution as discussed in Section 3.4 shows that the characteristics of limit orders such as limit-order price and size, and the state of an order book such as the best bid-offer spread and volatility, affect the completion time of a limit order and its probability of completion (c.f., Al-Suhaibani and Kryzanowsky (2000), Cho and Nelling (2000), Omura, Tanigawa, and Jun (2000) and LMZ (2002)). Basing on the literature, in this section the explanatory variables that capture the characteristics of limit orders and order books are constructed.[109]

### *6.2.1 Explanatory Variables*

The explanatory variables are designed here to capture the characteristics of limit orders and order book conditions while simultaneously have no higher than 40% correlation across each other. The 40% maximum is chosen as the cut-off point, since under this cut-off point multicollinearity does not pose a problem in survival analysis as the literature reviewed in Subsection 5.3.2 suggested.

Let $P_l$ denote the limit-order price, $S_l$ the limit-order size, $P_b$ the best bid price, $S_b$ the best bid size, $P_o$ the best offer price, $S_o$ the best offer size, $P_q$ the mid-quote price and $P$ the market price of the most recent transaction.

The following are the explanatory variables for time-to-completion of buy limit orders (all variables are still measured at the submission time of a limit order).

[109] Given the computing facilities and technical support available, only some of initially proposed explanatory variables are selected. For example, a variable is proposed to capture the number of shares on the order book that has priority for execution over a submitted limit order. However it is impractical to extract this variable from the data.

PA is a measure of the relative distance between the limit-order price and the contemporary best bid price on the order book.[110] This variable is given by of Equation (6.1) below:

$$PA = \frac{P_b - P_l}{P_q} \times 100. \quad (6.1)$$

Since orders are 'queuing' and waiting to be executed with orders from the opposite side of the market, the length of the 'queue' is a possible variable that could affect the limit-order completion time. This variable captures the length of this 'queue' on the order book. In order to avoid a high correlation with variable BBOS in Equation (6.2), which captures the best bid-offer spread, PA measures only the distance between the limit-order price and the best bid price.[111]

PA captures the position and the price aggressiveness of a limit order in the current market condition. Demsetz (1968) suggests that aggressively priced limit orders should have shorter expected execution times. This idea has been empirically investigated by Potters and Bouchaud (2003). They use US data to investigate the typical lifetime of a limit order as a function of the distance from the best bid/offer price. They show that the lifetime of a given buy (sell) order increases as one moves away from the best bid (offer) price. Verhoeven, Ching and Ng (2004) find that execution probabilities of limit orders at the Australian Stock Exchange decrease gradually as limit orders are placed further away from the best bid/offer prices. Al-Suhaibani and Kryzanowsky (2000) find that in the Saudi stock market aggressively priced limit orders, on average, have shorter expected execution times and higher execution probabilities. In contrast to Al-Suhaibani and Kryzanowsky (2000), Cho and Nelling (2000) find the opposite in the US market. They find that the further away limit orders are placed from the best bid/offer prices, the higher the execution probability and the shorter the execution time.

PA is a relative scale measure which captures the position of a limit order in the current market condition in a better way than does an absolute scale measure. For example, as Figure 2.8 of Chapter 2 illustrates, the best bid price is 524 pence and the best offer

[110] PA refers to Price Aggressiveness.
[111] The distance between the limit-order price and the mid-quote price is the sum of the distance between the limit-order price and the best bid price and half of the best bid-offer spread. If PA captures the distance between the limit-order price and the mid-quote price, it is inevitable to be highly correlated with BBOS since PA has already captured half of BBOS.

price is 525 pence. Hence the mid-quote price is 524.5 pence for stock ABC. Assuming for stock DEF the mid-quote price is 52.45 pence. A limit order A for stock ABC is priced at 519.25 pence, which is 5.25 pence away from the mid-quote price of 524.5 pence. A limit order B for stock DEF is priced at 47.2 pence, which is also 5.25 pence away from the mid-quote price of 52.45 pence. Although both orders are 5.25 pence away from the mid-quote price, limit order A is far more aggressive and closer to the mid-quote price (about 1% away from the mid-quote price) than limit order B (about 10% away from the mid-quote price). For this reason a relative scale measure will better differentiate the aggressiveness of limit orders than an absolute scale measure.

PA captures the price aggressiveness of a limit order and the position of this order in the current market condition. Large values of PA indicate that traders are 'patient' and price their orders far below the best bid price. This indicates that these traders are waiting for better opportunities to execute their orders. In the above example, if a trader submits a buy market order, this order will be executed at the best offer price (525 pence) assuming that liquidity is available. If he submits a limit order priced at 519.25 pence, which is 5.25 (about 1%) pence away from the mid-quote price of 524.5 pence, and later this order is executed as the market moves in favour of him this trader then gains about 1% of execution price (compared to the market order execution price of 525 pence). Large values of PA also indicate that this limit order does not improve on the current best bid price. For example, as Figure 2.8 of Chapter 2 illustrates, the best bid price is 524 pence and the best offer price is 525 pence. Hence, the mid-quote price is 524.5 pence. For a limit order, a value of 10% for PA means that the limit-order price is 472.05 pence, which is far below the best bid price (524 pence) meaning this limit order does not improve on the current best bid price (524 pence) and does not have any priority for execution over the current best bid order. Therefore large values of PA indicate low execution probability and long execution time, because orders are executed according to the price priority discussed in Chapter 2.

BBOS represents the relative best bid-offer spread.[112] This variable is given by of Equation (6.2) below:

[112] BBOS refers to the Best Bid-Offer Spread. The best bid-offer spread is also known as the 'touch' in the literature.

$$BBOS = \frac{P_o - P_b}{P_q} \times 100. \quad (6.2)$$

Since narrow spreads induce market order submission as Foucault, Kadan and Kandel (2003) show and unexecuted orders on the order book will be executed with market orders from the opposite side of the market, the wideness of the spread is a possible variable that could affect the limit-order completion time.

Chung *et al*. (1997) find that in the NYSE the intra-day variation in the best bid-offer spreads is significantly related to the intra-day variation in limit-order placement and execution. Cho and Nelling (2000) and Elull *et al.* (2003) find that in the US stock market execution probabilities of limit orders depend on the best bid-offer spreads. Omura, Tanigawa, and Jun (2000) provide similar evidence for the Japanese stock market. Al-Suhaibani and Kryzanowsky (2000) find that in the Saudi stock market orders placed when the spread is wide are expected to have longer execution times.

BBOS is a relative scale measure and therefore captures the spread in a better way than would an absolute scale measure. For instance, a 10 pence (0.1%) bid-offer spread around a mid-quote price of 1000 pence is significantly different from a 10 pence (1%) bid-offer spread around a mid-quote price of 100 pence on a relative scale.

As discussed in Chapter 5, a wider spread increases the trading cost and hence discourages trading. Thus when the spread is wide, traders intend to submit limit orders rather than market orders. Since unexecuted limit orders on the order book will be executed with market orders (or marketable limit orders) from the opposite side, the fewer market orders are available from the opposite side of the order book, the less easily limit orders will be executed. Hence, larger values of BBOS indicate lower execution probability and longer execution time.

The bid-offer spread is one of the early research topics in market microstructure.[113] Cohen *et al*. (1981) investigate the bid-offer spread in a quote-driven market and argue that it would be driven to zero by the law of one price in a perfect market with full information in which by definition supply and demand for stocks facing each market maker are infinitely price elastic. Levin and Wright (2004) argue that the presence of positive bid-offer spreads necessarily means that the demand and supply for stock facing market makers cannot be infinitely price elastic. Instead of analysing the bid-offer spread, this study will investigate the effects of the bid-offer spread on limit-order completion time.

BSIDS is an indicator variable, which is a simplified version of BSID in Equation (5.2) and retains the underlying economic interpretation of BSID. This variable is given by of Equation (6.3) below:

$$BSIDS = \begin{cases} 1 & \text{if } P > P_q \\ 0 & \text{otherwise} \end{cases}. \tag{6.3}$$

BSIDS indicates whether or not the prior trade was buy-side initiated (BSIDS equals to one). This variable also captures order imbalance on the order book. If we assume that the order book is balanced (the same number and size of limit orders on both side of the book), then a trade initiated by buyers would result in a temporary buy order imbalance on the order book (more buy orders than sell orders). Assuming that no sell orders are submitted at this point. As a result, buy limit orders will compete with each other for limited liquidity.

Handa, Schwartz and Tiwari (1998) argue that when the buy order imbalance is large, there is high non-execution risk for buy orders. Accordingly buy orders are expected to

[113] The bid-offer spread has been decomposed into four components. The first component is the order processing cost as Demsetz (1968) suggests. This includes the administrative costs incurred by market makers who prepare to trade throughout the trading day. The second component is adverse selection cost, which compensates market makers for the losses they suffer when trading with well-informed traders. It allows them to recoup from uninformed traders. The third component is the inventory-holding cost as Biais (1993) suggests. This is the cost for market makers to hold unbalanced portfolios. The fourth component is the market maker rent or economic profit as Huang and Stoll (1996) suggests. With respect to this component, it is important to make a clear distinction between ‘normal’ and ‘economic’ profit. Normal profit is a necessary cost for the smooth functioning of competitive markets because it is by definition the minimum profit required in order to hold resources in the industry. In contrast, economic profit is profit in excess of normal profit and is not regarded as a cost because it is not required to maintain resources in the industry. Levin and Wright (2004) find that the existence of economic profit indicates the presence of imperfect competition and signals the need for regulatory attention.

have longer execution times as Handa & Schwartz (1996) argue. Al-Suhaibani and Kryzanowsky (2000) provide empirical evidence to support this prediction. Hence the side that initiated the prior trade is a possible variable that could affect the limit-order completion time.

BSIDS also captures some information about the state of a limit-order book. For instance, for buy orders, a positive value of BSIDS indicates that the prior trade has occurred on the sell side of the order book. Thus the number of available sell orders on the order book has decreased assuming that no sell orders are submitted at this point. When buyers notice the decreased liquidity on the sell side, they will price their orders more aggressively to compete for the liquidity. Hence more aggressively priced buy orders will enter the order book. As a result, the unexecuted buy orders on the order book will be 'pushed to the back of the queue' and are expected to have lower execution probabilities and longer execution times.

NL30MT, like TURN in Equation (5.8), is a trading activity measure that provides an absolute measure of volatility.[114] This variable is given by of Equation (6.4) below:

$$NL30MT = \log(1 + \text{number of trades during last 30 minutes}) . \qquad (6.4)$$

As discussed in Chapter 5, higher volatility will generally increase the probability of limit-order completion and hence reduce the limit-order completion time since a limit order placed away from the current market price will be executed when the limit-order price is first met. As a result, the more active and volatile the market becomes, the higher the order completion probability and the shorter the expected completion time. Hence the volatility of the market is a possible variable that could affect the limit-order completion time.

Cho and Nelling (2000) find that in the US market execution probabilities of limit orders depend on market volatility. LMZ (2002) also find that in the US market limit-order execution times depend on market volatility.

[114] NL30MT refers to the Number of Last 30 Minutes Trades. Adding one to the number of trades during the last 30 minutes can avoid invalid NL30MT when no trade occurs in the last 30 minutes. To some extent volatility is a proxy of the rate of arrival of information, as Easley, Kiefer, and O'Hara (1997) argue that low trading activity is related to the lack of valuation-relevant information.

NL30MT captures the interaction between STKV in Equation (5.7) and TURN in Equation (5.8) as

$$e^{NL30MT} = STKV \times e^{TURN} + 1 \, . \tag{6.5}$$

This variable captures the number of trades, rather than transaction size, in the preceding 30 minutes. It is high if the volatility in the market in the preceding 30 minutes is high. The more active a market becomes the higher the limit-order execution probability and the shorter the expected completion time.

LOS captures the limit-order size.[115] This variable is given by of Equation (6.6) below:

$$LOS = \log(S_l) \, . \tag{6.6}$$

Intuitively it is more difficult and takes a longer period of time to complete a large order. The larger a limit order, the lower the order completion probability and the longer the expected completion time. Hence the limit-order size is a possible variable that could affect the limit-order completion time.

Cho and Nelling (2000) find that in the US market execution probabilities of limit orders depend on the limit-order sizes. Also LMZ (2002) find that in the US market the limit-order size affects limit-order completion time.

One can argue that limit-order completion time for liquid stocks is less sensitive to this variable than that for illiquid stocks, as the more liquid a stock, the more speedily a limit order could be completed.[116] Since the stocks in the data sample used in this study are liquid FTSE 100 stocks, it is expected that limit-order completion time will be less sensitive to this variable. This will be discussed later in the next section.

LOS also captures the liquidity demanded by a limit order, as does SZSD in Equation (5.6). It is a simplified version of SZSD and designed to avoid high correlations among variables.

---

[115] LOS refers to the Logarithm of Order Size.

[116] Traders tend to break down a large limit order to small orders and large block of trades tend to be completed through a brokered market as discussed in Chapter 2. Hence, extremely large orders are rarely seen in a pure order-driven market such as SETS.

LS captures the depth (liquidity) of the sell side of the order book.[117] This variable is given by of Equation (6.7) below:

$$LS = \log(S_o)\,. \tag{6.7}$$

LS is a simplified version of MKD1 in Equation (5.5) and designed to avoid high correlations among variables. LS only captures the best offer size rather than the full depth of the sell side of the order book.[118]

The depth of the order book could be a proxy of liquidity on the order book as discussed in Chapter 5. The more liquidity is available from the opposite side of the market, the more easily a limit order will be completed. The depth of the order book could also be a proxy of traders' willingness to trade as discussed in Chapter 5. The more traders are willing to trade from the opposite side of the market, the more easily a limit order will be completed. Hence the depth of the order book is a possible variable that could affect the limit-order completion time.

Handa, Schwartz and Tiwari (2000) predict that the depth on the buy (sell) side of the book is a straightforward proxy of the proportion of buyers (sellers). They argue that a thicker depth on the sell (buy) side indicates that there are many sellers (buyers) competing to sell (buy). Omura, Tanigawa, and Jun (2000) find that execution probabilities of limit orders in the Japanese stock market depend on the depth of the opposite side of the book. Ranaldo (2003) finds that in the Swiss stock market the depth on the opposite side of an incoming order decreases the probability of that order being executed. Hence the more liquidity is available on the opposite side of the market, the higher the limit-order execution probability and the shorter the expected completion time.

[117] LS refers to the Logarithm of Size on the opposite side of the market. Initially a variable LSS capturing the depth on the same side of the order book was also constructed but was found to be highly correlated with the variable LOS. Since capturing the limit-order size is essential, LOS is selected and LSS is dropped. There is more detail in Table D.1 of Appendix D.

[118] Given the computing facilities and technical support available at the time of study, it is impractical to extract a variable that captures the full depth of the order book.

The following are redefined variables for the sell model, which retain the underlying economic interpretation:

$$PA = \frac{P_l - P_o}{P_q} \times 100\,. \tag{6.8}$$

$$BSIDS = \begin{cases} 1 & \text{if } P < P_q \\ 0 & \text{otherwise} \end{cases}. \tag{6.9}$$

$$LS = \log(S_b)\,. \tag{6.10}$$

### *6.2.2 Summary Statistics of Explanatory Variables*

Table 6.1 reports summary statistics of these explanatory variables. On average, there are considerable variations in the explanatory variables, but no significant differences between buy and sell orders. These variables vary across stocks as Table D.3 to Table D.14 in Appendix D present. Hence these variables capture cross-sectional difference across stocks. For this reason, this study does not use any additional variables to capture the cross-sectional difference.[119] Table 6.2 reports that the cross correlations of the explanatory variables are relatively low and under the 40% cut-off point.

The mean of PA indicates that the limit-order price of buy orders is about 0.15% below the best bid price and that of sell orders is 0.14% above the best offer price. This indicates that traders price their orders equally in terms of the distance from the best bid and offer prices. Hence the sample period covers a market that is balanced (i.e., neither a 'buy' market nor a 'sell' market) in terms of this distance.

The 5th and 50th percentiles of PA are zero for both buy and sell orders. These indicate that at least 50% of traders price their orders at the best bid/offer prices and are competing aggressively for execution. These also indicate that these traders do not improve on the best bid/offer prices and price their orders carefully, rather than randomly, at the exact best bid/offer prices on the order book they observe. The cluttering of orders at the best bid/offer prices could be due to the fact suggested by

[119] LMZ use LSO, LPR and LVO to capture the cross-sectional difference as discussed in Chapter 5.

Harris and Hasbrouck (1996) that limit orders placed at the prevailing best bid/offer prices perform better than market orders. Hence, the dominant order pricing strategy in this market during the sample period is to price buy orders at the best bid price and sell orders at the best offer price. This finding is consistent with Al-Suhaibani and Kryzanowsky (2000) who find that the limit orders in the Saudi stock market are frequently placed at the prevailing best bid/offer prices. This finding is also supportive of the imitation hypothesis (different traders imitate each other) proposed by Biais, Hillion and Spatt (1995).

Skewness of PA is positive for both buy and sell orders. Accordingly the right tail of the distribution is longer and the mass is concentrated on the left side of the mean. High kurtosis of PA for both buy and sell orders indicates that more of the variance of this variable is due to infrequent extreme deviations. For example, the 95th percentile of PA is 1.02% for buy orders and 0.88% for sell orders. This indicates that only a small proportion of buyers price their orders just over 1% below the best bid price and a small proportion of sellers price their orders just under 1% above the best offer price.

The above statistics of PA indicate that on average traders are ‘impatient’ and are willing to trade, since they price their orders at or close to the best bid/offer prices. As Table 6.2 presents, the highest correlation PA has with any other is 14% for buy orders and 11% for sell orders (between PA and LOS that captures the limit order size).

The mean of BBOS indicates that the relative best bid-offer spread, which is an important aspect of an order book, is about 0.33%. If the spread is too wide, traders will switch from market orders to aggressively priced limit order, which consequently narrows the spread. If the spread is too narrow, traders will change from limit orders to market orders, which consequently widens the spread, since market orders always executes with the best bid/offer orders. Hence the best bid-offer spread is dynamic and changing over time. This will be discussed in the next chapter.

The spread reported here is lower compared to those of other stock markets. For example, Angel (1997) reports an average relative spread of 0.65% for major market indices of 15 countries. Since the relative spread in the market studied in this thesis is relatively low, market liquidity, as measured by spread width, is relatively high. Thus traders can buy or sell at relatively low transaction costs. Levin and Wright (2004) report an average spread of 0.68% for FTSE 100 stocks of the first half of 1995.[120] It seems that since SETS was introduced in 1997 the spread has narrowed significantly. In this regard, one can argue that an order-driven market is superior to a quote-driven market. This finding supports previous research on cross-exchange comparisons, which show that trading costs in an order-driven market are lower than in a quote-driven market (c.f., De Jong, Nijman and Roell (1995) and Bessembinder and Kaufman (1997)). The narrower spread also indicates that the stocks in the sample used in this study are more liquid. Since the spread is the primary trading cost faced by traders, a narrower spread will encourage trading and hence shorten order execution time.

High kurtosis of BBOS for both buy and sell orders indicates that more of the variance of this variable is due to infrequent extreme deviations. For example, the 95th percentile of BBOS is 0.90% for buy orders and 0.89% for sell orders. Since the mean of PA for buy (sell) limit orders is about 0.15% (0.14%) below (above) the best bid (offer) price and the mean of BBOS is about 0.33%, then, on average buyers (sellers) in this market throughout the sample period price their orders just half of the best bid-offer spread below (above) the best bid (offer) price.

As Table 6.2 presents, the highest correlation BBOS has with any other is -26% for both buy and sell orders (between BBOS and NL30MT that captures the number of trades in the last 30 minutes). This negative correlation is expected, since the wider the spread the less likely trades are to occur.

The mean of BSIDS is 0.38 for buy orders and 0.40 for sell orders. This value and sign indicate that on average about 40% of trades are initiated by the same side of the order book. As Table 6.2 presents, the correlations between BSIDS and the remaining variables are quite low. This is expected, since BSIDS is a dummy variable that indicates whether it was a seller or a buyer who initiated the prior trade and most of the remaining variables capture market conditions at the time of order submission.

[120] In 1995, FTSE 100 stocks were still traded in a quote-driven market. These spreads were the spreads at the largest quote sizes.

The mean of NL30MT is 3.29 for buy orders and 3.28 for sell orders. This indicates that for the sample used in this study the average number of trades in the last 30 minutes prior to order submission is about 26 ($e^{3.285}-1$).

As Table 6.2 presents, the correlation between NL30MT and LOS, which captures the limit order size, is 28% for buy orders and 29% for sell orders. This medium correlation indicates that traders are likely to submit large orders when the market is active and volatile. This is expected, as the more active the market becomes the more easily orders could be executed. The correlation between NL30MT and LS, which captures the liquidity available from the opposite side of the market, is about 29% for both buy and sell orders. This is expected, as the more active the market becomes the more likely large orders will be submitted meaning the more liquidity is made available on the opposite side of the order book.[121]

The mean of LOS is 8.91 for buy orders and 8.90 for sell orders. This indicates that for the sample used in this study the average order size is about 7,369 ($e^{8.905}$) shares. This large order size is expected, as all stocks in the data sample used in this study are liquid FTSE 100 stocks.

As Table 6.2 presents, the highest correlation LOS has with any other is 31% for buy orders and 33% for sell orders (between LOS and LS that captures the liquidity available from the opposite side of the market). This medium correlation indicates that traders choose to submit large orders when the liquidity available on the opposite side of the order book is high. This is expected, as the more liquidity is available on the opposite side of the order book, the more likely large orders will be submitted to consume this liquidity.

The mean of LS is 9.52 for buy orders and 9.49 for sell orders. This indicates that for the sample used in this study the best bid/offer size is about 13,427 ($e^{9.505}$) shares, nearly double the average order size discussed above. Thus on average the quantities at the best bid/offer price equal to the size of about two incoming orders. This indicates

[121] The limit order submitted on the opposite side of the order book provides the liquidity.

that in this market, on average, for incoming orders the liquidity is available from the opposite side of the order book. Hence unavailable liquidity is not the reason behind some buyers (sellers) pricing their orders away from the best bid (offer) prices as discussed above. These traders are probably just waiting for better execution opportunities.

LS measures the depth of the order book. Since the depth is relative thick, market liquidity measured by depth is relatively high. This indicates that traders can buy or sell a large number of shares at relatively low costs. As Table 6.2 presents, the highest correlation LS has with any other is between LS and LOS as discussed above.

## 6.3 Model Specifications and Empirical Estimations

Two frequently used models for adjusting survival functions for the effects of explanatory variables are the Accelerated Failure Time (AFT) and Proportional Hazard (PH) models as discussed in Subsection 4.2.5 of Chapter 4. In this section, as LMZ (1999) suggested, both models are used to incorporate the explanatory variables discussed above.

