Algorithmic Trading: Model of Execution Probability and Order Placement Strategy

Chaiyakorn Yingsaeree

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Department of Computer Science
University College London

2012

Declaration

I, Chaiyakorn Yingsaeree, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

....

Chaiyakorn Yingsaeree

Abstract

Most equity and derivative exchanges around the world are nowadays organised as order-driven markets where market participants trade against each other without the help of market makers or other intermediaries as in quote-driven markets. In these markets, traders have a choice either to execute their trade immediately at the best available price by submitting market orders or to trade patiently by submitting limit orders to execute a trade at a more favourable price. Consequently, determining an appropriate order type and price for a particular trade is a fundamental problem faced everyday by all traders in such markets. On one hand, traders would prefer to place their orders far away from the current best price to increase their pay-offs. On the other hand, the farther away from the current best price the lower the chance that their orders will be executed. As a result, traders need to find a right trade-off between these two opposite choices to execute their trades effectively. Undoubtedly, one of the most important factors in valuing such trade-off is a model of execution probability as the expected profit of traders who decide to trade via limit orders is an increasing function of the execution probability.

Although a model of execution probability is a crucial component for making this decision, the research into how to model this probability is still limited and requires further investigation. The objective of this research is, hence, to extend this literature by investigating various ways in which the execution probability can be modelled with the aim to find a suitable model for modelling this probability as well as a way to utilise these models to make order placement decisions in algorithmic trading systems. To achieve this, this thesis is separated into four main experiments:

1. The first experiment analyses the behaviour of previously proposed execution probability models in a controlled environment by using data generated from simulation models of order-driven markets with the aim to identify the advantage, disadvantage and limitation of each method.

2. The second experiment analyses the relationship between execution probabilities and price fluctuations as well as a method for predicting execution probabilities based on previous price fluctuations and other related variables.

3. The third experiment investigates a way to estimate the execution probability in the simulation model utilised in the first experiment without resorting to computer simulation by deriving a model for describing the dynamic of asset price in this simulation model and utilising the derived model to estimate the execution probability.

4. The final experiment assesses the performance of utilising the developed execution probability models when applying them to make order placement decisions for liquidity traders who must fill his order before some specific deadline.

The experiments with previous models indicate that survival analysis is the most appropriate method for modelling the execution probability because of its ability to handle censored observations caused by unexecuted and cancelled orders. However, standard survival analysis models (i.e. the proportional hazards model and accelerated failure time model) are not flexible enough to model the effect of explanatory variables such as limit order price and bid-ask spread. Moreover, the amount of the data required to fit these models at several price levels simultaneously grows linearly with the number of price levels. This might cause a problem when we want to model the execution probability at all possible price levels. To amend this problem, the second experiment purposes to model the execution probability during a specified time horizon from the maximum price fluctuations during the specified period. This model not only reduces the amount of the data required to fit the model in such situation, but it also provides a natural way to apply traditional time series analysis techniques to model the execution probability. Additionally, it also enables us to empirically illustrate that future execution probabilities are strongly correlated to past execution probabilities. In the third experiment, we propose a framework to model the dynamic of asset price from the stochastic properties of order arrival and cancellation processes. This establishes the relationship between microscopic dynamic of the limit order book and a long-term dynamic of the asset price process. Unlike traditional methods that model asset price dynamic using one-dimensional stochastic process, the proposed framework models this dynamic using a two dimensional stochastic process where the additional dimension represents information about the last price change. Finally, the results from the last experiment indicate that the proposed framework for making order placement decision based on the developed execution probability model outperform naive order placement strategy and the best static strategy in most situations.

Acknowledgements

I knew from the beginning that pursuing doctoral study is a difficult and challenging task. Throughout this long, sometimes painful, but definitely enlightening journey, I have received a lot of support and guidance from several individuals and institutions to whom I would like to express my thanks. Among them, I want to thank my principal supervisor, Professor Philip Treleaven, for his critical reviews of my work, his highly appreciated guidance and his friendship during this work. The good advice, support and friendship from him have been invaluable on both an academic and a personal level, for which I am extremely grateful. I am also grateful to my second supervisor, Dr. Giuseppe Nuti, for introducing me to the topic as well as giving me idea and inspiration throughout my Ph.D. program. The discussions with him always made me understand more about the topic.

I also benefit from the intense research atmosphere at the Department of Computer Science of the University College London. I also acknowledge the support and encouragement of fellow PhD students and other colleagues in the department, without which my life here would not have been so rich and enjoyable. I would also like to acknowledge the financial support from the National Electronics and Computer Technology Center Thailand. Without this financial support I would not be able to pursue this PhD study in the first place.

Finally, I want to dedicate this thesis to my parents for their love and encouragement throughout my education career. They are not only loving parents but also my mentors and best friends. Their intellectual curiosity and high standards of research have been to me a constant source of inspiration and strength.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The objective of this chapter is to present an overview of this thesis by discussing the motivation behind the research problem, the objectives and contributions of this study and the structure of this thesis. The chapter starts by introducing some background information on order-driven markets and pointing out a reason why execution probability is an important component for making order placement decisions in such markets. It then briefly reviews previous methods for modelling this probability and discusses the reason why a new model is needed. The chapter then concludes with the objectives and contributions of this work and the thesis structure.

1.1 Motivations from the literature and industry

Most equity and derivative exchanges around the world are nowadays organised as order-driven markets where traders trade against each other using market and limit orders. Traders who supply liquidity to the market submit *limit orders* (i.e. requests to buy a specific quantity at a price not exceeding some specified maximum, or to sell a specified quantity at a price not less than some specified minimum) to indicate the terms at which they want to trade. Unless it can be executed against a pre-existing order in the order book, a new limit order joins the queue in the limit order book and remains there until it is amended, cancelled, or executed against subsequent orders. On the other hand, traders who take liquidity accept those terms by submitting *market orders* (i.e. requests to transact a specified quantity at the best available price) to execute the trades at the best available price. Market orders are generally executed immediately and as fully as possible. Any unexecuted part may then be converted to limit orders at the same price or executed at the next best available price which will result in partial executions at progressively worse price until the order is fully executed. Liquidity takers can also execute their trades immediately by submitting marketable limit orders (i.e. limit orders to buy at or above the best available price, or to sell at or below the best available price). Since both market orders and marketable limit orders result in immediate execution, we do not make a distinction between them and refer to both of them as market orders.

The main difference between market orders and limit orders is the price at which the order is executed and the probability of execution as well as the execution time. When market conditions permit (i.e. enough liquidity), a market order provides immediate execution but the execution price is not certain[1]. On the other hand, a limit order guarantees the execution price, but it may sometimes be executed only partially or not at all. Although, through the use of limit orders, traders can improve their execution price relative to market orders, this improvement comes with a risk of non-execution and adverse selection cost[2] inherent in limit orders if they do not monitor the market continuously. Consequently, determining an appropriate order placement strategy for a particular trade is a fundamental problem faced everyday by all traders participating in such markets, and the solution to this problem is significant not only to all traders, particularly to institutional investors who frequently trade large volumes of shares representing a quarter or more of the whole market volume, but also to market microstructure literature that analyse the rational for, and the profitability of, limit order trading as well as the characteristics and dynamic behaviours of a limit order market. Additionally, a methodology to solve such problem can also be utilised as a building block to solve many decision problems in algorithmic trading systems.

The desire to understand order placement strategies of traders in these markets has inspired a wide range of theoretical and empirical research. On the theoretical side, many order placement models have been proposed and examined to analyse the rationale for, and the profitability of, limit order trading, as well as the characteristics and the dynamic behaviour of order-driven markets (e.g. [19, 32, 33, 40, 41, 76]). Empirical approaches, on the other hand, analyse the history of trades and quotes that occur in these exchanges to achieve the same goals. Although recent empirical studies [8, 9, 17, 37, 38, 39, 62, 67, 82, 91] indicate that traders' decision about when to submit each order type is significantly influenced by the state of the order book (e.g. the queue volume, the market depth, and the bid-ask spread) as well as its dynamic (i.e. recent changes to the order book), there is very little academic research that focuses on utilising this information to make order placement decisions. Notable exceptions are Nevmyvaka et al. [72, 73] who propose a quantitative method that allows traders to optimally price their limit orders with the aim of minimising trading costs based on the state of the order book. While their results indicate that incorporating market conditions into this decision could greatly reduce trading costs, their works are loosely related to traditional order placement models as they utilise reinforcement learning to find the optimal trading policy. As a result, the main drawback of their approach is that, when a trader's trading objective changes, new reinforcement learning model has to be constructed and trained to get an appropriate trading policy. To avoid this inconvenience, it might be more appropriate to incorporate market conditions into traditional models so that, after the model is calibrated, traders can utilise the model regardless of their objectives.

---

[1] The uncertainty of execution price is usually caused by rapid changes to the limit order book during a period between order submission and trade execution. In an electronic market, multiple events can happen within a millisecond, and our execution price may be affected by the submission of market orders from competing traders as well the revision of the price and volume at the best quote.

[2] Adverse selection cost, also known as picking-off risk or winner's curse problem, is associated with the concept that limit orders are free options to other traders [21] and these options will become mispriced as soon as more information about the price of the asset is made available. Hence, traders who submit limit orders may expose potentially large losses if they do not constantly and promptly update their orders to reflect these changes.

Although traders' order placement decisions can be explained by several factors, theoretical models generally view these decisions as a trade-off between the expected profit and the free-trading option. The expected profit depends on the execution price and the execution probability of a limit order, while the value of free trading depends on the arrival probability of adverse information which may move the price through the submitted limit order. Undoubtedly, one of the most important factors in valuing such trade-offs is a model of limit order execution times and the associated execution probability [2, 19, 32, 40, 48, 76, 86]. The main reason for this is that the expected profit of traders who decide to trade via limit orders is an increasing function of the execution probability. The larger the execution probability, the shorter the expected waiting time, and thus the smaller the expected adverse selection cost. In addition, recent empirical findings indicate that there is a strong relationship between this probability and the state of the order book. In particular, Omura et al. [74] reports that the probability of execution of limit orders on the Tokyo Stock Exchange (TSE) is lower when there are open ticks between the bid-ask spread and when the depth of limit orders at the best price of the same side is high. Conversely, the execution probability is higher when the depth of the opposite side of the book is high. This is in accordance with Biais et al. [9] who indicate that, for the Paris Bourse CAC system, order flow is concentrated near the best quotes, with depth somewhat larger at nearby quotes. When depth at the best quote is large, traders rapidly place limit orders within the spread, while traders place market orders when the spread is small. All of these suggest that it is sensible to model the execution probability of limit orders using the state of the order book and utilise this model to derive the optimal order placement strategy. However, to the best of our knowledge, no research effort on this topic has been reported in the literature before. Consequently, the objective of this research is to extend this literature by investigating various ways in which the execution probability can be modeled with aim to find a suitable model for predicting this probability as well as a way to utilise these models to make order placement decisions in algorithmic trading system.

1.2 Research objectives

The main objective of this research is to develop a computational model of limit-order executions in an order driven market that can be utilised to predict the probability that a given limit order will be executed within a specified period of time. To achieve this, we firstly conduct several experiments to analyse the behaviour of previously proposed models in a controlled environment by utilising data generated from simulation models of order driven market with the aim to identify the limitation of these models. We then proposed two alternative methods for modelling this probability. The first method is an empirical model for modelling the execution probability given a specified trading horizon from the fluctuation of asset price during the interested trading period. The second method is a theoretical model linking the relationship between the order arrival/cancellation process and the asset price dynamic in an order driven market.

Another focus of this research is on the trade implementation problem faced by traders who want to trade a financial instrument in an order driven market. Specifically, the problem these traders face is whether to trade aggressively by submitting a market order or trade patiently by placing a limit order. In addition, there is the question of how to dynamically update this decision based on the changing market condition in order to execute the trade at the best price. Although recent empirical research indicate that this decision is significantly influenced by the state of the order book and its dynamics, little attention has been paid to developing an order submission model that utilises this information to optimise trade execution. To fill the gap in the literature, we are interested in extending traditional order submission models to utilise this information. Particularly, we want to model the execution probability of limit orders using this information and utilise this model to make order placement decision in algorithmic trading systems.

1.3 Major contributions

This thesis focuses on the development of the execution probability model that can be utilised to predict the probability that a given limit order will be executed within a specified period of time and the way to utilise the developed models to make order placement decisions in algorithmic trading systems. The principal contributions of this thesis are as follows:

In the first part of this thesis, we develop software for simulating order flows in an order-driven market, where the arrival of limit orders, market orders and order cancellations are characterized by independent Poisson processes, and utilise the software to generate data for comparing the performance of previously proposed execution probability models in a controlled environment. The result demonstrates that survival analysis is the most appropriate method for modelling the execution probability because of its ability to handle censored observations in the form of unexecuted and cancelled orders. However, standard survival analysis techniques utilised in previous works (i.e. the proportional hazards model and the accelerated failure time model) are not flexible enough to model the effect of explanatory variables such as limit order price, bid-ask spread and the number of orders in the order book. Additionally, the amount of the data required to fit these models at several price levels simultaneously depends linearly on the number of price levels desired. This is not a desirable property as we generally need the execution probability at all possible price levels to determine the best price for submitting the orders.

To reduce the amount of data required to fit the model, we propose a new framework for modelling the execution probability at a specified trading horizon from the distribution of asset price fluctuation during the interested period. The major advantage of this approach over traditional models is that it requires only one record per sample while traditional models might require n records per sample to model the execution probability for n price levels. Moreover, it also provides a natural way to apply traditional time series analysis techniques to model the execution probability. By applying the proposed approach to the historical dataset obtained from the Multi Commodity Exchange of India and the New York Stock Exchange, we can empirically demonstrate that future execution probability is strongly correlated to past execution probability, and the execution probability also has intraday seasonality patterns whose forms mainly differ between the individual exchanges and less between the different assets traded in the same exchange. Furthermore, we also find evidence of asymmetry between the execution probabilities of buy and sell orders, which suggest that we might need to model them separately.

To find a suitable method for modelling the execution probability under the new framework, we investigate several ways in which the unconditional and conditional distributions of asset price fluctuation can be modelled. For the unconditional distribution, we derive the unconditional distribution of price fluctuation when the asset price is assumed to follow the arithmetic Brownian motion as well as fit several distributions with non-negative support (i.e. the exponential, Weibull, gamma, generalised gamma, generalised F and Burr distribution) to the historical dataset. As the empirical distribution of price fluctuation generally has probability mass at zero and is discrete in nature, we propose estimating the parameters of these distributions by maximising the likelihood of the discrete distribution implied by the considered distribution rather than maximising the likelihood of these distributions directly. The result indicates that the distribution estimated by the proposed method is generally better than the one estimated by a traditional maximum likelihood estimator, and the generalized F distribution is the best distribution for modelling the unconditional distribution of asset price fluctuations. For conditional distributions, we perform an experiment to utilise three major time series analysis techniques (i.e. the autoregressive moving average model, the generalised autoregressive conditional heteroskedasticity and the autoregressive conditional duration model) to model the conditional distributions by maximising both original likelihood functions and the modified likelihood functions that account for the discreteness and non-negativity properties of the price fluctuation distribution. The result demonstrates that the autoregressive conditional duration model estimated by maximising the modified likelihood function is the best performing model.

Last but not least, we propose a new framework for making order placement decisions based on the trade-off between the profit gained from a better execution price and the risk of non-execution that utilises the proposed execution probability model to balance this trade-off. The result obtained from applying the proposed framework to make order placement decision for liquidity traders who need to execute their order before the end of the deadline in historical simulation indicates that the proposed framework has better performance than simple static strategies that execute a trade at the beginning and the end of the period in all cases. Although the proposed framework can beat the best static strategy only in eight out of twelve cases, the improvement gained from the proposed framework when it does beat the best static strategy is significant.

1.4 Thesis outline

The overall structure of this thesis is organised as follows:

Chapter 2 - *Background.* We briefly present background information on a number of key concepts in the areas that this research spans to give a reader a clear view of the problems and environments studied in this thesis. Particularly, information about trading mechanism frequently used in financial markets is presented with the main emphasis on the limit order book markets which are the markets studied in this research. The broad area of algorithmic trading is also reviewed to place our work in a well defined context and to outline a rich picture of business reality for which the extended version of this research could be considered in future work. The chapter ends with a review of survival analysis which is the most widely used technique for modelling execution probability.

Chapter 3 - *Simulation model of a pure double auction market.* We give an overview of the simulation models employed for studying the behaviour of the execution probability as well as assessing the prediction performance of the execution probability model studied in the subsequent chapters. The models employed here are models of agent behaviour in continuous double auction markets that contain two main types of agents (i.e. impatient agent and patient agent). Impatient agents place market orders randomly according to some predefined stochastic process, while patient agents place limit orders randomly both in time and in price. Additionally, unexecuted limit orders are assumed to be cancelled according to some predefined stochastic processes. By controlling the properties of these orders submission and cancellation processes, several realisations of the order book dynamic that have similar stochastic properties can be generated. This enables us to evaluate the performance of execution probability models in a controlled environment before applying them to the data generated from real markets.

Chapter 4 - *Execution probability model.* We present an in-depth review of execution probability models together with performance comparison in a controlled environment based on the data generated from the simulation model of an order-driven market presented in the previous section. The results indicate that among all models considered, the execution time model that utilises techniques from survival analysis to handle cancelled orders is the best performing methods both form theoretical and empirical point of views. However, the experiment in applying survival analysis techniques to model the determinants of the execution probability indicates that traditional techniques, which are the proportional hazards model and the accelerated failure time model, are not flexible enough to model this probability. Consequently, a new method that does not suffer from this limitation is required to model the execution probability properly.

Chapter 5 - *Execution probability and price fluctuation.* We propose a new framework for modelling the execution probability at a specified time period from the distribution of asset price fluctuations during the interested period. The advantage of this approach over traditional techniques is that it requires less data to model the execution probability at all price levels as it requires only one record per sample while traditional models generally require n records per sample to model the execution probability at n price levels. Additionally, it also provides a natural way to apply traditional time series analysis techniques to model the execution probability. By applying the proposed approach to the historical dataset obtained from the Multi Commodity Exchange of India and the New York Stock Exchange, we can empirically demonstrate that future execution probability is strongly correlated to past execution probability, and the execution probability also has intraday seasonality patterns. To find a suitable method to model the execution probability under this new framework, we perform several experiments to compare the performance of applying major probability distributions with non-negative support (e.g. the generalised gamma, the generalised F and the Burr distribution) as well as three major time series analysis techniques (i.e. the autoregressive moving average model, the generalised autoregressive conditional heteroskedasticity and the autoregressive conditional duration model) to model the unconditional and conditional distributions of price fluctuations. The result indicates that the generalised F distribution is the best distribution for modelling the unconditional distribution of price fluctuations, while the autoregressive conditional duration model is the most appropriate method for modelling the conditional distribution, and, thus, the best model for modelling the execution probability.

Chapter 6 - *Asset price dynamics in continuous double auction market.* We propose a stochastic model of asset prices in an order-driven market whose dynamics are described by the incoming flow of markets orders, limit orders and order cancellation processes. In particularly, we introduce a framework to model the dynamics of asset prices given the statistical properties of those processes; thus, establishing the relation between the microscopic dynamics of the limit order book and the long-term dynamics of the asset price process. Unlike traditional methods that model asset price dynamics using a one-dimensional stochastic process, the proposed framework models these dynamics using a two-dimensional stochastic process where the additional dimension represents information about the latest price change. Using dynamic programming methods, we are able to estimate several interesting properties of the asset price dynamics (i.e. volatility, occupation probability and first-passage probability), conditioning on the trading horizon, without resorting to simulation.

Chapter 7 - *Order placement strategy.*We propose a new framework for making order placement decision based on the trade-off between the profit gain from the better execution price and the risk of non-execution that utilise the developed execution probability model to balance this trade-off. The result obtained from applying the proposed framework to make order placement for liquidity traders who need to execute their order before the end of the deadline in the historical dataset obtained from the Multi Commodity of India and the New York Stock Exchange indicates that the proposed framework has better performance than the best static order placement strategy for all instruments in the Multi Commodity of India, while it beat the best static strategy only in two out of six cases studied in the New York Stock Exchange. Although the proposed framework cannot beat the best static strategy in all cases, the improvement gained from the proposed framework when it can beat the best static strategy is very significant.

Chapter 8 - *Conclusion.* We summarise the key points of this work, what guided us in this direction and what can be learned from our models and experiments. We then review our contributions and academic achievements, and suggest some direct applications for practitioners.

Chapter 2

Background

This chapter presents background information on a number of key concepts in the areas that this research spans. Particularly, the information about trading mechanisms frequently used in financial markets is presented with the main emphasis on the limit order book markets which is the main market studied in this research. The broad area of algorithmic trading is also reviewed to place our work in a well defined context and to outline a rich picture of business reality for which the extended version of this research could be considered in the future work. The chapter then ends with a review of survival analysis which is the most widely used technique for modelling the execution probability.

2.1 Market architecture

A financial market is a place where firms and individuals enter into contracts to buy or sell financial instruments such as stocks, bonds, options and futures. It basically provides a forum where traders meet to arrange trades according to some predefined rules that govern its trading mechanism. During the last few decades, financial markets have evolved significantly from traditional floor markets, where traders come in contact and agree on a price physically on the floor of an exchange, to electronic markets, where traders submit orders, via a computerised screen-based system, to a central order book, and trades are created according to a specific matching algorithm. As a result, despite their original mechanisms, most financial markets nowadays are actually hybrids, involving a limit order book and other trading mechanisms including dealers, clearing, and auctions. This section briefly describes the main properties of these trading mechanisms. Note that the detail presented in this section are summarised from [44], and, thus, more thorough information on the topic can be obtained from there.

2.1.1 Limit order markets

Most financial markets have at least one electronic limit order book. A limit order is an order to buy a specific quantity of a financial instrument at no more (or sell at no less) than a specific price. In a limit order market, orders arrive randomly in time. The limit price of a newly arrived order is compared to the orders that are already stored in the system to determine whether there is a match or not. If there is a match, the trade will occur at the price set by the order previously stored in the system. The sequence in which these orders are executed is governed by some specific priority rules specified by the exchange.

Price normally is the most important rule, and a buy (sell) order with the highest (lowest) limit price will have the highest trading priority. There is also a time rule, which states that if two orders have the same limit price, the one which was entered into the system first will have more priority. Other rules related to an order quantity are also used in some markets (e.g. LIFFE). A list of unexecuted limit orders stored in the system constitutes a limit order book. Since limit orders can be modified or cancelled at any time, the order book is dynamic and can sometimes change rapidly especially in active markets.

Instead of using limit orders, a trader may require that an order is executed at the market price, i.e., at the best available price. To achieve this, the trader can submit a market order, an unpriced order which will be executed immediately against the best order, in the market. If the order quantity is larger than the quantity available at the single best price on the book, the order will either walk the book, resulting in partial executions at progressively worse prices until the order is fully filled, or be converted to a limit order at the executed price.

A market might have multiple limit order books, each managed by different broker. Limit order books might also be used in conjunction with other mechanisms. When all trading occurrs through a single book, the market is said to be organized as a consolidated limit order book (CLOB) which is used for actively traded stocks in most Asian and European markets.

2.1.2 Dealers

Dealer markets

A dealer is basically an intermediary who is willing to act as counterparty for the trades of his customers. A trade in a dealer market, such as the FX market, usually starts with a customer calling a dealer to ask for their price quotes (i.e. the dealer's bid and ask price), and, then, the customer may buy at the dealer's ask, sell at the dealer's bid, or do nothing. Unlike limit order markets, where a buyer who thinks the best price in the book is unreasonable can place his or her own bid, a buyer in dealer markets does not have the opportunity to do that. Dealer markets are also usually characterised by low transparency since dealers usually provide quotes only in response to customer inquiries and these are not publicly visible.

In addition to dealer-customer interactions, interdealer trading is also important for conducting dealer business. Since the incoming buy and sell orders that a particular dealer sees are usually imbalanced, accommodating these customer needs may leave the dealer with an undesired long or short position. In such case, the dealer may attempt to sell or buy in the interdealer market to balance its position. Nowadays, some of these interdealer markets (e.g. FX market) are conducted via a limit order book, such as Electronic Broking Services (EBS) and Reuters Dealing 3000 Spot Matching (D2).

Dealers in hybrid markets

Dealers can make markets work where they might otherwise fail. For example, a limit order market where customers directly trade against each other generally has difficulty with small stocks for which trading interest is insufficient to sustain continuous trading. In such a case, a dealer may make continuous trading possible by actively supplying bids and offers. Although this may well be occurring in actively traded securities, the potential dealer's costs of continuously monitoring bids and offers of low activity securities may be too large to recover from the relatively infrequent trades. In these instances, continuous liquidity requires that a dealer be designated (by the market authority) and provided with additional incentives. The best-known designated dealer is possibly the NYSE specialist who has many roles and responsibilities but an important one is to maintain a two-sided market when there is nothing on the limit order book and no one else on the floor bidding or offering.

With the advent of the electronic order management system, the competitive position of dealers and other intermediaries has weakened since, nowadays, customers can update and revise their limit orders rapidly enough to respond to market conditions. Hence, they can quickly supply liquidity when it is profitable to do so and quickly withdraw their bids and offers when markets are volatile. As a result, the presence of a dealer to maintain a two-sided market has considerably diminished in most financial markets. However, dealers today serve another useful function in facilitating large (block) trades especially in the block market, also called the upstairs market. When an institution contacts a dealer to fill a large order, the dealer can act as a counterparty, try to locate a counterparty for the full amount, work the order over time, or some combination of these. The dealer's advantage here thus lies in access to capital, knowledge of potential counterparties, and expertise in executing large order overtimes.

2.1.3 Auctions

When multiple buyers and sellers are concentrated in one venue at one time, trades may not need to be coordinated since agents can contact each other sequentially to strike bilateral bargains. However, the result obtained from such approaches may not be economically efficient since many participants will execute their trades at prices worse than the best price realized over the entire set of trades. To avoid this problem, a single-price clearing, which is generally implemented with a single-price double-sided auction could be employed. In this mechanism, supply and demand curves are constructed by ranking bids and offers from all participants, and the clearing price is usually determined by maximising the feasible trading volume. The double-sided auction is widely used in security markets, especially for low activity securities. The Euronext markets, for instance, conduct auctions once or twice per day, depending on the level of interest. Double-sided auctions are also usually used to open continuous trading sessions (e.g. Euronext, Tokyo Stock Exchange, and NYSE) and, also, at the close of continuous trading sessions.

Although most auctions in secondary markets are double-sided, single-sided auctions are widely used in primary markets. These include the U.S. Treasury debt markets, and most U.S. municipal bond offerings. They are also used, though not as often, for initial issues of equity.

2.1.4 Summary

In summary, financial markets have various architectures. Some of the main characteristics that distinguish them are the presence or lack of intermediation and continuous or periodic trading. Intermediated markets employ market markers, dealers, or specialist, who determine price quotes and act as counterparties in each trade. Consequently, these markets are usually referred to as quote-driven markets due to the quote setting function of the dealers. In non-intermediated markets, trading does not involve intermediaries but submitted orders are stored, matched, and executed via the limit order book. These markets are usually referred to as order-driven markets since the whole trading process is determined by submitted orders. The second characteristic determines if trades are executed continuously during a trading session or only at certain points in time. These two models are called a continuous double auction and a periodic auction respectively.

2.2 Algorithmic trading

As financial markets become more competitive, financial institutions and investors have started to turn to automated trading, the computerisation execution of financial instruments following some specified rules and guidelines, to gain competitive advantage. With the ability to communicate electronically with exchanges and other electronic trading venues, it is now possible to construct automated trading systems to analyse the changing market data and place orders when certain criteria are met. These systems can be customised to execute almost any trading strategy. Some aim to detect fleeting price anomalies and arbitrage opportunities in order to take a position and make a profit when such situations occur. Others slice up a large trade into smaller trades to manage market impact and timing risk as well as to mask intentions and prevent their rivals from squeezing the price. This section presents an overview of this fast growing area by presenting its definition together with some aspects concerning the development of such system.

Although different people might utilise algorithmic trading systems to achieve different objectives, in general, algorithmic trading can be described as trading with some of its processes being performed by an algorithm running on a computer with little or no human intervention. The trading process could be roughly separated into three main steps: trading signal generation, trading decision, and trade execution. The signal generation usually involves the analysis of changing market information to detect the trading opportunities within the market, and the result is a trading signal indicating when to buy and when to sell a particular financial instrument. The generated trading signals are then analysed, usually by humans, to confirm the trading decision in the second step. After the trading decision is finalised, the last step is to execute the trading decision by sending the corresponding order to financial markets. Although this simplified description of the trading process does not cover some traders (e.g. the market makers), it illustrates that the trading process can be divided into different steps, each of which can be separately programmed and executed by an algorithm running on a computer system. According to this simplified trading process, algorithmic trading system might be categorised into four main types, as described by Idvall and Jonsson [50], which are:

- Systems that automate the first step of the trading process, namely the trading signal generation. Thus, human intervention is required for the last two tasks of the trading process, which are the trading decision and the execution of the trade.
- Systems that automate the trade execution, which is last step of the trading process. The aim of the execution algorithm is often focused on placing and managing orders in the market in order to minimize the trading cost. Using execution algorithms leaves the first two steps to the human trader.

- Systems that combine the first two categories but leaving the trade execution to the human trader.
- Fully automated systems, often referred to as black-box trading systems, that automate all steps in the trading process.

Hence, most algorithmic trading systems consist of two main parts: determining *when to trade* and *how to trade*. Determining when to trade is the analytical part of the strategy which revolves around watching the changing market data and detecting opportunities within the market. For example, consider a pair trading strategy that examines pairs of financial instruments that are known to be statistically correlated. Normally, statistically correlated instruments are likely to move together. When these instruments break correlation, the trader may buy on and sell the other at premium with the hope to gain profit when both instruments become correlated again. In this case, the algorithm involves monitoring for any changes in the price of both instruments and then recalculating various analytics to detect a break in correlation. Another example is a market making strategy which tries to place a limit order to sell above the current market price or buy a limit order below the current price in order to benefit from the bid-ask spread. In addition, any sort of pattern recognition or predictive model can also be used to initiate the trade. Neural networks and genetic programming have been extensively used to create these models.

Determining how to trade focuses on placing and managing orders in the market. At the lowest level, this involves determining a suitable choice of order type (i.e. limit and market order) for each trade. This choice is no simple matter and requires some sophistication since market orders are executed immediately but incur substantial price impact while limit orders incur no price impact but may not be executed immediately, if at all. A higher level problem involves breaking up a large order into smaller orders and placing them into the market over time. The benefit of this is that large orders have a major impact in moving the market while smaller orders are more likely to flow under the market's radar, and subsequently have less impact on the market. In addition, when an instrument is traded in multiple exchanges, an execution strategy also needs to determine where the order should be submitted to. Since this research is more related to this issue, more detailed information about it will be discussed in the next section.

2.3 Trade execution strategies

An investor, or an algorithmic trading system, who wants to buy or sell shares of a particular financial instrument faces a number of choices. After the trading decision have been finalised (i.e. the financial instruments for buying and selling have already been picked), the main problem to be solved is trade execution with constraints on transaction costs and trading duration. To execute the trade, an order has to be submitted to a trading venue with the choice depending on the selected financial instrument, order size, hours of operation and other factors. If an order requires a small number of shares, comparing to the available liquidity, it can be executed by a submission of a single market order. Alternatively, if the number of shares required is larger than what is available in the market, an order may be broken up into a sequence of smaller orders which will be submitted to the market over a specific period of time. In addition, a trader also needs to decide on the preferred order type. If the trader is patient, he may choose to submit a limit order to obtain price improvement. On the other hand, an information motivated trader may choose to submit a market order to achieve an immediate execution. The following sections briefly discuss these three problems. More detailed information can be found in reference [10].

2.3.1 Choice of trading venue

Some financial instruments may be traded on more than one financial market. To execute a trade for these instruments, a trader needs to determine the market the order has to be submitted to. Normally, the trader may want to submit the order to the market whose characteristics suit his requirements most. Some of the most important characteristics the trader usually considers are liquidity, trading mechanism and degree of trader's anonymity.

A financial instrument in a particular market is considered liquid if the volume of trades and orders of that instrument is large. Liquidity is important because a high liquid market is usually associated with fast trade execution and low transaction costs. Thus, all other things being equal, the trader would prefer to submit his orders to the market with the most liquidity.

A trading mechanism employed in the market is also an important characteristic the trader usually considers before making trade execution decision, since each mechanism has its own advantages and disadvantages. As discussed in Section 2.1, trades in a continuous double auction market are executed continuously during a trading session, while trades in a periodic auction are executed only at certain points in time. As a result, it is more appropriate to trade in a continuous double auction market when immediacy is required. However, trades in a periodic auction have lower price volatility when compared to trades in the continuous double auction [25].

In the case of a trader's order being too large to be executed instantaneously without an unwanted price impact, the trader's action will be influenced by his trading motivations. If trading is information motivated, it might be more appropriate to carry out in a market that offers anonymity. In addition, trader may also break up the large order into a sequence of smaller orders and submit them to the market over a period of time with the aim of reducing price impact by hiding from other participants the fact that all those orders were originated by the same trader. On the other hand, a liquidity-motivated trader whose motivation is not information related is not necessary to do that and may submit his order to an upstairs market directly.

2.3.2 Choice of trade schedule

As previously discussed, a price impact of a single market order for a particular financial instrument will be minimal if the order size does not exceed the volume available at the best quote. However, if the size of the order is too large to execute without an unwanted price impact, it would be more efficient to break the order down into several smaller orders which are then submitted into the market over a period of time. The benefit of this is that large orders have a major impact in moving the market while smaller orders are more likely to flow under the market's radar, and subsequently have less impact to the market. Although smaller orders will have a lower price impact, delayed execution may expose them to potential adverse price movements as well as an opportunity cost. Thus, the problem of generating an optimal trade schedule which will achieve a desired balance between price impact and opportunity cost is another important problem for traders whose position is usually larger than the depth of the market.

During the last decades, there is a growing interest in developing models to solve such decision problems (See [3, 16, 54, 4, 5] for example). The optimal trade schedule generated from these models usually depends on several factors including trader's objectives, market impact and the dynamics of future market prices. Typically there are two main steps in specifying the trading objective. The first step is to define execution cost by defining the specification of transactional cost and choosing the desired benchmark price (e.g. previous close, opening price, arrival price, VWAP, TWAP, and future close). The benchmark price is investor specific and depends on investment objectives (e.g. a mutual fund may desire execution at the closing price to coincide with valuation of the fund while an indexer may desire execution that achieves VWAP as an indication of fair prices for the day). The second step is to specify the degree of risk-aversion (i.e. how much we penalise variance relative to expected cost) which indicates the level of trading aggressiveness or passiveness. Aggressive trading is associated with higher cost and less risk while passive trading is associated with lower market impact and higher risk. Market impact, or the degree to which an order affects the market price, consists of permanent impact cost due to information leakage of the order and temporary impact cost due to the liquidity and immediacy needs of the investor. These market impacts are usually approximated by fitting some parametric functions (e.g. linear and power laws function) using historical data. In addition, these functions can be both time dependent and time independent. To specify the dynamics of future market prices, arithmetic random walk is the most popular model. Giving specifications of all these factors, an optimal trading strategy for a specific trading objective may be obtained by solving the corresponding stochastic dynamic optimisation problem.

2.3.3 Choice of order type

As discussion in Section 2.1.1, there are two main order types that a trader can submit to an order-driven market which are a market order and a limit order. A market order is an order to buy/sell a pre-specified quantity of a financial instrument at the best available price placed by previously submitted limit orders that make up a limit order book. In contrast, a limit order is an order to buy/sell a pre-specified quantity of a financial instrument at a specific price. Unexecuted limit orders are stored in the limit order book until they are cancelled or triggered by incoming market orders.

The main differences between market orders and limit orders are the price at which the order is executed and the probability of execution. When market conditions permit (i.e. enough liquidity), a market order provides immediate execution but the execution price is not certain[1]. On the other hand, a limit order guarantees the execution price, but the order may sometimes be executed only partially, or not at all. In addition, a trader who submits limit orders may also offset the price by the picking-off risk[2] if he does not monitor the market continuously. Although with limit orders traders can improve

[1] The uncertainty of execution price is usually caused by rapid changes to the limit order book during a period between order submission and trade execution. In an electronic market, multiple events can happen within a millisecond, and our execution price may be affected by the submission of market orders from competing traders as well the revision of the price and volume at the best quote.

[2] Picking-off risk, also known as adverse selection cost and winner's curse problem, is associated with the well-known concept that limit orders are free options to other traders and these options will become mispriced as soon as the fundamental of the instrument is changed. Hence, traders who submit limit orders may expose potentially large losses if they do not constantly update

their execution price relative to market orders, this improvement is offset by the risk of non-execution and adverse selection cost inherent in limit orders. Thus, to determine an appropriate order submission strategy, traders have to find the right trade-off between price improvement, execution probability and adverse selection cost. Note that this decision problem is the main focus of this research and more detail information about how to model it will be discussed in Chapter 7.

2.4 Survival analysis

Survival analysis is a class of statistical methods for analysing the occurrence and timing of events, which is usually referred to as *survival time* or *failure time*. Examples of survival times include the lifetimes of machine components in industrial reliability, the duration or period of unemployment in economics, the survival times of patients in a clinical trial or, in our case, the waiting times until limit orders are executed. A special difficulty in analysing these data is that failure time information for some individuals in a dataset may be incomplete, or so-called censored, which generally occurred when some individuals can not be observed for the full time until the event of interest is happened. For instance, at the close of the market, not all limit orders may have executed and we will not be able to observe the time-to-execution of those orders. What makes survival analysis differ from other methods is a unique and natural approach for accommodating these censored observations by maximally extracting partial information from censored observations rather than just including or excluding them from the dataset which generally cause biases in statistical inference. This section presents a brief review of survival analysis techniques. Readers interested in a more detail exposition should consult Cox and Oakes [22] and Kalbfleisch and Prentice [52].

2.4.1 Basic definitions

Let T be a nonnegative random variable representing the survival times of individuals in some population. Let $F(.)$ denote the *cumulative distribution function* (c.d.f) of T with the corresponding *probability density function* (p.d.f) $f(.)$. Since $T \geq 0$ we have

$$F(t) = Pr\{T \leq t\} = \int_0^t f(x)dx. \tag{2.1}$$

The probability that an individual's survival time will be at least t is given by the *survivor function*, $S(t)$, which could be defined by

$$S(t) = Pr\{T \geq t\} = 1 - F(t) = \int_t^\infty f(x)dx. \tag{2.2}$$

Note that this function is a monotonic decreasing function with $S(0) = 1$ and $S(\infty) = \lim_{t\to\infty} S(t) = 0$. Conversely, we can express the p.d.f. as

$$f(t) = \lim_{\Delta t \to 0^+} \frac{Pr\{t \leq T < t + \Delta t\}}{\Delta t} = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt} \tag{2.3}$$

their orders to reflect these changes.

Additionally, the *hazard rate* of T at time t which is the instantaneous failure rate of failure at $T = t$ given that the individual survived up to time t is defined as

$$h(t) = \lim_{\Delta t \to 0^+} \frac{Pr\{t \leq T < t + \Delta t \mid T \geq t\}}{\Delta t} = \frac{f(t)}{S(t)}. \tag{2.4}$$

When the distribution has an atom f_j of probability at time a_j, the hazard function $h(t)$ will contain a component $h_j \delta(t - a_j)$, where

$$h_j = f_j / S(a_j). \tag{2.5}$$

Consequently, the hazard function for a purely discrete distribution with atoms $\{f_j\}$ at points $\{a_j\}$ where $a_1 < a_2 < \cdots$, is specified by

$$h(t) = \sum h_j \delta(t - a_j), \tag{2.6}$$

where

$$\begin{aligned} h_j &= f_j / S(a_j) \\ &= f_j / (f_j + f_{j+1} + \ldots). \end{aligned}$$

For continuous distributions in Equation (2.3) and (2.4) we have

$$h(t) = -\frac{dS(t)/dt}{S(t)} = -\frac{d \log\left(S(t)\right)}{dt}. \tag{2.7}$$

This indicates that $h(t)$ completely specifies the distribution of T since we can recover the distribution of $S(t)$, $f(t)$ and $F(t)$ by knowing only $h(t)$. To see this, the above equation can be rearranged into

$$\begin{aligned} S(t) &= \exp\left(-\int_0^t h(u)du\right), \\ &= \exp\big(-H(t)\big), \end{aligned} \tag{2.8}$$

where $H(t) = \int_0^t h(u)du$ is the cumulative hazard function. Consequently, the p.d.f of T can be expressed as

$$\begin{aligned} f(t) &= h(t) \exp\left(-\int_0^t h(u)du\right), \\ &= h(t) \exp\big(-H(t)\big). \end{aligned} \tag{2.9}$$

The above equations illustrate that any one of these four functions (i.e. the p.d.f, the c.d.f, the survivor function and the hazard function) uniquely determines the other three, and, thus, if one function is known, the rest can be derived mathematically. Consequently the focus of survival analysis is to estimate one of these functions from the dataset and the methods for achieving this will be described in Section 2.4.3.

2.4.2 Censoring

Before discussing the methods for estimating the survival probability, let us firstly describe the concept of censoring and censored observations. Censoring comes in many forms and occurs for many different reasons. Fundamentally, censoring occurs when we have some information about the individual survival time, but the survival time is not exactly known. This can generally be categorised into three main types which are right censoring, left censoring and interval censoring.

An observation on a variable T is right censored if all we know about T is that it is greater than some value c. This typically happens when observation is terminated before the event occurs. As an example, consider the execution time of limit orders submitted to the market. If a given order is cancelled or expired before execution, then the execution time of this order is considered censored. Although we do not know the execution time of this order, we know that the execution time of this order is at least as long as the its cancellation or expiration time.

Conversely, left censoring occurs when all we know about an observation on a variable T is that it is less than some values. This is most likely to occur when we start to observe a sample at a time when some of the individuals may have already experienced the event. For example, consider a situation when we are monitoring the first-passage time that a stock price reaches a particular level in an electronic market and the network connection is accidentally broken. If after the network is back to live we find out that the stock price already crosses that level, then this first-passage time will be considered censored. Although we do not know the exact first-passage time of this order, we know that this first-passage time must be less than the time we come back to live.

Finally, interval censoring combines both right and left censoring. Particularly, an observation on a variable T is interval censored if all we know about T is that $a < T < b$, for some values of a and b. This type of censoring is likely to occur when observations are made at infrequent intervals and there is no way to get retrospective information on the exact timing of events. For example, consider the situation when the execution time of limit order cannot be directly observed. We may estimate the execution time by the first time that the stock price reaches and crosses the order price. Although we do not know the exact execution time of this order, we know that the execution time of this order must be greater than the first time that the stock price reaches the order price while it must be less than the first time that the stock price crosses the order price.

Since we will consider only right censored data in this research, the rest of this section will discuss only a technique for right censored data.

2.4.3 Estimation methods

Given a dataset containing information about the lifetime and the censored time of an individual observation, a survival distribution can be estimated using both parametric and nonparametric methods. Parametric methods generally assume a specific parametric family for the distribution of failure times (e.g. exponential distribution, Weibull distribution and gamma distribution) and estimate its parameters using maximum likelihood estimation. On the other hand, nonparametric methods estimate the survivor function nonparametrically, without resorting to any parametric assumptions. This section briefly reviews concepts behind these methods, with an emphasis on the methods that will be utilised in the rest of this research.

Let T denote a lifetime with c.d.f. $F(t)$ and survivor function $S_f(t)$ and C denote a random censor time with c.d.f. $G(t)$, p.d.f. $g(t)$ and survivor function $S_g(t)$. Each individual has a lifetime T_i and a censor time C_i. Instead of directly observing T_i and C_i, we observe the pair (Y_i, δ_i) where

$$Y_i = \min(T_i, C_i), \quad \text{and} \quad \delta_i = \begin{cases} 1 & \text{if } T_i \leq C_i \\ 0 & \text{if } C_i < T_i \end{cases}. \tag{2.10}$$

Under the assumption that times T_i and C_i are independent of each other, the likelihood of censored observations can be computed from

$$\begin{aligned} Pr\{Y = y, \delta = 0\} &= Pr\{C = y, C < T\} = Pr\{C = y, y < T\} \\ &= Pr\{C = y\}Pr\{y < T\} \text{ by independence} \\ &= g(y)S_f(y), \end{aligned} \tag{2.11}$$

while the likelihood of noncensored observations is

$$\begin{aligned} Pr\{Y = y, \delta = 1\} &= Pr\{T = y, T < C\} = Pr\{T = y, y < C\} \\ &= Pr\{T = y\}Pr\{y < C\} \\ &= f(y)S_g(y). \end{aligned} \tag{2.12}$$

Combining these two above equations, the likelihood function for n independent and identically distribution (i.i.d.) random pairs (Y_i, δ_i) is given by

$$\begin{aligned} L &= \prod_{i=1}^{n} \left(f(y_i)S_g(y_i)\right)^{\delta_i} \cdot \left(g(y_i)S_f(y_i)\right)^{1-\delta_i} \\ &= \left(\prod_{i=1}^{n} S_g(y_i)^{\delta_i} g(y_i)^{1-\delta_i}\right) \cdot \left(\prod_{i=1}^{n} f(y_i)^{\delta_i} S_f(y_i)^{1-\delta_i}\right) \end{aligned} \tag{2.13}$$

When the distribution of C does not involve any parameters of interest, which is the case in this research, the first factor in the above equation plays no role in the maximisation process, and, hence, the likelihood function can be reduced to

$$\begin{aligned} L &= \prod_{i=1}^{n} f(y_i)^{\delta_i} S_f(y_i)^{1-\delta_i} \\ &= \prod_{U} f(y_i) \prod_{C} S_f(y_i), \end{aligned} \tag{2.14}$$

where U and C denote the indexes of the uncensored and censored observations, respectively. Given the likelihood function in Equation (2.14), the parameters of the distribution T can be estimated via maximum likelihood estimation.

Parametric methods

Let us firstly discuss parametric methods for estimating survival distributions from the dataset containing n i.i.d. random pairs as discussed above. To achieve this, parametric methods generally assume a specific parametric family of survival distribution and estimate its parameters from the dataset using the maximum likelihood estimator. Particularly, parametric methods assume that T_1,T_2,...,T_n are i.i.d. from a known distribution with p.d.f. $f(t|\theta)$ and survivor function $S(t|\theta)$ where θ is a vector of parameters of the specific distribution which belongs to some parameter space Ω. The likelihood function from n i.i.d. random pairs in Equation (2.14) is now given by

$$L(\theta) = \prod_{i=1}^{n} f(y_i|\theta)^{\delta_i} S(y_i|\theta)^{1-\delta_i} = \prod_{U} f(y_i|\theta) \prod_{C} S(y_i|\theta) \tag{2.15}$$

The maximum likelihood estimator (MLE), denoted by $\hat{\theta}$, is the value of θ in Ω that maximises $L(\theta)$ or, equivalently, maximises the log-likelihood

$$\log L(\theta) = \sum_{U} \log f(y_i|\theta) + \sum_{C} \log S(y_i|\theta) \tag{2.16}$$

Although any distributions over nonnegative values can be utilised to model the survival time distribution, we will only focus on six widely used distributions, which are the exponential distribution, Weibull distribution, standard gamma distribution, generalised gamma distribution, log-normal distribution and log-logistic distribution. The rest of this section will cover the analytical expressions of these six distributions in more detail.

Exponential distribution

An exponential distribution with the rate parameter $\lambda > 0$ has

$$f(t|\lambda) = \lambda e^{-\lambda t} \tag{2.17}$$

$$S(t|\lambda) = e^{-\lambda t}, \tag{2.18}$$

$$h(t|\lambda) = \lambda. \tag{2.19}$$

The outstanding simplicity of this model is reflected in its constant hazard rate which reflects the lack of memory property. Particularly, for any $t \geq 0$, the conditional distribution of $T > t + s$, given $T > s$, is the same as the unconditional distribution of $T > t$.

Weibull distribution

A Weibull distribution with scale parameter $\lambda > 0$ and shape parameter $\kappa > 0$ has

$$f(t|\lambda, \kappa) = \lambda\kappa(\lambda t)^{\kappa-1} e^{-(\lambda t)^{\kappa}}, \tag{2.20}$$

$$S(t|\lambda, \kappa) = e^{-(\lambda t)^{\kappa}}, \tag{2.21}$$

$$h(t|\lambda, \kappa) = \lambda\kappa(\lambda t)^{\kappa-1}. \tag{2.22}$$

Since $S(t|\lambda,\kappa) = e^{-\lambda^\kappa t^\kappa}$, it follows that T^κ has an exponential distribution with rate parameter λ^κ. Note that the Weibull distribution reduces to the exponential distribution when $\kappa = 1$. Additionally, the Weibull hazard function is monotonically increasing when $\kappa > 1$ and monotonically decreasing when $\kappa < 1$. Hence, the parameter κ is called the shape parameter, as the shape of the distribution depends on the value of κ. On the other hand, the λ is called a scale parameter because the effect of different value of λ is just to change the scale on the horizontal axis, not the basic shape of the distribution.

Standard gamma distribution

A standard gamma distribution with scale parameter $\lambda > 0$ and shape parameter $\kappa > 0$ has

$$\begin{aligned} f(t|\lambda,\kappa) &= \frac{\lambda(\lambda t)^{\kappa-1}e^{-\lambda t}}{\Gamma(\kappa)}, && (2.23)\\ S(t|\lambda,\kappa) &= 1-\frac{\gamma(\kappa,\lambda t)}{\Gamma(\kappa)}, && (2.24)\\ h(t|\lambda,\kappa) &= \frac{\lambda(\lambda t)^{\kappa-1}e^{-\lambda t}}{\Gamma(\kappa)-\gamma(\kappa,\lambda t)}, && (2.25)\end{aligned}$$

where $\gamma(\kappa,\lambda t)$ is the lower incomplete gamma function, which is defined by

$$\gamma(\kappa,x) = \int_0^x t^{\kappa-1}e^{-t}dt, \tag{2.26}$$

and $\Gamma(\kappa)$ is a gamma function, which is defined as

$$\Gamma(\kappa) = \int_0^\infty t^{\kappa-1}e^{-t}dt. \tag{2.27}$$

Like the Weibull distribution, the gamma distribution contains the exponential distribution as a special case when $\kappa = 1$. Its hazard function is monotone increasing from zero when $\kappa > 1$ and monotone decreasing from ∞ if $\kappa < 1$. Unlike the Weibull distribution, where the hazard function increases without limit, the gamma hazard function approaches λ as an upper limit. Similarly, when the decreasing Weibull hazard approaches zero as a lower limit, the decreasing gamma hazard has a lower limit of λ.

Log-normal distribution

Possible distributions for T can also be obtained by specifying for $\log(T)$ any convenient family of distributions on the real line. The simplest possibility is to take $\log(T)$ normally distributed with mean μ and variance σ^2, which leads to the log-normal distribution with

$$\begin{aligned} f(t|\lambda,\alpha) &= (2\pi)^{-\frac{1}{2}}\alpha t^{-1}\exp\left(\frac{-\alpha^2(\log(\lambda t))^2}{2}\right), && (2.28)\\ S(t|\lambda,\alpha) &= 1-\Phi\big(\alpha\log(\lambda t)\big), && (2.29)\\ h(t|\lambda,\alpha) &= \frac{(2\pi)^{-\frac{1}{2}}\alpha t^{-1}}{1-\Phi\big(\alpha\log(\lambda t)\big)}\exp\left(\frac{-\alpha^2(\log(\lambda t))^2}{2}\right), && (2.30)\end{aligned}$$

where $\mu = -\log(\lambda)$, $\sigma = \alpha^{-1}$ and $\Phi(.)$ is the cumulative distribution function of the standard normal distribution. Unlike the exponential and Weibull distributions, the hazard function associated with log-normal distribution is non-monotonic. In fact, log-normal hazard function has an inverted U-shape, in which the hazard increases from 0 at $t = 0$ to a maximum at the mean value and then decreases and approaches zero as t become large.

Log-logistic distribution

Another convenient distributions for $\log(T)$ is the continuous logistic density with location parameter μ and scale parameter σ which leads to log-logistic distribution with

$$f(t|\lambda, \alpha) = \frac{\alpha t^{\alpha-1}\lambda^{\alpha}}{\left(1 + (\lambda t)^{\alpha}\right)^2}, \quad (2.31)$$

$$S(t|\lambda, \alpha) = \frac{1}{1 + (\lambda t)^{\alpha}}, \quad (2.32)$$

$$h(t|\lambda, \alpha) = \frac{\alpha t^{\alpha-1}\lambda^{\alpha}}{1 + (\lambda t)^{\alpha}}, \quad (2.33)$$

where $\mu = -\log(\lambda)$ and $\sigma = \alpha^{-1}$ as in the case of log-normal distribution. Similar to the log-normal distribution, the log-logistic hazard function also has an inverted U-shape. Particularly, if $\alpha > 1$, the hazard function has a single maximum, while if $\alpha < 1$ the hazard function is decreasing.

Generalised gamma distribution

A generalised gamma distribution with a scale parameter $\lambda > 0$ and two shape parameters $\kappa > 0$ and $\rho > 0$ has

$$f(t|\lambda, \kappa, \rho) = \frac{\rho\lambda(\lambda t)^{\rho\kappa-1}e^{-(\lambda t)^{\rho}}}{\Gamma(\kappa)}, \quad (2.34)$$

$$S(t|\lambda, \kappa, \rho) = 1 - \frac{\gamma(\kappa, (\lambda t)^{\rho})}{\Gamma(\kappa)}, \quad (2.35)$$

$$h(t|\lambda, \kappa, \rho) = \frac{\rho\lambda(\lambda t)^{\rho\kappa-1}e^{-(\lambda t)^{\rho}}}{\Gamma(\kappa) - \gamma(\kappa, (\lambda t)^{\rho})}. \quad (2.36)$$

Since the generalised gamma distribution has one more parameter than other distributions, its hazard function can take on a wide variety of shapes. In fact, the exponential, Weibull, standard gamma, and log-normal distributions are all special cases of the generalised gamma model. In particular, the generalised gamma distribution reduces to the exponential distribution when $\kappa = 1$ and $\rho = 1$, reduces to the Weibull distribution when $\kappa = 1$, reduces to standard gamma distribution when $\rho = 1$ and reduces to log-normal distribution when $\kappa \to \infty$. Additionally, the generalised gamma distribution also allows the hazard function with U or bathtub shapes, in which the hazard declines, reaches a minimum, and then increases.

Non-parametric methods

Unlike parametric methods, non-parametric methods estimate the survivor function without resorting to any parametric assumptions. This can be achieved by assuming that the possibly improper distribution for the survival time is discrete, with atom f_i at finitely many specified points $a_1 < a_2 < \cdots < a_g$. Since the distribution obtained from non-parametric methods is purely discrete, this section will firstly derive the survivor function when the distribution is purely discrete and then discusses the method for estimating this discrete distribution from the dataset.

As described in Section 2.4.1, the hazard function for a purely discrete distribution with atoms $\{f_j\}$ at points $\{a_j\}$ where $a_1 < a_2 < \cdots$ can be specified by

$$h(t) = \sum h_j \delta(t - a_j),$$

where

$$\begin{aligned} h_j &= f_j / S(a_j) \\ &= f_j / (f_j + f_{j+1} + \ldots). \end{aligned}$$

Consequently, we have

$$h_j = f_j / \big(f_j + S(a_{j+1})\big), \tag{2.37}$$

or equivalently,

$$S(a_{j+1}) = \frac{f_j}{h_j} - f_j = \frac{f_j}{h_j}(1 - h_j) = S(a_j)(1 - h_j) \tag{2.38}$$

By applying Equation (2.38) recursively or by a direct application of the product law of probabilities, we have

$$S(t) = \prod_{a_j < t} (1 - h_j) = \prod_{(t)} (1 - h_j), \tag{2.39}$$

where $\prod_{(t)}$, and subsequently $\sum_{(t)}$, denote product and summation over $j, a_j < t$. Consequently, in terms of h_j, the f_j may be written in the form

$$f_j = h_j S(a_j) = h_j \prod_{k<j} (1 - h_k). \tag{2.40}$$

To derive the full likelihood function from a sample of n i.i.d. random pairs, we first collect all information related to the atom a_j. If there are d_j failures among the r_j individuals in view at a_j, the contribution to the total likelihood is

$$h_j^{d_j} (1 - h_j)^{r_j - d_j}. \tag{2.41}$$

The total log-likelihood is then

$$L = \sum_j \left[d_j \log h_j + (r_j - d_j) \log(1 - h_j) \right]. \tag{2.42}$$

Consequently, a non-parametric estimator of the survivor function can be specified by

$$\hat{S(t)} = \prod_{(t)} (1 - \hat{h}_j), \tag{2.43}$$

where $\hat{h}_j$ is the maximum likelihood estimator of the h_j, and is the solution of

$$\frac{\partial L}{\partial h_j} = \frac{d_j}{h_j} - \frac{r_j - d_j}{1 - h_j} = 0,$$

i.e. $\hat{h}_j = d_j/r_j$. Hence, the corresponding non-parametric estimator of the survivor function, generally called the Kaplan-Meier or product-limit estimator [53], is then

$$\hat{S}(t) = \prod_{(t)} \left(1 - \frac{d_j}{r_j}\right) = \prod_{(t)} \left(\frac{r_j - d_j}{r_j}\right). \tag{2.44}$$

Note that any terms in the product which $d_j = 0$ can be omitted without affecting the estimation result. As a result, the estimate $\hat{S}(t)$ is formally independent of the selection of point a_j for which the observed number of failures is zero, and is therefore a function of the data only. Although minus of the logarithm of the Kaplan-Meier estimator could also be used to estimate the cumulative hazard function, it is more usual to take

$$\hat{H}(t) = \sum_{(t)} h_j = \sum_{(t)} d_j/r_j, \tag{2.45}$$

which often called the Nelson estimators [71].

2.4.4 Dependency on explanatory variables

In the previous section, we review methods for the relatively simple problem involving a single distribution with no explanatory variables. However, in real situations, we are also interested in the effect that each explanatory variable has on the survival time. For example, we may be interested in the effect that the limit order size has on the limit order execution time. Although this can be analysed by estimating survivor functions for each value of explanatory variables separately and then making a qualitative comparison between the estimated survivor functions. More complicated analysis is best handled by comprehensive models in which the effects from explanatory variables are represented as parameters in the model. This section presents two widely used models for modelling the relationship between survival time and explanatory variables. These models are the accelerated failure time model and the proportional hazards model.

Accelerated failure time model

One way to extend the analysis from the previous section to handle explanatory variables is to assume that the difference between two individuals is the rate at which they age. Specifically, let T denote failure time and $X \equiv \big(X_1, \ldots X_n\big)$ represent a vector of explanatory variables. Accelerated failure time model assumes that there is a function $\psi(X)$ such that the survival time of each individual is given by

$$T = T_0/\psi(X), \tag{2.46}$$

where T_0 is the baseline survival time which is basically the survival time under standard condition when $\psi(X) = 1$. Consequently, T is a scaled transformation of the baseline survival time T_0, where the scaling is determined by the explanatory variables. With this assumption, the survivor function, density function and hazard function are, respectively,

$$\begin{aligned} S(t|X) &= S_0\big(t\psi(X)\big), \\ f(t|X) &= f_0\big(t\psi(X)\big)\psi(X), \qquad (2.47) \\ h(t|X) &= h_0\big(t\psi(X)\big)\psi(X),, \end{aligned}$$

where $S_0(.)$, $f_0(.)$ and $h_0(.)$ are the survivor function, density function and hazard function of the baseline survival time T_0 respectively. Although the choice of $\psi(X)$ can be any arbitrary function satisfying $\psi(X) > 0$, a natural candidate is

$$\psi(X|\beta) = e^{\beta^T X}, \qquad (2.48)$$

where β is a vector of parameters. Consequently, Equation (2.46) can be rewritten as

$$\begin{aligned} T &= T_0 e^{-\beta^T X}, \\ \log T &= \mu_0 - \beta^T X + \epsilon, \qquad (2.49) \end{aligned}$$

where $\mu_0 = E(\log T_0)$ and ϵ is a random variable of zero mean, whose distribution does not depend on X. The above equation is very similar in form to a linear regression model and the main difference is that the dependent is in a logged form. With a specified parametric form of the baseline distribution, this parameter vector can be directly incorporated into the likelihood function by substituting Equation (2.47) into Equation (2.16), and, thus, its values can be directly estimated by maximum likelihood methods as described in the previous section.

Model checking

The central property of the accelerated failure time model can be re-expressed in various ways that can be utilised as a basis for testing the adequacy of the model. From Equation (2.49), the distribution of $\log T$ at various values of X differ only by translation. Consequently, in a two-sample problem where the value of the explanatory variable can take only two values, we can compare quantiles and utilise quantile-quantile plot to verify the assumption of the accelerated failure time model. Particularly, let $S_0(t)$ and $S_1(t)$ be the survivor function of the two-sample distribution, and define t_0^p and t_1^p for $0 < p < 1$, by

$$p = S_0(t_0^p), \qquad t_0^p = S_0^{-1}(p), \qquad (2.50)$$

$$p = S_1(t_1^p), \qquad t_1^p = S_1^{-1}(p). \qquad (2.51)$$

Since under Equation (2.46), t_1^p must equal to t_0^p/ψ, i.e. $t_1^p = t_0^p/\psi$, the plot between t_1^p and t_0^p must be a straight lines through the origin and the evidence against this is an indication of the violation of the accelerated failure time assumption. Alternatively, we can also verify this assumption by plotting $\log t_1^p$ and $\log t_2^p$ versus p on the same chart. If the assumption is satisfied, these two curves should be parallel to each other since they differ only by translation, and thus the evidence against this is also an indication of the violation of this assumption.

To assess this assumption quantitatively, we can utilise a method presented by Yang [94] to estimate the timescale factor, $\hat{\psi}$ from the dataset and then utilise the log-rank test proposed by Mantel and Haenszel [65] to determine whether $S_1(t)$ and $S_0(\hat{\psi}t)$ is similar to each other or not. Specifically, the timescale factor, $\hat{\psi}$, is defined as a minimum of

$$D(\psi) = \int_0^\infty W(t|\psi)\big\{\hat{H}_0(t) - \hat{H}_1(\psi t)\big\}dt, \tag{2.52}$$

where $\hat{H}_0(t)$ and $\hat{H}_1$ are the corresponding two-sample cumulative hazard functions estimated from the Nelson estimator, and $W(t|\psi)$ is a data-dependent weight function defined by

$$W(t|\psi) = \big\{1 - S(t|\psi)\big\}^{\epsilon}\big\{S(t|\psi)\big\}^{\delta}/t, \tag{2.53}$$

where $\epsilon > 0$ and $\delta > 0$ are arbitrary constants, and $S(t|\psi)$ is the Kaplan-Meier estimator of the combined dataset construct from T_0 and T_1/ψ. Because the function $D(\psi)$ is not smooth, the ordinary Newton-Raphson iteration is not reliable for finding ψ. Consequently, we will utilise the optimisation algorithms for general non-smooth functions (e.g. grid search, random search and simulated annealing algorithms) to locate $\hat{\psi}$. After obtaining an estimate of $\hat{\psi}$, the log-rank test will be utilised to determine whether $S_0(\psi t)$ and $S_1(t)$ have the same distribution or not, and any evidence that these two distributions are not similar is an evidence against the validity of the accelerated failure time assumption. Let $a_1 < \ldots < a_m$ be distinct survival times in the dataset, r_{kj} is the number of individual in view at a_j in group k, and d_{kj} is the number of failures at a_j in group k. The log-rank statistic for the null hypothesis that these two populations come from the same distribution can be written as

$$\left(\frac{\sum_{j=1}^{m}\left(d_{1j} - E_{1j}\right)}{\sqrt{\sum_{j=1}^{m} V_j}}\right)^2 \tag{2.54}$$

where

$$\begin{aligned}
E_{1j} &= \frac{D_j r_{1j}}{R_j}, && (2.55)\\
V_j &= \frac{D_j\big(r_{1j}/R_j\big)\big(1 - r_{1j}/R_j\big)\big(R_j - D_j\big)}{R_j - 1}, && (2.56)\\
R_j &= r_{1j} + r_{2j}, && (2.57)\\
D_j &= d_{1j} + d_{2j}. && (2.58)
\end{aligned}$$

Under the null hypothesis, the distribution of the square root of this statistic is approximately standard normal and hence this statistic is distributed as chi-square with one degree of freedom. This chi-square distribution is then integrated on the right of the statistic value to obtain the p-value, which is equal to the probability of getting a statistic equal to or larger than that observed under the null hypothesis. Consequently, the lower the p-value, the less likely that the null hypothesis is true. One often rejects the null hypothesis if the p-value is less than 0.05 or 0.01, corresponding to 5% or 1% chance of observing an outcome at least that extreme, given the null hypothesis.

Proportional hazards model

Another way to model the relationship between explanatory variables and survival time is to assume that the hazard of each observation is specified by

$$h(t|X) = \psi(X)h_0(t), \tag{2.59}$$

where $h_0(t)$ is an unspecified baseline hazard function which is free of the explanatory variables. Hence, in this assumption, the explanatory variables act multiplicatively on the hazard. At two different point X_1 and X_2, the proportion

$$\frac{h(t|X_1)}{h(t|X_2)} = \frac{\psi(X_1)}{\psi(X_2)}, \tag{2.60}$$

called the *hazard ratio*, is constant with respect to time t. Under this assumption, the survivor function and density are given by

$$\begin{aligned} S(t|X) &= S_0(t)^{\psi(X)}, && (2.61) \\ f(t|X) &= \psi(X)f_0(t)S_0(t)^{\psi(X)-1} && (2.62) \end{aligned}$$

Although $\psi(X)$ fulfills the same role as in the accelerated failure time model, it does not have precisely the same interpretation. Similar to the previous section, $\psi(X)$ can be parameterized as $\psi(X|\beta)$ and the most important special case is again

$$\psi(X|\beta) = e^{\beta^T X} \tag{2.63}$$

Although we can estimate β by maximizing the likelihood obtained from substituting the likelihood (2.16) by some specific forms of baseline hazard functions as in the accelerated failure time model, such choices are unnecessary in the proportional hazards model. To see this, let us firstly consider the situation when all data are uncensored. Let $\tau_1 < \tau_2 < \ldots < \tau_n$ denote the order failure times of the n individuals, ξ_j denote the label of the individual that fails at τ_j so that $\xi_j = i$ only when $t_i = \tau_j$ and $R(\tau_j) = \{i|t_i \geq \tau_j\}$ denote the risk set just before the j-th ordered failure time. The conditional probability that $\xi_j = i$ given the entire history up to the j-th ordered failure time τ_j (i.e. $\mathbb{H}_j = \{\tau_1, \tau_2, \ldots, \tau_j, i_1, i_2, \ldots, i_{j-1}\}$) is basicially the conditional probability that i fails at τ_j given that one individual from the risk set $R(\tau_j)$ fails at τ_j, which equal to

$$\frac{h(\tau_j|X_i)}{\sum_{k\in R(\tau_j} h(\tau_j|X_k)} = \frac{\psi(X_i)h_0(\tau_j)}{\sum_{k\in R(\tau_j)} \psi(X_k)h_0(\tau_j)} = \frac{\psi(X_i)}{\sum_{k\in R(\tau_j)} \psi(X_k)}. \tag{2.64}$$

Note the baseline hazard function $h_0(\tau_j)$ cancelling because of the multiplicative assumption in Equation (2.59). Although Equation (2.64) was derived as the conditional probability that $\xi_j = i$ given the entire history $\mathbb{H}_j$, it is functionally independent of $\tau_1, \tau_2, \ldots, \tau_j$ and depended only on $i_1, i_2, \ldots, i_{j-1}$. Consequently, the joint distribution $\Pr\left\{\xi_1 = i_1, \ldots, \xi_n = i_n\right\}$ over the set of all possible permutations of $(1, 2, \ldots, n)$ can be obtained by the usual chain rule of conditional probabilities as

$$\begin{aligned}\Pr\left\{\xi_1 = i_1, \ldots, \xi_n = i_n\right\} &= \prod_{j=1}^{n} \Pr\left\{\xi_j = i_j | \xi_1 = i_1, \ldots, \xi_{j-1} = i_{j-1}\right\}, \\ &= \prod_{j=1}^{n} \frac{\psi(X_{i_j})}{\sum_{k \in R(\tau_j)} \psi(X_k)}. \end{aligned} \tag{2.65}$$

In the presence of censoring, a similar argument applies when we assume that censoring can only occur immediately after failures. Particularly, suppose that there are d observed failures from the sample of size n, and let the ordered observed failure times be $\tau_1, \tau_2, \ldots, \tau_d$. The likelihood of the dataset can be described by

$$\begin{aligned} L &= \prod_{j=1}^{d} \frac{\psi(X_{i_j})}{\sum_{k \in R(\tau_j)} \psi(X_k)}, \\ &= \prod_{i \in \mathbb{D}} \frac{\psi(X_i)}{\sum_{k \in R(t_i)} \psi(X_k)}, \end{aligned} \tag{2.66}$$

where $\mathbb{D}$ denotes the set of individuals who fail. The corresponding log-likelihood is then

$$l = \sum_{i \in \mathbb{D}} \left(\log\left(\psi(X_i)\right) - \log\left(\sum_{k \in R(t_i)} \psi(X_k) \right) \right) = \sum_{i \in \mathbb{D}} l_i. \tag{2.67}$$

Consequently, we can estimate the value of β by maximizing this likelihood function instead of Equation (2.16). Since this function does not depend on baseline hazard function, the relation between explanatory variables and the survival time can be estimated without any assumption about the baseline hazard function, and this is the reason why the proportional hazards model is a semi-parametric model. When $\psi(X) = e^{\beta^T X}$, this log-likelihood reduces to

$$l(\beta) = \sum_{i \in \mathbb{D}} \left(\beta^T X_i - \log\left(\sum_{k \in R(t_j)} \exp\left(\beta^T X_k\right) \right) \right). \tag{2.68}$$

The derivative of this log-likelihood with respect to the r-th explanatory variable is

$$\begin{aligned} \frac{\partial l}{\partial \beta_r} &= \sum_{i \in \mathbb{D}} \left(X_{ir} - \frac{\sum_{k \in R(t_i)} X_{kr} e^{\beta^T X_k}}{\sum_{k \in R(t_i)} e^{\beta^T X_k}} \right), \\ &= \sum_{i \in \mathbb{D}} \left(X_{ir} - E_{ir}(\beta) \right), \end{aligned} \tag{2.69}$$

where $E_{ir}(\beta) = \frac{\sum_{k \in R(t_i)} X_{kr} e^{\beta^T X_k}}{\sum_{k \in R(t_i)} e^{\beta^T X_k}}$ is the weighted average of the r-th variables over the risk set at time t_i, i.e. $R(t_i)$. Accordingly the maximum likelihood estimator of β_r is a solution of $\frac{\partial l}{\partial \beta_r} = 0$.

Model checking

Similar to the accelerated failure time model, we can assess the validity of the proportional hazards model both by graphical and quantitative methods. From Equation (2.61), $\log(-\log(S(t|X)))$ at various values of X differ only by translation since

$$\begin{aligned} \log(-\log S(t|X))) &= \log\left(-\log\left(S_0(t)^{\psi(X)}\right)\right), \\ &= \log(-\psi(X)\log(S_0(t))), \\ &= \log(\psi(X)) + \log(-S_0(t)). \end{aligned} \tag{2.70}$$

Consequently, we can verify this assumption in a two-sample problem by plotting $\log(-\log(S(t)))$ versus t for the two sample on the same chart. If the assumption is satisfied, these two curves should be parallel to each other since they differ only by translation, and thus the evidence against this is also an indication of the violation of this assumption.

To assess this assumption quantitatively, Grambsch and Therneau [36] developed a test for proportional hazard assumption by considering the time-varying coefficients

$$\beta(t) = \beta + \theta g(t), \tag{2.71}$$

where $g(t)$ is a predictable processes. Under the proportional hazards assumption, $\beta(t)$ should be independent of time and, hence, θ should be equal to zero. Consequently, they developed a score test for a null hypothesis that $\theta = 0$ based on the generalised least squares estimator of θ. In particular, they define

$$S^{(r)}(\beta, t) = \sum_{i \in R(t)} \exp\left\{\beta^T X_i(t)\right\} X_i(t)^{\otimes r} \quad \text{for } r = 0, 1, 2 \tag{2.72}$$

where, for a column vector a, $a^{\otimes 2}$ denotes the outer product aa^T, $a^{\otimes 1}$ denotes the vector a and $a^{\otimes 0}$ denotes the scalar 1. Grambsch and Therneau illustrated that the expectation of the Schoenfeld residual [85], defined as $r_i(\beta) = X_i - M(\beta, t)$ is characterised by

$$E\{r_i(\beta)\} = V(\beta, t_i)G(t_i)\theta, \tag{2.73}$$

where $G(t_i)$ is a diagonal matrix with kk element equal to $g_k(t_i)$, and

$$\begin{aligned} M(\beta, t) &= S^{(1)}(\beta, t)/S^{(0)}(\beta, t), \\ V(\beta, t) &= \frac{S^{(2)}(\beta, t)}{S^{(0)}(\beta, t)} - \left\{\frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)}\right\}^{\otimes 2}. \end{aligned}$$

Consequently, define $r_i^*(\beta) = V^{-1}(\beta, t_i)r_i(\beta)$ as the scaled Schoenfeld residual so that $E\{r_i^*(\beta)\} = G(t_i)\theta$. Let $\hat{\beta}$ be the maximum partial likelihood estimate under the null hypothesis that $\theta = 0$ from the likelihood in Equation (2.68), $\hat{V}(t) = V(\hat{\beta}, t)$ and $\hat{r}_i = r_i(\hat{\beta})$. Generalized least squares gives

$$\hat{\theta} = D^{-1} \sum_{i \in \mathbb{D}} G(t_i) \hat{r}_i, \tag{2.74}$$

with

$$D = \sum_{i \in \mathbb{D}} G(t_i) \hat{V}(t_i) G(t_i)^T - \left(\sum_{i \in \mathbb{D}} G(t_i) \hat{V}(t_i) \right) \left(\sum_{i \in \mathbb{D}} \hat{V}(t_i) \right)^{-1} \left(\sum_{i \in \mathbb{D}} G(t_i) \hat{V}(t_i) \right)^T . \tag{2.75}$$

Under the null hypothesis, the asymptotic variance of $n^{-1/2} \sum G(t_i) \hat{r}_i$ can be consistently estimated by $n^{-1} D$, leading to an asymptotic χ^2 test statistic on p degree of freedom:

$$T(G) = \left(\sum_{i \in \mathbb{D}} G(t_i) \hat{r}_i \right)^T D^{-1} \left(\sum_{i \in \mathbb{D}} G(t_i) \hat{r}_i \right), \tag{2.76}$$

where p is the number of explanatory variables.

2.5 Summary

This chapter gives an overview of a number of key concepts in the areas that this research will span. In particular, the chapter starts with a brief review of trading mechanisms utilised in major financial markets around the world. It then introduces the definition and components of an algorithmic trading system with emphasis on trading execution, which is the major problem studied in this research. Finally, it introduces survival analysis as a statistical modelling technique for modelling the execution time of limit orders.

Chapter 3

Simulation model of a pure double auction market

This chapter gives an overview of the simulation models employed for studying the behaviour of the execution probability and assessing the predictive performance of the execution probability model studied in the subsequence chapters. The models employed here are models of agent behaviour in continuous double auction markets that contain two main types of agents (i.e. impatient and patient agent). Impatient agents place market orders randomly according to some predefined stochastic process, while patient agents place limit orders randomly both in time and in price. Additionally, unexecuted limit orders are assumed to be cancelled according to some predefined stochastic processes. By controlling the properties of these order submission and cancellation processes, several realisations of the order book dynamic that have similar stochastic properties can be generated. This enables us to study the properties of the execution probability models in a controlled environment before applying them to the data generated from real markets.

This chapter is organised as follows. Section 3.1 describes the characteristics of the continuous double auction mechanism. Previous models of double auction markets are summarised in Section 3.2. Section 3.3 describes a stylised model of the dynamics of a double auction market which can be utilised to implement several previous models. The implementation of the proposed model is described in Section 3.4, while examples of result obtained from the proposed model is presented and analysed in Section 3.5. Finally, we end the chapter with a conclusion in Section 3.6.

3.1 The continuous double auction

The continuous double auction is the most widely used trading mechanism employed by major financial markets around the world. This type of market has gained popularity in recent years over *quote-driven* markets where liquidity is provided by market makers or designated dealers. Examples of such equity markets include the Electronic Communication Networks in the United States, the Toronto Stock Exchange, the Stockholm Stock Exchange, the Australian Stock Exchange, the Shanghai Stock Exchange and the Tokyo Stock Exchange. Order-driven markets for derivative instruments have also gained popu-

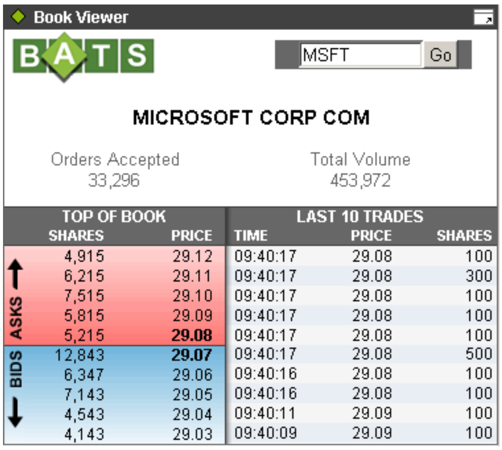


Figure 3.1: Example of a limit order book, for the Microsoft Corporation. The best bid and offer are respectively to buy 12,843 shares at 29.07$ and sell 5,215 shares at 29.08$. Snapshot from Batstrading.com

larity in recent years over the traditional open-outcry auctions, and many derivative exchanges, including the Chicago Mercantile Exchange, the International Petroleum Exchange of London, the Sydney Futures Exchange, and the Hong Kong Futures Exchange, are nowadays organised in this fashion.

Continuous double auction markets are generally characterised by the presence of a limit order book, where unexecuted and partially executed orders are stored and wait for future execution. The limit order book normally consists of two queues: the bid side for buy orders and the ask (or offer) side for sell orders. The highest buy price at a particular time is called *the best bid*, while the lowest sell price is called *the best ask*. The difference between the best ask and the best bid is called *the bid-ask spread*, while their average is called *the mid price*. The quantity of limit orders at the best price is sometimes called *the depth* of the market. Orders queued in the book are generally sorted by price, time of arrival, and volume, with variation from market to market. Limit buy and sell orders are entered into the book by market participants throughout a trading day, with prices and sizes of their choice. When a new buy (respectively sell) orders reaches the book, it either triggers a trade if its limit price is higher than the best ask (respectively lower than the best bid), or stored in the book. A trade in this market also is triggered when a trader submits a market order, an unpriced order which is executed immediately against the best order. If the order quantity is larger than the quantity available at the best price in the book, the order will walk the book, resulting in partial executions at progressively worse prices until the order is fully executed. On some exchanges, however, market orders are implemented via limit orders priced for immediate execution, which are known as *marketable limit orders*.

A real world example of the order book is illustrated in Figure 3.1, for Microsoft Corporation stock, listed on the BATS Exchange. Notice the two queues utilised as repositories for the outstanding limit orders. Orders in this market are sorted and aggregated by price level. At each price level, orders are then sorted by the arrival time, with the oldest orders given priority. Other useful indicators include the book depth, the last trade executed, with its time and price, the number of the order submitted to the exchange, and the total volume of orders executed so far during the day.

3.2 Previous work

With the world wide proliferation of continuous double auction markets, various studies have focused on modelling the dynamic behaviour in these markets with the aim of providing more insight into price formation and the stochastic properties of price fluctuations. This research can be classified into two main approaches. The first approach tries to model this dynamic with the interaction between heterogeneous agents who trade against each other in a continuous double auction market. These agents are normally assumed to act for their own best interest and place their orders individually, according to some predefined trading strategies which can be ranging from fundamentalist, chartist and noise traders. Although many studies [34, 18] indicate that such models exhibit similar features to real markets, the parameters of these models are generally difficult to estimate since they normally contain unobservable parameters such as the true value of the asset price and the distribution of the traders, i.e. the number of traders in each category. Examples of research along this line are Parlour [76], Chiarella and Iori [18], Foucault et al. [33], Ghoulmie et al. [34] and Rosu [83]. Whilst traders participating in the market may make their decision in an extremely complex environment, the end result of these decisions is reduced to the simple action of placing and cancelling trading orders. Consequently, instead of attempting to anticipate how traders will behave, the second approach starts by assuming that their combined effect is to generate flows of order submission and cancellation with known distributions for limit price and size, and then determines the quantities of interest based on this assumption. The advantage of this approach is that all model parameters can be directly estimated from historical data while still be able to generate several stylized facts observed in the real markets [30, 69]. Example of research along this line includes Domowitz and Wang [23], Luckock [63], Smith et al. [88], Mike and Farmer [69] and Const et al. [20].

Apart from the above classification, we can also classify these models into discrete-time models and continuous-time models. In discrete-time model, agents normally make their decision only at discrete point in time. This includes turn-based models where agents take turns to make their decision (e.g. Parlour [76] and Foucault [32]) and models where decisions are modelled in event time not calendar time (e.g. Ghoulmie et al. [34], Preis et al. [79] and Mike and Farmer [69]). On the other hand, continuous-time models generally involve event-driven models, where agents sleep after performing actions and then wake up at a predefined time or as a result of certain events (e.g. their orders get filled or new information is arrival in the market). Examples of such models include Domowitz and Wang [23], Smith et al. [88], Luckock [63] and Const et al. [20].

Since financial markets operate in continuous time, it is more appropriate to model them in continuous time rather than in discrete time. Consequently, the model utilised in this study will be a continuous-time aggregated order flow model. More detail about this model will be given in the next section.

3.3 The model

The simulation model utilised in this study is adapted from limit-order book models of Cont et al. [20], Preis et al. [79] and Smith et al. [88]. The model consists of a market where limit orders are placed on an integer price grid $p \in \{1, ..., n\}$ where p represents a multiple of a price tick. Using notations similar to the one utilised by Const et al. [20], the state of the order book at a particular time t will be represented by $X(t) \equiv (X_1(t), ..., X_n(t))_{t \geq 0}$, where $|X_p(t)|$ is the number of unexecuted limit orders at price p, $1 \leq p \leq n$, and the sign of $X_p(t)$ indicates the side of the orders; particularly, there will be $X_p(t)$ sell orders at price p when $X_p(t)$ is positive, while there will be $-X_p(t)$ buy orders at price p when $X_p(t)$ is negative (see Figure 3.2 for example). Using this notation, the *best ask price*, $p_A(t)$, which is the lowest selling price offered at a particular time t, can be defined by

$$p_A(t) \equiv \inf \{p = 1, \ldots, n \mid X_p(t) > 0\} \wedge (n + 1). \tag{3.1}$$

Similarly, the *best bid price*, $p_B(t)$ which is the highest buying price at a particular time t, can be defined by

$$p_B(t) \equiv \sup \{p = 1, \ldots, n \mid X_p(t) < 0\} \vee 0. \tag{3.2}$$

Notice that, when there are no sell orders in the book, the best ask is forced to be $n + 1$, while the best bid is forced to be zero when there are no buy orders in the book. From the definition of the best bid and the best ask, the *bid-ask spread*, $s(t)$ which measure the gap between the best bid price and the best ask price can be defined by

$$s(t) = p_A(t) - p_B(t). \tag{3.3}$$

Accordingly, the *mid-price*, $p_M(t)$, which is the average between the best bid price and the best ask price can be defined by

$$p_M(t) = \frac{p_A(t) + p_B(t)}{2} \tag{3.4}$$

The dynamics of the order book are assumed to be driven by two different types of agents who place orders randomly according to independent Poisson processes. Impatient agents place market buy orders and market sell orders randomly with an independent Poisson rate of μ_B and μ_A shares per unit time respectively. On the other hand, patient agents place limit orders randomly both in time and in price.

BIDS		ASKS	
SHARES	PRICE	PRICE	SHARES
23,661	50	52	25,700
24,610	49	53	65,784
65,400	48	54	11,704

p	...	48	49	50	51	52	53	54	...
Xp	...	-65,400	-24,610	-23,661	0	25,700	65,784	11,704	...

Figure 3.2: Example of limit order book together with the corresponding $X_p(t)$ where buy orders are represented by negative number while sell orders are represented by positive number.

Parameter	Description	Dimensions
μ_B	arrival rate of market buy orders	shares/time
μ_A	arrival rate of market sell orders	shares/time
$\alpha_B(i)$	arrival rate of limit buy orders i ticks away from the best ask	shares/time
$\alpha_A(i)$	arrival rate of limit sell orders i ticks away from the best bid	shares/time
$\delta_B(i)$	cancelation rate of buy orders i ticks away from the best ask	1/time
$\delta_A(i)$	cancelation rate of sell orders i ticks away from the best bid	1/time

Table 3.1: The six parameters characterising the proposed simulation model. Market buy and sell orders arrive at an exponential time with rate of μ_B and μ_A respectively. Limit buy (sell) orders at a distance of i ticks from the opposite best price arrive at an exponential time with rate of $\alpha_B(i)$ ($\alpha_A(i)$). Outstanding buy (sell) orders at a distance of i ticks from the opposite best price are cancelled with rate of $\delta_B(i)$ ($\delta_A(i)$).

Limit buy orders are placed in the interval $1 \leq p < p_A(t)$, and the limit buy orders at the distance of $i = p_A(t) - p$ ticks from the best ask price is assumed to arrive with a Poisson rate of $\alpha_B(i)$ shares per unit time. Similarly, limit sell orders, which must be placed in the interval $p_B(t) < p \leq n$, at the distance of $i = p - p_B(t)$ ticks from the best bid price are assumed to arrive with a Poisson rate of $\alpha_A(i)$ shares per unit time. Queued limit buy and sell orders are cancelled according to a Poisson process with a rate of $\delta_B(i)$ and $\delta_A(i)$ per unit time depending on the distance between their limit price and the opposite best quote i. Assuming that all orders are of unit size, the order book process $X(t)$ is a continuous-time Markov chain with state space $\mathbb{Z}^n$ and transition rates given by:

$$
\begin{array}{llll}
x_p \to x_p - 1 & \text{with rate} & \alpha_B(p_A(t) - p) & \text{for} \quad p < p_A(t), \\
x_p \to x_p + 1 & \text{with rate} & \alpha_A(p - p_B(t)) & \text{for} \quad p > p_B(t), \\
x_{p_A(t)} \to x_{p_A(t)} - 1 & \text{with rate} & \mu_B & \\
x_{p_B(t)} \to x_{p_B(t)} + 1 & \text{with rate} & \mu_A & \\
x_p \to x_p + 1 & \text{with rate} & \delta_B(p_A(t) - p)|x_p(t)| & \text{for} \quad p < p_A(t), \\
x_p \to x_p - 1 & \text{with rate} & \delta_A(p - p_B(t))|x_p(t)| & \text{for} \quad p > p_B(t),
\end{array}
$$

Consequently, the dynamic behaviour and statistical properties of this Markov chain are completely specified by the six parameters characterising the model as summarised in Table 3.1. By controlling these parameters, this order book model can be utilised to implement several previous order book models, such as those proposed by Smith et al. [88], and Cont et al [20]. The information about these models, together with parameters for implementing them, will be briefly described in Section 3.5.

3.4 Model implementation

The order book model proposed in the previous section is implemented using a discrete-event simulation which is based on an event-oriented approach[1]. This approach generally consists of a main subroutine and separate subroutines for each event type. The main subroutine is responsible for maintaining an event list and handling each event in order of increasing time by calling the corresponding subroutines. Since result of an event may alter the schedule of other events (e.g. the cancellation time of an outstanding

[1]Although it is tempted to simulate the proposed model directly from a continuous-time Markov chain representation, such approach will not provide information about each individual order which is one of the main information for investigating the property of execution probability.

order needs to be rescheduled when the best price changes), the main subroutine also needs to reschedule the affected events when necessary.

The pseudo-code for the main subroutine for implementing the proposed order book model is illustrated in Figure 3.3. Firstly, the order book is initialised with some arbitrary initial conditions[2]. Then, the event list is initialised. This list contains information about event types and event times (the times at which the events will occur). The events considered are:

- The submission of a market buy order, which arrives at an exponential rate of μ_B.
- The submission of a market sell order, which arrives at an exponential rate of μ_A.
- The submission of a limit buy order at i tick below the best ask price, which arrives at an exponential rate of $\alpha_B(i)$, for $1 \leq i \leq L$.
- The submission of a limit sell order at i tick below the best ask price, which arrives at an exponential rate of $\alpha_A(i)$, for $1 \leq i \leq L$.
- Cancellation of every outstanding limit buy order at an exponential time with a rate of $\delta_B(i)$, where i is the difference between the best ask price and their limit price.
- Cancellation of every outstanding limit sell order at an exponential time with a rate of $\delta_A(i)$, where i is the difference between their limit price and the best bid price.

Although impatient agents in real markets can place limit orders at any price they want, it is clearly impossible to simulate order arrivals at every price level. A reasonable simplification is to consider only order arrivals and cancellations in a moving band of price levels centred around current best price. This is confirmed in recent empirical results (e.g. Mike and Farmer [69]) which indicate that most trading activities occurred around the mid price. Accordingly, we will consider only orders arrivals of up to L ticks from the opposite best quote ($1 \leq i \leq L$). After the event list is initialised, the main subroutine then moves on to the main simulation process which will continue until stopping conditions are met. For example, we may stop the simulation when time reaches or exceeds some certain point, or once the order book reaches some particular states (e.g. the mid-price move for a certain range). The main simulation process consists of four main steps: i) collecting statistics from the current order book state, ii) removing the first event from the event list iii) calling the associated subroutine to handle the event, and iv) rescheduling the affected events if necessary. The last step is generally involved with the rescheduling of order cancellation events when the best price changes. In particular, if the best bid price changes at time t, the cancellation of all outstanding sell orders will be rescheduled to $t + e$, with $e \sim \text{Exp}(\delta_A(i))$, while the cancellation of all outstanding buy orders will be rescheduled to $t + e$, with $e \sim \text{Exp}(\delta_B(i))$, when the best ask price changes.

Unlike the main subroutine, the role of event subroutines is to update the system state and to schedule new events into the event list. Accordingly, each event subroutine will consists of two main steps,

[2]The initial state of the book is not important as long as we wait a sufficient length of time. For most of the simulations studied here we choose the initial book so that there are ten orders on the best bid, and ten orders on the best ask, and ran the simulation for 100,000 iterations before sampling.

```
1    Initialise the order book with some arbitrary orders
2    Initialise the event list
3    While (simulation is not finished)
         3.1    Collect statistics from current order book state
         3.2    Remove the first event from the event list
         3.3    Call the associated subroutine to handle it
         3.4    Reschedule the affected events if necessary
```

Figure 3.3: Overview of the main subroutine for discrete-event simulation whose main responsibility is to maintain an event list and handling each event in order of increasing time by calling the corresponding subroutine as well as collect the require statistics from the state of the order book.

which are: i) handling the event and ii) inserting new events into the event list. For example, the event subroutine for the market buy order submission event at time t will submit a market buy order to the order book, which will result in the execution at the best ask price, and schedules the next market buy order submission events at time $t + e$, with $e \sim \text{Exp}(\mu_B)$. The subroutine for the submission of limit buy orders at i ticks below the best ask price at time t will firstly submit the limit buy order at i tick below the best ask price, and then schedules the next limit buy order submission event at time $t + e$ with $e \sim \text{Exp}(\alpha_B(i))$ as well as scheduling the cancellation event of the submitted order at time $t + e$ with $e \sim \text{Exp}(\delta_B(i))$. Unlike other events, the subroutine for order cancellation will only cancel the associated order from the order book without scheduling any new events. The subroutines for market and limit sell order submission operate in a manner similar to the corresponding order arrival rate.

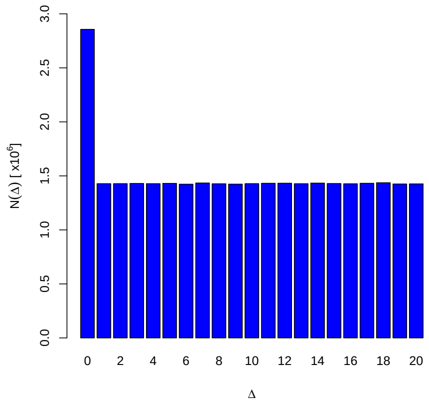
3.5 Simulation results

This section illustrates several ways in which the proposed model can be utilised to implement several previous order book models such as those proposed by Smith et al. [88] and Cont et al. [20]. The simulation results produced by these models are also analysed to provide more insight into the properties of each model.

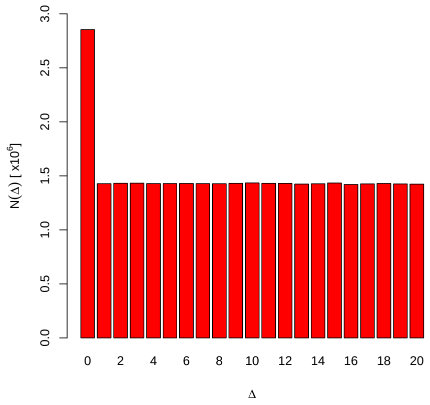
3.5.1 The SFGK model

The simplest model considered here is proposed by Smith, Farmer, Gillemot and Krishnamurthy [88], hereinafter referred to as the SFGK model. In this model, market orders arrive at a rate of μ shares per unit time with equal probability to be a buy and sell order; thus, the rate at which buy and sell orders arrive individually is $\mu/2$. Limit orders at each price level arrive at a rate of α shares per unit price and per unit time for both buy and sell orders. Queued limit orders are assumed to be removed randomly with constant probability of δ per unit time. Thus, the model has three parameters which are: the market order arrival rate, μ, the limit order arrival rate per unit price, α, and the rate of limit order decays, δ. This model can be easily implemented in our framework by setting $\mu_A = \mu_B = \mu/2$, $\alpha_A(i) = \alpha_B(i) = \alpha$ and $\delta_A(i) = \delta_B(i) = \delta$.

As an example, let us analyse a simulation run produced from the SFGK model with parameters $\mu = 2$, $\alpha = 0.5$, $\delta = 0.025$ and $L = 20$. The result reported here is generated by initialising the order

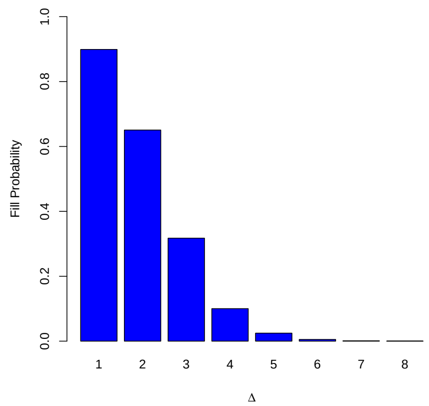


(a) Number of buy orders submitted

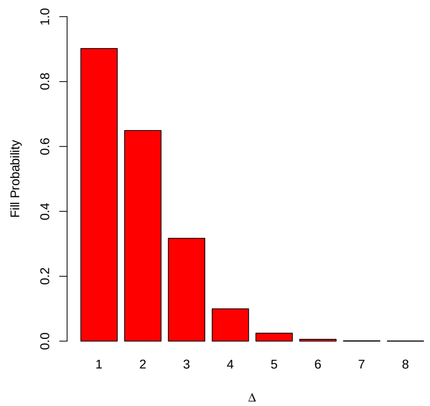


(b) Number of sell orders submitted

Figure 3.4: The number of buy orders (a) and sell orders (b) submitted to the market as a function of the distance from the opposite best price during a simulation run of the SFGK model.



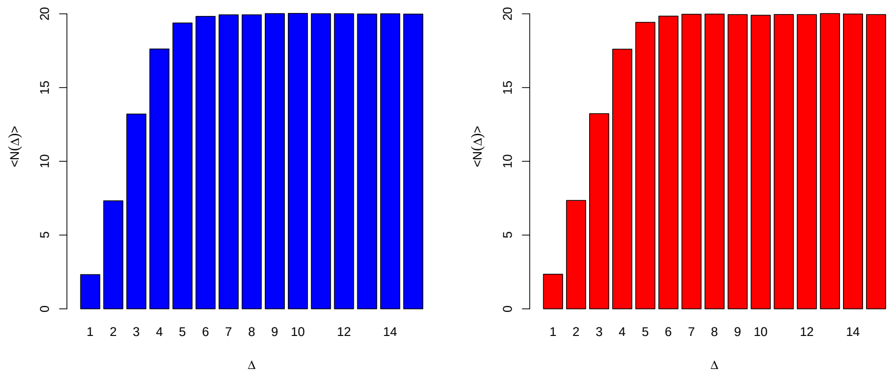
(a) Fill probability of submitted buy orders



(b) Fill probability of submitted sell orders

Figure 3.5: The fill probability of buy orders (a) and sell orders (b) as a function of the distance from the opposite best price generated from a simulation run of the SFGK model.

book so that the bid-ask spread is equal to one tick with ten orders at the best bid and ten orders at the best ask, and running the simulation for 1×10^7 events after a burn-in period of 2×10^6 events. During the sampling period, a total of $6,285,499$ orders were sent to the order book. A fraction of $f_b = 50.007\%$ were buy orders and a fraction of $f_s = 49.993\%$ were sell orders. This is inline with the assumption of the model which assumes that the order arrival and cancellation rate of buy and sell orders are equal. The number of orders submitted at each price level and the fill probability, or the probability that the orders are executed before they are cancelled, are also illustrated in Figure 3.4 and 3.5 respectively. As expected from a constant order arrival rate, the number of limit orders submitted at each price level is roughly the same both for buy and sell orders. Conversely, the fill probability is a decreasing function of the distance from the opposite best price with only the first two levels greater than fifty percent and only the first five levels greater than one percent. This indicates that the orders far away from the opposite



(a) Average book depth of buy orders

(b) Average book depth of sell orders

Figure 3.6: Average book depth as a function of the distance from the opposite best price, for the buy (a) and sell (b) orders generated from a simulation run of the SFGK model.

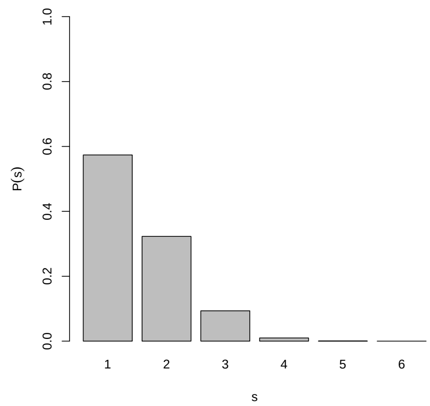
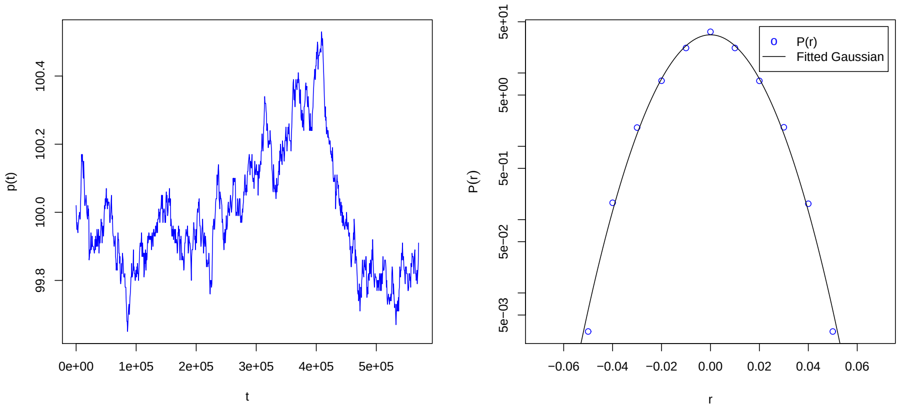


Figure 3.7: The probability mass function of the bid-ask spread generated from a simulation run of the SFGK model.

best price are generally cancelled before they get executed, and, thus, it may not be appropriate to utilise them to model the execution probability which requires the execution time of each order. More detail about this issue will be further analysed in Section 4.4.

The average book depth for buy and sell orders as a function of the distance from the opposite best price is illustrated in Figure 3.6. Although the limit order arrival rate at each price level is constant, the average book depth at each price level is not. Actually, the average book depth is an increasing function of the distance from the opposite best price with the maximum volume converging to a theoretical value computed from $\alpha/\delta = 0.5/0.025 = 20$. The deviation from this value in the first few ticks from the opposite best price results from the arrival of market orders that keeps removing the orders near the opposite best price. It is this interaction that makes the maximum volume far away from the opposite best price, as observed in real markets.

The distribution of the bid-ask spread, the difference between the best ask price and the best bid



(a) Price

(b) Probability density function of return $P(r)$

Figure 3.8: Price trajectory (a) and the corresponding distribution of returns (b) generated from a simulation run of the SFGK model. The distribution of return $P(r)$ plotted against a fitted Gaussian indicates that the return can well be approximated by a Gaussian distribution with a mean of -1.7512×10^{-7} and a standard deviation of 1.2023×10^{-2}.

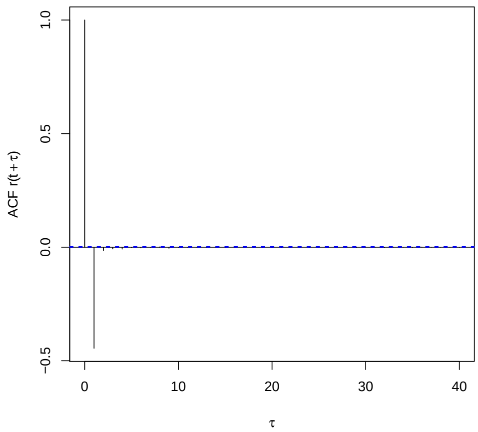


Figure 3.9: Autocorrelation function of the tick-by-tick returns generated from a simulation run of the SFGK model. After a few trades corresponding to the bid-ask bounce, successive returns do not exhibit any correlation.

price, which is a part of the transaction cost the traders will incur when they submit a marketable order, is illustrated in Figure 3.7. The distribution is extremely skewed, as reported in real markets. This is an emergent property of the interplay between the order flow and the order book. In particular, the spread can be affected by the cancellation of orders at the best price, the submission of market orders which remove the order at the best price and the submission of limit orders inside the spread.

Figure 3.8 displays the tick-by-tick price and the corresponding return distribution. Although the price trajectory look comparable to the one generally observed in real markets, the return distribution exhibit no fat tails but can rather well be approximated by a Gaussian distribution. The autocorrelation function of tick-by-tick returns has negative first-order autocorrelation as can be observed in real markets.

As illustrated in Figure 3.9, after a few negative values corresponding to the bid-ask bounce, we observe a quick convergence to the noise level which indicates that the raw returns are memoryless, as expected from our model.

Although the dynamics generated from this model do not exhibit features observed in real markets (e.g. the return distribution does not exhibit a fat tails), the fact that the model depended only on three parameters makes it the best candidate for analysing the relation between the order arrival/cancellation rate and the execution probability which will be investigated in Section 4.4.

3.5.2 The CST model

Another model considered here is the model of Cont, Stoikov and Talreja (CST) [20]. Similar to the SFGK model, the CST model is a symmetrical model, where the rate of order arrival and cancelation of buy and sell orders is the same. In this model, market orders are assumed to arrive at an independent exponential time with rate μ. Limit orders at a distance of i ticks from the opposite best price arrive at an independent exponential time with rate $\lambda(i)$, which is assumed to follow a power law function of the form

$$\lambda(i) = \frac{k}{i^{\alpha}}$$

as suggested by Bouchaud et al. [14] and Zovko and Farmer [95]. A queued limit order at a distance of i ticks from the opposite best price is cancelled at an independent exponential time with rate $\theta(i)$ which does not have a functional form but can be directly estimated from time-stamped sequences of trades and quotes using the equation

$$\hat{\theta}(i) = \frac{N_c(i)}{TQ_i}\frac{S_c}{S_l},$$

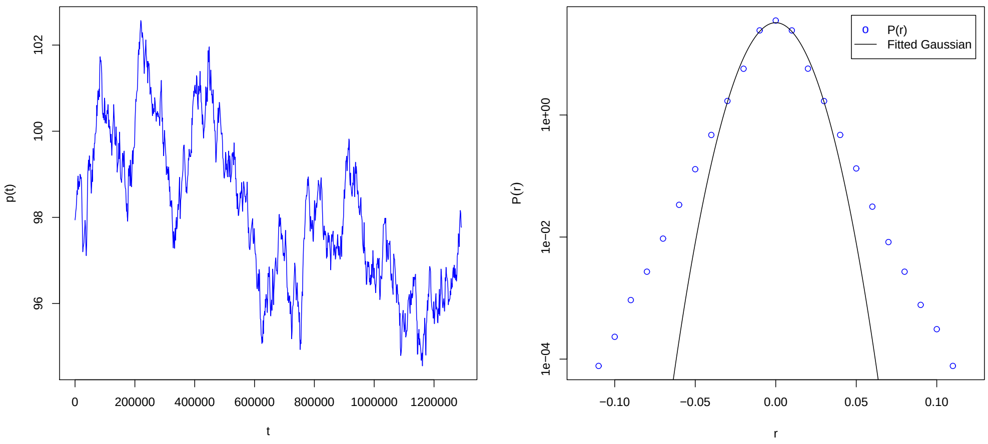
where $N_c(i)$ is the number of times that a quote at distance i decreases in size by cancellation, T is the length of the sample, Q_i is the average number of orders at a distance of i ticks from the opposite best quote, S_c is the average size of cancelled orders, and S_l the average size of limit order. Thus, the model has four parameters which are μ, the market order arrival rate; $\theta(i)$ order cancellation rate; k and α, which specifies the limit order arrival rate. One can implement this model in our framework by setting $\mu_A = \mu_B = \mu$, $\alpha_A(i) = \alpha_B(i) = \lambda(i) = \frac{k}{i^{\alpha}}$, and $\delta_A(i) = \delta_B(i) = \theta(i)$.

As an example, let's consider the price trajectory generated from 4×10^7 events of a simulation run with parameters $\mu = 0.91$, $k = 1.92$, $\alpha = 0.52$, $L = 20$ and $\theta(i)$, as illustrated in Table 3.2. Figure 3.10 displays the tick-by-tick price and the corresponding return distribution. The figure illustrates that the price trajectory and return distribution look comparable to the one generally observed in real market as it exhibits clear fat tails.

Like the SFGK model, the autocorrelation function of tick-by-tick returns also has negative first-order autocorrelation as can be observed in real markets as displayed in Figure 3.11. The distributions

	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i \geq 5$
$\theta(i)$	0.71	0.81	0.68	0.56	0.47

Table 3.2: Parameters of $\theta(i)$ used in the simulation of the CST model.



(a) Price **(b)** Probability density function of return $P(r)$

Figure 3.10: Price trajectory (a) and the corresponding distribution of returns (b) generated from 4×10^7 events of a simulation run of the CST model with parameters $\mu = 0.91$, $k = 1.92$, $\alpha = 0.52$, $L = 20$ and $\theta(i)$ as illustrated in Table 3.2. The distribution of return $P(r)$ plotted against a fitted Gaussian clearly exhibit the fat tails.

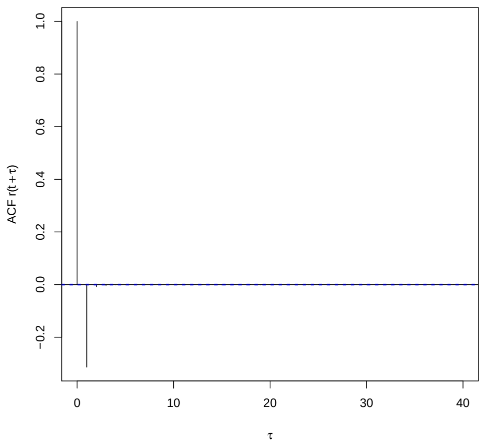


Figure 3.11: Autocorrelation function of the tick-by-tick returns generated from 4×10^7 events of a simulation run of the CST model. After a few trades corresponding to the bid-ask bounce, successive returns do not exhibit any correlation.

of the spread and the average book depth are illustrated in Figure 3.12 and Figure 3.13 respectively. Compared to the SFGK model, the results obtained from the CST model are more realistic than the SFGK model.

3.6 Summary

This chapter presents an overview of the simulation models employed for studying the behaviour of the execution probability as well as assessing the prediction performance of the execution probability studied in the subsequent chapters. The models employed here are models of agent behaviour in continuous

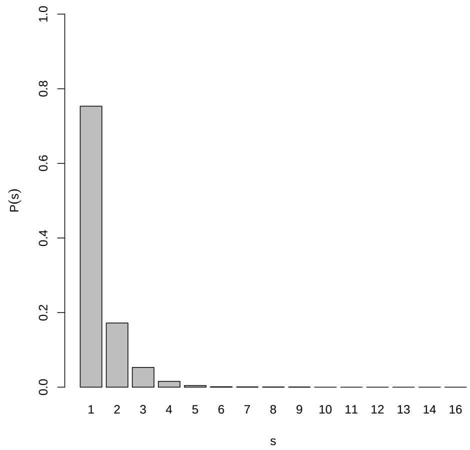
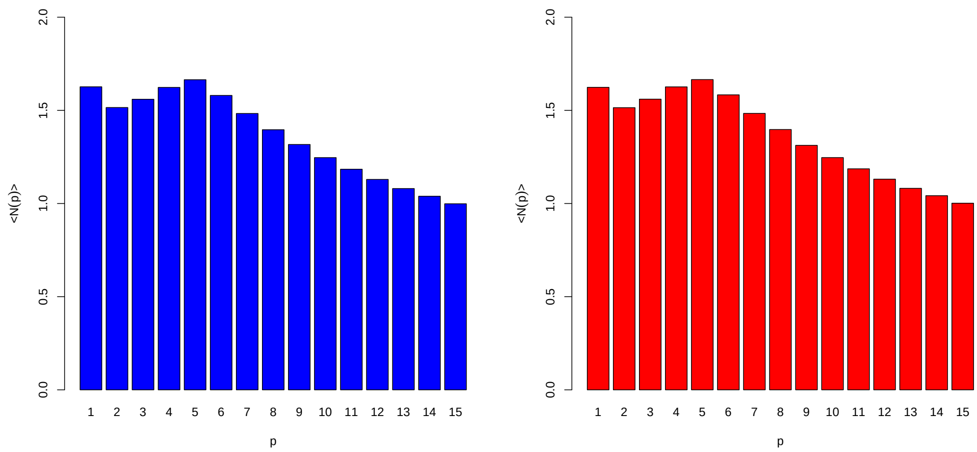


Figure 3.12: The probability mass function of the bid-ask spread generated from 4×10^7 events of a simulation run of the CST model.



(a) Buy orders **(b)** Sell orders

Figure 3.13: Average book depth as a function of the distance from the opposite best price, for the buy (a) and sell (b) orders generated from 4×10^7 events of a simulation run of the CST model.

double auction markets that contain two main types of agents (i.e. impatient agent and patient agent). Impatient agents place market orders randomly according to some predefined stochastic process, while patient agents place limit orders randomly both in time and in price. Additionally, unexecuted limit orders are assumed to be cancelled according to some predefined stochastic processes. By controlling the properties of these orders submission and cancellation processes, several realisations of the order book dynamic that have similar stochastic properties can be generated. This enables us to evaluate the developed model in a controlled environment before applying them to the data generated from real markets.

Chapter 4

Execution probability model

This chapter presents an in-depth review of execution probability models together with performance comparison in a controlled environment based on the data generated from the simulation model of an order-driven market presented in the previous section. The results indicate that among all models considered, the execution time model that utilises techniques from survival analysis to handle cancelled orders is the best performing methods both form theoretical and empirical point of views. However, the experiment in applying survival analysis techniques to model the determinants of the execution probability indicates that traditional techniques which are the proportional hazards model and accelerated failure time model are not flexible enough to model this probability.

4.1 Introduction

A model of execution probability that can be utilised to estimate the probability of execution (i.e. a probability that a given limit order will be executed in a specified period of time) is one of the most important components for determining an appropriate order type for a trader in an order-driven market since the expected payoff the trader gets from submitting a limit order is largely depend on this probability. Although several methods have been proposed to model this probability; we believe that they have several limitations that prevent their use in real situations. This chapter presents a recent review of the methods for modelling this probability, together with experiments to compare the advantages and disadvantages of each method in a controlled environment, by utilising the data generated from simulation models of order-driven markets described in the previous chapter. The limitation of each method is also highlighted in order to guide the direction for further developments. Additionally, we also investigate the effect of explanatory variables such as the bid-ask spread, the market depth and the order arrival/cancel trade to the execution probability.

The rest of this chapter is organised as follows: Section 4.2 gives a basic definition of probability execution and related variables. Section 4.3 reviews several ways in which execution probability can be estimated. The experiment comparing the performance of each method in a controlled environment is analysed in Section 4.4. Section 4.5 analyses the relationship between the execution probability and other variables, while Section 4.6 investigates the best parametric distribution for modelling the execution probability, Finally, the chapter ends with a conclusion in Section 4.7

4.2 Overview of execution probability

After submission, a limit order will stay in the order book until it is cancelled or traded. At each time instant, each limit order can be in one of three possible states, which are: fully executed, partially executed and unexecuted. For a fully executed order, the elapsed time between its placement and its complete execution is called *time-to-fill (TTF)*. Since orders are usually not fully executed by a single transaction, one can separately define *time-to-first-fill (TTFF)*[1] as the time from order placement to the first transaction this order participates in. Additionally, for cancelled orders, the time between order placement and its cancellation can also be defined as *time-to-cancel (TTC)*. Theoretically, a limit order will be executed only when enough market orders arrive during the remainder of the trading horizon to execute all preceding orders in the order book; thus, the probability of execution depends on both the state of the book (e.g. book depths and spread) when traders submit their orders and the future incoming flow of market and limit orders. Hence, a good execution probability model should incorporate this information into consideration. To model the probability of execution, one can either directly model the probability distribution of those three outcomes or model the probability of execution time and then utilise it to derive the execution probability. Generally, the relationships between the probability of these outcomes and the execution time distribution can be described by:

$$P_{FE}(t) = Pr\left\{TTF \leq t\right\}, \tag{4.1}$$

$$P_{UE}(t) = 1 - Pr\left\{TTFF \leq t\right\}, \tag{4.2}$$

$$P_{PE}(t) = Pr\left\{TTFF \leq t\right\} - Pr\left\{TTF \leq t\right\}, \tag{4.3}$$

where $P_{FE}(t)$, $P_{PE}(t)$, and $P_{UE}(t)$ represent the probability that the order is fully executed, partially executed and unexecuted at time t. Equation (4.1) and (4.2) are directly derived from the definition of fully executed and unexecuted. Equation (4.3) is derived from the fact that the summation of the probability of these outcomes must equal to one, and, thus, $P_{PE}(t) = 1 - P_{FE}(t) - P_{UE}(t)$. When time-to-fill and time-to-first-fill are equal[2], $P_{PE}(t)$ will always equal to zero and the probability that the order is unexecuted is reduced to $P_{UE}(t) = 1 - P_{FE}(t)$. Hence, knowing only $P_{FE}(t)$, the probability that the order is fully executed with in a specific time t, is enough to derive other interested quantities, and, throughout the rest of this study, we will refer to this probability as a probability of execution, unless stated otherwise.

4.3 Previous work

Previous models for modelling the probability of execution can be categorised into two main categories which are i) execution probability models, which model the execution probability directly, and ii) execution time models, which model the execution time and utilise it to calculate the execution probability.

[1] Note that time-to-first-fill is always smaller or equal to time-to-fill.

[2] This is usually be the case for small order which can be fully executed in one transaction

4.3.1 Execution probability model

This section describes methods that model the execution probability directly. These methods can be classified into two main categories, which are theoretical models and empirical models.

Theoretical models

Since the execution probability depends on future traders' order submission, an order submission model that describes traders' behaviour can be utilised to determine the execution probability. Given the model of the market together with the distribution of traders and their valuation, the execution probability in an equilibrium[3] can be derived (see [76, 32, 33] for examples). Specifically, assuming that all traders in the market are rational and use the same optimal order submission strategy, the execution probability of the limit order can be estimated by calculating the probability that other traders will submit an order to trigger our limit order.

For example, Foucault [32] studies a market for a single risky asset whose trading day is divided into discrete time intervals denoted $t = 1, 2, ..., T$, where T is a random stopping time, and, at each t, the probability that trading will terminate is $(1 - \rho)$. When the trading terminates, the payoff of the asset, whose value is constant and equal to v, will be realised. At each time t, a trader who is characterised by the reservation price, $R_t = v + y_t$, arrives. y_t can take two value $y_h = +L$ or $y_l = -L$ with probability k and $(1 - k)$, respectively. Thus, in this market, there are two types of traders: the y_h type, who only places buy orders and the y_l type, who only places sell orders. Consider a trader of type y_h who arrives at time t. Let B be the bid price this trader chooses if he posts a buy limit order. This order will be executed only if (i) the game does not stop before the arrival of the next trader (with probability ρ), (ii) the next trader is type y_l (with probability $(1 - k)$), and (iii) the next trader submits a market order. The probability of the last event is endogenous and depends on the bid price. In particular, if the bid price is too low, the y_l-type trader will be better off posting a sell limit order. However, if the bid price is greater than the price that y_l-type trader is indifferent between market order and a limit order, this buy limit order will be executed with probability $\rho(1 - k)$. Foucault illustrates that this threshold price is $\theta_h = v - L + \frac{1-\rho(1-k)}{1-\rho^2 k(1-k)}(2L)$.

Another example is an equilibrium model of Parlour [76]. In her model, there is an asset that is traded on day 1 and pays off a certain amount V on day 2. On day 1, at each times $t = 1, 2, ..., T$, a randomly drawn agent arrives at the market. Agents are characterised as potential buyers or sellers with probability π_b and π_s respectively. A potential buyer has an endowment of cash that can be used to purchase one share. On the other hand, a potential seller holds one share which can be sold for cash. Agent t's preferences are given by a utility function $U(C_1, C_2; \beta_t) = C_1 + \beta_t C_2$, where C_1 is agent's consumption on day 1, C_2 is agent's consumption on day 2 and β_t is a trader's personal trade-off between C_1 and C_2. The parameter β_t determines an agent's willingness to trade and is assumed to be randomly distributed over an interval $(\underline{\beta}, \overline{\beta})$ with some continuous distribution $F(.)$. Normally, a potential seller with a low value of β_t will be eager to sell, while a potential buyer with a low value of β_t will be disinclined to buy. All agents have only one opportunity to submit orders, and, once they submit

[3] In an equilibrium, optimal order submission strategies are determined based on the execution probability which is computed by assuming that all traders follow the same strategy

an order, the order cannot be withdrawn. The market, in this setting, is characterised by the designated bid and ask price B, A and a limit order book b_t. The limit order book is described by the number of shares on the bid and ask sides, b_t^B and b_t^A, immediately prior to the arrival of agent t. The optimal strategy for each agent in this market can be determined by recursively working backward from time T. In particular, the agent at time T has the options of trading using a limit order or doing nothing. If the agent is a potential seller, he will compare the utility of doing nothing ($\beta_T V$) with the utility of selling at the bid (B). If $\beta_T V < B$, then he will enter a market sell order. Giving a particular distribution for β_T, the probability that the agent T will enter a market sell order can be computed. A similar computation can be utilised to compute the probability of a market buy order at time T. For the agent arrive at time $T-1$, if he is a buyer and $b_{T-1}^B = 0$, he can enter a buy limit order which will be executed if agent T enter a market sell order, the probability of which we just computed. If $b_{T-1}^B \geq 1$, agent $T-1$'s limit buy will not be first in the execution queue, and, thus, cannot be executed in the one remaining period. Given agent $T-1$'s direction, the limit order book b_{T-1}, and the limit order execution probability, Parlour illustrates that there are cutoffs $\underline{\beta}_{Limit}^{Buy}$ and $\overline{\beta}_{Limit}^{Buy}$ such that if $\beta_{T-1} < \underline{\beta}_{Limit}^{Buy}$, agent $T-1$ will do nothing; if $\underline{\beta}_{Limit}^{Buy} < \beta_{T-1} < \overline{\beta}_{Limit}^{Buy}$, he enters a limit buy order; and if $\beta_{T-1} > \overline{\beta}_{Limit}^{Buy}$, he will enter a market buy order. Given a distribution of β_{T-1}, the probability of these events can be computed and utilised to define the execution probability for time $T-2$, which defines agent $T-2$ optimal strategies, and so on.

Although the results obtained from these analytical studies may not be appropriate for real situations since they depend on assumptions about the market model which is usually a lot simpler than the real market, these results provide us with an understanding of the relation between the execution probability and other related variables.

Empirical models

Apart from using analytical methods, we can also utilise historical data on trades and quotes to estimate the execution probability. For a specific limit of time, the execution probability can be defined as a ratio of the number of limit orders that are executed within that time to the total number of limit orders considered[4]. Using this definition, the execution probability can be easily estimated from a limit-order dataset that contains information about time-to-fill of each limit order. Unfortunately, this information is not available for orders that are cancelled before they get executed. To solve this problem, previous works (e.g. Smith et al. [88]) usually assume that all cancelled orders are unexecuted order and estimated the execution probability at time t with the ratio of the number of limit orders that are executed within time t and the total number of orders in the dataset. Particularly, let the time-to-fill of limit orders in the dataset containing N orders be $TTF_1, TTF_2, ..., TTF_N$ where $TTF_i = \infty$ for unexecuted and cancelled orders. The execution probability can be estimated by

$$P_{EPL}(t) = \frac{\sum_{i=1}^{N} \mathbf{1}(TTF_i \leq t)}{N}, \tag{4.4}$$

[4]This definition has been utilised by several authors including Omura et al. [74] and Hollifield et al. [48]

where $\mathbf{1}(a)$ is an indicator function which will equal to 1 when a is $true$ and 0 otherwise. Although this can solve the problem of cancelled order, the estimated result normally underestimates the real execution probability since cancelled orders may get executed if they are not removed from the order book.

Another approach to solve the problem is to discard cancelled orders from the estimation (e.g. Hollifield et al. [48]). However, the fact that a limit order is cancelled after, say, 10 minutes provides a piece of useful information (the order is unexecuted for at least 10 minutes) and ignoring it would clearly bias the distribution toward a higher execution probability (since we are discarding the orders that are not executed from the analysis). Thus, it is better to incorporate them in the estimation by making use of the information we have up to the time the order is cancelled. Specifically, let the time-to-cancel of each orders in the dataset be $TTC_1, TTC_2, ..., TTC_N$ where $TTC_i = \infty$ for executed and unexecuted order. The execution probability is estimated by

$$P_{EPU}(t) = \frac{\sum_{i=1}^{N} \mathbf{1}(TTF_i \leq t)}{\sum_{i=1}^{N} \mathbf{1}(TTC_i > t)}, \tag{4.5}$$

where the numerator in the above equation is the number of limit orders in the dataset that are executed within time t, and the denominator is the number of limit orders whose status at time t is exactly known, i.e., its cancellation time is greater than t. Although this estimator utilises information from cancelled orders up to the time of cancellation, it still discards the cancelled orders from the analysis when analysing the execution probability at a trading horizon larger than the cancellation time. As a result, the estimation result will still overestimate the real execution probability because of the reasons discussed above.

Although these two estimators are biased, they can be utilised as an upper and lower bound of the execution probability where the lower bound is determined by $P_{EPL}(t)$ while the upper bound is determined by $P_{EPU}(t)$. Additionally, the main drawback of this approach is that it depends on the specific time period. If the trading period of the trader changes, the execution probability must be recalculated.

4.3.2 Execution time models

Instead of directly modelling the probability of execution, we can also model the distribution of execution time and utilise it to estimate the execution probability. The main advantage of this approach, with respect to the former one, is that the trading time is the parameter of the model. Thus, if traders' trading horizon changes, they do not need to recalculate the model to obtain the execution probability.

Related methods for modelling such distributions can be separated into two main categories which are first-passage time models and empirical execution time models.

First-passage time models

The execution time of a limit order can be approximated by a first-passage time (FPT), the first time that the price of an asset reaches or crosses the limit order price. Particularly, let the price of an asset at time $t = 0$ be p_0. The first-passage time through a prescribed level $p_0 + \Delta$ for a fixed distance $\Delta > 0$ is defined as the first time t when $p(t) \geq p_0 + \Delta$. Similarly, a first-passage time for a level $p_0 - \Delta$ is defined as the first time t when $p(t) \leq p_0 - \Delta$. Theoretically, the first passage-time can be thought of as a lower bound of time-to-first-fill, and it will equal to actual time-to-first-fill only when the buy (sell) order is at the top of the queue, at the limit price, or close enough to the top so that it is filled with the first incoming market sell (buy) order. The first time the asset price falls below (rise above) the limit buy (sell) price can also be thought of as an upper bound of the time-to-fill since the asset price can cross the limit price only when all orders at the limit price are executed. Consequently, the relation between first-passage time, time-to-first-fill, and time-to-fill of a limit sell order at Δ ticks above the current price can be summarised as:

$$FPT^{+}(\Delta) \leq TTFF \leq TTF \leq FPT^{+}(\Delta + \epsilon), \tag{4.6}$$

where ϵ is the tick size of the asset. Consequently, the execution probability at time t can be approximated from the first-passage time distribution by

$$P_{FE}(t) = Pr\left\{TTF \leq t\right\} \approx Pr\left\{FPT \leq t\right\}, \tag{4.7}$$

where $x \approx y$ means x is approximately equal to y. However, the probability obtained from this estimation usually overestimates the actual execution probability since the estimated first-passage time is typically lower than the actual time-to-fill[5].

The first-passage time distribution can be modelled both by theoretical and empirical approaches. On the theoretical side, this distribution can be explicitly derived if a stochastic property of the asset price process is given. For example, if the dynamics of the asset price $p(t)$ are given by an arithmetic Brownian motion with drift which has the form

$$dp(t) = \alpha dt + \sigma dz(t), \tag{4.8}$$

where $z(t)$ is a standard Brownian motion and α and σ are constants. The probability density function of the first-passage time to a price level $p_0 + \Delta^{+}$ and $p_0 - \Delta^{-}$ is given by (see [51] page 353–354 and chapter 10 of [93] for example)

$$f(t; \Delta^{+}) = \frac{\Delta^{+}}{\sqrt{2\pi\sigma^2 t^3}} \exp\left(-\frac{(\Delta^{+} - \alpha t)^2}{2\sigma^2 t}\right), \tag{4.9}$$

$$f(t; \Delta^{-}) = \frac{\Delta^{-}}{\sqrt{2\pi\sigma^2 t^3}} \exp\left(-\frac{(\Delta^{-} + \alpha t)^2}{2\sigma^2 t}\right). \tag{4.10}$$

Using the above equation, the probability that a limit sell order at price $p_0 + \Delta^{+}$ will be executed

[5]The reason why the estimated first-passage time is typically lower than the actual time-to-fill is that when the asset price first reaches the limit price the order will be executed only when it is the first order in the queue. Thus, the time-to-fill is usually longer than the first-passage time.

within time t can be approximated by

$$\begin{aligned} P_{FE}(t;\Delta^+) &\approx Pr\left\{FPT \le t;\Delta^+\right\} \\ &= \int_0^t f(t;\Delta^+)dt \\ &= 1-\Phi\left(\frac{\Delta^+ - \alpha t}{\sigma\sqrt{t}}\right) + \exp\left(\frac{2\alpha\Delta^+}{\sigma^2}\right)\Phi\left(\frac{-\Delta^+ - \alpha t}{\sigma\sqrt{t}}\right), \end{aligned} \tag{4.11}$$

where $\Phi(.)$ is the standard normal cumulative distribution function. Similarly, the execution probability of a limit buy order at price $p_0 - \Delta^-$ can be approximated by

$$\begin{aligned} P_{FE}(t;\Delta^-) &\approx Pr\left\{FPT \le t;\Delta^-\right\} \\ &= \int_0^t f(t;\Delta^-)dt \\ &= 1-\Phi\left(\frac{\Delta^- + \alpha t}{\sigma\sqrt{t}}\right) + \exp\left(\frac{2\alpha\Delta^-}{\sigma^2}\right)\Phi\left(\frac{-\Delta^- + \alpha t}{\sigma\sqrt{t}}\right). \end{aligned} \tag{4.12}$$

To utilise these approximations, the estimation of parameter α and σ is required. Given historical data, these parameters can be easily estimated via the maximum likelihood estimator using the equation:

$$\hat{\alpha} = \frac{1}{N\tau}\sum_{i=1}^{N} r_i, \tag{4.13}$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{(r_i - \hat{\alpha}\tau)^2}{\tau}, \tag{4.14}$$

where N is the number of observations in the sample, $r_i = p_i - p_{i-1}$ is the return of the asset over a time interval of τ unit, and τ is a fixed sampling interval.

Although many stochastic processes (e.g. geometric Brownian motion and Markov processes) can be applied in this context, the estimated result is largely dependent on the asset price model. If this specification is not appropriate, the estimation error can be incredibly large. For example, if the asset price exhibits a short-term mean reversion but a geometric Brownian motion is utilised to model the execution time, the predicted execution time will greatly underestimate the real execution time as reported in [61]. To amend the problem, empirical approaches directly model the first-passage time distribution using historical time series of transactions data (see [6, 41] for example)[6]. Given a time series of the asset price $p(t)$, the first-passage time distribution can be estimated by sampling the asset price every τ unit of time and recording the first time that the asset price increases (decreases) by Δ ticks. Unfortunately, we may not be able to evaluate the first-passage time for some observation since the asset price may never reach the expected level in the sampling period. Eliminating these observations from the analysis would clearly bias the empirical distribution towards a shorter first-passage time. Fortunately, a well-known method for handling this observation has been developed by Kaplan and Meier [53], which is a non-parametric

[6]The primary advantages of the empirical approach to the theoretical approach is that if the stochastic process for the asset prices exhibits mean revision or more complex forms of temporal dependence and heterogeneity, this will be automatically incorporated into the empirical model

method for modelling the survival time distribution as described in Section 2.4.3. Specifically, let the observed first-passage time of N observations be $t_1 \leq t_2 \leq ... \leq t_N$, and corresponding to each t_i is n_i the number of orders that are still active before time t_i, and d_i the number of order executed at time t_i. The survival probability, $S(t)$, that describes the probability that the lifetime of the order exceeds t can be estimated with

$$S(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i} \tag{4.15}$$

Using this equation, the execution probability can be estimated with

$$P_{FE}(t) \approx Pr\left\{FPT \leq t\right\} = 1 - S(t) \tag{4.16}$$

Whilst this empirical approach can solve the model specification problem, the first-passage time model still suffers from several other important limitations. The most obvious weakness is the assumption that the order is executed when the limit price is first attained; hence, the model can not be easily modified to handle the variation of limit order sizes as well as the distinction between time-to-first-fill and time-to-fill. Thus, although the first-passage time model is a natural theoretical framework for modelling the order executions, it leaves much to be desired from a practical point of view.

Execution time models

To resolve the central weakness of the first-passage time models, it may be more appropriate to construct models of limit-order execution using the actual execution time, i.e., an elapsed time between the order placement and its complete execution for time-to-fill, and the time from order placement to the first transaction this order participate in for time-to-first-fill. To achieve this, we need information about the time when an order is submitted and when it is executed. In particular, all information about every action relating to the order during its lifetime must be time-stamped and recorded to make it possible to calculate the execution time. Thus, more data is required to utilise this approach than to utilise the first-passage time models that require only historical time series of transactional data.

After obtaining the information about time-to-fill of each executed order, one may directly utilise this information to model the execution time distribution (see [26] for example). Unfortunately, discarding information about unexecuted orders from the model would clearly bias the empirical distribution towards shorter execution times because unexecuted orders are usually orders that have to wait for a considerable amount of time for their execution so that they are usually cancelled before they get executed. Nevertheless, since these orders are not executed, it is not possible to directly calculate the execution time distribution for these unexecuted orders. To solve this problem Lo et al. [61] utilise survival analysis, a form of conditional logistic regression analysis that allows censored observations[7], to model the execution time distribution. The advantage of this approach is that information from unexecuted orders can be easily and correctly accommodated. The idea behind this is that although we can not calculate the exact execution time of unexecuted orders, we know that if these orders were executed, their execution time should be at least as long as their lifetime. In addition, survival analysis also enables us to estimate

[7] In this content, censored observations are unexecuted orders that expire or are cancelled before they are executed.

the execution time distribution as a function of dependent variables such as information about the order (e.g. limit price and order size), state of the order book and market conditions as discussed in Section 2.4.4. This property is very important since recent empirical research suggests that the execution probability is largely dependent on these variables [74], and, without the ability to incorporate these variables into the model, the result may not be accurate. Although empirical execution time models based on survival analysis seem to be a good candidate to model the execution probability, as it can solve all the aforementioned problems, they still have some limitations, as will be pointed out later in this chapter.

4.4 Comparison of previous models

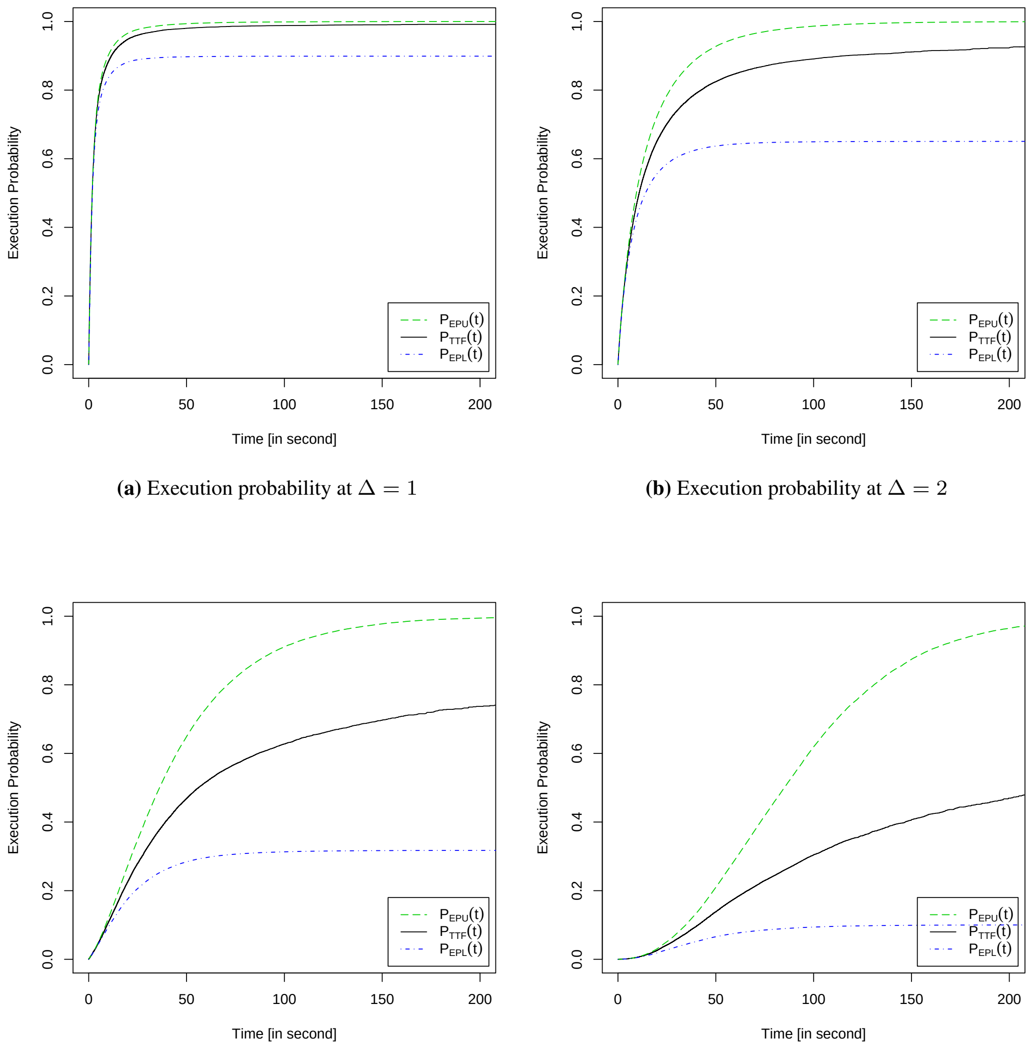
This section illustrates the result of applying the estimation methods discussed in previous sections with the data generated from the SFGK model described in Section 3.5.1 with the aim of analysing the advantages and disadvantages of each method as well as to find the most appropriate methods for modelling the execution probability in the rest of this study.

4.4.1 Execution probability model and execution time model

Let us start by comparing the results from the empirical execution probability model and the empirical execution time model using the data generated from 1×10^7 events of the SFGK model with parameters $\mu = 2$, $\alpha = 0.5$ and $\delta = 0.025$ as described in Section 3.5.1. The lifetime information of each limit order is recorded in order to compute the time-to-fill and the time-to-cancel which are the main components for estimating the execution probability by empirical models. The models compared here are the execution probability model assuming all cancelled order are unexecuted, $P_{EPL}(t)$, described in Equation (4.4), the execution probability model discarding cancel orders from the estimation, $P_{EPU}(t)$, described in Equation (4.5) and the execution time model utilising the Kaplan-Meier estimator, $P_{TTF}(t)$, described in Equation (4.15).

Figure 4.1 illustrates the execution probability of limit buy orders at a distance of $\Delta = 1, 2, 3$ and 4 ticks from the best ask price as estimated from these three methods. The results at all price levels indicate that these three methods produce similar results for the first few seconds, when the number of cancelled orders is still small, and they start to produce different results when the time is increased. These differences are resulted from the different ways in which cancelled orders are handled in each method. Hence the differences are increased when the number of cancelled orders is increased. Consequently, these differences are also an increasing function of the distance from the best ask price, since, at the same time horizon, the larger the distance, the larger the number of cancelled orders as illustrated in Figure 3.5.

The results from all price levels also confirm the bias of execution probability models discussed in Section 4.3.1. Specifically, the execution probability estimated by assuming that all cancelled orders are unexecuted is always lower than that obtained from other methods, which indicates the tendency to underestimate the real execution probability of this approach. Conversely, the execution probability obtained by discarding all cancelled orders from the estimation is generally higher than the results obtained from other methods, which signifies the tendency to overestimate the real execution probability of



(a) Execution probability at $\Delta = 1$

(b) Execution probability at $\Delta = 2$

(c) Execution probability at $\Delta = 3$

(d) Execution probability at $\Delta = 4$

Figure 4.1: Execution probability of a buy order at a distance $\Delta = 1, 2, 3$ and 4 ticks away from the best ask as estimated from an execution probability model that assumes all cancelled orders are unexecuted, $P_{EPL}(t)$, an execution probability model that discards cancel orders from the estimation, $P_{EPU}(t)$ and an execution time model that utilises Kaplan and Meier estimator, $P_{TTF}(t)$.

this approach. The execution probability obtained from execution time model which handles cancelled orders using the Kaplan-Meier estimator is nicely lies between the upper bound and lower bound formed by the above two methods. As a result, among these three estimators, the execution time method seem to be the best methods for estimating the real execution probability both from theoretical and empirical point of view.

4.4.2 Execution time model and first-passage time model

The results from the previous section suggest that it is more appropriate to utilise the execution time model to estimate the execution probability rather than the empirical execution probability model since the execution time model handles cancelled orders better. This section further compares the two execution time models described in Section 4.3.2 - the empirical first-passage time model and the empirical execution time model - using the data generated from 1×10^7 events of the SFGK model as described in the previous section. The empirical execution time model is estimated as described in the previous section. To estimate the empirical first-passage time model, we utilise the difference between the first time that the asset price reaches or crosses the limit price after each order is submitted and the order submission time as a proxy for the execution time. This first-passage time is then utilised as an input in the Kaplan-Meier estimator. Note that a censor observation in this case is the situation when the asset price never reaches the limit price at the end of the simulation.