### *6.3.1 Empirical Estimations of the GG AFT Model*

This study is one of a few applications of survival models in finance. In this subsection, the AFT model is used to incorporate the explanatory variables. Since the AFT model requires the specification of the baseline survival time distribution that is still unknown, this study follows the approach suggested by LMZ, as their study is the leading research in this field.

A GG AFT model suggested by LMZ and discussed in Section 5.3 is used to incorporate these explanatory variables, as the generalised gamma distribution is extremely flexible and allows for a large number of possible shapes that nest the exponential, Weibull and lognormal distributions as special cases.

Once the shape of the baseline survival time distribution is known, it is possible to identify the distribution of the completion time of a limit order. Then it is possible to know the probability of the limit-order completion, which is the foundation of developing dynamic order submission strategies.

Table 6.3 shows the results of estimating the GG AFT model. The estimates of the parameters associated with the conditioning variables generally are statistically significant for both models of buy and sell orders.[122] These coefficients reflect the sensitivities of limit-order completion time to these explanatory variables.

*The Coefficients of PA*

The estimate of the coefficient on PA is 5.52 with z-statistic of 190.54 for buy orders and 5.34 with z-statistic of 184.07 for sell orders. According to the AFT model discussed in Subsection 4.2.5 of Chapter 4, if a buy order is priced 1% of the mid-quote price below the best bid price (PA value increases by one unit holding the remaining variables constant), the expected completion time will be extended by 250 ($e^{5.52}$) times. Also if a sell order is priced 1% of the mid-quote price above the best offer price, the expected completion time will be extended by 209 ($e^{5.34}$) times. Hence, limit-order completion time is very sensitive to this variable, which captures the position of a limit order relative to the best bid/offer price. This is expected since orders are 'queuing up' on the order book for subsequent execution and PA captures the length of this 'queue'. In addition, completion time of buy orders is slightly more sensitive to this variable than that of sell orders. This could be due to the fact that this market seems to be a 'buy' market in which more buy limit orders are competing for execution as Table 4.1 of Chapter 4 presents. The high sensitivity to this variable indicates that traders can shorten expected limit-order completion times significantly by only improving the limit-order price slightly (closer to the best bid/offer prices). These estimates also quantify trade-offs between gains from execution prices and losses from execution times predicted by Cohen *et al.* (1981). These estimates may also explain the pricing strategies revealed above and show that traders are aware of the impact of the position of a limit order on its completion time and hence deliberately price their orders at the

[122] The standard error of a statistic depends on the sample size. In general, the larger the sample size, the smaller the standard error and the larger the z-statistics. In this study the coefficient estimates, hence, are expected to be significant due to the large size of the data sample.

best bid/offer price to reduce the expected limit-order completion time and the associated opportunity cost.

The positive sign of the coefficients of PA indicates that the lower (higher) the buy (sell) limit-order price below (above) the best bid (offer) price, the longer the expected completion time. This finding is consistent with the evidence Potters and Bouchaud (2003) and Al-Suhaibani and Kryzanowsky (2000) provide for non-UK stock markets. This positive sign could be a mechanical consequence of the price priority in an order-driven market discussed in Chapter 2. It could also be due to aggressively priced orders inducing latent demand for trade, since aggressively priced limit orders could encourage traders on the opposite side of the market to submit market orders rather than limit orders. For example, as Figure 2.8 of Chapter 2 illustrates, the best bid price is 524 pence and the best offer price is 525 pence. At 10:00 am, a limit order priced at 524.5 pence is submitted. Since the limit-order price is higher than the best bid price, it becomes the best bid price, which is the execution price for sell market orders. This order becomes the best bid order, which has the highest priority for execution. Accordingly the best bid-offer spread is narrowed to 0.5 pence from 1 penny. Assume at 10:00am a seller, who could access the best bid/offer price in real time, is also planning to sell and is thinking of choosing between market and limit orders. The narrowed spread and improved execution price for sellers (the improved best bid price) could induce this seller to submit sell market orders rather than limit orders. This aggressively priced buy limit order consequently rewards this seller for submitting sell market orders, as he pays a narrower spread and executes his order at a higher price. As a result, sellers could be more likely to submit market orders rather than limit orders and consequently this buy order would have a shorter expected completion time.

These estimates discussed above have the following implications. In order to minimise the expected waiting time of order completion, traders should price their orders as close to the best bid/offer price as possible. Exchanges should encourage traders to submit aggressively priced orders. Exchanges could reduce the minimum tick size to encourage traders to compete with each other.

*The Coefficients of BBOS*

The estimate of the coefficient on BBOS is 2.29 with z-statistic of 144.63 for buy orders and 2.35 with z-statistic of 143.43 for sell orders. These indicate that if the best bid-offer spread is widened by 1% of the mid-quote price (BBOS value increases by one unit holding the remaining variables constant) the expected completion time for buy orders will be extended by 9.87 ($e^{2.29}$) times and for sell orders will be extended by 10.48 ($e^{2.35}$) times. Thus limit-order completion time is very sensitive to the best bid-offer spread. The high sensitivity is expected since narrow spreads will encourage trading, as the best bid-offer spread is one of major trading costs faced by traders.

The positive sign of the coefficients of BBOS indicates that the wider the best bid-offer spread, the longer the expected completion time and the lower the expected completion probability. This finding is consistent with the evidences Harris and Hasbrouck (1996), Al-Suhaibani and Kryzanowsky (2000) and Elull *et al.* (2003) provide for none-UK stock markets. The positive sign could be due to narrow spreads inducing latent demand for trade, since narrow spreads encourage traders to submit market orders as the example above illustrates. Foucault, Kadan and Kandel (2003) provide empirical evidence showing that narrow spreads induce market order submission. The more market orders are available, the higher the expected limit-order execution probability and the shorter the expected limit-order execution time. Hence for exchanges narrowing the best bid-offer spread could facilitate limit-order execution.

These estimates discussed above have the following implications. Traders should submit limit orders when the market spread becomes narrower. Exchanges should minimise the market spread to facilitate the order execution. Exchanges should prefer a hybrid (order-driven/quote-driven) market to a pure order-driven market. In a hybrid (order-driven/quote-driven) market, as discussed in Chapter 2, dealers are obliged to compete with the order book for business. Hence increased competition should narrow the spread. In this case, the hybrid market may be superior to a pure order-driven market in terms of lowering transaction cost by narrowing the spread and consequently shortening limit-order completion time. Hence exchanges could switch from a pure order-driven market to a hybrid market to reduce the market spread and then the order completion time.

*The Coefficients of BSIDS*

The estimate of the coefficient on BSIDS is 0.80 with z-statistic of 103.28 for buy orders and 0.83 with z-statistic of 106.17 for sell orders. These indicate that if the prior trade was initiated from the same side of the order book (BSIDS value is one holding the remaining variables constant), the expected completion time for buy orders will be extended by 2.23 ($e^{0.80}$) times and by 2.29 ($e^{0.83}$) times for sell orders. Compared to the two variables just discussed limit-order completion time is not as sensitive to this variable. This low sensitivity could be due to the fact that BSIDS only approximates the side initiating the previous trade.[123]

The positive sign of the coefficients of BSIDS is as expected. For example, as Figure 2.8 of Chapter 2 illustrates, the best bid price is 524 pence and the best offer price is 525 pence. Hence the mid-quote price is 524.5 pence. If the prior trade price was 524.75 pence, then for a buy order the value of BSIDS is one (524.75>524.5). Thus it was a buyer who submits a market order and consequently initiates the prior trade. This could have reduced the number of shares available on the sell side of the order book and cause order imbalances, as Handa and Schwartz (1996) predict, assuming that no sell orders are submitted at this point. As a result, buy limit orders will compete with each other for limited liquidity and consequently have longer expected completion times. This finding is consistent with the evidence Al-Suhaibani and Kryzanowsky (2000) provide for the Saudi market. They find that if the same side of the market initiates most trades in the last 30 minutes, a longer time-to-execution is expected.

As discussed in Chapter 5, informed traders will use market orders to take advantage of their private information. Assume that buyers use market orders to initiate the previous transactions. This may indicate that buyers could have some private information suggesting that the share price is going to increase. Assume that more and more buyers have access to this information. As a result, more buyers could enter the market and submit more market orders. Then the market becomes a 'buy' market since buyers are more willing to trade. Since more buyers are submitting market orders and these market orders will be executed with sell limit orders, the expected time-to-completion for these sell orders will be shorter.

[123] Given the data sample used in this study, it is impossible to know whether buyers or sellers submit the previous market orders and hence initiate the previous trades.

These estimates discussed above indicate that traders should monitor activities on the opposite side of the market. Assume that these traders notice an inflow of buy market orders. This indicates that the price may increase. These traders should think carefully before rushing to sell any shares since their orders may be ‘picked up’ quickly by informed buyers. However this may present an opportunity, since it would be easier to execute sell limit orders and the waiting time could also be reduced.

*The Coefficients of NL30MT*

The estimate of the coefficient on NL30MT is -0.73 with z-statistic of -159.21 for buy orders and -0.70 with z-statistic of -148.76 for sell orders. These indicate that if NL30MT value decreases by one unit holding the remaining variables constant, the expected completion time for buy orders will be extended by 2.08 ($e^{0.73}$) times and for sell orders will be extended by 2.01 ($e^{0.70}$) times. Compared to PA and BBOS just discussed limit-order completion time is not as sensitive to this variable.

The negative sign of the coefficient of NL30MT indicates that the higher the number of trades in the last 30 minutes, the more active and volatile the market becomes and hence the shorter the expected completion time. This finding is consistent with the evidences Cho and Nelling (2000), Al-Suhaibani and Kryzanowsky (2000) and LMZ (2002) provide for none-UK stock markets.

As discussed in Chapter 5, higher volatility will generally increase the probability of limit-order completion and hence reduce the limit-order completion time (benefit) since a limit order placed away from the current market price will fill when the limit-order price is first met, but may also increase the probability of being ‘picked up’ by an informed counterpart from the opposite side of the market (cost). The potential cost of being ‘picked up’ could discourage the use of limit orders. In addition, if the market moves further away from a submitted limit order due to the higher volatility, traders may have to switch from limit orders to market orders to chase the price. Hence the completion probability of existing limit orders will be improved since these orders will be executed with market orders from the opposite side of the market. From what have been discussed, the negative signs on these coefficients of NL30MT are expected. As discussed in Chapter 5, although higher volatility will generally reduce the expected waiting time of limit-order completion, traders should be always cautious and careful when submitting a limit order in a volatile market since the cost could exceed the benefit.

As discussed in Chapter 5, the probability and time of limit-order completion depend on the overall volatility and the cost of being ‘picked up’ depends on the ‘information-driven’ volatility since the probability of being ‘picked up’ depends on the arrival of informed traders from the opposite side of the market. Hence exchanges should encourage the ‘mechanic’ volatility in order to facilitate limit-order execution, but be as transparent as possible to reduce the information-oriented volatility. For example, exchanges should publish the real-time order-book information to traders. In this case, volatility driven by the order-book information will be reduced since no one has private information about the state of the order book. Exchanges should also improve their trade systems to allow traders to submit and cancel their orders more quickly. In this case, the ‘mechanic’ volatility will increase.

As the estimates discussed above show, the higher volatility will reduce limit-order completion time and hence facilitate the limit-order execution. These estimates show that traders should submit limit orders in a ‘moderate’ volatile market, since the waiting time of limit-order completion could be reduced.

*The Coefficients of LOS*

The estimate of the coefficient on LOS is 0.17 with z-statistic of 66.86 for buy orders and 0.13 with z-statistic of 49.33 for sell orders. These indicate that if LOS value increases by one unit holding the remaining variables constant, the expected completion time will be extended by 1.19 ($e^{0.17}$) times for buy orders and by 1.14 ($e^{0.13}$) times for sell orders. Compared to PA and BBOS just discussed limit-order completion time is not as sensitive to this variable. This low sensitivity is consistent with the evidence Omura, Tanigawa and Jun (2000) provide for the Japanese stock market. They find that small limit orders do not necessarily have higher execution probabilities and longer execution times. This finding implies that traditionally breaking a large limit order into a number of small orders will not shorten order completion time significantly. This finding suggests that breaking a large order into a few of small orders may not benefit traders, since the cost of monitoring these small orders could be higher than the benefit of reducing completion time.[124]

The positive sign of the coefficient of LOS indicates that the larger a limit order, the longer the expected completion time. This is expected since the larger a limit order, the more difficult to fill this order. This finding is consistent with the evidence Cho and Nelling (2000) provide for the US stock market.

The positive sign revealed above could be due to the competition among traders. Assume that the submission of a large order indicates traders' willingness to trade. For example, a large buy order is submitted at 10:00am. This may indicate that buyers are more willing to trade and buy orders will enter the market. As a result, competition among buyers increases and the expected buy-order completion time will be longer.

*The Coefficients of LS*

The estimate of the coefficient on LS is -0.12 with z-statistic of 253.45 for buy orders and -0.10 with z-statistic of 249.98 for sell orders. These indicate that if the LS value decreases by one unit holding the remaining variables constant, the expected completion time for buy orders will be extended by 1.13 ($e^{0.12}$) times and for sell orders will be extended by 1.11 ($e^{0.10}$) times. Compared to PA and BBOS just discussed limit-order completion time is also not as sensitive to this variable.

The negative sign of the coefficient of LS indicates that the larger the best offer (bid) size, the shorter the expected buy-order (sell-order) completion time. This is expected, as the more liquidity is available from the opposite side of the market, the more easily a limit order will be completed. This finding is supportive of the prediction of Handa, Schwartz and Tiwari (2000) and consistent with the evidence Omura, Tanigawa and Jun (2000) provide for the Japanese stock market.

[124] Improving trading systems could reduce the cost of monitoring unexecuted orders.

As discussed in Chapter 5, the depth of the order book could be a proxy of traders' willingness to trade. For example, assume that at 11:00am the order book is empty. From 11:00am to 11:30am, 30 sell limit orders and five buy limit orders arrive at the order book. As a result, more liquidity is available on the sell side of the market. It can be argued that sellers are more willing to trade than buyers. It is possible that some sellers may become impatient, cancel their existing orders and submit market orders instead. Hence the expected completion times of existing buy limit orders will be shorter. This finding indicates that in order to reduce the waiting time traders should submit limit orders when more liquidity is available. If traders behave in this way, exchanges should disclose the full depth of the order book and restrict the use of 'hidden' orders, e.g. iceberg orders. This move could improve the limit-order execution.

As Table 4.3 shows, the average order completion time in the data sample used in this study is over nine minutes. For traders the opportunity cost of waiting for these nine minutes could be significant as discussed in Chapter 4. In order to reduce this opportunity cost, traders have to minimise the expected waiting time of limit-order completion.

The above empirical findings show that the limit-order completion time is quite sensitive to the variables of PA and BBOS. Hence in order to minimise the expected waiting time traders should price their orders more aggressively (closer to the best bid/offer price) or only submit these orders when the market spread becomes narrower.

Simplifications of the generalised gamma to the lognormal (kappa is zero), Weibull (kappa is one) or exponential (both kappa and sigma are one) distribution are strongly rejected. As Table 6.3 presents, the coefficient of Kappa is 0.11 with a standard error of 0.01 in the buy model and 0.14 with a standard error of 0.01 in the sell model. Accordingly, the estimated shape parameter (kappa) for buy and sell orders is more than 80 standard errors from one and 10 standard errors from zero. The coefficient of sigma is 2.26 with a standard error of less than 0.01 in the buy model and 2.23 with a standard error of less than 0.01 in the sell model. Accordingly the estimated scale parameter (sigma) for buy and sell orders is at least more than 120 standard errors from one.

To check the goodness-of-fit of the GG AFT model estimated above, the graphical diagnostic (Q-Q plot) discussed in Section 4.2.6 is used. Figure 6.1 shows Q-Q plots for both models and these plots obviously deviate from the 45-degree line. Compared to Figure 5.1, the goodness-of-fit is improved considerably, as the Q-Q plots are closer to the 45-degree line. Hence, the explanatory variables constructed in this chapter capture the UK market conditions and the characteristics of limit orders submitted at the LSE in a much better way than the explanatory variables suggested by LMZ, which have been empirically investigated in Chapter 5. However compared to the Q-Q plots reported in the LMZ study these Q-Q plots are still not close to the 45-degree line. Hence it is believed that there may be room for the goodness-of-fit of these models to be further improved. This is the subject of the investigation presented in the next chapter.

### *6.3.2 Empirical Estimations of the Cox PH Model*

In the PH model discussed in Subsection 4.2.5, the baseline hazard function $h_0(t)$ may be left unspecified, yielding the Cox PH model. The Cox PH model is the most frequently employed regression tool in survival analysis.[125] Its unique feature is that one could estimate the relationship between the hazard function $h(t)$ and explanatory variables X without any assumptions about the shape of the baseline hazard function. This is the main motivation for investigating the adequacy of a Cox PH model in addition to the AFT model of the previous section.

Table 6.4 presents the estimation results of the Cox PH model. The estimates associated with the conditioning variables generally are statistically significant for both models of buy and sell orders. As a PH model illustrates, these hazard ratios reflect the sensitivities of the hazard function $h(t)$ to explanatory variables X.[126] As presented in Subsection 4.2.1, the hazard function $h(t)$ is defined as the probability of completion of a limit order during a very small time interval, assuming that the limit order has survived (remained on the order book) to the beginning of the interval. If the hazard function of a limit order is lower, a longer completion time is expected.

[125] It is sometimes referred to as a semi-parametric approach.
[126] See Subsection 4.2.6 for detail about hazard ratio.

*The Estimated Hazard Ratio of PA*

The estimated hazard ratio of PA ($e^{\hat{\beta}_{PA}}$) is 0.04 for buy and sell orders. According to the Cox PH model presented in Subsection 4.2.5 of Chapter 4, if a buy order is priced 1% of the mid-quote price below the best bid price (PA value increases by one unit holding the remaining variables constant) the hazard function will decrease to 4%. If a sell order is priced 1% of the mid-quote price above the best offer price, the hazard function will also decrease to 4%. Accordingly in both cases a longer completion time is expected. This finding is consistent with the coefficient estimates on PA in the GG AFT model discussed above, which indicate that an increase of PA value will extend the expected limit-order completion time. Therefore, the underlying economic interpretation of the coefficient estimates on PA remains the same as presented above.

In Subsection 6.3.1, the coefficient estimates on PA in the GG AFT models show that completion time of buy orders is slightly more sensitive to this variable than that of sell orders. This is not consistent with these estimates discussed here, which show that the estimated hazard ratios of PA are the same for buy and sell orders. This implies that the Cox PH model estimated here could be wrongly specified. This issue will be discussed later.

These estimates revealed above are expected since orders are 'queuing up' on the order book for subsequent execution and PA captures the length of this 'queue'. These estimates also quantify trade-offs between gains from execution prices and losses from execution probability predicted by Cohen *et al.* (1981). These estimates may also explain the pricing strategies discussed in Subsection 6.2.2. These estimates show that traders are aware of the impact of the position of a limit order on its completion probability and hence intentionally price their orders at the best bid/offer price to improve the limit-order completion probability. These estimates imply that in order to improve the probability of order completion, traders need to price their orders as close to the best bid/offer price as possible. These estimates also imply that exchanges should encourage traders to submit aggressively priced orders, since this may improve the limit-order completion probability and hence facilitate the limit-order execution.

*The Estimated Hazard Ratio of BBOS*

The estimated hazard ratio of BBOS ($e^{\hat{\beta}_{BBOD}}$) is 0.22 for buy and sell orders. These indicate that if the best bid-offer spread is widened by 1% of the mid-quote price (BBOS value increases by one unit holding the remaining variables constant) the hazard function will decrease to 22%. Accordingly, a longer completion time is expected. This finding is consistent with the coefficient estimates on BBOS in the GG AFT model discussed above, which indicate that an increase of BBOS value will extend the expected limit-order completion time. Therefore, the underlying economic interpretation of the coefficient estimates on BBOS remains the same as presented above.

These estimates imply that traders should submit limit orders when the market spread becomes narrower. These estimates also imply that exchanges should minimise the market spread to facilitate the order execution. For example, as suggested in Subsection 6.3.1, exchanges should prefer a hybrid (order-driven/quote-driven) market to a pure order-driven market. Since in the hybrid market dealers are obliged to compete with the order book for business, increased competition should narrow the spread. Hence the hybrid market may be superior to a pure order-driven market in terms of lowering transaction cost by narrowing the spread and consequently improving limit-order completion probability.

*The Estimated Hazard Ratio of BSIDS*

The estimated hazard ratio of BSIDS ($e^{\hat{\beta}_{BSIDS}}$) is 0.68 for buy orders and 0.67 for sell orders. These indicate that if the prior trade was initiated from the same side of the order book (BSIDS value is one holding the remaining variables constant), the hazard function will decrease to 68% for buy orders and 67% for sell orders. Accordingly, a longer completion time is expected. This finding is consistent with the coefficient estimates on BSIDS in the GG AFT model discussed above, which indicate that an increase of BSIDS value from zero to one will extend the expected limit-order completion time. Therefore the underlying economic interpretation of the coefficient estimates on BSIDS remains the same as presented above.

These estimates discussed above imply that traders should pay attention to activities on the opposite side of the market. Assume that these traders notice an inflow of sell market orders. This indicates that the price may fall. Hence these traders have to think carefully before rushing to buy any shares since their orders may be 'picked up' by informed sellers. However this could also be an opportunity, since it would be easier to complete buy limit orders and the waiting time could also be reduced.

*The Estimated Hazard Ratio of NL30NT*

The estimated hazard ratio of NL30NT ($e^{\hat{\beta}_{NL30NT}}$) is 1.47 for buy and sell orders. These indicate that if NL30MT value increases by one unit holding the remaining variables constant, the hazard function will increase to 147%. Accordingly a shorter completion time is expected. This finding is consistent with the coefficient estimates on NL30NT in the GG AFT model discussed above, which indicate that an increase of NL30NT value will shorten the expected limit-order completion time. Therefore the underlying economic interpretation of the coefficient estimates on NL30NT remains the same as presented above.

As discussed in Subsection 6.3.1, the probability of limit-order completion depends on the overall volatility and the cost of being 'picked up' depends on the 'information-driven' component of the volatility since the probability of being 'picked up' depends on the arrival of traders with private information. Hence in order to facilitate limit-order execution exchanges should increase the 'mechanic' volatility by improving their trade systems. However in order to reduce the information-oriented volatility exchanges should disclose the real-time order-book information to all traders. These estimates also show that traders should submit limit orders in a 'moderate' volatile market, since the probability of limit-order completion will be improved. Although higher volatility will generally improve the probability of limit-order completion, traders should not submit limit orders in a highly volatile market since the cost could be far higher than the benefit as discussed in Subsection 6.3.1.

*The Estimated Hazard Ratio of LOS*

The estimated hazard ratio of LOS ($e^{\hat{\beta}_{LOS}}$) is 0.87 for buy orders and 0.97 for sell orders. These indicate that if LOS value increases by one unit holding the remaining variables constant, the hazard function will decrease to 87% for buy orders and 97% for sell orders. Accordingly a shorter completion time is expected. This finding is unexpected and opposite to the coefficient estimates on LOS in the GG AFT model discussed above, which indicate that an increase of LOS value will extend the expected limit-order completion time. Therefore the underlying economic interpretation of the coefficient estimates on LOS is opposite to what has been presented above. The unexpected coefficient estimates indicate that this model could be wrongly specified. This will be confirmed by the Q-Q plots presented below.

The estimates discussed above indicate that the larger a limit order, the higher the completion probability. This is unexpected since intuitively the larger a limit order, the more difficult to complete this order. Assume that these estimates are correct. In this case, traders would have an incentive to pool small orders into a large order. This may affect the inflow of new orders and the continuous function of the market. In some extreme cases, the market may fail to function since no orders are submitted to the order book. Intuitively this could not happen and these estimates could not be true.

*The Estimated Hazard Ratio of LS*

The estimated hazard ratio of LS ($e^{\hat{\beta}_{LS}}$) is 1.07 for buy orders and 1.06 for sell orders. These indicate that if LS value increases by one unit holding the remaining variables constant, the hazard function will increase to 107% for buy orders and 106% for sell orders. Accordingly a longer completion time is expected. This finding is unexpected and opposite to the coefficient estimates on LS in the GG AFT model discussed above, which indicate that an increase of LS value will shorten the expected limit-order completion time. Therefore the underlying economic interpretation of the coefficient estimates on LS is opposite to what has been presented above. The unexpected coefficient estimates also indicate that this model could be wrongly specified. This will be confirmed by the Q-Q plots presented below.

This finding indicates that traders should always monitor the liquidity available on the opposite side of the order book and submit limit orders when less liquidity is available. Intuitively this could not be true. Hence these estimates could not be true.

To check the goodness-of-fit of the Cox PH model estimated above the graphical diagnostic (Q-Q plot) is used. Figure 6.2 shows Q-Q plots for Cox PH models of buy and sell orders. Obviously these plots deviate markedly from the 45-degree line. These plots indicate that the Cox PH model is wrongly specified relative to the AFT model.

Since the explanatory variables are the same the better performance of the GG AFT model presented above can be driven by the better interaction between explanatory variables and the baseline survival time distribution. In other words, rescaling baseline survival time, which has the form of a generalised gamma distribution, captures the effects of explanatory variables in a much better way than rescaling hazard function. For this reason, the GG AFT model is used in the remainder of this thesis.

## 6.4 Summary and Conclusions

In this chapter explanatory variables are designed to capture the characteristics of limit orders and order book conditions while simultaneously have no higher than 40% correlation across each other. And separate models of limit-order completion time are constructed for buy and sell orders.

In this chapter both the GG AFT and Cox PH models are used to incorporate the explanatory variables.[127] The empirical estimations of the GG AFT model (Table 6.3) show that all the estimated coefficients have the expected sign. They indicate that limit-order completion time is more sensitive to some variables, such as PA and BBOS, but not to other variables, such as LOS and LS. The Q-Q plots of the GG AFT model

[127] This study also attempts three frequently used parametric PH models: exponential PH model ($h_0(t)=1$), Weibull PH model ($h_0(t)= pt^{p-1}$) and Gompertz PH model ($h_0(t)=e^{rt}$). The estimates of these models are similar to and have the same interpretations as those of the Cox PH model discussed above. Q-Q plots of these models also markedly deviate from the 45-degree line as those of the Cox PH model. Hence, these models are also wrongly specified and cannot be selected. There is more detail in Appendix E.

indicate that the explanatory variables constructed in this chapter capture the UK market conditions and the characteristics of limit orders submitted at the LSE in a much better way than the explanatory variables suggested by LMZ, which have been empirically investigated in Chapter 5. The empirical estimations of the Cox PH model as Table 6.4 presents show that the estimated hazard ratios of two variables (LOS and LS) are unexpected. The Q-Q plots of the Cox PH model deviate markedly from the 45-degree line. Basing on the Q-Q plots closeness to the 45-degree line and all estimated coefficients having expected signs the GG AFT model performs better the Cox PH model and is therefore used in the remaining chapters of this thesis.[128] However, compared to the Q-Q plots reported in the LMZ study, the Q-Q plots of the GG AFT model are still not close to the 45-degree line. Hence it is believed that there may be room for the goodness-of-fit of these models to be further improved. This is the subject of the investigation presented in the next chapter.

All variables suggested by LMZ and constructed in Chapter 5, those constructed in this chapter and those used in the literature reviewed in Subsection 3.4.2 are measured at a single fixed time during an order's 'life', namely at the time of order submission. Thus these variables and the associated models can be labelled as static. However, many variables, especially those that describe the state of the market, such as the bid-offer spread, are dynamic and change over time. These time-varying variables and the associated models can be labelled as dynamic. The next chapter presents an investigation into the relative adequacy of time-varying version of the variables constructed in this chapter. This is looked upon as a move from static to dynamic models.

[128] Table D.2 in Appendix D presents the empirical estimations of the GG AFT model with all explanatory variables including LSS. Compared to the empirical estimations when LSS is not included as Table 6.3 presents, the estimate of LOS decreases about 124% from 0.17 to -0.04 in the buy model and 162% from 0.13 to -0.08 in the sell model. Hence this model suffers from the effects of multicollinearity as expected.

**Table 6.1: Summary Statistics of the Explanatory Variables**

This table contains summary statistics of the explanatory variables for limit orders of a pooled sample of 38 stocks, throughout a sample period from October 2000 to December 2000. The variables have the interpretation as follows. PA measures the relative distance between the limit-order price and the contemporary best bid/offer price on the order book. BBOS represents the relative best bid-offer spread. BSIDS is an indicator of whether or not the prior trade was same-side initiated. NL30MT is a trading activity measure that provides an absolute measure of volatility. LOS captures the limit-order size. LS captures the depth of the opposite side of the order book. T refers to trades and S refers to shares.

| **Variables** | **Units** | **Mean** | **Std. Dev.** | **Skewness** | **Normal Kurtosis** | **5th Percentile** | **50th Percentile** | **95th Percentile** |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | | |
| **Buy Orders** | | | | | | | | |
| **PA** | % | 0.15 | 0.40 | 3.38 | 14.89 | 0.00 | 0.00 | 1.02 |
| **BBOS** | % | 0.33 | 0.33 | 3.80 | 27.92 | 0.07 | 0.24 | 0.90 |
| **BSIDS** | N/A | 0.38 | 0.48 | 0.51 | 1.26 | 0.00 | 0.00 | 1.00 |
| **NL30MT** | LogT | 3.29 | 0.88 | -0.43 | 3.96 | 1.79 | 3.33 | 4.67 |
| **LOS** | LogS | 8.91 | 1.46 | -0.57 | 4.16 | 6.45 | 9.02 | 10.90 |
| **LS** | LogS | 9.52 | 1.53 | -0.53 | 4.16 | 6.91 | 9.62 | 11.74 |
| | | | | | | | | |
| **Sell Orders** | | | | | | | | |
| **PA** | % | 0.14 | 0.37 | 3.73 | 17.88 | 0.00 | 0.00 | 0.88 |
| **BBOS** | % | 0.33 | 0.32 | 3.67 | 26.75 | 0.07 | 0.24 | 0.89 |
| **BSIDS** | N/A | 0.40 | 0.49 | 0.40 | 1.16 | 0.00 | 0.00 | 1.00 |
| **NL30MT** | LogT | 3.28 | 0.88 | -0.40 | 3.87 | 1.79 | 3.33 | 4.66 |
| **LOS** | LogS | 8.90 | 1.46 | -0.65 | 4.97 | 6.55 | 8.99 | 10.95 |
| **LS** | LogS | 9.49 | 1.54 | -0.54 | 4.13 | 6.87 | 9.62 | 11.71 |

**Table 6.2: Correlation Matrices of the Explanatory Variables**

This table contains correlation matrices of the explanatory variables for limit orders of a pooled sample of 38 stocks, throughout a sample period from October 2000 to December 2000. The variables have the interpretation as follows. PA measures the relative distance between the limit-order price and the contemporary best bid/offer price on the order book. BBOS represents the relative best bid-offer spread. BSIDS is an indicator of whether or not the prior trade was same-side initiated. NL30MT is a trading activity measure that provides an absolute measure of volatility. LOS captures the limit-order size. LS captures the depth of the opposite side of the order book.

| | Sell Orders | | | | | | |
|---|---|---|---|---|---|---|---|
| | | **PA** | **BBOS** | **BSIDS** | **NL30MT** | **LOS** | **LS** |
| **Buy Orders** | **PA** | 100% | -8% | 6% | 4% | 11% | 6% |
| | **BBOS** | -10% | 100% | 6% | -26% | -11% | -11% |
| | **BSIDS** | 1% | 6% | 100% | -6% | -9% | -8% |
| | **NL30MT** | 7% | -26% | -4% | 100% | 29% | 29% |
| | **LOS** | 14% | -11% | -8% | 28% | 100% | 33% |
| | **LS** | 8% | -0.11% | -7% | 29% | 31% | 100% |

**Table 6.3: Parameter Estimates of the GG AFT Model**

The table below shows parameter estimates of the accelerated failure time specification of limit-order completion under the generalised gamma distribution for limit orders of a pooled sample of 38 stocks, throughout a sample period from October 2000 to December 2000. The variable '_cons' denotes the intercept and the definitions of the remaining explanatory variables are given as follows. PA measures the relative distance between the limit-order price and the contemporary best bid/offer price on the order book. BBOS represents the relative best bid-offer spread. BSIDS is an indicator of whether or not the prior trade was same-side initiated. NL30MT is a trading activity measure that provides an absolute measure of volatility. LOS captures the limit-order size. LS captures the depth of the opposite side of the order book. Z-statistics are asymptotically standard normal under the null hypothesis that the corresponding coefficient is zero. The explanatory variables are measured at the time of order submission.

| Variable | Coef. | Std. Err. | z | P>z | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| | | | | | | |
| **Buy Orders** | | | | | | |
| **PA** | 5.52 | 0.03 | 190.54 | 0.00 | 5.46 | 5.57 |
| **BBOS** | 2.29 | 0.02 | 144.63 | 0.00 | 2.26 | 2.32 |
| **BSIDS** | 0.80 | 0.01 | 103.28 | 0.00 | 0.78 | 0.81 |
| **NL30MT** | -0.73 | 0.00 | -159.21 | 0.00 | -0.73 | -0.72 |
| **LOS** | 0.17 | 0.00 | 66.86 | 0.00 | 0.16 | 0.17 |
| **LS** | -0.12 | 0.00 | -49.29 | 0.00 | -0.13 | -0.12 |
| **_cons** | 7.31 | 0.03 | 253.45 | 0.00 | 7.26 | 7.37 |
| | | | | | | |
| **/ln_sig** | 0.81 | 0.00 | 409.42 | 0.00 | 0.81 | 0.82 |
| **/kappa** | 0.11 | 0.01 | 19.58 | 0.00 | 0.10 | 0.12 |
| | | | | | | |
| **sigma** | 2.26 | 0.00 | | | 2.25 | 2.27 |
| | | | | | | |
| **Sell Orders** | | | | | | |
| **PA** | 5.34 | 0.03 | 184.07 | 0.00 | 5.28 | 5.39 |
| **BBOS** | 2.35 | 0.02 | 143.43 | 0.00 | 2.32 | 2.38 |
| **BSIDS** | 0.83 | 0.01 | 106.17 | 0.00 | 0.81 | 0.85 |
| **NL30MT** | -0.70 | 0.00 | -148.76 | 0.00 | -0.70 | -0.69 |
| **LOS** | 0.13 | 0.00 | 49.33 | 0.00 | 0.12 | 0.13 |
| **LS** | -0.10 | 0.00 | -40.20 | 0.00 | -0.11 | -0.10 |
| **_cons** | 7.37 | 0.03 | 249.98 | 0.00 | 7.31 | 7.42 |
| | | | | | | |
| **/ln_sig** | 0.80 | 0.00 | 388.07 | 0.00 | 0.80 | 0.81 |
| **/kappa** | 0.14 | 0.01 | 23.20 | 0.00 | 0.12 | 0.15 |
| | | | | | | |
| **sigma** | 2.23 | 0.00 | | | 2.22 | 2.24 |

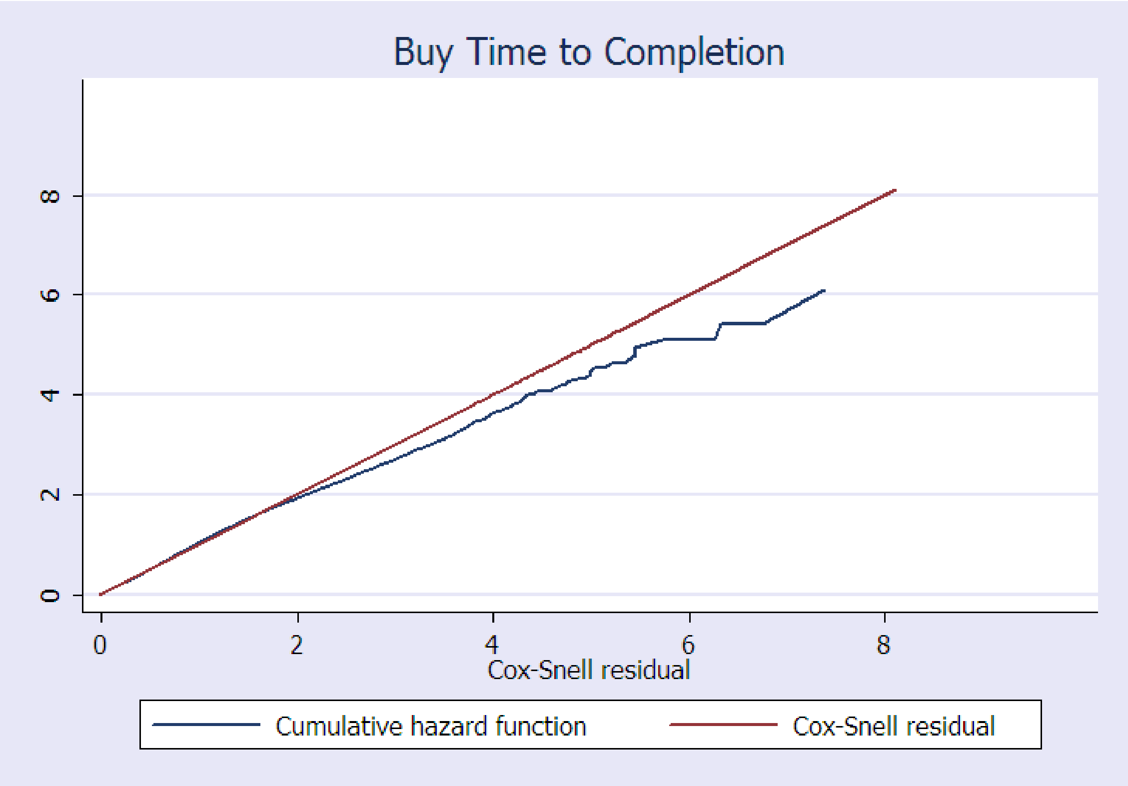
**Table 6.4: Parameter Estimates of the Cox PH Model**

The table below shows parameter estimates of the Cox proportional hazard model of limit-order completion for limit orders of a pooled sample of 38 stocks, throughout a sample period from October 2000 to December 2000. 'Haz. Ratio' denotes the hazard ratio. The definitions of the explanatory variables are given as follows. PA measures the relative distance between the limit-order price and the contemporary best bid/offer price on the order book. BBOS represents the relative best bid-offer spread. BSIDS is an indicator of whether or not the prior trade was same-side initiated. NL30MT is a trading activity measure that provides an absolute measure of volatility. LOS captures the limit-order size. LS captures the depth of the opposite side of the order book. Z-statistics are asymptotically standard normal under the null hypothesis that the corresponding hazard ratio is zero. The explanatory variables are measured at the time of order submission.

| Variable | Haz. Ratio | Std. Err. | z | P>z | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| | | | | | | |
| **Buy Orders** | | | | | | |
| **PA** | 0.04 | 0.00 | -181.61 | 0.00 | 0.03 | 0.04 |
| **BBOS** | 0.22 | 0.00 | -149.25 | 0.00 | 0.22 | 0.22 |
| **BSIDS** | 0.68 | 0.00 | -87.14 | 0.00 | 0.68 | 0.69 |
| **NL30MT** | 1.47 | 0.00 | 147.91 | 0.00 | 1.47 | 1.48 |
| **LOS** | 0.89 | 0.00 | -85.68 | 0.00 | 0.89 | 0.89 |
| **LS** | 1.07 | 0.00 | 50.80 | 0.00 | 1.07 | 1.08 |
| | | | | | | |
| **Sell Orders** | | | | | | |
| **PA** | 0.04 | 0.00 | -163.11 | 0.00 | 0.03 | 0.04 |
| **BBOS** | 0.22 | 0.00 | -142.45 | 0.00 | 0.22 | 0.23 |
| **BSIDS** | 0.67 | 0.00 | -89.58 | 0.00 | 0.67 | 0.68 |
| **NL30MT** | 1.47 | 0.00 | 140.22 | 0.00 | 1.46 | 1.47 |
| **LOS** | 0.91 | 0.00 | -69.23 | 0.00 | 0.91 | 0.91 |
| **LS** | 1.06 | 0.00 | 42.03 | 0.00 | 1.06 | 1.06 |

**Figure 6.1: Q-Q Plots of the GG AFT Model**

The following graph displays Q-Q plots of empirical estimates of the cumulative hazard functions of Cox-Snell residuals for limit orders of a pooled sample of 38 stocks, throughout a sample period from October 2000 to December 2000.



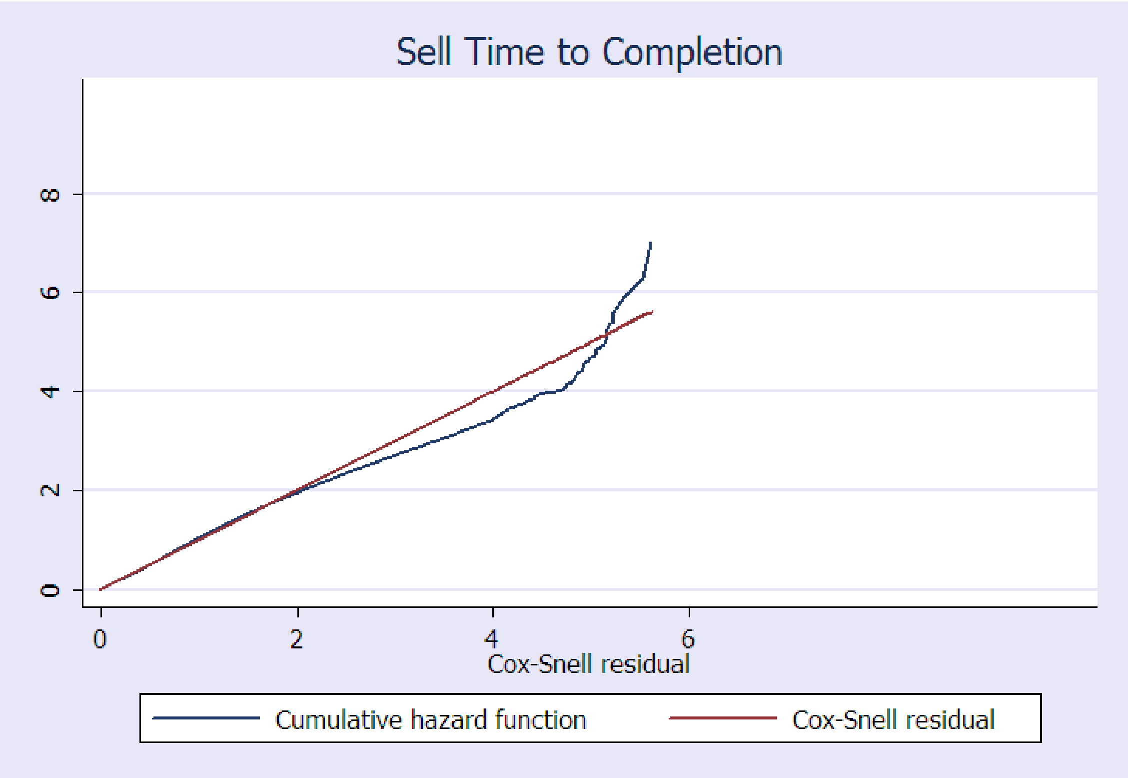
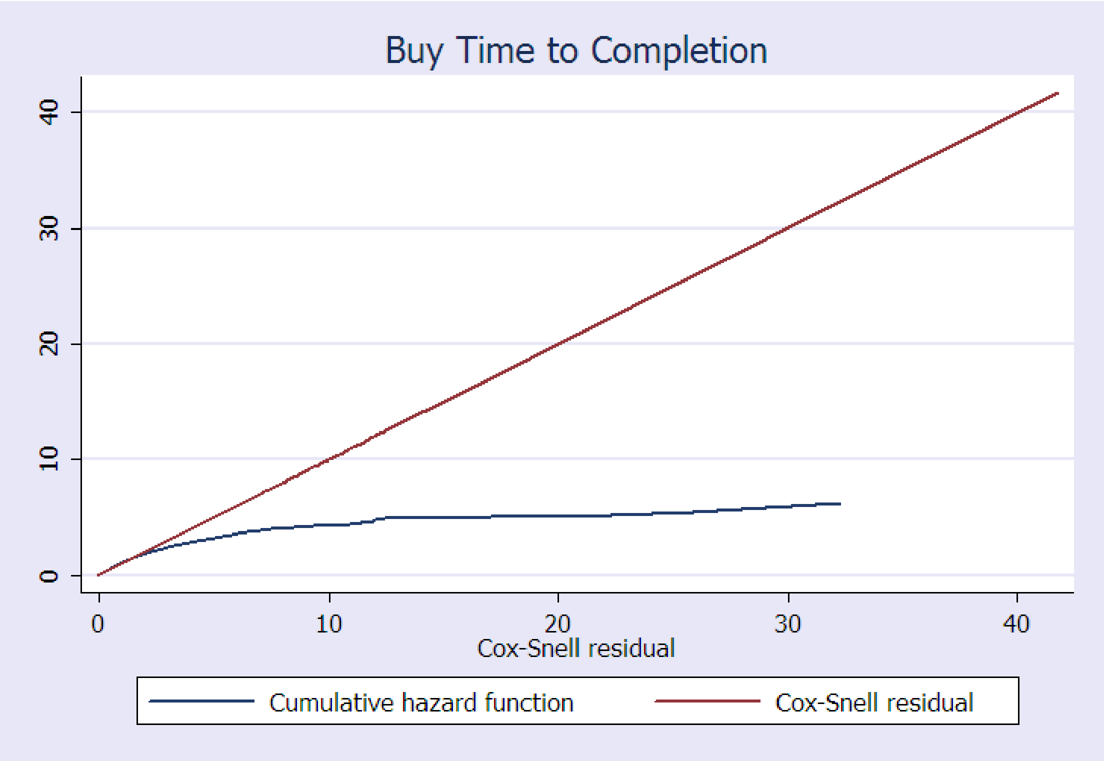
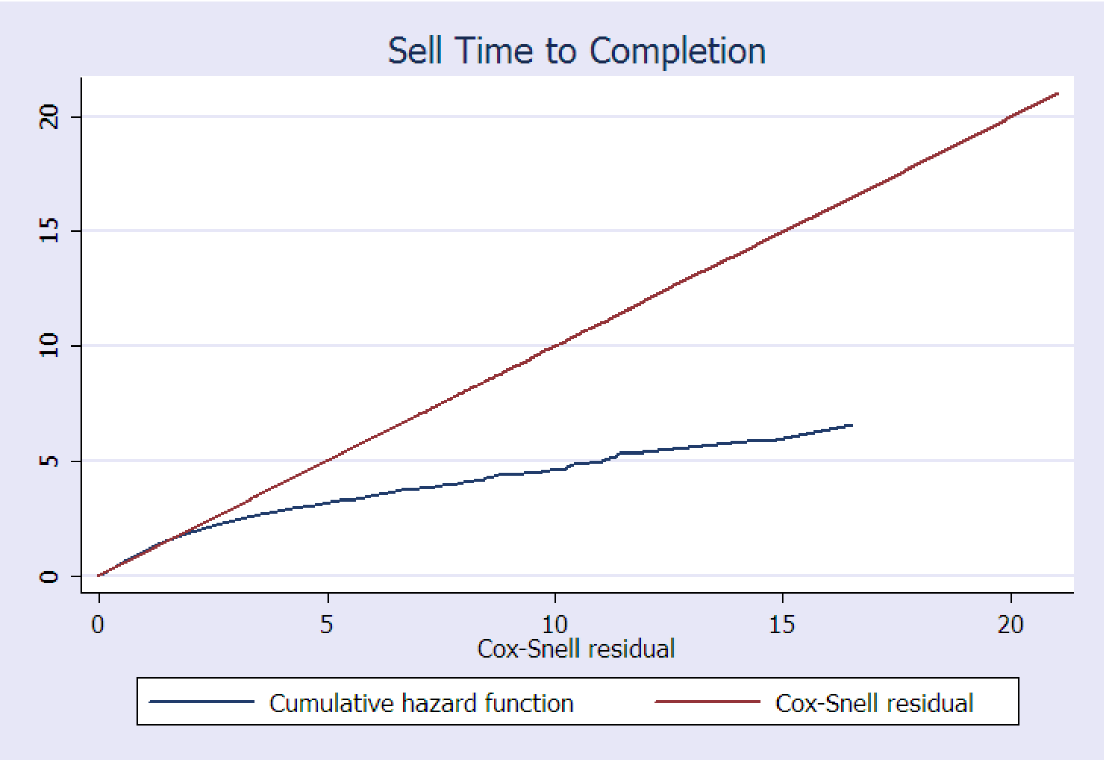