The comparison between the execution probability of limit buy orders at a distance of $\Delta = 1, 2, 3, 4$ and 5 ticks away from the best ask price as estimated from the empirical first-passage time model and the empirical execution time model is illustrated in Figure 4.2. As discussed in Section 4.3.2, the execution probability of limit buy orders at a distance of Δ ticks away from the best ask price estimated from the empirical execution time model, $P_{TTF}(t)$, is generally bounded by the execution probability estimated from the empirical first-passage time model, $P_{FPT}(t)$, at a distance of Δ and $\Delta + 1$ ticks away from the best ask price. This is conformed to theoretical explanation that the first-passage time is a lower bound of time-to-fill while the first time that the price breaks through the limit price is an upper bound of time-to-fill. Although some inconsistencies do occur when time is greater than two hundred seconds for $\Delta = 2, 4$ and 5, these inconsistencies are caused by a large standard error of execution time models when time is getting larger as shown in Figure 4.3a where dot lines in the figure display 95% confidence interval of the estimated execution probability. This large standard error results from the lack of observations when time gets larger which is caused by the fact that the lifetime of limit orders in this simulation is exponentially distributed with a rate of $\delta = 0.025$ and the expected lifetime of $1/\delta = 40$ seconds. Consequently, most of the limit orders will be cancelled before their lifetime has reached the expected lifetime, and only $e^{-0.025 \times 200} = 0.0067\%$ of limit orders will have a lifetime longer than two hundred

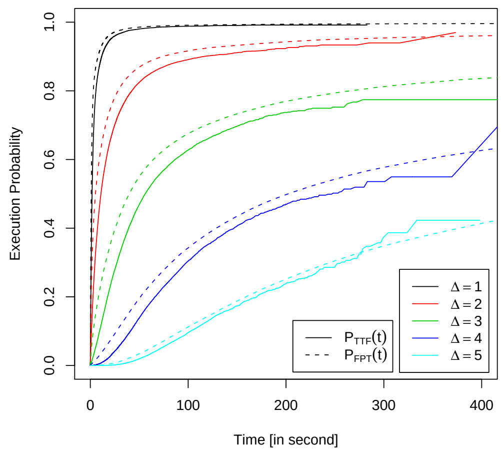
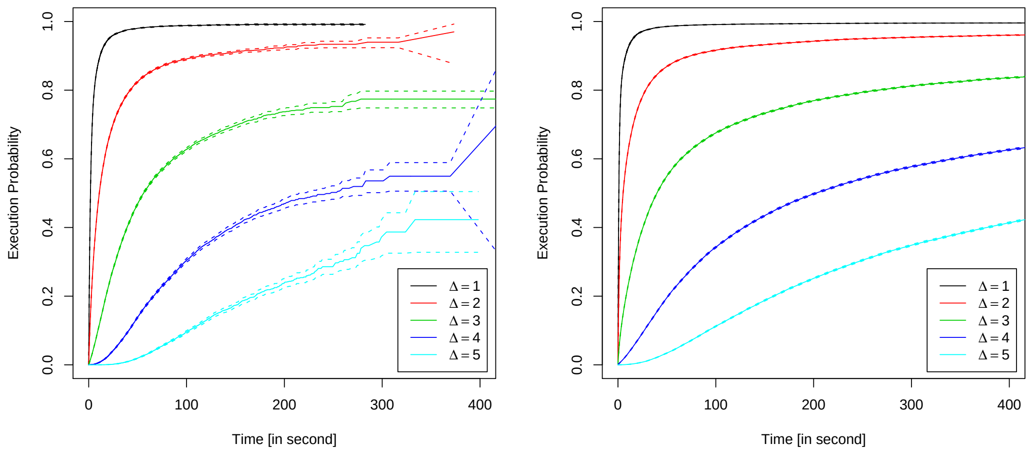


Figure 4.2: Execution probability of a buy order at a distance $\Delta = 1, 2, 3, 4$ and 5 ticks away from the best ask price as estimated from an empirical execution time model, $P_{TTF}(t)$, and an empirical first-passage time model, $P_{FPT}(t)$.



(a) Confidence interval of $P_{TTF}(t)$ **(b)** Confidence interval of $P_{FPT}(t)$

Figure 4.3: Confidence interval of the estimated execution probability from an empirical execution time model, $P_{TTF}(t)$, and an empirical first-passage time model, $P_{FPT}(t)$.

seconds. In general, most of the limit orders in this simulation will be cancelled before they get executed, as illustrated by the fill probability displayed in Figure 3.5, and there will be only a few observations left when we want to analyse the execution probability at the time greater than the expected lifetime. As a result, it might not be appropriate to utilise the execution time model that requires information about the lifetime of each limit order to estimate the execution probability when the time that we want to model is greater than the expected lifetime. A better alternative is to utilise the first-passage time model which does not suffer from this limitation since it does not require information about the lifetime of each limit order but estimates the execution time from the first time that the asset price reaches the limit price as illustrated in Figure 4.3b.

Although it is more appropriate to utilise the first-passage time model to estimate the execution probability over a long time horizon, the result obtained from this model generally overestimates the real execution probability and will equal the real execution probability only when the order considered is the first order in the order book at the limit price. Thus these two models have their own advantages and disadvantages, and the decision of when to utilise which models is totally dependent on the problem faced. Specifically, the execution time model provide a better estimation of the execution probability but the standard error when analysing the execution probability over a long time horizon can be very large. On the other hand, the first-passage time model has small standard error over all time horizons but the estimated execution probability generally overestimates the real execution probability.

4.4.3 Empirical and theoretical first-passage time model

This section compares the results from an empirical first-passage time model and a theoretical first-passage time model when the asset price is assumed to follow the arithmetic first-passage Brownian motion as discussed in Section 4.3.2. The result from the empirical model is obtained by applying the Kaplan-Meier estimator to the first-passage time data set obtained from a simulation run of 1×10^7

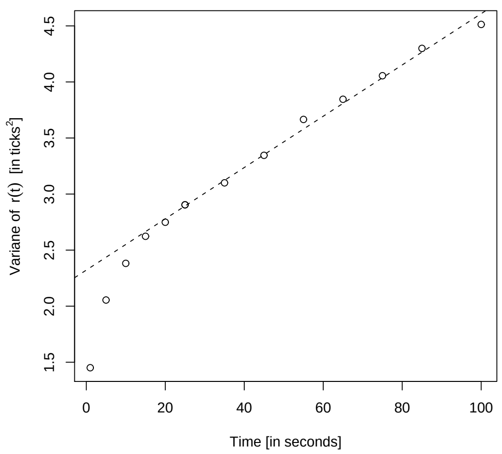
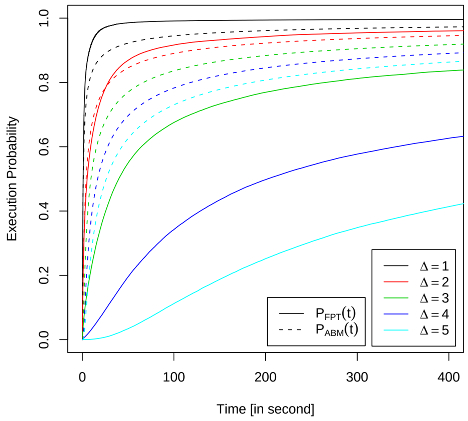


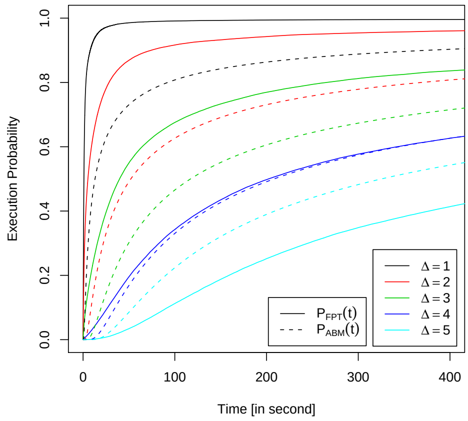
Figure 4.4: Variance of the returns estimated from several sampling periods. The estimated variance is not linear in time as expected from the arithmetic Brownian motion.

events as discussed in the previous section, while the result from the theoretical model is obtained by firstly estimating the time-normalised mean and the time-normalised variance of returns generated from the same simulation run using Equation (4.13) and (4.14), and, then, utilising Equation (4.12) to estimate the execution probability.

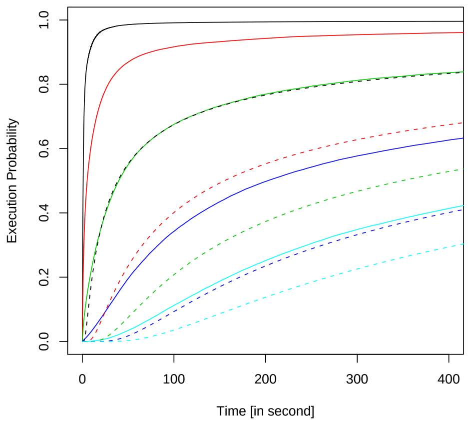
Figure 4.4 displays the sample variance of the returns estimated from this simulation run at several sampling periods. The result indicates that the asset price dynamic generated from this simulation does not follow the arithmetic Brownian motion since the variance is not linear in time as expected from the Brownian motion. Consequently, this poses a problem in utilising this model to predict the execution probability as the time-normalised variances computed from different sampling periods are different. The comparison between the execution probability of limit buy orders at a distance of $\Delta = 1, 2, 3, 4$ and 5 ticks away from the best ask price as estimated from the empirical first-passage time model and the theoretical first-passage time model using the time-normalised mean and variance computed at four different sampling periods (i.e. 1 second, 5 seconds, 10 seconds and 20 seconds) is illustrated in Figure 4.5. Since the time-normalised variance obtained from each sampling period is different, the estimated execution probability of the theoretical model obtained from each sampling period is also different. Additionally, all of these results deviate significantly from the one obtained from the empirical model, which further confirm that the price dynamic generated from the SFGK model does not follow the arithmetic Brownian motion. Consequently, this illustrates that the execution probability estimated from theoretical first-passage time model can be quite different from empirical first-passage time model especially when the assumption about the asset price dynamic is incorrect, and, thus, it might not be appropriate to utilise the theoretical model to study the execution probability unless the correct model of asset price dynamic is known. This problem leads us to study asset price dynamics in this simulation model, which will be the main topic of Chapter 6.



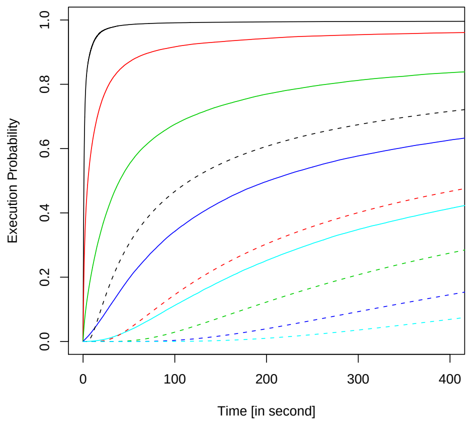
(a) 1 second



(b) 5 seconds



(c) 10 seconds



(d) 20 seconds

Figure 4.5: Execution probability of a buy order at a distance of $\Delta = 1, 2, 3, 4$ and 5 ticks away from the best ask as estimated from an empirical first-passage time model, $P_{FPT}(t)$, and an theoretical first-passage time model, $P_{ABM}(t)$ using the time-normalised mean and variance computed at four different sampling periods which are 1, 5, 10 and 20 seconds.

4.4.4 Summary

To sum up, the experiments in this section indicate that the execution probability model and the execution time model produce similar results when the number of cancelled orders is small, and they start to produce different results when there are more cancelled orders. Among all models considered, the execution time model that utilises the Kaplan-Meier estimator to handle cancelled orders seem to be the best performing methods both from a theoretical and an empirical point of view. The choice between first-passage time and execution time model depends on the problem faced. Specifically, the execution time model generally provides better estimation compared to the first-passage time model since the first-passage time model generally overestimates the real execution probability and will equal the real execution probability only when the considered order is at the top of the queue. However, the result obtained from execution time model might have a large standard error when analysing the execution probability over a long time horizon, while the result from first-passage time has small standard error over all time horizons. The experiment with theoretical first-passage time model indicates that the estimated execution probability can be quite different from empirical first-passage time model when the assumption about the asset price dynamic is incorrect.

4.5 Parameter of execution probability

To gain more insight into the relationship between the execution probability and other variables, this section further analyses the effect that each variable has on the execution probability by utilising the data generated from the SFGK model. The reason why we utilise the data generated from this simulation model instead of the data obtained from a real market is that the simulation allow us to analyse the effect of each determinant separately by running the simulation model with different initial conditions which is not possible when utilising the data generated from the real market. Consequently, the determinants studied in this section will be variables that affect the execution probability in the SFGK model. These include the limit price of the order, the state of the order book, which is generally summarised by the bid-ask spread and the quantity of orders at each price level and the order arrival and cancellation rates, which are the main parameters of the SFGK model. Note that unlike previous works (e.g. [61] and [74]) whose aim is to study the effect that each parameter has on the execution probability, the main objective of this section is to find the most appropriate model for modelling these effects not the effect itself. Since the results from the previous section suggest that the most appropriate method for analysing the execution probability is the Kaplan-Meier estimator, which is a non-parametric survival analysis technique, this section will utilise survival analysis technique as a main tool for investigating this relationship. Specifically, the Kaplan-Meier estimator will be utilised to estimate the execution probability in each situation in order to study the basic effect that each determinant has on the execution probability. After that we will study the suitability of applying the two most popular methods for handling explanatory variables in survival analysis literature, which are a proportional hazards model and an accelerated failure time model, with the aim of determining the most appropriate techniques for modelling these effects by utilising tests and graphical diagnostics methods discussed in Section 2.4.4. In particular, the plot of log-log survival curve and the Grambsch and Therneau test [36], hereafter referred to GT test, will be utilised to test the proportional hazards assumption while the plot of quantiles and the log-rank statistic for testing the different between two survival curves will be utilised to test the accelerated life time assumption.

4.5.1 Distance from the opposite best price

Let us firstly study the effect of the limit order price, as measured by the distance from the opposite best price, on the execution probability by analysing the data generated from a simulation of the SFGK model. To achieve this, we simulate the SFGK model with parameter $\mu = 2$, $\alpha = 0.5$ and $\delta = 0.025$ for $10,000$ rounds. In each round, the initial bid-ask spread is set to one tick, while the number of orders at all price levels is set to $\alpha/\delta = 0.5/0.025 = 20$, which is the expected number of the orders at price level far away from the opposite best price as illustrated in Figure 3.6. The simulation is run until the simulation time reaches eight hours and the first time that the transactional price reaches or crosses each

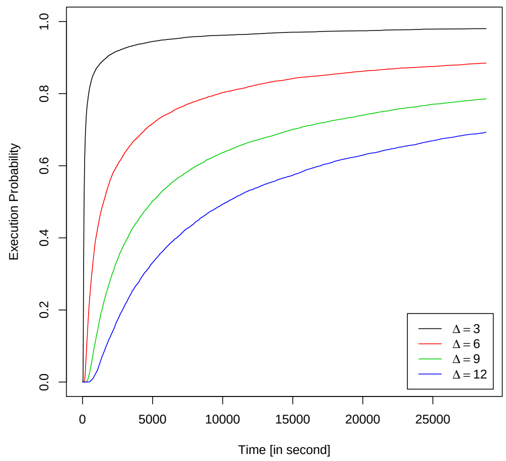


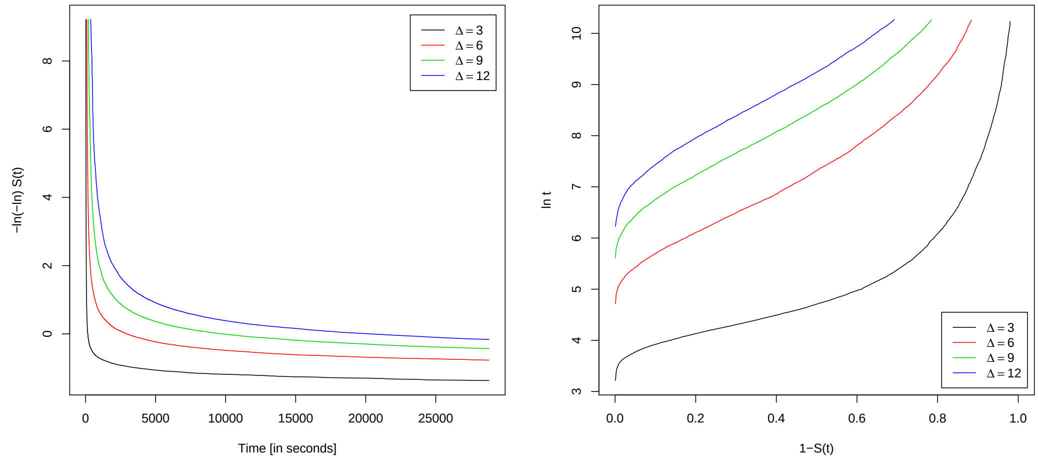
Figure 4.6: Execution probability of limit buy orders at a distance of $\Delta = 3, 6, 9$ and 12 ticks away from the best ask price as estimated from the Kaplan-Meier estimator.

price level will be utilised as an estimation of the execution time of a limit buy order at that price level.

The execution probability at a distance of three, six, nine and twenty ticks away from the best ask price obtained by applying the Kaplan-Meier estimator to the data generated from the above simulation are displayed in Figure 4.6. These results illustrate that the execution probability is negatively correlated with the distance from the opposite best price since the execution probability is generally lower when the distance from the opposite best price is larger. This is because the larger the distance from the opposite best price, the more time it takes the transactional price to reaches that price level and thus the lower the execution probability.

To determine whether the proportional hazards model is appropriate for modelling this effect or not, we plot the log-log survival curves at those four price levels in Figure 4.7a. The result indicates that these four curves seem to be unparallel to each other, which indicates the violation of the proportional hazards assumption. To confirm this, we further utilise the GT test to check the proportional hazards assumption quantitatively. As discussed in Section 2.4.4, this can be achieved by firstly fitting the proportional hazards model to the data and then computing the Schoenfeld residuals for each non-censored individual from the fitted model. Since, under the proportional hazards assumption, the Schoenfeld residuals are independent of time, the slope of the Schoenfeld residuals against a function of time should be zero and the deviation from this is the evidence against the proportional hazards assumption. To make the analysis independent from the functional form of the covariate, we will apply the test in two-sample setting so that the covariate can be represented by a binary variable indicating the category each individual belongs to. The results from applying the GT test to six combinations of limit order price displayed in Table 4.1 further support the violation of the proportional hazards assumption since the p-value of all combinations is less than 5% critical value. Consequently, these results indicate that the proportional hazards model may not be an appropriate model for modelling the effect of limit order price on the execution probability.

To determine whether the accelerated failure time model is appropriate for modelling this effect or



(a) Log-log survival curve

(b) Log quantiles plot

Figure 4.7: The plot of log-log survival curve and the log quantiles associated with the estimated execution probability of limit buy orders at a distance of $\Delta = 3, 6, 9$ and 12 ticks away from the best ask price estimated from the Kaplan-Meier estimator.

not, we plot the quantiles for a limit buy order at each price level in Figure 4.7b. The result provides mixed evidence, since the quantiles of a limit buy order when $\Delta = 3$ is clearly unparallel to other curves while the quantiles of a limit buy order when $\Delta = 6, 9$ and 12 are roughly parallel to each other. To clarify this, we further applying log rank tests for the different between two survival curves to check the accelerated failure time assumption quantitatively. As discussed in Section 2.4.4, this can be achieved by firstly estimating the failure time scale parameter ψ of the two samples, and then checking whether the associated normalised survival curve of the two samples is similar to each other or not using the log-rank tests. The results reported in Table 4.1 indicates the same results as the graphical diagnostics. Particularly, all p-values of combinations involving with $\Delta = 3$ is less than 5% critical value, while the p-value of other combinations are higher than 5% critical value. This indicates that the accelerated life time assumption is violated for all combinations involving with $\Delta = 3$, and is not violated in other combinations. This suggests that we might be able to divide the effect of limit order price into two regimes, one of which can be modelled by the accelerated life time model while the other cannot.

To further investigate this, we apply the test to all price levels from one tick to twenty ticks, and the result displayed in Table 4.2 indicates that all combinations that involve the case when limit order price is less than five ticks are generally not suitable for accelerated failure time model. The combinations that involve the case when limit order price is equal to five ticks provide mixed results; most of the combinations have p-value greater than 5% critical value with only three combinations (i.e. 5-10, 5-11 and 5-12) less than 5%. Additionally, the p-value of all other combinations is well above the 5% critical value indicating that the accelerated failure time assumption is satisfied for all limit order price greater than five ticks away from the best ask price. Consequently, these results support our hypothesis that the effect of limit order price can be divided into two regimes, one of which can be modelled by the

Group	Propotional hazard model			Accelerated life time model		
	β	Chisq	p-value	ψ	Chisq	p-value
$\Delta = 3, 6$	-1.23	3472.27	0.00	22.00	216.42	0.00
$\Delta = 3, 9$	-1.68	3245.90	0.00	62.15	96.11	0.00
$\Delta = 3, 12$	-1.97	2787.16	0.00	107.21	18.85	0.00
$\Delta = 6, 9$	-0.56	1095.17	0.00	3.34	0.37	0.54
$\Delta = 6, 12$	-0.91	2050.86	0.00	6.87	0.96	0.33
$\Delta = 9, 12$	-0.38	425.28	0.00	2.08	0.46	0.50

Table 4.1: The estimated parameters of the proportional hazards model and the accelerated life time model together with the test statistics for the execution probability of limit buy orders at different limit price.

Group	ψ	p-value	Group	ψ	p-value	Group	ψ	p-value
$\Delta = 1, 2$	30.16	0.00	$\Delta = 1, 3$	994.36	0.00	$\Delta = 1, 4$	2759.67	0.00
$\Delta = 1, 5$	4644.96	0.00	$\Delta = 1, 6$	6370.65	0.00	$\Delta = 1, 7$	8602.67	0.48
$\Delta = 1, 8$	10543.55	0.03	$\Delta = 1, 9$	12448.86	0.00	$\Delta = 1, 10$	14276.95	0.00
$\Delta = 1, 11$	16477.12	0.00	$\Delta = 1, 12$	18785.14	0.00	$\Delta = 1, 13$	21065.24	0.00
$\Delta = 1, 14$	23421.34	0.00	$\Delta = 1, 15$	26021.00	0.00	$\Delta = 1, 16$	29233.96	0.00
$\Delta = 1, 17$	32522.13	0.00	$\Delta = 1, 18$	36316.32	0.00	$\Delta = 1, 19$	40095.71	0.00
$\Delta = 2, 3$	27.33	0.00	$\Delta = 2, 4$	86.78	0.00	$\Delta = 2, 5$	149.66	0.00
$\Delta = 2, 6$	210.38	0.00	$\Delta = 2, 7$	281.53	0.00	$\Delta = 2, 8$	342.10	0.00
$\Delta = 2, 9$	399.35	0.00	$\Delta = 2, 10$	440.17	0.00	$\Delta = 2, 11$	490.74	0.91
$\Delta = 2, 12$	537.94	0.01	$\Delta = 2, 13$	570.14	0.00	$\Delta = 2, 14$	611.87	0.00
$\Delta = 2, 15$	651.18	0.00	$\Delta = 2, 16$	690.02	0.00	$\Delta = 2, 17$	727.03	0.00
$\Delta = 2, 18$	761.30	0.00	$\Delta = 2, 19$	792.93	0.00	-	-	-
$\Delta = 3, 4$	4.80	0.00	$\Delta = 3, 5$	11.95	0.00	$\Delta = 3, 6$	22.00	0.00
$\Delta = 3, 7$	35.12	0.00	$\Delta = 3, 8$	48.65	0.00	$\Delta = 3, 9$	62.15	0.00
$\Delta = 3, 10$	76.49	0.00	$\Delta = 3, 11$	91.16	0.00	$\Delta = 3, 12$	107.21	0.00
$\Delta = 3, 13$	121.63	0.02	$\Delta = 3, 14$	135.98	0.68	$\Delta = 3, 15$	151.47	0.39
$\Delta = 3, 16$	167.14	0.05	$\Delta = 3, 17$	179.69	0.00	$\Delta = 3, 18$	193.62	0.00
$\Delta = 3, 19$	206.56	0.00	-	-	-	-	-	-
$\Delta = 4, 5$	2.38	0.01	$\Delta = 4, 6$	4.53	0.00	$\Delta = 4, 7$	7.65	0.00
$\Delta = 4, 8$	11.32	0.00	$\Delta = 4, 9$	15.40	0.00	$\Delta = 4, 10$	19.87	0.00
$\Delta = 4, 11$	24.92	0.00	$\Delta = 4, 12$	30.49	0.00	$\Delta = 4, 13$	36.04	0.01
$\Delta = 4, 14$	41.69	0.14	$\Delta = 4, 15$	48.09	0.24	$\Delta = 4, 16$	55.37	0.24
$\Delta = 4, 17$	61.96	0.62	$\Delta = 4, 18$	68.48	0.64	$\Delta = 4, 19$	74.64	0.42
$\Delta = 5, 6$	1.87	0.65	$\Delta = 5, 7$	3.15	0.11	$\Delta = 5, 8$	4.66	0.06
$\Delta = 5, 9$	6.32	0.06	$\Delta = 5, 10$	8.26	0.03	$\Delta = 5, 11$	10.63	0.02
$\Delta = 5, 12$	13.09	0.04	$\Delta = 5, 13$	15.81	0.08	$\Delta = 5, 14$	18.64	0.12
$\Delta = 5, 15$	21.69	0.21	$\Delta = 5, 16$	25.03	0.36	$\Delta = 5, 17$	28.15	0.76
$\Delta = 5, 18$	31.44	0.73	$\Delta = 5, 19$	35.00	0.76	-	-	-
$\Delta = 6, 7$	1.67	0.56	$\Delta = 6, 8$	2.46	0.62	$\Delta = 6, 9$	3.34	0.54
$\Delta = 6, 10$	4.33	0.55	$\Delta = 6, 11$	5.56	0.49	$\Delta = 6, 12$	6.87	0.33
$\Delta = 6, 13$	8.39	0.28	$\Delta = 6, 14$	10.00	0.33	$\Delta = 6, 15$	11.69	0.35
$\Delta = 6, 16$	13.64	0.58	$\Delta = 6, 17$	15.63	0.49	$\Delta = 6, 18$	17.71	0.44
$\Delta = 6, 19$	19.75	0.46	-	-	-	-	-	-
$\Delta = 7, 8$	1.48	0.83	$\Delta = 7, 9$	2.02	0.80	$\Delta = 7, 10$	2.62	0.88
$\Delta = 7, 11$	3.38	0.70	$\Delta = 7, 12$	4.20	0.52	$\Delta = 7, 13$	5.10	0.70
$\Delta = 7, 14$	6.09	0.63	$\Delta = 7, 15$	7.14	0.48	$\Delta = 7, 16$	8.39	0.47
$\Delta = 7, 17$	9.62	0.48	$\Delta = 7, 18$	10.97	0.61	$\Delta = 7, 19$	12.20	0.88
$\Delta = 8, 10$	1.78	0.69	$\Delta = 8, 11$	2.29	0.61	$\Delta = 8, 12$	2.84	0.68
$\Delta = 8, 13$	3.45	0.72	$\Delta = 8, 14$	4.12	0.70	$\Delta = 8, 15$	4.83	0.63
$\Delta = 8, 16$	5.64	0.69	$\Delta = 8, 17$	6.46	0.78	$\Delta = 8, 18$	7.36	0.78
$\Delta = 8, 19$	8.23	0.77	-	-	-	-	-	-

Table 4.2: The estimated parameters of the accelerated life time model together with the test statistics for the execution probability of limit buy orders at different limit price.

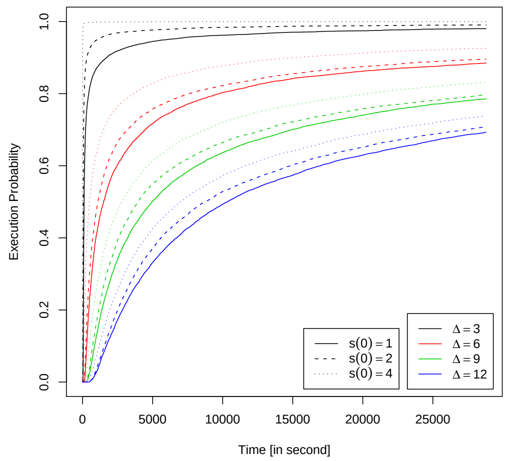


Figure 4.8: Execution probability of limit buy orders at a distance of $\Delta = 3, 6, 9$ and 12 ticks away from the best ask price when the initial bid-ask spread equals to $1, 2$ and 4 ticks as estimated from the Kaplan-Meier estimator.

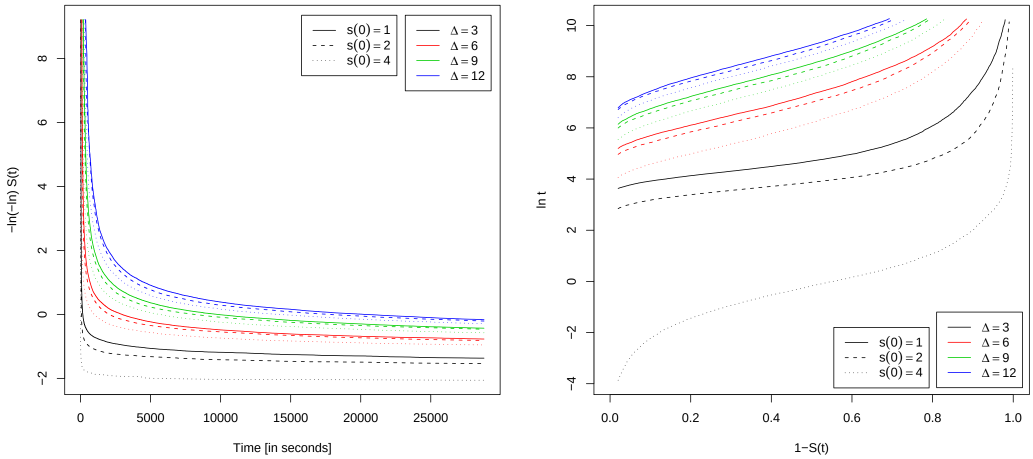
accelerated failure time model, and the appropriate threshold for separating this, in this case, is a limit order price of six ticks away from the best ask price.

Briefly, the execution probability is negatively correlated with the distance from the opposite best price. Both graphical plots and test statistics suggest that it is more appropriate to utilise the accelerated failure time model to model the effect of limit order price rather than the proportional hazards model since the proportional hazards assumption is violated at all price levels, while we can divide the effect of limit order price into two regimes, one of which satisfies the accelerated failure time assumption and another does not.

4.5.2 Bid-ask spread

Let us now analyse the effect of the bid-ask spread on the execution probability. To accomplish this, we simulate the SFGK model as described in the previous section with initial bid-ask spread equal to one, two and four ticks. The execution probability obtained by applying the Kaplan-Meier estimator to the data generated from these simulations are illustrated in Figure 4.8. The result shows that the execution probability is generally higher when the initial bid-ask spread is larger, indicating a positive correlation between the bid-ask spread and the execution probability.

To determine whether the proportional hazards model is an appropriate candidate for modelling the effect of the bid-ask spread on the execution probability or not, we plot the log-log survival curve for each value of initial bid-ask spread at four different price levels in Figure 4.9a. These log-log curves are clearly not parallel to each other, indicating that the hazards of different bid-ask spreads are not proportional to each other. This is further confirmed by the GT test reported in Table 4.3, where the p-value at all price levels is less than the 5% critical value. Consequently, the effect of the bid-ask spread does not satisfy the proportional hazards assumption and the proportional hazards model should not be utilised to model this effect.



(a) Log-log survival curve

(b) Log quantiles plot

Figure 4.9: The plot of log-log survival curve and the log quantiles associated with the estimated execution probability of limit buy orders at a distance of $\Delta = 3, 6, 9$ and 12 ticks away from the best ask price when the initial bid-ask spread equals to $1, 2$ and 4 ticks as estimated from the Kaplan-Meier estimator.

Group		Propotional hazard model			Accelerated life time model		
		β	Chisq	p-value	ψ	Chisq	p-value
$\Delta = 3$	$s(0) = 1, 2$	0.69	2354.80	0.00	0.1827	988.34	0.00
	$s(0) = 1, 4$	2.78	1889.08	0.00	0.0043	76.12	0.00
	$s(0) = 2, 4$	2.31	2736.37	0.00	0.0209	109.82	0.00
$\Delta = 6$	$s(0) = 1, 2$	0.13	122.45	0.00	0.7624	0.01	0.91
	$s(0) = 1, 4$	0.45	1379.42	0.00	0.3562	0.64	0.42
	$s(0) = 2, 4$	0.34	861.38	0.00	0.4677	0.59	0.44
$\Delta = 9$	$s(0) = 1, 2$	0.08	56.73	0.00	0.8365	0.19	0.67
	$s(0) = 1, 4$	0.26	361.30	0.00	0.5749	0.07	0.79
	$s(0) = 2, 4$	0.18	149.56	0.00	0.6808	0.06	0.81
$\Delta = 12$	$s(0) = 1, 2$	0.07	23.31	0.00	0.8662	0.28	0.60
	$s(0) = 1, 4$	0.18	153.57	0.00	0.6827	0.53	0.47
	$s(0) = 2, 4$	0.12	63.44	0.00	0.7957	0.01	0.94

Table 4.3: The estimated parameters of the proportional hazards model and the accelerated life time model together with the test statistics for the execution probability of limit buy orders at a distance of three, six, nine and twelve ticks away from the best ask price when varying the initial bid-ask spread between one, two and four.

Group		ψ	p-value	Group		ψ	p-value	Group		ψ	p-value
Δ	$s(0)$			Δ	$s(0)$			Δ	$s(0)$		
1	1, 2	0.8394	0.00	2	1, 2	0.0648	0.00	3	1, 2	0.1827	0.00
1	1, 4	0.7589	0.00	2	1, 4	0.0462	0.00	3	1, 4	0.0043	0.00
1	2, 4	0.9107	0.71	2	2, 4	0.7512	0.18	3	2, 4	0.0210	0.00
4	1, 2	0.5153	0.00	5	1, 2	0.6840	0.36	6	1, 2	0.7624	0.92
4	1, 4	0.0047	0.00	5	1, 4	0.1969	0.83	6	1, 4	0.3563	0.42
4	2, 4	0.0090	0.00	5	2, 4	0.2929	0.12	6	2, 4	0.4677	0.44
7	1, 2	0.7901	0.80	8	1, 2	0.8213	0.96	9	1, 2	0.8365	0.67
7	1, 4	0.4531	0.52	8	1, 4	0.5106	0.94	9	1, 4	0.5749	0.79
7	2, 4	0.5750	0.82	8	2, 4	0.6135	0.53	9	2, 4	0.6808	0.81

Table 4.4: The estimated parameters of the accelerated life time model together with the test statistics for the execution probability of limit buy orders when varying the initial bid-ask spread between one, two and four.

The test for the accelerated failure time model provides more positive results. In particular, the quantiles curve for each value of the initial bid-ask spread at different price levels displayed in Figure 4.9b are somewhat parallel when $\Delta = 6, 9$ and 13. This suggest a possibility to utilise accelerated failure time model to model this effect at those price level; however, the clearly unparallel curves when $\Delta = 3$ suggest that the accelerated failure time assumption does not hold for all price levels and we may need to divide the effect from the bid-ask spread into two regimes as in the previous section. To confirm this, we apply the log-rank test to all price levels from one tick to nine ticks, and the result displayed in Table 4.4 indicates that a suitable criteria for separating the bid-ask spread into two regimes is a limit order price value of five ticks since all p-value when Δ is greater or equal to five ticks are well above the 5% critical value while the p-value when Δ is less than five ticks are not. The dependency on the limit order price together with the fact that the estimated time scale factor, ψ, at each price level is largely different also suggests an interaction effect between the bid-ask spread and the limit order price on the execution probability, and thus this effect needs to be addressed when developing a full execution probability model.

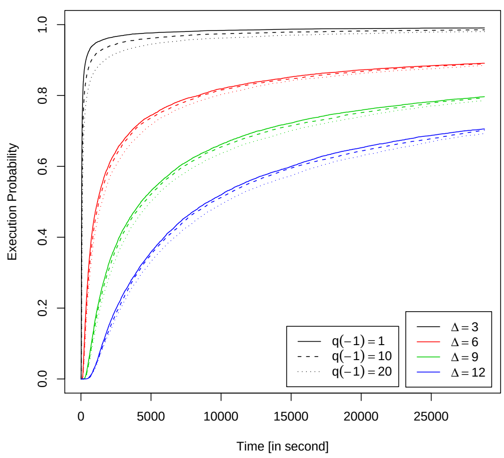
In conclusion, the execution probability is positively correlated with the bid-ask spread and it is more appropriate to utilise the accelerated failure time model to model this effect rather than the proportional hazards model. Unfortunately the accelerated failure time assumption is not valid at all price levels but is satisfied only when the limit order price is larger than some specified value. The dependency on limit order price suggests that there is an interaction effect between the bid-ask spread and limit order price which should be addressed properly when developing a full execution probability model.

4.5.3 Number of orders at each price level

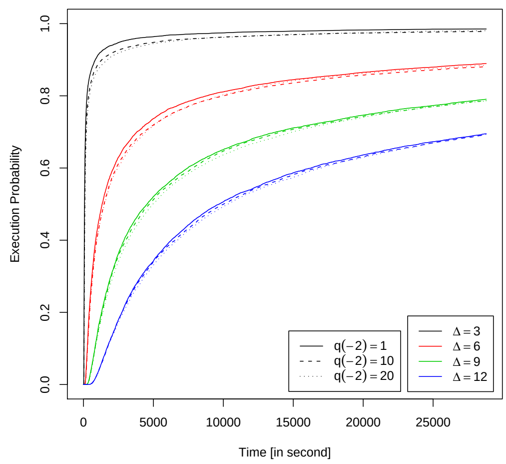
This section investigates the effect of the number of orders at each price level on the execution probability. Since the order book has two sides (i.e the bids and the asks), this section will investigate these two sides separately.

Number of buy orders

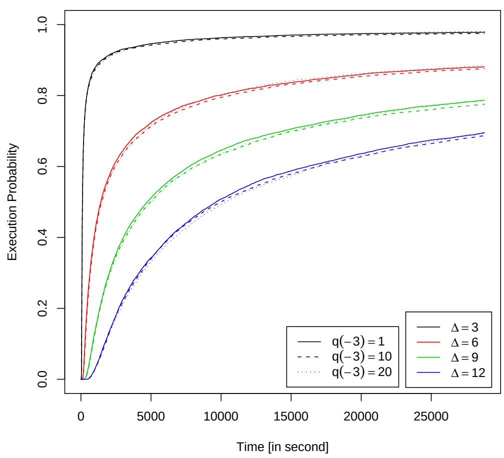
Let us now analyse the effect of the number of buy orders on the execution probability of limit buy orders. Similar to previous sections, we will utilise the data generated from the simulation of the SFGK model with the same setting as described in Section 4.5.1 but vary the number of buy orders at some specified levels between one, ten and twenty. The execution probability obtained by applying Kaplan-Meier estimator to the data generated from the above simulation when varying the number of buy orders at a distance of one, two, three and four ticks away from the best ask price are illustrated in Figure 4.10. These results indicates that the execution probability is negatively correlated with the number of buy orders in the order book since the execution probability is higher when the number of buy orders is lower in all cases. Additionally, these results also suggest that the number of buy orders at one tick away from the best ask price has bigger effect on the execution probability than the number of buy orders at other prices since the difference between the curves at each price level in Figure 4.10a is larger than those in Figure 4.10b, 4.10c, and 4.10d. Using the same argument, we can conclude that the number of buy orders at two ticks away from the best ask price has more effect than the number of buy orders at three



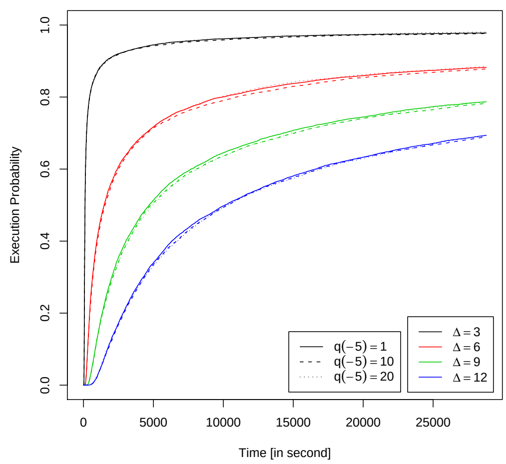
(a) One tick away from the best ask price



(b) Two ticks away from the best ask price



(c) Three tick away from the best ask price

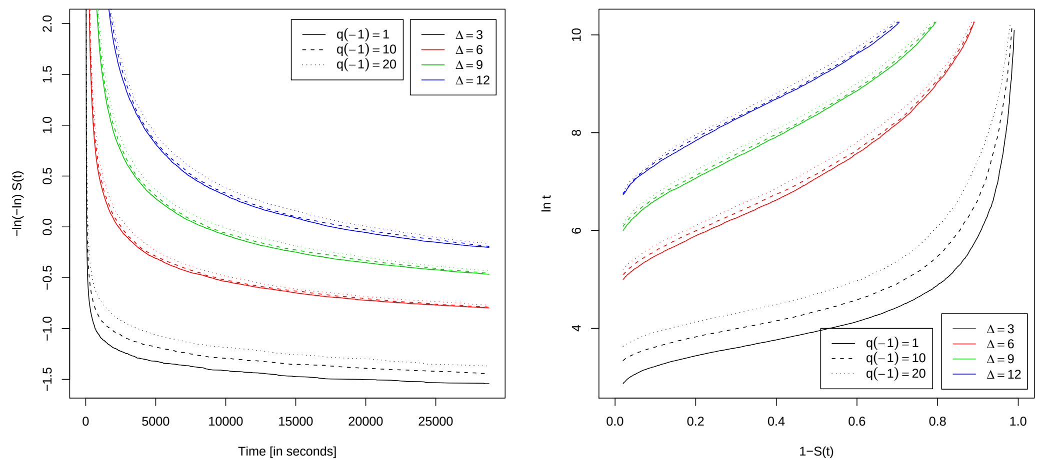


(d) Four ticks away from the best ask price

Figure 4.10: Execution probability of limit buy orders at a distance of $\Delta = 3, 6, 9$ and 12 ticks away from the best ask price when the number of buy orders at one, two, three and four ticks away from the best ask price are varied between one, ten and twenty as estimated from the Kaplan-Meier estimator.

and four ticks away from the best ask price. These indicate that the number of buy orders at a price level near the best ask price has a bigger effect on the execution probability than the number of orders at price levels further away, and the number of buy orders at the best bid price has the largest effect on the execution probability. Consequently the rest of this section will study only the effect of the number of orders at the best bid price.

To determine whether the effect from the number of buy orders satisfies the proportional hazards assumption or not, we plot the log-log survival curves obtained by varying the number of buy orders at the best ask price in Figure 4.11a. The result indicates that the log-log curve for each number of buy orders at the same price level is clearly not parallel to each other. This suggests that the proportional hazards assumption might be violated in this situation. To confirm this, we further apply the GT test, and



(a) Log-log survival curve

(b) Log quantiles plot

Figure 4.11: The plot of log-log survival curve and the log quantiles associated with the estimated execution probability of limit buy orders at a distance of $\Delta = 3, 6, 9$ and 12 ticks away from the best ask price when the number of buy orders at the best bid price is one, ten and twenty as estimated from the Kaplan-Meier estimator.

Group		Propotional hazard model			Accelerated life time model		
		β	Chisq	p-value	ψ	Chisq	p-value
$\Delta = 3$	$q(-1) = 1, 10$	-0.3632	824.17	0.00	2.1937	246.72	0.00
	$q(-1) = 1, 20$	-0.6334	2020.01	0.00	4.7735	831.19	0.00
	$q(-1) = 10, 20$	-0.2879	510.85	0.00	2.1765	282.28	0.00
$\Delta = 6$	$q(-1) = 1, 10$	-0.0352	19.39	0.00	1.0566	0.50	0.48
	$q(-1) = 1, 20$	-0.0911	85.74	0.00	1.2017	0.36	0.55
	$q(-1) = 10, 20$	-0.0566	24.61	0.00	1.1318	0.01	0.94
$\Delta = 9$	$q(-1) = 1, 10$	-0.0234	4.77	0.03	1.0558	0.04	0.85
	$q(-1) = 1, 20$	-0.0681	29.41	0.00	1.1680	0.19	0.66
	$q(-1) = 10, 20$	-0.0448	10.78	0.00	1.1066	0.09	0.76
$\Delta = 12$	$q(-1) = 1, 10$	-0.0184	2.55	0.11	1.0384	0.00	0.98
	$q(-1) = 1, 20$	-0.0584	14.49	0.00	1.1363	0.27	0.61
	$q(-1) = 10, 20$	-0.0399	5.01	0.02	1.0904	0.39	0.53

Table 4.5: The estimated parameters of the proportional hazards model and the accelerated life time model together with the test statistics for the execution probability of a limit buy order at a distance of three, six, nine and twelve ticks away from the best ask price when varying the number of buy order at the best bid price between one, ten and twenty.

the result displayed in Table 4.5 also indicates that proportional hazards assumption is violated in most of the case with some exceptions when comparing between the cases when the number of buy orders is one and ten at a limit order price of twelve ticks. The reason why we obtain a positive result in this situation is that the execution probability in these two situations is nearly identical as displayed in Figure 4.10a. Consequently, the proportional hazards model seems not to be a good candidate for modelling this effect as in two previous sections.

The plot of quantiles for different value of the number of buy orders at the best bid price, illustrated in Figure 4.11b, is somewhat parallel to each other when $\Delta = 6, 9$ and 12 but seem to be unparallel when $\Delta = 3$. The p-value obtained from the log-rank test reported in Table 4.6 has a similar result as all

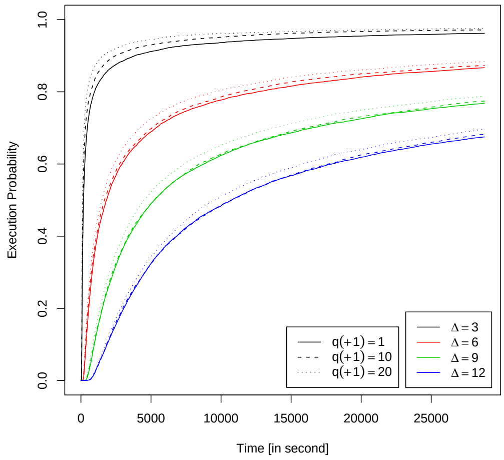
Group Δ	$q(-1)$	ψ	p-value	Group Δ	$q(-1)$	ψ	p-value	Group Δ	$q(-1)$	ψ	p-value
1	1, 10	0.9959	0.48	2	1, 10	4.0050	0.00	3	1, 10	2.1937	0.00
1	1, 20	1.0111	0.48	2	1, 20	8.3688	0.00	3	1, 20	4.7735	0.00
1	10, 20	1.0135	0.19	2	10, 20	2.0689	0.00	3	10, 20	2.1765	0.00
4	1, 10	1.2788	0.54	5	1, 10	1.1236	0.15	6	1, 10	1.0566	0.48
4	1, 20	1.7394	0.03	5	1, 20	1.3154	0.24	6	1, 20	1.2017	0.55
4	10, 20	1.2875	0.45	5	10, 20	1.1743	0.69	6	10, 20	1.1318	0.94
7	1, 10	1.0557	0.95	8	1, 10	1.0582	0.77	9	1, 10	1.0558	0.85
7	1, 20	1.2245	0.62	8	1, 20	1.1906	0.73	9	1, 20	1.1680	0.66
7	10, 20	1.1530	0.61	8	10, 20	1.1246	0.89	9	10, 20	1.1066	0.76
10	1, 10	1.0683	0.79	11	1, 10	1.0419	0.99	12	1, 10	1.0384	0.98
10	1, 20	1.1512	0.63	11	1, 20	1.1493	0.55	12	1, 20	1.1363	0.61
10	10, 20	1.0802	0.76	11	10, 20	1.1047	0.48	12	10, 20	1.0904	0.53
13	1, 10	1.0430	0.92	14	1, 10	1.0375	0.79	15	1, 10	1.0268	0.89
13	1, 20	1.1293	0.73	14	1, 20	1.1060	0.83	15	1, 20	1.0932	0.76
13	10, 20	1.0793	0.57	14	10, 20	1.0638	0.65	15	10, 20	1.0620	0.65
16	1, 10	1.0253	0.94	17	1, 10	1.0356	0.97	18	1, 10	1.0352	0.90
16	1, 20	1.0945	0.66	17	1, 20	1.0894	0.70	18	1, 20	1.0893	0.84
16	10, 20	1.0675	0.76	17	10, 20	1.0543	0.69	18	10, 20	1.0547	0.69

Table 4.6: The estimated parameters of the accelerated life time model together with the test statistics for the execution probability of a limit buy order when varying the number of buy order at the best bid price between one, ten and twenty.

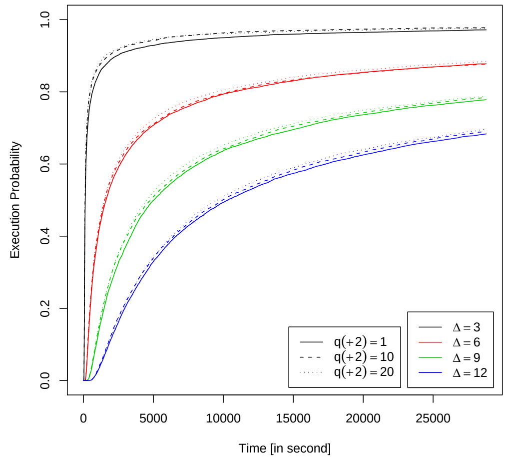
p-values when $\Delta = 6, 9$ and 12 are higher than 5% critical value while all p-values when $\Delta = 3$ is less than 5%. This indicates that the accelerated failure time assumption is not satisfied at all price levels. However, as discussed in the previous section, we might be able to divide this effect into two regimes, one of which can be modelled by the accelerated failure time model and the other cannot. To confirm this, we further apply the log-rank test to all price levels from one tick to eighteen ticks and the results reported in Table 4.6 indicate that a suitable criterion for separating this effect into two regimes is a limit order price of five ticks since all p-value when Δ is greater than or equal to five ticks are all larger than the 5% critical value. Furthermore, this result also suggests that there is an interaction effect between the number of buy orders at the best ask price and the limit order price on the execution probability, as the estimated time scale factor ψ at each price level is largely different. Accordingly, this interaction effect needs to be properly addressed when developing a full model of the execution probability.

Number of sell orders

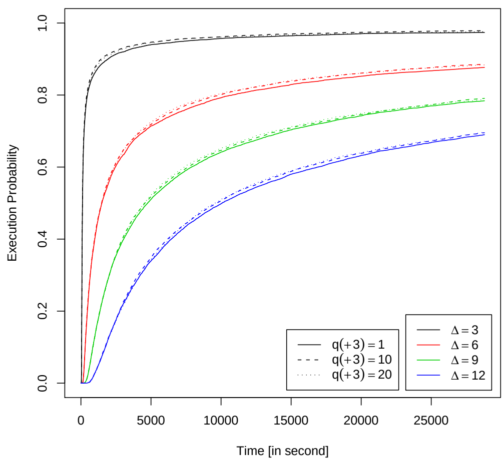
Let us now analyse the effect of the number of sell orders on the execution probability of limit buy orders. Similar to the previous section, we will utilise the data generated by simulating the SFGK model with the same setting as in Section 4.5.1, but varying the number of sell orders at some specified price levels between one, ten and twenty. The execution probability obtained by applying the Kaplan-Meier estimator to the data generated from the above simulation is illustrated in Figure 4.12. These results indicate that the execution probability of a buy order is positively correlated with the number of sell orders in the order book since the execution probability is higher when the number of sell orders is higher in all cases. Additionally the results also suggest that the number of sell orders at one tick away from the best bid price has a bigger effect on the execution probability than the number of sell orders at other price levels since the difference between the curves at each price level in Figure 4.12a is larger than



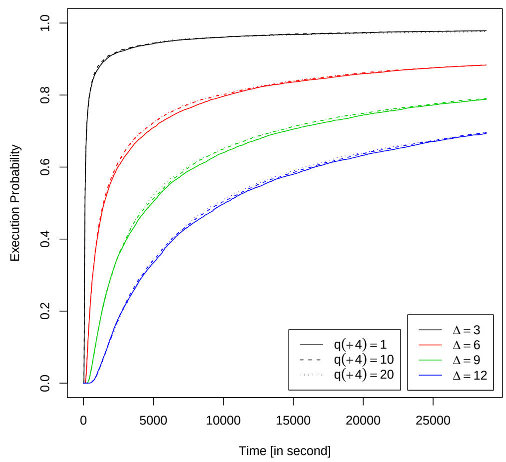
(a) One tick away from the best ask price



(b) Two ticks away from the best ask price



(c) Three tick away from the best ask price

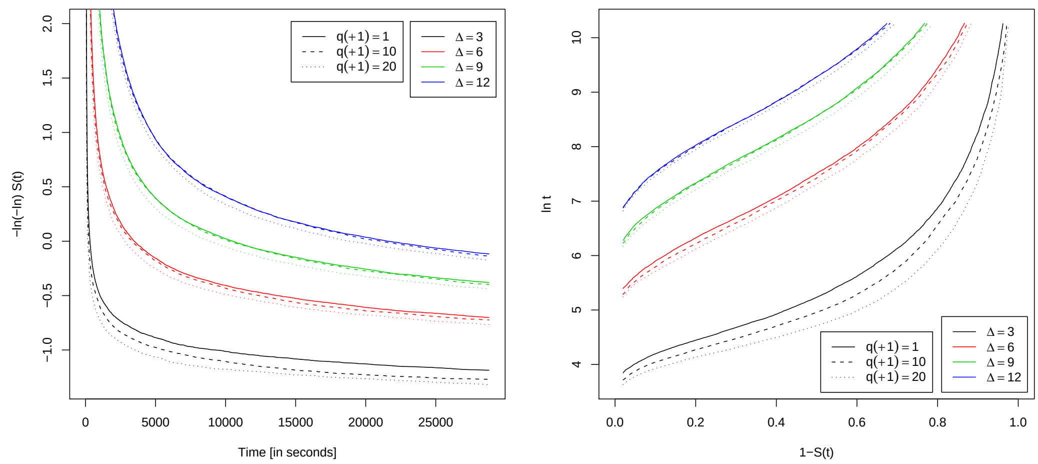


(d) Four ticks away from the best ask price

Figure 4.12: Execution probability of limit buy orders at a distance of $\Delta = 3, 6, 9$ and 12 ticks away from the best ask price when the number of sell orders at one, two, three and four ticks away from the best bid price are varied between one, ten and twenty as estimated from the Kaplan-Meier estimator.

those in Figure 4.12b, 4.12c, and 4.12d. Using the same argument, we can conclude that the number of sell orders at two ticks away from the best bid price has more effect than the number of sell orders at three and four ticks away from the best bid price. As a result, these indicate that the number of sell orders at a price level near the best bid price has more effect on the execution probability than the number of orders at price levels further away, and the number of sell orders at the best ask price has the biggest effect on the execution probability. Consequently, the rest of this section will study only the effect of the number of sell orders at the best ask price as in the previous section.

To investigate the validity of the proportional hazards assumption, we plot the log-log survival curves for different values of the number of sell orders at the best ask price in Figure 4.13a. This figure illustrates that the log-log survival curve for different values of sell orders at the same price level seem



(a) Log-log survival curve

(b) Log quantiles plot

Figure 4.13: The plot of log-log survival curve and the log quantiles associated with the estimated execution probability of limit buy orders at a distance of $\Delta = 3, 6, 9$ and 12 ticks away from the best ask price when the number of sell orders at the best ask price is one, ten and twenty as estimated from the Kaplan-Meier estimator.

Group		Propotional hazard model			Accelerated life time model		
		β	Chisq	p-value	ψ	Chisq	p-value
$\Delta = 3$	$q(+1) = 1, 10$	0.1624	108.52	0.00	0.6601	21.53	0.00
	$q(+1) = 1, 20$	0.3257	428.46	0.00	0.4270	126.01	0.00
	$q(+1) = 10, 20$	0.1635	107.70	0.00	0.6444	53.37	0.00
$\Delta = 6$	$q(+1) = 1, 10$	0.0373	9.12	0.00	0.9304	0.14	0.71
	$q(+1) = 1, 20$	0.1046	40.08	0.00	0.7693	1.35	0.255
	$q(+1) = 10, 20$	0.0669	10.25	0.00	0.8374	0.93	0.34
$\Delta = 9$	$q(+1) = 1, 10$	0.0155	0.10	0.75	0.9778	0.10	0.75
	$q(+1) = 1, 20$	0.0783	13.42	0.00	0.8437	0.19	0.66
	$q(+1) = 10, 20$	0.0628	10.72	0.00	0.8630	0.49	0.48
$\Delta = 12$	$q(+1) = 1, 10$	0.0140	0.57	0.45	0.9938	0.36	0.55
	$q(+1) = 1, 20$	0.0636	1.14	0.28	0.8905	0.04	0.85
	$q(+1) = 10, 20$	0.0498	3.27	0.07	0.9003	0.36	0.55

Table 4.7: The estimated parameters of the proportional hazards model and the accelerated life time model together with the test statistics for the execution probability of limit buy orders at a distance of three, six, nine and twelve ticks away from the best ask price when varying the number of sell order at the best ask price between one, ten and twenty.

to be unparallel to each other. The p-value obtained from the GT test reported in Table 4.7 indicates that most of the p-value are less than 5% critical value except when $\Delta = 12$. Although this suggests that we might be able to divide the effect of the number of sell orders into two regimes, one of which satisfies the proportional hazards assumption while the other does not, we decide not to investigate this in more detail since the proportional hazards model seem not to be a good candidate for modelling the effect of other determinants as reported in previous sections.

To determine whether the effect of the number of sell orders satisfies the accelerated failure time assumption or not, we plot the quantiles for different value of the number of sell orders at the best ask price in Figure 4.13b. The result indicates that the quantile plot when $\Delta = 3$ is clearly unparallel to each

Group Δ	Group $q(+1)$	ψ	p-value	Group Δ	Group $q(+1)$	ψ	p-value	Group Δ	Group $q(+1)$	ψ	p-value
1	1, 10	0.6012	0.00	2	1, 10	0.4158	0.00	3	1, 10	0.6601	0.00
1	1, 20	0.5982	0.00	2	1, 20	0.2143	0.00	3	1, 20	0.4270	0.00
1	10, 20	1.0110	0.07	2	10, 20	0.5243	0.00	3	10, 20	0.6444	0.00
4	1, 10	0.8280	0.87	5	1, 10	0.9107	0.56	6	1, 10	0.9304	0.71
4	1, 20	0.6410	0.04	5	1, 20	0.7474	0.53	6	1, 20	0.7693	0.25
4	10, 20	0.7720	0.02	5	10, 20	0.8202	0.27	6	10, 20	0.8374	0.34
7	1, 10	0.9441	0.76	8	1, 10	0.9607	0.70	9	1, 10	0.9778	0.75
7	1, 20	0.8111	0.68	8	1, 20	0.8312	0.94	9	1, 20	0.8437	0.66
7	10, 20	0.8613	0.57	8	10, 20	0.8602	0.58	9	10, 20	0.8630	0.48
10	1, 10	0.9976	0.61	11	1, 10	1.0000	0.58	12	1, 10	0.9938	0.55
10	1, 20	0.8835	0.86	11	1, 20	0.8920	0.95	12	1, 20	0.8905	0.85
10	10, 20	0.8901	0.59	11	10, 20	0.8930	0.67	12	10, 20	0.9003	0.55
13	1, 10	0.9934	0.75	14	1, 10	0.9909	0.62	15	1, 10	0.9945	0.68
13	1, 20	0.9051	0.75	14	1, 20	0.8963	0.89	15	1, 20	0.9169	0.89
13	10, 20	0.9078	0.92	14	10, 20	0.9050	0.62	15	10, 20	0.9212	0.80
16	1, 10	1.0014	0.53	17	1, 10	0.9933	0.60	18	1, 10	0.9927	0.73
16	1, 20	0.9229	0.76	17	1, 20	0.9155	0.87	18	1, 20	0.9257	0.94
16	10, 20	0.9218	0.74	17	10, 20	0.9208	0.69	18	10, 20	0.9294	0.71

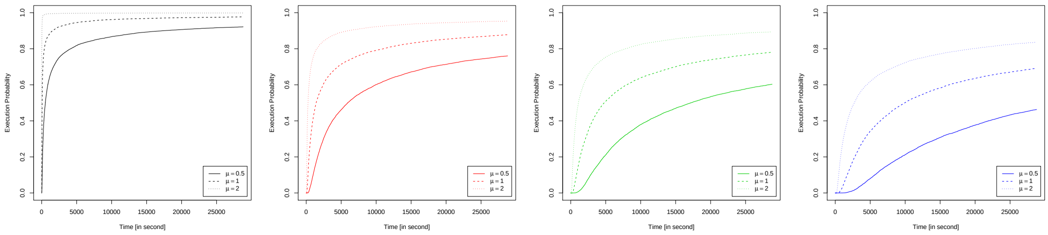
Table 4.8: The estimated parameters of the accelerated life time model together with the test statistics for the execution probability of limit buy orders when varying the number of sell order at the best ask price between one, ten and twenty.

other while the plot at other price levels seem to be parallel to each other. The p-value obtained from the log-rank test reported in Table 4.7 suggests a similar result. This indicates the possibility to divide the effect from the number of sell orders into two regimes as in the previous section. To confirm this, we further apply the log-rank test to all price level from one tick to eighteen ticks and the result reported in Table 4.8 suggests that a suitable threshold for separating this effect into two regimes is a limit order price value of five ticks as all p-value when $\Delta \geq 5$ is greater than 5% critical value. Similar to the previous section, this suggests that there is an interaction effect between the number of sell orders and the limit order price which should be addressed properly when developing a full execution probability model.

In summary, the execution probability of limit buy orders is negatively correlated with the number of buy orders in the order book, while it is positively correlated with the number of sell orders in the order book. Both graphical plots and test statistics indicates that it is more appropriate to utilise the accelerated failure time model to model this effect rather than the proportional hazards model. Unfortunately, the accelerated failure time assumption is not valid at all limit price levels, but is satisfied only when limit order price is greater than or equal to five ticks. This suggests that there is an interaction effect between the number of orders and the limit order price which must be properly modelled in the full execution probability model.

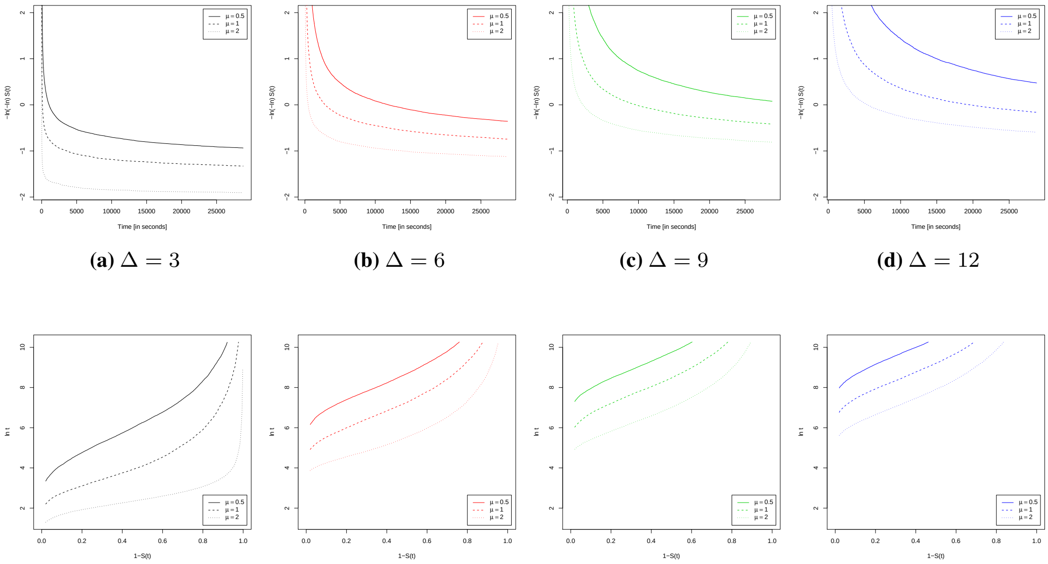
4.5.4 Arrival rate of market orders

This section investigates the effect of market order arrival rate on the execution probability by analysing the execution probability estimated from the SFGK model with the same parameters as described in Section 4.5.1 but varies the market order arrival rate between 0.5, 1.0 and 2.0. The execution probability of a buy order at four different price levels as estimated from the Kaplan-Meier estimator illustrated in



(a) $\Delta = 3$ **(b)** $\Delta = 6$ **(c)** $\Delta = 9$ **(d)** $\Delta = 12$

Figure 4.14: Execution probability of limit buy orders at a distance of $\Delta = 3, 6, 9$ and 12 ticks away from the best ask price when the market order arrival rate is 0.5, 1 and 2 as estimated from the Kaplan-Meier estimator.



(a) $\Delta = 3$ **(b)** $\Delta = 6$ **(c)** $\Delta = 9$ **(d)** $\Delta = 12$

(e) $\Delta = 3$ **(f)** $\Delta = 6$ **(g)** $\Delta = 9$ **(h)** $\Delta = 12$

Figure 4.15: The plot of log-log survival curve and the log quantiles associated with the estimated execution probability of limit buy orders at a distance of $\Delta = 3, 6, 9$ and 12 ticks away from the best ask price when the market order arrival rate is 0.5, 1.0 and 2.0 as estimated from the Kaplan-Meier estimator.

Figure 4.14 indicates that the execution probability is positively correlated with the market order arrival rate since all the execution probabilities displayed in the figure are higher when the market order arrival rate is larger. This result can be explained if we consider the situation when the limit order arrival rate and order cancellation rate is fixed. As the market order arrival rate increases, more limit orders are removed from the book. Consequently the execution probability is also higher.

The log-log survival curves for a limit buy order at different market order arrival rates illustrated in Figure 4.15a to 4.15d are clearly unparallel to each other, suggesting that the effect of market order arrival rate does not satisfy the proportional hazards assumption. This is further confirmd by the GT test summarised in Table 4.9, where the p-values at all price levels are less than 5% critical value. Thus the effect of the bid-ask spread does not satisfy the proportional hazards assumption and the proportional hazards model should not be utilised to model this effect.

Group		Propotional hazard model			Accelerated life time model		
		β	Chisq	p-value	ψ	Chisq	p-value
$\Delta = 3$	$\alpha = 0.5, 1.0$	0.8390	1881.29	0.00	0.0891	60.22	0.00
	$\alpha = 0.5, 2.0$	2.4892	1570.23	0.00	0.0037	959.47	0.00
	$\alpha = 1.0, 2.0$	1.5157	1261.20	0.00	0.0212	2257.87	0.00
$\Delta = 6$	$\alpha = 0.5, 1.0$	0.6249	1407.74	0.00	0.2456	0.37	0.54
	$\alpha = 0.5, 2.0$	1.3198	3275.56	0.00	0.0393	10.33	0.00
	$\alpha = 1.0, 2.0$	0.7563	1957.18	0.00	0.1433	33.67	0.00
$\Delta = 9$	$\alpha = 0.5, 1.0$	0.6679	1108.37	0.00	0.2835	0.00	0.97
	$\alpha = 0.5, 2.0$	1.2612	2801.02	0.00	0.0652	0.68	0.41
	$\alpha = 1.0, 2.0$	0.6543	1524.72	0.00	0.2264	0.79	0.37
$\Delta = 12$	$\alpha = 0.5, 1.0$	0.7693	944.42	0.00	0.2927	0.00	1.00
	$\alpha = 0.5, 2.0$	1.3378	2296.56	0.00	0.0799	0.04	0.85
	$\alpha = 1.0, 2.0$	0.6300	1234.50	0.00	0.2748	0.08	0.78

Table 4.9: The estimated parameters of the proportional hazards model and the accelerated life time model together with the test statistics for the execution probability of limit buy orders at a distance of three, six, nine and twelve ticks away from the best ask price when varying the market order arrival rate between 1.0, 2.0 and 4.0.

The test of the accelerated failure time model is more positive. The quantile curves for each value of market order arrival rate displayed in Figure 4.15e to 4.15h seem to be parallel to each other when $\Delta = 9$ and 12, while clearly unparallel when $\Delta = 3$ and 6. This suggests that it might be possible to divide the effect of market order arrival rate into two regimes as in previous sections. To verify this, we apply the log-rank test to all price levels from one tick to eighteen ticks, and the result reported in Table 4.10 indicates that a suitable criteria for separating this effect into two regimes is a limit order price value of eight ticks since all p-value when Δ is greater or equal to eight ticks is well above the 5% critical value while the p-value when Δ is less than eight ticks are not. Additionally, the dependency on the limit price together with the fact that the estimated time scale factor, ψ, at each price level is largely different suggests that there is an interaction effect between the market order arrival rate and the limit order price on the execution probability.

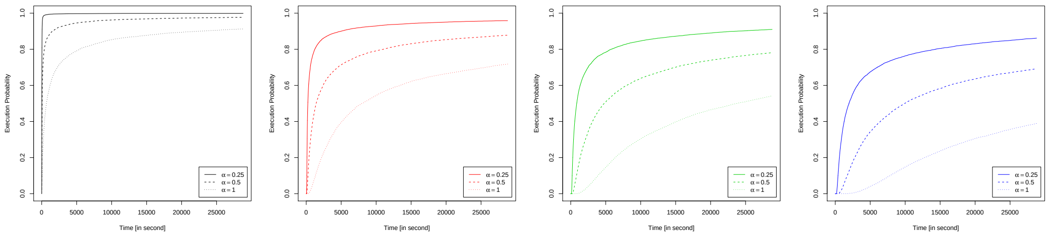
To sum up, the execution probability is positively correlated with the market order arrival rate, and it is more appropriate to utilise the accelerated failure time model to model this effect rather than the proportional hazards model. Though the accelerated failure time assumption is not valid at all price levels, we can divide this effect into two regimes by the limit order price so that one of these regimes can be modelled by the accelerated failure time model. The dependency on the limit order price suggests an interaction effect between the market order arrival rate and the limit order price on the execution probability.

4.5.5 Arrival rate of limit orders

Let us now study the effect of the limit order arrival rate on the execution probability. To achieve this, we simulate the SFGK model with the same parameters setting as described in Section 4.5.1 but varying the limit order arrival rate between 0.25, 0.5 and 1.0. The execution probabilities obtained by applying the Kaplan-Meier estimator to the data generated from the above simulations are illustrated in Figure 4.16. The result indicates that the execution probability is lower when the limit order arrival rate is higher, indicating a negative relation between the execution probability and the limit order arrival rate. This is

Group		ψ	p-value	Group		ψ	p-value	Group		ψ	p-value
Δ	α			Δ	α			Δ	α		
1	0.5, 1.0	0.4776	0.41	2	0.5, 1.0	0.0262	0.00	3	0.5, 1.0	0.0891	0.00
1	0.5, 2.0	0.2415	0.51	2	0.5, 2.0	0.0037	0.00	3	0.5, 2.0	0.0037	0.00
1	1.0, 2.0	0.5056	0.71	2	1.0, 2.0	0.1130	0.00	3	1.0, 2.0	0.0212	0.00
4	0.5, 1.0	0.1690	0.03	5	0.5, 1.0	0.2191	0.31	6	0.5, 1.0	0.2456	0.54
4	0.5, 2.0	0.0109	0.00	5	0.5, 2.0	0.0233	0.00	6	0.5, 2.0	0.0393	0.00
4	1.0, 2.0	0.0456	0.00	5	1.0, 2.0	0.0894	0.00	6	1.0, 2.0	0.1433	0.00
7	0.5, 1.0	0.2696	0.95	8	0.5, 1.0	0.2749	0.97	9	0.5, 1.0	0.2835	0.97
7	0.5, 2.0	0.0507	0.11	8	0.5, 2.0	0.0581	0.16	9	0.5, 2.0	0.0652	0.41
7	1.0, 2.0	0.1783	0.01	8	1.0, 2.0	0.2052	0.11	9	1.0, 2.0	0.2264	0.37
10	0.5, 1.0	0.2912	0.91	11	0.5, 1.0	0.2903	0.82	12	0.5, 1.0	0.2927	1.00
10	0.5, 2.0	0.0727	0.56	11	0.5, 2.0	0.0762	0.69	12	0.5, 2.0	0.0799	0.85
10	1.0, 2.0	0.2485	0.55	11	1.0, 2.0	0.2593	0.69	12	1.0, 2.0	0.2748	0.78
13	0.5, 1.0	0.2988	0.84	14	0.5, 1.0	0.3045	0.92	15	0.5, 1.0	0.3052	0.92
13	0.5, 2.0	0.0840	0.83	14	0.5, 2.0	0.0870	0.80	15	0.5, 2.0	0.0886	0.79
13	1.0, 2.0	0.2782	0.95	14	1.0, 2.0	0.2809	0.82	15	1.0, 2.0	0.2841	0.98
16	0.5, 1.0	0.3067	0.98	17	0.5, 1.0	0.3130	0.90	18	0.5, 1.0	0.3115	0.95
16	0.5, 2.0	0.0909	0.87	17	0.5, 2.0	0.0933	0.96	18	0.5, 2.0	0.0949	0.69
16	1.0, 2.0	0.2883	0.97	17	1.0, 2.0	0.2934	0.99	18	1.0, 2.0	0.3008	0.96

Table 4.10: The estimated parameters of the accelerated life time model together with the test statistics for the execution probability of limit buy orders when varying the market order arrival rate between 1.0, 2.0 and 4.0.



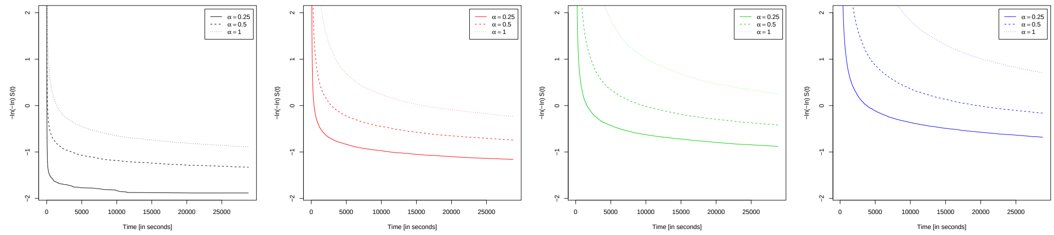
(a) $\Delta = 3$ **(b)** $\Delta = 6$ **(c)** $\Delta = 9$ **(d)** $\Delta = 12$

Figure 4.16: Execution probability of limit buy orders at a distance of $\Delta = 3, 6, 9$ and 12 ticks away from the best ask price when the limit order arrival rate is 0.25, 0.5 and 1 as estimated from the Kaplan-Meier estimator.

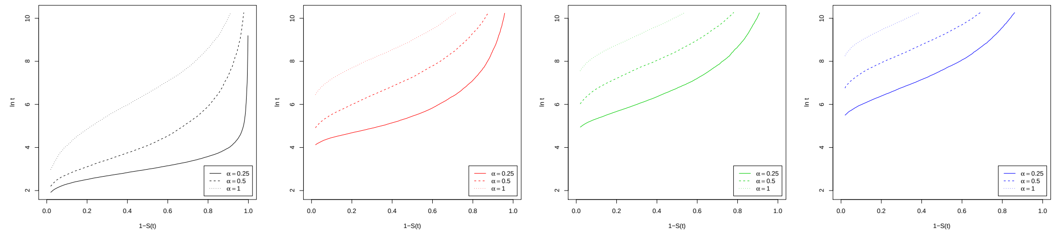
because the larger the limit order arrival rate, the more limit orders are submitted to the order book and the more market orders are required to move the price. As a result, when the market order arrival rate and order cancellation rate are fixed, the execution probability will decrease as the limit order arrival rate increases.

To determine whether the proportional hazards model is an appropriate candidate for modelling the effect of limit order arrival rate on the execution probability or not, we plot the log-log survival curve for three values of limit order arrival rate at four different price levels in Figure 4.17a to 4.17d. These log-log survival curves are clearly unparallel to each other, indicating that the proportional hazards assumption might be unsatisfied in this case. This is further confirmed by the GT test reported in Table 4.11 where all p-value reported are less than 5% critical value.

To access the appropriateness of the accelerated failure time model, we plot the quantile curve for each value of limit order arrival in Figure 4.17e to 4.17h. Although these curves are clearly unparallel



(a) $\Delta = 3$ **(b)** $\Delta = 6$ **(c)** $\Delta = 9$ **(d)** $\Delta = 12$



(e) $\Delta = 3$ **(f)** $\Delta = 6$ **(g)** $\Delta = 9$ **(h)** $\Delta = 12$

Figure 4.17: The plot of log-log survival curve and the log quantiles associated with the estimated execution probability of limit buy orders at a distance of $\Delta = 3, 6, 9$ and 12 ticks away from the best ask price when the limit order arrival rate is $0.25, 0.5$ and 1.0 as estimated from the Kaplan-Meier estimator.

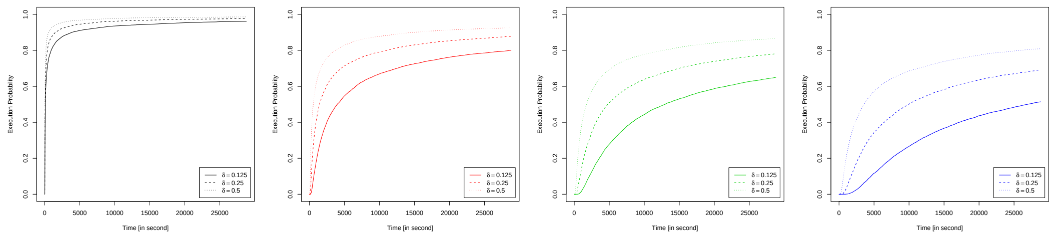
Group		Propotional hazard model			Accelerated life time model		
		β	Chisq	p-value	ψ	Chisq	p-value
$\Delta = 3$	$\alpha = 0.25, 0.5$	-1.1862	146.47	0.00	30.6049	2486.15	0.00
	$\alpha = 0.25, 1.0$	-2.3622	948.88	0.00	191.6476	802.67	0.00
	$\alpha = 0.5, 1.0$	-0.9034	1617.86	0.00	14.5622	63.30	0.00
$\Delta = 6$	$\alpha = 0.25, 0.5$	-0.8202	2028.45	0.00	8.4352	89.98	0.00
	$\alpha = 0.25, 1.0$	-1.5183	3274.64	0.00	38.3713	21.09	0.00
	$\alpha = 0.5, 1.0$	-0.7743	1770.31	0.00	5.5544	0.03	0.86
$\Delta = 9$	$\alpha = 0.25, 0.5$	-0.7683	1917.68	0.00	5.9837	7.08	0.01
	$\alpha = 0.25, 1.0$	-1.5316	2869.82	0.00	26.2853	0.02	0.89
	$\alpha = 0.5, 1.0$	-0.8579	1446.49	0.00	4.8651	0.05	0.82
$\Delta = 12$	$\alpha = 0.25, 0.5$	-0.7682	1705.27	0.00	5.1122	0.97	0.32
	$\alpha = 0.25, 1.0$	-1.6796	2357.44	0.00	22.1503	1.99	0.16
	$\alpha = 0.5, 1.0$	-1.0126	1214.94	0.00	4.6430	0.05	0.82

Table 4.11: The estimated parameters of the proportional hazards model and the accelerated life time model together with the test statistics for the execution probability of limit buy orders at a distance of three, six, nine and twelve ticks away from the best ask price when varying the limit order arrival rate between 0.25, 0.5 and 1.0.

when $\Delta = 3, 6$ and 9, the curve when $\Delta = 12$ seem to be parallel to each other. This suggests that it might be possible to divide the effect of limit order arrival rate into two regimes, one of which can be modelled by the accelerated failure time model while the other cannot. To further investigate this issue, we apply the log-rank test to all price levels from one tick to eighteen ticks, and the result displayed in Table 4.12 indicates that all p-value when limit order price is larger than or equal to ten ticks are always larger than 5% critical value. This supports our hypothesis that the effect of limit order arrival rate can be divided into two regimes, and an appropriate threshold for dividing this is a limit order price of ten ticks. Since the estimated time scale factor, ψ, at each price level is largely different, there must be an

Group		ψ	p-value	Group		ψ	p-value	Group		ψ	p-value
Δ	α			Δ	α			Δ	α		
1	0.25, 0.5	1.03	0.60	2	0.25, 0.5	4.72	0.00	3	0.25, 0.5	30.60	0.00
1	0.25, 1.0	1.05	0.73	2	0.25, 1.0	185.94	0.00	3	0.25, 1.0	191.65	0.00
1	0.5, 1.0	1.02	0.59	2	0.5, 1.0	49.44	0.00	3	0.5, 1.0	14.56	0.00
4	0.25, 0.5	19.64	0.00	5	0.25, 0.5	11.69	0.00	6	0.25, 0.5	8.44	0.00
4	0.25, 1.0	89.83	0.00	5	0.25, 1.0	52.44	0.00	6	0.25, 1.0	38.37	0.00
4	0.5, 1.0	7.84	0.04	5	0.5, 1.0	6.07	0.75	6	0.5, 1.0	5.55	0.86
7	0.25, 0.5	7.23	0.00	8	0.25, 0.5	6.40	0.00	9	0.25, 0.5	5.98	0.01
7	0.25, 1.0	31.81	0.04	8	0.25, 1.0	28.33	0.12	9	0.25, 1.0	26.29	0.89
7	0.5, 1.0	5.16	0.99	8	0.5, 1.0	4.94	0.87	9	0.5, 1.0	4.87	0.82
10	0.25, 0.5	5.60	0.07	11	0.25, 0.5	5.43	0.15	12	0.25, 0.5	5.11	0.32
10	0.25, 1.0	24.46	0.52	11	0.25, 1.0	23.73	0.46	12	0.25, 1.0	22.15	0.16
10	0.5, 1.0	4.75	0.87	11	0.5, 1.0	4.71	0.92	12	0.5, 1.0	4.64	0.82
13	0.25, 0.5	5.05	0.64	14	0.25, 0.5	4.87	0.98	15	0.25, 0.5	4.86	0.94
13	0.25, 1.0	21.50	0.07	14	0.25, 1.0	20.76	0.14	15	0.25, 1.0	20.38	0.20
13	0.5, 1.0	4.57	0.90	14	0.5, 1.0	4.53	0.90	15	0.5, 1.0	4.51	0.88
16	0.25, 0.5	4.82	0.71	17	0.25, 0.5	4.81	0.74	18	0.25, 0.5	4.78	0.95
16	0.25, 1.0	19.85	0.08	17	0.25, 1.0	19.84	0.12	18	0.25, 1.0	19.59	0.08
16	0.5, 1.0	4.48	0.82	17	0.5, 1.0	4.45	0.56	18	0.5, 1.0	4.44	0.74

Table 4.12: The estimated parameters of the accelerated life time model together with the test statistics for the execution probability of limit buy orders when varying the limit order arrival rate between 0.25, 0.5 and 1.0.



(a) $\Delta = 3$ **(b)** $\Delta = 6$ **(c)** $\Delta = 9$ **(d)** $\Delta = 12$

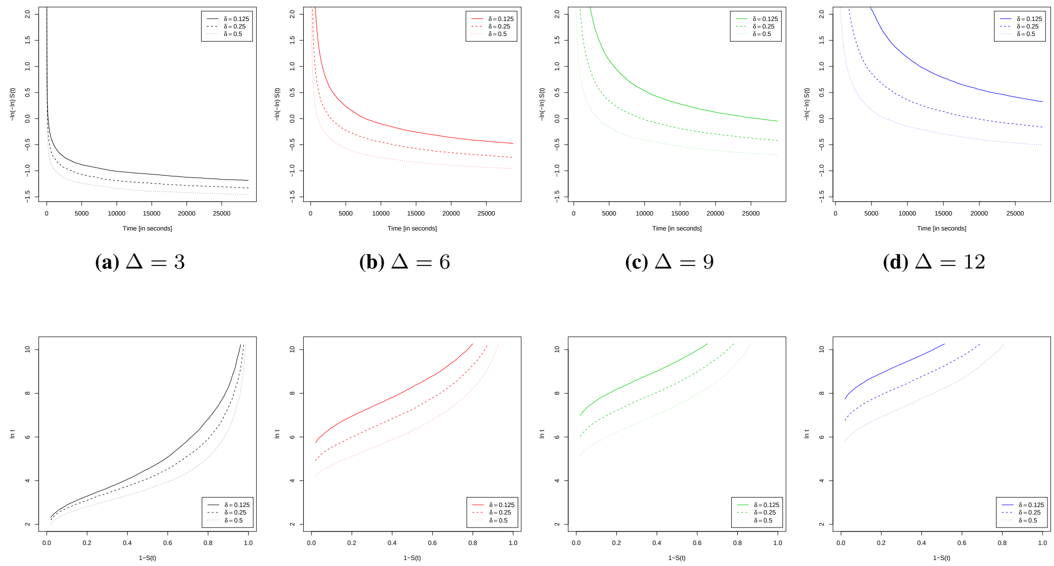
Figure 4.18: Execution probability of limit buy orders at a distance of $\Delta = 3, 6, 9$ and 12 ticks away from the best ask price when the order cancellation rate is 0.125, 0.25 and 0.5 as estimated from the Kaplan-Meier estimator.

interaction effect between the limit order arrival rate and limit price which should be handled properly when developing the full model of the execution probability.

In conclusion, the execution probability is negatively correlated with the limit order arrival rate. This effect can be divided into two regimes by limit order price, one of which can be modelled by the accelerated failure time model while the other cannot. The dependency on the limit order price suggests an interaction effect between the limit order arrival rate and the limit order price on the execution probability.

4.5.6 Cancellation rate of limit orders

This section studies the effect of the limit order cancellation rate on the execution probability by analysing the data generated from the SFGK model with the same parameters setting as described in Section 4.5.1 but varies the limit order cancellation rate between 0.125, 0.25 and 0.5. The execution probabilities obtained by applying the Kaplan-Meier estimator to the data generated from the above sim-



(a) $\Delta = 3$ (b) $\Delta = 6$ (c) $\Delta = 9$ (d) $\Delta = 12$

(e) $\Delta = 3$ (f) $\Delta = 6$ (g) $\Delta = 9$ (h) $\Delta = 12$

Figure 4.19: The plot of log-log survival curve and the log quantiles associated with the estimated execution probability of limit buy orders at a distance of $\Delta = 3, 6, 9$ and 12 ticks away from the best ask price when the order cancellation rate is $0.125, 0.25$ and 0.5 as estimated from the Kaplan-Meier estimator.

Group		Propotional hazard model			Accelerated life time model		
		β	Chisq	p-value	ψ	Chisq	p-value
$\Delta = 3$	$\delta = 0.125, 0.25$	0.2313	18.43	0.00	2.5395	105.44	0.00
	$\delta = 0.125, 0.5$	0.5170	240.17	0.00	6.7252	404.12	0.00
	$\delta = 0.25, 0.5$	0.2932	148.42	0.00	2.7462	180.42	0.00
$\Delta = 6$	$\delta = 0.125, 0.25$	0.4373	724.12	0.00	2.7282	0.70	0.40
	$\delta = 0.125, 0.5$	0.8391	2147.17	0.00	7.5699	2.36	0.12
	$\delta = 0.25, 0.5$	0.4230	770.58	0.00	2.8976	4.70	0.03
$\Delta = 9$	$\delta = 0.125, 0.25$	0.5099	726.54	0.00	2.6843	0.05	0.82
	$\delta = 0.125, 0.5$	0.9460	2190.16	0.00	7.4257	0.01	0.90
	$\delta = 0.25, 0.5$	0.4665	871.41	0.00	2.8329	0.37	0.54
$\Delta = 12$	$\delta = 0.125, 0.25$	0.6057	715.49	0.00	2.6948	0.12	0.73
	$\delta = 0.125, 0.5$	1.0735	2038.03	0.00	7.4253	0.12	0.73
	$\delta = 0.25, 0.5$	0.5043	854.31	0.00	2.7806	0.07	0.80

Table 4.13: The estimated parameters of the proportional hazards model and the accelerated life time model together with the test statistics for the execution probability of limit buy orders at a distance of three, six, nine and twelve ticks away from the best ask price when varying the order cancellation rate between 0.125, 0.25 and 0.5.

ulations are displayed in Figure 4.18. The result indicates that the execution probability is positively correlated with the limit order cancellation rate since all the execution probabilities displayed in this figure are higher when the order cancellation rate is larger. This is because when the order cancellation rate increases, the number of limit orders in the book decreases, and the number of market orders required to move the price also decreases. As a result, when market and limit order arrival rates are fixed, the execution probability will increase as the order cancellation rate increases.

The test of proportional hazards assumption by the plot of log-log survival curve suggest that this assumption might not be satisfied since the log-log curve at four different limit prices illustrated in

Group Δ	Group α	ψ	p-value	Group Δ	Group α	ψ	p-value	Group Δ	Group α	ψ	p-value
1	0.125, 0.25	1.04	0.47	2	0.125, 0.25	1.25	0.00	3	0.125, 0.25	2.54	0.00
1	0.125, 0.5	1.04	0.59	2	0.125, 0.5	1.75	0.00	3	0.125, 0.5	6.73	0.00
1	0.25, 0.5	1.00	0.92	2	0.25, 0.5	1.37	0.00	3	0.25, 0.5	2.75	0.00
4	0.125, 0.25	2.89	0.00	5	0.125, 0.25	2.71	0.14	6	0.125, 0.25	2.73	0.40
4	0.125, 0.5	7.48	0.00	5	0.125, 0.5	7.47	0.01	6	0.125, 0.5	7.57	0.12
4	0.25, 0.5	2.68	0.00	5	0.25, 0.5	2.89	0.00	6	0.25, 0.5	2.90	0.03
7	0.125, 0.25	2.70	0.57	8	0.125, 0.25	2.66	0.65	9	0.125, 0.25	2.68	0.82
7	0.125, 0.5	7.58	0.37	8	0.125, 0.5	7.48	0.58	9	0.125, 0.5	7.43	0.90
7	0.25, 0.5	2.91	0.16	8	0.25, 0.5	2.89	0.35	9	0.25, 0.5	2.83	0.54
10	0.125, 0.25	2.65	0.84	11	0.125, 0.25	2.67	0.96	12	0.125, 0.25	2.69	0.73
10	0.125, 0.5	7.44	0.86	11	0.125, 0.5	7.46	0.91	12	0.125, 0.5	7.43	0.73
10	0.25, 0.5	2.83	0.84	11	0.25, 0.5	2.82	0.82	12	0.25, 0.5	2.78	0.80
13	0.125, 0.25	2.65	0.48	14	0.125, 0.25	2.70	0.57	15	0.125, 0.25	2.69	0.58
13	0.125, 0.5	7.40	0.47	14	0.125, 0.5	7.45	0.47	15	0.125, 0.5	7.43	0.40
13	0.25, 0.5	2.83	0.95	14	0.25, 0.5	2.79	1.00	15	0.25, 0.5	2.79	0.82
16	0.125, 0.25	2.68	0.73	17	0.125, 0.25	2.66	0.58	18	0.125, 0.25	2.68	0.54
16	0.125, 0.5	7.42	0.40	17	0.125, 0.5	7.43	0.28	18	0.125, 0.5	7.44	0.38
16	0.25, 0.5	2.79	0.93	17	0.25, 0.5	2.81	0.97	18	0.25, 0.5	2.79	0.82

Table 4.14: The estimated parameters of the accelerated life time model together with the test statistics for the execution probability of limit buy orders when varying the order cancellation rate between 0.125, 0.25 and 0.5.

Figure 4.19a to Figure 4.19d are clearly unparallel to each other. This is further confirmed by the GT test reported in Table 4.13 where the p-values at all price levels are less than 5% critical value. Hence, the effect of order cancellation rate does not satisfy the proportional hazards assumption and the proportional hazards model should not be utilised to model this effect.

The test of accelerated failure time assumption by the plot of quantiles suggests that this assumption is valid at some price levels since the quantiles plot at four different price levels displayed in Figure 4.19e to Figure 4.19h is clearly unparallel when $\Delta = 3$ and 6 but somewhat parallel when $\Delta = 9$ and 12. To further examine this, we apply the log-rank test to all price levels from one tick to eighteen ticks and the results reported in Table 4.14 indicates that all p-value when limit order price is higher than or equal to seven ticks are all higher than 5% critical value. This indicates that it might be possible to divide the effect of order cancellation rate into two regimes, as in previous sections. Additionally, the fact that the estimated time scale factor, ψ, at each price level is largely different also suggests an interaction effect between the order cancellation rate and the limit order price on the execution probability, which must be properly addressed when developing a full model of the execution probability.

In summary, the execution probability is positively correlated with the order cancellation rate and it is more appropriate to utilise the accelerated failure time model to model this effect rather than the proportional hazards model. Nevertheless, the accelerated failure time assumption is not satisfied at all price levels, but we can divide this effect into two regimes by limit order price so that one of which satisfies the accelerated failure time assumption while the other cannot. The dependency on the limit order price suggests an interaction effect between the order cancellation rate and the limit order price on the execution probability, which should be addressed properly when developing a full model of the execution probability.

4.6 Distribution of execution probability

Unlike the previous two sections that estimate the execution probability using non-parametric methods, this section will estimate the execution probability using parametric models with the aim to identify the most suitable distribution for modelling the execution probability. Since the result in Section 4.4 suggests that the most appropriate method for analysing the execution probability is the Kaplan-Meier estimator, which is a non-parametric survival analysis method, this section will utilise parametric survival analysis methods to model the execution probability.

As described in Section 2.4.3, parametric methods generally assume a specific parametric family of survival distribution and estimate its parameters from the dataset using maximum likelihood methods. Although any distribution over nonnegative values can be utilised to model the survival time distribution, this section will focus only on the four most widely used distributions which are Weibull, log-normal, log-logistic, and generalised gamma distribution. Among these four distributions, generalised gamma distribution is the most complicated model which has exponential, Weilbull, log-normal and gamma distribution as its special cases, and, thus, we expect it to be the most appropriate model for modelling the execution probability.

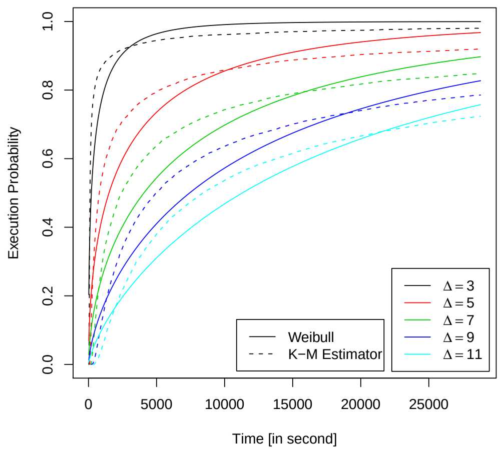
Let us now analyse the distribution of the execution probability in the SFGK model. Since the execution probability depends on several factors, we will analyse this distribution when all determinants of the execution probability described in the previous section are kept fixed at some particular value so that the data utilised in this analysis is not depend on any variables. To achieve this, we simulate the SFGK model with parameter $\mu = 2, \alpha = 0.5$ and $\delta = 0.025$ for $10,000$ rounds as described in Section 4.5.1. Particularly, in each round, the initial bid-ask spread is set to one tick while the number of order at all price levels is set to 20, and the simulation is run until the simulation time reaches eight hours. The first time that the transactional price reaches or crosses each price level is then recorded and utilised as an estimation of the execution time of limit orders at the corresponding price level.

The results obtained from fitting Weibull, log-normal, log-logistic and generalised gamma distributions to the data generated from the above simulations at each price level separately are displayed in Figure 4.20. As expected, the results indicate that among these four distributions, generalised gamma distribution is the best distribution for modelling the distribution of the execution probability, since its curve is the closest to the result from the Kaplan-Meier estimator.

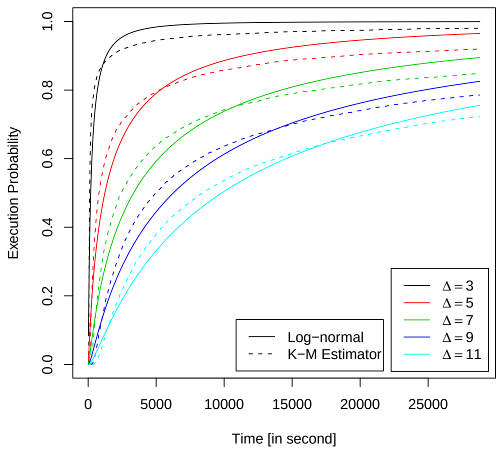
To confirm that the accelerated failure time model is not appropriate for modelling the execution probability in the SFGK model since the effect of limit price does not satisfy the accelerate failure time assumption, we further fit the accelerated failure time model with generalised gamma distribution as a baseline distribution. Since the effect from the limit order price may not be linear, this dependency is modelled by Chebyshev polynomial degree eight where Chebyshev polynomial degree n, $T_n(x)$, is defined by

$$T_n(x) = \cosh(n \operatorname{arccosh}(x)). \tag{4.17}$$

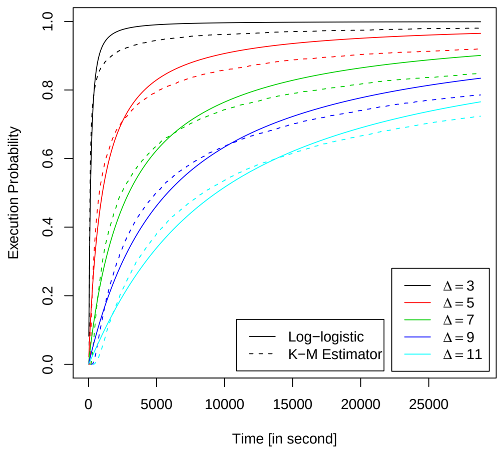
Figure 4.21a illustrates the execution probability obtained by fitting the accelerated failure time model to the execution time of limit buy orders at one to forty ticks away from the best ask price. The result



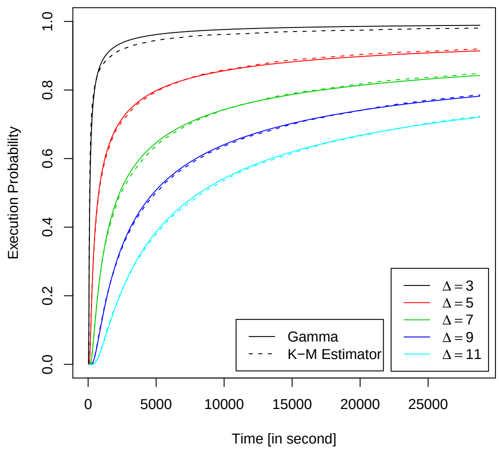
(a) Weibull distribution



(b) Log-normal distribution



(c) Log-logistic distribution

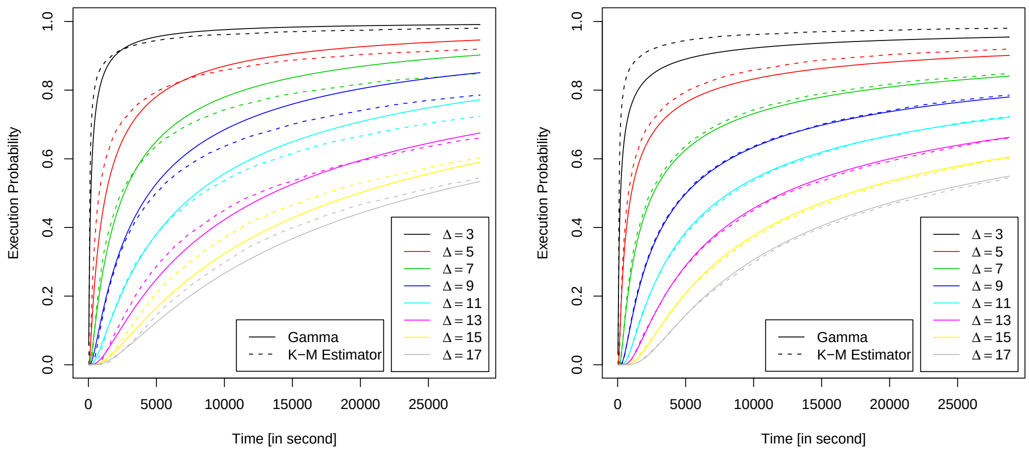


(d) Gamma distribution

Figure 4.20: Comparison between execution probability estimated from the Kaplan-Meier estimator and the parametric methods assuming Weibull (a), log-normal (b), log-logistic (c) and generalised gamma distribution.

clearly indicates the inability to fit the data of the accelerated failure time model as discussed in Section 4.5.1. However, the result when fitting the model to the execution time of limit buy orders at six to forty ticks[8] away from the best ask price displayed in Figure 4.21b lies nicely with the one obtained from the Kaplan-Meier estimator when $\Delta \geq 6$. This further confirms the fact that the effect of limit order price can be divided into two regimes, one of which can be modelled by the accelerated failure time model and the other cannot.

[8]We utilised the data from six ticks because the result in Section 4.5.1 indicate that effect of limit order price can be divided into two regimes one of which satisfies the accelerated failure time assumption and a suitable threshold dividing this a limit order price of six.



(a) Using all price levels levels

(b) Using only price levels greater or equal to six ticks distribution

Figure 4.21: Comparision between the execution probability estimated from the Kaplan Meier estimator and the accelerated failure time model when fit the model with the execution time data of limit buy orders at all price levels (a) and all price levels greater than or equal to six ticks away from the best ask price (b).

4.7 Summary

This chapter presents an in-depth review of previously proposed methods for modelling the execution probability. The experiment with data generated from the SFGK model in Section 4.4 indicates that the execution probability model and the execution time model produce similar results when the number of cancelled orders is small, and they start to produce different results when there are more cancelled orders. Among all models considered, the execution time model that utilises the Kaplan-Meier estimator to handle cancelled orders seems to be the best performing method from both theoretical and empirical points of view. The choice between the first-passage time and the execution time model depends on the problem faced. Specifically, the execution time model generally provides a better estimation of the real execution probability comparing to the first-passage time model since the first-passage time model generally overestimates the real execution probability and will equal to the real execution probability only when the considered order is at the top of the queue. However, the result obtained from the execution time model might have a large standard error when analysing the execution probability over a long time horizon, while the result from first-passage time has a small standard error over all time horizons. The experiment with theoretical first-passage time model indicates that the estimated execution probability can be quite different from empirical first-passage time model when the assumption about the asset price dynamic is incorrect.

Section 4.5 analyses the relationship between the execution probability and other variables in the SFGK model. The result indicates that the execution probability of a limit buy order is positively correlated with bid-ask spread, number of sell orders in the order book, market order arrival rate and order cancellation rate, while it is negatively correlated with the distance from the opposite best price, the number of buy orders in the order book and the limit order arrival rate. Both graphical diagnostics and test statistics suggest that it is more appropriate to model the effects from these determinants by the accelerated failure time model rather than the proportional hazards model, since these effects do not satisfy the proportional hazards assumption but satisfy the accelerated failure time assumption. Unfortunately, the accelerated failure time assumption is not satisfied at all price levels but is satisfied only when the distance from the opposite best price is higher than a threshold value which varies from effect to effect. This limitation makes it inappropriate to directly apply the accelerated failure time model to model the execution probability at all price levels and, hence, other alternatives are required to model the execution probability at all price levels properly. Additionally, the dependency on the limit order price also suggests an interaction effect between the limit order price and other determinants on the execution probability. This suggests that the full model of the execution probability must also needs to handle these interaction effects properly in order to obtain a good result.

In Section 4.6, we perform the experiment to determine the most appropriate distribution for modelling the execution probability. The result indicates that among the four most widely used distributions (i.e. Weibull, log-normal, log-logistic and generalised gamma distribution), the generalised gamma distribution is the most appropriate distribution for modelling the execution probability. We also confirm the fact that the accelerated failure time assumption is not satisfied at all price levels but satisfied only when the distance from the opposite best price is higher than a threshold value by fitting the accelerated failure time model with the execution time at all price levels and the execution time at all price levels is greater than six ticks. The result clearly indicates that the first model does not fit the data while the second models fit the data very well especially when the limit order price is larger than six ticks.

In conclusion, the experiments in this section indicate that although survival analysis is the most appropriate method for modelling the execution probability both from theoretical and empirical point of view, directly applying traditional survival analysis techniques (i.e. the proportional hazards model and the accelerated failure time model) to model the execution time data may not be appropriate since the effect from all determinants does not satisfy the assumptions of those techniques. Consequently, a new method that does not suffer from this limitation is required to model the execution probability more properly, and this will be the main subject of the two following chapters.

Chapter 5

Execution probability and price fluctuation

This chapter proposes a new framework for modelling the execution probability at a specified time period from the distribution of asset price fluctuations during the interested period. The advantage of this approach over traditional techniques is that it requires less data, as it requires only one record per sample while traditional models generally require n records per sample to model the execution probability at n price levels. Additionally, it also provides a natural way to apply traditional time series analysis techniques to model the execution probability. By applying the proposed approach to the historical dataset obtained from the Multi Commodity Exchange of India and the New York Stock Exchange, we can empirically demonstrate that future execution probability is strongly correlated to past execution probability, and the execution probability also has intraday seasonality patterns. To find a suitable method to model the execution probability under this new framework, we perform several experiments to compare the performance of applying major probability distributions with non-negative support (e.g. the generalised gamma, the generalised F and the Burr distribution), as well as three major time series analysis techniques (i.e. the autoregressive moving average model, the generalise autoregressive conditional heteroskedasticity model and the autoregressive conditional duration model) to model the unconditional and conditional distributions of price fluctuations. The result indicates that the generalised F distribution is the best distribution for modelling the unconditional distribution of price fluctuations, while the autoregressive conditional duration model is the most appropriate method for modelling the conditional distribution, and, thus, the best model for modelling the execution probability.

5.1 Introduction

Most equity and derivative exchanges around the world are nowadays organised as order-driven markets where traders execute their trades by submitting either market orders or limit orders. Consequently, the decision whether to submit market orders or limit orders to execute a trade is a fundamental problem faced everyday by a trader in such markets. Although this decision can be modelled from many perspectives, the most natural approach is to view these decisions as a trade-off between the payoffs associated with limit orders and the risk of non-execution. On one hand, traders would prefer to place their orders very far from the best price since this will increase their payoff; on the other hand, the greater the distance from the best price, the greater the chance that the order will not be executed. Accordingly, traders have to find the right trade-off between these two opposite choices in order to maximise the expected profit obtained from the trade. Undoubtedly, one of the most important factors in valuing such trade-off is a model of execution probability, as the expected profit that traders will get from limit orders is an increasing function of the execution probability.

Although the execution probability is one of the most important components for valuing such a trade-off, in our opinion, the research into how to model this probability is still very limited and requires further investigation. Consequently, this chapter proposes a new method for modelling the probability that a limit order, at a given price level, will be executed within a specified trading horizon from price fluctuation during the interested period. The main advantage of this approach over traditional approaches is that it requires less data to fit the model, especially when we want to model the execution probability at several price levels simultaneously. Additionally, it also provides a natural way to apply traditional time series analysis techniques to model the execution probability. Last but not least, it also enables us to empirically illustrate that future execution probability is strongly correlated to past execution probability and the execution probability also exhibits intraday seasonality.

To achieve this, the chapter starts by firstly analysing the relationship between price fluctuation and the probability of execution in Section 5.2. Section 5.3 then further investigates the empirical properties of price fluctuations using the historical dataset collected from the Multi Commodity Exchange of India and the New York Stock Exchange. In Section 5.4, we analyse the unconditional distribution price fluctuations both theoretically and empirically. In particular, we derive the unconditional model of price fluctuation when the asset price is assumed to follow the arithmetic Brownian motion, and fit the historical dataset to the derived distribution as well as several well known probability distributions with non-negative support including the exponential distribution, the Weibull distribution, the Gamma distribution, the generalised Gamma distribution, the Burr distribution, and the generalised F distribution. Section 5.5 then compares the performance of applying three major time series analysis techniques (i.e. the autoregressive moving average model, the generalised autoregressive conditional heteroskedasticity and the autoregressive conditional duration model) to model the price fluctuation dataset. Since the result indicates that the most appropriate time series analysis technique for modelling price fluctuation is the autoregressive conditional duration (ACD) model, Section 5.6 further investigates the performance of several extensions of the ACD model with the aim of finding the most appropriate model for modelling the conditional distribution of price fluctuation. Finally, a summary of the result obtained in this chapter is given in Section 5.7.

5.2 Price fluctuation and execution probability

This section establishes the relationship between price fluctuation and execution probability. Consider a situation where we want to estimate the probability that a limit buy order submitted at a distance of Δ

ticks away from the best ask price will be executed within a specified time T, denoted by $P_E(\Delta;T)$[1]. Let the best ask price at time $t = 0$ be p_0. This probability can be estimated from the first-passage time that the asset price reaches or crosses this limit order price, which equals $p_0 - \Delta$. In particular, this probability can be estimated from:

$$\begin{aligned} P_E(\Delta;T) &= \mathbf{Pr}\big\{ \inf\{t;\ p(t) \le p_0 - \Delta\} \le T \big\}, \\ &= \mathbf{Pr}\big\{ \inf\{p(t);\ 0 \le t \le T\} \le p_0 - \Delta \big\}, \\ &= \mathbf{Pr}\big\{ \sup\{p_0 - p(t);\ 0 \le t \le T\} \ge \Delta \big\}, \end{aligned} \tag{5.1}$$

where $p(t)$ is the asset price at time t. Consequently, we define the price fluctuation during time T, denoted by M_T, as the difference between the initial price level and the lowest price level reached during time T, or equivalently

$$M_T = \sup\{p_0 - p(t);\ 0 \le t \le T\}. \tag{5.2}$$

Inserting Equation (5.2) into (5.1), we have

$$P_E(\Delta;T) = \mathbf{Pr}\big\{ M_t \ge \Delta \big\} = \int_{\Delta}^{\infty} f_{M_T}(p)dp = 1 - F_{M_T}(\Delta), \tag{5.3}$$

where $f_{M_T}(.)$ and $F_{M_T}(.)$ are the probability density function (p.d.f.) and the cumulative distribution function (c.d.f.) of M_T, respectively. Rearranging Equation (5.3), we obtain

$$F_{M_T}(\Delta) = 1 - P_E(\Delta;T), \tag{5.4}$$

and

$$f_{M_T}(\Delta) = \frac{d}{d\Delta} F_{M_T}(\Delta;T) = -\frac{d}{d\Delta} P_E(\Delta;T). \tag{5.5}$$

Consequently, Equation (5.3), (5.4) and (5.5) describe the relationship between the execution probability and the distribution of price fluctuation. It also illustrates that any one of these three functions (i.e. the execution probability, the p.d.f. and the c.d.f. of price fluctuation) uniquely determines the other two and thus if one function is known the rest can be derived mathematically. While the focus of execution probability discussed in previous chapters is to model the execution probability either directly or from the execution time distribution, this chapter will focus on modelling the distribution of price fluctuation and utilise these relations to obtain the execution probability.

5.3 Statistical properties of price fluctuation

This section analyses the statistical properties of price fluctuation using the historical dataset obtained from the Multi Commodity Exchange of India (MCX) and the New York Stock Exchange (NYSE). Section 5.3.1 starts by giving a detailed description of the dataset utilised in this study. The statistical

[1]The execution probability of a limit sell order can be estimated in the same manner; hence, this section will only focus on the execution probability of a limit buy order

properties of price fluctuation are then empirically analysed in Section 5.3.2, while the dependency between price fluctuation, return and volatility are empirically analysed in Section 5.3.3.

5.3.1 Price fluctuation databases and data preparation

MCX trading

Commodity futures are one of the most actively traded futures contracts, and the Multi Commodity Exchange of India (MCX) is one of the major exchanges for such contracts. During 2009, MCX ranked first in silver, second in gold, copper and natural gas, and third in crude oil in terms of number of futures contracts traded in the world, as per data compiled from the exchange's website. The commodity futures trading at MCX is based on an electronic screen-based trading system. It is a continuous auction system with automatic electronic order matching where traders can submit either limit orders or market orders to execute their trades. The regular trading hours at MCX generally start at 10:00 and end at 23:30 from Monday to Friday and 14:00 on Saturday. The ending time is extended to 23:55 during day light savings period, which is typically between November and March of the following year.

The data sets utilised in this study contain time stamped records of any change in quantities of limit orders at the best bid and the best ask price as well as all transactions happening during trading hours, which were manually recorded from the Reuters 3000 Xtra platform during 11/08/2008 to 03/03/2009. Accordingly, this database allows us to compute price fluctuations directly from the definition in Equation (5.2). Particularly, let $a(t)$ be the best ask at time t, $b(t)$ be the best bid at time t and $p(t)$ be the last transactional price at time t. Bid price fluctuations, denoted by M_T^B, and ask price fluctuations, denoted by M_T^A, during time period T compute at time t_0 can be computed from

$$M_T^B = a(t_0) - \min\{p(t);\ t_0 \leq t \leq t_0 + T\},$$

and

$$M_T^A = \max\{p(t);\ t_0 \leq t \leq t_0 + T\} - b(t_0).$$

Consequently, the price fluctuations data sets utilised in this study are generated from this database by applying the above equations at time $t_0 = \{$10:20, 10:20+T, 10:20+2$T, \ldots,$ 23:20$-T\}$ when $T = 5, 10$ and 30 minutes, and the resulting dataset is, thus, a time series of price fluctuations computed at three different time frames.

NYSE trading

Trading at the New York Stock Exchange (NYSE) is based on the so-called hybrid system, i.e. the trading mechanism combines a market maker system with an order book system. For each stock, one market maker (specialist) has to manage the trading and quote process and has to guarantee the provision of liquidity, when necessary, by taking the other side of the market. Regular trading at NYSE starts at 9:30 and ends at 16:00.

We utilise a historical data set provided by Dukascopy, which contains information about transactional prices (i.e. opening price, closing price, highest price and lowest price) and trading volumes sampling at every ten minutes from January 2006 to January 2010. Since this data set consists only of the transactional price and contain no information about the order book, no best bid or best ask prices are available. Consequently, the exact price fluctuation cannot be directly observed and need to be estimated by using the transactional price at the beginning of the period instead of the best bid and the best ask price. Specifically, let $p(t)$ be the transactional price at time t. The bid price fluctuation and the ask price fluctuation during time period T compute at time t_0 can be estimated from

$$M_T^B = p(t_0) - \min\{p(t);\ t_0 \leq t \leq t_0 + T\},$$

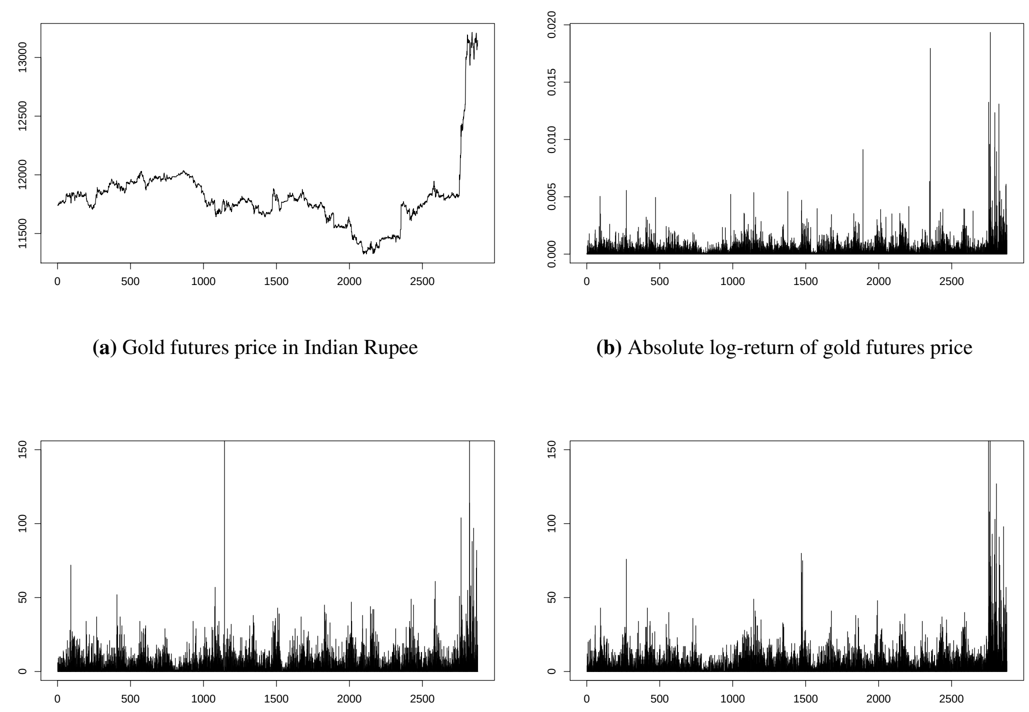
and

$$M_T^A = \max\{p(t);\ t_0 \leq t \leq t_0 + T\} - p(t_0).$$

Similar to the previous section, the price fluctuations data sets are then generated by applying the above equations at time $t_0 = \{$9:30, 9:30+T, 9:30+2$T, \ldots,$ 16:00$-T\}$ when $T = 10, 30$ and 60 minutes.

Example of price fluctuation series

Before discussing the statistical properties of price fluctuation time series in the next section, this section illustrates an example of buy and sell price fluctuations computed from the database discussed in the previous section.



(a) Gold futures price in Indian Rupee

(b) Absolute log-return of gold futures price

(c) Buy price fluctutations

(d) Sell price fluctutations

Figure 5.1: Gold futures prices (a) from 25 August to 27 September 2008 together with the corresponding absolute log-return (b), buy price fluctuations (c), and sell price fluctuations (d). The x-axis denotes time in five-minute units.

Figure 5.1 shows nearly 3000 consecutive five-minute gold futures prices, covering a period of one month from 25 August to 27 September 2008, together with the corresponding return and price fluctuation series. During this period, the gold futures price begins at about 11,740 Rs, raises to just under 12,030 Rs during the following six trading days, and then decreases to 11,320 in the next eight trading days. The futures price then rises steadily to just below 12,000 Rs before rising sharply to 13,150 Rs in one trading day. This price series clearly illustrates that there are large changes in the volatility, since the gold futures price changes very quickly at the end of the series but changes much slower at the beginning of the series. Additionally, the time series of absolute log-return displayed in Figure 5.1b also exhibits a characteristic known as volatility clustering, which indicates that large price changes tend to be followed by large changes and small changes tend to be followed by small changes. This implies that future absolute returns are not independent of their history, and the plot of their autocorrelation function will generally display a positive and significant autocorrelation at several lags ranging from a few minutes to several weeks. This characteristic generally occurs not only on absolute return series but also on series of other quantities that can be utilised as a proxy for volatility such as square returns and the ranges between the highest and the lowest price.

Since the summation of bid price fluctuations and ask price fluctuations is approximately equal to the ranges between the highest and the lowest price, we expect that the properties of these price fluctuations should somewhat resemble the properties of the volatility mentioned above. In fact, the plot of bid price fluctuations and ask fluctuations, displayed in Figure 5.1c and Figure 5.1d, provide a good evidence for supporting this expectation as they clearly indicate that a large fluctuation is generally followed by large fluctuations and a small fluctuation is generally followed by small fluctuations. However, we do not expect their properties to be completely similar, as the volatility is not related to the direction of price changes but this direction seems to be relevant for bid and ask price fluctuations. This is because the plot of bid and ask price fluctuations suggests that ask price fluctuations tend to be larger than bid price fluctuations when the price increases as can be observed during the sharp increase at the end of the period, while bid price fluctuations tend to be larger than ask price fluctuations when the price decreases as can be seen during the down turn period when the gold futures price decreases from 12,030 Rs to 11,320 Rs. In the next section, we will test these features quantitatively.

5.3.2 Statistical properties of price fluctuation

In this section we present some of the stylised facts about financial price fluctuation time series data. We focus here on the gold, silver and natural gas futures contracts traded at the MCX and the GE, IBM and Microsoft stock traded at NYSE.

Table 5.1 shows descriptive statistics of bid and ask price fluctuations for MCX trading and NYSE trading. The statistics suggest that both bid and ask price fluctuations are generally higher when the period is larger, indicating that the longer the order stays in the order book, the higher the probability that it will be executed. The statistics also suggest that the 5%, 25% and 50% quantiles of bid and ask price fluctuations are roughly equal to each other in all situations, while 75% and 95% quantiles are somewhat different in some situations. The p-value from the bootstrap Kolmogorov-Smirnov test

Table 5.1: Descriptive statistics of price fluctuations (number of observations, mean, standard deviation, maximum, quantiles, Ljung-Box $(\chi^2(20))$ statistic), Lo's rescale R/S statistic and the p-value from the bootstrap Kolmogorov-Smirnov test based on trading on MCX and NYSE.

	Obs	Mean	S.D.	MAX	0.05q	0.25q	0.50q	0.75q	0.95q	LB(20)	R/S	KS
Gold futures price fluctuations, MCX (11/08/2008 to 03/03/2009)												
BID,T=5	20097	10.82	13.43	391.00	0.00	3.00	7.00	14.00	34.00	9124	3.767	**0.036**
ASK,T=5	20097	10.95	13.03	474.00	0.00	3.00	7.00	15.00	34.00	10257	3.456	-
BID,T=10	10123	15.53	18.84	391.00	0.00	4.00	10.00	20.00	49.00	3385	3.137	0.257
ASK,T=10	10123	15.64	18.35	474.00	0.00	4.00	10.00	21.00	49.00	3440	2.712	-
Silver futures price fluctuations, MCX (11/08/2008 to 03/03/2009)												
BID,T=5	20083	23.98	29.51	571.00	0.00	5.00	15.00	32.00	77.00	8918	3.952	0.680
ASK,T=5	20083	23.28	26.88	465.00	0.00	5.00	15.00	31.00	74.00	8680	3.579	-
BID,T=10	10116	33.65	39.14	571.00	0.00	9.00	22.00	45.00	101.00	3444	3.307	0.433
ASK,T=10	10116	32.51	35.74	465.00	0.00	9.00	22.00	44.00	101.00	3524	2.820	-
Natural gas futures price fluctuations, MCX (11/08/2008 to 03/03/2009)												
BID,T=5	18565	0.57	0.96	22.70	0.00	0.00	0.20	0.70	2.20	5931	3.154	**0.026**
ASK,T=5	18565	0.52	0.84	16.60	0.00	0.00	0.20	0.70	2.00	6023	2.902	-
BID,T=10	9637	0.80	1.19	23.10	0.00	0.10	0.40	1.00	2.90	2732	2.363	**0.016**
ASK,T=10	9637	0.72	1.00	12.70	0.00	0.10	0.40	1.00	2.60	3501	2.247	-
GE price fluctuations, NYSE (01/01/2006 to 01/01/2010)												
BID,T=10	41431	0.05	0.07	3.70	0.00	0.01	0.03	0.06	0.14	38503	11.556	**0.041**
ASK,T=10	41431	0.04	0.06	2.63	0.00	0.01	0.03	0.06	0.14	50495	10.630	-
BID,T=30	14084	0.07	0.11	3.71	0.00	0.01	0.04	0.09	0.24	16085	8.295	0.150
ASK,T=30	14084	0.07	0.11	2.89	0.00	0.01	0.04	0.09	0.25	17259	7.487	-
IBM price fluctuations, NYSE (01/01/2006 to 01/01/2010)												
BID,T=10	42693	0.16	0.22	5.49	0.00	0.04	0.10	0.20	0.51	34888	10.760	0.138
ASK,T=10	42693	0.16	0.22	7.87	0.01	0.04	0.10	0.20	0.50	29764	11.181	-
BID,T=30	14237	0.27	0.37	7.27	0.01	0.06	0.17	0.35	0.90	11206	7.848	0.301
ASK,T=30	14237	0.27	0.36	7.87	0.01	0.06	0.17	0.34	0.87	7993	7.874	-
Microsoft price fluctuations, NYSE (01/01/2006 to 01/01/2010)												
BID,T=10	42655	0.05	0.06	3.03	0.00	0.02	0.04	0.06	0.14	14393	10.391	0.378
ASK,T=10	42655	0.05	0.06	3.86	0.00	0.02	0.04	0.06	0.14	12686	10.166	-
BID,T=30	14226	0.08	0.10	3.04	0.01	0.03	0.06	0.11	0.25	4649	7.539	0.987
ASK,T=30	14226	0.08	0.10	3.86	0.01	0.03	0.06	0.10	0.26	3976	7.248	-

[89, 78] can reject the hypothesis that the observed bid and ask price fluctuations are drawn from the same distribution with 95% confidence only for natural gas futures, gold futures (when T=5) and GE (when T=10). Consequently, this provides evidence that the distributions of bid and ask price fluctuations are not necessary symmetrical and we may need to model them separately.

Focusing on distributional aspects of the price fluctuation, we observe that the standard deviation of price fluctuations series is greater than the mean estimated from the same series in all datasets. This indicates an overdispersion of the observed distribution relative to an exponential distribution, whose standard deviation must always be lower than its mean. Consequently, this evidence suggests that the exponential distribution might not be a good candidate for modelling the price fluctuation distribution. This overdispersion effect is also reflected in the distributional shape of the price fluctuations. Figure 5.2 and 5.3 show kernel density plots [46] of the price fluctuation for MCX and NYSE trading respectively. The density depicts a strong right-skewed shape, indicating a high occurrence of relatively low price fluctuations and a strongly declining proportion of higher price fluctuations. The density also illustrates that the modes of these distributions are generally higher than zero in most of the situations, which is yet more evidence against the exponential distribution.

The Ljung-Box (LB) statistics [59] reported in Table 5.1 formally tests the null hypothesis that the first 20 autocorrelations are zero and are $\chi^2(20)$ distributed with a critical value of 31.41 at the 5% significance level. Based on these statistics, the null hypothesis of no autocorrelation is easily rejected

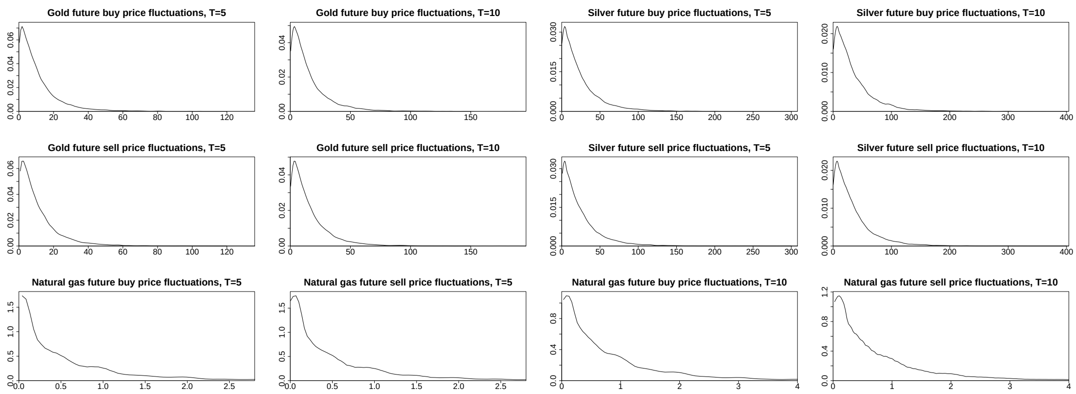


Figure 5.2: Kernel density plots (Epanechnikov kernel with optimal bandwidth) of price fluctuations of gold, silver and natural gas futures contracts based on MCX trading.

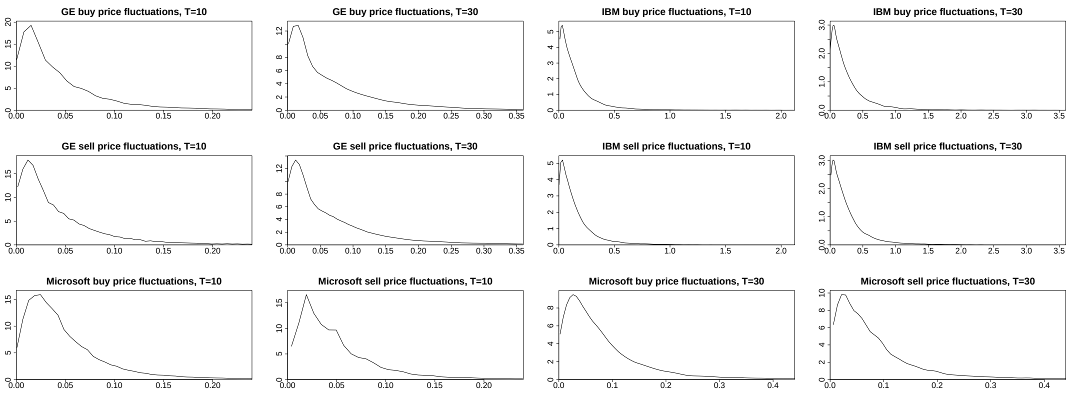


Figure 5.3: Kernel density plots (Epanechnikov kernel with optimal bandwidth) of price fluctuations of GE, IBM and Microsoft based on NYSE trading.

for all price fluctuations studied in the table.

Figure 5.4 and 5.5 show the autocorrelation function (ACF) of bid and ask price fluctuations for MCX and NYSE trading respectively. In general, we observe that the patterns of the ACF in these two markets are quite different, while the ACF for instruments traded in the same market are less different. Price fluctuations on the MCX (see Figure 5.4) have relatively low autocorrelations compared to those on the NYSE (see Figure 5.5). This indicates that price fluctuations on the NYSE generally have a stronger clustering of price fluctuations than those on the MCX. In all cases, the autocorrelation functions have clear seasonality pattern, as the peaks in the autocorrelation functions are associated with time period that dates back to previous trading days. Additionally, these price fluctuation processes are very persistent since the autocorrelation functions decay with a slow, hyperbolic rate, which is typical for long memory processes. Lo's rescaled R/S statistics [60] in Table 5.1 formally test the null hypothesis that the process is short-memory, and we can reject this null hypothesis with 95% confidence when the statistics are outside the interval $[0.809, 1.862]$. Based on these statistics, the null hypothesis of short range dependency is easily rejected for all situations. Consequently, these results represent strong evidence supporting long-memory in price fluctuation processes.

Figure 5.6 and 5.7 show the intraday seasonality patterns of price fluctuations based on cubic spline

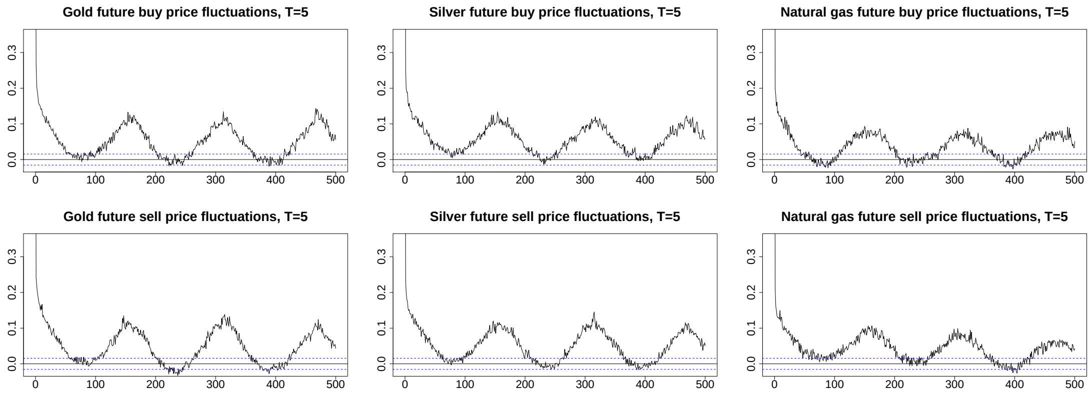


Figure 5.4: Autocorrelation function of price fluctuations for gold futures, silver futures and natural gas futures contracts based on MCX trading. Dotted line represents 99% confidence interval. The x-axis denotes the lag.

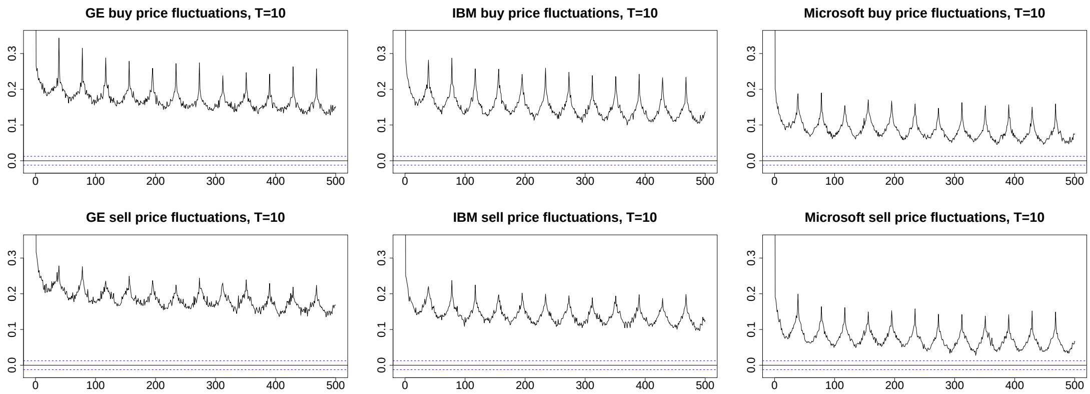


Figure 5.5: Autocorrelation function of price fluctuations for GE, IBM and Microsoft based on NYSE trading. Dotted line represents 99% confidence interval. The x-axis denotes the lag.

regressions. For MCX trading (see Figure 5.6), we find the intraday seasonality of price fluctuations in the MCX has an inverted U-shaped pattern. Specifically, the market starts with relatively low price fluctuations and increases steadily throughout the trading day with a peak around 19:00 which corresponding to the opening time of most American exchanges during the daylight saving period. The price fluctuations then decrease and finish the day with a slightly higher level than at the beginning of the day. Interestingly, the intraday seasonality pattern of the NYSE trading (see Figure 5.7) presents a completely different picture. In contrast to trading at the MCX, intraday seasonality pattern in the NYSE exhibits U-shaped pattern where high price fluctuations are observed after opening and before the closing of the market with the lowest level around lunch time period.

Summarising from these findings, we can conclude that the form of market seems to have a strong impact on the dynamics of the resulting price fluctuations, since the strength and persistence of serial dependency in price fluctuations mainly differ between the individual exchanges and less between the different assets traded in the same exchange. We also find evidence of asymmetry between bid and ask price fluctuations which suggest that we might need to model them separately. Additionally, these price fluctuation processes seem to have long range dependency with clear intraday seasonality patterns.

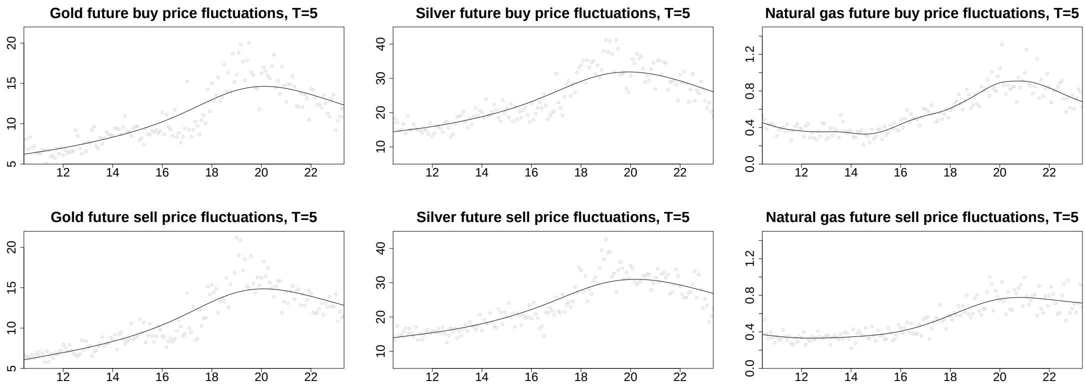


Figure 5.6: Cubic spline function of price fluctuations for gold futures, silver futures and natural gas futures based on MCX trading. The x-axis denotes the local calendar time.

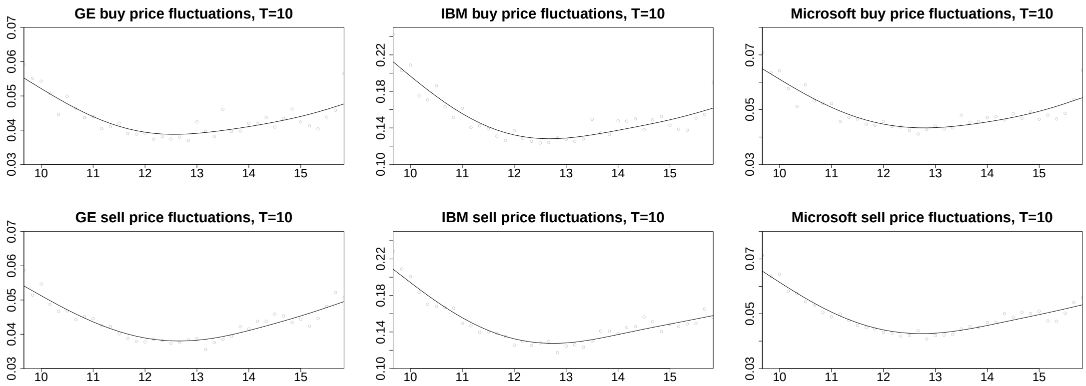


Figure 5.7: Cubic spline function of price fluctuations for GE, IBM and Microsoft based on NYSE trading. The x-axis denotes the local calendar time.

5.3.3 Dependency between price fluctuations, returns and volatilities

This section studies the dependency between bid price fluctuation, ask price fluctuation, return and volatility using the same dataset analysed in the previous section. Particularly, we will focus on the gold, silver, and natural gas futures contracts traded at the MCX and the GE, IBM and Microsoft stocks traded at the NYSE.

Table 5.2 shows the correlation between price fluctuations, return and volatility as calculated by the three correlation measures: Pearson product-moment correlation coefficient, Spearman's rank correlation coefficient and Kendall tau rank correlation coefficient. The difference between these three measures is that Pearson's coefficient detects only linear dependency between two variables while Spearman's coefficient assesses how well the relationship between two variables can be described using a monotonic function and Kendall's coefficient measures the similarity of the orderings of the data when ranked by each quantity. The return in the MCX is calculated from the change of the logarithm of the mid-price, while the return in the NYSE is calculated from the change of the logarithm of the opening price. Since the volatility cannot be directly observed from the historical price process, we will utilise the absolute log-return and the range between the highest and the lowest price as its proxy. The correlation coefficients obtained from these three methods suggest that there is a strong dependency between price fluctuations,

Table 5.2: Correlation between bid price fluctuations, ask price fluctuations, return and volatility as measured by Pearson's product-moment correlation coefficient, Kendall's tau rank correlation coefficient and Spearman's rank correlation coefficient.

		Pearson's coefficient					Spearman's coefficient					Kendall's coefficient				
		M_T^B	M_T^A	r	$\|r\|$	$h-l$	M_T^B	M_T^A	r	$\|r\|$	$h-l$	M_T^B	M_T^A	r	$\|r\|$	$h-l$
Gold futures price fluctuations, MCX (11/08/2008 to 03/03/2009)																
$T=5$	M_T^B	-	-0.13	-0.64	0.47	0.67	-	-0.32	-0.67	0.27	0.51	-	-0.23	-0.51	0.20	0.40
	M_T^A	-0.13	-	0.61	0.46	0.65	-0.32	-	0.68	0.29	0.52	-0.23	-	0.52	0.22	0.41
$T=10$	M_T^B	-	-0.16	-0.65	0.47	0.66	-	-0.36	-0.69	0.25	0.50	-	-0.25	-0.53	0.19	0.39
	M_T^A	-0.16	-	0.63	0.46	0.64	-0.36	-	0.69	0.29	0.50	-0.25	-	0.53	0.22	0.39
Silver futures price fluctuations, MCX (11/08/2008 to 03/03/2009)																
$T=5$	M_T^B	-	-0.10	-0.61	0.46	0.71	-	-0.22	-0.61	0.27	0.58	-	-0.15	-0.45	0.20	0.44
	M_T^A	-0.10	-	0.58	0.40	0.63	-0.22	-	0.62	0.26	0.56	-0.15	-	0.46	0.19	0.43
$T=10$	M_T^B	-	-0.14	-0.64	0.46	0.69	-	-0.27	-0.65	0.26	0.55	-	-0.19	-0.48	0.19	0.42
	M_T^A	-0.14	-	0.61	0.40	0.62	-0.27	-	0.64	0.26	0.53	-0.19	-	0.48	0.19	0.40
Natural gas futures price fluctuations, MCX (11/08/2008 to 03/03/2009)																
$T=5$	M_T^B	-	0.04	-0.49	0.40	0.76	-	-0.04	-0.50	0.25	0.65	-	-0.03	-0.38	0.19	0.53
	M_T^A	0.04	-	0.44	0.35	0.67	-0.04	-	0.47	0.23	0.62	-0.03	-	0.35	0.17	0.50
$T=10$	M_T^B	-	0.02	-0.52	0.42	0.77	-	-0.05	-0.51	0.28	0.66	-	-0.04	-0.38	0.21	0.52
	M_T^A	0.02	-	0.47	0.34	0.65	-0.05	-	0.49	0.23	0.61	-0.04	-	0.37	0.17	0.48
GE price fluctuations, NYSE (01/01/2006 to 01/01/2010)																
$T=10$	M_T^B	-	-0.10	-0.66	0.59	0.70	-	-0.27	-0.64	0.34	0.53	-	-0.20	-0.49	0.26	0.42
	M_T^A	-0.10	-	0.60	0.50	0.64	-0.27	-	0.64	0.32	0.53	-0.20	-	0.49	0.24	0.42
$T=30$	M_T^B	-	-0.11	-0.66	0.59	0.69	-	-0.23	-0.65	0.37	0.54	-	-0.17	-0.50	0.28	0.43
	M_T^A	-0.11	-	0.60	0.49	0.65	-0.23	-	0.63	0.33	0.53	-0.17	-	0.48	0.25	0.42
IBM price fluctuations, NYSE (01/01/2006 to 01/01/2010)																
$T=10$	M_T^B	-	-0.04	-0.62	0.55	0.69	-	-0.12	-0.60	0.39	0.60	-	-0.09	-0.46	0.29	0.47
	M_T^A	-0.04	-	0.65	0.60	0.70	-0.12	-	0.61	0.39	0.58	-0.09	-	0.46	0.29	0.46
$T=30$	M_T^B	-	-0.05	-0.63	0.55	0.69	-	-0.12	-0.61	0.38	0.61	-	-0.09	-0.46	0.29	0.48
	M_T^A	-0.05	-	0.65	0.58	0.69	-0.12	-	0.61	0.38	0.57	-0.09	-	0.46	0.28	0.45
Microsoft price fluctuations, NYSE (01/01/2006 to 01/01/2010)																
$T=10$	M_T^B	-	-0.14	-0.69	0.58	0.66	-	-0.32	-0.65	0.30	0.51	-	-0.24	-0.51	0.23	0.41
	M_T^A	-0.14	-	0.66	0.53	0.65	-0.32	-	0.66	0.28	0.50	-0.24	-	0.51	0.22	0.40
$T=30$	M_T^B	-	-0.14	-0.70	0.58	0.66	-	-0.29	-0.67	0.32	0.52	-	-0.21	-0.51	0.24	0.41
	M_T^A	-0.14	-	0.66	0.53	0.65	-0.29	-	0.66	0.31	0.51	-0.21	-	0.51	0.23	0.41

return and volatility. Specifically, we found that price fluctuation is highly correlated with return, since the correlation coefficient between price fluctuation and log-return has a high absolute value, with large negative value for bid price fluctuation and large positive value for ask price fluctuation. This suggests that price fluctuation is dependent on the direction of the price change in the sense that bid price fluctuations tend to have small (large) value while ask price fluctuation tend to have large (small) value when the price increases (decreases) which is consistent with the observation discussed in Section 5.3.1. Since bid price fluctuation is negatively correlated with return, while ask price fluctuation is positively correlated with return, one might expect bid and ask price fluctuation to be negatively correlated with each other. However, the result indicates that this correlation is generally weaker than the correlation between price fluctuation and return, especially in the case of natural gas futures where the correlation coefficient is near zero. Additionally, the result also suggests that price fluctuation is positively correlated with volatility, as the correlation coefficient between price fluctuation and absolute log-return as well as the range between the highest and the lowest price has high positive value in all cases. This indicates that price fluctuation is generally high when the volatility is high and vice versa.