**Figure 6.2: Q-Q Plots of the Cox PH Model**

The following graph displays Q-Q plots of empirical estimates of the cumulative hazard functions of Cox-Snell residuals for limit orders of a pooled sample of 38 stocks, throughout a sample period from October 2000 to December 2000.



# CHAPTER 7 – THE EFFECT OF TIME-VARYING DYNAMICS OF THE ORDER BOOK ON LIMIT-ORDER COMPLETION TIME IN THE UK MARKET

## 7.1 Introduction

In Chapter 6, the static GG AFT models are used to investigate limit-order completion time in the UK market. The empirical estimations reported in Table 6.3 show that limit-order completion time is more sensitive to some variables (e.g. PA and BBOS) than others (e.g. LOS and LS). The Q-Q plots show that these variables produce models that fit the data better than the LMZ variables discussed in Chapter 5.

All explanatory variables used in Chapters 5 and 6 are measured at the time of order submission. The state of an order book (e.g. the best bid-offer spread), however, changes frequently. Thus, static explanatory variables would not capture the dynamics of an order book and would reflect order book information that is out-of-date as soon as the associated limit order were submitted. In this chapter, time-varying explanatory variables that capture these dynamics are proposed and used to investigate limit-order completion time in the UK market. Incorporating time-varying explanatory variables in survival models is not a novel concept; it has been investigated by statisticians (c.f., Leemis, Shih and Reynertson (1990), Lin and Ying (1995) and Satten, Datta and Robins (2001), and Jenkins (2004)) and applied in medical studies (c.f., McCullough *et al.* (2003)), but it is believed that this study is the first attempt to apply this approach to finance.

Much of the novelty of the present analysis is attributable to the fact that time-varying variables are used to look at how the limit-order completion time is affected by changes in market condition after the order is submitted. Another novelty is that this time-varying approach captures the 'real' dynamics of the order book and frees the assumption of the distribution of explanatory variables, since in most of the literature mentioned above the distribution of explanatory variables is assumed to have a certain form in order to incorporate the time-varying dynamics into the models.

The remainder of this chapter is organised as follows. Section 7.2 explains how the survival times are split to update the explanatory variables more frequently following order submission. Section 7.3 presents the empirical estimations of the dynamic GG AFT model. Section 7.4 presents the empirical estimations of a modified version of this dynamic GG AFT model. Section 7.5 presents an investigation into the predicted accuracy of this model. Section 7.6 offers a summary and conclusions, and introduces the work in the subsequent chapter.

## 7.2 Splitting Survival Times

In Chapters 5 and 6 all explanatory variables are measured and fixed at the time of order submission, so it is assumed that explanatory variables are all constant over time and remain so throughout the life of an order.[129] However, market conditions are constantly changing, therefore the longer a limit order remains on the order book, the less influence these static variables will have on the expected completion time. At some point, these static variables will fail to capture the extant state of the order book. For example, as Figure 2.8 of Chapter 2 illustrates, the best bid price is 524 pence and the best offer price is 525 pence. Let us assume that at 10:00am a buy limit order for stock ABC is submitted at a price of 524.5 pence. Since the limit-order price is higher than the prevailing best bid price, this limit-order price becomes the new prevailing best bid price. Consequently, this limit order will have the highest execution priority as soon as it is submitted. Accordingly the relative position of this limit order on the order book is captured by the variable PA, as discussed in Chapter 6.[130] Now let us assume that the market moved against this limit order and left it unexecuted on the order book and that by 11:00am the best bid price and the best offer price became 530 pence and 535 pence, respectively. As a result, this limit order is 'pushed back down the queue' and no longer has the highest priority for execution. Accordingly, the variable PA that was measured

[129] Operationally the data sample is organised as single-record survival data, having one row for each limit order as shown in Table 7.1.

[130] PA measures the relative distance between the limit-order price and the contemporary best bid (offer) price on the order book for buy (sell) orders. It captures the position of a limit order in the current market condition.

at 10:00am no longer captures the precise position of this order relative to the prevailing current best bid price. In order to capture the changing dynamics of the order book, time-varying explanatory variables are allowed to vary over time. These are then incorporated into the GG AFT model discussed in Chapter 6.

Technically, the estimation of survival models with time-varying explanatory variables requires that the survival time of each limit order be split into sub-periods. Within each of these sub-periods explanatory variables are assumed to be constant, but across the totality of the sub-periods explanatory variables can be re-measured and, hence, can take on different values. In this analysis, it is assumed that the explanatory variables are held constant for one hour.[131] Each limit order's behaviour is traced following submission. After each hour, if the order is still on the order book, the market conditions (such as spread costs) are re-evaluated and the corresponding explanatory variables are recalculated. Hence, all the explanatory variables, except LOS, are updated hourly.[132] For limit orders completed with one fill, LOS is constant over their lifetimes. For limit orders completed with multiple fills, LOS decreases after each fill.[133] Thus in this study LOS captures the limit-order size at the submission time only and is assumed to be constant over time, since this study concentrates on completed limit orders only.

In survival analysis, when an explanatory variable is allowed to have more than one entry, it is said to have switched from single-record to multiple-record survival data. The following is a theoretical example of changing single-record survival data into multiple-record survival data, which have multiple rows for each limit order. Table 7.1 presents some of the original single-record survival data used in Chapter 6.

[131] Table F.1 to Table F.2 in Appendix F show that only 3% of orders remain on the order book for more than one hour. Under this assumption, the survival times of these orders will be split into sub-periods of one hour. Given the computing facilities and technical support available, it is impractical to split them into further sub-periods of less than one hour.
[132] LOS captures the limit-order size, which is not altered following submission. A trader can change the size of his order by cancelling the existing order and resubmitting a new order.
[133] Given the computing facilities and technical support available, it is impractical to update LOS after every fill.

**Table 7.1: Single-Record Survival Data**

| Order | Survival time (in minutes) | Events | PA | BBOS | BSIDS | NL30MT | LOS | LS |
|---|---|---|---|---|---|---|---|---|
| **A** | 90 | Completion | 0 | 1 | 1 | 1 | 1 | 1 |
| **B** | 30 | Completion | 0 | 2 | 0 | 2 | 2 | 2 |
| **C** | 90 | Cancellation | 1 | 2 | 1 | 2 | 2 | 2 |

**Table 7.2: Multiple-Record Survival Data**

| Order | Survival time (in minutes) | Events | PA | BBOS | BSIDS | NL30MT | LOS | LS |
|---|---|---|---|---|---|---|---|---|
| **A** | 60 | Staying | 0 | 1 | 1 | 1 | 1 | 1 |
| **A** | 90 | Completion | 1 | 1 | 0 | 0 | 1 | 1 |
| **B** | 30 | Completion | 0 | 2 | 0 | 2 | 2 | 2 |
| **C** | 60 | Staying | 1 | 2 | 1 | 2 | 2 | 2 |
| **C** | 90 | Cancellation | 0 | 1 | 0 | 1 | 2 | 2 |

Each limit order is followed over its lifetime, and explanatory variables are updated every hour following submission. Accordingly, the single-record survival data as shown in Table 7.1 are changed into the multiple-record survival data of Table 7.2.

The interpretation of the records for order A in Table 7.2 is as follows:

*Interval (0, 60): PA=0, BBOS=1, BSIDS=1, NL30MT =1, LOS =1 and LS =1.*

These explanatory variables are measured at t=0 (the time of order submission). Since this order remains unexecuted on the order book at t=60 (one hour after submission), it is treated as a censored order at the end of this interval. In the second interval,

*Interval (60, 90): PA=1, BBOS=1, BSIDS=0, NL30MT =0, LOS =1 and LS =1*

the explanatory variables are updated and measured at t=60 (one hour after submission), and since this order is completed at t=90, it is treated as a completed order at the end of this interval. As a result, the single record for order A presented in Table 7.1 is split into two records for the same order, as presented in Table 7.2.[134] Accordingly, robust variance estimators are to be used, which will be discussed later in this section.

Similarly, the records for orders B and C that appear in Table 7.2 are interpreted as follows:

[134] The single record is split into two records for the same order, but not two new orders.

Order B

*Interval (0, 30): PA=0, BBOS=2, BSIDS=0, NL30MT =2, LOS =2 and LS =2.*

These explanatory variables are measured at t=0 (the time of order submission), and since this order is completed at t=30, it is treated as a completed order.

Order C

*Interval (0, 60): PA=1, BBOS=2, BSIDS=1, NL30MT =2, LOS =2 and LS =2.*

These explanatory variables are measured at t=0 (the time of order submission), and since this order remains unexecuted on the order book at t=60 (one hour after submission), it is treated as a censored order at the end of this interval. In the second interval,

*Interval (60, 90): PA=0, BBOS=1, BSIDS=0, NL30MT =1, LOS =2 and LS =2*

the explanatory variables are updated and measured at t=60 (one hour after submission), and since this order is cancelled at t=90, it is treated as a censored order.

Thus, in the multiple-record survival data, if a limit order remains on the order book by the end of each hour after the submission, it is treated as a censored order. Accordingly, the completion and censoring times do not exceed one hour.

Obviously, the change of survival data from single-record to multiple-record results in repeated observations for the same limit order, thereby distorting the conventional estimate of variance. Accordingly, robust variance estimators are used for the variance-covariance matrix of the coefficients (and hence, the reported standard errors) when analysing multiple-record survival data (c.f., Huber (1967), Kish and Frankel (1974), Fuller (1975), White (1980, 1982), Kent (1982), Binder (1983), Royall (1986), Gail, Tan, and Piantadosi (1988), Lin and Wei (1989), Froot (1989), Rogers (1993), Williams (2000) and Wooldridge (2002)).

## 7.3 Empirical Estimation of the Dynamic GG AFT Model

The GG AFT model discussed in Chapter 6 is used to incorporate the time-varying explanatory variables and shall henceforth be referred to as the dynamic GG AFT model for the remainder of this thesis. Table 7.3 presents the estimation results of the dynamic GG AFT model. The estimates of the parameters associated with the conditioning variables generally are statistically significant for both models. These coefficients reflect the sensitivities of limit-order completion time to the corresponding explanatory variables. All estimates have the same signs and, except ones of kappa, have similar magnitudes to those of the static GG AFT model reported in Table 6.3 and discussed in Subsection 6.3.1. Hence, these estimates have the same economic interpretations as those discussed in Subsection 6.3.1.

PA captures the position of a limit order in the current market condition. The estimate of the coefficient on PA is 5.29 with z-statistic of 123.27 for buy orders and 5.20 with z-statistic of 120.15 for sell orders. The positive sign of the coefficients of PA indicates that the lower (higher) the buy (sell) limit-order price below (above) the best bid (offer) price, the longer the expected completion time. This positive sign could be a mechanical consequence of the price priority in an order-driven market discussed in Chapter 2. It could also be due to aggressively priced orders inducing latent demand for trade, since aggressively priced limit orders could encourage traders on the opposite side of the market to submit market orders rather than limit orders as discussed in Chapter 6. These estimates discussed above imply that traders need to price their orders as close to the best bid/offer price as possible in order to minimise the expected waiting time of order completion and exchanges should encourage traders to submit aggressively priced orders in order to facilitate the limit-order execution.

BBOS represents the relative best bid-offer spread. The estimate of the coefficient on BBOS is 2.38 with z-statistic of 107.49 for buy orders and 2.44 with z-statistic of 110.05 for sell orders. The positive sign of the coefficients of BBOS indicates that the wider the best bid-offer spread, the longer the expected completion time and the lower the expected completion probability. It could also be due to narrow spreads inducing latent demand for trade, since narrow spreads encourage traders to submit market orders as discussed in Chapter 6. These estimates discussed above imply that traders should submit limit orders when the market spread becomes narrower and exchanges should minimise the market spread to facilitate the order execution.

BSIDS is an indicator of whether or not the prior trade was same-side initiated (BSIDS equals to one). The estimate of the coefficient on BSIDS is 0.77 with z-statistic of 101.51 for buy orders and 0.80 with z-statistic of 104.28 for sell orders. These estimates show that if the prior trade was initiated from the same side of the order book (BSIDS value is one holding the remaining variables constant), the expected completion time will be extended. The positive sign of the coefficients of BSIDS is expected as discussed in Chapter 6. These estimates discussed above indicate that traders should develop their order submission strategies basing on activities on the opposite side of the market.

NL30MT is a trading activity measure that provides an absolute measure of volatility. The estimate of the coefficient on NL30MT is -0.71 with z-statistic of -154.09 for buy orders and -0.68 with z-statistic of -145.24 for sell orders. The negative sign of the coefficients of NL30MT indicates that the higher the number of trades in the last 30 minutes, the more active and volatile the market becomes and hence the shorter the expected completion time. The estimates discussed above show that higher volatility would reduce limit-order completion time and hence facilitate the limit-order execution. These estimates also show that traders should submit limit orders in a volatile market, since the waiting time of limit-order completion could be reduced as discussed in Chapter 6.

LOS captures the limit-order size. The estimate of the coefficient on LOS is 0.17 with z-statistic of 66.15 for buy orders and 0.13 with z-statistic of 49.18 for sell orders. The positive sign of the coefficients of LOS indicates that the larger a limit order, the longer the expected completion time. This is expected since the larger a limit order, the more difficult to fill this order. The positive sign revealed above could be due to the competition among traders as discussed in Chapter 6. This finding implies that breaking a large limit order into a number of small orders would not reduce order completion time significantly. Hence this may not benefit traders, since the cost of monitoring these small orders could be much higher than the benefit of reducing completion time as discussed in Chapter 6.

LS captures the depth (liquidity) on the opposite side of an order book. It captures the best offer size for buy orders and the best bid size for sell orders. The estimate of the coefficient on LS is -0.12 with z-statistic of -49.44 for buy orders and -0.10 with z-statistic of 40.07 for sell orders. The negative sign of the coefficients of LS indicates that the larger the best offer (bid) size, the shorter the expected buy-order (sell-order) completion time. This is expected, as the more liquidity is available from the opposite side of the market, the more easily a limit order will be completed. This finding indicates that traders should submit limit orders when more liquidity is available in order to reduce the waiting time of completion.

The estimate of kappa has changed significantly. It increases around 82% from 0.11 to 0.20 in the buy model and 64% from 0.14 to 0.23 in the sell model. This indicates that the estimated baseline survival time distribution differs from that estimated in Chapter 6. Hence the estimated completion time distribution would differ from that estimated in Chapter 6 and the prediction based on these models discussed above would also differ from that based on those estimated in Chapter 6.

Simplifications of the generalised gamma to the lognormal (kappa is zero), Weibull (kappa is one) or exponential (both kappa and sigma are one) distribution are strongly rejected. As Table 7.3 presents, the coefficient of kappa is 0.20 with a standard error of 0.01 in the buy model and 0.23 with a standard error of 0.01 in the sell model. Hence, the estimated shape parameter (kappa) for buy and sell models is more than 70 standard errors from one and 20 standard errors from zero. The coefficient of sigma is 2.17 with a standard error of 0.01 in the buy model and 2.13 with a standard error of 0.01 in the sell model. Hence, the estimated scale parameter (sigma) for buy and sell models is at least more than 110 standard errors from one.

To check the goodness-of-fit of the dynamic GG AFT model estimated above, the graphical diagnostic (Q-Q plot) discussed in Subsection 4.2.6 is used.[135] Figure 7.1 shows Q-Q plots for both models of buy and sell orders, and these plots obviously deviate from the 45-degree line. Compared to Figure 6.1, Figure 7.1 shows that the goodness-of-fit is improved, as the Q-Q plots are closer to the 45-degree line. Hence, time-varying variables capture the state of an order book in a better way than static variables discussed in Chapter 6. However, these Q-Q plots still markedly deviate from the 45-degree line. Hence, it is believed that there may be room for the goodness-of-fit of these models to be further improved. This is the subject of the investigation presented in the next section.

## 7.4 Empirical Estimation of the Modified Dynamic GG AFT Model

The models presented in Chapters 5 and 6 assume a fixed shape of the baseline survival time distribution. Namely, the shape of the baseline distribution is assumed to be constant for different limit orders. As a result, all the ancillary parameters (sigma and kappa) are estimated as constants. LMZ (2002), however, report different shapes of survival functions for three randomly selected limit orders in their data sample. One can argue that these may indicate the shapes of the baseline survival time distribution could vary for individual limit orders. Smith and Vreeland (2003) suggest that modelling the ancillary parameter as a function of explanatory variables can capture the varying shape of the baseline survival time distribution.[136] Accordingly, in this section, the ancillary parameter (sigma) of the GG AFT model is modelled as a function of the variable LOS, which captures the characteristics of a limit order - the limit-order size, since the remaining variables mainly capture the market conditions.[137] Incorporating LOS into the sigma estimation of the AFT model is described next.

[135] In these Q-Q plots of the multiple-record survival data, the Cox-Snell residual is cumulative, which is based on calculating the Cox-Snell residual for each record and then summing them. Similar to the Cox-Snell residual in the Q-Q plots of the single-record survival data discussed in Chapters 5 and 6, one value per limit order is measured. All the Cox-Snell residuals in Q-Q plots of the dynamic GG AFT model used in the remainder of the thesis are cumulative.

[136] Smith and Vreeland (2003) analyse the 'survival time' of political leaders. They use the Weibull PH model and model the ancillary parameter as a function of one of their explanatory variables.

[137] This study also attempts to model the sigma as a function of the remaining explanatory variables. Appendix G presents empirical estimations and Q-Q plots of these attempts. Modelling the sigma as a function of LOS is the only one that improves the goodness-of-fit of this model as the Q-Q plots illustrate. The remaining variables, therefore, seem to affect the logarithm of limit-order completion time only

Equation (4.18) may be re-written as,

$$\ln(T) = \beta X^{'} + \sigma W \ . \tag{7.1}$$

Where X is a vector of explanatory variables and $\beta$ is a parameter vector. $W = \varepsilon / \sigma$ is an error term with density function $f(w)$ and $\sigma$ is a scale factor and captures the variance of the residual distribution.

Distributional assumptions about W determine the AFT model describing the random variable T. When the error term W has a suitable distribution, e.g. extreme value, generalised extreme value, normal or logistic, it leads to Weibull, generalised gamma, log-normal or log-logistic AFT models for T.

In this section, a modified version of the AFT model, which assumes that sigma depends on the variable LOS as Equation (7.2) illustrates, is used to incorporate the explanatory variables and investigate limit-order completion time in the UK market.

$$\sigma = c_{LOS} + \lambda_{LOS} LOS \ . \tag{7.2}$$

This model will be referred to as the modified AFT model. It is believed that this study is unique in applying this modified AFT model to finance.

If $\sigma$ depends on one of the explanatory variable, then the variance of the residual distribution varies across limit orders. Hence the heteroscedasticity exists. In this way, the heteroscedasticity can be tested and addressed. As far as we know, this study is the first to test and address the heteroscedasticity on survival data.

Table 7.4 presents the estimation results. The estimated parameters are statistically significant for both buy and sell orders. These coefficients reflect the sensitivities of limit-order completion time to the corresponding explanatory variables. All estimates have the same signs and similar magnitudes to those of the dynamic GG AFT model reported in Table 7.3 and discussed above, and, hence, have the same economic interpretations.