Although the above result suggests that price fluctuation is highly correlated with return and volatility during the same time period, this has nothing to do with when we want to forecast next price fluctuation from past return and volatility. To investigate this, we further compute the correlation between price

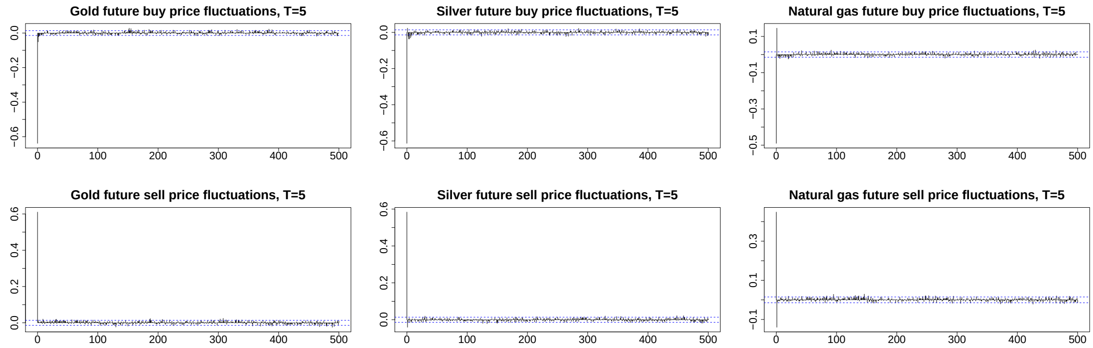


Figure 5.8: Correlation between price fluctuations and return at several time lags for gold futures, silver futures and natural gas futures contracts based on MCX trading. Dotted line represents 99% confidence interval. The x-axis denotes the lag.

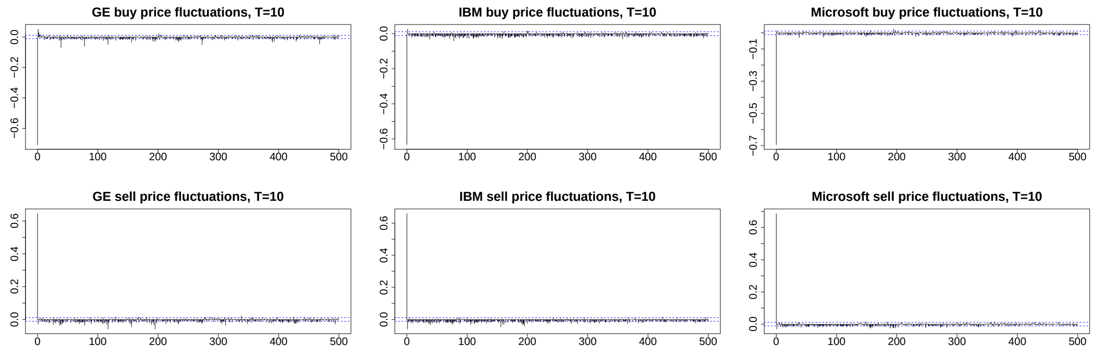


Figure 5.9: Correlation between price fluctuations and return at several time lags for GE, IBM and Microsoft based on NYSE trading. Dotted line represents 99% confidence interval. The x-axis denotes the lag.

fluctuation, return and volatility at several time lags. The correlation between price fluctuation and return displayed in Figure 5.8 and 5.9 indicates that correlation between price fluctuation and return is significant only for the first few lags and has the highest value at zero lag. In most situations, the correlation at the first lag generally has a different sign than the correlation at the zero lag. This reflects the autocorrelation property of return series which generally has negative correlation for the first few lags. However, the correlation at other lags is typically weak and might not have influence on future price fluctuation except for the case of natural gas futures contract where the first lag is larger than 0.1. The correlation between price fluctuation and volatility illustrated in Figure 5.10 and 5.11 is significant at all time lags considered, and has a clear seasonality pattern since the higher correlation is associated with time period that dates back to previous trading days. Consequently, this correlation has characteristics similar to the autocorrelation function of price fluctuations, which further supports the idea that the properties of price fluctuation are similar to the properties of volatility.

Briefly, the result reported in this section indicates that price fluctuation is heavily dependent on the direction of return during the same period, in the sense that buy price fluctuation is negatively correlated with return while sell price fluctuation is positively correlated with return. However, the correlation between price fluctuation and previous return is typically weak and might not be viable for predicting

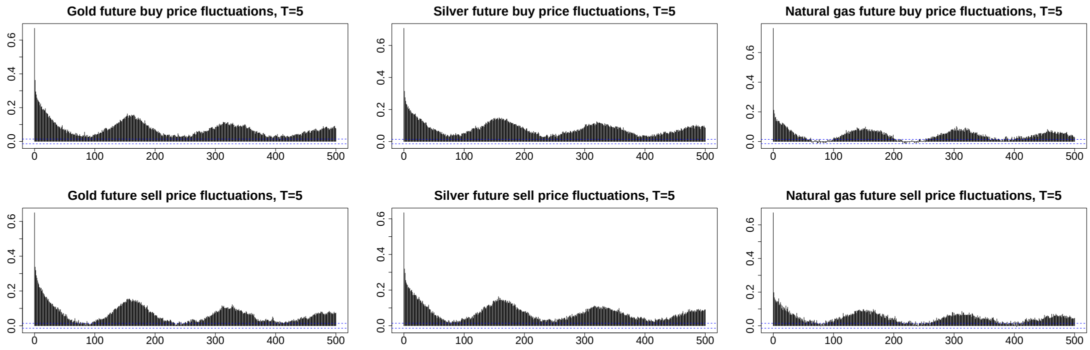


Figure 5.10: Correlation between price fluctuations and volatility at several time lags for gold futures, silver futures and natural gas futures contracts based on MCX trading. Dotted line represents 99% confidence interval. The x-axis denotes the lag.

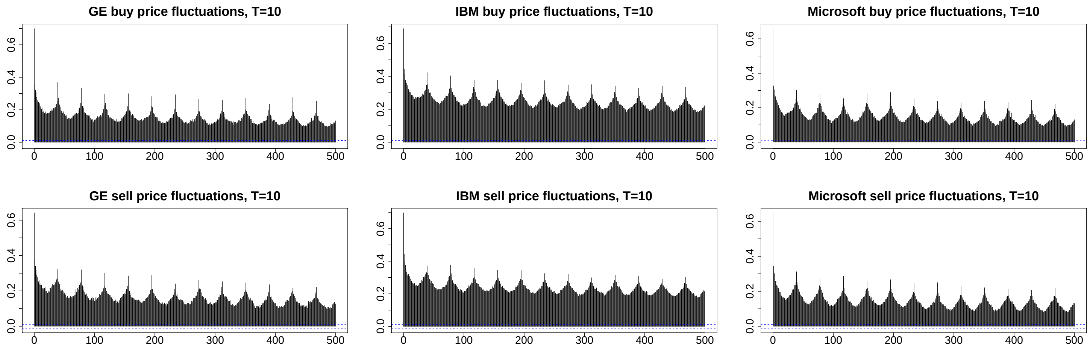


Figure 5.11: Correlation between price fluctuations and volatility at several time lags for GE, IBM and Microsoft based on NYSE trading. Dotted line represents 99% confidence interval. The x-axis denotes the lag.

future price fluctuation. Moreover, the result indicates that price fluctuation is strongly correlated to volatility, as estimated by the range between the highest and lowest price, since the correlation between price fluctuation and volatility has similar properties to the autocorrelation of price fluctuation.

5.4 Unconditional distribution of price fluctuation

Before studying the conditional model of price fluctuations, this section firstly analyses the unconditional distribution of price fluctuations in order to gain more insight into the method for estimating this distribution from the dataset, as well as identify a suitable distribution for use as a baseline when we estimate the conditional model in the next section. To achieve this, we firstly derive the unconditional distribution of price fluctuation when the asset price is assumed to follow the arithmetic Brownian motion, and fit the derived distribution to the price fluctuation dataset described in Section 5.3.1. The result suggests that this distribution is not flexible enough to model the price fluctuation dataset and other alternative models might be required. To search for an alternative model, we perform an experiment to fit the price fluctuation dataset to several continuous distributions with non-negative support including the exponential, Weibull, gamma, generalised gamma, generalised F, and Burr distribution. However, the result obtained from these distributions still does not provide satisfactory results and this is caused by the fact that there is a lot of probability mass at zero which is not supported by these distributions. To solve this problem, we propose to model this dataset using a discretised version of the above distributions, which not only allow a probability mass at zero but also directly account for the discreteness characteristic of the dataset. The result obtained by fitting this discrete version to the price fluctuation dataset indicates that these discretised distributions provide better fit than their continuous counterparts.

This section is organised as follows. Section 5.4.1 studies the unconditional distribution of price fluctuation when the asset price is assumed to follow arithmetic Brownian motion. The experiment to fit the price fluctuation dataset to several continuous distributions is discussed in Section 5.4.2. Section 5.4.3 proposes a new method to fit this dataset by using a discretized version of the above continuous distributions. In Section 5.4.4, the goodness-of-fit of the proposed model to the price fluctuation dataset is investigated, and a comparison is drawn between the continuous distribution and their discretised version. Finally, there is a brief summary and discussion in Section 5.4.5.

5.4.1 Distribution implied by the arithmetic Brownian motion

This section studies the unconditional distribution of price fluctuations when the asset price is assumed to follow the arithmetic Brownian motion with drift, which has the form:

$$dp(t) = \sigma dW_t + \mu dt,$$

where $p(t)$ is the asset price at time t, W_t is a Wiener process, σ is the constant volatility and μ is the constant growth rate (or drift). Assuming that the asset price starts at some specified value $p_0 > 0$ at $t = 0$, one can derive the probability that the asset price hits a price level $p = 0$ for the first time at time t (see [51] page 353–354 and chapter 10 of [93] for example) as

$$f(t; p_0, \mu, \sigma) = \frac{p_0}{\sqrt{2\pi\sigma^2 t^3}} \exp\left(-\frac{(p_0 + \mu t)^2}{2\sigma^2 t}\right).$$

Since the probability of price change under this assumption is not dependent on price level, this probability can be thought of as the probability that a limit buy order submitted at p_0 ticks below the current price will be executed at time t. Consequently, the probability that a limit buy order, submitted at a distance of Δ ticks away from the best price, will be executed within time T is given by the cumulative distribution function of the above probability density function which can be computed from one minus the probability of the asset price not hitting the desired level. Using Harrison ([43], page 14, equation 11), this gives us:

$$\begin{aligned} P_E(\Delta; T, \mu, \sigma) &= \int_0^T f(t; \Delta, \mu, \sigma) dt \\ &= 1 - \int_T^\infty f(t; \Delta, \mu, \sigma) dt \\ &= \Phi\left(\frac{-\Delta - \mu T}{\sigma\sqrt{t}}\right) + \exp\left(\frac{-2\mu\Delta}{\sigma^2}\right) \Phi\left(\frac{-\Delta + \mu T}{\sigma\sqrt{t}}\right), \end{aligned} \tag{5.6}$$

where $\Phi(.)$ is the cumulative distribution function of a standard normal distribution. Inserting Equation (5.6) into (5.4) and (5.5), we have

$$F_{M_T}(\Delta;\mu,\sigma) = 1 - \Phi\left(\frac{-\Delta-\mu T}{\sigma\sqrt{T}}\right) - \exp\left(\frac{-2\mu\Delta}{\sigma^2}\right)\Phi\left(\frac{-\Delta+\mu T}{\sigma\sqrt{T}}\right), \tag{5.7}$$

and

$$\begin{aligned} f_{M_T}(\Delta;\mu,\sigma) &= \frac{d}{d\Delta}F_{M_T}(\Delta), \\ &= \phi\left(\frac{-\Delta-\mu T}{\sigma\sqrt{T}}\right)\left(\frac{1}{\sigma\sqrt{T}}\right) + \exp\left(\frac{-2\mu\Delta}{\sigma^2}\right)\phi\left(\frac{-\Delta+\mu T}{\sigma\sqrt{T}}\right)\left(\frac{1}{\sigma\sqrt{T}}\right) \\ &\quad + \exp\left(\frac{-2\mu\Delta}{\sigma^2}\right)\Phi\left(\frac{-\Delta+\mu T}{\sigma\sqrt{T}}\right)\left(\frac{2\mu}{\sigma^2}\right), \end{aligned} \tag{5.8}$$

where $\phi(.)$ is the probability density function of a standard normal distribution. When $\mu = 0$, which is corresponding to the case of the arithmetic Brownian motion with no drift, the above equations reduce to

$$F_{M_T}(\Delta;\mu=0,\sigma) = 1 - \Phi\left(\frac{-\Delta}{\sigma\sqrt{t}}\right) - \Phi\left(\frac{-\Delta}{\sigma\sqrt{T}}\right) = 1 - 2\Phi\left(\frac{-\Delta}{\sigma\sqrt{T}}\right), \tag{5.9}$$

and

$$\begin{aligned} f_{M_T}(\Delta;\mu=0,\sigma) &= \phi\left(\frac{-\Delta}{\sigma\sqrt{T}}\right)\left(\frac{1}{\sigma\sqrt{T}}\right) + \phi\left(\frac{-\Delta}{\sigma\sqrt{T}}\right)\left(\frac{1}{\sigma\sqrt{T}}\right), \\ &= \frac{2}{\sigma\sqrt{T}}\phi\left(\frac{-\Delta}{\sigma\sqrt{T}}\right), \\ &= \frac{2}{\sqrt{2\pi\sigma^2 T}}\exp\left(-\frac{\Delta^2}{2\sigma^2 T}\right), \end{aligned} \tag{5.10}$$

which is basically a half-normal distribution with zero mean and variance equals to $\sigma^2 T$. Consequently, the expectation and variance of price fluctuation, in this case, is given by $\sqrt{2\sigma^2 T/\pi}$ and $\sigma^2 T\,(1-2/\pi)$, respectively.

To gain more insight into the properties of this distribution, Figure 5.12 plots the probability density function and the implied execution probability of this distribution in several parameter settings. The figures in the first column illustrate the distribution when we vary the drift parameter. The result indicates that more probability mass will move towards zero when the drift parameter increases, while it will move away from zero when the drift parameter decreases. Consequently, this suggests that the probability of executing at lower price levels will be low when the drift parameter is high, while this probability will be high when the drift parameter is low. The figuries in the second column convey a different result for the volatility parameter, since it indicates that there will be more probability mass near zero when the volatility parameter is decreasing. This suggests that the probability of executing at lower price levels will be lower when the volatility parameter is lower. Finally, the figures in the last column illustrate that the longer the time horizon, the higher chance that the order will be executed.

Although we can estimate the parameters of this model by firstly estimating the parameters μ and σ of the arithmetic Brownian motion from the asset price dynamics, and then using the estimated param-

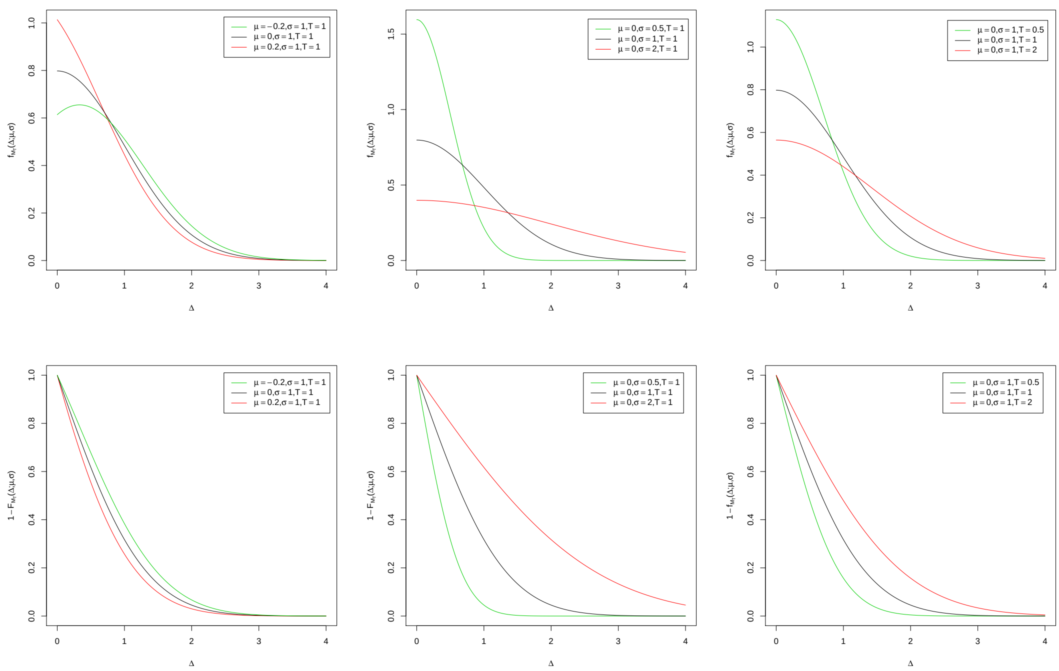


Figure 5.12: The probability density function of price fluctuation (the top row) and the implied execution probability (the bottom row) when the asset price dynamic is assumed to follow the arithmetic Brownian motion at several parameter settings.

eters to estimate the distribution of the price fluctuation, it might be better to estimate these parameters directly from historical price fluctuation dataset. Additionally, the dependency on the time period T of the distribution in Equation (5.8) is somewhat unpleasant. We can remove this variable from the equation by reparameterisation with parameters ν and ρ where

$$\nu = \frac{\mu T}{\sigma\sqrt{T}} \quad \text{and} \quad \rho = \frac{1}{\sigma\sqrt{T}},$$

so that the distribution in Equation (5.8) reduces to

$$f_{M_T}(\Delta; \nu, \rho) = \phi(-\rho\Delta - \nu)\rho + \exp(-2\rho\nu)\phi(-\rho\Delta + \nu)\rho + \exp(-2\rho\nu)\Phi(-\rho\Delta + \nu)(2\rho\nu). \quad (5.11)$$

Accordingly, the maximum likelihood estimator of these two parameters, denoted by $\hat{\nu}$ and $\hat{\rho}$, given a sample of price fluctuations $\mathbf{\Delta} = (\Delta_1, \ldots, \Delta_N)$ is given by

$$(\hat{\nu}, \hat{\rho}) = \underset{(\nu,\rho)}{\operatorname{argmax}} \sum_{i=1}^{N} \log f_{M_T}(\Delta_i; \nu, \rho). \quad (5.12)$$

For a special case when $\mu = 0$, or equivalently $\nu = 0$, the maximum likelihood estimator of ρ reduces to

$$\hat{\rho} = \underset{\rho}{\operatorname{argmax}} \sum_{i=1}^{N} \log\left(2\rho\phi(\rho\Delta)\right) = \underset{\rho}{\operatorname{argmax}}\; N \log \rho - \sum_{i=1}^{N} (\rho\Delta)^2/2.$$

The derivative of this likelihood function with respect to the parameter ρ is

$$\frac{d}{d\rho}\left(N\log\rho - \sum_{i=1}^{N}\frac{(\rho\Delta)^2}{2}\right) = \frac{N}{\rho} - \rho\sum_{i=1}^{N}\Delta^2,$$

which will equal to zero when $\rho = \sqrt{N/\sum_{i=1}^{N}\Delta^2}$. Consequently, the maximum likelihood estimator of ρ when $\nu = 0$ is

$$\hat{\rho} = \sqrt{\frac{N}{\sum_{i=1}^{N}\Delta^2}}. \tag{5.13}$$

To measure the goodness of fit of the above distributions to the price fluctuation dataset described in Section 5.3.1, we fit the above distributions to the buy price fluctuations dataset at three different trading horizons, and measure the goodness of fit using Pearson's χ^2 goodness-of-fit test, which is a statistical test generally used to measure the departure of the data from the reference model. This test statistic is constructed from the difference between the observed frequency and the theoretical frequency implied by the reference model. Particularly, if all possible outcomes are classified into K different categories, the χ^2 goodness of fit statistics can be computed from

$$X^2 = \sum_{k=1}^{K}\frac{(O_k - E_k)^2}{E_k}, \tag{5.14}$$

where O_k is the observed frequency of the observation belonging to the k-th category and E_k is the theoretical frequency of the k-th category. In our setting, K is set at one plus the price level that contained the first 98% of the observations, or equivalently

$$K = 1 + \min\left\{k \in \mathbb{N} \mid \Pr\{\Delta_i \leq \delta k\} < 0.98\right\},$$

where δ is the tick size of the instrument considered. Accordingly, the observed frequency and theoretical frequency for the k-th category can be computed from

$$O_k = \sum_{i=1}^{N} I\{\Delta_i = \delta k\},$$

and

$$E_k = N\left[F_{M_T}\big(\delta(k+1)\big) - F_{M_T}\big(\delta k\big)\right],$$

for $k = 0, \ldots, K-1$; and

$$O_K = \sum_{i=1}^{N} I\{\Delta_i \geq \delta K\},$$

and

$$E_K = N\big[1 - F_{M_T}(\delta K)\big],$$

for the last category. The asymptotic distribution of this test statistic X^2 is a χ^2 distribution with $K-p-1$ degree of freedom.

Table 5.3: Maximum log-likelihood estimates of the distribution implied by the arithmetic Brownian motion both with and without drift for the buy price fluctuation dataset together with the maximum log-likelihood, Pearson's χ^2 goodness of fit statistic and the associated p-value.

Model	T	LOGLIK	X^2	p-value	ν	ρ
Gold futures buy price fluctuations, MCX						
w drift	60	-6955	547.96	0.00	-0.12	0.03
w/o drift	60	-6984	627.83	0.00	-	0.03
w drift	30	-17725	1172.02	0.00	-0.11	0.05
w/o drift	30	-17852	1216.47	0.00	-	0.05
w drift	10	-31884	5358.55	0.00	-0.10	0.07
w/o drift	10	-32180	4689.79	0.00	-	0.07
Silver futures buy price fluctuations, MCX						
w drift	30	-7919	493.54	0.00	-0.08	0.02
w/o drift	30	-7976	481.34	0.00	-	0.02
w drift	10	-20837	1610.12	0.00	-0.08	0.03
w/o drift	10	-20966	1616.05	0.00	-	0.03
w drift	5	-38427	5383.03	0.00	-0.08	0.04
w/o drift	5	-38798	4680.62	0.00	-	0.04
Natural gas futures buy price fluctuations, MCX						
w drift	30	-1918	117.32	0.00	6.53	0.06
w/o drift	30	-2196	545.11	0.00	-	0.54
w drift	10	-3042	500.63	0.00	6.55	0.11
w/o drift	10	-4201	4624.11	0.00	-	0.86
w drift	5	-2645	2102.01	0.00	6.71	0.15
w/o drift	5	-5772	22545.07	0.00	-	1.12
GE buy price fluctuations, NYSE						
w drift	60	5899	173.88	0.00	6.31	0.77
w/o drift	60	5097	1104.82	0.00	-	6.20
w drift	30	13436	618.90	0.00	6.52	0.99
w/o drift	30	11991	2318.46	0.00	-	8.30
w drift	10	52759	4051.63	0.00	6.45	1.62
w/o drift	10	48863	5948.27	0.00	-	13.67
IBM buy price fluctuations, NYSE						
w drift	60	-295	898.92	0.00	6.58	0.20
w/o drift	60	-1755	8048.05	0.00	-	1.55
w drift	30	2920	912.99	0.00	6.68	0.26
w/o drift	30	130	10410.09	0.00	-	2.09
w drift	10	27502	2364.12	0.00	6.63	0.46
w/o drift	10	18501	30673.54	0.00	-	3.55
Microsoft buy price fluctuations, NYSE						
w drift	60	7068	285.87	0.00	6.45	0.67
w/o drift	60	5879	3018.46	0.00	-	5.38
w drift	30	16505	834.00	0.00	6.70	0.86
w/o drift	30	14336	3541.65	0.00	-	7.25
w drift	10	66675	5203.09	0.00	6.48	1.48
w/o drift	10	60877	10725.49	0.00	-	12.22

Table 5.3 displays the results obtained from fitting the above distributions with the buy price fluctuation dataset using a maximum log-likelihood estimator. The result indicates that the distribution implied by the arithmetic Brownian motion with drift has higher log-likelihood than the distribution implied by the arithmetic Brownian motion without drift, and the drift parameter is far away from zero in all situations. This suggests that it is more appropriate to model this dataset using the model with drift rather than the model without drift. However, this distribution is still not flexible enough to model the buy price fluctuation dataset since the p-value obtained from the Pearson χ^2 goodness of fit test are zero in all cases. Consequently, more complicated models are required if we want to model the price fluctuation dataset correctly. To achieve this, the rest of this section will focuses on finding the best candidate model for modelling this dataset by trying to fit several popular distributions with non-negative support and ranking them according to the Pearson χ^2 goodness of fit test statistic.

5.4.2 Continuous distribution for price fluctuation

In the search to find better distributions for modelling the price fluctuation dataset, this section tries to fit the price fluctuation dataset to several well known continuous distributions, which is generally utilised to model non-negative random variables. These distributions include the exponential, Weibull, gamma, generalised gamma, generalised F and Burr distribution. Since we have already discussed the exponential, Weibull, gamma and generalised gamma distribution in Section 2.4, this section will review only the generalised F and Burr distributions. We then analyse the results obtained from fitting the price fluctuation dataset to these distributions.

The generalised F distribution

The generalised F distribution, introduced by Prentice [80], is a four-parameter distribution that generalises the central F distribution with non-integer degrees of freedom $(2m_1, 2m_2)$ by adding location (μ) and scale ($\sigma > 0$) parameters. Particularly, let $W = (\ln \Delta - \mu)/\sigma$ be a logarithm of a random variable having the central F distribution with $2m_1$ and $2m_2$ degrees of freedom. The random variable Δ will have a generalised F distribution with the probability density function specified by

$$\begin{aligned} f_{GF}(\Delta) &= \frac{1}{\Delta\sigma}\frac{(m_1/m_2)^{m_1}\, e^{wm_1}}{B(m_1,m_2)\,(1+(m_1/m_2)\,e^{w})^{m_1+m_2}}, \\ &= \frac{(m_1/m_2)^{m_1}\, e^{-m_1\mu/\sigma}\Delta^{(m_1/\sigma)-1}}{\sigma B(m_1,m_2)\left(1+(m_1/m_2)\,(e^{-\mu}\Delta)^{1/\sigma}\right)^{m_1+m_2}}, \end{aligned} \tag{5.15}$$

where $B(m_1, m_2) = \Gamma(m_1)\Gamma(m_2)/\Gamma(m_1 + m_2)$ is the beta function evaluated at $m_1, m_2 > 0$. The cumulative distribution function of Δ is specified by

$$F_{GF}(\Delta) = \frac{1}{B(m_1,m_2)}B\left(\frac{m_1(e^{-\mu}\Delta)^{1/\sigma}}{m_2+m_1(e^{-\mu}\Delta)^{1/\sigma}};m_1,m_2\right) \tag{5.16}$$

where $B(x; a, b) = \int_0^x u^{a-1}(1-u)^{b-1}du$ is the incomplete beta function. The generalised F distribution is considered one of the most generalised models for modelling non-negative random variables since it includes many commonly used distributions as its special cases. Specifically, it reduces to the generalised log-logistic distribution when $m_1 = m_2 = m$ and the log-logistic distribution when $m_1 = m_2 = 1$. When $m_2 \to \infty$, this distribution reduces to the generalised gamma distribution which further reduces to the exponential distribution when $\mu_1 = 1$ and $\sigma_1 = 1$, the Weibull distribution when $\mu_1 = 1$, the gamma distribution when $\sigma = 1$, and the log-normal distribution when $m_1 \to \infty$. Additionally, it also contains the Burr type III distribution and the Burr type XII distribution when $m_2 = 1$ and $m_1 = 1$, respectively. To produce a well-behaved likelihood for the limiting case of the generalised gamma distribution, Prentice [80] proposed an alternative parameterisation by replacing m_1 and m_2 by alternative parameters

$$q = \left(\frac{1}{m_1} - \frac{1}{m_2}\right)\left(\frac{1}{m_1} + \frac{1}{m_2}\right)^{-\frac{1}{2}}, \text{ and } p = \frac{2}{m_1+m_2},$$

so that $-\infty < q < \infty$ and $p > 0$, and the original parameters can be reconstructed from

$$m_1 = 2\left(q^2 + 2p + q(q^2+2p)^{1/2}\right)^{-1}, \text{ and } m_2 = 2\left(q^2 + 2p - q(q^2+2p)^{1/2}\right)^{-1}.$$

With this new parameterisation, the generalised gamma distribution is given by the limiting case when $p = 0$ with $q = \lambda$ as the shape parameter of the generalised gamma distribution, while the limiting case when $p = q = 0$ is the log-normal distribution. The case when $q = 0$ defines the generalised log-logistic distribution with $m = 1/p$, which will further reduce to the log-logistic distribution when $p = 1$. Accordingly, the Burr type XII distribution is defined by $q = (1-p)[2/(2-p)]^{1/2}$, while the Burr type III is defined by $q = -(1-p)[2/(2-p)]^{1/2}$.

Given a sample of price fluctuations $\mathbf{\Delta} = (\Delta_1, \ldots, \Delta_N)$, the maximum likelihood estimator for $\theta = (\mu, \sigma, p, q)$ can be obtained by maximising the log-likelihood function

$$\begin{aligned}\ln L(\mathbf{\Delta};\theta) &= \sum_{i=1}^{N} \ln f_{GF}\left(\Delta_i; \mu, \sigma, p, q\right) \\ &= \sum_{i=1}^{N} -\ln\sigma - \ln B(m_1, m_2) + m_1 \ln(m_1/m_2) - \frac{m_1\mu}{\sigma} \\ &\quad + \left(\frac{m_1}{\sigma} - 1\right)\ln\Delta_i - (m_1+m_2)\ln\left(1 + \frac{m_1}{m_2}\exp\left\{\frac{\ln\Delta_i - \mu}{\sigma}\right\}\right). \quad (5.17)\end{aligned}$$

The Burr distribution

Unlike the generalised F distribution, the Burr distribution is a three-parameter function that can be derived as a gamma mixture of Weibull distributions (see Lancaster [55] for example). This distribution contains the exponential, Weibull and log-logistic distribution as a special case. The probability density function of this distribution is specified by

$$f_{Burr}(\Delta) = \frac{a}{\lambda}\left(\frac{\Delta}{\lambda}\right)^{a-1}\left[1 + \eta\left(\frac{\Delta}{\lambda}\right)^{a},\right]^{-(1+\eta^{-1})} \quad (5.18)$$

where $\lambda > 0$ is a scale parameter, while $a > 0$ and $\eta > 0$ are the shape parameters. The cumulative distribution function of the Burr distribution is specified by

$$F_{Burr}(\Delta) = 1 - \left(1 + \eta\lambda^{-a}\Delta^{a}\right)^{-1/\eta}. \quad (5.19)$$

It is easy to see that this distribution will reduce to the log-logistic distribution when $\eta = 1$, while for $\eta \to \infty$, this distribution will converge to the Weibull distribution.

Given a sample of price fluctuations $\mathbf{\Delta} = (\Delta_1, \ldots, \Delta_N)$, the maximum likelihood estimator for $\theta = (\lambda, a, \eta)$ can be obtained by maximising the log-likelihood function

$$\begin{aligned}\ln L(\mathbf{\Delta};\theta) &= \sum_{i=1}^{N} \ln f_{Burr}\left(\Delta_i; \lambda, a, \eta\right) \\ &= \ln a - a\ln\lambda + (a-1)\ln\Delta - (1+\eta^{-1})\ln(1 + \eta\lambda^{-a}\Delta^{a}). \quad (5.20)\end{aligned}$$

The results

This section analyses the results obtained from fitting the above distributions to the buy price fluctuation dataset using a maximum likelihood estimator. Since some of the above distributions may produce an infinite log-likelihood when the price fluctuation is zero (for example the Weibull distribution will produces ∞ log-likelihood when the shape parameter is less than one, and it will produce $-\infty$ log-likelihood when the shape parameter is greater than one), we will replace all observations with zero price fluctuations by a very small positive number (i.e. 1×10^{-10}) so that the log-likelihood is finite in all situations. In order to rank the model with a different number of parameters, we also report the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) together with the maximum log-likelihood. The result reported in Table 5.4 illustrates that the log-likelihood obtained from more complicated models is generally higher than the less complicated models, i.e. the log-likelihood of the generalised F distribution and the Burr distribution is typically higher than the models they encompass. Among all models considered, the generalised F distribution has the highest log-likelihood and lowest AIC and BIC value in all situations. However, its Pearson's χ^2 goodness of fit test statistic is higher than that of the exponential distribution in all situations, indicating that the exponential distribution provides a better fit to the price fluctuations dataset than the generalised F distribution in all cases. In fact, the exponential distribution, which generally has the lowest log-likelihood, seems to be the best candidate to model the price fluctuation dataset according to the Pearson's χ^2 test statistics. This indicates that, in our situations, the distribution with higher log-likelihood does not necessarily provide better fit to the data than the distribution with lower log-likelihood. Consequently, it might not be appropriate to utilise the maximum log-likelihood estimator to estimate the parameters of these models when our objective is to obtained the distribution that provides the best fit to our dataset.

To understand why the distribution with higher log-likelihood does not necessarily mean a better fit to the dataset, we plot the example of the estimated exponential, Weibull, generalised gamma and generalised F distributions together with the empirical distribution of the dataset in Figure 5.13. The figure clearly indicates that the reason why the Weibull, generalised gamma, and generalised F distribution have higher log-likelihood than the exponential distribution is not because they fit the dataset better than the exponential distribution, but mainly because they converge to the distribution that has large probability density at zero. Accordingly, this further confirms the inappropriateness of utilising the maximum log-likelihood estimator to estimate model parameters from this dataset and requires us to develop a new criteria for selecting model parameters that can prevent this from happening which will be the main topic of the next section.

5.4.3 Discrete distribution for price fluctuations

To solve the problem discussed in the previous section, this section presents a new method for fitting any continuous distribution with non-negative support to the price fluctuation dataset by maximising the likelihood of the discrete distribution implied by the distribution considered rather than maximising the likelihood of the distribution directly. The idea behind this approach is that, while the original distribution might produce undesirable results when the density at zero can be infinite, the probability

Table 5.4: Maximum log-likelihood estimates of the exponential, Weibull, gamma, generalised gamma, generalised F and Burr distribution for the buy price fluctuation dataset together with the maximum log-likelihood, Akaike information criterion, Bayesian information criterion, Pearson's χ^2 goodness of fit statistic and the associated p-value.

Model	T	LOGLIK	AIC	BIC	X^2	p-value
GE buy price fluctuations, NYSE						
Brownian Motion	60	5899 (6)	-11795 (6)	-11782 (6)	173.88 (2)	0.00
Exponential	60	5889 (7)	-11775 (7)	-11769 (7)	170.77 (1)	0.00
Weibull	60	6725 (4)	-13446 (4)	-13433 (4)	1546.57 (5)	0.00
Gamma	60	7243 (3)	-14482 (3)	-14469 (3)	1706.60 (6)	0.00
Generalised Gamma	60	7416 (2)	-14825 (2)	-14806 (2)	1475.05 (4)	0.00
Generalised F	60	7597 (1)	-15185 (1)	-15159 (1)	1254.90 (3)	0.00
Burr	60	6721 (5)	-13437 (5)	-13417 (5)	1709.20 (7)	0.00
Brownian Motion	30	13436 (6)	-26868 (6)	-26854 (6)	618.90 (2)	0.00
Exponential	30	13414 (7)	-26825 (7)	-26818 (7)	614.50 (1)	0.00
Weibull	30	14572 (4)	-29141 (4)	-29127 (4)	3094.56 (5)	0.00
Gamma	30	15435 (3)	-30867 (3)	-30853 (3)	3564.03 (7)	0.00
Generalised Gamma	30	15752 (2)	-31497 (2)	-31476 (2)	3141.56 (6)	0.00
Generalised F	30	16106 (1)	-32205 (1)	-32177 (1)	2614.01 (3)	0.00
Burr	30	14572 (5)	-29139 (5)	-29117 (5)	3094.28 (4)	0.00
Brownian Motion	10	52759 (6)	-105514 (6)	-105498 (6)	4051.63 (2)	0.00
Exponential	10	52705 (7)	-105407 (7)	-105399 (7)	4041.40 (1)	0.00
Weibull	10	56591 (4)	-113179 (4)	-113162 (4)	15061.08 (5)	0.00
Gamma	10	59810 (3)	-119617 (3)	-119600 (3)	16685.64 (7)	0.00
Generalised Gamma	10	61031 (2)	-122056 (2)	-122031 (2)	14304.91 (4)	0.00
Generalised F	10	62396 (1)	-124784 (1)	-124751 (1)	11559.40 (3)	0.00
Burr	10	56572 (5)	-113138 (5)	-113113 (5)	16296.24 (6)	0.00
IBM buy price fluctuations, NYSE						
Brownian Motion	60	-295 (7)	593 (7)	607 (7)	898.92 (2)	0.00
Exponential	60	-295 (6)	591 (6)	598 (6)	898.92 (1)	0.00
Weibull	60	1505 (4)	-3006 (4)	-2992 (4)	1379.65 (5)	0.00
Gamma	60	2141 (3)	-4277 (3)	-4264 (3)	1440.85 (7)	0.00
Generalised Gamma	60	2269 (2)	-4532 (2)	-4512 (2)	1388.07 (6)	0.00
Generalised F	60	2389 (1)	-4770 (1)	-4743 (1)	1229.28 (4)	0.00
Burr	60	1498 (5)	-2989 (5)	-2969 (5)	1197.87 (3)	0.00
Brownian Motion	30	2920 (6)	-5835 (6)	-5821 (7)	912.99 (2)	0.00
Exponential	30	2918 (7)	-5835 (7)	-5827 (6)	906.40 (1)	0.00
Weibull	30	5582 (4)	-11160 (4)	-11145 (4)	2179.52 (5)	0.00
Gamma	30	6596 (3)	-13189 (3)	-13174 (3)	2431.99 (7)	0.00
Generalised Gamma	30	6792 (2)	-13577 (2)	-13555 (2)	2368.36 (6)	0.00
Generalised F	30	7079 (1)	-14150 (1)	-14120 (1)	1994.37 (3)	0.00
Burr	30	5581 (5)	-11156 (5)	-11134 (5)	2157.03 (4)	0.00
Brownian Motion	10	27502 (6)	-55001 (6)	-54984 (6)	2364.12 (2)	0.00
Exponential	10	27482 (7)	-54962 (7)	-54953 (7)	2309.90 (1)	0.00
Weibull	10	37083 (4)	-74161 (4)	-74144 (4)	9397.49 (4)	0.00
Gamma	10	40859 (3)	-81714 (3)	-81697 (3)	10187.83 (7)	0.00
Generalised Gamma	10	41521 (2)	-83035 (2)	-83010 (2)	9865.25 (6)	0.00
Generalised F	10	42513 (1)	-85018 (1)	-84985 (1)	8247.30 (3)	0.00
Burr	10	37083 (5)	-74159 (5)	-74134 (5)	9418.79 (5)	0.00
Microsoft buy price fluctuations, NYSE						
Brownian Motion	60	7068 (6)	-14132 (7)	-14118 (7)	285.87 (2)	0.00
Exponential	60	7067 (7)	-14132 (6)	-14125 (6)	282.10 (1)	0.00
Weibull	60	8312 (4)	-16621 (4)	-16607 (4)	1989.14 (5)	0.00
Gamma	60	8988 (3)	-17972 (3)	-17958 (3)	2208.12 (7)	0.00
Generalised Gamma	60	9174 (2)	-18342 (2)	-18322 (2)	2031.94 (6)	0.00
Generalised F	60	9379 (1)	-18749 (1)	-18722 (1)	1715.17 (3)	0.00
Burr	60	8312 (5)	-16619 (5)	-16598 (5)	1974.03 (4)	0.00
Brownian Motion	30	16505 (6)	-33005 (6)	-32991 (6)	834.00 (2)	0.00
Exponential	30	16495 (7)	-32988 (7)	-32981 (7)	831.53 (1)	0.00
Weibull	30	18127 (4)	-36250 (4)	-36235 (4)	4018.71 (5)	0.00
Gamma	30	19216 (3)	-38427 (3)	-38413 (3)	4646.12 (7)	0.00
Generalised Gamma	30	19547 (2)	-39087 (2)	-39065 (2)	4271.36 (6)	0.00
Generalised F	30	20025 (1)	-40043 (1)	-40013 (1)	3501.38 (3)	0.00
Burr	30	18126 (5)	-36247 (5)	-36225 (5)	3992.24 (4)	0.00

Table 5.4 continued: Maximum log-likelihood estimates of the exponential, Weibull, gamma, generalised gamma, generalised F and Burr distribution for the buy price fluctuation dataset together with the maximum log-likelihood, Akaike information criterion, Bayesian information criterion, Pearson's χ^2 goodness of fit statistic and the associated p-value.

Model	T	LOGLIK	AIC	BIC	X^2	p-value
Microsoft buy price fluctuations, NYSE						
Brownian Motion	10	66675 (6)	-133347 (6)	-133330 (6)	5203.09 (2)	0.00
Exponential	10	66615 (7)	-133229 (7)	-133220 (7)	5181.42 (1)	0.00
Weibull	10	71113 (4)	-142223 (4)	-142206 (4)	17531.95 (5)	0.00
Gamma	10	74793 (3)	-149582 (3)	-149565 (3)	19947.84 (7)	0.00
Generalised Gamma	10	76094 (2)	-152182 (2)	-152157 (2)	17775.51 (6)	0.00
Generalised F	10	77855 (1)	-155702 (1)	-155668 (1)	14200.88 (3)	0.00
Burr	10	71113 (5)	-142221 (5)	-142195 (5)	17524.34 (4)	0.00
Gold futures buy price fluctuations, MCX						
Brownian Motion	30	-6955 (7)	13913 (7)	13924 (7)	547.96 (6)	0.00
Exponential	30	-6851 (6)	13705 (6)	13710 (6)	146.29 (1)	0.04
Weibull	30	-6687 (4)	13379 (4)	13389 (4)	477.17 (3)	0.00
Gamma	30	-6545 (3)	13095 (3)	13105 (3)	565.78 (7)	0.00
Generalised Gamma	30	-6459 (2)	12925 (2)	12941 (2)	491.70 (4)	0.00
Generalised F	30	-6427 (1)	12862 (1)	12883 (1)	436.24 (2)	0.00
Burr	30	-6688 (5)	13383 (5)	13399 (5)	520.23 (5)	0.00
Brownian Motion	10	-17725 (7)	35453 (7)	35466 (7)	1172.02 (2)	0.00
Exponential	10	-17327 (6)	34656 (6)	34662 (6)	194.98 (1)	0.00
Weibull	10	-15482 (4)	30969 (4)	30982 (4)	2849.62 (6)	0.00
Gamma	10	-14568 (3)	29140 (3)	29153 (3)	2699.72 (5)	0.00
Generalised Gamma	10	-14244 (2)	28494 (2)	28513 (2)	2190.93 (4)	0.00
Generalised F	10	-14155 (1)	28317 (1)	28343 (1)	1988.02 (3)	0.00
Burr	10	-15486 (5)	30978 (5)	30998 (5)	3094.12 (7)	0.00
Brownian Motion	5	-31884 (7)	63772 (7)	63787 (7)	5358.55 (3)	0.00
Exponential	5	-30995 (6)	61993 (6)	62000 (6)	328.15 (1)	0.00
Weibull	5	-24245 (4)	48494 (4)	48508 (4)	8954.45 (7)	0.00
Gamma	5	-21794 (3)	43593 (3)	43607 (3)	7294.32 (5)	0.00
Generalised Gamma	5	-21101 (2)	42209 (2)	42230 (2)	5876.52 (4)	0.00
Generalised F	5	-20903 (1)	41814 (1)	41843 (1)	5268.73 (2)	0.00
Burr	5	-24245 (5)	48497 (5)	48518 (5)	8877.27 (6)	0.00
Silver futures buy price fluctuations, MCX						
Brownian Motion	30	-7919 (7)	15842 (7)	15853 (7)	493.54 (2)	0.00
Exponential	30	-7854 (6)	15710 (6)	15716 (6)	342.26 (1)	0.00
Weibull	30	-7572 (4)	15149 (4)	15159 (4)	904.51 (5)	0.00
Gamma	30	-7353 (3)	14710 (3)	14721 (3)	969.76 (6)	0.00
Generalised Gamma	30	-7244 (2)	14494 (2)	14510 (2)	778.45 (4)	0.00
Generalised F	30	-7188 (1)	14384 (1)	14405 (1)	735.21 (3)	0.00
Burr	30	-7577 (5)	15161 (5)	15177 (5)	1014.17 (7)	0.00
Brownian Motion	10	-20837 (7)	41678 (7)	41690 (7)	1610.12 (2)	0.00
Exponential	10	-20477 (6)	40955 (6)	40962 (6)	550.39 (1)	0.00
Weibull	10	-17569 (4)	35141 (4)	35154 (4)	4088.54 (6)	0.00
Gamma	10	-16401 (3)	32807 (3)	32820 (3)	3435.51 (5)	0.00
Generalised Gamma	10	-16027 (2)	32061 (2)	32080 (2)	2676.24 (4)	0.00
Generalised F	10	-15913 (1)	31834 (1)	31860 (1)	2422.89 (3)	0.00
Burr	10	-17608 (5)	35223 (5)	35242 (5)	4909.12 (7)	0.00
Brownian Motion	5	-38427 (7)	76858 (7)	76872 (7)	5383.03 (2)	0.00
Exponential	5	-37453 (6)	74907 (6)	74914 (6)	1066.38 (1)	0.00
Weibull	5	-26885 (4)	53774 (4)	53788 (4)	11217.17 (7)	0.00
Gamma	5	-24004 (3)	48012 (3)	48026 (3)	8046.08 (5)	0.00
Generalised Gamma	5	-23302 (2)	46610 (2)	46632 (2)	6397.39 (4)	0.00
Generalised F	5	-23094 (1)	46195 (1)	46224 (1)	5884.20 (3)	0.00
Burr	5	-26885 (5)	53776 (5)	53798 (5)	11192.30 (6)	0.00
Natural gas futures buy price fluctuations, MCX						
Brownian Motion	30	-1918 (6)	3839 (7)	3850 (7)	117.32 (2)	0.00
Exponential	30	-1918 (7)	3837 (6)	3843 (6)	117.32 (1)	0.00
Weibull	30	-768 (4)	1539 (4)	1550 (4)	1440.50 (6)	0.00
Gamma	30	-372 (3)	747 (3)	758 (3)	1185.71 (5)	0.00
Generalised Gamma	30	-276 (2)	558 (2)	575 (2)	994.09 (4)	0.00
Generalised F	30	-236 (1)	480 (1)	502 (1)	901.80 (3)	0.00
Burr	30	-775 (5)	1555 (5)	1572 (5)	1697.40 (7)	0.00

Table 5.4 continued: Maximum log-likelihood estimates of the exponential, Weibull, gamma, generalised gamma, generalised F and Burr distribution for the buy price fluctuation dataset together with the maximum log-likelihood, Akaike information criterion, Bayesian information criterion, Pearson's χ^2 goodness of fit statistic and the associated p-value.

Model	T	LOGLIK	AIC	BIC	X^2	p-value
		Natural gas futures buy price fluctuations, MCX				
Brownian Motion	10	-3042 (6)	6088 (7)	6101 (7)	500.63 (1)	0.00
Exponential	10	-3042 (7)	6086 (6)	6092 (6)	500.63 (2)	0.00
Weibull	10	6972 (4)	-13940 (4)	-13927 (4)	8238.05 (6)	0.00
Gamma	10	8549 (3)	-17095 (3)	-17082 (3)	4949.69 (4)	0.00
Generalised Gamma	10	8822 (2)	-17638 (2)	-17618 (2)	4179.30 (3)	0.00
Generalised F	10	8906 (1)	-17805 (1)	-17779 (1)	5467.13 (5)	0.00
Burr	10	6833 (5)	-13659 (5)	-13640 (5)	8271.63 (7)	0.00
Brownian Motion	5	-2645 (6)	5294 (6)	5309 (7)	2102.01 (2)	0.00
Exponential	5	-2649 (7)	5301 (7)	5308 (6)	2035.08 (1)	0.00
Weibull	5	29503 (4)	-59001 (4)	-58987 (4)	20360.80 (6)	0.00
Gamma	5	32624 (3)	-65244 (3)	-65229 (3)	10586.32 (5)	0.00
Generalised Gamma	5	33064 (2)	-66121 (2)	-66100 (2)	9140.45 (4)	0.00
Generalised F	5	33217 (1)	-66426 (1)	-66398 (1)	8545.60 (3)	0.00
Burr	5	29503 (5)	-58100 (5)	-58978 (5)	20360.80 (7)	0.00

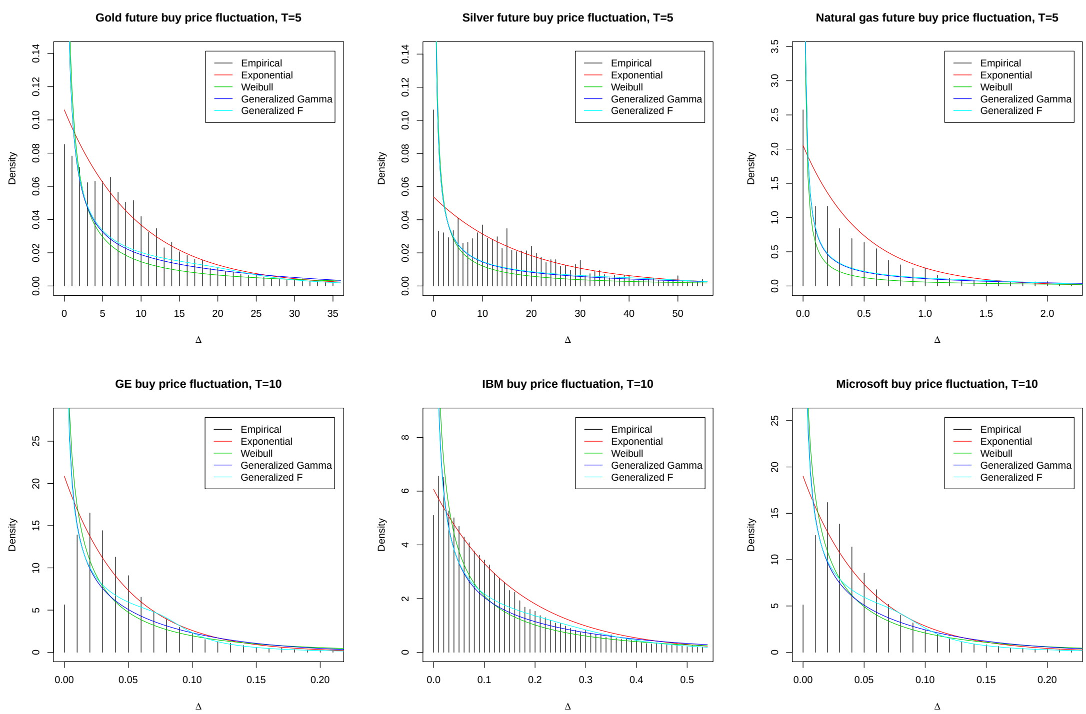


Figure 5.13: Examples of the empirical distribution of buy price fluctuation and the estimated exponential, Weibull, generalised gamma and generalised F distributions obtained from maximum log-likelihood estimator.

mass at zero of the implied discrete distribution will never be larger than one and thus will not suffer from the same problem as their continuous counterpart. Additionally, modelling the price fluctuations using discrete distributions might be more appropriate than continuous distributions, since the price fluctuation dataset is discrete in nature.

To achieve this, we construct a discrete distribution from a specified continuous distribution as follows: Let δ be a minimum tick size, the smallest increment by which the price of financial instruments can move, of the instrument that we want to model the price fluctuation distribution, and $F_C(\Delta)$ be the cumulative distribution function of the selected continuous distribution. The probability mass function of the implied discrete distribution is defined by

$$f_D(k\delta) = \begin{cases} F_C\big(\delta\big) & \text{for } k = 0 \\ F_C\big((k+1)\delta\big) - F_C(k\delta) & \text{for } k = 1, 2, 3, \dots, \end{cases}$$

or equivalently,

$$f_D(\Delta) = \begin{cases} F_C\big(\delta\big) & \text{for } \Delta = 0 \\ F_C\big(\Delta + \delta\big) - F_C\big(\Delta\big) & \text{otherwise} \end{cases}, \tag{5.21}$$

where $\Delta \in \{0, \delta, 2\delta, \dots\}$ is a price fluctuation. Note that this probability has the same form as the probability utilised to calculate the Pearson's χ^2 goodness of fit statistic in the previous section. Consequently, the cumulative distribution of the implied discrete distribution at price level $\Delta = k\delta$ can be defined by

$$\begin{aligned} F_D(\Delta) &= \sum_{i=0}^{k} f_D(i\delta) \\ &= F_C(\delta) + \sum_{i=1}^{k} \big[F_C\big((i+1)\delta\big) - F_C\big(i\delta\big)\big] \\ &= F_C(\Delta + \delta) \end{aligned} \tag{5.22}$$

Given a sample of price fluctuations $\boldsymbol{\Delta} = (\Delta_1, \dots, \Delta_N)$, the maximum likelihood estimator for the parameter $\theta = (\theta_1, \dots, \theta_M)$ of this distribution can be obtained by maximising the log-likelihood function

$$\begin{aligned} \ln L(\boldsymbol{\Delta}; \theta) &= \sum_{i=1}^{N} \ln f_D(\Delta_i; \theta) \\ &= \sum_{i=1}^{N} \ln \left(F_C\left(\Delta_i + \delta; \theta\right) - F_C(\Delta_i; \theta)\right). \end{aligned} \tag{5.23}$$

The derivative of this log-likelihood function with respect to the parameter θ_j can be computed from

$$\frac{\partial}{\partial \theta_j} \ln L(\boldsymbol{\Delta}; \theta) = \sum_{i=1}^{N} \left[\frac{1}{f_D(\Delta_i)} \left(\frac{\partial}{\partial \theta_j} F_C(\Delta_i + \delta; \theta) - \frac{\partial}{\partial \theta_j} F_C(\Delta_i; \theta) \right) \right].$$

5.4.4 Experiment results

To compare the performance of the proposed method for estimating the model parameters from the price fluctuation dataset and the traditional maximum log-likelihood estimator, this section analyses the result obtained from fitting the above distributions to the buy price fluctuation dataset using the proposed methods and compares it with the one obtained in Section 5.4.2. The result reported in Table 5.5 indicates that the distributions estimated from the proposed method have smaller Pearson's χ^2 goodness of fit test statistic than the one estimated in the previous section in most of the cases except when we fit the distribution derived from the arithmetic Brownian motion to Microsoft's ten minute buy price fluctuation dataset. The improvement obtained for the exponential distribution and the distribution derived from the arithmetic Brownian motion is generally lower than the gain obtained from other distributions. This is because these two distributions do not suffer from the problem discussed in the previous section. Among all distributions considered, the generalised F distribution seem to be the best performing distribution since it has the highest log-likelihood and the lowest AIC, BIC and χ^2 test statistic in most situations. The second best performing distribution is generally the Burr distribution, while third place belongs to the generalised gamma distribution. The result from Pearson's χ^2 goodness of fit test indicates that the test cannot reject the hypothesis that the empirical and the estimated distribution are similar only in two situations (i.e. Microsoft's 60 minute buy price fluctuation and Gold futures 30 minute buy price fluctuation). To see how bad our estimated distributions fit the dataset in the situation when the Pearson's χ^2 test rejected this hypothesis, we plot the estimated generalised F distribution and the empirical distribution of this dataset in Figure 5.14. The result suggests that whilst the goodness of fit test rejects these distributions, the plot indicates that the generalised F-distribution provides a reasonable estimation of the empirical distribution with a large error only at some price levels.

5.4.5 Summary

In this section, we studied several methods for modelling the unconditional distribution of price fluctuations. In particular, we derived the unconditional distribution of price fluctuation when the asset price is assumed to follow the arithmetic Brownian motion. Additionally, we also fitted several distributions with non-negative support including the exponential, Weibull, gamma, generalised gamma, generalised F and Burr distribution to the buy price fluctuation dataset using maximum likelihood estimator. The result indicated that maximum likelihood estimator is not a good method for estimating model parameters from this dataset since the estimated distribution converge to the distribution that has large probability density at zero rather than the distribution that provide a good fit to the dataset. To solve the problem, we proposed to estimate model parameters by maximising the likelihood of the discrete distribution implied by the considered distribution rather than maximising the likelihood of the distribution directly. The experiment results indicated that the distribution estimated by the proposed method does not suffer from this problem and is able to estimate the empirical distribution reasonably well. Among all considered models the generalised F distribution was the best performing distribution while the Burr distribution and the generalised gamma distribution are the second and third best models respectively.

Table 5.5: Maximum log-likelihood estimates of the implied discrete distributions of the exponential, Weibull, gamma, generalised gamma, generalised F and Burr distribution for the buy price fluctuation dataset together with the maximum log-likelihood, Akaike information criterion, Bayesian information criterion, Pearson's χ^2 goodness of fit statistic and the associated p-value together with the test statistics obtained from traditional log-likelihood estimator and the improvement gained.

Model	T	LOGLIK	AIC	BIC	X_C^2	X_D^2	p-value	$X_C^2 - X_D^2$
				GE buy price fluctuations, NYSE				
Brownian Motion	60	-15696 (5)	31396 (6)	31409 (6)	173.88 (2)	143.33 (5)	0.00	30.55
Exponential	60	-15697 (6)	31396 (5)	31402 (5)	170.77 (1)	143.4 (6)	0.00	27.37
Weibull	60	-15692 (4)	31388 (4)	31400 (4)	1546.57 (5)	95.57 (4)	0.00	1451
Gamma	60	-15703 (7)	31409 (7)	31422 (7)	1706.6 (6)	147.3 (7)	0.00	1559.3
Generalised Gamma	60	-15615 (3)	31236 (3)	31255 (3)	1475.05 (4)	49.24 (1)	0.01	1425.81
Generalised F	60	-15608 (1)	31224 (2)	31250 (2)	1254.9 (3)	49.31 (2)	0.01	1205.59
Burr	60	-15609 (2)	31224 (1)	31243 (1)	1709.2 (7)	52.6 (3)	0.00	1656.59
Brownian Motion	30	-26815 (7)	53633 (7)	53648 (7)	618.9 (2)	563.38 (7)	0.00	55.52
Exponential	30	-26815 (6)	53631 (6)	53638 (6)	614.5 (1)	563.27 (6)	0.00	51.23
Weibull	30	-26711 (5)	53426 (5)	53440 (5)	3094.56 (5)	296.06 (5)	0.00	2798.5
Gamma	30	-26628 (4)	53260 (4)	53274 (4)	3564.03 (7)	169.17 (4)	0.00	3394.86
Generalised Gamma	30	-26453 (3)	52912 (3)	52933 (3)	3141.56 (6)	35.83 (2)	0.02	3105.73
Generalised F	30	-26444 (1)	52895 (1)	52923 (1)	2614.01 (3)	33.5 (1)	0.02	2580.5
Burr	30	-26452 (2)	52911 (2)	52932 (2)	3094.28 (4)	53.45 (3)	0.00	3040.83
Brownian Motion	10	-68913 (7)	137830 (7)	137847 (7)	4051.63 (2)	4000.59 (6)	0.00	51.05
Exponential	10	-68905 (6)	137812 (6)	137820 (6)	4041.4 (1)	4004.12 (7)	0.00	37.27
Weibull	10	-67940 (5)	135884 (5)	135900 (5)	15061.08 (5)	1605.93 (5)	0.00	13455.15
Gamma	10	-67338 (4)	134680 (4)	134697 (4)	16685.64 (7)	707.63 (4)	0.00	15978.01
Generalised Gamma	10	-66550 (3)	133107 (3)	133131 (3)	14304.91 (4)	98.59 (3)	0.00	14206.32
Generalised F	10	-66477 (1)	132961 (1)	132994 (1)	11559.4 (3)	19.38 (1)	0.04	11540.02
Burr	10	-66487 (2)	132981 (2)	133005 (2)	16296.24 (6)	38.6 (2)	0.00	16257.64
				IBM buy price fluctuations, NYSE				
Brownian Motion	60	-28686 (7)	57377 (7)	57390 (7)	898.92 (2)	501.29 (6)	0.00	397.63
Exponential	60	-28686 (6)	57374 (6)	57381 (6)	898.92 (1)	501.32 (7)	0.00	397.6
Weibull	60	-28485 (4)	56975 (4)	56988 (4)	1379.65 (5)	195.08 (3)	0.00	1184.57
Gamma	60	-28516 (5)	57036 (5)	57049 (5)	1440.85 (7)	213.55 (5)	0.00	1227.3
Generalised Gamma	60	-28478 (3)	56962 (3)	56982 (3)	1388.07 (6)	195.31 (4)	0.00	1192.76
Generalised F	60	-28463 (1)	56934 (1)	56961 (1)	1229.28 (4)	173.74 (1)	0.01	1055.54
Burr	60	-28470 (2)	56946 (2)	56966 (2)	1197.87 (3)	187.65 (2)	0.00	1010.22
Brownian Motion	30	-50020 (7)	100044 (7)	100058 (7)	912.99 (2)	474.28 (7)	0.00	438.71
Exponential	30	-49852 (6)	99707 (6)	99714 (6)	906.4 (1)	429.25 (6)	0.00	477.15
Weibull	30	-49629 (4)	99262 (4)	99276 (4)	2179.52 (5)	223.47 (4)	0.00	1956.04
Gamma	30	-49700 (5)	99403 (5)	99418 (5)	2431.99 (7)	261.37 (5)	0.00	2170.62
Generalised Gamma	30	-49558 (3)	99121 (3)	99143 (3)	2368.36 (6)	180.8 (3)	0.00	2187.56
Generalised F	30	-49537 (1)	99083 (1)	99112 (2)	1994.37 (3)	167.14 (1)	0.00	1827.23
Burr	30	-49539 (2)	99084 (2)	99106 (1)	2157.03 (4)	167.55 (2)	0.00	1989.48
Brownian Motion	10	-131362 (7)	262728 (7)	262745 (7)	2364.12 (2)	874.12 (5)	0.00	1490
Exponential	10	-131273 (6)	262547 (6)	262556 (6)	2309.9 (1)	887.49 (6)	0.00	1422.42
Weibull	10	-130916 (4)	261836 (4)	261853 (4)	9397.49 (4)	803.52 (4)	0.00	8593.97
Gamma	10	-131128 (5)	262260 (5)	262277 (5)	10187.83 (7)	890.39 (7)	0.00	9297.44
Generalised Gamma	10	-130298 (3)	260603 (3)	260628 (3)	9865.25 (6)	157.84 (3)	0.00	9707.41
Generalised F	10	-130251 (1)	260509 (1)	260543 (2)	8247.3 (3)	113.44 (1)	0.00	8133.86
Burr	10	-130255 (2)	260516 (2)	260541 (1)	9418.79 (5)	122.15 (2)	0.00	9296.64
				Microsoft buy price fluctuations, NYSE				
Brownian Motion	60	-21543 (7)	43090 (7)	43103 (7)	285.87 (2)	200.65 (7)	0.00	85.23
Exponential	60	-21502 (6)	43006 (5)	43013 (5)	282.1 (1)	170.53 (6)	0.00	111.58
Weibull	60	-21502 (5)	43008 (6)	43022 (6)	1989.14 (5)	168.59 (5)	0.00	1820.55
Gamma	60	-21492 (4)	42989 (4)	43002 (4)	2208.12 (7)	127.59 (4)	0.00	2080.53
Generalised Gamma	60	-21366 (3)	42739 (3)	42759 (3)	2031.94 (6)	29.98 (3)	0.67	2001.97
Generalised F	60	-21354 (1)	42715 (2)	42742 (2)	1715.17 (3)	18.01 (1)	0.98	1697.17
Burr	60	-21354 (2)	42714 (1)	42735 (1)	1974.03 (4)	19.45 (2)	0.98	1954.58
Brownian Motion	30	-36712 (7)	73427 (7)	73442 (7)	834 (2)	769.55 (6)	0.00	64.45
Exponential	30	-36712 (6)	73425 (6)	73433 (6)	831.53 (1)	770.53 (7)	0.00	61.01
Weibull	30	-36670 (5)	73343 (5)	73358 (5)	4018.71 (5)	560.62 (5)	0.00	3458.09
Gamma	30	-36607 (4)	73217 (4)	73232 (4)	4646.12 (7)	481.88 (4)	0.00	4164.24
Generalised Gamma	30	-36202 (3)	72410 (3)	72432 (3)	4271.36 (6)	65.79 (3)	0.00	4205.57
Generalised F	30	-36172 (1)	72351 (1)	72381 (2)	3501.38 (3)	34.09 (1)	0.06	3467.29
Burr	30	-36176 (2)	72357 (2)	72379 (1)	3992.24 (4)	41.39 (2)	0.02	3950.85

Table 5.5 continued: Maximum log-likelihood estimates of the implied discrete distributions of the exponential, Weibull, gamma, generalised gamma, generalised F and Burr distribution for the buy price fluctuation dataset together with the maximum log-likelihood, Akaike information criterion, Bayesian information criterion, Pearson's χ^2 goodness of fit statistic and the associated p-value together with the test statistics obtained from traditional log-likelihood estimator and the improvement gained.

Model	T	LOGLIK	AIC	BIC	X_C^2	X_D^2	p-value	$X_C^2 - X_D^2$
Microsoft buy price fluctuations, NYSE								
Brownian Motion	10	-94229 (7)	188463 (7)	188480 (7)	5203.09 (2)	5214.52 (7)	0.00	-11.43
Exponential	10	-94179 (6)	188360 (6)	188369 (6)	5181.42 (1)	5125.65 (6)	0.00	55.77
Weibull	10	-93308 (5)	186620 (5)	186637 (5)	17531.95 (5)	2631.24 (5)	0.00	14900.71
Gamma	10	-92513 (4)	185030 (4)	185047 (4)	19947.84 (7)	1391.08 (4)	0.00	18556.76
Generalised Gamma	10	-92159 (3)	184323 (3)	184348 (3)	17775.51 (6)	1021.87 (3)	0.00	16753.63
Generalised F	10	-90979 (1)	181966 (1)	182000 (1)	14200.88 (3)	35.85 (1)	0.00	14165.04
Burr	10	-90999 (2)	182005 (2)	182030 (2)	17524.34 (4)	62.93 (2)	0.00	17461.41
Gold futures buy price fluctuations, MCX								
Brownian Motion	30	-7001 (7)	14006 (7)	14017 (7)	547.96 (6)	203.2 (7)	0.00	344.76
Exponential	30	-6885 (6)	13773 (6)	13778 (6)	146.29 (1)	108.93 (6)	0.00	37.36
Weibull	30	-6881 (5)	13765 (5)	13776 (5)	477.17 (3)	92.07 (5)	0.04	385.1
Gamma	30	-6878 (4)	13761 (4)	13772 (2)	565.78 (7)	89.28 (3)	0.06	476.5
Generalised Gamma	30	-6876 (3)	13759 (3)	13775 (4)	491.7 (4)	91.28 (4)	0.04	400.41
Generalised F	30	-6871 (1)	13751 (1)	13772 (3)	436.24 (2)	82.95 (1)	0.10	353.28
Burr	30	-6873 (2)	13752 (2)	13769 (1)	520.23 (5)	87.29 (2)	0.07	432.93
Brownian Motion	10	-17928 (7)	35860 (7)	35873 (7)	1172.02 (2)	560.51 (7)	0.00	611.51
Exponential	10	-17501 (6)	35004 (6)	35011 (6)	194.98 (1)	160.78 (6)	0.00	34.2
Weibull	10	-17497 (5)	34997 (5)	35010 (5)	2849.62 (6)	140.75 (5)	0.00	2708.87
Gamma	10	-17492 (4)	34989 (4)	35002 (4)	2699.72 (5)	132.75 (3)	0.00	2566.96
Generalised Gamma	10	-17483 (3)	34973 (3)	34992 (3)	2190.93 (4)	132.92 (4)	0.00	2058.01
Generalised F	10	-17460 (1)	34928 (1)	34954 (1)	1988.02 (3)	86.19 (1)	0.00	1901.83
Burr	10	-17470 (2)	34946 (2)	34966 (2)	3094.12 (7)	114.03 (2)	0.00	2980.09
Brownian Motion	5	-32346 (7)	64695 (7)	64710 (7)	5358.55 (3)	937.03 (7)	0.00	4421.52
Exponential	5	-31485 (6)	62972 (6)	62980 (6)	328.15 (1)	226.46 (6)	0.00	101.69
Weibull	5	-31478 (5)	62961 (5)	62975 (5)	8954.45 (7)	186.37 (5)	0.00	8768.08
Gamma	5	-31470 (4)	62945 (4)	62959 (4)	7294.32 (5)	170.27 (3)	0.00	7124.05
Generalised Gamma	5	-31451 (3)	62908 (3)	62930 (3)	5876.52 (4)	183.16 (4)	0.00	5693.36
Generalised F	5	-31382 (1)	62773 (1)	62801 (1)	5268.73 (2)	58.59 (1)	0.00	5210.14
Burr	5	-31419 (2)	62844 (2)	62865 (2)	8877.27 (6)	144.19 (2)	0.00	8733.07
Silver futures buy price fluctuations, MCX								
Brownian Motion	30	-7964 (7)	15933 (7)	15943 (7)	493.54 (2)	303.78 (7)	0.00	189.76
Exponential	30	-7873 (6)	15748 (6)	15753 (2)	342.26 (1)	264.89 (5)	0.00	77.37
Weibull	30	-7869 (4)	15743 (3)	15754 (3)	904.51 (5)	257.19 (3)	0.00	647.32
Gamma	30	-7871 (5)	15745 (5)	15756 (5)	969.76 (6)	264.45 (4)	0.00	705.31
Generalised Gamma	30	-7869 (3)	15743 (4)	15759 (6)	778.45 (4)	245.11 (2)	0.00	533.34
Generalised F	30	-7841 (1)	15689 (1)	15711 (1)	735.21 (3)	206.33 (1)	0.00	528.88
Burr	30	-7867 (2)	15739 (2)	15756 (4)	1014.17 (7)	270.22 (6)	0.00	743.95
Brownian Motion	10	-21002 (7)	42009 (7)	42022 (7)	1610.12 (2)	1017.1 (7)	0.00	593.01
Exponential	10	-20568 (6)	41137 (6)	41144 (4)	550.39 (1)	488.51 (6)	0.00	61.87
Weibull	10	-20564 (5)	41132 (4)	41145 (5)	4088.54 (6)	473.85 (4)	0.00	3614.69
Gamma	10	-20559 (3)	41122 (3)	41135 (3)	3435.51 (5)	453.06 (3)	0.00	2982.44
Generalised Gamma	10	-20548 (2)	41102 (2)	41121 (2)	2676.24 (4)	396.33 (2)	0.00	2279.91
Generalised F	10	-20470 (1)	40949 (1)	40975 (1)	2422.89 (3)	272.87 (1)	0.00	2150.02
Burr	10	-20564 (4)	41134 (5)	41153 (6)	4909.12 (7)	476.09 (5)	0.00	4433.03
Brownian Motion	5	-38832 (7)	77668 (7)	77682 (7)	5383.03 (2)	2707.34 (7)	0.00	2675.69
Exponential	5	-37704 (6)	75411 (6)	75418 (6)	1066.38 (1)	1018.65 (6)	0.00	47.73
Weibull	5	-37636 (4)	75276 (4)	75291 (4)	11217.17 (7)	829.31 (4)	0.00	10387.86
Gamma	5	-37610 (3)	75225 (3)	75239 (3)	8046.08 (5)	752.09 (3)	0.00	7293.99
Generalised Gamma	5	-37595 (2)	75197 (2)	75218 (2)	6397.39 (4)	674.72 (2)	0.00	5722.68
Generalised F	5	-37449 (1)	74907 (1)	74935 (1)	5884.2 (3)	459.89 (1)	0.00	5424.3
Burr	5	-37636 (5)	75279 (5)	75300 (5)	11192.3 (6)	830.54 (5)	0.00	10361.76

Table 5.5 continued: Maximum log-likelihood estimates of the implied discrete distributions of the exponential, Weibull, gamma, generalised gamma, generalised F and Burr distribution for the buy price fluctuation dataset together with the maximum log-likelihood, Akaike information criterion, Bayesian information criterion, Pearson's χ^2 goodness of fit statistic and the associated p-value together with the test statistics obtained from traditional log-likelihood estimator and the improvement gained.

Model	T	LOGLIK	AIC	BIC	X_C^2	X_D^2	p-value	$X_C^2 - X_D^2$
Natural gas futures buy price fluctuations, MCX								
Brownian Motion	30	-5761 (7)	11526 (7)	11537 (7)	117.32 (2)	79.43 (4)	0.00	37.89
Exponential	30	-5761 (6)	11524 (5)	11530 (3)	117.32 (1)	79.43 (3)	0.00	37.89
Weibull	30	-5759 (4)	11522 (4)	11533 (5)	1440.5 (6)	83.33 (6)	0.00	1357.17
Gamma	30	-5761 (5)	11525 (6)	11536 (6)	1185.71 (5)	81.25 (5)	0.00	1104.45
Generalised Gamma	30	-5755 (3)	11517 (3)	11533 (4)	994.09 (4)	85.1 (7)	0.00	908.99
Generalised F	30	-5739 (1)	11486 (1)	11507 (1)	901.8 (3)	60.53 (1)	0.00	841.27
Burr	30	-5748 (2)	11502 (2)	11518 (2)	1697.4 (7)	78.65 (2)	0.00	1618.75
Brownian Motion	10	-14386 (7)	28777 (7)	28790 (7)	500.63 (1)	282.1 (7)	0.00	218.53
Exponential	10	-14386 (6)	28775 (6)	28781 (6)	500.63 (2)	282.1 (6)	0.00	218.54
Weibull	10	-14298 (4)	28601 (4)	28614 (3)	8238.05 (6)	197.95 (3)	0.00	8040.1
Gamma	10	-14308 (5)	28620 (5)	28633 (5)	4949.69 (4)	190.03 (2)	0.00	4759.66
Generalised Gamma	10	-14297 (3)	28600 (3)	28620 (4)	4179.3 (3)	203.55 (5)	0.00	3975.75
Generalised F	10	-14257 (1)	28522 (1)	28548 (1)	5467.13 (5)	139.61 (1)	0.00	5327.52
Burr	10	-14288 (2)	28583 (2)	28602 (2)	8271.63 (7)	201.26 (4)	0.00	8070.38
Brownian Motion	5	-25383 (7)	50769 (7)	50784 (7)	2102.01 (2)	668.36 (7)	0.00	1433.65
Exponential	5	-25382 (6)	50766 (6)	50773 (6)	2035.08 (1)	661.02 (6)	0.00	1374.06
Weibull	5	-24994 (4)	49992 (4)	50006 (4)	20360.8 (7)	163.8 (3)	0.00	20197
Gamma	5	-25033 (5)	50069 (5)	50083 (5)	10586.32 (5)	160.72 (2)	0.00	10425.59
Generalised Gamma	5	-24989 (3)	49984 (3)	50005 (3)	9140.45 (4)	175.95 (5)	0.00	8964.5
Generalised F	5	-24918 (1)	49843 (1)	49872 (1)	8545.6 (3)	68.76 (1)	0.00	8476.84
Burr	5	-24972 (2)	49950 (2)	49972 (2)	19797.52 (6)	167.48 (4)	0.00	19630.05

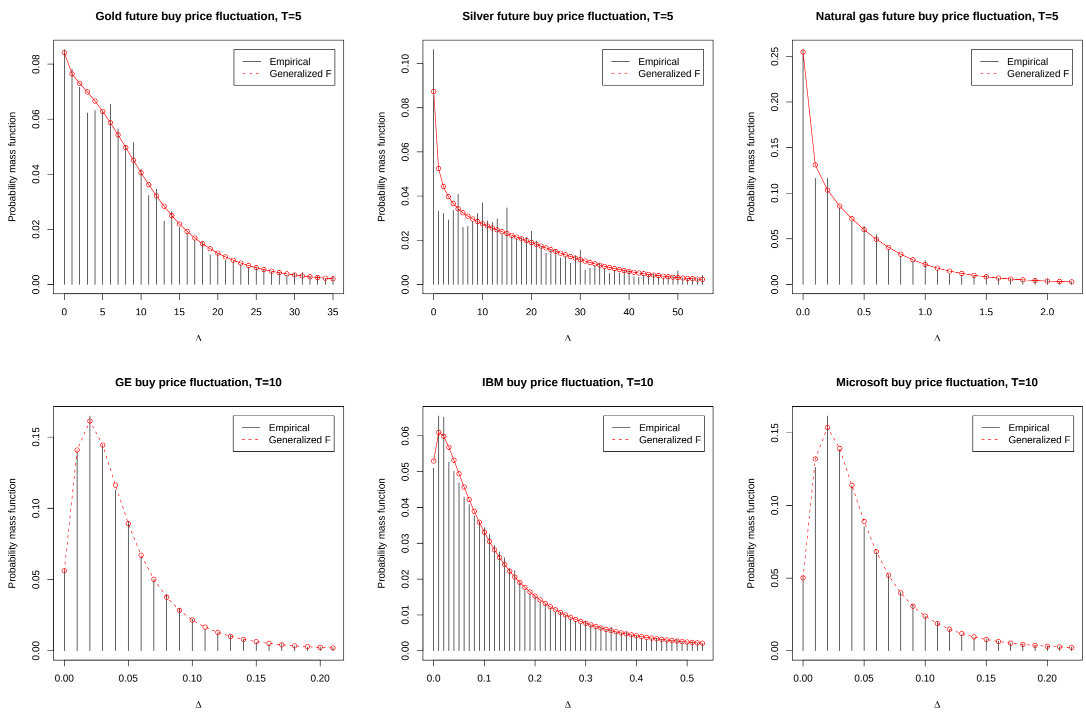


Figure 5.14: Examples of empirical distribution of buy price fluctuation and the estimated exponential, Weibull, generalised gamma and generalised F distributions obtained from maximum log-likelihood estimator of the discrete distribution implied by these distributions.

5.5 Conditional distribution for price fluctuations

Whilst the models discussed in the previous section mainly focus on the unconditional distribution of price fluctuations, it may be more useful to focus on the conditional distribution of price fluctuations since the autocorrelation function of price fluctuations studied in Section 5.3.2 indicates strong correlation between current and past price fluctuations. To achieve this, this section studies the possibility of using time series analysis techniques to model this conditional probability by fitting the price fluctuation dataset to three major time series models: the autoregressive moving average (ARMA) model [75], the generalised autoregressive conditional heteroskedasticity (GARCH) model [12] and the autoregressive conditional duration (ACD) model [28]. Since the result in the previous section indicates that the traditional maximum likelihood estimator is not a good candidate to estimate model parameters from price fluctuation dataset, we also try to estimate the parameters of these models by maximising the likelihood of the discrete distribution implied by the model, rather than maximising the likelihood of the model, as in the previous section. An experiment is then conducted to compare the performance of each model as well as the performance of the proposed estimator and the traditional maximum likelihood estimator.

This section is organised as follows. Section 5.5.1 discusses the ARMA model for price fluctuations as a natural starting point. In Section 5.5.2, an ARMA-GARCH type model that models the conditional mean using the ARMA model and the conditional variance using the GARCH model is discussed. The ACD model, which is a time series model for non-negative random variables, is introduced in Section 5.5.3. The result obtained from fitting these models to the price fluctuation dataset will be presented and analysed in Section 5.5.4. Finally, Section 5.5.5 provides a brief summary of the results obtained in this study.

5.5.1 ARMA model for price fluctuations

The first model considered here is the autoregressive moving average model for price fluctuations that models the price fluctuation at the i-th time step by

$$\Delta_i = \omega + \sum_{j=1}^{p} \alpha_j \Delta_{i-1} + \sum_{j=1}^{q} \beta_j z_{i-j} + z_i \quad (5.24)$$

$$z_i = \sigma \epsilon_i, \quad (5.25)$$

where $\{\epsilon_i\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with zero mean and unit variance, and $\theta = (\sigma, \omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$ is a parameter of the model. To account for the intraday seasonality effects discussed in Section 5.3.2, one common solution is to generate seasonally adjusted series by differencing out the time-of-day effects. In our case, we decompose the price fluctuations into a deterministic and stochastic component by assuming that the deterministic seasonality effects act additively, thus

$$\Delta_i = \tilde{\Delta}_i + s(t_i), \quad (5.26)$$

where $\tilde{\Delta}_i$ denotes the seasonally adjusted price fluctuation and $s(t_i)$ is the seasonality component at time t_i. The seasonality component can be specified by a time-of-day equation

$$s(t_i) = \nu_1 I_1(t_i) + \nu_2 I_2(t_i) + \ldots + \nu_s I_s(t_i), \tag{5.27}$$

where $(\nu_1, \ldots, \nu_s)$ are the parameters to be estimated and $I_k(t_i)$ is an indicator function whose value can be either one or zero indicating whether the time t_i is in a particular time interval of the day or not. Although this seasonality function can be jointly estimated with the ARMA parameters, it is more common to employ a two-step estimation approach where, in the first step, the price fluctuations are seasonally filtered and, in the second step, parameters of the ARMA model are estimated from the deseasonalised price fluctuation time series. Consequently, the seasonally adjusted price fluctuation series are modelled using the ARMA model, so that the price fluctuations at the i-th time step is characterised by

$$\Delta_i - s(t_i) = \omega + \sum_{j=1}^{p} \alpha_j \big(\Delta_{i-1} - s(t_i)\big) + \sum_{j=1}^{q} \beta_j z_{i-j} + z_i. \tag{5.28}$$

If $\{\epsilon_i\}$ has a probability density function $f_\epsilon(.)$, the conditional distribution of price fluctuation at the i-th time step for this ARMA model will be given by

$$f(\Delta_i|\mathcal{F}_{i-1}) = \frac{1}{\sigma} f_\epsilon\left(\frac{z_i}{\sigma}\right) \tag{5.29}$$

where $\mathcal{F}_i$ is the information set available after the i-th time step. To account for the discreteness and non-negativity of price fluctuations, we apply the method proposed in Section 5.4.3 to model this conditional distribution. Particularly, if the tick size of the considered asset is δ, the conditional distribution of the price fluctuation at the i-th time step can be estimated by

$$f_{ARMA}(\Delta_i|\mathcal{F}_{i-1}) = \begin{cases} F_\epsilon\big((z_i+\delta)/\sigma\big) & \text{if } \Delta_i = 0 \\ F_\epsilon\big((z_i+\delta)/\sigma\big) - F_\epsilon\big(z_i/\sigma\big) & \text{otherwise} \end{cases}, \tag{5.30}$$

where $F_\epsilon(.)$ is the cumulative distribution function of $\{\epsilon_i\}$.

Clearly, the ARMA model can be specified based on any distribution defined on a real value support. In time series forecasting literature, a standard way is to utilise the standard normal distribution whose probability density function and cumulative distribution function are specified by

$$f_N(\epsilon) = \exp\left(-\epsilon^2/2\right)/\sqrt{2\pi}, \tag{5.31}$$

$$F_N(\epsilon) = \big(1 + \text{erf}(\epsilon/\sqrt{2})\big)/2, \tag{5.32}$$

where $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ is the Gauss error function. However, this distribution might not be an appropriate candidate to model the residuals of the price fluctuation time series, which is highly skewed. To account for this, we also utilise the asymmetric Laplace distribution [90] to model price fluctuation time series. The probability density function and the cumulative distribution function of this distribution are of the form

$$f_{AL}(\epsilon;\kappa) = \begin{cases} \frac{\kappa\sqrt{2}}{1+\kappa^2}\exp\left(-\sqrt{2}\kappa(\epsilon-\theta)\right) & \text{if } \epsilon \geq \theta \\ \frac{\kappa\sqrt{2}}{1+\kappa^2}\exp\left(\sqrt{2}\kappa^{-1}(\epsilon-\theta)\right) & \text{if } \epsilon < \theta \end{cases}, \tag{5.33}$$

$$F_{AL}(\epsilon;\kappa) = \begin{cases} \frac{1}{1+k^2}\left(1+\kappa^2-\exp\left(-\sqrt{2}\kappa(\epsilon-\theta)\right)\right) & \text{if } \epsilon \geq \theta \\ \frac{k^2}{1+k^2}\exp\left(\sqrt{2}\kappa^{-1}(\epsilon-\theta)\right) & \text{if } \epsilon < \theta \end{cases}, \tag{5.34}$$

$$\theta = \frac{1}{\sqrt{2}}\left(\kappa-\kappa^{-1}\right). \tag{5.35}$$

Given a time series of price fluctuations $\Delta = (\Delta_1, \ldots, \Delta_N)$, the parameter of this model will be estimated using two different approaches. The first method estimates model parameters by maximising the conditional log-likelihood computed from Equation (5.29), while the second approach estimates the parameters by maximising the conditional log-likelihood computed from Equation (5.30). In particular, the maximum likelihood estimator of the first approach is obtained by maximising the conditional log-likelihood function

$$\ln \mathcal{L}(\Delta;\theta) = \sum_{i=p+1}^{N} \ln f_\epsilon\left(\frac{z_i}{\sigma}\right) - N\ln\sigma, \tag{5.36}$$

where z_i are computed from Equation (5.28) using the observations $\Delta_1, \ldots, \Delta_N$ with $z_i = \mathbb{E}Z_i = 0$, for $i = \min(p-q+1, p), \ldots, p$. Similarly, the maximum likelihood estimator of the second approach can be obtained by maximising the conditional log-likelihood function

$$\ln \mathcal{L}(\Delta;\theta) = \sum_{i=p+1}^{N} \ln f_{ARMA}(\Delta_i|\mathcal{F}_{i-1}). \tag{5.37}$$

5.5.2 ARMA-GARCH model for price fluctuations

More sophisticated specifications for price fluctuations might be obtained by considering the non-linear class of autoregressive conditionally heteroscedastic model introduced by Engle [27]. This class was then generalised by Bollerslev [12] to the GARCH process. Accordingly, this section considers ARMA models whose residual is driven by the GARCH process so that the price fluctuation at the i-th time step is modelled by

$$\Delta_i = \omega + \sum_{j=1}^{p}\alpha_j\Delta_{i-1} + \sum_{j=1}^{q}\beta_j z_{i-j} + z_i, \tag{5.38}$$

$$\sigma_i^2 = \gamma + \sum_{j=1}^{P}\phi_j z_{i-j}^2 + \sum_{j=1}^{Q}\lambda_j\sigma_{i-j}^2, \tag{5.39}$$

$$z_i = \sigma_i\epsilon_i, \tag{5.40}$$

where $\theta = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \gamma, \phi_1, \ldots, \phi_P, \lambda_1, \ldots, \lambda_Q)$ is a parameter of this model, and $\{\epsilon_i\}$ is a sequence of i.i.d. random variables with zero mean and unit variance. By using the method discussed in the previous section to account for the intraday seasonality effect, the conditional mean function in Equation (5.38) is transformed to

$$\Delta_i - s(t_i) = \omega + \sum_{j=1}^{p} \alpha_j \big(\Delta_{i-1} - s(t_i)\big) + \sum_{j=1}^{q} \beta_j z_{i-j} + z_i, \tag{5.41}$$

where $s(t_i)$ is the seasonality component at time t_i computed from Equation (5.27). Accordingly, the conditional distribution of price fluctuation at the i-th time step implied from this model is given by

$$f(\Delta_i|\mathcal{F}_{i-1}) \quad = \quad \frac{1}{\sigma_i} f_\epsilon \left(\frac{z_i}{\sigma_i}\right), \tag{5.42}$$

while the conditional distribution of price fluctuation at the i-th time step from the extension to account for the discreteness and non-negativity of price fluctuations is given by

$$f_{GARCH}(\Delta_i|\mathcal{F}_{i-1}) = \begin{cases} F_\epsilon\big((z_i+\delta)/\sigma_i\big) & \text{if } \Delta_i = 0 \\ F_\epsilon\big((z_i+\delta)/\sigma_i\big) - F_\epsilon\big(z_i/\sigma_i\big) & \text{otherwise} \end{cases}, \tag{5.43}$$

where $f_\epsilon(.)$ and $F_\epsilon(.)$ is the probability density function and cumulative distribution function of $\{\epsilon_i\}$ respectively.

Given a time series of price fluctuations $\Delta = (\Delta_1, \ldots, \Delta_N)$, the parameter of this model will be estimated using two different approaches. The first method estimates this parameter by maximising the conditional log-likelihood computed from Equation (5.42) while the second approach estimates the parameters by maximising the conditional log-likelihood computed from Equation (5.43). In particular, the maximum likelihood estimator of the first approach is obtained by maximising the conditional log-likelihood function

$$\ln \mathcal{L}(\Delta;\theta) = \sum_{i=p+1}^{N} \left(\ln f_\epsilon \left(\frac{z_i}{\sigma_i}\right) - \ln \sigma_i \right), \tag{5.44}$$

where z_i is computed from Equation (5.41) using the observations $\Delta_1, \ldots, \Delta_N$ with $z_i = \mathbb{E}Z_i = 0$ for $i \leq p$, while σ_i are computed recursively from Equation (5.39) and (5.40), with $\sigma_i^2 = \hat{\sigma}^2$ for all $i \leq 0$ when $\hat{\sigma}^2$ is the sample variance of the GARCH residual $\{z_i\}$. Similarly, the maximum likelihood estimator of the second approach can be obtained by maximising the conditional log-likelihood function

$$\ln \mathcal{L}(\Delta;\theta) = \sum_{i=p+1}^{N} \ln f_{GARCH}(\Delta_i|\mathcal{F}_{i-1}). \tag{5.45}$$

5.5.3 ACD model for price fluctuations

Another time series analysis technique that we can utilise to model price fluctuation time series is the autoregressive conditional duration (ACD) model originally proposed by Engle and Russel [28] to model trade durations. The basic idea behind the ACD model is a dynamic parameterisation of the conditional mean function. Specifically, let ψ_i be the expectation of the i-th observation, the basic ACD model for price fluctuations can be specified by

$$\Delta_i = \psi_i \epsilon_i, \tag{5.46}$$

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j \Delta_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j}, \tag{5.47}$$

where $\{\epsilon_i\}$ is an independent identically distributed non-negative random variable with mean one and finite second moment. To account for the intraday seasonality effects, a common solution is to generate seasonally adjusted series by partialling out the time-of-day effects. In this case, we decompose the price fluctuations into a deterministic and stochastic component by assuming that the deterministic seasonality effects act multiplicatively so that

$$\Delta_i = \tilde{\Delta}_i s(t_i), \tag{5.48}$$

where $s(t_i)$ is the seasonality component at time t_i computed from Equation (5.27). Accordingly, the seasonal adjusted ACD model is characterised by

$$\frac{\Delta_i}{s(t_i)} = \psi_i \epsilon_i, \tag{5.49}$$

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j \frac{\Delta_{i-j}}{s(t_{i-j})} + \sum_{j=1}^{q} \beta_j \psi_{i-j}. \tag{5.50}$$

If $\{\epsilon_i\}$ has a probability density function $f_\epsilon(.)$, the conditional distribution of price fluctuation at the i-th time step for this ACD model will be given by

$$f(\Delta_i|\mathcal{F}_{i-1}) = \frac{1}{\psi_i s(t_i)} f_\epsilon\left(\frac{\Delta_i}{\psi_i s(t_i)}\right) \tag{5.51}$$

where $\mathcal{F}_i$ is the information set available after the i-th time step. To account for the discreteness and non-negativity of price fluctuations, we apply the method proposed in Section 5.4.3 to model this conditional distribution. In particular, if the tick size of the considered asset is δ, the conditional distribution of the price fluctuation at the i-th time step will be given by

$$f_{ACD}(\Delta_i|\mathcal{F}_{i-1}) = \begin{cases} F_\epsilon\left(\frac{\Delta_i+\delta}{\psi_i s(t_i)}\right) & \text{if } \Delta_i = 0 \\ F_\epsilon\left(\frac{\Delta_i+\delta}{\psi_i s(t_i)}\right) - F_\epsilon\left(\frac{\Delta_i}{\psi_i s(t_i)}\right) & \text{otherwise} \end{cases}, \tag{5.52}$$

where $F_\epsilon(.)$ is the cumulative distribution function of $\{\epsilon_i\}$.

Since the ACD model is a model for non-negative random variables, the model must be specified based on a distribution defined on non-negative support. Accordingly, the exponential [28], Weibull [28], and generalised gamma [64] distribution have all been previously used in the context of the ACD model. The probability density function and the cumulative distribution of these distributions are characterised by

$$\begin{aligned}
f_E(\epsilon) &= \exp(-\epsilon), && (5.53)\\
F_E(\epsilon) &= 1-\exp(-\epsilon), && (5.54)\\
f_W(\epsilon;\kappa) &= \frac{\kappa}{\epsilon}\left(\epsilon\Gamma(1+1/\kappa)\right)^{\kappa}\exp\left(-\left(\epsilon\Gamma(1+1/\kappa)\right)^{\kappa}\right), && (5.55)\\
F_W(\epsilon;\kappa) &= 1-\exp\left(-\left(\epsilon\Gamma(1+1/\kappa)\right)^{\kappa}\right), && (5.56)\\
f_{GG}(\epsilon;\kappa,\lambda) &= \frac{\kappa}{\epsilon\Gamma(\lambda)}\left(\frac{\epsilon\Gamma(\lambda+1/\kappa)}{\Gamma(\lambda)}\right)^{\kappa\lambda}\exp\left[-\left(\frac{\epsilon\Gamma(\lambda+1/\kappa)}{\Gamma(\lambda)}\right)^{\kappa}\right], && (5.57)\\
F_{GG}(\epsilon;\kappa,\lambda) &= \frac{\gamma\left(\lambda,\left[\epsilon\Gamma(\lambda+1/\kappa)/\Gamma(\lambda)\right]^{\kappa}\right)}{\Gamma(\lambda)}, && (5.58)
\end{aligned}$$

where $\kappa > 0$ and $\lambda > 0$ are shape parameters of the Weibull and generalised gamma distribution, $\gamma(\lambda, x) = \int_0^x t^{\lambda-1}e^{-t}dt$ is the lower incomplete gamma function, and $\Gamma(\lambda) = \int_0^\infty t^{\lambda-1}e^{-t}dt$ is the gamma function. It is not difficult to see that these three distributions are nested since the generalised gamma distribution reduces to the Weibull distribution when $\lambda = 1$, and to the exponential distribution when $\lambda = \kappa = 1$. As a result, straight forward parametric tests can be employed to select the distribution that better describes the dataset.

Given a time series of price fluctuations $\Delta = (\Delta_1, \dots, \Delta_N)$, the parameters of the ACD model will be estimated using two different approaches. The first method estimates model parameters by maximising the conditional log-likelihood computed from Equation (5.51), while the second approach estimates the parameters by maximising the conditional log-likelihood computed from Equation (5.52). In particular, the maximum likelihood estimator of the first approach is obtained by maximising the conditional log-likelihood function

$$\ln\mathcal{L}(\Delta;\theta) = \sum_{i=p+1}^{N}\left(\ln f_\epsilon\left(\frac{\Delta_i}{\psi_i s(t_i)}\right) - \ln\psi_i - \ln s(t_i)\right), \tag{5.59}$$

where ψ_i are computed from Equation (5.50) using the observations $\Delta_1, \dots, \Delta_N$ with $\psi_i = \hat{\Delta}$ for all $i \leq 0$ when $\hat{\Delta}$ is a sample mean of $(\Delta_1, \dots, \Delta_N)$. Similarly, the maximum likelihood estimator of the second approach can be obtained by maximising the conditional log-likelihood function

$$\ln\mathcal{L}(\Delta;\theta) = \sum_{i=p+1}^{N}\ln f_{ACD}(\Delta_i|\mathcal{F}_{i-1}). \tag{5.60}$$

5.5.4 Experimental results

To determine the most suitable model for modelling the conditional distribution of price fluctuations, this section tries to fit these three models to the buy price fluctuations dataset described in Section 5.3.1 by maximising both the original likelihood function and the modified likelihood function that account for the discreteness of the dataset.

Performance measures

The results obtained from fitting the above models will be evaluated against two different performance measures, which are Brier's score [15] and Epstein's score [29].

The Brier score is a proper score function that measures the accuracy of a set of probability assessment using the average squared deviation between predicted probabilities for a set of events and their outcomes. In particular, when we consider a categorical variable whose sample space consists of a finite number K of mutually exclusive events, the Brier score is given by

$$B_s = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}(p_{ki} - d_{ki})^2, \tag{5.61}$$

where p_{ki} is the predicted probability of category k at the i-th time step, while d_{ki} is an indicator variable which takes the value 1 if the outcome at the i-th time step falls in category k and 0 otherwise. The range for the Brier score is usually stated as $0 \leq B_s \leq 2$, though the extreme values can be obtained in extreme circumstances when all probability is assigned to a single category and the outcome either does or does not fall into it. In general, there is a non-zero lower bound that corresponds to the best fit, and thus a lower score represents higher accuracy. In our application, K is set at one plus the price level that contained the first 98% of the observations, or equivalently

$$K = 1 + \min\{k \in \mathbb{N} | \Pr\{\Delta_i \leq \delta k\} < 0.98\}, \tag{5.62}$$

where δ is the tick size of the considered instrument. The category k for $1 \leq k < K$ is corresponding to the situation when the price fluctuation is equal to $k\delta$, and the category $k = K$ is corresponding to the situation when the price fluctuation is greater than $K\delta$.

Since the Brier score is indifferent to the ordering of the categories, it might be more useful to consider a score that also considers the ordering of the categories. To achieve this, Epstein [29] propose to replace the density functions implicit in the Brier score with their corresponding cumulative distribution functions defined as

$$P_{ki} = \sum_{j=1}^{k} p_{ji}, \text{ and } D_{ki} = \sum_{j=1}^{k} d_{ji}, \tag{5.63}$$

where $k = 1, \ldots, K$ with $P_{ki} = D_{ki} = 1$. The ranked probability score is then calculated from

$$E_s = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}(P_{ki} - D_{ki})^2. \tag{5.64}$$

With this equation, the Epstein's score penalises forecasts less severely when their probabilities are close to the actual outcome and more severely when their probabilities are further away from the actual outcome.

Table 5.6: List of conditional models for price fluctuations utilised to fit the buy price fluctuations dataset.

Short name	Model	Distribution
GF	-	Generalised F
NARMA	ARMA(3,3)	Normal
ALARAMA	ARMA(3,3)	Asymmetric laplace
NGARCH	ARMA(3,3)-GARCH(3,3)	Normal
AGARCH	ARMA(3,3)-GARCH(3,3)	Asymmetric laplace
EACD	ACD(3,3)	Exponential
WACD	ACD(3,3)	Weibull
GGACD	ACD(3,3)	Generalised gamma

Results

For each instrument and for each time frame of price fluctuations, we estimate the ARMA, ARMA-GARCH and ACD models summarised in Table 5.6, and compute the Brier's score, Epstein's score and the $\{\epsilon_i\}$ series implied by each model. The parameters of these models are estimated by maximising the likelihood of the original model as well as the likelihood of the extension to account for the discreteness and non-negativity of price fluctuation. We specify the autoregressive models with three lags as our primarily result at the longest time frame indicates that the best model according to the AIC and BIC criterion never utilises more than three lags. The evaluation will be done using both in-sample and out-of-sample result by using the first 80% of the sample as a training dataset. To establish a benchmark, we also evaluate the Brier's score and Epstein's score based on the best unconditional models (i.e. the generalised F distribution) for price fluctuations described in the previous section.

The results obtained from fitting these models to the buy price fluctuation dataset are illustrated in Table 5.7 and 5.8. Table 5.7 reports the in-sample Brier's score, Epstein's score, and p-values of the Ljung-Box Q-statistic for ϵ and ϵ^2 using the first 10 autocorrelations for each estimator as well as the improvement gained from maximising the modified likelihood function rather than the original likelihood function as measured by Brier's score and Epstein's score. The result obtained from maximising the modified likelihood function that accounts for the discreetness and non-negativity of the price fluctuations dataset, reported in the last two columns of the table, indicates that the modified likelihood provides better fit to the dataset than maximising the original likelihood function in most situations since the improvements reported are generally positive. A closer inspection reveals that, in case of the ACD model, the modified approach always provide better result for the ACD model with the Weibull and generalised gamma distribution having the problem as discussed in Section 5.4.2. As expected, the improvement gained in the case of the exponential distribution is somewhat limited since it does not suffer from the same problem. For the ARMA and ARMA-GARCH model, the improvement gained in the case of the normal distribution is generally higher than that of the asymmetric Laplace distribution. However, the gains obtained in this case are not consistent as in the ACD model. This indicates that it is better to utilise the modified likelihood to estimate the parameters of the ACD model, while we might need to apply both estimators to estimate the parameters of the ARMA and ARMA-GARCH model.

Among these three models, the ACD model seem to be the best model for modelling the price

Table 5.7: In-sample result obtained from a traditional maximum likelihood and the modified maximum likelihood estimated of several time series analysis models together with the maximum log-likelihood, the Brier's score, the Epstein's score and the p-value of Ljung-Box Q-statistic based on the first 10 autocorrelations of the residual and squared residual sequences.

Model	T	Original likelihood				Modified likelihood				ΔBs	ΔEs
		Bs	Es	$Q(10)$	$Q^2(10)$	Bs	Es	$Q(10)$	$Q^2(10)$		
GE buy price fluctuations, NYSE											
GF	60	0.9988 (13)	5.22 (16)	-	-	0.9533 (5)	5.02 (13)	-	-	0.0455	0.21
NARMA	60	1.0048 (14)	4.89 (11)	29.81	100.00	0.9917 (11)	4.84 (10)	16.09	100.00	0.0131	0.05
ALARMA	60	0.9685 (9)	4.82 (9)	0.00	99.99	0.9671 (8)	4.81 (7)	0.00	100.00	0.0014	0.02
NGARCH	60	1.0058 (15)	5.02 (14)	0.00	100.00	0.9828 (10)	4.81 (8)	68.31	100.00	0.0230	0.20
ALGARCH	60	0.9669 (7)	4.78 (6)	65.12	100.00	0.9655 (6)	4.76 (5)	84.53	100.00	0.0014	0.02
EACD	60	0.9518 (4)	4.65 (4)	98.71	100.00	0.9515 (3)	4.61 (3)	99.98	100.00	0.0003	0.03
WACD	60	0.9962 (12)	5.01 (12)	93.69	100.00	0.9480 (2)	4.61 (2)	99.55	100.00	0.0482	0.40
GGACD	60	1.0110 (16)	5.05 (15)	0.00	100.00	0.9468 (1)	4.58 (1)	99.99	100.00	0.0642	0.47
GF	30	0.9926 (14)	3.83 (16)	-	-	0.9374 (3)	3.65 (12)	-	-	0.0552	0.19
NARMA	30	0.9948 (15)	3.58 (11)	21.99	100.00	0.9761 (11)	3.54 (10)	2.27	100.00	0.0187	0.04
ALARMA	30	0.9485 (9)	3.53 (9)	0.00	100.00	0.9466 (8)	3.50 (7)	0.00	100.00	0.0019	0.02
NGARCH	30	0.9866 (12)	3.70 (14)	0.00	100.00	0.9681 (10)	3.51 (8)	45.61	100.00	0.0184	0.19
ALGARCH	30	0.9454 (7)	3.48 (6)	24.90	100.00	0.9428 (6)	3.46 (5)	27.49	100.00	0.0025	0.01
EACD	30	0.9422 (5)	3.41 (4)	45.51	100.00	0.9405 (4)	3.39 (3)	96.58	100.00	0.0017	0.02
WACD	30	0.9918 (13)	3.66 (13)	77.55	100.00	0.9312 (2)	3.35 (2)	98.26	100.00	0.0607	0.32
GGACD	30	1.0078 (16)	3.72 (15)	0.00	100.00	0.9297 (1)	3.35 (1)	94.63	100.00	0.0781	0.38
GF	10	0.9964 (14)	2.31 (16)	-	-	0.8981 (4)	2.09 (12)	-	-	0.0983	0.22
NARMA	10	0.9609 (13)	2.06 (11)	9.31	99.98	0.9333 (11)	2.03 (9)	0.05	99.99	0.0277	0.03
ALARMA	10	0.9078 (7)	2.03 (10)	0.00	100.00	0.9021 (6)	2.01 (8)	0.00	100.00	0.0057	0.02
NGARCH	10	0.9503 (12)	2.09 (13)	0.00	82.99	0.9208 (9)	1.99 (6)	98.56	100.00	0.0295	0.10
ALGARCH	10	0.9020 (5)	1.99 (5)	13.40	100.00	0.8952 (3)	1.96 (3)	74.38	100.00	0.0067	0.03
EACD	10	0.9216 (10)	2.00 (7)	70.98	100.00	0.9158 (8)	1.98 (4)	98.68	100.00	0.0058	0.03
WACD	10	1.0137 (15)	2.24 (14)	33.85	100.00	0.8877 (2)	1.91 (2)	99.01	100.00	0.126	0.33
GGACD	10	1.0287 (16)	2.28 (15)	0.00	100.00	0.8851 (1)	1.91 (1)	98.79	100.00	0.1436	0.37
IBM buy price fluctuations, NYSE											
GF	60	1.0012 (9)	21.62 (16)	-	-	0.9843 (5)	21.29 (15)	-	-	0.0169	0.33
NARMA	60	1.0428 (16)	19.51 (12)	0.02	0.00	1.0378 (15)	19.24 (11)	0.00	0.00	0.0050	0.27
ALARMA	60	1.0026 (11)	19.84 (14)	0.00	0.00	1.0063 (12)	19.55 (13)	0.00	0.00	-0.0037	0.29
NGARCH	60	1.0144 (13)	18.67 (7)	0.00	1.58	1.0201 (14)	18.74 (8)	0.00	17.22	-0.0057	-0.07
ALGARCH	60	0.9970 (7)	18.65 (6)	0.00	0.69	0.9988 (8)	18.55 (5)	0.00	0.37	-0.0018	0.10
EACD	60	0.9813 (4)	17.92 (4)	0.02	0.46	0.9813 (3)	17.91 (3)	0.01	0.23	0.0000	0.02
WACD	60	0.9929 (6)	19.05 (10)	0.00	4.25	0.9813 (2)	17.91 (2)	0.01	0.23	0.0116	1.14
GGACD	60	1.0014 (10)	18.92 (9)	0.00	0.83	0.9813 (1)	17.89 (1)	0.02	21.55	0.0201	1.02
GF	30	0.9990 (12)	15.58 (16)	-	-	0.9801 (5)	15.35 (15)	-	-	0.0189	0.23
NARMA	30	1.0411 (16)	14.01 (12)	1.62	0.00	1.0313 (15)	13.81 (11)	0.05	0.00	0.0097	0.20
ALARMA	30	0.9969 (9)	14.19 (14)	0.00	0.00	0.9989 (11)	14.07 (13)	0.00	0.00	-0.002	0.12
NGARCH	30	1.0177 (14)	13.51 (8)	45.50	77.73	1.0150 (13)	13.42 (7)	93.20	60.43	0.0027	0.09
ALGARCH	30	0.9936 (7)	13.32 (6)	1.46	94.57	0.9944 (8)	13.29 (5)	4.18	95.29	-0.0008	0.03
EACD	30	0.9754 (4)	12.84 (4)	90.02	99.97	0.9753 (3)	12.82 (3)	78.73	99.97	0.0000	0.01
WACD	30	0.9894 (6)	13.60 (10)	0.00	100.00	0.9752 (1)	12.79 (2)	65.01	99.98	0.0142	0.81
GGACD	30	0.9985 (10)	13.53 (9)	0.00	99.97	0.9753 (2)	12.78 (1)	89.52	99.97	0.0232	0.75
GF	10	1.0033 (12)	9.01 (16)	-	-	0.9668 (5)	8.79 (15)	-	-	0.0365	0.23
NARMA	10	1.0287 (16)	8.08 (13)	0.00	0.00	1.0176 (15)	7.94 (11)	0.00	0.00	0.0112	0.13
ALARMA	10	0.9787 (8)	8.18 (14)	0.00	0.00	0.9803 (9)	8.08 (12)	0.00	0.00	-0.0016	0.10
NGARCH	10	1.0035 (13)	7.70 (8)	0.00	99.92	0.9932 (10)	7.62 (7)	0.02	99.96	0.0102	0.08
ALGARCH	10	0.9732 (7)	7.58 (6)	0.00	97.51	0.9729 (6)	7.54 (5)	0.00	99.56	0.0003	0.03
EACD	10	0.9585 (4)	7.31 (4)	0.42	100.00	0.9581 (3)	7.30 (3)	96.30	100.00	0.0004	0.01
WACD	10	0.9973 (11)	7.91 (9)	0.68	100.00	0.9559 (2)	7.27 (2)	95.35	100.00	0.0414	0.64
GGACD	10	1.0074 (14)	7.92 (10)	0.00	100.00	0.9556 (1)	7.27 (1)	93.11	100.00	0.0519	0.64

Table 5.7 continued: In-sample result obtained from a traditional maximum likelihood and the modified maximum likelihood estimated of several time series analysis models together with the maximum log-likelihood, the Brier's score, the Epstein's score and the p-value of Ljung-Box Q-statistic based on the first 10 autocorrelations of the residual and squared residual sequences.

Model	T	Original likelihood				Modified likelihood				ΔBs	ΔEs
		Bs	Es	$Q(10)$	$Q^2(10)$	Bs	Es	$Q(10)$	$Q^2(10)$		
Microsoft buy price fluctuations, NYSE											
GF	60	1.0001 (12)	6.04 (16)	-	-	0.9567 (5)	5.82 (15)	-	-	0.0434	0.22
NARMA	60	1.0166 (16)	5.64 (11)	5.14	66.87	1.0046 (14)	5.58 (8)	2.53	67.91	0.0119	0.06
ALARMA	60	0.9729 (9)	5.65 (12)	0.00	0.89	0.9728 (8)	5.61 (10)	0.00	4.38	0.0001	0.04
NGARCH	60	1.0041 (13)	5.59 (9)	0.94	99.99	0.9902 (10)	5.51 (7)	14.73	100.00	0.0139	0.08
ALGARCH	60	0.9693 (7)	5.46 (6)	19.83	100.00	0.9686 (6)	5.44 (5)	25.46	100.00	0.0007	0.03
EACD	60	0.9566 (4)	5.27 (4)	32.62	100.00	0.9561 (3)	5.26 (3)	35.12	100.00	0.0006	0.01
WACD	60	0.9966 (11)	5.66 (13)	4.72	100.00	0.9525 (2)	5.23 (2)	18.23	100.00	0.0441	0.43
GGACD	60	1.0113 (15)	5.67 (14)	0.00	99.99	0.9524 (1)	5.23 (1)	32.23	100.00	0.0589	0.44
GF	30	0.9930 (14)	4.30 (16)	-	-	0.9420 (3)	4.11 (15)	-	-	0.051	0.19
NARMA	30	1.0075 (16)	4.01 (12)	6.27	42.59	0.9882 (11)	3.96 (8)	0.17	39.74	0.0193	0.05
ALARMA	30	0.9586 (9)	3.99 (11)	0.00	0.70	0.9566 (8)	3.97 (10)	0.00	2.55	0.002	0.02
NGARCH	30	0.9913 (13)	3.96 (9)	0.00	100.00	0.9752 (10)	3.90 (7)	1.19	100.00	0.0161	0.06
ALGARCH	30	0.9549 (7)	3.88 (6)	19.00	100.00	0.9521 (6)	3.86 (5)	42.24	100.00	0.0028	0.02
EACD	30	0.9469 (5)	3.74 (4)	18.15	100.00	0.9454 (4)	3.73 (3)	30.48	100.00	0.0015	0.01
WACD	30	0.9907 (12)	4.02 (13)	0.00	99.99	0.9363 (2)	3.68 (2)	11.23	100.00	0.0544	0.33
GGACD	30	1.0058 (15)	4.03 (14)	0.00	99.99	0.9354 (1)	3.68 (1)	29.69	100.00	0.0704	0.35
GF	10	0.9907 (14)	2.55 (16)	-	-	0.9044 (3)	2.35 (13)	-	-	0.0863	0.20
NARMA	10	0.9756 (13)	2.29 (12)	60.45	3.90	0.9486 (12)	2.26 (10)	0.00	4.20	0.027	0.03
ALARMA	10	0.9182 (7)	2.28 (11)	0.00	0.14	0.9131 (6)	2.25 (8)	0.00	0.61	0.0051	0.02
NGARCH	10	0.9472 (11)	2.25 (9)	0.00	100.00	0.9290 (10)	2.21 (7)	0.01	100.00	0.0182	0.05
ALGARCH	10	0.9108 (5)	2.21 (6)	24.50	100.00	0.9045 (4)	2.19 (5)	3.43	100.00	0.0063	0.02
EACD	10	0.9277 (9)	2.19 (4)	7.07	100.00	0.9217 (8)	2.17 (3)	10.73	100.00	0.006	0.02
WACD	10	0.9998 (15)	2.41 (14)	0.02	100.00	0.8954 (2)	2.11 (2)	26.41	100.00	0.1044	0.30
GGACD	10	1.0158 (16)	2.44 (15)	0.00	100.00	0.8935 (1)	2.11 (1)	13.47	100.00	0.1223	0.33
Gold future buy price fluctuations, MCX											
GF	30	0.9961 (11)	11.66 (13)	-	-	0.9797 (5)	11.34 (5)	-	-	0.0165	0.32
NARMA	30	1.0111 (15)	11.63 (12)	90.63	0.00	1.0109 (14)	11.60 (11)	63.63	0.00	0.0001	0.02
ALARMA	30	0.9949 (10)	11.51 (10)	3.15	0.00	1.0573 (16)	16.59 (16)	0.00	0.00	-0.0624	-5.08
NGARCH	30	1.0099 (13)	11.73 (15)	4.65	0.00	1.0091 (12)	11.67 (14)	28.91	0.00	0.0008	0.06
ALGARCH	30	0.9936 (8)	11.48 (9)	27.95	46.72	0.9938 (9)	11.40 (6)	53.01	38.93	-0.0002	0.08
EACD	30	0.9757 (4)	11.10 (4)	95.41	61.39	0.9757 (3)	11.09 (3)	95.80	60.14	0.0001	0.01
WACD	30	0.9845 (6)	11.45 (8)	92.17	26.80	0.9752 (2)	11.07 (2)	95.96	60.33	0.0093	0.38
GGACD	30	0.9925 (7)	11.43 (7)	98.48	78.87	0.9752 (1)	11.07 (1)	95.87	60.20	0.0173	0.31
GF	10	1.0230 (15)	7.04 (15)	-	-	0.9644 (5)	6.56 (5)	-	-	0.0586	0.49
NARMA	10	0.9954 (13)	6.75 (13)	88.61	0.00	0.9953 (12)	6.73 (12)	89.91	0.00	0.0001	0.02
ALARMA	10	0.9823 (9)	6.67 (11)	0.00	0.00	0.9804 (8)	6.65 (8)	0.00	0.00	0.002	0.01
NGARCH	10	0.9882 (10)	6.65 (10)	1.32	35.62	0.9896 (11)	6.65 (9)	0.71	58.17	-0.0014	0.00
ALGARCH	10	0.9799 (7)	6.63 (7)	73.68	71.38	0.9761 (6)	6.61 (6)	79.97	71.11	0.0038	0.03
EACD	10	0.9608 (4)	6.42 (4)	88.20	78.14	0.9605 (3)	6.41 (3)	87.51	71.88	0.0002	0.01
WACD	10	1.0108 (14)	7.04 (16)	54.86	35.50	0.9600 (2)	6.40 (2)	88.13	75.13	0.0508	0.64
GGACD	10	1.0247 (16)	6.96 (14)	27.91	77.21	0.9599 (1)	6.40 (1)	88.33	76.80	0.0647	0.56
GF	5	1.0516 (15)	5.22 (14)	-	-	0.9499 (5)	4.67 (10)	-	-	0.1017	0.55
NARMA	5	0.9776 (12)	4.73 (13)	74.79	0.00	0.9793 (13)	4.72 (12)	2.96	0.00	-0.0017	0.01
ALARMA	5	0.9662 (9)	4.71 (11)	0.00	0.00	0.9623 (7)	4.66 (7)	0.00	0.00	0.0039	0.05
NGARCH	5	0.9682 (10)	4.66 (8)	1.77	32.81	0.9701 (11)	4.66 (9)	3.37	34.85	-0.0019	0.00
ALGARCH	5	0.9624 (8)	4.65 (6)	3.83	29.97	0.9582 (6)	4.62 (5)	22.17	54.68	0.0043	0.02
EACD	5	0.9450 (4)	4.52 (4)	55.16	65.15	0.9444 (3)	4.51 (3)	42.94	48.64	0.0005	0.01
WACD	5	1.0498 (14)	5.25 (16)	53.11	10.47	0.9437 (2)	4.50 (2)	44.18	52.42	0.106	0.74
GGACD	5	1.0627 (16)	5.22 (15)	12.22	66.23	0.9437 (1)	4.50 (1)	44.45	53.23	0.119	0.72

Table 5.7 continued: In-sample result obtained from a traditional maximum likelihood and the modified maximum likelihood estimated of several time series analysis models together with the maximum log-likelihood, the Brier's score, the Epstein's score and the p-value of Ljung-Box Q-statistic based on the first 10 autocorrelations of the residual and squared residual sequences.

Model	T	Original likelihood				Modified likelihood				ΔBs	ΔEs
		Bs	Es	$Q(10)$	$Q^2(10)$	Bs	Es	$Q(10)$	$Q^2(10)$		
Silver future buy price fluctuations, MCX											
GF	30	1.0018 (13)	20.79 (16)	-	-	0.9888 (5)	19.76 (11)	-	-	0.0130	1.03
NARMA	30	0.9981 (8)	19.48 (8)	19.71	0.00	1.0122 (16)	19.45 (5)	13.15	99.86	-0.0140	0.03
ALARMA	30	0.9999 (11)	19.52 (9)	10.33	91.76	0.9992 (9)	19.48 (6)	39.59	98.37	0.0007	0.04
NGARCH	30	1.0099 (15)	19.92 (13)	96.26	99.68	1.0064 (14)	19.81 (12)	33.59	0.14	0.0035	0.11
ALGARCH	30	0.9999 (12)	19.54 (10)	13.12	94.85	0.9998 (10)	19.48 (7)	3.67	86.58	0.0001	0.06
EACD	30	0.9859 (4)	19.01 (4)	95.10	82.25	0.9859 (3)	19.00 (3)	95.16	82.21	0.0000	0.01
WACD	30	0.9934 (6)	20.17 (15)	92.79	96.78	0.9857 (2)	18.89 (2)	95.21	96.74	0.0077	1.27
GGACD	30	0.9980 (7)	19.99 (14)	63.10	78.89	0.9853 (1)	18.89 (1)	93.92	96.21	0.0127	1.10
GF	10	1.0255 (15)	13.82 (14)	-	-	0.9801 (7)	12.58 (8)	-	-	0.0455	1.25
NARMA	10	1.0122 (13)	19.73 (16)	99.52	99.95	1.0007 (12)	12.70 (12)	1.02	0.00	0.0116	7.03
ALARMA	10	0.9781 (6)	12.23 (6)	58.24	3.76	0.9781 (5)	12.22 (4)	58.19	3.75	0.0000	0.01
NGARCH	10	0.9908 (10)	12.64 (11)	50.92	96.46	0.9923 (11)	12.58 (9)	27.37	98.00	-0.0015	0.06
ALGARCH	10	0.9874 (9)	12.60 (10)	14.03	74.53	0.9861 (8)	12.52 (7)	20.20	79.91	0.0013	0.08
EACD	10	0.9781 (4)	12.23 (5)	58.24	3.76	0.9781 (3)	12.22 (3)	58.19	3.75	0.0000	0.01
WACD	10	1.0135 (14)	13.86 (15)	4.77	8.54	0.9781 (2)	12.22 (2)	58.18	3.72	0.0355	1.64
GGACD	10	1.0264 (16)	13.73 (13)	41.12	53.48	0.9768 (1)	12.21 (1)	54.71	9.73	0.0496	1.52
GF	5	1.0493 (15)	10.62 (14)	-	-	0.9702 (6)	9.44 (11)	-	-	0.0791	1.18
NARMA	5	0.9841 (12)	9.57 (13)	64.19	0.00	0.9886 (13)	9.45 (12)	25.50	0.00	-0.0045	0.12
ALARMA	5	0.9682 (5)	9.13 (3)	30.39	0.17	0.9774 (10)	9.38 (7)	0.00	0.00	-0.0092	-0.25
NGARCH	5	0.9772 (9)	9.43 (10)	27.28	98.31	0.9800 (11)	9.38 (8)	27.39	97.21	-0.0027	0.05
ALGARCH	5	0.9758 (8)	9.39 (9)	32.33	79.90	0.9731 (7)	9.31 (6)	49.29	89.43	0.0027	0.08
EACD	5	0.9682 (4)	9.13 (2)	30.39	0.17	0.9682 (3)	9.12 (1)	26.87	0.16	0.0000	0.01
WACD	5	1.0457 (14)	10.80 (16)	0.00	0.00	0.9675 (2)	9.14 (5)	27.93	0.16	0.0782	1.65
GGACD	5	1.0592 (16)	10.79 (15)	0.66	1.03	0.9655 (1)	9.14 (4)	5.94	0.42	0.0937	1.65
Natural gas future buy price fluctuations, MCX											
GF	30	1.0397 (14)	6.51 (14)	-	-	0.9579 (5)	5.94 (11)	-	-	0.0818	0.57
NARMA	30	0.9847 (12)	5.99 (12)	88.39	0.02	0.9866 (13)	5.93 (10)	82.77	0.04	-0.0019	0.06
ALARMA	30	0.9726 (9)	5.90 (7)	0.12	0.18	0.9700 (8)	5.87 (6)	3.38	0.21	0.0025	0.03
NGARCH	30	0.9769 (10)	6.00 (13)	20.47	99.11	0.9790 (11)	5.91 (9)	76.32	99.81	-0.0021	0.09
ALGARCH	30	0.9688 (7)	5.90 (8)	81.43	99.70	0.9653 (6)	5.85 (5)	94.20	99.51	0.0035	0.05
EACD	30	0.9497 (4)	5.61 (4)	94.21	92.49	0.9493 (3)	5.60 (3)	90.04	87.29	0.0004	0.01
WACD	30	1.0480 (15)	6.61 (16)	70.50	64.04	0.9491 (2)	5.59 (2)	90.12	87.41	0.099	1.01
GGACD	30	1.0665 (16)	6.59 (15)	2.24	3.32	0.9484 (1)	5.57 (1)	73.75	93.34	0.1181	1.02
GF	10	1.3311 (16)	4.25 (14)	-	-	0.9226 (9)	3.85 (11)	-	-	0.4084	0.4
NARMA	10	0.9295 (12)	3.90 (13)	11.22	0.00	0.9353 (13)	3.83 (10)	0.08	0.00	-0.0057	0.07
ALARMA	10	0.9245 (10)	3.86 (12)	0.00	0.00	0.9222 (8)	3.80 (7)	0.00	0.00	0.0023	0.05
NGARCH	10	0.9200 (6)	3.80 (8)	16.07	78.87	0.9249 (11)	3.78 (6)	16.16	98.53	-0.0049	0.02
ALGARCH	10	0.9212 (7)	3.81 (9)	47.99	66.70	0.9159 (5)	3.75 (5)	72.98	93.68	0.0053	0.07
EACD	10	0.9093 (3)	3.66 (4)	94.55	97.72	0.9095 (4)	3.66 (3)	83.87	96.72	-0.0001	0.01
WACD	10	1.1624 (15)	4.87 (16)	0.70	100.00	0.9080 (2)	3.66 (2)	87.24	96.68	0.2544	1.21
GGACD	10	1.1376 (14)	4.82 (15)	13.68	94.94	0.9072 (1)	3.66 (1)	78.20	97.65	0.2304	1.16
GF	5	1.0561 (14)	3.32 (14)	-	-	0.8834 (13)	2.88 (6)	-	-	0.1727	0.44
NARMA	5	0.8734 (10)	3.01 (13)	38.07	0.00	0.8802 (12)	2.92 (9)	7.98	0.00	-0.0068	0.08
ALARMA	5	0.8748 (11)	2.97 (12)	0.00	0.00	0.8684 (7)	2.91 (8)	0.00	0.00	0.0064	0.06
NGARCH	5	0.8665 (6)	2.94 (11)	0.90	98.10	0.8695 (8)	2.90 (7)	10.40	99.62	-0.003	0.03
ALGARCH	5	0.8721 (9)	2.93 (10)	68.18	83.34	0.8626 (3)	2.87 (5)	53.99	99.36	0.0095	0.07
EACD	5	0.8649 (4)	2.83 (4)	95.09	96.48	0.8665 (5)	2.83 (3)	57.29	91.69	-0.0016	0.00
WACD	5	1.1515 (16)	5.33 (16)	0.00	32.49	0.8590 (2)	2.82 (2)	81.03	91.71	0.2925	2.51
GGACD	5	1.1339 (15)	3.88 (15)	86.51	95.72	0.8583 (1)	2.82 (1)	72.47	94.94	0.2756	1.06

fluctuation dataset, since the best models according to both Brier's score and Epstein's score always belong to this class, with the generalised gamma distribution being the best distribution and the Weibull distribution being the second best distribution. The results from the ARMA-GARCH model are always better than the one obtained from the ARMA model with respect to the same distributional assumption, while the result from the asymmetric Laplace distribution is better than the one from the standard normal distribution. Interestingly, the resulted obtained from the unconditional model using the generalised F distribution regularly beat the ARMA and ARMA-GARCH model in terms of Brier's score, which does not consider the order of the categories.

An easy-to-compute in-sample test for correct specification consists of calculating sample autocorrelations of residuals for the estimated models and comparing it to the assumption of the models. Consequently, if the model is correctly specified, there must be no unexplained structure left in the residuals and, thus, the residuals time series must be independent. To determine this, we calculate the Lijung-Box statistics at 10 lags for both residuals and square residuals. The p-value reported in Table 5.7 indicates that the best model generally passes this test except in the case of IBM's 60 minute buy price fluctuations which we can reject the independent assumption at a convenient level.

The out-of-sample result reported in Table 5.8 suggest similar results. In particular, the best model among these three models is still the ACD model since the best performing model, in all datasets, belongs to this class, and the modified likelihood estimator always provides better results for the Weibull and generalised gamma distribution. This further confirms that the ACD estimated by maximising the modified likelihood model is the best candidate for modelling the price fluctuations dataset.

5.5.5 Summary

In this section, we studied several approaches for modelling the conditional distribution of price fluctuation time series. We started this section by giving an overview of three important models for forecasting time series in financial econometrics literatures. These included the autoregressive moving average (ARMA) model, the generalised autoregressive conditional heteroskedasticity (GARCH) model and the autoregressive conditional duration (ACD) model. For each of these models, we derived a modified likelihood function that account for the discreteness and non-negativity of the price fluctuations, and fit these models to the price fluctuation dataset both by maximising the original likelihood function and the modified likelihood function. Since these models can be specified based on several distributional assumptions, we utilise the normal and asymmetric Laplace distribution in case of the ARMA and ARMA-GARCH models, while we utilise the exponential, Weibull and generalised gamma distribution in case of the ACD model. The experimental results indicate that the modified likelihood function always improve the fitted result only for the ACD model under the Weibull and the generalised gamma distribution, which allow the density function to be infinite value at zero, while the gain obtained from the exponential distribution is quite limited. However, the improvement in the case of the ARMA and ARMA-GARCH model is not consistent, indicating that one might need to estimate these models using both original likelihood and modified likelihood in order to find the best model parameters. Among all models considered, the ACD model, with generalised gamma distribution estimated by maximising the modified likelihood function,

Table 5.8: Out-of-sample result obtained from a traditional maximum likelihood and the modified maximum likelihood estimated of several time series analysis models together with the maximum log-likelihood, the Brier's score, the Epstein's score and the p-value of Ljung-Box Q-statistic based on the first 10 autocorrelations of the residual and squared residual sequences.

Model	T	Original likelihood				Modified likelihood				ΔBs	ΔEs
		Bs	Es	$Q(10)$	$Q^2(10)$	Bs	Es	$Q(10)$	$Q^2(10)$		
GE buy price fluctuations, NYSE											
GF	60	1.0215 (15)	12.91 (16)	-	-	0.9822 (8)	12.32 (15)	-	-	0.0393	0.59
NARMA	60	0.9870 (11)	10.59 (9)	0.00	0.00	0.9839 (10)	10.78 (11)	0.00	0.00	0.0031	-0.19
ALARMA	60	0.9819 (7)	11.19 (14)	0.00	0.00	0.9813 (6)	11.06 (12)	0.00	0.00	0.0006	0.13
NGARCH	60	1.0814 (16)	11.11 (13)	0.71	7.78	0.9973 (13)	10.23 (6)	0.53	0.23	0.0841	0.87
ALGARCH	60	0.9823 (9)	10.35 (7)	0.00	0.00	0.9798 (5)	10.17 (5)	0.00	0.00	0.0025	0.18
EACD	60	0.9731 (4)	10.16 (4)	7.19	1.54	0.9710 (3)	9.72 (2)	11.22	1.87	0.0021	0.44
WACD	60	0.9963 (12)	10.53 (8)	0.10	1.32	0.9696 (2)	9.76 (3)	0.60	0.00	0.0267	0.76
GGACD	60	1.0013 (14)	10.73 (10)	0.00	0.00	0.9694 (1)	9.68 (1)	13.55	1.81	0.0319	1.05
GF	30	1.0244 (15)	9.15 (16)	-	-	0.9768 (10)	8.65 (15)	-	-	0.0476	0.5
NARMA	30	0.9810 (11)	7.34 (8)	0.66	70.44	0.9763 (9)	7.46 (10)	0.49	69.84	0.0047	-0.13
ALARMA	30	0.9742 (8)	7.84 (14)	0.00	46.99	0.9728 (7)	7.66 (12)	0.62	59.49	0.0014	0.17
NGARCH	30	1.0695 (16)	7.64 (11)	0.00	81.44	0.9955 (12)	7.21 (7)	5.05	56.25	0.0739	0.43
ALGARCH	30	0.9687 (5)	7.13 (6)	12.69	50.01	0.9690 (6)	7.08 (5)	13.49	43.40	-0.0003	0.05
EACD	30	0.9663 (4)	7.04 (4)	56.83	64.93	0.9641 (3)	6.85 (3)	74.57	59.42	0.0022	0.19
WACD	30	0.9968 (13)	7.36 (9)	9.73	43.61	0.9604 (2)	6.79 (1)	64.88	56.06	0.0364	0.57
GGACD	30	1.0180 (14)	7.78 (13)	8.50	75.68	0.9602 (1)	6.81 (2)	73.10	58.32	0.0578	0.97
GF	10	1.0379 (14)	5.11 (16)	-	-	0.9552 (10)	4.67 (15)	-	-	0.0827	0.45
NARMA	10	0.9610 (11)	3.96 (8)	0.11	78.45	0.9547 (9)	4.03 (9)	2.20	79.27	0.0063	-0.07
ALARMA	10	0.9544 (8)	4.28 (13)	0.00	39.70	0.9518 (7)	4.15 (11)	0.00	54.41	0.0026	0.13
NGARCH	10	1.0644 (16)	4.26 (12)	0.00	0.00	0.9649 (12)	3.87 (6)	30.35	99.93	0.0995	0.39
ALGARCH	10	0.9434 (5)	3.92 (7)	1.09	99.99	0.9425 (3)	3.84 (5)	56.10	99.90	0.0009	0.09
EACD	10	0.9469 (6)	3.80 (4)	2.61	61.63	0.9434 (4)	3.75 (3)	21.87	43.70	0.0035	0.06
WACD	10	1.0062 (13)	4.14 (10)	0.00	12.04	0.9320 (2)	3.70 (1)	35.88	35.20	0.0742	0.43
GGACD	10	1.0408 (15)	4.44 (14)	0.23	27.95	0.9314 (1)	3.71 (2)	22.21	49.68	0.1094	0.73
IBM buy price fluctuations, NYSE											
GF	60	0.9994 (10)	16.56 (16)	-	-	0.9834 (5)	16.28 (15)	-	-	0.016	0.27
NARMA	60	1.0402 (16)	16.10 (14)	68.28	97.04	1.0345 (15)	15.70 (11)	95.55	98.00	0.0057	0.4
ALARMA	60	0.9990 (9)	15.53 (8)	99.65	99.30	1.0023 (12)	15.46 (7)	97.65	99.08	-0.0033	0.07
NGARCH	60	1.0234 (14)	15.80 (12)	98.51	95.30	1.0209 (13)	15.63 (10)	98.88	98.22	0.0026	0.17
ALGARCH	60	0.9951 (7)	15.39 (6)	85.80	98.92	0.9968 (8)	15.32 (5)	92.94	99.11	-0.0016	0.06
EACD	60	0.9805 (4)	14.88 (4)	78.86	98.37	0.9805 (3)	14.87 (3)	88.70	98.31	0	0.01
WACD	60	0.9922 (6)	15.86 (13)	13.63	93.99	0.9805 (2)	14.87 (2)	88.71	98.30	0.0117	0.99
GGACD	60	0.9998 (11)	15.60 (9)	46.69	98.90	0.9803 (1)	14.86 (1)	58.76	99.13	0.0195	0.74
GF	30	0.9976 (11)	11.82 (15)	-	-	0.9787 (5)	11.63 (14)	-	-	0.0189	0.19
NARMA	30	1.0407 (16)	11.59 (13)	88.82	100.00	1.0299 (14)	11.28 (12)	93.11	100.00	0.0108	0.31
ALARMA	30	0.9935 (7)	11.13 (9)	91.36	100.00	0.9956 (10)	11.11 (8)	92.99	100.00	-0.0021	0.02
NGARCH	30	1.0400 (15)	12.23 (16)	0.00	100.00	1.0162 (13)	11.20 (10)	63.85	100.00	0.0238	1.03
ALGARCH	30	0.9942 (8)	11.09 (6)	79.15	99.99	0.9949 (9)	11.06 (5)	77.84	100.00	-0.0008	0.02
EACD	30	0.9756 (4)	10.63 (4)	56.47	97.46	0.9755 (3)	10.62 (3)	62.47	97.19	0.0001	0.01
WACD	30	0.9904 (6)	11.28 (11)	0.95	99.57	0.9753 (1)	10.59 (2)	63.78	96.94	0.0151	0.69
GGACD	30	0.9988 (12)	11.11 (7)	22.35	99.53	0.9754 (2)	10.58 (1)	78.57	98.08	0.0234	0.53
GF	10	1.0028 (13)	6.74 (16)	-	-	0.9644 (5)	6.53 (14)	-	-	0.0384	0.2
NARMA	10	1.0313 (16)	6.58 (15)	0.82	97.09	1.0184 (15)	6.38 (11)	26.29	98.10	0.0129	0.2
ALARMA	10	0.9768 (8)	6.28 (9)	1.13	96.11	0.9785 (9)	6.27 (8)	65.71	98.15	-0.0017	0.02
NGARCH	10	1.0024 (12)	6.32 (10)	0.13	100.00	0.9954 (10)	6.22 (7)	3.79	100.00	0.007	0.1
ALGARCH	10	0.9742 (7)	6.17 (6)	1.68	100.00	0.9740 (6)	6.14 (5)	9.39	100.00	0.0002	0.03
EACD	10	0.9596 (4)	5.98 (4)	36.46	100.00	0.9596 (3)	5.96 (3)	65.71	99.99	0	0.02
WACD	10	1.0002 (11)	6.50 (13)	0.54	100.00	0.9574 (2)	5.93 (2)	71.59	99.98	0.0428	0.57
GGACD	10	1.0109 (14)	6.43 (12)	0.00	99.99	0.9571 (1)	5.93 (1)	58.50	99.99	0.0538	0.49

Table 5.8 continued: Out-of-sample result obtained from a traditional maximum likelihood and the modified maximum likelihood estimated of several time series analysis models together with the maximum log-likelihood, the Brier's score, the Epstein's score and the p-value of Ljung-Box Q-statistic based on the first 10 autocorrelations of the residual and squared residual sequences.