---

linearly by having significant coefficients in the expected part of Equation (7.1), but have no non-linear effects on the logarithm of completion time through the volatility of the unexpected component.

The economic interpretations can be summarised as follows. The positive sign of the coefficients of PA indicates that the higher (lower) the sell (buy) limit-order price above (below) the best offer (bid) price, the longer the expected waiting time of completion. The estimates indicate that traders should price their orders aggressively to reduce the expected waiting time of order completion and exchanges should encourage traders to submit aggressively priced orders in order to facilitate the limit-order execution. The positive sign of the coefficients of BBOS indicates that the narrower the market spread, the shorter the expected completion time and the higher the expected completion probability. The estimates suggest that traders should submit limit orders when the market spread becomes narrower. The positive sign of the coefficients of BSIDS indicates if the prior trade was initiated from the opposite side of the order book (BSIDS value is zero holding the remaining variables constant), the expected waiting time of completion will be reduced. The estimates imply that market activities on the opposite side of the market could affect the expected limit-order completion time. The negative sign of the coefficients of NL30MT indicates that the lower the number of trades in the last 30 minutes, the longer the expected waiting time of completion. The estimates show that traders should submit limit orders in a volatile market. The positive sign of the coefficients of LOS indicates that the smaller a limit order, the shorter the expected waiting time of completion. This finding suggests that breaking a large limit order into a number of small orders may not reduce the expected limit-order completion time considerably. The negative sign of the coefficients of LS indicates that the smaller the best offer (bid) size, the longer the expected buy-order (sell-order) completion time. This finding indicates that traders should submit limit orders when more liquidity is available in order to reduce the waiting time.

The empirical results also show that sigma depends on LOS. The estimate of $\lambda_{LOS}$ is 0.04 with z-statistic of around 42 for buy and sell orders. This indicates that the larger the limit order, the larger the ‘scale’ of the baseline survival time distribution as Equation (7.1) presents. In other words, the larger the limit order, the higher the volatility of the unexpected component in Equation (7.1). This also indicates that heteroscedasticity exists, since the variance of the residuals varies across the limit orders.

To check the goodness-of-fit of the modified dynamic GG AFT model estimated above, the graphical diagnostic (Q-Q plot) discussed in Subsection 4.2.6 is used. Figure 7.2 shows Q-Q plots for both models. Compared to Figure 7.1, Figure 7.2 shows that the goodness-of-fit is improved, as the Q-Q plots are much closer to the 45-degree line. Hence, this model is better specified than that discussed in Section 7.3. There are only a few limit orders with Cox-Snell residuals greater than five in each model and hence are suspected to be the main cause of the deviation from the 45-degree line. Figure 7.3 shows Q-Q plots after excluding these orders. They result in a line closer to the 45-degree line and represent a very good fit and a well-specified model. Hence, in terms of goodness-of-fit, this modified dynamic GG AFT model performs best and will be used in the remainder of the thesis.

The empirical findings and improved goodness-of-fit discussed above indicate that the variance of the residuals varies across the limit orders and the heteroscedasticity exists in the pooled data.[138] This study shows that modelling the ancillary parameters as a function of explanatory variables could be a new approach to test and address heteroscedasticity on survival data, since any traditional heteroscedasticity tests cannot handle one aspect of survival data: censoring.[139]

The estimation results presented above seem to indicate that order size is not only directly related to the expected logarithm of completion times ($\hat{\beta}X$) but also important in capturing variations in unexpected completion times across orders, through a change in the shape of baseline distribution. In other words, order size affects the logarithm of completion time both in linear fashion (on the level, or expected value) and non-linear fashion (on the shape of the residual distribution). These results also indicate that if order size is a measure of types of market activities, then such activities seem to be important in determining variations in completion times across orders.

---

**138** Due to the size of the data sample used in this study, it is expected that $\hat{\lambda}_{LOS}$ will be statistical significant. Without improving goodness-of-fit, these estimates alone cannot justify the existence of heteroscedasticity. This is the reason why this heteroscedasticity issue is left to be addressed here but not in Chapter 5, as the newly constructed time-varying explanatory variables need to be incorporated into the models first.

**139** ARCH and GARCH models can handle the heteroscedasticity but still cannot handle the censoring aspect of survival data. For this reason ARCH and GARCH models cannot be used in this study.

Although Figure 7.2 shows a good fit, it does so for the pooled data and says nothing about the performance of the model for individual stocks. Q-Q plots of the 38 individual stocks are listed by ticker symbols in Appendix H. The Q-Q plots are constructed stock-by-stock by calculating Cox-Snell residuals for each stock using the modified dynamic GG AFT model estimated with the pooled data, and stock-specific limit-order data.[140] They show that although there are some variations in the goodness-of-fit of the models across stocks, the pooled models also fit individual stock limit-order data quite well. The good fit of the pooled models at the individual stock level also indicates that the explanatory variables do a good job at capturing the cross-sectional difference.

The good fit of the modified dynamic GG AFT model on the pooled data and stock-specific limit-order data says nothing about the prediction accuracy of this model. Thus the next section presents an investigation of its prediction accuracy.

### 7.5 Prediction Accuracy of the Modified Dynamic GG AFT Model

A point prediction, which is a single value forecast for survival time, is used to examine the prediction accuracy of the modified dynamic model discussed above. In this section, for each completed limit order a predicted median survival time is computed and compared to the actual survival time. The median survival time is used more often than the mean because distributions of survival times often tend to be skewed.[141]

Parkes (2000) provides a method of measuring point prediction accuracy. He defines that a predicted median survival time is in a 'serious error' range if it is either less than half the actual survival time or more than twice the actual survival time. He also defines that a prediction is 'pessimistic' if the predicted median survival time is less than half the actual survival time and 'optimistic' if the predicted median survival time is more than twice the actual survival time. This section presents an investigation of the

[140] The modified dynamic GG AFT model is not re-estimated for each stock. Instead, the Cox-Snell residuals for each stock are extracted directly from those for the pooled sample.

[141] One median survival time for each completed limit order will be computed on the explanatory variables measured at the submission time. Table C.3 in Appendix C shows how skewed the survival-time distributions are for the limit orders sampled in this study.

percentage of predicted median survival times of limit orders in the data sample used in this study that fell into the serious error range discussed above.

Table 7.5 presents the percentage of median survival times predicted by the modified dynamic GG AFT model that fall within the serious error range discussed above. As Table 7.5 shows, predictions are optimistic for 65.79% of buy orders and 65.23% of sell orders, and predictions are pessimistic for 10.87% of buy orders and 11.29% of sell orders. There is, therefore, a tendency to be too optimistic.

Although Parkes' definition of serious error gives a generous range of predicted median survival times, about 75% of statistical predictions are still in the error range as shown by Table 7.5. The poor prediction performance is no surprise. Henderson, Jones and Stare (2001) demonstrate that poor prediction accuracy is inherent in survival analysis. They investigate the prediction accuracy of some commonly-used survival models, such as the lognormal and Weibull AFT models. They demonstrate that mathematically for these survival models at least 50% of cases will fall into Parkes' serious error range even without the additional complication of errors in sampling and estimation. They also provide empirical evidence to support this argument. They conclude that at the individual level even if complete information pertaining to explanatory variables is available and all parameters are fully known poor point predictions should be expected for these well known and much used survival models. They do not, however, cover the GG AFT model.

In addition, the modified dynamic GG AFT model is estimated using the data sample in which 64% of orders are censored as Table 4.1 presents. Since censored orders do not provide information about the time that orders are completed, the high proportion of censored orders in the data sample used in this study could very well be the reason behind the poor prediction accuracy. It is possible that the poor prediction accuracy discussed above is an unavoidable consequence of the data used in this study.

## 7.6 Summary and Conclusions

In this chapter, time-varying explanatory variables that capture the dynamics of an order book are used to investigate limit-order completion time in the UK market. Technically, estimation of survival models with time-varying explanatory variables requires that the survival times of limit orders be split. The survival time of each limit order is split into sub-periods within each of which explanatory variables are assumed to be constant but across each of which explanatory variables can be re-measured and, hence, can assume different values. In this analysis, it is assumed that the explanatory variables are held constant for one hour. Accordingly, single-record survival data used in Chapter 6 are changed into multiple-record survival data.

In this chapter, the GG AFT model is used to incorporate time-varying explanatory variables. Table 7.3 presents the estimation results of the dynamic GG AFT model. The estimates of the parameters associated with the conditioning variables generally are statistically significant for both models. Figure 7.1 shows Q-Q plots for both models of buy and sell orders. Compared to Figure 6.1, Figure 7.1 shows that the goodness-of-fit is improved, as the Q-Q plots are closer to the 45-degree line. Hence, time-varying variables capture the state of an order book in a better way than the static variables discussed in Chapter 6.

In this chapter, a modified version of the dynamic GG AFT model, which assumes that sigma depends on LOS, is also used to incorporate the explanatory variables. Table 7.4 presents the estimation results. The estimated parameters are statistically significant for both buy and sell orders. Figure 7.2 shows Q-Q plots for both models. Compared to Figure 7.1, Figure 7.2 shows that the goodness-of-fit is improved, as the Q-Q plots are much closer to the 45-degree line. Hence, in terms of goodness-of-fit, this modified dynamic GG AFT model performs best and will be used in the remainder of the thesis.

The Q-Q plots are constructed stock-by-stock by calculating Cox-Snell residuals for each stock using the modified dynamic GG AFT model estimated with the pooled data, and stock-specific limit-order data. They show that although there are some variations in the goodness-of-fit of the models across stocks, the pooled models also fit individual stock limit-order data quite well.

A point prediction is used to examine the prediction accuracy of the modified dynamic GG AFT model. Table 7.5 shows that predictions are optimistic - that is, more than twice the actual survival time - for 65.79% of buy orders and 65.23% of sell orders. Predictions are pessimistic - less than half the actual survival time - for 10.87% of buy orders and 11.29% of sell orders. There is, therefore, a tendency to be too optimistic.

Greater trading activities at certain times of day are expected to substantially increase the probability of execution of a limit order. Hence, the submission time of a limit order could affect its completion probability and completion time. In pursuing the object of investigating the effects of the time of order submission on limit-order completion time in the UK market the next chapter presents a set of modified dynamic GG AFT models that capture these time-of-day effects.

**Table 7.3: Parameter Estimates of the Dynamic GG AFT Model**

The table below shows parameter estimates of the accelerated failure time specification of limit-order completion under the generalised gamma distribution for limit orders of a pooled sample of 38 stocks, throughout a sample period from October 2000 to December 2000. The variable '_cons' denotes the intercept and the definitions of the remaining explanatory variables are given as follows. PA measures the relative distance between the limit-order price and the contemporary best bid/offer price on the order book. BBOS represents the relative best bid-offer spread. BSIDS is an indicator of whether or not the prior trade was same-side initiated. NL30MT is a trading activity measure that provides an absolute measure of volatility. LOS captures the limit-order size. LS captures the depth of the opposite side of the order book. Z-statistics are asymptotically standard normal under the null hypothesis that the corresponding coefficient is zero. The explanatory variables, except LOS, are measured hourly after submission.

| Variable | Coef. | Robust Std. Err. | z | P>z | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| | | | | | | |
| **Buy Orders** | | | | | | |
| **PA** | 5.29 | 0.04 | 123.27 | 0.00 | 5.20 | 5.37 |
| **BBOS** | 2.38 | 0.02 | 107.49 | 0.00 | 2.34 | 2.42 |
| **BSIDS** | 0.77 | 0.01 | 101.51 | 0.00 | 0.75 | 0.78 |
| **NL30MT** | -0.71 | 0.00 | -154.09 | 0.00 | -0.72 | -0.70 |
| **LOS** | 0.17 | 0.00 | 66.15 | 0.00 | 0.16 | 0.17 |
| **LS** | -0.12 | 0.00 | -49.44 | 0.00 | -0.13 | -0.12 |
| **_cons** | 7.29 | 0.03 | 249.27 | 0.00 | 7.23 | 7.35 |
| | | | | | | |
| **/ln_sig** | 0.77 | 0.00 | 330.43 | 0.00 | 0.77 | 0.78 |
| **/kappa** | 0.20 | 0.01 | 31.26 | 0.00 | 0.19 | 0.21 |
| | | | | | | |
| **sigma** | 2.17 | 0.01 | | | 2.16 | 2.18 |
| | | | | | | |
| **Sell Orders** | | | | | | |
| **PA** | 5.20 | 0.04 | 120.15 | 0.00 | 5.11 | 5.28 |
| **BBOS** | 2.44 | 0.02 | 110.05 | 0.00 | 2.39 | 2.48 |
| **BSIDS** | 0.80 | 0.01 | 104.28 | 0.00 | 0.78 | 0.81 |
| **NL30MT** | -0.68 | 0.00 | -145.24 | 0.00 | -0.69 | -0.67 |
| **LOS** | 0.13 | 0.00 | 49.18 | 0.00 | 0.12 | 0.13 |
| **LS** | -0.10 | 0.00 | -40.07 | 0.00 | -0.11 | -0.10 |
| **_cons** | 7.34 | 0.03 | 247.23 | 0.00 | 7.28 | 7.40 |
| | | | | | | |
| **/ln_sig** | 0.76 | 0.00 | 309.25 | 0.00 | 0.75 | 0.76 |
| **/kappa** | 0.23 | 0.01 | 36.29 | 0.00 | 0.22 | 0.24 |
| | | | | | | |
| **sigma** | 2.13 | 0.01 | | | 2.12 | 2.15 |

**Table 7.4: Parameter Estimates of the Modified Dynamic GG AFT Model**

The table below shows parameter estimates of the accelerated failure time specification of limit-order completion under the generalised gamma distribution for limit orders of a pooled sample of 38 stocks, throughout a sample period from October 2000 to December 2000. The variable '_cons' denotes the intercept and the definitions of the remaining explanatory variables are given as follows. PA measures the relative distance between the limit-order price and the contemporary best bid/offer price on the order book. BBOS represents the relative best bid-offer spread. BSIDS is an indicator of whether or not the prior trade was same-side initiated. NL30MT is a trading activity measure that provides an absolute measure of volatility. LOS captures the limit-order size. LS captures the depth of the opposite side of the order book. Z-statistics are asymptotically standard normal under the null hypothesis that the corresponding coefficient is zero. The explanatory variables, except LOS, are measured hourly after submission.

| Variable | Coef. | Robust Std. Err. | z | P>z | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| | | | | | | |
| **Buy Orders** | | | | | | |
| **PA** | 5.31 | 0.04 | 122.12 | 0.00 | 5.23 | 5.40 |
| **BBOS** | 2.38 | 0.02 | 108.17 | 0.00 | 2.33 | 2.42 |
| **BSIDS** | 0.75 | 0.01 | 99.78 | 0.00 | 0.73 | 0.76 |
| **NL30MT** | -0.70 | 0.00 | -151.70 | 0.00 | -0.71 | -0.69 |
| **LOS** | 0.20 | 0.00 | 79.15 | 0.00 | 0.20 | 0.21 |
| **LS** | -0.12 | 0.00 | -48.23 | 0.00 | -0.12 | -0.11 |
| **_cons** | 6.89 | 0.03 | 224.72 | 0.00 | 6.83 | 6.95 |
| | | | | | | |
| **ln_sig** | | | | | | |
| **LOS** | 0.04 | 0.00 | 41.25 | 0.00 | 0.04 | 0.04 |
| **_cons** | 0.44 | 0.01 | 51.42 | 0.00 | 0.42 | 0.45 |
| | | | | | | |
| **kappa** | | | | | | |
| **_cons** | 0.21 | 0.01 | 33.98 | 0.00 | 0.20 | 0.23 |
| | | | | | | |
| **Sell Orders** | | | | | | |
| **PA** | 4.91 | 0.04 | 124.14 | 0.00 | 4.84 | 4.99 |
| **BBOS** | 2.40 | 0.02 | 113.41 | 0.00 | 2.36 | 2.44 |
| **BSIDS** | 0.78 | 0.01 | 104.00 | 0.00 | 0.77 | 0.80 |
| **NL30MT** | -0.68 | 0.00 | -146.11 | 0.00 | -0.69 | -0.67 |
| **LOS** | 0.17 | 0.00 | 65.62 | 0.00 | 0.16 | 0.17 |
| **LS** | -0.10 | 0.00 | -39.75 | 0.00 | -0.10 | -0.09 |
| **_cons** | 6.92 | 0.03 | 229.64 | 0.00 | 6.86 | 6.98 |
| | | | | | | |
| **ln_sig** | | | | | | |
| **LOS** | 0.04 | 0.00 | 42.49 | 0.00 | 0.04 | 0.04 |
| **_cons** | 0.39 | 0.01 | 44.51 | 0.00 | 0.38 | 0.41 |
| | | | | | | |
| **kappa** | | | | | | |
| **_cons** | 0.23 | 0.01 | 37.47 | 0.00 | 0.22 | 0.25 |

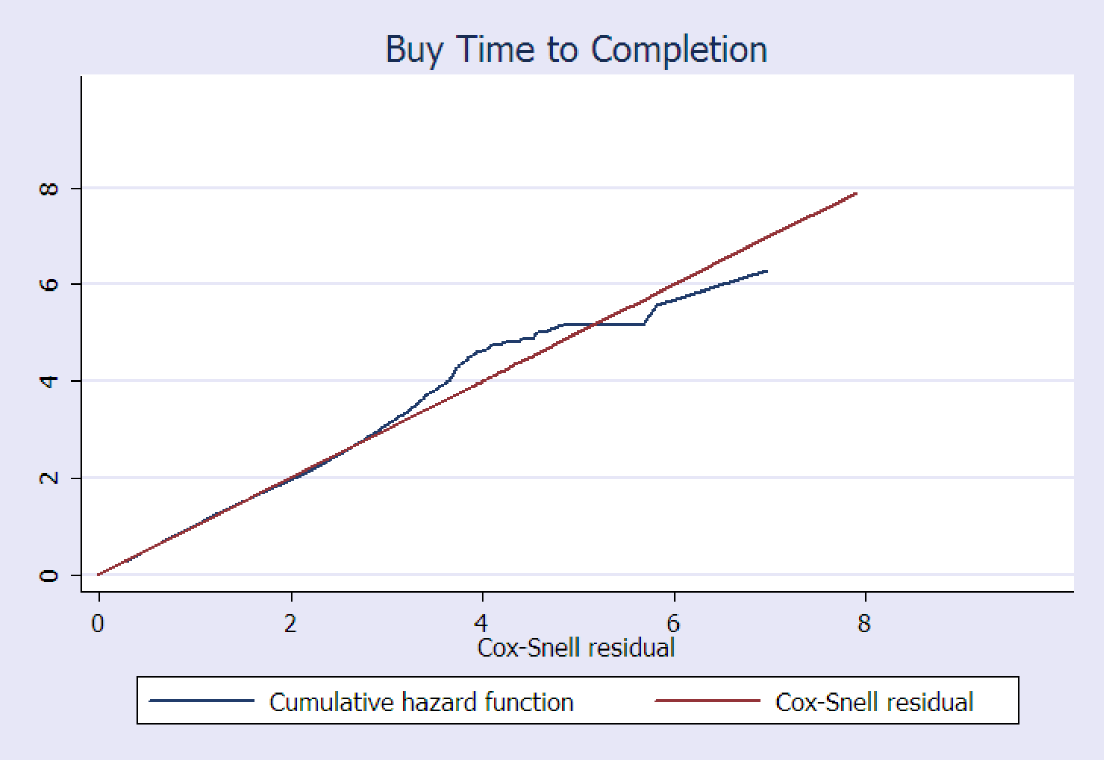
**Table 7.5: Prediction Errors of the Modified Dynamic GG AFT Model**

This table shows point prediction errors of completed limit orders of a pooled sample of 38 stocks throughout a sample period from October 2000 to December 2000. PMST is the abbreviation for the predicted median survival time, while AST is the abbreviation for the actual survival time. ‘No serious error’ refers to the percentage of limit orders whose predicted median survival times fall between half of actual survival times and twice actual survival times. ‘Serious error 1’ refers to the percentage of limit orders whose predicted median survival times are less than half of actual survival times. ‘Serious error 2’ refers to the percentage of limit orders whose predicted median survival times are more than twice actual survival times.

| | **Serious error 1** | **No serious error** | **Serious error 2** |
|---|---|---|---|
| | **PMST<0.5AST** | **PMST between 0.5AST and 2AST** | **PMST>2AST** |
| **Buy Orders** | 10.87% | 23.33% | 65.79% |
| **Sell Orders** | 11.29% | 23.48% | 65.23% |

**Figure 7.1: Q-Q Plots of the Dynamic GG AFT Model**

The following graph displays Q-Q plots of empirical estimates of the cumulative hazard functions of Cox-Snell residuals for limit orders of a pooled sample of 38 stocks, throughout a sample period from October 2000 to December 2000. In order to compare with figures presented in Chapter 6, the scale is kept the same.