Model	T	Original likelihood				Modified likelihood				ΔBs	ΔEs
		Bs	Es	$Q(10)$	$Q^2(10)$	Bs	Es	$Q(10)$	$Q^2(10)$		
Microsoft buy price fluctuations, NYSE											
GF	60	0.9971 (12)	4.78 (16)	-	-	0.9516 (4)	4.61 (11)	-	-	0.0455	0.17
NARMA	60	1.0158 (16)	4.67 (15)	74.19	99.99	1.0014 (14)	4.56 (10)	84.07	99.99	0.0144	0.11
ALARMA	60	0.9654 (8)	4.44 (7)	76.15	100.00	0.9656 (9)	4.45 (8)	88.04	100.00	-0.0002	-0.01
NGARCH	60	0.9997 (13)	4.65 (13)	10.13	100.00	0.9891 (10)	4.52 (9)	26.27	100.00	0.0106	0.14
ALGARCH	60	0.9628 (7)	4.40 (6)	77.65	100.00	0.9626 (6)	4.40 (5)	86.76	100.00	0.0002	0
EACD	60	0.9519 (5)	4.28 (4)	80.28	99.93	0.9513 (3)	4.27 (3)	84.95	99.94	0.0007	0.01
WACD	60	0.9965 (11)	4.66 (14)	68.01	99.87	0.9468 (2)	4.23 (2)	81.33	99.95	0.0497	0.44
GGACD	60	1.0104 (15)	4.63 (12)	36.15	99.85	0.9466 (1)	4.22 (1)	83.30	99.95	0.0638	0.41
GF	30	0.9881 (12)	3.56 (16)	-	-	0.9349 (3)	3.42 (11)	-	-	0.0532	0.14
NARMA	30	1.0098 (16)	3.50 (15)	37.50	100.00	0.9869 (11)	3.41 (10)	42.72	100.00	0.0229	0.09
ALARMA	30	0.9510 (9)	3.33 (6)	85.97	100.00	0.9492 (8)	3.33 (7)	91.58	100.00	0.0018	0
NGARCH	30	0.9886 (13)	3.43 (12)	2.92	100.00	0.9742 (10)	3.39 (9)	19.73	100.00	0.0145	0.05
ALGARCH	30	0.9486 (7)	3.33 (8)	73.47	100.00	0.9461 (6)	3.32 (5)	80.50	100.00	0.0025	0.01
EACD	30	0.9423 (5)	3.22 (4)	52.31	99.43	0.9406 (4)	3.21 (3)	71.68	99.62	0.0017	0.01
WACD	30	0.9901 (14)	3.48 (14)	3.41	99.17	0.9304 (2)	3.16 (2)	69.19	99.82	0.0597	0.32
GGACD	30	1.0049 (15)	3.47 (13)	0.11	98.68	0.9293 (1)	3.16 (1)	68.43	99.61	0.0756	0.31
GF	10	0.9841 (14)	2.11 (14)	-	-	0.8944 (3)	1.95 (12)	-	-	0.0897	0.16
NARMA	10	0.9759 (13)	1.99 (13)	8.97	1.08	0.9443 (12)	1.94 (10)	0.96	0.30	0.0316	0.05
ALARMA	10	0.9087 (7)	1.89 (6)	0.00	0.00	0.9031 (5)	1.89 (7)	0.00	0.00	0.0056	0
NGARCH	10	0.9402 (11)	1.94 (11)	0.00	100.00	0.9220 (9)	1.89 (8)	84.43	100.00	0.0182	0.04
ALGARCH	10	0.9032 (6)	1.88 (5)	0.11	99.99	0.8965 (4)	1.86 (3)	33.78	100.00	0.0067	0.02
EACD	10	0.9240 (10)	1.90 (9)	50.76	99.99	0.9175 (8)	1.88 (4)	79.19	99.98	0.0065	0.02
WACD	10	1.0013 (15)	2.11 (15)	3.88	100.00	0.8881 (2)	1.80 (2)	78.29	99.98	0.1131	0.31
GGACD	10	1.0161 (16)	2.12 (16)	0.00	80.98	0.8860 (1)	1.80 (1)	72.29	99.98	0.1301	0.32
Gold future buy price fluctuations, MCX											
GF	30	0.9979 (11)	16.06 (12)	-	-	0.9832 (6)	15.57 (5)	-	-	0.0147	0.49
NARMA	30	1.0014 (12)	16.27 (15)	0.04	0.00	1.0030 (13)	16.03 (11)	7.93	0.00	-0.0016	0.24
ALARMA	30	0.9898 (8)	15.85 (10)	10.23	0.00	1.0480 (16)	19.75 (16)	0.02	1.07	-0.0583	-3.9
NGARCH	30	1.0278 (15)	16.26 (14)	0.04	0.00	1.0276 (14)	16.25 (13)	0.04	0.00	0.0002	0.01
ALGARCH	30	0.9909 (9)	15.72 (7)	52.34	92.49	0.9914 (10)	15.83 (9)	1.26	0.00	-0.0005	-0.11
EACD	30	0.9746 (4)	15.36 (4)	61.59	65.02	0.9745 (3)	15.33 (3)	61.36	63.45	0.0001	0.03
WACD	30	0.9807 (5)	15.63 (6)	81.06	94.99	0.9744 (2)	15.31 (2)	62.44	65.43	0.0064	0.32
GGACD	30	0.9883 (7)	15.80 (8)	67.72	78.05	0.9744 (1)	15.31 (1)	61.62	63.91	0.0139	0.49
GF	10	1.0278 (16)	9.72 (16)	-	-	0.9713 (5)	9.03 (13)	-	-	0.0565	0.69
NARMA	10	0.9907 (10)	8.84 (9)	5.46	0.00	0.9912 (11)	8.85 (10)	3.09	0.00	-0.0005	-0.01
ALARMA	10	0.9843 (9)	8.98 (12)	0.00	0.00	0.9813 (8)	8.90 (11)	0.01	0.00	0.0031	0.08
NGARCH	10	0.9987 (13)	8.78 (8)	66.80	99.54	0.9927 (12)	8.70 (7)	56.52	99.69	0.0061	0.08
ALGARCH	10	0.9803 (7)	8.70 (6)	59.50	97.45	0.9781 (6)	8.66 (5)	58.12	99.25	0.0022	0.05
EACD	10	0.9663 (4)	8.51 (4)	41.33	97.10	0.9661 (3)	8.49 (3)	42.60	97.14	0.0002	0.02
WACD	10	1.0062 (14)	9.29 (15)	28.64	97.62	0.9659 (2)	8.48 (2)	42.31	97.06	0.0403	0.81
GGACD	10	1.0219 (15)	9.26 (14)	22.68	93.92	0.9658 (1)	8.48 (1)	42.11	97.02	0.056	0.78
GF	5	1.0615 (15)	6.87 (16)	-	-	0.9596 (5)	6.11 (13)	-	-	0.1019	0.76
NARMA	5	0.9766 (10)	5.92 (10)	3.89	0.00	0.9774 (11)	5.91 (8)	4.61	0.00	-0.0008	0.01
ALARMA	5	0.9764 (9)	6.09 (12)	0.00	0.00	0.9690 (8)	5.96 (11)	0.00	0.00	0.0074	0.14
NGARCH	5	0.9821 (13)	5.91 (9)	90.52	67.91	0.9804 (12)	5.86 (6)	95.02	93.78	0.0016	0.05
ALGARCH	5	0.9675 (7)	5.87 (7)	69.86	89.84	0.9661 (6)	5.83 (5)	81.72	91.98	0.0014	0.04
EACD	5	0.9549 (4)	5.73 (4)	40.04	93.14	0.9545 (3)	5.71 (3)	85.11	96.45	0.0004	0.02
WACD	5	1.0468 (14)	6.61 (15)	12.05	84.98	0.9541 (2)	5.70 (2)	85.49	96.51	0.0927	0.91
GGACD	5	1.0638 (16)	6.61 (14)	89.00	96.96	0.9541 (1)	5.70 (1)	85.58	96.52	0.1097	0.91

Table 5.8 continued: Out-of-sample result obtained from a traditional maximum likelihood and the modified maximum likelihood estimated of several time series analysis models together with the maximum log-likelihood, the Brier's score, the Epstein's score and the p-value of Ljung-Box Q-statistic based on the first 10 autocorrelations of the residual and squared residual sequences.

Model	T	Original likelihood				Modified likelihood				ΔBs	ΔEs
		Bs	Es	$Q(10)$	$Q^2(10)$	Bs	Es	$Q(10)$	$Q^2(10)$		
Silver future buy price fluctuations, MCX											
GF	30	1.0102 (12)	33.67 (16)	-	-	0.9922 (6)	31.87 (7)	-	-	0.018	1.8
NARMA	30	1.0058 (11)	31.86 (6)	59.05	0.00	1.0161 (14)	33.45 (15)	18.09	0.00	-0.0104	-1.59
ALARMA	30	0.9981 (7)	32.93 (13)	8.14	0.00	0.9982 (8)	32.55 (11)	16.30	0.00	-0.0001	0.38
NGARCH	30	1.0564 (16)	33.20 (14)	57.67	54.26	1.0335 (15)	31.63 (5)	99.67	99.87	0.0229	1.57
ALGARCH	30	1.0026 (10)	31.96 (9)	0.92	0.00	1.0149 (13)	31.90 (8)	1.23	0.00	-0.0124	0.06
EACD	30	0.9844 (4)	31.29 (4)	98.92	79.80	0.9843 (3)	31.24 (3)	98.85	79.62	0	0.05
WACD	30	0.9915 (5)	32.44 (10)	93.76	64.89	0.9839 (1)	31.10 (1)	95.95	58.31	0.0076	1.34
GGACD	30	0.9996 (9)	32.85 (12)	98.52	66.92	0.9840 (2)	31.16 (2)	97.87	68.27	0.0156	1.69
GF	10	1.0427 (15)	20.05 (16)	-	-	0.9858 (7)	18.10 (13)	-	-	0.0569	1.96
NARMA	10	0.9988 (10)	17.65 (11)	23.64	78.68	1.0000 (11)	17.59 (10)	3.65	74.40	-0.0012	0.06
ALARMA	10	0.9821 (6)	17.13 (6)	88.40	100.00	0.9820 (4)	17.11 (4)	86.66	100.00	0	0.02
NGARCH	10	1.0095 (12)	17.74 (12)	67.53	100.00	1.0097 (13)	17.43 (7)	92.00	100.00	-0.0002	0.31
ALGARCH	10	0.9926 (8)	17.58 (9)	86.82	100.00	0.9976 (9)	17.52 (8)	77.34	100.00	-0.0049	0.06
EACD	10	0.9821 (5)	17.13 (5)	88.40	100.00	0.9820 (3)	17.11 (3)	86.66	100.00	0	0.02
WACD	10	1.0234 (14)	19.48 (14)	72.58	100.00	0.9820 (2)	17.11 (2)	86.60	100.00	0.0414	2.37
GGACD	10	1.0433 (16)	19.91 (15)	80.41	100.00	0.9817 (1)	17.09 (1)	94.56	100.00	0.0616	2.82
GF	5	1.0769 (15)	14.04 (15)	-	-	0.9787 (6)	12.26 (12)	-	-	0.0982	1.78
NARMA	5	0.9959 (10)	12.15 (11)	41.93	70.40	0.9998 (12)	12.11 (9)	43.65	73.27	-0.0038	0.05
ALARMA	5	0.9763 (5)	11.74 (3)	98.06	100.00	0.9921 (9)	12.27 (13)	0.07	62.53	-0.0158	-0.54
NGARCH	5	0.9988 (11)	12.11 (10)	99.29	100.00	1.0024 (13)	12.05 (7)	99.38	100.00	-0.0036	0.06
ALGARCH	5	0.9859 (7)	12.08 (8)	99.28	100.00	0.9880 (8)	11.98 (6)	98.80	100.00	-0.0021	0.09
EACD	5	0.9763 (4)	11.74 (2)	98.06	100.00	0.9762 (3)	11.73 (1)	97.77	100.00	0	0.01
WACD	5	1.0644 (14)	13.96 (14)	20.61	100.00	0.9762 (2)	11.75 (4)	97.88	100.00	0.0881	2.21
GGACD	5	1.0896 (16)	14.35 (16)	83.49	100.00	0.9757 (1)	11.76 (5)	93.80	100.00	0.1138	2.59
Natural gas future buy price fluctuations, MCX											
GF	30	0.9921 (14)	4.16 (15)	-	-	0.9390 (6)	4.42 (16)	-	-	0.0531	-0.26
NARMA	30	0.9517 (11)	4.09 (14)	23.38	30.55	0.9537 (12)	3.93 (9)	61.07	35.76	-0.002	0.16
ALARMA	30	0.9429 (9)	4.07 (12)	70.91	64.36	0.9392 (7)	4.05 (11)	75.06	59.15	0.0037	0.02
NGARCH	30	0.9508 (10)	4.07 (13)	18.15	99.97	0.9540 (13)	4.05 (10)	50.86	99.94	-0.0032	0.01
ALGARCH	30	0.9410 (8)	3.84 (7)	77.23	99.67	0.9368 (5)	3.81 (6)	85.87	99.88	0.0042	0.03
EACD	30	0.9095 (4)	3.31 (1)	49.36	85.82	0.9095 (3)	3.32 (3)	58.11	85.22	-0.0001	-0.01
WACD	30	1.0109 (15)	3.90 (8)	94.96	96.55	0.9095 (2)	3.32 (2)	59.82	85.68	0.1014	0.59
GGACD	30	1.0187 (16)	3.76 (5)	88.24	97.57	0.9093 (1)	3.33 (4)	70.32	86.80	0.1094	0.43
GF	10	1.2377 (15)	2.66 (13)	-	-	0.8905 (6)	2.68 (14)	-	-	0.3472	-0.02
NARMA	10	0.9099 (12)	2.63 (12)	30.31	35.83	0.9175 (13)	2.52 (10)	13.02	40.82	-0.0076	0.12
ALARMA	10	0.8923 (7)	2.55 (11)	13.00	89.12	0.8946 (10)	2.50 (9)	30.03	70.78	-0.0023	0.05
NGARCH	10	0.8932 (9)	2.39 (8)	84.42	96.18	0.8994 (11)	2.37 (7)	50.68	96.12	-0.0062	0.02
ALGARCH	10	0.8930 (8)	2.37 (6)	47.93	98.67	0.8851 (5)	2.34 (5)	78.21	98.01	0.0079	0.03
EACD	10	0.8588 (1)	2.11 (1)	61.70	65.87	0.8596 (2)	2.14 (3)	58.84	58.70	-0.0009	-0.03
WACD	10	1.2431 (16)	3.98 (16)	0.00	46.19	0.8605 (3)	2.14 (2)	59.16	57.87	0.3826	1.84
GGACD	10	1.1136 (14)	2.74 (15)	52.79	44.72	0.8621 (4)	2.15 (4)	61.46	64.28	0.2515	0.59
GF	5	0.9895 (14)	2.00 (13)	-	-	0.8424 (9)	1.96 (11)	-	-	0.1471	0.04
NARMA	5	0.8507 (11)	2.02 (14)	50.08	35.19	0.8631 (13)	1.88 (9)	41.85	53.80	-0.0123	0.14
ALARMA	5	0.8390 (6)	1.97 (12)	0.01	58.26	0.8399 (8)	1.88 (10)	45.50	52.14	-0.0009	0.09
NGARCH	5	0.8396 (7)	1.76 (6)	88.72	100.00	0.8528 (12)	1.75 (5)	93.15	100.00	-0.0132	0.01
ALGARCH	5	0.8439 (10)	1.82 (8)	89.27	100.00	0.8327 (5)	1.77 (7)	99.97	100.00	0.0112	0.05
EACD	5	0.8130 (1)	1.65 (1)	89.54	99.85	0.8135 (2)	1.67 (4)	83.54	99.79	-0.0004	-0.03
WACD	5	1.1229 (16)	2.98 (16)	0.00	97.88	0.8143 (3)	1.67 (3)	83.37	99.78	0.3086	1.32
GGACD	5	1.1161 (15)	2.27 (15)	78.77	99.70	0.8154 (4)	1.67 (2)	85.02	99.79	0.3007	0.6

is the best performing model both from in-sample and out-of-sample tests.

5.6 Alternative ACD model for price fluctuations

Although the experiment performed in the previous section indicates that the autoregressive conditional duration is the most appropriate model for analysing price fluctuations dataset, the basic ACD model considered in the previous section still has several limitations since it assumes that the conditional mean adjusts proportionally to recent price fluctuations and the effects of these shocks decay exponentially over time which might not be compatible with the price fluctuations process. To further investigate this issue, this section tries to fit several extensions of the basic ACD model to the price fluctuations dataset with the aim of examining whether some extensions perform better than others, and which extension is particularly suited, or inadequate, for modelling the price fluctuations dataset. The extensions considered here are models that try to generalise the linear parameterisation of the basic ACD model in three main directions. The first approach generalises the basic model by adding an additive innovation component into the conditional mean function, while the second approach achieves this by applying the conditional mean function on the transformation of price fluctuations rather than in the plain price fluctuation. Lastly, the third approach generalises this by introducing an asymmetric response to positive and negative shocks into the conditional mean equation. Since the result in the previous section indicates that it is more appropriate to estimate ACD model parameters, by maximising the likelihood of the implied discrete distribution, the parameter of each model will be estimated solely based on this approach.

This section is organised as follows. Section 5.6.1 firstly reviews the concepts behind the ACD model. In Section 5.6.2, several extensions to the basic ACD parameterisation are discussed. The results obtained from fitting these extensions to the price fluctuation dataset are analysed and discussed in Section 5.6.3. Finally, Section 5.6.4 gives a brief discussion on the result obtained in this section.

5.6.1 The basic ACD framework

This section summarises the basic ideas behind the autoregressive conditional duration model originally proposed by Engle and Russell [28] to model the duration between two consecutive trades. In its original form, the ACD model is a stochastic process where the observation at the $i-$th time step is modelled by

$$\Delta_i = \psi_i \epsilon_i,$$

where $\mathcal{F}_{i-1}$ is the set of all information available at the i-th time step, $\{\epsilon_i\}$ is an independent and identically distributed nonnegative process with unit mean and finite second moment, $\psi_i = \mathbb{E}(\Delta_i|\mathcal{F}_{i-1})$ is a conditional mean process which is assumed to be stochastically independent of the i.i.d. sequence formed by ϵ_i and $\mathcal{F}_{i-1}$. This setup is very general and allows us to construct a variety of models by choosing different specifications for the conditional mean function, ψ_i, and different distributions for $\{\epsilon_i\}$. The conditional mean function is generally assumed to depend on its own lags as well as past observations, which can be characterised by

$$\psi_i = \psi(\Delta_{i-1}, \dots, \Delta_{i-p}, \psi_{i-1}, \dots, \psi_{i-q}). \tag{5.65}$$

Since this conditional mean function depends on the last p lags of the observed durations and on the last q lags of expected durations, this model is generally called ACD(p,q). One example of this specification is the basic model considered in the previous section, which models this conditional mean function by a linear parameterisation of the form

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j \Delta_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j}. \tag{5.66}$$

If $f_\epsilon(\epsilon)$ and $F_\epsilon(\epsilon)$ are the probability density function and cumulative distribution function of $\{\epsilon_i\}$ respectively, the conditional density of Δ_i will be given by

$$f(\Delta_i|\mathcal{F}_{i-1}) = \frac{1}{\psi_i} f_\epsilon\left(\frac{\Delta_i}{\psi_i}\right),$$

while the implied discrete distribution when the tick size of the considered asset is δ will be given by

$$f(\Delta_i|\mathcal{F}_{i-1}) = F_\epsilon\left(\frac{\Delta_i+\delta}{\psi_i}\right) - F_\epsilon\left(\frac{\Delta_i}{\psi_i}\right).$$

Consequently, giving a time series of price fluctuations $\Delta = (\Delta_1, \ldots \Delta_N)$, the parameters of the ACD(p,q) model can be estimated by maximising the log-likelihood computed from the implied discrete distribution, or, equivalently, by maximising the conditional log-likelihood function

$$\ln \mathcal{L}(\Delta;\theta) = \sum_{i=p+1}^{N} \left(F_\epsilon\left(\frac{\Delta_i+\delta}{\psi_i}\right) - F_\epsilon\left(\frac{\Delta_i}{\psi_i}\right)\right),$$

where ψ_i are computed from the observations $\Delta_1, \ldots, \Delta_N$ with $\psi_i = \hat{\Delta}$ for all $i < 0$ when $\hat{\Delta}$ is a sample mean of $(\Delta_1, \ldots, \Delta_N)$. This log-likelihood function can then be optimised using numerical optimisation methods.

5.6.2 Extensions of the ACD framework

Although the results obtained in the previous section indicates that the linear parameterisation performs quite well, as it passes the independence test in most of the cases, this parameterisation might not be the best model for describing the price fluctuations dataset. To further investigate this issue, this section briefly discusses several extensions to the standard ACD model which could improve the fitting to the price fluctuation dataset. In particular, we consider extensions in three directions. The first direction extends the basic ACD model by allowing both additive and multiplicative stochastic components, where lagged innovations enter the conditional mean function both additively and multiplicatively, while the second approach generalises the basic ACD model by applying the conditional mean function on the transformation of price fluctuations rather than on the plain price fluctuation. Lastly, the third approach generalises this by introducing an asymmetric response to positive and negative shocks into the conditional mean equation.

Additive and multiplicative ACD (AMACD) model

The first extension considered here is an additive and multiplicative ACD model proposed by Hautsch [45]. This specification is a generalisation of the basic ACD model specified by Equation (5.66) and an Additive ACD (AACD) model whose conditional mean function is specified by

$$\psi_i = \omega + \sum_{j=1}^{p} \lambda_j \epsilon_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j}.$$

Unlike the basic ACD model, the innovation, ϵ_i, enters the conditional mean function in an additive form without any interaction with previous conditional mean values ψ_i. Consequently, a more general specification that encompasses these two models can be specified by

$$\begin{aligned} \psi_i &= \omega + \sum_{j=1}^{p} (\alpha_j \psi_{i-j} + \lambda_j) \epsilon_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j}, \\ &= \omega + \sum_{j=1}^{p} \alpha_j \Delta_{i-j} + \sum_{j=1}^{p} \lambda_j \epsilon_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j}. \end{aligned} \tag{5.67}$$

In this specification, the lagged innovation enters the conditional mean function both additively and multiplicatively. In this sense, this model is more flexible and nest the basic ACD model when λ_i = 0, and the additive ACD model when $\alpha_i = 0$.

Logarithmic ACD (LACD) model

In the basic ACD(p, q) model, described in Equation (5.66), sufficient conditions on model parameters are required to ensure the positivity of the conditional mean function. Since these conditions might be violated when we added explanatory variables, which have negative effects into the conditional mean function, Bauwens and Giot [7] introduce a more flexible model in which the autoregressive equation is specified based on the logarithmic transformation of the conditional mean value, ψ_i. Specifically, Bauwens and Giot propose two parameterisations of the conditional mean function which can be characterised by

$$\begin{aligned} \ln \psi_i &= \omega + \sum_{j=1}^{p} \alpha_j \ln \Delta_{i-j} + \sum_{j=1}^{q} \beta_j \ln \psi_{i-j} \\ &= \omega + \sum_{j=1}^{p} \alpha_j \ln \epsilon_{i-j} + \sum_{j=1}^{q} (\beta_j - \alpha_j) \ln \psi_{i-j}, \end{aligned}$$

and

$$\ln \psi_i = \omega + \sum_{j=1}^{p} \lambda_j \epsilon_{i-j} + \sum_{j=1}^{q} \beta_j \ln \psi_{i-j}. \tag{5.68}$$

Unlike the standard ACD(p, q) model, no non-negativity restrictions on the parameters of the autoregressive equation are needed to ensure the positivity of these two conditional mean function. Since the innovation in our situations can be zero, the first parameterisation might not be a good candidate for modelling price fluctuation time series. Consequently, we will utilise only the second parameterisation to model the price fluctuation time series in the rest of this study.

Logarithmic additive and multiplicative ACD (LAMACD) model

Since the lagged innovations in the LACD model discussed in the previous section enter the conditional mean function only in an additive form, we can extend this model by introducing a multiplicative component into the conditional mean function, resulting in a conditional mean function of the form

$$\ln \psi_i = \omega + \sum_{j=1}^{p} (\alpha_j \ln \psi_{i-j} + \lambda_j)\epsilon_{i-j} + \sum_{j=1}^{q} \beta_j \ln \psi_{i-j}. \tag{5.69}$$

Consequently, in this specification, the lagged innovation enters the conditional mean function both additively and multiplicatively. In this sense, this model is more flexible than the LACD model since it nests the LACD model when $\alpha_j = 0$ for all $1 \leq j \leq p$.

Box-Cox ACD (BACD) model

Another specification that generalises the basic ACD model, by using a transformation, could be obtained by applying a Box-Cox transformation as suggested by Hentschel [47], Hautsch [45], and Fernandes and Gramming [31], giving way to

$$\frac{\psi_i^{\delta_1} - 1}{\delta_1} = \tilde{\omega} + \sum_{j=1}^{p} \tilde{\lambda}_j \frac{\epsilon_{i-j}^{\delta_2} - 1}{\delta_2} + \sum_{j=1}^{q} \beta_j \frac{\psi_{i-j}^{\delta_1} - 1}{\delta_1},$$

where δ_1 and δ_2 are parameters of the Box-Cox transformation, and the Box-Cox transformation will be concave when $\delta_i \leq 1$ and convex when $\delta_i \geq 1$. The BACD model can then be ensured by rewriting the above equation as

$$\psi_i^{\delta_1} = \omega + \sum_{j=1}^{p} \lambda_j \epsilon_{i-j}^{\delta_2} + \sum_{j=1}^{q} \beta_j \psi_{i-j}^{\delta_1}, \tag{5.70}$$

where $\omega = 1 + \delta_1 \tilde{\omega} - \sum_{j=1}^{p} \delta_1 \tilde{\lambda}_j / \delta_2 - \sum_{j=1}^{q} \beta_j$, and $\lambda_j = \delta_1 \tilde{\lambda_j} / \delta_2$. Consequently, this specification allows for concave, convex as well as linear conditional mean functions. It nests the ACD model when $\delta_1 = \delta_2 = 1$, the LACD model when $\delta_1 \to 0$ and $\delta_2 = 1$. When $\delta_1 \to 1$, it coincides with a Box-Cox ACD specification proposed by Dufour and Engle [24]. Note that although the LACD is a encompassed by the BACD model, the Box-Cox transformation does not necessarily guarantee the non-negativity of the conditional mean function. In fact, the non-negativity of the conditional mean function is guaranteed only at some value of δ_1, and, thus, can cause a problem when we optimise the parameters of this model using local search techniques as the process might generate a negative conditional mean value when δ_1 changes value. To solve this problem, we will optimise this model only in the region that guarantees the non-negativity of the conditional mean function for all values of δ_1.

Box-Cox additive and multiplicative ACD (BAMACD) model

Similar to the LACD model, the BACD model discussed in the previous section allows the innovations to enter the conditional mean function only in an additive form. Accordingly, we can generalise the BACD model by adding multiplicative components into the conditional mean function so that the conditional mean function becomes

$$\psi_i^{\delta_1} = \omega + \sum_{j=1}^{p}(\alpha_j \psi_{i-j}^{\delta_1} + \lambda_j)\epsilon_{i-j}^{\delta_2} + \sum_{j=1}^{q}\beta_j \psi_{i-j}^{\delta_1}. \tag{5.71}$$

With this specification, the lagged information enters the conditional mean function both additively and multiplicatively, and the BACD model can be obtained by setting $\alpha_j = 0$ for all $1 \leq j \leq p$.

EXponential ACD (EXACD) model

Instead of applying a transformation to the conditional mean value, we can also generalise the basic ACD model by accounting for the asymmetry between the impact of large and small innovations. To achieve this, Dufour and Engle [24] introduce the so called EXponential ACD model that captures features of the EGARCH specification proposed by Nelson [70] by utilising a pair-wise linear conditional mean function. This model allows the conditional mean function that is kinked at $\epsilon_{i-j} = 1$ and is characterised by

$$\ln \psi_i = \omega + \sum_{j=1}^{p}\left(\lambda_j(\epsilon_{i-j} - 1) + c_j|\epsilon_{i-j} - 1|\right) + \sum_{j=1}^{q}\beta_j \ln \psi_{i-j}. \tag{5.72}$$

Hence, for observations less than the conditional mean ($\epsilon_{i-j} < 1$), the impact from these observations will be $(\alpha_j - c_j)\epsilon_{i-j}$, while for observations larger than the conditional mean ($\epsilon_{i-j} > 1$), the impact will be $(\alpha_j + c_j)\epsilon_{i-j}$ respectively. Accordingly, this model is more flexible and nests the LACD model when $c_j = 0$ for all $1 \leq j \leq p$.

Augmented logarithmic additive and multiplicative ACD (ALAMACD) model

The EXACD model discussed in the previous section can be extended in two directions. Firstly, since the EXACD model allows the lagged innovations to enter the conditional mean function only in an additive form, we can extend this model by adding the multiplicative component into the conditional mean function like we did for the LAMACD and BAMACD models. Secondly, while the conditional mean function of the EXACD model is kinked at $\epsilon_{i-j} = 1$, a valuable generalisation is to parameterise the position of the kink. Using the parameterisation for modelling an asymmetric GARCH process suggested by Hentschel [47], the specification that we call the ALAMACD model can be specified by

$$\ln \psi_i = \omega + \sum_{j=1}^{p}\alpha_j \ln \psi_{i-j}\left(|\epsilon_{i-j} - b| + c_j(\epsilon_{i-j} - b)\right) + \sum_{j=1}^{p}\lambda_j\left(|\epsilon_{i-j} - b| + c_j(\epsilon_{i-j} - b)\right) + \sum_{j=1}^{q}\beta_j \ln \psi_{i-j}. \tag{5.73}$$

In this specification, the parameter b gives the position of the kink. It nests the EXACD model when $b = 1$, and $\alpha_j = 0$ for all $1 \leq j \leq p$, and encompasses the LAMACD model when $b = 0$ and $c_j = 0$ for all $1 \leq j \leq p$.

Augmented Box-Cox additive and multiplicative ACD (ABAMACD) model

The last extension considered in this section is the augmented Box-Cox additive and multiplicative ACD model, originally proposed as Augmented ACD model in [45]. This specification is constructed by utilising the Box-Cox transformation instead of logarithmic transformation as in the ALAMACD model. Accordingly, the conditional mean function of the ABAMACD model can be specified by

$$\psi_i^{\delta_1} = \omega + \sum_{j=1}^{p} \alpha_j \psi_{i-j}^{\delta_1} \left(|\epsilon_{i-j} - b| + c_j(\epsilon_{i-j} - b)\right)^{\delta_2} + \sum_{j=1}^{p} \lambda_j \left(|\epsilon_{i-j} - b| + c_j(\epsilon_{i-j} - b)\right)^{\delta_2} + \sum_{j=1}^{q} \beta_j \psi_{i-j}^{\delta_1}. \quad (5.74)$$

This specification nests all specifications outlined above. Particularly, it encompasses all specifications nested by the ALAMACD model, as well as all other models based on additive and multiplicative stochastic components as it nests the AMACD model for $\delta_1 = \delta_2 = 1$, $b = c_j = 0$ and the BAMACD model for $b = c_j = 0$ for all $1 \leq j \leq p$. Even though this specification allows for more flexibility, it has one major drawback since the parameter restriction $|c_j| \leq 1$ has to be imposed in order to circumvent a complex value whenever $\delta_2 \neq 1$.

5.6.3 Experimental results

To compare the performance of the basic ACD model and its eight extensions discussed in Section 5.6.2, this section analyses the results obtained from fitting these models to the buy price fluctuations dataset by maximising the likelihood of the implied discrete distribution. For each instrument and for each time frame of buy price fluctuation time series, we compute the maximum likelihood, the Brier's score, Epstein's score and the $\{\epsilon_i\}$ series implied by each model. We specify the autoregressive model with three lags as discussed in Section 5.5.4. The evaluation will be done using both in-sample and out-of-sample results by using the first 80% of the sample as a training dataset.

Table 5.9 reports the in-sample and out-of-sample log-likelihood, Brier's score, Epstein's score and the p-value of the Ljung-Box Q-statistic for ϵ and ϵ^2 using the first 10 autocorrelations for each of the nine models in both training and testing dataset. The results indicate that among all models considered, the ABAMACD seems to be the best performing model, according to in-sample log-likelihood and Epstein's score, as it was the maximum log-likelihood and minimum Epstien's score in 13 out of the 18 cases considered. When the ABAMACD is not the best model, the ALAMACD is the model with the best performance. This result is not surprising since the ABAMACD is the most general model that encompasses all models considered here, while ALAMACD is the most general model that encompasses all models that utilise a logarithmic transform. Although, theoretically, the ABAMACD contains the ALAMACD as its special cases, the reason why the ALAMACD beat the ABAMACD in some cases is simply because the best model obtained from ALAMACD model has negative conditional mean (before applying the transformation), which we are not allowed to have in the ABAMACD model as discussed in the previous section. However, when we consider the out-of-sample results, no clear winning model can be identified, as the best model varies from case to case. To simplify the analysis, we calculated the average rank score of each model by taking the arithmetic mean of the rank in each model. The result reported in Table 5.10 indicates that the ABAMACD is the best performing model in both training and testing dataset according to all measures. Additionally, we observe the strongest increase of the log-likelihood function when the model is extended to include an additive stochastic component, i.e. when the ACD model is extended to the AMACD model, the BACD model is extended to the BAMACD

Table 5.9: In-sample result and out-of-sample result obtained from the maximum log-likelihood estimate of several non-linear ACD models. The statistics reported include the maximum log-likelihood, Brier's score, Epstein's score and the p-value of Ljung-Box Q statistic based on the first 10 autocorrelations of the residual and squared residuals sequence.

		Training dataset					Testing dataset				
Model	T	LL	Bs	Es	$Q(10)$	$Q^2(10)$	LL	Bs	Es	$Q(10)$	$Q^2(10)$
GE buy price fluctuations, NYSE											
ACD	60	-15187 (4)	0.9468 (4)	4.579 (4)	99.99	100.00	-4577 (2)	0.9694 (2)	9.681 (4)	13.55	1.81
LACD	60	-15264 (9)	0.9497 (9)	4.647 (9)	98.43	100.00	-4605 (8)	0.9703 (9)	9.865 (6)	0.00	0.00
BACD	60	-15231 (5)	0.9472 (5)	4.608 (5)	99.98	100.00	-4605 (9)	0.9701 (8)	9.869 (7)	0.39	0.00
AMACD	60	-15183 (3)	0.9468 (2)	4.571 (3)	100.00	100.00	-4576 (1)	0.9694 (1)	9.677 (2)	14.53	3.43
LAMACD	60	-15240 (7)	0.9485 (7)	4.609 (6)	95.52	100.00	-4594 (6)	0.9697 (5)	9.889 (8)	0.08	0.00
BAMACD	60	-15182 (2)	0.9468 (3)	4.569 (2)	100.00	100.00	-4578 (3)	0.9694 (3)	9.681 (3)	11.82	1.39
EXACD	60	-15247 (8)	0.9481 (6)	4.623 (8)	99.55	100.00	-4602 (7)	0.9700 (7)	9.822 (5)	0.00	0.00
ALAMACD	60	-15240 (6)	0.9485 (8)	4.609 (7)	94.94	100.00	-4589 (5)	0.9697 (6)	9.894 (9)	1.67	0.23
ABAMACD	60	-15176 (1)	0.9467 (1)	4.562 (1)	100.00	100.00	-4580 (4)	0.9695 (4)	9.675 (1)	13.79	0.96
ACD	30	-25779 (6)	0.9297 (7)	3.346 (6)	94.63	100.00	-7815 (1)	0.9602 (5)	6.805 (3)	73.10	58.32
LACD	30	-25864 (9)	0.9325 (9)	3.378 (9)	98.25	100.00	-7856 (9)	0.9616 (9)	6.901 (9)	2.08	7.41
BACD	30	-25765 (5)	0.9286 (4)	3.345 (4)	94.06	100.00	-7848 (7)	0.9604 (6)	6.890 (8)	7.68	8.36
AMACD	30	-25762 (4)	0.9290 (5)	3.345 (5)	81.15	100.00	-7817 (3)	0.9600 (4)	6.803 (2)	35.58	52.00
LAMACD	30	-25801 (8)	0.9307 (8)	3.351 (7)	94.33	100.00	-7825 (6)	0.9607 (8)	6.879 (7)	63.81	44.48
BAMACD	30	-25752 (2)	0.9284 (2)	3.338 (2)	71.71	100.00	-7817 (2)	0.9600 (3)	6.800 (1)	58.13	59.39
EXACD	30	-25799 (7)	0.9291 (6)	3.357 (8)	91.60	100.00	-7850 (8)	0.9605 (7)	6.860 (6)	1.22	0.21
ALAMACD	30	-25762 (3)	0.9284 (3)	3.345 (3)	80.51	100.00	-7824 (5)	0.9599 (1)	6.830 (5)	5.34	37.46
ABAMACD	30	-25745 (1)	0.9282 (1)	3.337 (1)	75.35	100.00	-7820 (4)	0.9599 (2)	6.807 (4)	42.73	56.10
ACD	10	-64256 (5)	0.8851 (7)	1.908 (5)	98.79	100.00	-19895 (4)	0.9314 (5)	3.709 (4)	22.21	49.68
LACD	10	-64576 (8)	0.8865 (9)	1.926 (8)	4.19	100.00	-20120 (9)	0.9330 (8)	3.814 (9)	4.26	68.79
BACD	10	-64173 (4)	0.8843 (3)	1.905 (4)	97.14	100.00	-19956 (7)	0.9317 (6)	3.741 (7)	0.00	0.06
AMACD	10	-64161 (3)	0.8844 (4)	1.904 (3)	97.84	100.00	-19881 (3)	0.9309 (3)	3.705 (3)	2.83	44.05
LAMACD	10	-64576 (9)	0.8865 (8)	1.926 (9)	4.19	100.00	-19995 (8)	0.9359 (9)	3.802 (8)	0.34	12.84
BAMACD	10	-64144 (2)	0.8841 (2)	1.902 (2)	93.83	100.00	-19880 (2)	0.9309 (2)	3.703 (2)	0.50	9.12
EXACD	10	-64338 (7)	0.8850 (6)	1.915 (7)	26.93	100.00	-19956 (6)	0.9320 (7)	3.736 (6)	0.00	0.00
ALAMACD	10	-64299 (6)	0.8848 (5)	1.912 (6)	6.77	100.00	-19901 (5)	0.9312 (4)	3.714 (5)	0.04	0.00
ABAMACD	10	-64144 (1)	0.8841 (1)	1.902 (1)	93.83	100.00	-19878 (1)	0.9309 (1)	3.703 (1)	0.50	9.12
IBM buy price fluctuations, NYSE											
ACD	60	-27556 (7)	0.9813 (1)	17.894 (5)	0.02	21.55	-6754 (5)	0.9803 (5)	14.862 (6)	58.76	99.13
LACD	60	-27560 (9)	0.9813 (2)	17.909 (8)	0.01	22.77	-6754 (6)	0.9804 (7)	14.863 (7)	78.36	99.30
BACD	60	-27560 (8)	0.9813 (4)	17.923 (9)	0.02	55.14	-6753 (3)	0.9803 (6)	14.849 (4)	88.31	98.81
AMACD	60	-27550 (5)	0.9813 (3)	17.880 (2)	0.02	30.05	-6754 (8)	0.9803 (4)	14.862 (5)	57.63	99.30
LAMACD	60	-27555 (6)	0.9814 (5)	17.897 (6)	0.03	15.48	-6754 (4)	0.9804 (8)	14.865 (9)	65.39	98.99
BAMACD	60	-27543 (3)	0.9815 (9)	17.899 (7)	0.15	61.62	-6755 (9)	0.9804 (9)	14.865 (8)	66.46	98.34
EXACD	60	-27543 (4)	0.9814 (7)	17.889 (4)	0.39	88.62	-6752 (2)	0.9803 (2)	14.819 (1)	90.75	96.90
ALAMACD	60	-27534 (1)	0.9814 (6)	17.867 (1)	0.93	50.52	-6752 (1)	0.9802 (1)	14.819 (2)	86.10	93.42
ABAMACD	60	-27537 (2)	0.9814 (8)	17.884 (3)	0.07	10.26	-6754 (7)	0.9803 (3)	14.848 (3)	80.68	97.85
ACD	30	-47753 (6)	0.9753 (7)	12.777 (6)	89.52	99.97	-11668 (7)	0.9754 (7)	10.579 (7)	78.57	98.08
LACD	30	-47755 (8)	0.9755 (9)	12.791 (9)	75.21	99.95	-11686 (8)	0.9757 (8)	10.653 (9)	55.68	98.12
BACD	30	-47764 (9)	0.9751 (3)	12.782 (7)	87.16	100.00	-11668 (6)	0.9752 (5)	10.577 (6)	75.09	98.15
AMACD	30	-47727 (4)	0.9751 (4)	12.761 (4)	92.75	100.00	-11666 (5)	0.9752 (6)	10.570 (4)	60.33	97.62
LAMACD	30	-47755 (7)	0.9755 (8)	12.791 (8)	75.21	99.95	-11686 (9)	0.9757 (9)	10.653 (8)	55.68	98.12
BAMACD	30	-47726 (3)	0.9749 (1)	12.753 (3)	91.23	100.00	-11664 (1)	0.9751 (2)	10.563 (2)	62.50	98.41
EXACD	30	-47733 (5)	0.9753 (6)	12.775 (5)	83.38	100.00	-11664 (3)	0.9751 (4)	10.558 (1)	67.89	97.85
ALAMACD	30	-47702 (1)	0.9752 (5)	12.750 (1)	97.50	99.99	-11664 (2)	0.9751 (1)	10.573 (5)	48.61	97.41
ABAMACD	30	-47725 (2)	0.9749 (2)	12.752 (2)	91.79	100.00	-11665 (4)	0.9751 (3)	10.564 (3)	60.94	98.39
ACD	10	-124224 (7)	0.9556 (6)	7.272 (6)	93.11	100.00	-30234 (9)	0.9571 (7)	5.933 (8)	58.50	99.99
LACD	10	-124319 (9)	0.9571 (9)	7.298 (9)	97.77	100.00	-30224 (6)	0.9580 (9)	5.923 (7)	64.77	99.50
BACD	10	-123994 (4)	0.9547 (3)	7.237 (3)	91.26	100.00	-30185 (4)	0.9561 (3)	5.908 (5)	30.43	100.00
AMACD	10	-123992 (3)	0.9549 (4)	7.241 (4)	99.97	100.00	-30174 (3)	0.9563 (4)	5.896 (3)	26.08	100.00
LAMACD	10	-124319 (8)	0.9571 (8)	7.298 (8)	97.77	100.00	-30224 (7)	0.9580 (8)	5.923 (6)	64.77	99.50
BAMACD	10	-123931 (2)	0.9546 (2)	7.225 (2)	97.45	100.00	-30166 (2)	0.9560 (2)	5.893 (1)	14.52	99.99
EXACD	10	-124171 (6)	0.9558 (7)	7.275 (7)	99.59	100.00	-30232 (8)	0.9571 (6)	5.936 (9)	84.61	99.99
ALAMACD	10	-124033 (5)	0.9555 (5)	7.246 (5)	59.12	100.00	-30187 (5)	0.9566 (5)	5.905 (4)	74.74	99.99
ABAMACD	10	-123916 (1)	0.9545 (1)	7.222 (1)	92.77	100.00	-30161 (1)	0.9559 (1)	5.893 (2)	10.94	100.00

Table 5.9 continued: In-sample result and out-of-sample result obtained from the maximum log-likelihood estimate of several non-linear ACD models. The statistics reported include the maximum log-likelihood, Brier's score, Epstein's score and the p-value of Ljung-Box Q statistic based on the first 10 autocorrelations of the residual and squared residuals sequence.

		Training dataset					Testing dataset				
Model	T	LL	Bs	Es	$Q(10)$	$Q^2(10)$	LL	Bs	Es	$Q(10)$	$Q^2(10)$
Microsoft buy price fluctuations, NYSE											
ACD	60	-20810 (6)	0.9524 (1)	5.226 (5)	32.23	100.00	-4982 (5)	0.9466 (1)	4.224 (3)	83.30	99.95
LACD	60	-20817 (8)	0.9526 (6)	5.238 (8)	31.57	100.00	-4984 (7)	0.9468 (4)	4.234 (8)	74.60	99.94
BACD	60	-20825 (9)	0.9527 (8)	5.242 (9)	0.19	100.00	-4986 (9)	0.9470 (7)	4.229 (6)	49.43	99.95
AMACD	60	-20798 (4)	0.9524 (2)	5.223 (4)	12.31	100.00	-4980 (4)	0.9467 (2)	4.228 (4)	66.96	99.89
LAMACD	60	-20804 (5)	0.9526 (7)	5.230 (6)	14.81	100.00	-4982 (6)	0.9469 (6)	4.236 (9)	60.49	99.91
BAMACD	60	-20796 (3)	0.9525 (4)	5.221 (3)	12.97	100.00	-4980 (3)	0.9468 (5)	4.228 (5)	71.50	99.89
EXACD	60	-20812 (7)	0.9525 (5)	5.235 (7)	33.88	100.00	-4985 (8)	0.9467 (3)	4.231 (7)	83.13	99.98
ALAMACD	60	-20792 (2)	0.9529 (9)	5.220 (2)	11.26	100.00	-4978 (2)	0.9471 (9)	4.218 (2)	79.21	99.92
ABAMACD	60	-20790 (1)	0.9525 (3)	5.216 (1)	12.20	100.00	-4977 (1)	0.9471 (8)	4.211 (1)	68.44	99.91
ACD	30	-35136 (5)	0.9354 (6)	3.681 (4)	29.69	100.00	-8444 (4)	0.9293 (6)	3.161 (4)	68.43	99.61
LACD	30	-35153 (9)	0.9358 (8)	3.692 (9)	20.22	100.00	-8447 (9)	0.9295 (7)	3.168 (9)	80.91	99.75
BACD	30	-35138 (7)	0.9349 (3)	3.684 (5)	30.40	100.00	-8445 (6)	0.9289 (3)	3.165 (7)	70.15	99.84
AMACD	30	-35120 (3)	0.9351 (4)	3.680 (3)	30.75	100.00	-8443 (3)	0.9291 (5)	3.162 (5)	62.36	99.67
LAMACD	30	-35136 (6)	0.9358 (7)	3.687 (6)	23.17	100.00	-8446 (8)	0.9296 (8)	3.167 (8)	75.49	99.76
BAMACD	30	-35116 (2)	0.9346 (1)	3.678 (2)	28.99	100.00	-8440 (2)	0.9286 (1)	3.159 (2)	65.61	99.85
EXACD	30	-35141 (8)	0.9353 (5)	3.689 (8)	20.81	100.00	-8445 (5)	0.9290 (4)	3.161 (3)	80.93	99.87
ALAMACD	30	-35135 (4)	0.9360 (9)	3.688 (7)	17.54	100.00	-8445 (7)	0.9297 (9)	3.165 (6)	78.11	99.63
ABAMACD	30	-35115 (1)	0.9348 (2)	3.678 (1)	28.28	100.00	-8440 (1)	0.9287 (2)	3.158 (1)	66.47	99.84
ACD	10	-87717 (5)	0.8935 (7)	2.107 (5)	13.47	100.00	-21010 (9)	0.8860 (6)	1.802 (6)	72.29	99.98
LACD	10	-87772 (8)	0.8938 (8)	2.112 (8)	10.72	100.00	-21009 (7)	0.8861 (8)	1.802 (8)	64.61	99.98
BACD	10	-87645 (4)	0.8926 (4)	2.104 (4)	28.33	100.00	-20998 (4)	0.8852 (1)	1.800 (4)	68.77	100.00
AMACD	10	-87608 (3)	0.8925 (2)	2.103 (3)	42.58	100.00	-20994 (3)	0.8853 (3)	1.800 (3)	65.20	99.99
LAMACD	10	-87772 (9)	0.8938 (9)	2.112 (9)	10.72	100.00	-21009 (6)	0.8861 (7)	1.802 (7)	64.61	99.98
BAMACD	10	-87601 (2)	0.8926 (3)	2.101 (2)	14.90	100.00	-20991 (1)	0.8853 (2)	1.799 (1)	73.39	100.00
EXACD	10	-87733 (6)	0.8934 (5)	2.110 (6)	3.89	100.00	-21010 (8)	0.8857 (5)	1.801 (5)	48.71	99.96
ALAMACD	10	-87733 (7)	0.8934 (6)	2.110 (7)	3.89	100.00	-21007 (5)	0.8861 (9)	1.803 (9)	90.42	99.99
ABAMACD	10	-87597 (1)	0.8925 (1)	2.101 (1)	21.24	100.00	-20994 (2)	0.8853 (4)	1.799 (2)	72.32	100.00
Gold future buy price fluctuations, MCX											
ACD	30	-6224 (8)	0.9752 (7)	11.068 (6)	95.87	60.20	-1664 (5)	0.9744 (6)	15.312 (6)	61.62	63.91
LACD	30	-6227 (9)	0.9753 (8)	11.110 (9)	97.93	70.67	-1674 (9)	0.9749 (9)	15.578 (8)	4.58	1.84
BACD	30	-6221 (6)	0.9751 (4)	11.080 (8)	93.76	72.12	-1663 (4)	0.9744 (4)	15.295 (4)	68.34	87.33
AMACD	30	-6219 (3)	0.9750 (1)	11.046 (2)	62.14	55.49	-1663 (3)	0.9744 (5)	15.276 (3)	66.99	78.20
LAMACD	30	-6220 (5)	0.9754 (9)	11.068 (7)	99.31	79.54	-1667 (7)	0.9744 (3)	15.380 (7)	53.80	25.89
BAMACD	30	-6219 (4)	0.9750 (2)	11.046 (3)	62.14	55.49	-1664 (6)	0.9744 (7)	15.311 (5)	68.84	86.01
EXACD	30	-6222 (7)	0.9752 (6)	11.067 (5)	94.71	72.18	-1662 (2)	0.9744 (2)	15.267 (2)	68.33	83.78
ALAMACD	30	-6218 (1)	0.9751 (3)	11.037 (1)	85.28	63.66	-1671 (8)	0.9749 (8)	15.652 (9)	87.63	95.53
ABAMACD	30	-6219 (2)	0.9752 (5)	11.064 (4)	96.17	87.93	-1662 (1)	0.9743 (1)	15.241 (1)	58.14	57.14
ACD	10	-15726 (9)	0.9599 (9)	6.401 (9)	88.33	76.80	-4246 (5)	0.9658 (8)	8.475 (3)	42.11	97.02
LACD	10	-15718 (8)	0.9597 (5)	6.389 (8)	10.94	9.75	-4243 (2)	0.9658 (2)	8.470 (2)	43.43	95.39
BACD	10	-15715 (6)	0.9598 (8)	6.383 (6)	89.62	84.56	-4241 (1)	0.9658 (6)	8.457 (1)	61.14	97.07
AMACD	10	-15716 (7)	0.9598 (7)	6.383 (7)	75.53	85.00	-4250 (8)	0.9658 (3)	8.514 (9)	45.98	96.06
LAMACD	10	-15714 (3)	0.9596 (2)	6.377 (4)	54.48	42.76	-4247 (6)	0.9658 (5)	8.507 (7)	54.63	95.69
BAMACD	10	-15714 (4)	0.9596 (4)	6.377 (2)	54.46	42.61	-4252 (9)	0.9659 (9)	8.508 (8)	35.02	95.93
EXACD	10	-15715 (5)	0.9595 (1)	6.380 (5)	10.55	15.84	-4244 (3)	0.9658 (7)	8.476 (5)	32.89	89.99
ALAMACD	10	-15714 (2)	0.9596 (3)	6.377 (3)	54.48	42.75	-4249 (7)	0.9658 (4)	8.484 (6)	20.41	93.79
ABAMACD	10	-15709 (1)	0.9597 (6)	6.376 (1)	59.90	94.42	-4246 (4)	0.9656 (1)	8.476 (4)	43.35	90.49
ACD	5	-28221 (9)	0.9437 (9)	4.503 (9)	44.45	53.23	-7603 (1)	0.9541 (9)	5.702 (1)	85.58	96.52
LACD	5	-28218 (8)	0.9437 (7)	4.503 (8)	22.94	58.53	-7604 (3)	0.9540 (6)	5.704 (3)	92.68	96.44
BACD	5	-28218 (7)	0.9437 (6)	4.503 (7)	22.94	58.52	-7604 (2)	0.9540 (5)	5.704 (2)	92.68	96.44
AMACD	5	-28204 (3)	0.9434 (3)	4.496 (3)	54.83	43.62	-7620 (9)	0.9540 (4)	5.727 (8)	24.34	88.68
LAMACD	5	-28205 (5)	0.9434 (1)	4.497 (5)	31.92	31.71	-7618 (6)	0.9540 (1)	5.728 (9)	66.56	97.41
BAMACD	5	-28201 (2)	0.9434 (5)	4.494 (2)	61.20	73.08	-7616 (5)	0.9540 (7)	5.719 (5)	33.79	90.30
EXACD	5	-28217 (6)	0.9437 (8)	4.503 (6)	22.62	70.77	-7605 (4)	0.9541 (8)	5.706 (4)	93.40	95.16
ALAMACD	5	-28205 (4)	0.9434 (4)	4.497 (4)	28.97	47.46	-7618 (7)	0.9540 (3)	5.727 (7)	71.78	94.91
ABAMACD	5	-28198 (1)	0.9434 (2)	4.494 (1)	50.13	56.05	-7618 (8)	0.9540 (2)	5.723 (6)	31.19	86.25

Table 5.9 continued: In-sample result and out-of-sample result obtained from the maximum log-likelihood estimate of several non-linear ACD models. The statistics reported include the maximum log-likelihood, Brier's score, Epstein's score and the p-value of Ljung-Box Q statistic based on the first 10 autocorrelations of the residual and squared residuals sequence.

		Training dataset					Testing dataset				
Model	T	LL	Bs	Es	$Q(10)$	$Q^2(10)$	LL	Bs	Es	$Q(10)$	$Q^2(10)$
Silver future buy price fluctuations, MCX											
ACD	30	-7095 (9)	0.9853 (8)	18.889 (9)	93.92	96.21	-1957 (5)	0.9840 (6)	31.164 (6)	97.87	68.27
LACD	30	-7079 (2)	0.9850 (5)	18.687 (2)	78.18	71.81	-1961 (7)	0.9840 (5)	31.086 (3)	91.19	68.44
BACD	30	-7079 (7)	0.9850 (2)	18.687 (7)	78.68	72.89	-1954 (2)	0.9835 (1)	30.898 (2)	97.08	77.46
AMACD	30	-7090 (8)	0.9854 (9)	18.865 (8)	69.49	80.86	-1959 (6)	0.9841 (8)	31.231 (8)	96.37	61.52
LAMACD	30	-7079 (3)	0.9850 (4)	18.687 (3)	78.18	71.81	-1984 (9)	0.9845 (9)	32.386 (9)	97.00	94.08
BAMACD	30	-7079 (6)	0.9850 (3)	18.687 (6)	78.68	72.89	-1950 (1)	0.9838 (3)	30.790 (1)	99.89	96.45
EXACD	30	-7079 (4)	0.9850 (6)	18.687 (4)	78.18	71.81	-1962 (8)	0.9838 (4)	31.204 (7)	80.88	50.38
ALAMACD	30	-7079 (5)	0.9850 (7)	18.687 (5)	78.18	71.81	-1955 (3)	0.9838 (2)	31.114 (5)	99.76	90.80
ABAMACD	30	-7078 (1)	0.9850 (1)	18.671 (1)	78.68	72.89	-1956 (4)	0.9840 (7)	31.096 (4)	98.91	91.78
ACD	10	-18472 (9)	0.9768 (7)	12.212 (9)	54.71	9.73	-5018 (1)	0.9817 (1)	17.085 (1)	94.56	100.00
LACD	10	-18471 (8)	0.9768 (5)	12.210 (8)	56.75	17.55	-5021 (2)	0.9818 (2)	17.146 (2)	96.68	100.00
BACD	10	-18463 (4)	0.9768 (9)	12.198 (3)	65.65	88.50	-5030 (4)	0.9820 (3)	17.223 (4)	32.03	100.00
AMACD	10	-18465 (7)	0.9767 (2)	12.204 (6)	44.58	39.25	-5034 (5)	0.9821 (6)	17.287 (5)	84.00	100.00
LAMACD	10	-18465 (6)	0.9767 (1)	12.204 (4)	55.44	60.84	-5040 (8)	0.9823 (8)	17.387 (9)	4.37	100.00
BAMACD	10	-18459 (3)	0.9767 (4)	12.197 (2)	69.72	78.00	-5036 (6)	0.9821 (5)	17.314 (6)	89.25	100.00
EXACD	10	-18465 (5)	0.9768 (8)	12.206 (7)	69.34	41.07	-5023 (3)	0.9821 (4)	17.191 (3)	99.14	100.00
ALAMACD	10	-18459 (2)	0.9767 (3)	12.204 (5)	58.90	60.85	-5038 (7)	0.9824 (9)	17.381 (8)	18.67	100.00
ABAMACD	10	-18455 (1)	0.9768 (6)	12.193 (1)	47.55	43.11	-5040 (9)	0.9822 (7)	17.355 (7)	66.11	100.00
ACD	5	-33689 (7)	0.9655 (5)	9.137 (5)	5.94	0.42	-9108 (1)	0.9757 (1)	11.760 (1)	93.80	100.00
LACD	5	-33697 (9)	0.9655 (2)	9.146 (9)	70.32	9.62	-9130 (9)	0.9768 (8)	11.904 (8)	96.03	100.00
BACD	5	-33696 (8)	0.9655 (4)	9.146 (8)	53.28	11.09	-9128 (7)	0.9764 (7)	11.873 (7)	93.42	100.00
AMACD	5	-33687 (5)	0.9655 (1)	9.132 (1)	8.50	0.31	-9111 (2)	0.9758 (2)	11.771 (2)	91.90	100.00
LAMACD	5	-33689 (6)	0.9655 (3)	9.136 (4)	6.77	0.44	-9116 (4)	0.9764 (6)	11.837 (5)	95.82	100.00
BAMACD	5	-33680 (3)	0.9656 (7)	9.135 (3)	7.01	1.46	-9116 (3)	0.9760 (3)	11.813 (3)	92.88	100.00
EXACD	5	-33681 (4)	0.9656 (8)	9.139 (7)	4.31	6.17	-9119 (5)	0.9763 (5)	11.843 (6)	96.23	100.00
ALAMACD	5	-33677 (2)	0.9655 (6)	9.138 (6)	4.46	7.46	-9129 (8)	0.9770 (9)	11.919 (9)	97.14	100.00
ABAMACD	5	-33675 (1)	0.9656 (9)	9.133 (2)	9.22	5.96	-9128 (6)	0.9761 (4)	11.821 (4)	94.01	100.00
Natural gas future buy price fluctuations, MCX											
ACD	30	-5052 (6)	0.9484 (5)	5.573 (5)	73.75	93.34	-1079 (6)	0.9093 (6)	3.332 (5)	70.32	86.80
LACD	30	-5050 (2)	0.9483 (1)	5.566 (1)	72.81	80.87	-1088 (8)	0.9117 (8)	3.454 (8)	45.80	98.67
BACD	30	-5062 (9)	0.9489 (8)	5.602 (9)	79.76	96.06	-1075 (1)	0.9086 (1)	3.306 (2)	84.00	89.70
AMACD	30	-5052 (5)	0.9484 (6)	5.573 (6)	73.75	93.34	-1075 (2)	0.9086 (2)	3.299 (1)	84.48	88.86
LAMACD	30	-5050 (3)	0.9483 (2)	5.566 (2)	72.81	80.87	-1086 (7)	0.9113 (7)	3.423 (7)	45.80	98.67
BAMACD	30	-5059 (8)	0.9490 (9)	5.591 (8)	97.64	98.38	-1076 (3)	0.9089 (4)	3.311 (3)	61.39	85.82
EXACD	30	-5050 (4)	0.9483 (3)	5.566 (3)	72.81	80.87	-1088 (9)	0.9117 (9)	3.454 (9)	45.80	98.67
ALAMACD	30	-5050 (1)	0.9483 (4)	5.566 (4)	72.83	80.86	-1077 (4)	0.9087 (3)	3.327 (4)	45.80	98.67
ABAMACD	30	-5052 (7)	0.9484 (7)	5.573 (7)	73.75	93.34	-1078 (5)	0.9091 (5)	3.335 (6)	83.33	95.90
ACD	10	-12489 (9)	0.9072 (1)	3.660 (9)	78.20	97.65	-2620 (7)	0.8621 (3)	2.148 (6)	61.46	64.28
LACD	10	-12485 (8)	0.9073 (5)	3.655 (5)	9.11	78.20	-2615 (1)	0.8619 (1)	2.136 (1)	74.12	76.95
BACD	10	-12481 (4)	0.9073 (6)	3.655 (4)	76.83	94.89	-2619 (4)	0.8629 (8)	2.147 (4)	50.58	81.28
AMACD	10	-12482 (7)	0.9072 (2)	3.656 (7)	52.54	93.73	-2622 (9)	0.8637 (9)	2.152 (8)	7.57	71.00
LAMACD	10	-12481 (5)	0.9073 (4)	3.657 (8)	68.38	95.89	-2619 (5)	0.8626 (6)	2.144 (3)	9.77	62.42
BAMACD	10	-12479 (3)	0.9073 (8)	3.655 (6)	73.88	95.44	-2621 (8)	0.8626 (5)	2.151 (7)	12.07	75.29
EXACD	10	-12482 (6)	0.9074 (9)	3.654 (3)	87.40	96.99	-2617 (2)	0.8620 (2)	2.139 (2)	63.62	92.04
ALAMACD	10	-12469 (2)	0.9073 (7)	3.650 (2)	74.86	96.04	-2620 (6)	0.8623 (4)	2.148 (5)	75.94	72.69
ABAMACD	10	-12465 (1)	0.9073 (3)	3.647 (1)	66.95	95.02	-2618 (3)	0.8627 (7)	2.152 (9)	47.70	88.51
ACD	5	-21343 (9)	0.8583 (4)	2.819 (9)	72.47	94.94	-4511 (4)	0.8154 (4)	1.669 (4)	85.02	99.79
LACD	5	-21332 (8)	0.8582 (2)	2.817 (6)	0.76	74.69	-4522 (8)	0.8179 (8)	1.692 (8)	93.94	99.94
BACD	5	-21322 (4)	0.8586 (5)	2.815 (5)	80.13	99.97	-4513 (5)	0.8161 (5)	1.671 (5)	95.46	99.96
AMACD	5	-21330 (6)	0.8582 (1)	2.815 (3)	60.81	98.46	-4510 (2)	0.8153 (3)	1.668 (3)	73.26	99.87
LAMACD	5	-21332 (7)	0.8582 (3)	2.817 (7)	0.76	74.69	-4557 (9)	0.8209 (9)	1.726 (9)	0.06	99.68
BAMACD	5	-21319 (3)	0.8586 (6)	2.813 (2)	75.52	99.93	-4511 (3)	0.8152 (1)	1.666 (2)	80.99	99.90
EXACD	5	-21323 (5)	0.8591 (7)	2.817 (8)	85.71	99.93	-4517 (6)	0.8170 (7)	1.679 (6)	97.62	99.97
ALAMACD	5	-21313 (2)	0.8591 (8)	2.815 (4)	54.30	99.97	-4518 (7)	0.8170 (6)	1.681 (7)	89.00	99.92
ABAMACD	5	-21308 (1)	0.8593 (9)	2.810 (1)	45.67	99.67	-4497 (1)	0.8153 (2)	1.648 (1)	92.00	99.82

Table 5.10: The average ranking of the nine ACD models considered both in the in-sample and out-of-sample dataset.

	Training dataset			Testing dataset		
Model	LL	Bs	Es	LL	Bs	Es
ACD	6.88 (8)	5.71 (7)	6.29 (8)	4.53 (4)	4.94 (5)	4.35 (3)
LACD	7.71 (9)	6.29 (9)	7.47 (9)	6.47 (8)	6.47 (8)	6.29 (8)
BACD	6.24 (7)	4.94 (4)	6.00 (7)	4.71 (5)	4.71 (4)	4.71 (5)
AMACD	4.53 (4)	3.59 (2)	4.18 (4)	4.53 (3)	4.18 (2)	4.41 (4)
LAMACD	5.94 (6)	5.47 (5)	6.00 (6)	6.59 (9)	6.65 (9)	7.41 (9)
BAMACD	3.18 (2)	4.06 (3)	3.35 (2)	3.88 (2)	4.24 (3)	3.71 (2)
EXACD	5.82 (5)	6.00 (8)	5.88 (5)	5.35 (7)	5.06 (6)	4.76 (6)
ALAMACD	3.18 (3)	5.47 (6)	4.06 (3)	5.12 (6)	5.12 (7)	5.88 (7)
ABAMACD	1.53 (1)	3.47 (1)	1.76 (1)	3.82 (1)	3.65 (1)	3.47 (1)

model and the LACD model is extended to the LAMACD model. This result illustrates that for the price fluctuations dataset, it is crucial to account for both additive and multiplicative stochastic components. The p-value of the Lijung-Box Q-statistic, based on the first 10 autocorrelations of the residuals and square residuals reported in Table 5.9, indicates that all models pass this test in most cases except in the case of IBM's 60 minute buy price fluctuations.

5.6.4 Summary

This section compared the performance of several extensions of the ACD model in modelling the price fluctuation time series. These models extend the basic ACD model in three main directions which are i) introducing the additive innovation into the conditional mean function, ii) applying transformation to the conditional mean value, and iii) introducing an asymmetric response to positive and negative shocks into the conditional mean function. In particular, we compare the performance of nine ACD models including the linear ACD model, the BACD model, the LACD model, the AMACD model, the LAMACD model, the BAMACD model, the EXACD model, the ALAMACD model and the ABAMACD model. The experimentation results indicated that the ABAMACD model, which encompasses all other models as its special case, is the best performing model in both training and testing dataset according to all performance measures. Among all extensions considered, we find the strongest increase of log-likelihood function when the model is extended to include the additive innovation. As a result, it is crucial to account for both additive and multiplicative stochastic components when we apply the ACD model to model the price fluctuation time series.

5.7 Summary

This chapter proposed a new method for modelling the execution probability at a specified time period from the fluctuation of the asset price during the interested period. The advantage of this approach over traditional techniques is that it requires less data to model the execution probability at all price levels simultaneously since it requires only one record per sample while traditional techniques require n records per sample to model the execution probability for n price levels. Additionally, it provides a natural way to apply traditional time series analysis techniques to model the execution probability.

The statistical analysis of the price fluctuation dataset obtained from the Multi Commodity Exchange of India and the New York Stock Exchange in Section 5.3 indicated that the form of the market seems to have a strong impact on the dynamics of price fluctuations, as the strength and persistence of serial dependency in price fluctuations mainly differ between the individual exchanges and less between the different assets traded in the same exchange. The analysis also suggested that the price fluctuation process seem to have a long range dependency with a clear intraday seasonality pattern. The buy price fluctuation process and sell price fluctuation process of the same instrument are not necessarily identical and, thus one might need to model them separately. The analysis of the dependency between price fluctuation, return and volatility indicated that price fluctuation is highly correlated with the direction of return during the same period in the sense that buy price fluctuation is negatively correlated with return, while sell price fluctuation is positively correlated with return. However, the correlation between price fluctuation and previous returns is typically weak and it might not be useful for predicting future price fluctuation. Additionally, the results also indicated that price fluctuation is also strongly correlated to volatility, as estimated by the range between the highest and lowest price.

To find a suitable model for the price fluctuation process, Section 5.4 started the investigation by analysing the unconditional model of price fluctuations. In particular, we derived the unconditional distribution of price fluctuation when the asset price is assumed to follow the arithmetic Brownian motion. Moreover, we also fitted several distributions with non-negative support including the exponential, Weibull, gamma, generalised gamma, generalised F and Burr distribution to the buy price fluctuation dataset using the maximum likelihood estimator. The results indicated that the maximum likelihood estimator is not a good method for estimating model parameters of the price fluctuation process, as the estimated distribution converged to the distribution that had large probability density at zero rather the distribution that provided the best fit. To solve the problem, we proposed to estimate model parameters by maximising the likelihood of the discrete distribution implied by the distribution considered rather than maximising the likelihood of the distribution directly. The experiment results indicated that the distribution estimated by the proposed method does not suffer from this problem and is able to estimate the empirical distribution reasonably well. Among all considered models the generalised F distribution is the best performing distributions while the Burr distribution and the generalised gamma distribution are the second and third-best models, respectively.

In Section 5.5, we further investigatd this issue by applying three major time series analysis techniques, which are the autoregressive moving average (ARMA) model, the generalised autoregressive conditional heteroskedasticity (GARCH) model and the autoregressive conditional duration (ACD) model, to model price fluctuation processes. For each of these models, we derived a modified likelihood function that accounts for the discreteness and non-negativity of the price fluctuations dataset, and fit these models to the price fluctuation dataset both by maximising the original likelihood function and the modified likelihood function. Since these models can be specified based on several distributional assumptions, we utilised the normal and asymmetric Laplace distribution for the ARMA and ARMA-GARCH models, while we utilised the exponential, Weibull and generalised gamma distribution for the ACD model. The experimental results indicated that the modified likelihood function always provide improved results for the ACD model under the Weibull and generalised gamma distribution, which allow the density function to be infinite at zero, while the gain obtained from the exponential distribution is quite limited. However, the improvement in case of the ARMA and ARMA-GARCH models is not consistent, indicating that one might need to estimate these models using both original likelihood and modified likelihood in order to find the best model parameters. Among all models considered, the ACD model with generalised gamma distribution estimated by maximising the modified likelihood function is the best performing model both from in-sample and out-of-sample tests.

Although the experiment performed in Section 5.5 indicate that the ACD model is the most appropriate model for analysing price fluctuation processes, the assumption made by the basic ACD model is somewhat limited. To find a better model, Section 5.6 further applied several extensions of the basic ACD model to model the price fluctuation process. The experimentation result indicated that the ABAMACD model, which encompasses all other models as its special cases, is the best performing model in both training and testing dataset according to all performance measures. Consequently, we will utilise the ABAMACD model as a primary tool for estimating the probability that the limit order submitted at each price level will be executed in the rest of this study.

Chapter 6

Asset price dynamics in a continuous double auction market

This chapter presents a stochastic model of asset prices in an order-driven market whose dynamics are described by the incoming flow of market orders, limit orders and order cancellation processes. Particularly, we introduce a framework to model the dynamics of asset prices giving the statistical properties of those processes, thus, establishing the relationship between the microscopic dynamics of the limit order book and the long-term dynamics of the asset price process. Unlike traditional methods that model asset price dynamics using a one-dimensional stochastic process, the proposed framework models the dynamics using a two-dimensional stochastic process where the additional dimension represents information about the latest price change. Using dynamic programming methods, we are able to efficiently compute several interesting properties of the asset price dynamic (i.e. volatility, occupation probability and first-passage probability), conditioning on the trading horizon, without resorting to simulation.

6.1 Introduction

Many equity and derivative exchanges around the world are nowadays organised as *order-driven* markets where the instantaneous liquidity is provided through a limit order book, in which unexecuted or partially executed limit orders submitted by market participants are stored and waiting for possible execution. These types of market have gained in popularity in recent years over *quote-driven* markets where liquidity is provided by market makers or designated dealers. Examples of such equity markets include the Electronic Communication Networks in the United States, the Toronto Stock Exchange, the Stockholm Stock Exchange, the Australian Stock Exchange, the Shanghai Stock Exchange and the Tokyo Stock Exchange. Order-driven markets for derivative instruments have also gained in popularity in recent years over the traditional open-outcry auctions, and many derivative exchanges, including the Chicago Mercantile Exchange, the International Petroleum Exchange of London, the Sydney Futures Exchange, and the Hong Kong Futures Exchange, are nowadays organised in this fashion.

The growing popularity of order-driven markets clearly establishes a need for economic and statistical models of such markets. At a fundamental level, a good model should provide some insight into the interplay between order flows, liquidity, and price dynamics in these markets, while, at the level of application, such a model should also provide a quantitative framework for traders to optimise their execution strategies. Unfortunately, to the best of our knowledge, previous statistical models linking order flows to price dynamics (e.g. Bouchaud et al. [14], Mike and Farmer [68], and Boer et al. [11]) study this relation only by simulation, and, thus, may not be appropriate to employ in real-time applications where fast computation is a necessity. Although, Const et al. [20] recently proposed a stochastic model of a limit order book that allows fast computation of various interesting quantities without resorting to simulation, their model does not allow fast computation of the return distribution, probability of execution, and volatility which are important quantities for optimising trade execution strategies.

The main objective of this chapter is to develop a model for explaining the relation between order flows and price dynamics in order-driven markets that is simple enough to allow fast computation of such quantities. To achieve this, we derive a new stochastic model of price dynamics from microscopic behaviour of the limit order book, and present a procedure to estimate its parameters from order flow properties. Unlike traditional methods that model price dynamics using one-dimensional stochastic processes, we propose to model these dynamics using a two-dimensional stochastic process where the additional dimension represents information about the latest price change. This added dimension enables the model to reproduce the negative first-order autocorrelation property as can be observed in real markets. Under the independent and identical order flow assumption, the parameters of the proposed model and the above quantities can be estimated using numerical transformation techniques. A comparison with simulation results illustrates that our model can accurately predict the desired probabilities without resorting to simulation.

This chapter is organised as follows. In Section 6.2, the background information on a number of key concepts utilised in this study and related works will be reviewed in detail to give the reader a clear view of the problems and environments studied in this chapter. Section 6.3 presents the main result of this chapter by firstly describing a stylised model for the dynamics of a limit order book where the order flow is described by independent Poisson processes, and then deriving the price dynamics and the relation between them. In Section 6.4, simulation results are compared with the estimates from our model to assess the accuracy of the proposed model. Finally, a conclusion and scope for future work are given in Section 6.5.

6.2 Background

6.2.1 Order-driven markets and the limit order book

Most order-driven markets utilise the continuous double auction mechanism to match buyers and sellers during trading hours. This mechanism permits traders to provide or take liquidity at any time while the market is open. Although trading rules in these markets can vary considerably (i.e. by the types of orders that may be submitted and the way in which they are handled), most order-driven markets operate primarily as limit order markets where traders execute their trade by submitting either market orders or limit orders.