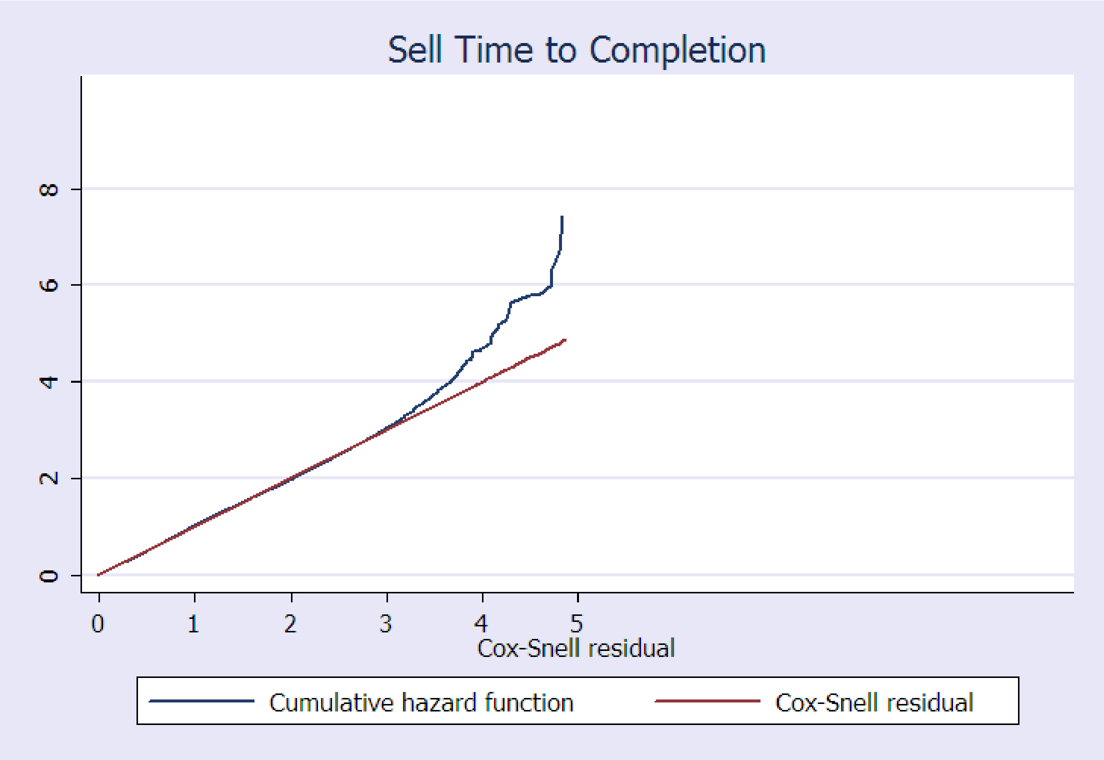
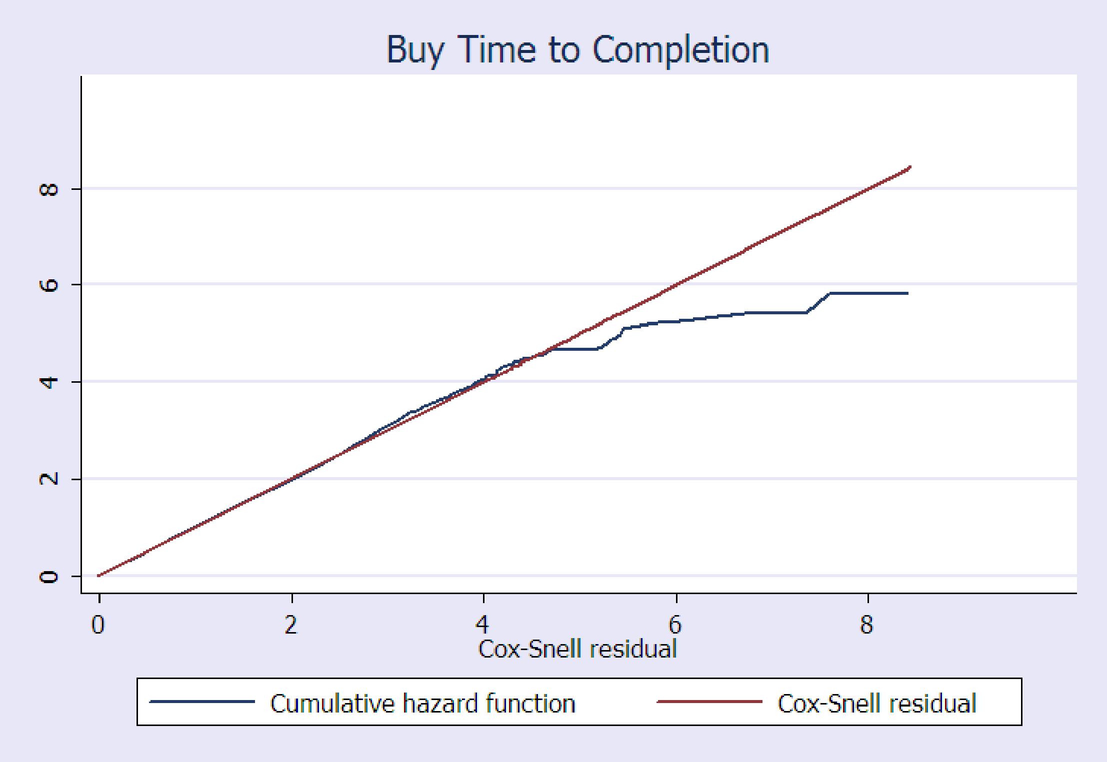



**Figure 7.2: Q-Q Plots of the Modified Dynamic GG AFT Model**

The following graph displays Q-Q plots of empirical estimates of the cumulative hazard functions of Cox-Snell residuals for limit orders of a pooled sample of 38 stocks, throughout a sample period from October 2000 to December 2000. In order to compare with figures presented in Chapter 6, the scale is kept the same.



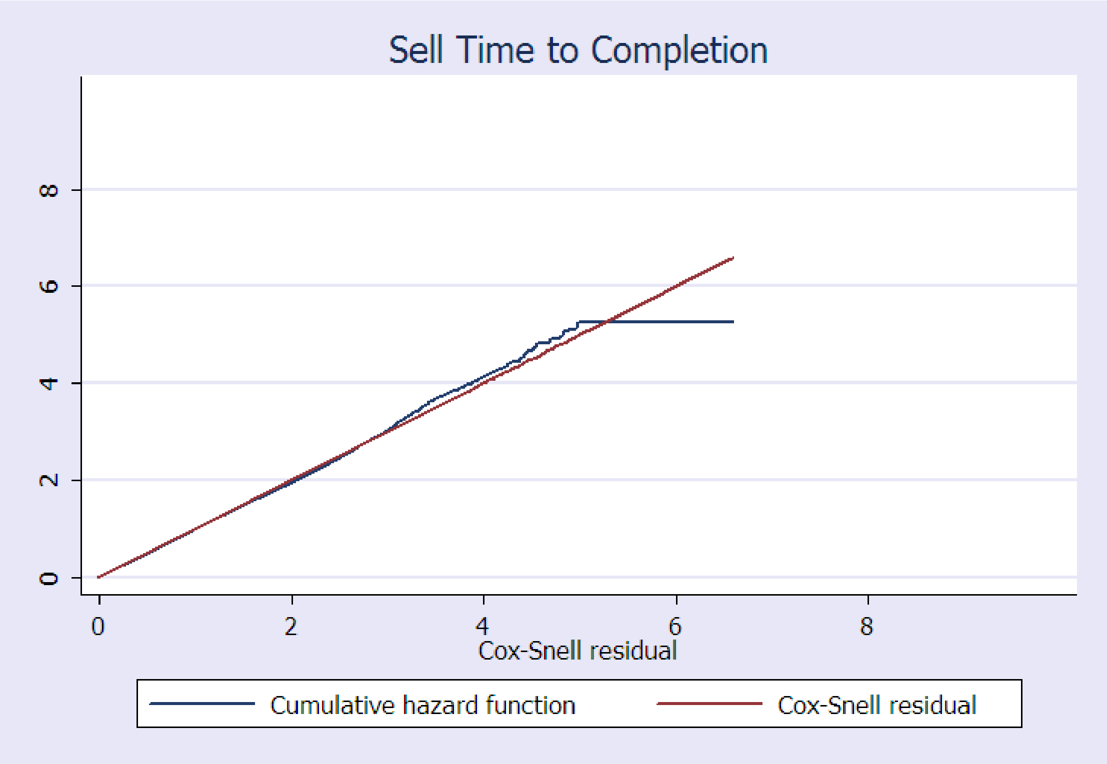
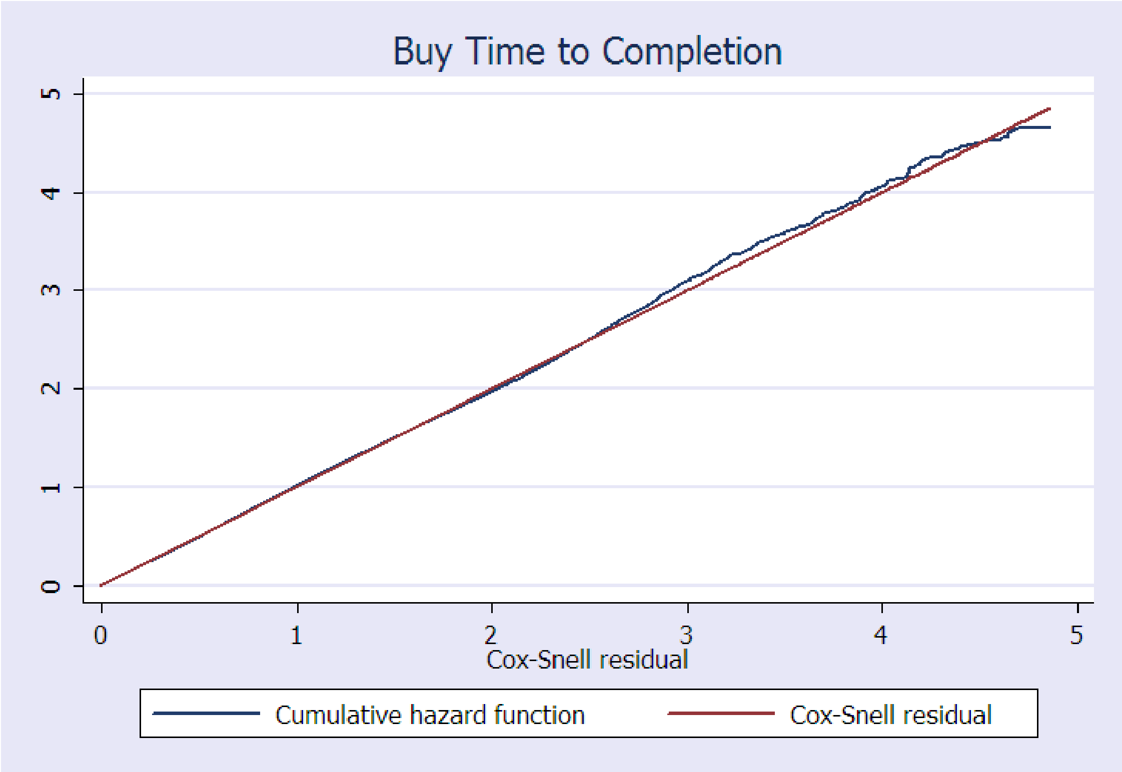
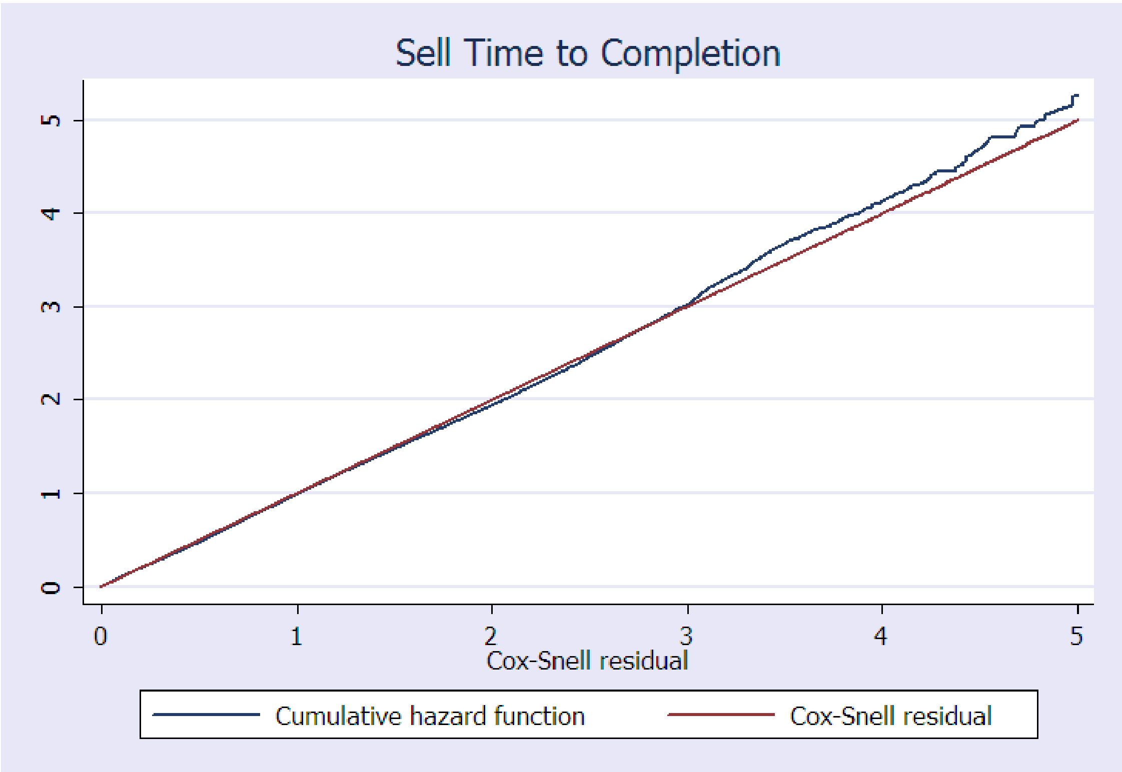



**Figure 7.3: Q-Q Plots of the Modified Dynamic GG AFT Model Excluding Observations with Cox-Snell Residuals Greater Than Five**

The following graph displays Q-Q plots of empirical estimates of the cumulative hazard functions of Cox-Snell residuals for limit orders of a pooled sample of 38 stocks, throughout a sample period from October 2000 to December 2000 excluding observations with Cox-Snell residuals greater than five.



# CHAPTER 8 – INTRA-DAILY SEASONALITY IN THE LIMIT-ORDER COMPLETION TIME IN THE UK MARKET

## 8.1 Introduction

Research on intra-day activities, which is carried out with transaction data from stock exchanges around the world, reveals the persistent u-shape patterns in returns, number of shares traded, volumes, bid-offer spreads, and volatility.[142] Intra-day patterns in order placements have also drawn the attention of research. Biais, Hillion and Spatt (1995) find that the placement of new orders in the Paris Stock Exchange tends to be concentrated in the morning while cancellations and large trades tend to occur late in the trading day. They also find that a larger proportion of small orders are executed at the opening, whereas larger proportions of large orders are executed during and at the end of the trading session. They put forward an explanation: small trades at the opening contribute to price discovery, while large trades tend to occur after price discovery has already occurred. Niemeyer and Sandas (1995) find a u-shape in the number of limit orders placed with a surprisingly high frequency of limit orders placed during the last minutes of the trading day using Stockholm Stock Exchange data. They find that 'patient' traders start to adjust their prices as the end of the session approaches in order to induce other traders to execute against them. Al-Suhaibani and Kryzanowsky (2000) find that the number of limit orders also exhibits a u-shape pattern during each continuous trading session in the Saudi stock market. They argue that the call market may be a contributing factor to the concentration at the opening and the high level of limit orders at the end of every trading session could result from limit-order price adjusting.[143] The concentration of order submissions around the opening and closing is observed in many stock markets (c.f., Jain and Joh (1988), McInish and Wood (1990, 1991), Gerety and Mulherin (1992) and Foster and Viswanathan (1993)). In general,

[142] Wood, McInish and Ord (1985), Jain and Joh (1988), McInish and Wood (1991, 1992), Handa (1992), Brock and Kleidon (1992), Gerety and Mulherin (1992), Foster and Viswanathan (1993), Chan, Christie and Schultz (1995) and Chan, Chung and Johnson (1995) find these u-shape patterns in the US markets. McInish and Wood (1990) and Lehmann and Modest (1994) report similar results for the Canadian stock market.

[143] Adjusting the limit-order price or quantity results in the order receiving a new time stamp. Accordingly, an order adjustment leads to two events: cancelling the existing order and submitting a new one.

existing literature shows that a large number of orders are placed and executed at the beginning and end of the trading day. The more orders are executed, the more active and volatile the market becomes and hence the higher the limit-order execution probability and the shorter the limit-order execution time. Thus limit-order submission time could potentially affect the execution probability and time.

Research on related time-of-day effects has also been carried out. Omura, Tanigawa, and Jun (2000) show that limit-order execution probabilities are higher for limit orders submitted during the earlier hours of the trading day. Gava (2006) uses survival analysis to investigate limit-order execution time in the Spanish stock market. She find that the orders placed during the first 30 minutes and the last hour of the trading day on the buy side of the market have shorter expected execution times and the orders placed during the last 90 minutes on the sell side also have shorter expected execution times.

This chapter presents an investigation on the time-of-day effects in the UK stock market using the modified dynamic GG AFT model discussed in Chapter 7. The remainder of the chapter is organised as follows. The next section presents the main part of the investigation on these time-of-day effects. Section 8.3 offers a summary and some conclusions.

## 8.2 Time-of-Day Patterns in Limit-Order Completion Times

In this section, the trading day is divided into three periods: morning, mid-day and afternoon.[144] The morning period is from 09:00 to 11:30, the mid-day period is from 11:30 to 14:00, and the afternoon period is from 14:00 to 16:30.

Table 8.1 presents the percentage breakdown of limit orders over the ‘average’ trading day. The table shows that, on average, 43.54% of buy orders and 42.99% of sell orders are submitted in the afternoon period, about 25% of orders are submitted in the mid-day period and about 30% of orders are submitted in the morning period. This table shows a

[144] It is believed that dividing a trading day into three periods is sufficient enough to reveal the intra-day patterns.

high frequency of limit-order placement in the morning and afternoon periods and a low frequency of limit-order placement in the mid-day period. This is consistent with the u-shape pattern revealed in the existing literature. A number of plausible justifications have been proposed: the opening of the US markets in the afternoon provides updated information, and could be the reason behind the high frequency of limit-order placement in the afternoon period. It could also be due to the adjustment of limit-order price, as argued by Al-Suhaibani and Kryzanowsky (2000). Another explanation could be that some traders are trying to meet their daily trading target before the market closes. Alternatively, it could also be due to the behaviour of informed traders.[145] The high frequency of order submission in the afternoon period is supportive of a model for explaining trade concentration.[146] The low frequency of limit-order placement in the mid-day period could also be due to the fact that traders are breaking for lunch, or awaiting the opening of the US market. Table 8.1 shows that more orders are placed in the afternoon than in the morning. This could be due to the fact that order submitted before 09:00am are excluded from the data sample used in this study, since a large number of orders are usually submitted just after the market opens as the literature discussed above shows.

Table 8.2 presents the summary statistics of the explanatory variables constructed in Chapter 6 over the 'average' trading day.[147] The table shows that the mean of PA decreases over the trading day. The following explanations can be proposed: first, traders price their orders more aggressively toward the end of the trading day. Second, informed traders submit aggressively priced limit orders toward the end of the trading day, as discussed above. Third, traders become 'impatient' toward the end of the trading day, since they need to meet their daily trading target before the market closes. Finally, 'patient' traders adjust their prices as the end of the trading day approaches, as argued by Niemeyer and Sandas (1995). The table shows that the mean of BBOS also decreases over the trading day. This could be due to overbidding and undercutting among traders, who price their orders more aggressively toward the end of the trading

[145] Bloomfield, O'Hara and Saar (2005) show that informed traders become suppliers of liquidity by submitting more limit orders towards the end of the trading day. This could be due to the fact that informed traders acquire updated information throughout the trading day and take advantage of this information by submitting aggressively priced limit orders toward the end of the trading day.

[146] Admati and Pfleiderer (1988) present a model where concentration of trading may be generated at a time of the day and explain that traders, who have to trade within a given time period, pool their trades in an effort to reduce their transaction costs.

[147] The summary statistics are computed on the multiple-record data presented in Chapter 7, since these data recorded updated market conditions.

day as just discussed, or the latent demand for trade being converted into order placement toward the end of the trading day. This finding is not consistent with the u-shape pattern of bid-offer spread revealed in the literature. This could be due to the fact that limit orders submitted in the opening hour are excluded from the data sample. The table shows that the mean of BSIDS exhibits a u-shape pattern for buy orders but not for sell orders. Perhaps a slight increase in the submission of sell market orders may occur during the mid-day period, since there is a small increase in the sell limit orders submitted in this period as Table 8.1 reports (assume that the arrival rate of sell limit orders is a proxy of the arrival rate of sell market orders). The table shows that the mean of NL30MT exhibits a u-shape pattern as posited in the literature. The u-shape pattern of limit order submission reported in Table 8.1 could be the reason behind the intra-day pattern of NL30MT. This finding confirms the correlation between order submission and trading activities. The table shows that the mean of LOS increases over the trading day for sell orders, but not for buy orders. This could be due to greater urgency toward the end of the trading day when traders are selling. The table shows that the mean of LS increases over the trading day. This indicates that more liquidity is available as the trading day progresses. This could be partly due to the fact that more limit orders are submitted in the afternoon period.

In order to identify time-of-day effects on limit-order completion time, two dummy variables are incorporated into the modified dynamic GG AFT model discussed in Chapter 7. These are:

TD2 = 1 when limit orders submitted in the mid-day period, otherwise 0, and (8.1)

TD3 = 1 when limit orders submitted in the afternoon period, otherwise 0. (8.2)

Table 8.3 presents the estimation results. The estimated parameters are statistically significant in both buy and sell models. These coefficients reflect the sensitivities of limit-order completion time to these explanatory variables. All coefficient estimates on the explanatory variables have the same signs as those of the modified dynamic GG AFT model reported in Table 7.4 and discussed in Section 7.3. Hence, these estimates have the same economic interpretations as those discussed in Section 7.3.

The positive estimates of the coefficients on TD2 indicate that the limit orders submitted in the mid-day period are expected to have longer completion times than those submitted in the morning period. The negative estimates of the coefficients on TD3 indicate that the limit order submitted in the afternoon period are expected to have shorter completion times than those submitted in the morning period. These estimates are as expected given the pattern of order submission reported in Table 8.1.

The high frequency of limit-order placement, as reported in Table 8.1, in the morning and afternoon periods indicates the willingness of traders to trade. The high volatilities, as reported in Table 8.2, during these two periods indicate the active nature of the market. Hence, limit orders submitted during these two periods should have higher completion probabilities and, consequently, shorter time-to-completion. In contrast, trading activity is relatively low in the mid-day period. Thus, limit orders submitted during this period should have lower completion probabilities and, consequently, longer time-to-completion. Thus, the estimates reported in Table 8.3 confirm the correlation between concentration of order submission and completion times.

The estimates presented above indicate that orders submitted in the afternoon period are expected to have the shortest completion times. This could be due to the 'deadline effect' posited in the economics literature.[148]

This finding suggests that traders should place their orders in the afternoon to shorten the expected limit-order completion times. If most traders place their orders in the morning and afternoon, trading could be arranged in two separate trading sessions (one in the morning and one in the afternoon) rather than a single continuous trading session through the whole day. In this case, it will be easier for buyers and sellers to find each other and hence orders can be executed quickly and easily.

The sensitivities of limit-order completion time to the explanatory variables over the three different periods are also investigated here. The morning, mid-day and afternoon

[148] Roth, Murnigham and Schoumaker (1988) argue that agreements are more likely to be reached at the last minute. The deadline effect refers to the high percentage of agreements reached just prior to the deadline.

models refer to those estimated with the dataset that includes limit orders submitted during each corresponding period only. The empirical results are reported in Table 8.4. All coefficient estimates on the explanatory variables have the same signs as those of the modified dynamic GG AFT model reported in Table 7.4 and discussed in Section 7.3. Hence, these estimates have the same economic interpretations as those discussed in Section 7.3.

*The Coefficients of PA*

PA measures the relative distance between the limit-order price and the contemporary best bid (offer) price on the order book for buy (sell) orders. It captures the position of a limit order in the current market condition. The estimates of the coefficients on PA are positive for all three periods. This positive sign could be a mechanical consequence of the price priority in an order-driven market discussed in Chapter 2.

The positive sign indicates that aggressively priced orders induce latent demand for trade as discussed in Subsection 6.3.1. These estimates imply that traders need to price their orders as close to the best bid/offer price as possible in order to reduce the expected waiting time of order completion and hence reduce the associated opportunity cost.

The coefficient of PA is 4.52 for buy orders and 4.60 for sell orders in the morning period, 4.97 for buy orders and 4.77 for sell orders in the mid-day period, and 6.66 for buy orders and 6.29 for sell orders in the afternoon period. According to the AFT model discussed in Subsection 4.2.5, if a buy order is priced 1% of the mid-quote price below the best bid price (PA value increases by one unit holding the remaining variables constant) then the expected completion time will be extended by 92 ($e^{4.52}$) times in the morning period, 144 ($e^{4.97}$) times in the mid-day period, and 780 ($e^{6.66}$) times in the afternoon period. According to the same model if a sell order is priced 1% of the mid-quote price above the best offer price, the expected completion time will be extended by 99 ($e^{4.60}$) times in the morning period, 118 ($e^{4.77}$) times in the mid-day period, and 539 ($e^{6.29}$) times in the afternoon period.

The sensitivity of limit-order completion time to this variable increases through the day. One possible explanation could be the strategic behaviour of day traders. Usually day traders analyse information and monitor the market in the morning, and then are more willing to trade later in the trading day in order to meet their trading target. Thus, aggressively priced orders could induce more latent demand for trade from these day traders over the trading day. In other words, aggressively priced orders induce traders on the opposite side of the market submit more market orders in the afternoon. Hence, limit-order completion time becomes more sensitive to this variable as the trading day progresses. This finding suggests that traders should place limit orders more aggressively in the afternoon period to take advantage of this higher sensitivity. This finding also indicates that exchanges should encourage traders to submit aggressively priced orders in the afternoon. This finding shows that aggressively priced orders are more effective in the afternoon. Hence traders should take this point into consideration when they develop their order submission strategies.

The trend these estimates exhibit could be due to the result of excluding orders submitted before 09:00am from the data sample used in this study. Since the market is usually very active in the first hour of trading, aggressively priced orders could be completed easily during this period of time. However, these aggressively priced orders could also be ‘picked up’ by informed traders as these informed traders might take advantage of their private information and start to trade as soon as the market opens.

*The Coefficients of BBOS*

BBOS represents the relative best bid-offer spread. The estimates of the coefficients on BBOS are positive for all three periods. The positive sign indicates that the wider the best bid-offer spread, the longer the expected completion time. This positive sign could be caused by the potential effect of narrow spreads inducing latent demand for trade as discussed in Subsection 6.3.1. These estimates suggest that traders should submit limit orders when the market spread becomes narrower and exchanges should minimise the market spread in order to facilitate the order execution.

The coefficient of BBOS is 2.31 for buy orders and 2.37 for sell orders in the morning period, 2.61 for buy orders and 2.79 for sell orders in the mid-day period, and 2.34 for buy orders and 2.39 for sell orders in the afternoon period. According to the AFT model discussed in Subsection 4.2.5, if the best bid-offer spread is widened by 1% of the mid-quote price (BBOS value increases by one unit holding the remaining variables constant) the expected completion time of buy orders will be extended by 10.07 ($e^{2.31}$) times in the morning period, 13.60 ($e^{2.61}$) times in the mid-day period, and 10.38 ($e^{2.34}$) times in the afternoon period. That of sell orders will be extended by 10.69 ($e^{2.37}$) times in the morning period, 16.28 ($e^{2.79}$) times in the mid-day period, and 10.91 ($e^{2.39}$) times in the afternoon period.

The estimates presented above exhibit an n-shape pattern. Thus, limit-order completion time is most sensitive to this variable in the mid-day period. As Tables 8.1 and 8.2 report, the frequency of limit-order placement and the volatility of the market exhibit u-shape patterns. Hence, the lack of market activities in the mid-day period, which induce less demand for trade, could be the reason behind the n-shape pattern of the estimates above. The n-shape pattern could also be due to the fact that narrower best bid-offer spreads induce more latent demand for trade from the opposite side of the market in the mid-day period. In other words, narrower best bid-offer spreads encourage traders from the opposite side of the market to submit more market orders in the mid-day period than in the other two periods. In addition, the market spread is usually wider in the mid-day period. Once the market spread becomes narrower in this period, some traders would take this 'rare' opportunity and submit market orders to execute against limit orders on the opposite side of the market. Hence limit-order completion time is more sensitive to this variable in the mid-day period.

This finding suggests that traders should monitor the best bid-offer spread more closely in the mid-day period and submit limit orders when the spread becomes narrower. This finding also indicates that limit orders are more effective in the mid-day period when the spread is narrower. This finding suggests that exchanges should minimise the market spread, especially in the mid-day period, since this could reduce the waiting time of order completion and accordingly facilitate limit-order execution.

These estimates show that limit-order completion time is less sensitive to BBOS in the morning than in the afternoon. This could be due to the fact that informed traders might use market orders to take advantage of their private information in the morning.

Excluding orders submitted before 09:00am from the data sample could also affect the estimates discussed above, since in the first hour of trading informed traders are active and these traders could use market orders without concerning about the market spread. These traders would use market orders to take advantage of their private information before this information becomes public.

*The Coefficients of BSIDS*

BSIDS is an indicator of whether or not the prior trade was same-side initiated (BSIDS equals to one). The estimates of the coefficients on BSIDS are positive and have similar magnitudes for all three periods. These estimates indicate that if the prior trade was initiated from the same side of the order book (BSIDS value is one holding the remaining variables constant), the expected limit-order completion time will be extended. These estimates indicate that activities on the opposite side of the market would affect the completion times of existing orders.

The coefficient of BSIDS is 0.77 for buy orders and 0.79 for sell orders in the morning period, 0.73 for buy orders and 0.78 for sell orders in the mid-day period, and 0.74 for buy orders and 0.73 for sell orders in the afternoon period. According to the AFT model discussed in Subsection 4.2.5, if the prior trade was initiated from the same side of the order book (BSIDS value is one holding the remaining variables constant) the expected completion time of buy orders will be extended by 2.16 ($e^{0.77}$) times in the morning period, 2.08 ($e^{0.73}$) times in the mid-day period, and 2.10 ($e^{0.74}$) times in the afternoon period. That of sell orders will be extended by 2.20 ($e^{0.79}$) times in the morning period, 2.18 ($e^{0.78}$) times in the mid-day period, and 2.08 ($e^{0.73}$) times in the afternoon period.

Generally the estimates presented above change only slightly over the trading day. The following justifications can be posited: first, the sensitivity of limit-order completion time to this variable is low as reported in Table 8.3. Second, this variable only approximates whether or not the prior trade was same-side initiated, given that the side initiating the prior trade is not discernable from the data sample used in this study. Finally, since the market is a 'sell' market for buy orders and a 'buy' market for sell orders when BSIDS equals zero as discussed in Subsection 6.3.1, this could be due to the constant effect of 'sell' markets improving buy limit-order completion and 'buy' markets improving sell limit-order completion over the trading day.[149] This finding indicates that traders monitor the market closely and continuously, and submit buy market orders when the market is already a 'buy' market and sell market orders when the market is already a 'sell' market throughout the trading day. This confirms that traders develop their order submission strategies basing on the order-book information. Excluding orders submitted before 09:00am from the data sample could also affect the estimates discussed above, since informed traders would not just follow the market and instead they would choose their orders basing on their private information.

*The Coefficients of NL30MT*

NL30MT is a trading activity measure that provides an absolute measure of volatility. The estimates of the coefficients on NL30MT are negative for all three periods. The negative sign indicates that the larger the number of trades in the preceding 30 minutes, the more active and volatile the market becomes and hence the shorter the expected completion time. This negative sign is expected as discussed in Chapter 6. These estimates suggest that traders should submit limit orders when market volatility increases as discussed in Chapter 6. These estimates also suggest that exchanges should reduce information-driven volatility by disclosing order-book information as discussed in Chapter 6.

The coefficient of NL30MT is -0.58 for buy orders and -0.50 for sell orders in the morning period, -0.62 for buy orders and -0.63 for sell orders in the mid-day period, and -0.73 for buy orders and -0.69 for sell orders in the afternoon period. According to the AFT model discussed in Subsection 4.2.5, if NL30MT value decreases by one unit holding the remaining variables constant, the expected completion time of buy orders will be extended by 1.79 ($e^{0.58}$) times in the morning period, 1.86 ($e^{0.62}$) times in the

[149] The 'buy'/'sell' market results from the arrival of buy/sell orders, which is triggered by economic events.

mid-day period, and 2.08 ($e^{0.73}$) times in the afternoon period. That of sell orders will be extended by 1.65 ($e^{0.50}$) times in the morning period, 1.88 ($e^{0.63}$) times in the mid-day period, and 1.99 ($e^{0.69}$) times in the afternoon period.

The estimates presented above decrease over the trading day. Hence, limit-order completion time becomes more sensitive to this variable as the trading day progresses. This could be due to the fact that volatility induces more demand for trade from the opposite side of the market over the trading day. Namely volatility encourages traders from the opposite side of the market to submit more market orders as the trading day progresses. Perhaps traders become 'impatient' over the trading day as discussed above. This could also be due to the fact that volatility becomes less information-driven through the day as private information can gradually become public and traders prefer mechanic-driven volatility to information-driven volatility.

Since the market is the most active in the afternoon as Table 8.2 reports and the coefficient of this variable is lowest (negative) in this period, the limit-order completion time is expected to be shortest as discussed above. This finding suggests that traders should submit limit orders in the afternoon.

Excluding orders submitted before 09:00am from the data sample could affect the estimates discussed above. Since informed traders would start to trade immediately after the market opens, the volatility in the first hour of trading is more likely to be information-driven. Hence traders would avoid submitting limit orders in this market, as these order are more likely to be 'picked up' by informed traders. Then limit-order completion time could be more sensitive to NL30MT during this period of time.

*The Coefficients of LOS*

LOS captures the limit-order size. The estimates of the coefficients on LOS are positive for all three periods. The positive sign indicates that the larger a limit order, the longer the expected completion time. Compared to PA and BBOS limit-order completion time is not as sensitive to this variable. These estimates indicate that splitting a large limit order into a number of small orders would not reduce order completion time significantly. Hence this may not benefit traders, since the cost of managing these small orders could be much higher than the benefit as discussed in Chapter 6.

The coefficient of LOS is 0.23 for buy orders and 0.17 for sell orders in the morning period, 0.21 for buy orders and 0.19 for sell orders in the mid-day period, and 0.17 for buy orders and 0.14 for sell orders in the afternoon period. According to the AFT model discussed in Subsection 4.2.5, if LOS value increases by one unit holding the remaining variables constant, the expected completion time of buy orders will be extended by 1.26 ($e^{0.23}$) times in the morning period, 1.23 ($e^{0.21}$) times in the mid-day period, and 1.19 ($e^{0.17}$) times in the afternoon period. That of sell orders will be extended by 1.19 ($e^{0.17}$) times in the morning period, 1.21 ($e^{0.19}$) times in the mid-day period, and 1.15 ($e^{0.14}$) times in the afternoon period.

The estimates presented above decrease slightly over the trading day. Hence, limit-order completion time is less sensitive to this variable in the afternoon period than in the other two periods. Perhaps larger orders induce more demand for trade from the opposite side of the market as the trading day progresses. This could also be due to the fact that traders are trying to meet their daily trading targets before the market closes as discussed above. This finding suggests that traders should place larger orders in the afternoon period to take advantage of this lower sensitivity. The summary statistics for LOS reported in Table 8.2, however, show that only sellers place larger orders in the afternoon period. This indicates that buyers do not take the opportunity to place larger orders in this period. This could be due to the fact that buyers are more patient than sellers. This finding suggests that traders should pool a number of small orders into a large order in the afternoon. This finding also suggests that exchanges should encourage large-order submission in the afternoon.

Since traders usually submit small orders in the early hours of trading as Al-Suhaibani and Kryzanowsky (2000) reveal, limit-order completion time could be more sensitive to the order size during this period of time. Hence the exclusion of orders submitted in the first hour of trading from the data sample could affect the estimates discussed above.

*The Coefficients of LS*

LS captures the depth (liquidity) of the opposite side of an order book. It captures the best offer size for buy orders and the best bid size for sell orders. The estimates of the coefficients on LS are negative and have similar magnitudes for all three periods. The negative sign indicates that the larger the best offer (bid) size, the shorter the expected buy-order (sell-order) completion time. This is expected, as the more liquidity is available from the opposite side of the market, the more easily a limit order will be completed. Compared to PA and BBOS limit-order completion time is also not as sensitive to this variable. These estimates imply that traders should submit limit orders when more liquidity is available in order to reduce the waiting time and exchanges should disclose the full depth of the order book as discussed in Chapter 6.

The coefficient of LS is -0.14 for buy orders and -0.12 for sell orders in the morning period, -0.12 for buy orders and -0.11 for sell orders in the mid-day period, and -0.11 for buy orders and -0.09 for sell orders in the afternoon period. According to the AFT model discussed in Subsection 4.2.5, if LS value decreases by one unit holding the remaining variables constant, the expected completion time of buy orders will be extended by 1.15 ($e^{0.14}$) times in the morning period, 1.13 ($e^{0.12}$) times in the mid-day period, and 1.12 ($e^{0.11}$) times in the afternoon period. That of sell orders will be extended by 1.13 ($e^{0.12}$) times in the morning period, 1.12 ($e^{0.11}$) times in the mid-day period, and 1.09 ($e^{0.09}$) times in the afternoon period.

The estimates presented above increase only slightly over the trading day. Hence, limit-order completion time is slightly less sensitive to this variable in the afternoon period than in the other two periods. Perhaps liquidity available from the opposite side of the market induces less demand for trade as the trading day progresses. Namely traders become reluctant to submit market orders to consume the liquidity available from the opposite side of the market. This could be due to the behaviour of informed traders as discussed above. The trend these estimates exhibit could also be due to the fact that traders become less concerned about the liquidity available from the opposite side of the market as the trading day progresses. Perhaps traders just become ‘impatient’ over the trading day or it is more difficult for traders to monitor the order book, as the market is more active in the afternoon.

Since informed traders would trade immediately after the market opens, they will be less concerned about the liquidity available from the opposite side of the market and hence limit-order completion time will be less sensitive to this variable. For this reason the exclusion of orders submitted in the first hour of trading from the data sample could affect the estimates discussed above.

Generally these empirical results show that the sensitivities of limit-order completion time to the explanatory variables vary over the trading day. Sensitivities to some variables, such as NL30MT, increase over the trading day, whilst sensitivities to other variables, such as LOS, decrease over the trading day. These indicate that traders behave differently over the trading day as Tables 8.1 and 8.2 report.

Table 8.5 presents the test statistics of the differences among the estimated coefficients reported in Table 8.4. The differences among the estimated coefficients of PA, NL30MT, LOS and LS are statistical significant. In the buy model, the difference between the estimated coefficient of BBOS of the morning period and that of the afternoon period is insignificant, and the difference between the estimated coefficient of BSIDS of the mid-day period and that of the afternoon period is also insignificant. In the sell model, the difference between the estimated coefficient of BBOS of the morning period and that of the afternoon period is insignificant, and the difference between the estimated coefficient of BSIDS of the mid-day period and that of the morning period is also insignificant.

## 8.3 Summary and Conclusions

In this chapter, a trading day is divided into three periods of time: morning, mid-day and afternoon. Time-of-day effects on limit-order completion time are investigated with the modified dynamic GG AFT model discussed in Chapter 7.

Table 8.1 presents the percentage breakdown of limit orders over a trading day. It shows that, on average, 43.54% of buy orders and 42.99% of sell orders are submitted in the afternoon period, about 25% of orders are submitted in the mid-day period and about 30% of orders are submitted in the morning period. Table 8.2 presents the summary statistics of the explanatory variables constructed in Chapter 6 over the ‘average’ trading day. This table shows that the mean of the explanatory variables exhibits intra-day patterns.

In order to identify time-of-day effects on limit-order completion time, two dummy variables are incorporated into the modified dynamic GG AFT model discussed in Chapter 7. Table 8.3 presents the estimation results. The estimated parameters are statistically significant in both buy and sell models. The positive estimates of the coefficients on TD2 indicate that the limit orders submitted in the mid-day period are expected to have longer completion times than those submitted in the morning period. The negative estimates of the coefficients on TD3 indicate that the limit orders submitted in the afternoon period are expected to have shorter completion times than those submitted in the morning period.

The sensitivities of limit-order completion time to the explanatory variables over the three different periods are also investigated in this chapter. The empirical results are reported in Table 8.4. Table 8.5 shows that in general the differences among the estimated coefficients reported in Table 8.4 are statistically significant. These results show that the sensitivities of limit-order completion time to the explanatory variables vary over the trading day. The next chapter provides general conclusions, a discussion of strengths and limitations of this study, and suggestions for future research.

**Table 8.1: Percentage Breakdown of Limit Orders over a Trading Day**

This table shows the percentage breakdown of the total number of limit orders for a pooled sample of 38 stocks (Pool). The sample period of the data is from October 2000 to December 2000. 'Morning' category includes limit orders submitted between 09:00 to 11:30. 'Mid-day' category includes limit orders submitted between 11:30 and 14:00. 'Afternoon' category includes limit orders submitted between 14:00 to 16:30.

| | **Buy Orders** | | | | **Sell Orders** | | |
|---|---|---|---|---|---|---|---|
| **Stock** | **Morning** | **Mid-day** | **Afternoon** | | **Morning** | **Mid-day** | **Afternoon** |
| **Pool** | 30.48% | 25.97% | 43.54% | | 30.49% | 26.52% | 42.99% |

**Table 8.2: Summary Statistics of the Explanatory Variables (Three Intra-Daily Periods)**

This table shows summary statistics of the explanatory variables for limit orders of a pooled sample of 38 stocks, throughout a sample period from October 2000 to December 2000. 'Morning' category includes limit orders submitted between 09:00 to 11:30. 'Mid-day' category includes limit orders submitted between 11:30 and 14:00. 'Afternoon' category includes limit orders submitted between 14:00 to 16:30. The definitions of the explanatory variables are given as follows. PA measures the relative distance between the limit-order price and the contemporary best bid/offer price on the order book. BBOS represents the relative best bid-offer spread. BSIDS is an indicator of whether or not the prior trade was same-side initiated. NL30MT is a trading activity measure that provides an absolute measure of volatility. LOS captures the limit-order size. LS captures the depth of the opposite side of the order book. T refers to trades and S refers to shares.

| | | Morning | | Mid-day | | Afternoon | |
|---|---|---|---|---|---|---|---|
| **Variable** | **Unites** | **Mean** | **Std. Dev.** | **Mean** | **Std. Dev.** | **Mean** | **Std. Dev.** |
| | | | | | | | |
| **Buy Orders** | | | | | | | |
| **PA** | % | 0.26 | 0.56 | 0.20 | 0.46 | 0.14 | 0.37 |
| **BBOS** | % | 0.37 | 0.37 | 0.32 | 0.31 | 0.30 | 0.29 |
| **BSIDS** | N/A | 0.38 | 0.49 | 0.37 | 0.48 | 0.38 | 0.48 |
| **NL30MT** | LogT | 3.04 | 0.93 | 2.99 | 0.86 | 3.54 | 0.82 |
| **LOS** | LogS | 8.96 | 1.47 | 8.95 | 1.46 | 8.95 | 1.45 |
| **LS** | LogS | 9.47 | 1.55 | 9.52 | 1.53 | 9.55 | 1.52 |
| | | | | | | | |
| **Sell Orders** | | | | | | | |
| **PA** | % | 0.24 | 0.56 | 0.18 | 0.44 | 0.13 | 0.35 |
| **BBOS** | % | 0.37 | 0.36 | 0.32 | 0.30 | 0.30 | 0.29 |
| **BSIDS** | N/A | 0.41 | 0.49 | 0.41 | 0.49 | 0.39 | 0.49 |
| **NL30MT** | LogT | 3.03 | 0.92 | 2.98 | 0.87 | 3.53 | 0.83 |
| **LOS** | LogS | 8.93 | 1.52 | 8.94 | 1.44 | 8.96 | 1.42 |
| **LS** | LogS | 9.44 | 1.53 | 9.51 | 1.55 | 9.53 | 1.55 |

**Table 8.3: Parameter Estimates of the Modified Dynamic GG AFT Model (Time-of-Day)**

The table below shows parameter estimates of the accelerated failure time specification of limit-order completion under the generalised gamma distribution for limit orders of a pooled sample of 38 stocks, throughout a sample period from October 2000 to December 2000. The variable '_cons' denotes the intercept and the definitions of the remaining explanatory variables are given as follows. PA measures the relative distance between the limit-order price and the contemporary best bid/offer price on the order book. BBOS represents the relative best bid-offer spread. BSIDS is an indicator of whether or not the prior trade was same-side initiated. NL30MT is a trading activity measure that provides an absolute measure of volatility. LOS captures the limit-order size. LS captures the depth of the opposite side of the order book. TD2 and TD3 are dummy variables that capture time-of-day effects. Z-statistics are asymptotically standard normal under the null hypothesis that the corresponding coefficient is zero. The explanatory variables, except LOS, are measured hourly after submission.

| Variable | Coef. | Robust Std. Err. | z | P>z | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| **Buy Orders** | | | | | | |
| **PA** | 5.27 | 0.04 | 121.71 | 0.00 | 5.18 | 5.35 |
| **BBOS** | 2.38 | 0.02 | 109.15 | 0.00 | 2.34 | 2.42 |
| **BSIDS** | 0.74 | 0.01 | 99.45 | 0.00 | 0.73 | 0.76 |
| **NL30MT** | -0.65 | 0.00 | -135.59 | 0.00 | -0.66 | -0.64 |
| **LOS** | 0.20 | 0.00 | 77.93 | 0.00 | 0.19 | 0.20 |
| **LS** | -0.12 | 0.00 | -50.40 | 0.00 | -0.13 | -0.12 |
| **TD2** | 0.07 | 0.01 | 6.99 | 0.00 | 0.05 | 0.08 |
| **TD3** | -0.24 | 0.01 | -28.50 | 0.00 | -0.26 | -0.23 |
| **_cons** | 6.91 | 0.03 | 222.23 | 0.00 | 6.85 | 6.97 |
| | | | | | | |
| **ln_sig** | | | | | | |
| **LOS** | 0.04 | 0.00 | 41.48 | 0.00 | 0.04 | 0.04 |
| **_cons** | 0.42 | 0.01 | 50.02 | 0.00 | 0.41 | 0.44 |
| | | | | | | |
| **kappa** | | | | | | |
| **_cons** | 0.24 | 0.01 | 37.80 | 0.00 | 0.23 | 0.25 |
| | | | | | | |
| **Sell Orders** | | | | | | |
| **PA** | 4.61 | 0.04 | 127.94 | 0.00 | 4.54 | 4.68 |
| **BBOS** | 2.40 | 0.02 | 114.31 | 0.00 | 2.36 | 2.44 |
| **BSIDS** | 0.78 | 0.01 | 104.30 | 0.00 | 0.76 | 0.79 |
| **NL30MT** | -0.61 | 0.00 | -129.09 | 0.00 | -0.62 | -0.61 |
| **LOS** | 0.16 | 0.00 | 65.30 | 0.00 | 0.16 | 0.17 |
| **LS** | -0.11 | 0.00 | -43.45 | 0.00 | -0.11 | -0.10 |
| **TD2** | 0.23 | 0.01 | 24.14 | 0.00 | 0.21 | 0.25 |
| **TD3** | -0.23 | 0.01 | -27.25 | 0.00 | -0.25 | -0.21 |
| **_cons** | 6.85 | 0.03 | 226.82 | 0.00 | 6.79 | 6.91 |
| | | | | | | |
| **ln_sig** | | | | | | |
| **LOS** | 0.04 | 0.00 | 42.68 | 0.00 | 0.04 | 0.04 |
| **_cons** | 0.38 | 0.01 | 42.51 | 0.00 | 0.36 | 0.39 |
| | | | | | | |
| **kappa** | | | | | | |
| **_cons** | 0.24 | 0.01 | 36.74 | 0.00 | 0.23 | 0.26 |

**Table 8.4: Parameter Estimates of the Modified Dynamic GG AFT Model (Three Intra-Daily Periods)**

The table below shows parameter estimates of the accelerated failure time specification of limit-order completion under the generalised gamma distribution for limit orders of a pooled sample of 38 stocks, throughout a sample period from October 2000 to December 2000 across a trading day. 'Morning' category includes orders submitted between 09:00 to 11:30. 'Mid-day' category includes orders submitted between 11:30 and 14:00. 'Afternoon' category includes orders submitted between 14:00 to 16:30. The variable '_cons' denotes the intercept and the definitions of the remaining explanatory variables are given as follows. PA measures the relative distance between the limit-order price and the contemporary best bid/offer price on the order book. BBOS represents the relative best bid-offer spread. BSIDS is an indicator of whether or not the prior trade was same-side initiated. NL30MT is a trading activity measure that provides an absolute measure of volatility. LOS captures the limit-order size. LS captures the depth of the opposite side of the order book. Z-statistics are asymptotically standard normal under the null hypothesis that the corresponding coefficient is zero. The explanatory variables, except LOS, are measured hourly after submission.

| | Morning | | | Mid-day | | | Afternoon | | |
|---|---|---|---|---|---|---|---|---|---|
| **Variable** | **Coef.** | **Robust Std. Err.** | **z** | **Coef.** | **Robust Std. Err.** | **z** | **Coef.** | **Robust Std. Err.** | **z** |
| | | | | | | | | | |
| **Buy Orders** | | | | | | | | | |
| **PA** | 4.52 | 0.06 | 79.20 | 4.97 | 0.08 | 64.21 | 6.66 | 0.10 | 66.02 |
| **BBOS** | 2.31 | 0.04 | 59.82 | 2.61 | 0.04 | 64.61 | 2.34 | 0.04 | 66.22 |
| **BSIDS** | 0.77 | 0.01 | 52.37 | 0.73 | 0.02 | 48.56 | 0.74 | 0.01 | 69.26 |
| **NL30MT** | -0.58 | 0.01 | -64.57 | -0.62 | 0.01 | -65.49 | -0.73 | 0.01 | -103.33 |
| **LOS** | 0.23 | 0.00 | 46.31 | 0.21 | 0.01 | 41.78 | 0.17 | 0.00 | 47.68 |
| **LS** | -0.14 | 0.00 | -29.53 | -0.12 | 0.00 | -25.28 | -0.11 | 0.00 | -31.54 |
| **_cons** | 6.67 | 0.06 | 112.30 | 6.84 | 0.06 | 114.64 | 6.94 | 0.04 | 156.56 |
| | | | | | | | | | |
| **ln_sig** | | | | | | | | | |
| **LOS** | 0.04 | 0.00 | 23.96 | 0.04 | 0.00 | 24.26 | 0.03 | 0.00 | 24.87 |
| **_cons** | 0.47 | 0.02 | 29.76 | 0.38 | 0.02 | 23.46 | 0.41 | 0.01 | 31.82 |
| | | | | | | | | | |
| **kappa** | | | | | | | | | |
| **_cons** | 0.22 | 0.01 | 18.17 | 0.33 | 0.01 | 28.79 | 0.19 | 0.01 | 18.35 |

**Table 8.4 continued …**

| Sell Orders | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **PA** | 4.60 | 0.06 | 77.98 | 4.77 | 0.08 | 59.16 | 6.29 | 0.09 | 66.99 |
| **BBOS** | 2.37 | 0.04 | 65.43 | 2.79 | 0.05 | 60.80 | 2.39 | 0.03 | 70.65 |
| **BSIDS** | 0.79 | 0.01 | 54.58 | 0.78 | 0.02 | 50.00 | 0.73 | 0.01 | 69.33 |
| **NL30MT** | -0.50 | 0.01 | -56.07 | -0.63 | 0.01 | -63.09 | -0.69 | 0.01 | -97.82 |
| **LOS** | 0.17 | 0.00 | 36.34 | 0.19 | 0.01 | 35.63 | 0.14 | 0.00 | 38.32 |
| **LS** | -0.12 | 0.00 | -25.41 | -0.11 | 0.01 | -21.52 | -0.09 | 0.00 | -25.87 |
| **_cons** | 6.73 | 0.06 | 118.34 | 6.98 | 0.06 | 112.48 | 6.87 | 0.04 | 155.51 |
| | | | | | | | | | |
| **ln_sig** | | | | | | | | | |
| **LOS** | 0.04 | 0.00 | 26.12 | 0.04 | 0.00 | 23.18 | 0.04 | 0.00 | 26.02 |
| **_cons** | 0.42 | 0.02 | 26.42 | 0.35 | 0.02 | 19.02 | 0.35 | 0.01 | 25.70 |
| | | | | | | | | | |
| **kappa** | | | | | | | | | |
| **_cons** | 0.25 | 0.01 | 21.58 | 0.38 | 0.01 | 31.39 | 0.26 | 0.01 | 27.22 |

**Table 8.5: Statistical Test of the Differences among Estimated Coefficients (Three Intra-Daily Periods)**

The table below shows the test statistics of the differences among the estimated coefficients reported in Table 8.4. These tests are based on the estimated coefficients and standard errors reported in Table 8.4. Ha is the alternative hypothesis, $H_0$ is the null hypothesis and $\hat{\sigma}_0$ is the estimated standard error of $H_0$. The test statistic is calculated as (Ha- $H_0$)/ $\hat{\sigma}_0$. Ha and $H_0$ are the estimated coefficients for each period reported in Table 8.4. Accordingly $\hat{\sigma}_0$ is the reported standard error in Table 8.4.

| | **Morning (Ha)** | | **Mid-day (Ha)** | | **Afternoon (Ha)** | |
|---|---|---|---|---|---|---|
| **Variable** | **Mid-day ($H_0$)** | **Afternoon ($H_0$)** | **Morning ($H_0$)** | **Afternoon ($H_0$)** | **Morning ($H_0$)** | **Mid-day ($H_0$)** |
| | | | | | | |
| **Buy Orders** | | | | | | |
| **PA** | -5.81 | -21.21 | 7.88 | -16.75 | 37.50 | 21.83 |
| **BBOS** | -7.43 | -0.85* | 7.77 | 7.64 | 0.78* | -6.68 |
| **BSIDS** | 2.66 | 2.81 | -2.72 | -0.94* | -2.04 | 0.67* |
| **NL30MT** | 4.23 | 21.23 | -4.45 | 15.57 | -16.70 | -11.62 |
| **LOS** | 3.98 | 16.83 | -4.03 | 11.22 | -12.08 | -7.96 |
| **LS** | -4.21 | -8.60 | 4.22 | -2.87 | 6.33 | 2.11 |
| | | | | | | |
| **Sell Orders** | | | | | | |
| **PA** | -2.11 | -18.00 | 2.88 | -16.19 | 28.65 | 18.85 |
| **BBOS** | -9.15 | -0.59* | 11.60 | 11.82 | 0.55* | -8.72 |
| **BSIDS** | 0.64* | 5.70 | -0.69* | 4.75 | -4.15 | -3.21 |
| **NL30MT** | 13.02 | 26.94 | -14.58 | 8.51 | -21.31 | -6.01 |
| **LOS** | -3.75 | 8.21 | 4.28 | 13.69 | -6.41 | -9.38 |
| **LS** | -1.96 | -8.62 | 2.12 | -5.75 | 6.35 | 3.91 |

# CHAPTER 9 – MAIN FINDINGS, STRENGTHS, LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

## 9.1 Introduction

This chapter aims to summary the different findings relating to the objectives of each empirical chapter and draws an overall conclusion. Furthermore, strengths and limitations of this study are outlined and suggestions for future research are made.

## 9.2 Main Findings

In this section, an outline of the main findings and overall conclusions of this study are presented. Subsections 9.2.1, 9.2.2, 9.2.3 and 9.2.4 summarise the results of Chapters 5, 6, 7 and 8 respectively. Subsection 9.2.5 presents the overall conclusions of this study.

### *9.2.1 Results of Chapter 5*

LMZ (2002) model limit-order execution time using survival analysis and US data. They find that the execution time is very sensitive to some explanatory variables, but not to others. Their models have not been investigated on UK data. The main aim of this chapter is to investigate the degree to which explanatory variables suggested by LMZ capture the state of the UK order book and affect limit-order completion time.

The findings of Chapter 5 can be summarised as follows:

- The data sample used in this study is significantly different from the LMZ data sample. This study uses LSE (SETS) stocks data while LMZ use NYSE stocks data. Moreover, in this study, the data sample includes all limit orders submitted to the order books for 38 FTSE 100 stocks from October 2000 to December 2000. In contrast, LMZ use the data provided by Investment

Technology Group (ITG) only. In addition to the difference discussed above, the minimum tick in the LMZ sample is 0.125 dollars, while in the data sample used in this study it is 0.01 pence.

- In this chapter, explanatory variables used in the LMZ study are constructed, and these variables are used to model time-to-completion in the UK market. Separate models are constructed for buy and sell orders. The LMZ variables have the interpretation as follows: MQLP is a measure of the distance between the limit-order price and current midpoint. BSID is an indicator of whether the prior trade was buyer-initiated or seller-initiated. MKD1 measures the number of shares that have higher priority for execution. MKD1X is an interactive term to capture non-linearity between market depth and market price relative to the limit-order price. MKD2 is a measure of liquidity available from the opposite side of the market. SZSD is a measure of liquidity demanded by the limit order. STKV is used to capture recent shifts in trading activity. TURN is an absolute measure of volatility. In the monthly updated variables, LSO is the logarithm of the number of shares outstanding, LPR is the logarithm of share price and LVO is the logarithm of average daily volume. They are 'primitive' variables included to capture differences across stocks. This chapter shows that the cross-correlations of these explanatory variables are generally high, with most being greater than 30% in magnitude.

- In this chapter, as LMZ (2002) suggest, an AFT model is used to incorporate a vector of explanatory variables and the generalised gamma distribution is chosen as the distribution of baseline survival times. This chapter presents the estimation results of the LMZ models, which show that the estimates of the parameters are statistically significant and have the expected signs, except for those on MQLP, for both models.

- To check the goodness-of-fit of the Generalised Gamma AFT model estimated above, a graphical diagnostic (Q-Q plot) is used. This chapter presents evidence showing that the Q-Q plots obviously deviate from the 45-degree line. This chapter also shows that after excluding the limit orders with Cox-Snell residuals greater than 10, the fitness is greatly improved.

- In order to investigate the extent of the multicollinearity effect on estimated parameters in the LMZ models, this study excludes, one at a time, the two most inter-correlated variables, SZSD and MKD1X, from the models. The empirical results show that the LMZ models suffer from the effects of multicollinearity even with the large dataset used here, as some coefficient estimates change significantly.

### *9.2.2 Results of Chapter 6*

This chapter presents a set of survival models that use explanatory variables constructed to solve the multicollinearity problem inherent in the LMZ variables. Both the Generalised Gamma (GG) AFT and the Cox PH models, which are the most frequently used survival models, are used to incorporate the constructed explanatory variables. Separate models are constructed for buy and sell orders.

The findings of Chapter 6 can be summarised as follows:

- In this chapter the explanatory variables are designed to capture the characteristics of limit orders and order book conditions while simultaneously having a correlation of no higher than 40% across one another. These variables are as follows: PA measures the relative distance between the limit-order price and the contemporary best bid/offer price on the order book. BBOS represents the relative best bid-offer spread. BSIDS is an indicator of whether or not the prior trade was same-side initiated. NL30MT is a trading activity measure that provides an absolute measure of volatility. LOS captures the limit-order size. LS captures the depth of the opposite side of the order book.
- The empirical estimations of the GG AFT model show that all coefficient estimates have the expected sign. These estimates indicate that limit-order completion time is more sensitive to some variables, such as PA and BBOS, but not to others, such as LOS and LS. The Q-Q plots of the GG AFT model estimated above indicate that the explanatory variables constructed in this chapter capture the UK market conditions and the characteristics of limit orders submitted at the LSE in a much better way than the explanatory variables suggested by LMZ, which have been empirically investigated in Chapter 5.
- The empirical estimations of the Cox PH model show that the estimated hazard ratios of two variables (LOS and LS) are unexpected. The Q-Q plots of the Cox

PH model estimated above deviate markedly from the 45-degree line. These plots indicate that the Cox PH model is wrongly specified relative to the GG AFT model.

- Basing on the Q-Q plots' closeness to the 45-degree line and all estimated coefficients having expected signs, the GG AFT model performs better than the Cox PH model and, hence, is used in the remaining chapters of this thesis.

### *9.2.3 Results of Chapter 7*

All proposed models of limit-order execution in the literature capture the state of an order book at the time of order submission only. They are 'snapshot' or 'static' models. However the state of an order book (e.g. the best bid-offer spread) changes frequently. Thus 'static' models may fail to capture the changing dynamics of an order book. This chapter provides an investigation of how these dynamics are incorporated into models of limit-order completion. This chapter also provides an examination of the prediction accuracy of the proposed models.

The findings of Chapter 7 can be summarised as follows:

- Time-varying explanatory variables that capture the dynamics of an order book are incorporated into the GG AFT model proposed in Chapter 6. Technically, estimation of survival models with time-varying explanatory variables requires the splitting of survival times of limit orders. The survival time of each limit order is split into sub-periods within each of which explanatory variables are assumed to be constant but across each of which explanatory variables can be re-measured and, hence, take on different values. In this analysis, it is assumed that the explanatory variables are held constant for one hour. Accordingly, single-record survival data used in Chapter 6 are changed into multiple-record survival data.
- This chapter presents the estimation results of the dynamic GG AFT model, which show that the parameter estimates generally are statistically significant for both models of buy and sell orders. The estimation results seem to indicate that order size is not only directly related to the expected logarithm of

completion times but also important in capturing variations in unexpected completion times across orders, through a change in the shape of baseline distribution. This empirical finding also indicates that heteroscedasticity exists in the pooled data, since the variance of the residuals varies across the limit orders. Hence modelling the ancillary parameters as a function of explanatory variables could be a new approach to test heteroscedasticity on survival data, since any traditional heteroscedasticity tests cannot handle one aspect of survival data: censoring. This chapter also presents the Q-Q plots of these models, which indicate that the goodness-of-fit is improved, as these plots are closer to the 45-degree line. Hence, time-varying variables capture the state of an order book in a better way than static variables discussed in Chapter 6.

- This chapter presents the estimation results of the modified dynamic GG AFT model, which show that the parameter estimates generally are statistically significant for both models of buy and sell orders. This chapter also presents the Q-Q plots of this model, which indicate that the goodness-of-fit is improved, as these plots are even closer to the 45-degree line. Hence, in terms of goodness-of-fit, the modified dynamic GG AFT model performs best and will be used in the remainder of the thesis.
- The Q-Q plots are constructed stock-by-stock by calculating Cox-Snell residuals for each stock using the modified dynamic GG AFT models estimated with the pooled data, and stock-specific limit-order data. These plots show that the pooled models also fit individual stock limit-order data quite well.
- A point prediction is used to examine the prediction accuracy of the modified dynamic GG AFT model. This chapter presents evidence showing that predictions are optimistic, namely more than twice the actual survival time, for 65.79% of buy orders and 65.23% of sell orders and predictions are pessimistic, less than half the actual survival time, for 10.87% of buy orders and 11.29% of sell orders.

### *9.2.4 Results of Chapter 8*

In this chapter, a trading day is divided into three periods of time: morning, mid-day and afternoon. The morning period is from 09:00 to 11:30, the mid-day period is from

11:30 to 14:00, and the afternoon period is from 14:00 to 16:30. This chapter provides an investigation of the intra-day patterns of order submission and time-of-day effects on limit-order completion time in the UK market with the modified dynamic GG AFT models proposed in Chapter 7.

The findings of Chapter 8 can be summarised as follows:

- This chapter presents the percentage breakdown of limit orders over a trading day. It shows that, on average, 43.54% of buy orders and 42.99% of sell orders are submitted in the afternoon period, about 25% of orders are submitted in the mid-day period and about 30% of orders are submitted in the morning period. This chapter also presents intra-day patterns of the explanatory variables.
- In order to identify time-of-day effects on limit-order completion time, two dummy variables are incorporated into the modified dynamic GG AFT models proposed in Chapter 7. This chapter presents the estimation results of these models. These results show that the limit orders submitted in the mid-day period are expected to have longer completion times than those submitted in the morning period and the limit orders submitted in the afternoon period are expected to have shorter completion times than those submitted in the morning period.
- The sensitivities of limit-order completion time to the explanatory variables over the three different periods are also investigated in this chapter. The empirical results confirm that the sensitivities of limit-order completion time to the explanatory variables vary over the trading day.

### *9.2.5 Overall Conclusions*

This subsection outlines the overall conclusions of this study. This study investigates limit-order completion time in the UK market and identifies the potential determinants. It also investigates time-of-day effects on limit-order completion time. First, this study provides empirical evidence showing that the LMZ explanatory variables do not capture the state of the UK order book well and the LMZ models suffer from the effects of multicollinearity. Second, this study shows that the explanatory variables constructed in

Chapter 6 capture the state of the UK order book in a better way than the LMZ ones do. Third, both the AFT and Cox PH models are used to incorporate the explanatory variables. This study provides evidence showing that the GG AFT model performs better than the Cox PH model. Fourth, this study provides evidence showing that time-varying variables capture the state of an order book in a better way than static ones discussed in Chapter 6 do and completion times are quite sensitive to some explanatory variables (e.g. the best bid-offer spread) but not to the others (e.g. the limit-order size). Fifth, this study proposes a modified dynamic GG AFT model, which fits both the pooled data and individual stock limit-order data quite well. The estimation results seem to indicate that order size is not only directly related to the expected logarithm of completion times but also important in capturing variations in unexpected completion times across orders, through a change in the shape of baseline distribution. Sixth, this study provides evidence showing that the prediction accuracy of the modified dynamic GG AFT model is poor and suggests that the high percentage of censoring orders in the data sample used in this study could be the reason behind this poor prediction accuracy. Finally, this study presents the intra-day pattern of order submission in the UK market This study also provides evidence showing that limit orders placed in the afternoon period are expected to have the shortest completion times while orders placed in the mid-day period are expected to have the longest completion times, and the sensitivities of limit-order completion time to the explanatory variables vary over the trading day.

These findings above support the possible formation of sophisticated dynamic order submission strategies that trade off the price impact of market orders against the waiting times and non-execution costs inherent in limit orders.

## 9.3 Strength of this Study

The strengths of this study lie in a number of factors including the following. First, this study is undertaken on 38 FTSE 100 stocks that traded from October 2000 to December 2000 at the LSE. Considering the entire universe of limit orders submitted to the order books frees this study from sample bias. In contrast, the majority of existing literature uses data from a single order routing system only. For example, LMZ (2002) analyse limit orders submitted through a broker firm only. Second, for comparison and robustness purposes, this study uses both the AFT and PH models to incorporate the explanatory variables. Estimation results are compared throughout Chapter 6. Third, all proposed models of limit-order completion in the literature capture the state of an order book at the time of order submission only. Thus these 'static' models may fail to capture the dynamics of an order book. In contrast, this study proposes a set of survival models that incorporate time-varying explanatory variables to capture the dynamics of the order book. Although incorporating time-varying explanatory variables in survival models is not a novel concept, it is believed that this study is the first attempt to apply this approach to finance. Accordingly all major regressions are carried out using both static and time-varying explanatory variables and results are compared throughout. Finally, this study also models the ancillary parameter (sigma) as a function of one of the explanatory variables (LOS). This frees this study from the assumption of a fixed shape of baseline survival time distribution. It is believed that this study is unique in applying this technique to finance.

## 9.4 Limitations of this Study

The main limitations of this study are as follows: First, the data sample used in this study includes limit orders that are executed, cancelled or expired in the same day of submission. It excludes limit orders that are executed immediately after submission or remain on the order book overnight. Second, in this study censoring is assumed to be non-informative. However, some traders may cancel their orders due to long expected completion times. In such case, the censoring is informative and violates the assumption above. Finally, this study uses the classical goodness-of-fit test for survival models (the Q-Q plot). This test is indicative rather than conclusive. This issue needs to be addressed by statisticians and mathematicians.

## 9.5 Suggestions for Future Research

This study identifies additional research areas and provides the basis upon which further research could be undertaken. The following are suggestions for future research. First, the modified dynamic GG AFT models proposed in Chapter 7 can be tested on data from stock exchanges other than the LSE. Second, future research can be carried out on limit-order completion probability, which is one of the primary interests to traders. Third, since a limit order can be completed with one fill or multiple fills and traders' aims are usually to complete their orders with as few fills as possible, future research can investigate time-to-one-fill-completion and time-to-multiple-fills-completion using the extension of survival analysis (competing risk analysis). Finally, future research can investigate the possible formation of dynamic order submission strategies.