Traders who provide liquidity submit *limit orders* (i.e. requests to buy a specific quantity at a price not exceeding some specified maximum, or to sell a specified quantity at a price not less than some specified minimum) to indicate the terms at which they want to trade. Unless it can be executed against a pre-existing order in the order book, a new limit order joins the queue in the limit order book and remains there until it is amended, cancelled, or executed against subsequent orders. Limit buy orders are called *bids*, and limit sell orders are called *asks*. The lowest selling price offered at any point in time is called the *best ask*, and the highest buying price, the *best bid*. The best prices may change as new orders arrive or old orders are cancelled. Prices in these markets are not continuous, but rather have discrete quanta called *ticks* with the minimum interval that prices change specified by the *tick size*. Throughout the rest of this chapter, all prices will be expressed as integers with the tick size equal to one.

Traders who take liquidity accept the above terms by submitting *market orders* (i.e. requests to transact a specified quantity at the best available price) to execute their trade at the best available price. A market order is normally executed immediately and as fully as possible. Any unexecuted part may then be converted to a limit order at the same price, or else executed at the next best available price resulting in partial executions at progressively worse price until the order is fully executed. Liquidity takers can also execute their trade immediately by submitting marketable limit orders, which are limit orders to buy (sell) at or above (below) the best available price. Since both market orders and marketable limit orders result in immediate execution, we do not make a distinction between them and refer to both of them as market orders in the rest of this chapter.

6.2.2 Related literatures

With the world-wide proliferation of order-driven markets, various studies have focused on modelling price dynamics in these markets with the aim of providing more insight into price formation and the stochastic properties of price fluctuations. Since the evolution of prices in such markets results from the interaction of incoming orders with existing orders in the limit order books, an understanding of this interaction is therefore required in order to understand these processes. To achieve this, a great deal of research has examined this relationship with recent studies focusing on understanding traders' decisions to submit more (or less) aggressive orders and how information in the limit order book affects these decisions. Recent empirical studies indicate that traders' decisions of when and how to trade are significantly influenced by the state of the order book (e.g. queued volume, the market depth, and the spread) as well as recent changes in both the order flow and the price [39]. For example, Biais, Hillion and Spatt [9] discover that traders place limit orders when the spread is large and the book is thin, and place market orders when the opposite holds true. Similarly, Ranaldo [82] shows that patient traders become more aggressive when the spread is wider, the volatility increases, and the own book is thicker as well as when the opposite book is thinner. Lo and Sapp [62] utilise an autoregressive conditional duration model to show that the execution of market orders, changes in the level of price uncertainty, and market depth impact the submission of both limit and market orders. Moreover, traders generally use market orders at times when execution risk for limit orders is higher, and use limit orders when the risk of unexpected price movements is highest. Verhoeven et al. [91] utilise logit regression to indicate that the spread, depth at the best price, price changes in five minutes and order imbalances are major determinants of the traders' decision to place market and limit orders.

Issues regarding the statistical properties of order flow have also gained more attention in recent years, and several stylised facts in these markets have been identified. These include: i) *Long-memory of order sign* [57, 13, 58], where the order flow appears to be a long memory process since the time series generated by replacing a buy order with +1 and a sell order with -1, exhibits a power-law decay both for market orders and limit orders; ii) *Power-law limit prices* [95, 14, 77], where the probability of a limit order placement depends significantly on the distance from the current best prices and this probability drops off asymptotically as a power law; iii) *Log-normal order size* [66, 13], where the order size distribution was reported to be very skewed, with tails possibly following a log-normal or power-law distribution; iv) *Non-exponential waiting time* [81, 84], where the waiting times between consecutive orders and between consecutive trades are not exponential. This refutes the hypothesis of constant trading activity during the day, as well as the modelling of this activity using a pure Poisson process.

Many analytical works have also attempted to explain some of these observed behaviours. Cohen et al. [19] consider traders' order submission strategies and discover that transaction cost causes bid-ask spread to be an equilibrium property of order-driven markets and the spread is negatively related to the order arrival rate. Glosten [35] analyses the nature of equilibrium in an idealised limit order book. His result indicates that the order book has a small positive bid-ask spread, and it is more profitable to trade in small size than in large size. Parlour [76] developed a dynamic model of a limit order market where traders' submission strategies are dependent on the state of the order book. In this model, all traders know that their order will effect the decisions of others, and, thus, the execution probability of their limit orders is endogenous. The result suggests that the optimal choice of either market or limit orders generates systematic patterns in any observed transaction data, and both sides of the book are important in determining a trader's choice. Foucault [32] describes a game theoretical model of price formation and order submission in a dynamic limit order market. His result indicates that the proportion of limit orders in the order flow is positively related to volatility, the ratio of filled limit orders to total number of limit orders is negatively related to volatility, and the proportion of limit orders is positively related to the spread. Foucault, Kadan and Landel [33] analyse a dynamic model of a limit order market populated by strategic liquidity traders of varying impatience who aim to optimise the trade off between the cost of delayed execution and the cost of immediacy. The result indicates that the proportion of patient traders in the population and the order arrival rate are the key determinants of the limit order book dynamics. In particular, traders submit aggressive limit orders when the order arrival rate is low or when the proportion of patient traders is larger. Lillo [56] considers the problem of the optimal limit order price in the framework of utility maximisation. The analytical solution of the problem gives insight into the origin of the empirically observed power law distribution of limit order prices. Although these models provide interesting insights into the price formation process, they contain unobservable parameters that govern agent preferences, and, thus, they are difficult to estimate and use in real applications.

Although traders make decisions in an extremely complex environment, in the end these decisions are reduced to the simple actions of placing and cancelling trading orders. Instead of attempting to anticipate how traders will behave, another approach is to start by assuming that their combined effect is to generate flows of order submission and cancellation with a known distribution, and then determine the quantities of interest based on this assumption. For example, Luckock's analysis [63] yields the stationary probability distribution for the best ask and best bid prices and the prices of actual trades when the arrival of new orders at each price level follows a Poisson process. Smith et al. [88] develop a microscopic dynamic statistical model of the order book under the assumption of independent and identical random order flow and analyse it using simulation, dimensional analysis, and mean field approximation. Their result provides testable predictions for basic properties of markets such as the depth of stored supply and demand versus price, the bid-ask spread, the price impact function, and the time and probability of filling orders in steady state. Mike and Farmer [68] develop a model of the order book based on empirical regularities of order flows in the London Stock Exchange and utilise it to simulate price formation which is then compared to those of real data. The result indicates that the prediction from the model is very good especially for small tick size stocks. Const et al. [20] propose a stochastic model of the continuous-time dynamics of a limit order book that can be utilised to compute the probability of various events, conditional on the state of the order book, including the probability of the mid-price increasing in the next move, the probability of executing an order at the best bid before the best ask move, and the probability of executing both a buy and a sell order at the best price before the price moves giving the state of the order book.

Whilst some of these works provide the interplay between order flows, liquidity and price dynamics (e.g. Bouchaud et al. [14], Mike and Farmer [68], and Boer et al. [11]), they study this relation only by simulation which may not be appropriate to employ in real applications where fast computation is necessity. Although, Const et al. [20] recently proposed a stochastic model of a limit order book that allows fast computation of various interesting quantities without resorting to simulation, their model does not allow fast computation of return distribution, probability of execution, and volatility which are important quantities for optimising trade execution strategies.

To fill this gap, this chapter aims to develop a model that is simple enough to allow fast computation of the interesting quantities. The model considered here is admittedly simpler in structure than other existing works since it does not incorporate strategic interaction of traders as in game theoretic approaches, nor does it account for long memory features of the order flow. However, it leads to an analytically tractable framework where several quantities of interest may be efficiently computed without resorting to simulation.

6.3 The Model

This section presents the main result of this chapter. We start by describing the model of the order book, derive the price dynamics, and the relationship between them.

6.3.1 Model of the order book

The limit order book model utilised in this chapter is adapted from the models of Smith et al. [88] and Const et al. [20] with some additional assumptions. It is constructed to be as analytically tractable as possible while capturing the key features of order driven markets. Particularly, we consider a limit order book model with the following assumptions:

A1: Limit orders are placed on an integer price grid $\mathbb{P} \equiv \{1, 2, ..., n\}$ which represents multiples of a price tick.

A2: There are large numbers of liquidity providers and liquidity takers acting independently of one another, with each individual only occasionally submitting an order to the exchange. This allows us to regard each order as originating from a different source and hence unrelated to any other order. The arrival of orders of any specified type is assumed to follow a Poisson process.

A3: All orders are of unit size. Hence all agents in the model need only to specify the price at which they want to trade. This eliminates the possibility of partial execution and the need for rules governing the handling of partially executed orders.

A4: Once submitted to the exchange, orders will be automatically cancelled when a specified lifetime is reached. This lifetime is also assumed to follow a Poisson process.

A5: Market participants prepare and submit their orders without making use of detailed information about the current state of the order book. Opportunistic traders are thus unable to take advantage of temporary anomalies in the order book.

A6: Liquidity providers will submit buy orders at price level p only when the price level $p - 1$ is previously occupied by other buy orders, while they will submit sell orders at price level p only when price level $p + 1$ is previously occupied by other sell orders.

Most of these assumption are similar to the one made in [88, 20]. The only difference is the assumption A6 which considerably simplifies the problem and appears to be indispensable to our analysis. Before discussing more details about this assumption, let us firstly define the notations utilised to described the state of the order book throughout the rest of this chapter.

Using notations similar to the one utilised by Const et al. [20], the state of the order book at a particular time t will be represented by $X(t) \equiv (X_1(t), ..., X_n(t))_{t\geq 0}$, where $|X_p(t)|$ is the number of unexecuted limit orders at price p, $1 \leq p \leq n$, and the sign of $X_p(t)$ indicates the side of the orders; particularly, there will be $X_p(t)$ sell orders at price p when $X_p(t)$ is positive, while there will be $-X_p(t)$ buy orders at price p when $X_p(t)$ is negative. Using this notation, the *best ask price*, $p_A(t)$, which is the lowest selling price offered at a particular time t, can be defined by

$$p_A(t) \equiv \inf \{p = 1, \dots, n \mid X_p(t) > 0\} \wedge (n + 1).$$

Similarly, the *best bid price*, $p_B(t)$ which is the highest buying price at a particular time t, can be defined by

$$p_B(t) \equiv \sup\{p = 1, \ldots, n \mid X_p(t) < 0\} \vee 0.$$

Notice that, when there are no sell orders in the book, the best ask is forced to be $n+1$, while the best bid is forced to be 0, when there is no buy orders in the book. From the definition of the best bid and the best ask, we can define the *mid-price*, $p_M(t)$, and the bid-ask spread, $s(t)$, by

$$p_M(t) \equiv \frac{p_A(t) + p_B(t)}{2} \text{ and } s(t) \equiv p_A(t) - p_B(t).$$

Under assumption A6, liquidity providers will submit limit buy orders at price level p only when price level $p-1$ is previously occupied by limit buy orders, while they will submit limit sell orders at price level p only when price level $p+1$ is previously occupied by limit sell orders. This suggests that the price level at which liquidity providers can submit limit buy and limit sell orders depends not only on the current state of the order book but also on its history. For example, if the order book has just changed from $(-5, 0, 0, -1, 4)$ to $(-5, 0, 0, 0, 4)$, liquidity providers can submit limit buy orders at all price levels from 1 to 4 while they can submit limit sell orders only at price level 4 and 5. Conversely, if the order book has just changed from $(-5, 1, 0, 0, 4)$ to $(-5, 0, 0, 0, 4)$, liquidity providers can submit limit buy orders only at price level 1 and 2 while they can submit limit sell orders at all price level from 2 to 5. If the order book has changed from $(-5, 0, -1, 1, 4)$ to $(-5, 0, 0, 1, 4)$ and then to $(-5, 0, 0, 0, 4)$, liquidity providers can submit limit buy orders from price level 1 to 4, while they can submit sell orders from price level 3 to 5.

Although we need the type of the order previously occupied at price level $p-1$ to determine whether we can submit limit buy orders at price level p or not, we do not need to determine this for all price levels. Particularly, knowing only the highest price level that is previously occupied by limit buy orders is enough to answer this question for all price levels since if price level p is previously occupied by buy orders, all price levels below p must also be previously occupied by buy orders as well. Similarly, knowing only the lowest price level that is previously occupied by sell orders is also enough to determine the price level at which liquidity providers can submit limit sell orders. Let us define the *reference ask price*, $r_A(t)$, as the lowest price level previously occupied by limit sell orders by

$$r_A(t) \equiv \inf\left\{p = 1, \ldots, n \mid X_p\left(\sup\left\{\hat{t} \leq t | X_p(\hat{t}) \neq 0\right\}\right) > 0 \ \wedge \ p > p_B(t)\right\}, \tag{6.1}$$

and the *reference bid price*, $r_B(t)$, as the highest price level that is previously occupied by limit buy orders at a particular time t by

$$r_B(t) \equiv \sup\left\{p = 1, \ldots, n \mid X_p\left(\sup\left\{\hat{t} \leq t | X_p(\hat{t}) \neq 0\right\}\right) < 0 \ \wedge \ p < p_A(t)\right\}. \tag{6.2}$$

Given these two reference prices, the set of all price levels at which liquidity traders can submit limit sell orders at a particular time t, $P_A(t)$, can be defined by

$$P_A(t) \equiv \{p = 1, \ldots, n \mid (p+1) \geq r_A(t) \ \wedge \ p > p_B(t)\}, \tag{6.3}$$

while the set of all price levels at which liquidity providers can submit limit buy orders at a particular time t, $P_B(t)$, can be defined by

$$P_B(t) \equiv \{p = 1, \ldots, n \mid (p-1) \leq r_B(t) \ \wedge \ p < p_A(t)\}. \tag{6.4}$$

From Equation (6.3) it is easy to see that the reference ask price can decrease by only one tick at a time since the lowest price level that liquidity providers can submit limit sell orders at a particular time t is always equal to or greater than $r_A(t) - 1$. Additionally the reference ask price can be increased by only one tick at a time, since this value will increase only when liquidity providers submit limit buy orders at a price equal to or greater than the current reference sell price and the highest price level that liquidity providers can submit limit buy orders at a particular time t is always equal to or less than $r_B(t) + 1$ as indicated in Equation (6.4). Similarly, when applying the same analysis to the reference bid price, one will find that its value can be changed by only one tick at a time as well. Since these reference prices can be changed by only one tick at a time, they are easier to model than other quantities (e.g. the best bid price, the best ask price and the mid-price) and thus we will utilise them as a proxy for the asset price in the rest of this chapter.

Let us now describe how the limit order book is updated by the incoming flow of market orders, limit orders and cancellation of limit orders at each price level. According to assumption A2 and A4, these flows are modelled as Poisson processes. Specifically, market buy and sell orders are assumed to arrive at a rate of μ orders per unit time, while limit buy and sell orders at each possible price level are assumed to arrive at a rate of α orders per unit time. In addition, all outstanding limit orders are cancelled randomly with a rate of δ per unit time. Assuming that all orders are of unit size (assumption A3),

- a market buy order decreases the quantity of sell orders at the best ask price : $X_{p_A(t)} \to X_{p_A(t)} - 1$
- a market sell order decreases the quantity of buy orders at the best bid price : $X_{p_B(t)} \to X_{p_B(t)} + 1$
- a limit buy order at price level $p \in P_B(t)$ increases the quantity of buy orders at price level p : $X_p \to X_p - 1$
- the arrival of a limit sell order at price level $p \in P_A(t)$ increases the quantity of sell orders at price level p : $X_p \to X_p + 1$
- a cancellation of an outstanding buy order at price level $p < p_A(t)$ decreases the quantity of buy orders at price p : $X_p \to X_p + 1$
- a cancellation of an outstanding sell order at price level $p > p_B(t)$ decreases the quantity of sell orders at price p : $X_p \to X_p - 1$

Given a sequence of these orders, the above rules completely determine the evolution of the order book. With the above assumptions, the order book process $X(t)$ is a stochastic process whose state space is a subset of $\mathbb{Z}^n$, and the dynamic behaviour and statistical properties of this process is completely specified by the three parameters characterising the model as summarised in Table 6.1. The next section will analyse this stochastic process to derive the model of the asset price dynamics from these three parameters.

6.3.2 Asset price dynamics

This section presents a framework for modelling the dynamics of the asset price from the order book model described in the previous section. Although the asset price can be represented by many quantities (e.g. the best price, the reference price and the mid-price), this chapter will utilise the reference price as a proxy for the asset price. This is because, under assumption A6, the reference price can be changed by only one tick at a time and thus is easier to model than other quantities which can be changed rapidly from one level to the others.

Before modelling the full dynamics of the reference price, let us firstly analyse a single-step transition from one reference price to the next. Under assumption A6, we know that the reference bid price and the reference ask price can be changed by only one tick at a time. Specifically, the reference bid price will increase from p to $p+1$ when liquidity providers submit limit buy orders at level $p+1$ and will decrease to $p-1$ when liquidity providers submit limit sell orders at price level p. Conversely, the reference ask price will increase from p to $p+1$ when liquidity providers submit limit buy orders at price level p and will decrease to $p-1$ when liquidity providers submit limit sell orders at price level $p-1$. Assume that the order book is initialised with the spread equal to one tick so that the reference bid price and the reference ask price are equal to the best bid price and the best ask price respectively. In this situation, the difference between these two reference prices will be one tick, and, hence, the submission of limit buy orders at price level $r_B(t)+1$ will increase both the reference bid price to level $r_B(t)+1$ and the reference ask price to level $r_A(t)+1$. Similarly, the submission of limit sell orders at price level $r_A(t)-1$ will decrease both the reference bid price to level $r_B(t)-1$ and the reference ask price to level $r_A(t)-1$. This illustrates that, when the order book is initialised with the spread equal to one, both reference prices will increase and decrease together and thus allow us to track their value by using either one of them.

Assume that the order book is initialised with the spread equal to one tick and the reference bid price is at level p. As time goes by, the order book will evolve according to the dynamics described in the previous section, and it will finally reach a situation when the reference bid price changes. Particularly, the reference bid price will decrease to $p-1$ when the orders at price level p change from buy orders

Parameter	Description	Dimensions
μ	arrival rate of market orders	shares/time
α	arrival rate of limit orders	shares/time
δ	cancellation rate of limit orders	1/time

Table 6.1: The six parameters that characterise this model.

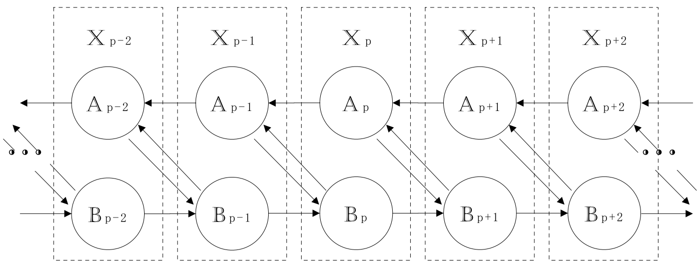


Figure 6.1: Dynamics of the reference price as modelled by the transitions between $\mathbb{A}_p$ and $\mathbb{B}_p$

to sell orders, while the reference bid price will increase to $p+1$ when the orders at price level $p+1$ change from buy orders to sell orders. After the reference price changes, the dynamics of the order book will continue from that price and will finally reach the situation when the reference price changes again. Define $\mathbb{A}_p$ as the situation when the reference bid price is at level p and there is only one sell order at price level $p+1$, $\mathbb{B}_p$ as the situation when the reference bid price is at level p and there is only one buy order at price level p and $\mathbb{X}_p$ as the situation when the reference bid price is at level p. The dynamics of the reference price can be illustrated in Figure 6.1. This diagram suggests that the dynamics of the reference price can be modelled as a two-dimensional stochastic process with state space $\mathbb{P} \times \{A, B\}$ and state transition diagram as shown in Figure 6.1. From any state $z = (p, l) \in \mathbb{P} \times \{A, B\}$, there are two possible transitions: to state $(p-1, A)$ when the reference price decreases by one tick, and to state $(p+1, B)$ when the reference price increases by one tick. Notice that the second dimension can be thought of as a representation of the latest price change since it will always be A when the reference price decreases and always be B when the reference price increases. To fully characterise this process, the stochastic behaviour of each state transition needs to be specified. Since these transitions are independent of the reference price, there will be only four main transition types, which are:

ρ_{A-}	the transition from (p, A) to $(p-1, A)$,
ρ_{A+}	the transition from (p, A) to $(p+1, B)$,
ρ_{B-}	the transition from (p, B) to $(p-1, A)$,
ρ_{B+}	the transition from (p, B) to $(p+1, B)$.

The stochastic behaviour of these transitions can be modelled from the dynamics of the order book at the reference bid price and the reference ask price. Particularly, let $T_{A|q_A}$ be the first time that the orders at the reference ask price change from q_A sell orders to buy orders, and $T_{B|q_B}$ be the first time that the orders at the reference bid price change from q_B buy orders to sell orders. The waiting time until the reference price changes when there are q_A sell orders at the reference ask price and q_B buy orders at the reference bid price, is

$$T_{W|q_A,q_B} = \min\left\{T_{A|q_A}, T_{B|q_B}\right\},$$

and its probability distribution can be computed from

$$\mathbb{P}\left\{T_{W|q_A,q_B} = t\right\} = \mathbb{P}\left\{T_{A|q_A} = t \wedge T_{B|q_B} > t\right\} + \mathbb{P}\left\{T_{B|q_B} = t \wedge T_{A|q_A} > t\right\}.$$

Since the dynamics of the orders at the reference bid price and the reference ask price are independent of each other, $T_{A|q_A}$ and $T_{B|q_B}$ are also independent and the above equation is reduced to

$$\mathbb{P}\left\{T_{W|q_A,q_B}=t\right\}=\mathbb{P}\left\{T_{A|q_A}=t\right\}\mathbb{P}\left\{T_{B|q_B}>t\right\}+\mathbb{P}\left\{T_{B|q_B}=t\right\}\mathbb{P}\left\{T_{A|q_A}>t\right\}.$$

Equivalently, the probability density function of $T_{W|q_A,q_B}$, denoted by $f_{W|q_A,q_B}(t)$, can be expressed by

$$f_{W|q_A,q_B}(t)=f_{A|q_A}(t)[1-F_{B|q_B}(t)]+f_{B|q_B}(t)[1-F_{A|q_A}(t)], \tag{6.5}$$

where $f_{A|q_A}(t)$ and $F_{A|q_A}(t)$ are the probability density function (p.d.f.) and cumulative distribution function (c.d.f.) of $T_{A|q_A}$, respectively, while $f_{B|q_B}(t)$ and $F_{B|a_B}(t)$ are the p.d.f. and c.d.f. of $T_{B|q_B}$ respectively. Since the order arrival and cancellation rate of buy and sell orders are similar, the distribution of $T_{A|q_A}$ and $T_{B|q_B}$ when $q_A=q_B$ must be similar as well and thus Equation (6.5) reduces to

$$f_{W|q_A,q_B}(t)=f_{T|q=q_A}(t)[1-F_{T|q=q_B}(t)]+f_{T|q=q_B}(t)[1-F_{T|q=q_A}(t)], \tag{6.6}$$

where $f_{T|q}(t)$ and $F_{T|q}(t)$ are respectively the p.d.f. and c.d.f. of the first time that the orders at the reference price change from q buy (sell) orders to sell (buy) orders. Additionally, the probability that the reference price will decrease to $\mathbb{A}_{p-1}$ at time t is given by

$$P_{B|q_A,q_B}(t)=\mathbb{P}\left\{T_{B|q_B}=t\ \wedge\ T_{W|q_A,q_B}=t\right\}=f_{T|q=q_B}(t)[1-F_{T|q=q_A}(t)]. \tag{6.7}$$

Similarly, the probability that the reference price will increase to $\mathbb{B}_{p-1}$ is given by

$$P_{A|q_A,q_B}(t)=\mathbb{P}\left\{T_{A|q_A}=t\ \wedge\ T_{W|q_A,q_B}=t\right\}=f_{T|q=q_A}(t)[1-F_{T|q=q_B}(t)]. \tag{6.8}$$

To apply the above equations to estimate the transition probability, the number of the orders at the reference price is required. Unfortunately, when the order book has just changed to $\mathbb{A}_p$ we know only the number of orders at the reference ask price, which must be equal to one, but not the number of orders at the reference bid price. Similarly, when the order book has just changed to $\mathbb{B}_p$, we know only the number of orders at the reference bid price but not the number of orders at the reference ask price. To solve the problem, we will assume that these unknown quantities are distributed according to some known probability mass function $f_q(x)$. Consequently, the distribution of the transition probability and the waiting time when the order book is in $\mathbb{A}_p$ can be estimated from

$$\begin{aligned}
\rho_{A-}(t) &= \sum_x P_{B|q_A=1,q_B=x}(t)f_q(x) = \sum_x f_{T|q=x}(t)[1-F_{T|q=1}(t)]f_q(x)\\
\rho_{A+}(t) &= \sum_x P_{A|q_A=1,q_B=x}(t)f_q(x) = \sum_x f_{T|q=1}(t)[q-F_{T|q=x}(t)]f_q(x)\\
f_{W|\mathbb{A}}(t) &= \sum_x f_{W|q_A=1,q_B=x}(t)f_q(x) = \rho_{A-}(t)+\rho_{A+}(t)
\end{aligned} \tag{6.9}$$

Similarly the distribution of the transition probability and the waiting time when the order book is in $\mathbb{B}_p$

can be estimated from

$$\begin{aligned}
\rho_{B-}(t) &= \sum_x P_{B|q_A=x,q_B=1}(t) f_q(x) = \sum_x f_{T|q=1}(t)[1 - F_{T|q=x}(t)] f_q(x) \\
\rho_{B+}(t) &= \sum_x P_{A|q_A=x,q_B=1}(t) f_q(x) = \sum_x f_{T|q=x}(t)[1 - F_{T|q=1}(t)] f_q(x) \qquad (6.10) \\
f_{W|\mathbb{B}}(t) &= \sum_x f_{W|q_A=x,q_B=1}(t) f_q(x) = \rho_{B-}(t) + \rho_{B+}(t)
\end{aligned}$$

Comparing Equation (6.9) and (6.10), we have

$$\rho_{A+}(t) = \rho_{B-}(t),\ \rho_{A-}(t) = \rho_{B}+(t) \text{ and } f_{W|\mathbb{A}}(t) = f_{W|\mathbb{B}}(t). \qquad (6.11)$$

Hence, given the distribution of $f_{T|q}(t)$, $F_{T|q}(t)$ and $f_q(x)$, we can derive the transition probability of the asset price dynamics model from Equation (6.9) and (6.10). The next question is how to estimate these three distributions from the three parameters of the order flow model described in the previous section, and this will be the main subject of the next section.

To avoid the corner condition and simplify the analysis, the rest of this chapter will model the difference between the reference price and the initial reference price, i.e. $r_B(t) - r_B(0)$, instead of directly modelling the reference price and further assume that this difference has no bound so that the domain of the first-dimension becomes $\mathbb{Z}$ rather than $\mathbb{P}$. Particularly we will model this dynamic using a two-dimensional stochastic process $\{Z(t) = [P(t), L(t)]\,; t \geq 0\}$ with state space $\mathbb{Z} \times \{A, B\}$ and initial conditions $P(0) = 0$, $L(0) = A$ with probability a_0 and $L(0) = B$ with probability b_0 where $a_0 + b_0 = 1$. The state $Z(t)$ of this process at time t is a two-dimensional vector with the first component representing the difference between the reference price at time t and the initial reference price, while the second component being the lastest price change. Since the state holding time of all states in the model have the same distribution, this dynamic process can be approximated by a discrete-time Markov chain whose state holding time can be estimated by

$$T_H \equiv \int_0^\infty t f_{W|\mathbb{A}}(t) dt = \int_0^\infty t f_{W|\mathbb{B}}(t) dt, \qquad (6.12)$$

and its transition probability characterised by:

$$\rho_{A+} \equiv \int_0^\infty \rho_{A+}(t) dt,\ \ \rho_{A-} \equiv \int_0^\infty \rho_{A-}(t) dt, \qquad (6.13)$$

$$\rho_{B+} \equiv \int_0^\infty \rho_{B+}(t) dt,\ \ \rho_{B-} \equiv \int_0^\infty \rho_{B-}(t) dt. \qquad (6.14)$$

The rest of this section will present a numerical method for estimating the quantities of interest from the asset price dynamics.

Occupancy probability, expected return and the volatility

Let $A(p,n)$ and $B(p,n)$ be the probabilities that the process is in state (p,A) and (p,B) at the n-th time step respectively. The evolution of these two probabilities is described by the master equation

$$A(p,n) = \rho_{A-}A(p+1,n-1) + \rho_{B-}B(p+1,n-1) \tag{6.15}$$

$$B(p,n) = \rho_{A+}A(p-1,n-1) + \rho_{B+}B(p-1,n-1) \tag{6.16}$$

with initial condition $A(0,0) = a_0$ and $B(0,0) = b_0$ where $a_0 + b_0 = 1$. Since the asset price can be changed by only one tick at a time, the asset price at the n-th time step must be in the closed interval $[-n,n]$ which is a finite set. This allows us to compute the value of $A(p,n)$ and $B(p,n)$ directly from the master equation by using a dynamic programming approach which requires $O(n)$ space and $O(n^2)$ running time. The probability that the reference price will be at level p at the n-th time step, $P(p,n)$, can be computed from

$$P(p,n) = A(p,n) + B(p,n) \tag{6.17}$$

Accordingly the expected return at the n-th time step, $\mu(n)$, can be estimated by

$$\mu(n) = \sum_{p=-n}^{n} pP(p,n), \tag{6.18}$$

while the volatility at the n-th time step, $\sigma(n)$, can be estimated by

$$\sigma(n) = \sum_{p=-n}^{n} p\left[P(p,n) - \mu(n)\right]^2 = \sum_{p=-n}^{n} pP(p,n)^2 - \mu(n)^2, \tag{6.19}$$

First-passage probability and survival probability

Since an important aspect of the first-passage phenomenon is the condition by which the process terminates when the target state is reached, we can compute the first-passage probability to state $z = (\hat{p},\hat{l})$ by the occupancy probability of a state z in a modified process where state z is an absorbing state. Let $F^{+}_{A,\hat{p}}(p,n)$ and $F^{+}_{B,\hat{p}}(p,n)$ be the probability that the modified process is in state (p,A) and (p,B) at the n-th time step when $\hat{p} > 0$ and $\hat{l} = B$ respectively. The evolution of these probabilities can be described by the master equations

$$F^{+}_{A,\hat{p}}(p,n) = \begin{cases} \rho_{A+}F^{+}_{A,\hat{p}}(p-1,n-1) + \rho_{B+}F^{+}_{B,\hat{p}}(p-1,n-1) & \text{if } p < \hat{p}-1 \\ 0 & \text{otherwise} \end{cases} \tag{6.20}$$

$$F^{+}_{B,\hat{p}}(p,n) = \begin{cases} \rho_{A-}F^{+}_{A,\hat{p}}(p+1,n-1) + \rho_{B-}F^{+}_{B,\hat{p}}(p+1,n-1) & \text{if } p \le \hat{p} \\ 0 & \text{otherwise} \end{cases} \tag{6.21}$$

with initial conditions $F^{+}_{A,\hat{p}}(0,0) = a_0$ and $F^{+}_{B,\hat{p}}(0,0) = b_0$ where $a_0 + b_0 = 1$. Similarly, let $F^{-}_{A,\hat{p}}(p,n)$ and $F^{-}_{B,\hat{p}}(p,n)$ be the probability that the modified process is in state (p,A) and (p,B) at the n-th time step when $\hat{p} < 0$ and $\hat{l} = A$. The evolution of this probability can be described by the master equations

$$F^-_{A,\hat{p}}(p,n) = \begin{cases} \rho_{A+}F^-_{A,\hat{p}}(p-1,n-1) + \rho_{B+}F^-_{B,\hat{p}}(p-1,n-1) & \text{if } p \geq \hat{p} \\ 0 & \text{otherwise} \end{cases} \tag{6.22}$$

$$F^-_{B,\hat{p}}(p,n) = \begin{cases} \rho_{A-}F^-_{A,\hat{p}}(p+1,n-1) + \rho_{B-}F^-_{B,\hat{p}}(p+1,n-1) & \text{if } p > \hat{p}+1 \\ 0 & \text{otherwise} \end{cases} \tag{6.23}$$

with similar initial conditions as before. These master equations allow us to compute the value of $F^+_{A,\hat{p}}(p,n)$, $F^+_{B,\hat{p}}(p,n)$, $F^-_{A,\hat{p}}(p,n)$ and $F^-_{B,\hat{p}}(p,n)$ using a dynamic programming approach which requires $O(n)$ space and $O(n^2)$ running time. Using these equations, the probability that the reference price reaches price level p for the first time at the n-th time step, $F(p,n)$ can be computed from

$$F(p,n) = \begin{cases} F^+_{B,p}(p,n) & \text{for } p > 0 \\ F^-_{A,p}(p,n) & \text{for } p < 0 \end{cases} \tag{6.24}$$

Accordingly, the survival probability, the probability that the reference price does not reach level p at the n-th time step, $S(p,n)$ can be computed from

$$S(p,n) = 1 - \sum_{i=1}^{p} F(i,n) \tag{6.25}$$

6.3.3 Parameter estimation

This section presents a method for estimating the two distributions required to derive the transition probability of the asset price dynamics model described in the previous section.

Distribution of waiting time until orders at reference price change side

We now present a method to estimate the distribution of the waiting time until the orders at the reference price change side. Let $T_{i,j}$ be a random variable representing the first-passage time that the state of the order book at the reference price changes from i to j when $i > j$. Note that the random variable we want to model is $T_{q,-1}$ which represents the waiting time until the orders at the reference bid (ask) price change from q buy (sell) orders to buy (sell) orders. This first passage time can be expressed in terms of the first passage time to neighbouring states as follows

$$T_{i,j} = T_{i,i-1} + T_{i-1,i-2} + \cdots + T_{j+1,j}, \tag{6.26}$$

where the random variables on the right-hand side are mutually independent. Let $f_{i,j}$ be the p.d.f. of $T_{i,j}$ and let $\widehat{f}_{i,j}$ be its Laplace transform, i.e.,

$$\widehat{f}_{i,j} \equiv \int_0^\infty e^{-st} f_{i,j}(t)dt \equiv Ee^{-sT_{i,j}}. \tag{6.27}$$

From Equation (6.26), we have

$$\widehat{f}_{i,j}(s) = \prod_{k=j+1}^{k=i} \widehat{f}_{k,k-1}(s). \tag{6.28}$$

Therefore, in order to compute $\widehat{f}_{i,j}$, it suffices to compute the simpler Laplace transform of the first passage time to a neighbouring state $\widehat{f}_{i,i-1}$. Let λ_i be the rate of the transition from state i to $i+1$, and let σ_i be the rate of the transition from state i to $i-1$. By considering the transition in state i, we have

$$T_{i,i-1} = \left(\frac{\sigma_i}{\lambda_i + \sigma_i}\right) T_i^w + \left(\frac{\lambda_i}{\lambda_i + \sigma_i}\right) \left(T_i^w + T_{i+1,i} + T_{i,i-1}\right),$$

where T_i^w is the state i holding time which is an exponential waiting time with rate $\lambda_i + \sigma_i$. Applying the Laplace transform to the above relation, we get

$$\begin{aligned}
\widehat{f}_{i,i-1}(s) &= \left(\frac{\sigma_i}{\lambda_i + \sigma_i}\right)\left(\frac{\lambda_i + \sigma_i}{\lambda_i + \sigma_i + s}\right) + \left(\frac{\lambda_i}{\lambda_i + \sigma_i}\right)\left[\left(\frac{\lambda_i + \sigma_i}{\lambda_i + \sigma_i + s}\right)\widehat{f}_{i+1,i}(s)\widehat{f}_{i,i-1}(s)\right] \\
&= \frac{\sigma_i}{\lambda_i + \sigma_i + s} + \frac{\lambda_i \widehat{f}_{i+1,i}(s)\widehat{f}_{i,i-1}(s)}{\lambda_i + \sigma_i + s}
\end{aligned}$$

Rearranging $\widehat{f}_{i,i-1}(s)$ to the left-hand side, we obtain

$$\widehat{f}_{i,i-1}(s) = \frac{\sigma_i}{\lambda_i + \sigma_i + s - \lambda_i \widehat{f}_{i+1,i}(s)} \tag{6.29}$$

Iterating on Equation (6.29) produces a continued fraction [1]

$$\widehat{f}_{i,i-1}(s) = -\frac{1}{\lambda_{i-1}} \Phi_{k=i}^{\infty} \frac{-\lambda_{k-1}\sigma_k}{\lambda_k + \sigma_k + s}, \tag{6.30}$$

where

$$\Phi_{k=1}^{\infty} \frac{a_k}{b_k} \equiv \frac{a_1}{b_1 +} \frac{a_2}{b_2 +} \frac{a_3}{b_3 +} \cdots . \tag{6.31}$$

Abate and Whitt [1] illustrate that when this continued fraction is convergent, its value can be approximated using a sample recursion for calculating the successive approximantions. Particularly, given the continued fractions in Equation (6.31), we have

$$w_n \;=\; \Phi_{k=1}^{n} \frac{a_k}{b_k} \;\approx\; \frac{P_n}{Q_n},$$

where $P_0 = 0$, $P_1 = a_1$, $Q_0 = 1$, $Q_1 = b_1$ and

$$\begin{aligned}
P_n &= b_n P_{n-1} + a_n P_{n-2} \\
Q_n &= b_n Q_{n-1} + a_n Q_{n-2}
\end{aligned}$$

for $n \geq 2$. To increase the quality of the approximation, they also suggest that it is prudent to renormalise these values after a couple of iterations by dividing the current value of P_k, Q_k, P_{k-1} and Q_{k-1} all by

Q_k. Combining Equation (6.28) and (6.30), we get

$$\widehat{f}_{i,j}(s) = \prod_{k=j+1}^{k=i} \left\{ -\frac{1}{\lambda_{i-1}} \Phi_{k=i}^{\infty} \frac{-\lambda_{k-1}\mu_k}{\lambda_k + \mu_k + s} \right\}. \tag{6.32}$$

Consequently, we can estimate $f_{T|q}(x)$ and $F_{T|q}(x)$ by applying the numerical inverse Laplace transform to $\widehat{f}_{q,-1}(s)$ and $\widehat{f}_{q,-1}(s)/s$ when the transition rate in each state is specified by

$$\lambda_i = \alpha,\ \sigma_i = \begin{cases} \mu + i\delta & \text{if } i > 0, \\ \alpha & \text{if } i = 0. \end{cases}$$

Distribution of the number of orders at the reference price

This section presents a method for approximating the number of orders at the new reference price when the reference price has just changed. Let us consider the situation when the reference bid price has just changed from level p to $p+1$. In this situation, the number of orders at the new reference bid price must be equal to one, while the number of orders at the new reference ask price is a random variable whose value is dependent on the dynamics of the order book at price level $p+2$. Under assumption A7, the dynamics of the orders at price level $p+2$ will evolve according to a birth-death process with birth rate α and death rate $i\delta$ in state $i \geq 1$, which have the same behaviour as that of the $M/M/\infty$ queuing system. Thus we will approximate this distribution using the steady-state probability of this process, which can be obtained from the following equation

$$f_q(x) = \frac{(\alpha/\delta)^x e^{-\alpha/\delta}}{x!} \tag{6.33}$$

6.4 Numerical Results

The proposed order book model allows one to compute various quantities of interest both by simulating the evolution of the order book as described in Section 6.3.1 and by using the estimation techniques in Section 6.3.2 and 6.3.3, based on the order flow parameters μ, α and δ. In this section, we compute these quantities from both methods and compare them to assess the precision of our estimation in several settings. Particularly, we will fix the market order arrival rate μ at 1, vary the limit order arrival rate α from $\{1/4, 1/2, 1, 2, 4\}$ and vary the limit order cancellation rate δ from $\{\alpha, \alpha/2, \alpha/4, \alpha/8\}$ which results in twenty unique parameter settings. In Section 6.4.1, we compare the parameter of the asset price model estimated from the simulation to the one obtained from the proposed estimation framework. Then Section 6.4.2 will compare the prediction from our model to the simulation results.

6.4.1 Parameter estimation

To access the accuracy of the proposed parameter estimation method, we compare the value of ρ_{A-} and T_H estimated by our model to the one obtained from Monte Carlo simulation. Table 6.2 gives these two parameters as computed using both simulation and the proposed method. The simulation results, reported as 95% confidence intervals, agree very well with the estimation results when α is large; however, there

μ	α	δ	Simulation		Estimated		Corrected $f_Q(x)$	
			ρ_{A-}	T_H	ρ_{A-}	T_H	ρ_{A-}	T_H
1.0000	0.2500	0.2500	0.5191 ± 0.0004	3.1333 ± 0.0021	0.5088	3.1931	0.5194	3.1333
1.0000	0.2500	0.1250	0.4816 ± 0.0004	3.5152 ± 0.0024	0.4432	3.7406	0.4816	3.5153
1.0000	0.2500	0.0625	0.4293 ± 0.0004	3.9325 ± 0.0028	0.3276	4.5361	0.4291	3.9328
1.0000	0.2500	0.0363	0.3727 ± 0.0004	4.3315 ± 0.0032	0.1751	5.5106	0.3726	4.3356
1.0000	0.5000	0.5000	0.5238 ± 0.0004	1.9685 ± 0.0015	0.5173	1.9951	0.5243	1.9699
1.0000	0.5000	0.2500	0.4650 ± 0.0005	2.4772 ± 0.0021	0.4351	2.6004	0.4654	2.4764
1.0000	0.5000	0.1250	0.3811 ± 0.0005	3.0992 ± 0.0029	0.2992	3.4740	0.3812	3.1006
1.0000	0.5000	0.0625	0.2906 ± 0.0005	3.7645 ± 0.0040	0.1443	4.5039	0.2903	3.7630
1.0000	1.0000	1.0000	0.5296 ± 0.0005	1.2460 ± 0.0011	0.5267	1.2532	0.5295	1.2468
1.0000	1.0000	0.5000	0.4556 ± 0.0006	1.8724 ± 0.0021	0.4389	1.9243	0.4553	1.8720
1.0000	1.0000	0.2500	0.3500 ± 0.0007	2.9322 ± 0.0043	0.3083	3.1262	0.3499	2.9318
1.0000	1.0000	0.1250	0.2412 ± 0.0008	4.6019 ± 0.0094	0.1804	5.0791	0.2420	4.6089
1.0000	2.0000	2.0000	0.5337 ± 0.0005	0.7598 ± 0.0008	0.5345	0.7596	0.5342	0.7601
1.0000	2.0000	1.0000	0.4555 ± 0.0007	1.4156 ± 0.0021	0.4494	1.4320	0.4557	1.4159
1.0000	2.0000	0.5000	0.3575 ± 0.0011	3.1717 ± 0.0078	0.3429	3.2639	0.3576	3.1802
1.0000	2.0000	0.2500	0.2863 ± 0.0017	9.1494 ± 0.0440	0.2726	9.4732	0.2865	9.2062
1.0000	4.0000	4.0000	0.5381 ± 0.0006	0.4366 ± 0.0005	0.5396	0.4349	0.5379	0.4364
1.0000	4.0000	2.0000	0.4609 ± 0.0009	0.9918 ± 0.0018	0.4595	0.9950	0.4610	0.9923
1.0000	4.0000	1.0000	0.3831 ± 0.0016	3.3720 ± 0.0128	0.3790	3.4010	0.3832	3.3738
1.0000	4.0000	0.5000	0.3618 ± 0.0041	23.6960 ± 0.2564	0.3615	23.8707	0.3635	23.7700

Table 6.2: The parameters of the asset pricing model obtained from simulation results (95% confidence intervals), the proposed estimatation method, and the proposed estimation method with corrected $f_Q(x)$.

are significant differences when the α is small. These differences result from the fact that when α is small relative to μ, the reference price will change so quickly that the estimation of $f_Q(x)$ with the steady state probability described in Section 6.3.3 is no longer accurate. Additionally, the difference also tends to be larger when δ is smaller. Although this may seem to contradict the above argument since smaller δ will lengthen the state holding time which should make the distribution $f_Q(x)$ more similar to the steady-state probability, the actual problem in this situation is that smaller δ is associated with smaller ρ_{A-} and ρ_{B+}, which causes the reference price to move back and forth between a particular price rather than extending the move to the new price. This make the $f_Q(x)$ look more similar to the dynamics of the order at the best price rather than the one analysed in Section 6.3.3.

To confirm that these errors are actually caused by the problem in the estimation of $f_Q(x)$, we re-estimate the parameters of the model by setting $f_Q(x)$ to the empirical distribution obtained from the simulation. The results agreed very well with the simulation results as illustrated in the last two columns of Table 6.2.

6.4.2 Model prediction

As discussed in the introduction, volatility, return distribution and probability of execution are the main quantities of interest for applications in algorithmic trading. A good asset pricing model should allow us to predict these quantities correctly. To assess the accuracy of the prediction obtained from our asset pricing model, this section compares the results obtained from simulations to the ones estimated from our model. Since we model the asset price dynamics in a discrete-time setting while the simulation is performed in a continuous-time setting, all results reported in this section will be in discrete-time and the time step for the simulation is the number of times the reference price changes.

Volatility

Table 6.3 gives the volatility computed using both simulation and our numerical method at several time steps. The simulation results agree very well with our numerical computations when α is large, while they differ substantially when the α is small, as in the previous section.

μ	α	δ	Simulation			Estimated		
			$n = 10$	$n = 50$	$n = 100$	$n = 10$	$n = 50$	$n = 100$
1.0000	0.2500	0.25000	10.39 ± 0.29	51.67 ± 1.43	101.65 ± 2.83	10.71	53.88	107.84
1.0000	0.2500	0.12500	8.33 ± 0.23	39.83 ± 1.10	79.89 ± 2.22	9.36	46.52	92.96
1.0000	0.2500	0.06250	5.79 ± 0.16	25.53 ± 0.71	49.24 ± 1.37	7.74	37.83	75.45
1.0000	0.2500	0.03125	3.59 ± 0.10	12.85 ± 0.37	-	6.27	30.03	59.74
1.0000	0.5000	0.50000	10.63 ± 0.29	52.93 ± 1.47	108.35 ± 3.02	10.89	54.89	109.88
1.0000	0.5000	0.25000	7.79 ± 0.22	37.57 ± 1.04	74.50 ± 2.08	8.81	43.57	87.03
1.0000	0.5000	0.12500	4.84 ± 0.13	21.87 ± 0.61	42.69 ± 1.19	6.47	31.10	61.90
1.0000	0.5000	0.06250	2.77 ± 0.08	10.92 ± 0.30	20.85 ± 0.58	4.51	20.89	41.37
1.0000	1.0000	1.00000	11.16 ± 0.31	55.08 ± 1.53	110.52 ± 3.09	11.12	56.15	112.44
1.0000	1.0000	0.50000	7.86 ± 0.22	38.53 ± 1.07	77.53 ± 2.17	8.52	41.99	83.83
1.0000	1.0000	0.25000	4.91 ± 0.14	22.69 ± 0.63	43.77 ± 1.23	5.74	27.28	54.21
1.0000	1.0000	0.12500	2.87 ± 0.08	12.07 ± 0.33	23.67 ± 0.67	3.63	16.35	32.24
1.0000	2.0000	2.00000	11.47 ± 0.32	57.80 ± 1.60	115.12 ± 3.23	11.29	57.08	114.32
1.0000	2.0000	1.00000	8.17 ± 0.23	41.08 ± 1.14	81.66 ± 2.29	8.52	41.98	83.82
1.0000	2.0000	0.50000	5.59 ± 0.15	25.89 ± 0.72	52.08 ± 1.47	5.91	28.17	55.99
1.0000	2.0000	0.25000	4.12 ± 0.11	19.62 ± 0.54	38.17 ± 1.09	4.43	20.48	40.54
1.0000	4.0000	4.00000	11.74 ± 0.33	58.02 ± 1.61	116.73 ± 3.29	11.47	58.08	116.33
1.0000	4.0000	2.00000	8.69 ± 0.24	41.76 ± 1.16	85.63 ± 2.41	8.69	42.89	85.64
1.0000	4.0000	1.00000	6.46 ± 0.18	30.59 ± 0.85	60.96 ± 1.73	6.52	31.36	62.41
1.0000	4.0000	0.50000	6.07 ± 0.17	28.63 ± 0.87	-	6.01	28.69	57.04

Table 6.3: Volatility obtained from simulation results (95% confidence intervals) and the proposed estimation method.

μ	α	δ	Simulation			Estimated		
			$n = 10$	$n = 50$	$n = 100$	$n = 10$	$n = 50$	$n = 100$
1.00000	0.25000	0.25000	0.77000 ± 0.01830	0.88558 ± 0.02786	0.91350 ± 0.08044	0.77427	0.88764	0.91314
1.00000	0.25000	0.12500	0.78640 ± 0.01919	0.89409 ± 0.02907	0.90828 ± 0.05065	0.77427	0.88765	0.91210
1.00000	0.25000	0.06250	0.79870 ± 0.01992	0.90398 ± 0.03071	0.91763 ± 0.03873	0.77234	0.88661	0.91434
1.00000	0.25000	0.03125	0.80700 ± 0.02045	0.90629 ± 0.03113	0.92449 ± 0.05409	0.76747	0.88398	0.90808
1.00000	0.50000	0.50000	0.77900 ± 0.01877	0.89038 ± 0.02853	0.90894 ± 0.03751	0.77418	0.88760	0.91311
1.00000	0.50000	0.25000	0.78570 ± 0.01915	0.89512 ± 0.02925	0.91574 ± 0.04627	0.77391	0.88745	0.91194
1.00000	0.50000	0.12500	0.79680 ± 0.01980	0.90137 ± 0.03026	0.92452 ± 0.07383	0.76839	0.88448	0.91171
1.00000	0.50000	0.06250	0.78750 ± 0.01925	0.89702 ± 0.02955	0.92174 ± 0.08367	0.75402	0.87662	0.90453
1.00000	1.00000	1.00000	0.77430 ± 0.01852	0.89041 ± 0.02853	0.91513 ± 0.07328	0.77405	0.88753	0.91305
1.00000	1.00000	0.50000	0.78060 ± 0.01886	0.89218 ± 0.02882	0.91584 ± 0.07088	0.77360	0.88729	0.91387
1.00000	1.00000	0.25000	0.77700 ± 0.01867	0.88900 ± 0.02834	0.92683 ± 0.18288	0.76460	0.88243	0.91308
1.00000	1.00000	0.12500	0.77370 ± 0.01849	0.88885 ± 0.02838	0.91078 ± 0.04661	0.74090	0.86928	0.90218
1.00000	2.00000	2.00000	0.77370 ± 0.01849	0.88930 ± 0.02839	0.91366 ± 0.05121	0.77395	0.88747	0.91402
1.00000	2.00000	1.00000	0.77430 ± 0.01852	0.88814 ± 0.02824	0.93590 ± 0.29412	0.77360	0.88729	0.91181
1.00000	2.00000	0.50000	0.77730 ± 0.01868	0.88909 ± 0.02838	0.91014 ± 0.04691	0.76561	0.88297	0.91156
1.00000	2.00000	0.25000	0.75560 ± 0.01758	0.87780 ± 0.02694	0.91086 ± 0.05785	0.75307	0.87610	0.90835
1.00000	4.00000	4.00000	0.77440 ± 0.01853	0.88570 ± 0.02788	0.90245 ± 0.03938	0.77382	0.88740	0.91295
1.00000	4.00000	2.00000	0.77690 ± 0.01866	0.89107 ± 0.02869	0.91490 ± 0.05087	0.77379	0.88739	0.91679
1.00000	4.00000	1.00000	0.76820 ± 0.01820	0.88292 ± 0.02754	0.91452 ± 0.07832	0.76860	0.88459	0.91470
1.00000	4.00000	0.50000	0.76770 ± 0.01817	0.88153 ± 0.02832	0.90880 ± 0.04317	0.76616	0.88327	0.91277

Table 6.4: First-passage probability to price level 1 obtained from simulation results (95% confidence intervals) and the proposed estimation method.

First-passage probability

Table 6.4-6.7 compare the first-passage probability to price levels $1, 2, 4$ and 6 from the simulations to the model-predicted probability. We computed these quantities using Monte Carlo simulation (using 10,000 replications) and the first-passage time model described in Section 6.3.2. The simulation results, reported as 95% confidence intervals, agree very well with the estimated results when p is equal to 1 and 2. However, the difference increases with p. Additionally, the results also illustrate a tendency toward a higher error when α is small and when δ is small, as discussed in the previous section. The results also indicate that the first-passage probability decreases when δ and α increase, and increases when δ and α decrease.

6.5 Summary

This chapter derived a stochastic model of asset prices in a stylised order-driven market whose dynamics are described by the incoming flow of market orders, limit orders and order cancellations, each of

μ	α	δ	Simulation			Estimated		
			$n = 10$	$n = 50$	$n = 100$	$n = 10$	$n = 50$	$n = 100$
1.00000	0.25000	0.25000	0.55000 ± 0.01106	0.77990 ± 0.01886	0.82023 ± 0.05378	0.55641	0.78373	0.83154
1.00000	0.25000	0.12500	0.54360 ± 0.01091	0.78331 ± 0.01903	0.81492 ± 0.03653	0.54105	0.77565	0.82100
1.00000	0.25000	0.06250	0.51660 ± 0.01034	0.76516 ± 0.01807	0.81683 ± 0.05481	0.51638	0.76227	0.81675
1.00000	0.25000	0.03125	0.46160 ± 0.00926	0.73342 ± 0.01661	0.77699 ± 0.02697	0.48472	0.74441	0.79577
1.00000	0.50000	0.50000	0.56590 ± 0.01142	0.78737 ± 0.01927	0.82030 ± 0.02631	0.55822	0.78467	0.83227
1.00000	0.50000	0.25000	0.53500 ± 0.01073	0.77854 ± 0.01878	0.81648 ± 0.03079	0.53363	0.77168	0.82206
1.00000	0.50000	0.12500	0.49560 ± 0.00991	0.75268 ± 0.01747	0.80270 ± 0.03402	0.48980	0.74733	0.80722
1.00000	0.50000	0.06250	0.39230 ± 0.00803	0.68993 ± 0.01494	0.74921 ± 0.04779	0.42724	0.70969	0.77544
1.00000	1.00000	1.00000	0.56110 ± 0.01131	0.78979 ± 0.01942	0.82142 ± 0.02997	0.56039	0.78580	0.83315
1.00000	1.00000	0.50000	0.52730 ± 0.01056	0.76620 ± 0.01814	0.80832 ± 0.04763	0.52926	0.76932	0.82422
1.00000	1.00000	0.25000	0.46170 ± 0.00926	0.73294 ± 0.01662	0.79532 ± 0.05293	0.47037	0.73603	0.80063
1.00000	1.00000	0.12500	0.37460 ± 0.00774	0.67694 ± 0.01454	0.73599 ± 0.02811	0.38413	0.68124	0.75552
1.00000	2.00000	2.00000	0.56270 ± 0.01134	0.78808 ± 0.01932	0.82286 ± 0.02980	0.56193	0.78659	0.83378
1.00000	2.00000	1.00000	0.52870 ± 0.01059	0.77226 ± 0.01849	0.81906 ± 0.04473	0.52925	0.76931	0.82020
1.00000	2.00000	0.50000	0.48370 ± 0.00968	0.74427 ± 0.01712	0.79095 ± 0.02599	0.47522	0.73889	0.80067
1.00000	2.00000	0.25000	0.41270 ± 0.00838	0.70589 ± 0.01560	0.76873 ± 0.03297	0.42382	0.70753	0.78537
1.00000	4.00000	4.00000	0.56780 ± 0.01146	0.78622 ± 0.01922	0.84592 ± 0.12635	0.56353	0.78742	0.83442
1.00000	4.00000	2.00000	0.53680 ± 0.01077	0.77613 ± 0.01870	0.82288 ± 0.04749	0.53178	0.77068	0.82716
1.00000	4.00000	1.00000	0.49010 ± 0.00980	0.74231 ± 0.01704	0.82072 ± 0.11285	0.49096	0.74799	0.81183
1.00000	4.00000	0.50000	0.47480 ± 0.00951	0.72723 ± 0.01729	0.79627 ± 0.02697	0.47796	0.74048	0.80613

Table 6.5: First-passage probability to price level 2 obtained from simulation results (95% confidence intervals) and the proposed estimation method.

μ	α	δ	Simulation			Estimated		
			$n = 10$	$n = 50$	$n = 100$	$n = 10$	$n = 50$	$n = 100$
1.00000	0.25000	0.25000	0.23650 ± 0.00557	0.57425 ± 0.01164	0.63955 ± 0.02132	0.23996	0.58660	0.67348
1.00000	0.25000	0.12500	0.19710 ± 0.00495	0.55791 ± 0.01125	0.63344 ± 0.02934	0.21362	0.56503	0.64789
1.00000	0.25000	0.06250	0.13710 ± 0.00399	0.48622 ± 0.00975	0.59842 ± 0.03328	0.17707	0.53180	0.63205
1.00000	0.25000	0.03125	0.06460 ± 0.00263	0.37558 ± 0.00777	0.54021 ± 0.08998	0.13834	0.49073	0.60968
1.00000	0.50000	0.50000	0.23900 ± 0.00560	0.58378 ± 0.01186	0.66789 ± 0.05121	0.24328	0.58921	0.67561
1.00000	0.50000	0.25000	0.18170 ± 0.00471	0.54909 ± 0.01106	0.63991 ± 0.04638	0.20195	0.55489	0.64732
1.00000	0.50000	0.12500	0.10950 ± 0.00351	0.45874 ± 0.00923	0.59870 ± 0.06815	0.14402	0.49724	0.60721
1.00000	0.50000	0.06250	0.03360 ± 0.00186	0.31080 ± 0.00673	0.43767 ± 0.03810	0.08591	0.41849	0.53525
1.00000	1.00000	1.00000	0.24830 ± 0.00575	0.59327 ± 0.01211	0.66384 ± 0.03332	0.24734	0.59235	0.68170
1.00000	1.00000	0.50000	0.18770 ± 0.00481	0.54469 ± 0.01097	0.64398 ± 0.05070	0.19537	0.54898	0.64999
1.00000	1.00000	0.25000	0.09970 ± 0.00333	0.44396 ± 0.00897	0.54066 ± 0.02580	0.12330	0.47249	0.59051
1.00000	1.00000	0.12500	0.03620 ± 0.00194	0.32635 ± 0.00701	0.44165 ± 0.02474	0.05824	0.36519	0.49704
1.00000	2.00000	2.00000	0.24560 ± 0.00571	0.59817 ± 0.01224	0.66533 ± 0.02774	0.25026	0.59460	0.68352
1.00000	2.00000	1.00000	0.19030 ± 0.00485	0.54724 ± 0.01104	0.62841 ± 0.02589	0.19535	0.54897	0.64240
1.00000	2.00000	0.50000	0.11990 ± 0.00369	0.47979 ± 0.00965	0.59260 ± 0.05586	0.12823	0.47864	0.61092
1.00000	2.00000	0.25000	0.07310 ± 0.00281	0.39920 ± 0.00822	0.51328 ± 0.02249	0.08340	0.41426	0.55744
1.00000	4.00000	4.00000	0.25690 ± 0.00588	0.60062 ± 0.01230	0.66683 ± 0.02215	0.25334	0.59695	0.68878
1.00000	4.00000	2.00000	0.19560 ± 0.00493	0.54149 ± 0.01092	0.81902 ± 0.70808	0.19913	0.55238	0.65634
1.00000	4.00000	1.00000	0.14280 ± 0.00408	0.49774 ± 0.01001	0.58946 ± 0.02382	0.14534	0.49873	0.61627
1.00000	4.00000	0.50000	0.12610 ± 0.00380	0.48898 ± 0.00985	0.62814 ± 0.07992	0.13107	0.48212	0.61727

Table 6.6: First-passage probability to price level 4 obtained from simulation results (95% confidence intervals) and the proposed estimation method.

μ	α	δ	Simulation			Estimated		
			$n = 10$	$n = 50$	$n = 100$	$n = 10$	$n = 50$	$n = 100$
1.00000	0.25000	0.25000	0.07030 ± 0.00275	0.40721 ± 0.00831	0.49192 ± 0.01595	0.07411	0.41592	0.53337
1.00000	0.25000	0.12500	0.04760 ± 0.00224	0.36920 ± 0.00767	0.46799 ± 0.02052	0.05762	0.38639	0.50229
1.00000	0.25000	0.06250	0.01880 ± 0.00138	0.27359 ± 0.00615	0.39742 ± 0.02909	0.03857	0.34296	0.46781
1.00000	0.25000	0.03125	0.00370 ± 0.00061	0.14511 ± 0.00413	0.22535 ± 0.01039	0.02295	0.29264	0.43568
1.00000	0.50000	0.50000	0.07590 ± 0.00287	0.41711 ± 0.00848	0.56250 ± 0.10918	0.07636	0.41956	0.53181
1.00000	0.50000	0.25000	0.04190 ± 0.00209	0.35377 ± 0.00742	0.44348 ± 0.01474	0.05107	0.37287	0.49008
1.00000	0.50000	0.12500	0.01390 ± 0.00119	0.24233 ± 0.00567	0.33465 ± 0.01231	0.02496	0.30036	0.43296
1.00000	0.50000	0.06250	0.00050 ± 0.00022	0.10276 ± 0.00340	0.20700 ± 0.02476	0.00867	0.21313	0.33981
1.00000	1.00000	1.00000	0.08000 ± 0.00295	0.42021 ± 0.00854	0.54504 ± 0.08220	0.07916	0.42398	0.54042
1.00000	1.00000	0.50000	0.04360 ± 0.00214	0.35511 ± 0.00745	0.50389 ± 0.10681	0.04757	0.36511	0.49301
1.00000	1.00000	0.25000	0.01220 ± 0.00111	0.23988 ± 0.00565	0.44531 ± 0.11454	0.01808	0.27146	0.41014
1.00000	1.00000	0.12500	0.00150 ± 0.00039	0.12168 ± 0.00376	0.33331 ± 0.15514	0.00402	0.16207	0.31885
1.00000	2.00000	2.00000	0.08290 ± 0.00301	0.42652 ± 0.00865	0.53889 ± 0.03480	0.08121	0.42715	0.54318
1.00000	2.00000	1.00000	0.04510 ± 0.00217	0.36330 ± 0.00759	0.46546 ± 0.01887	0.04756	0.36509	0.51568
1.00000	2.00000	0.50000	0.01730 ± 0.00133	0.28013 ± 0.00629	0.38502 ± 0.01779	0.01960	0.27852	0.43650
1.00000	2.00000	0.25000	0.00640 ± 0.00080	0.19897 ± 0.00505	0.32336 ± 0.01953	0.00817	0.20883	0.36622
1.00000	4.00000	4.00000	0.08860 ± 0.00312	0.43544 ± 0.00882	0.54966 ± 0.03066	0.08341	0.43047	0.55061
1.00000	4.00000	2.00000	0.04550 ± 0.00218	0.35538 ± 0.00747	0.49836 ± 0.04489	0.04955	0.36956	0.50178
1.00000	4.00000	1.00000	0.02820 ± 0.00170	0.30209 ± 0.00663	0.58014 ± 0.29036	0.02544	0.30214	0.44955
1.00000	4.00000	0.50000	0.02040 ± 0.00144	0.28476 ± 0.00637	0.59241 ± 0.24391	0.02052	0.28255	0.44491

Table 6.7: First-passage probability to price level 6 obtained from simulation results (95% confidence intervals) and the proposed estimation method.

which is assumed to be an independent Poisson process. This establishes the relationship between the microscopic dynamics of the limit order book and the long-term dynamics of the asset price process. Unlike traditional methods that model price dynamics using one-dimensional stochastic processes, the derived model is a two-dimensional stochastic process where the additional dimension represents the latest price change. The parameters of the proposed model can be estimated directly from the order arrival and cancellation rate describing the incoming flow of orders. Additionally, the model also allows us to efficiently compute several quantities of interest of the asset price dynamics (i.e. volatility, return distribution, and first-passage probability) without resorting to simulation.

The proposed parameter estimation method tends to have increased error when the arrival rate of limit orders is small when compared to the arrival rate of market orders as well as when the order cancellation rate is getting smaller. The merit of the proposed framework is the ability to accurately predict the long-term behaviour of the limit order market when we get the right parameters. Hence future work will be focused on developing a better method for estimating the distribution of the number of the orders at the reference price which is the cause of this error. Additionally, we also look forward to investigating the validity of the proposed model by estimating model parameters from a real data set and comparing the prediction with the empirical results.

Chapter 7

Order placement strategy

To illustrate a way to utilise the models of execution probability studied in this dissertation to make order placement decisions, this chapter proposes a new framework for making order placement decisions based on the trade-off between the profit gained from better execution prices and the risk of non-execution that uses the developed execution probability model to balance this trade-off. The result obtained from applying the proposed framework to make order placement for liquidity traders who need to transact their order before the end of a deadline in the historical dataset obtained from the Multi Commodity of India and the New York Stock Exchange indicates that the proposed framework has better performance than the best static order placement strategy for all instruments in the Multi Commodity of India, while it beat the best static strategy only in two out of six cases studied in the New York Stock Exchange. Although the proposed framework cannot beat the best static strategy in all cases, the improvement gained from the proposed framework when it can beat the best static strategy is very significant.

7.1 Introduction

To illustrate the application of the execution probability model developed in this study, this chapter presents and investigates an order placement strategy that uses the developed model to make order placement decisions. This decision is very important especially for traders who trade in limit-order markets, where traders can freely specify the price at which they want to trade. On one hand, traders would prefer to place their orders far away from the current best price as this will increase their payoff. On the other hand, the farther away from the best price, the lower the chance that the order will be executed. Consequently, traders have to find the right tradeoff between these two opposite choices in order to maximise the profit gained from the trade.

In reality, an order submission strategy that a trader selects normally depends on the trading problem he is trying to solve. As suggested by Harris [42], three main trading problems frequently faced by traders are: (i) *the liquidity trader problem* considers how a liquidity trader who must fill his order before some deadlines should trade, (ii) *the informed trader problem* considers how an informed trader who receives a single signal about asset value should trade before his information becomes obsolete and (iii) *the value-*

motivated trader problem considers how a trader who continuously estimates security value should trade. Specifically, liquidity traders must fill their order before some deadline which may arise when they need to invest or disinvest their cash flow. The main objective of these traders is to obtain the best price for their trades by carefully choosing their order submission strategies. On the other hand, informed traders, who have private information about the underlying value of the asset, want to profitably trade on their information. Although informed traders do have a trading deadline, which is the time their information becomes obsolete, they did not have to fill their orders before the deadline like liquidity traders. In fact, informed traders will trade only when it is profitable to do so. Like informed traders, value-motivated traders also have private information about the value of assets. However, unlike informed traders, they do not have a specific deadline and are assumed to trade repeatedly in the market since they receive continuous information about the values.

This chapter presents a framework to solve these decision problems based on the developed execution probability model. Section 7.2 starts the chapter by giving a brief review on previously proposed methods to solve this problem. In Section 7.3, the proposed framework that utilised the developed execution probability model to make order placement decision is presented. Section 7.3 then gives a detailed discussion on how to utilise the proposed framework to make order placement decisions for liquidity traders, as well as analysing the result obtained from applying the proposed framework to make order placement decisions in a historical dataset. Finally, the conclusion of the results investigated in this section are summarised in Section 7.5.

7.2 Previous work

Order placement strategies previously proposed in the literature can be classified into two main categories: (i) static order placement strategies and (ii) dynamic order placement strategies. Static order submission strategies view this problem as a one-shot game where traders can make their order decision only once. If they decide to submit a limit order, no additional change can be made to the order and it will stay in the order book until it is executed or the end of the trading period is reached. Conversely, dynamic order submission strategies allow traders to cancel or make changes to their orders before the order expires or is executed [42, 87]. Empirically, traders change their order submission as market conditions change. They continuously monitor the market and make appropriate changes to their orders whenever necessary. For example, to reduce the execution risk, they may convert their limit orders to market orders when the demand for immediacy increases. They may also reprice or cancel their limit orders when the underlying value of the asset changes to manage the adverse selection cost. Hence, it is more appropriate to model this decision with dynamic strategies than with static strategies.

This section briefly describes related work in order submission strategy. Static order submission strategies are presented in Section 7.2.1, while dynamic strategies are discussed in Section 7.2.2. The overview of all models discussed in this section is summarised in Table 7.1.

Table 7.1: Overview of previous work on order placement strategies. Each model is characterised by the trading problem it tries to solve, whether it is static or dynamic strategy and the market variables that it utilises.

Models	Problem Type			Strategy		Incorporated Variables
	I	II	III	Static	Dynamic	
Handa and Schwartz [41]	x			x		-
Parlour [76]		x		x		volume in the book
Foucault [32]		x		x		volatility
Hollifield et al. [49]		x		x		volatility, order quantity, volume at bid/ask, trading volume, time of day
Foucault et al. [33]		x		x		order arrival rate, spread
Nevmyvaka et al. [73]	x			x		order quantity, time of day, trading volume
Lillo [56]		x		x		volatility
Cohen et al. [19]		x			x	-
Harris [42]	x	x	x		x	volatility
Slive [87]		x			x	volatility, spread
Nevmyvaka et al. [72]	x				x	order quantity, spread, order imbalance, immediate cost, trading volume
Wang and Zhang [92]	x				x	order quantity, order imbalance

7.2.1 Static order submission strategies

As previously discussed, static order submission models consider the decision whether to submit a market order or to submit a limit order as a one-shot game where a trader can make their order decision only once. If the trader decides to submit a limit order, no additional change can be made to his order and it will stay in the order book until it is executed or the end of the trading period is reached. This formulation can be utilised to solve the problems of both liquidity traders [41, 73] and informed traders [19, 76, 32, 49, 33, 56]. The main difference between these two problems is that liquidity traders have to fill their order before the deadline; thus, if liquidity traders decide to submit a limit order and their orders are not executed, they have to submit a market order to execute the trade when their deadlines are approached. On the other hand, informed traders will submit market orders to fill the traders only when it is still profitable to do so.

Static strategies for liquidity traders

Consider the problem of liquidity traders who want to transact their orders before some specific deadlines. Normally, they can choose to submit their order using the following strategies: (i) submitting a market order at the beginning of the time period, (ii) submitting a market order at the end of the time period, and (iii) submitting a limit order at the beginning of the time period and a market order for unexecuted shares at the end of the time period.

Handa and Schwartz [41] analyse the profitability of the third strategy compared to the first one. The limit order strategy that they study is to submit a limit order placed l percent below the current price, where l is set to $0.5, 1, 2$ and 3. The limit order is followed until it executes or until the last price in the trading window is reached. If the limit order does not execute during the trading window, the stock is purchased at the opening price on the day following the trading window. The experimentation results indicate that returns of limit order conditional on execution are positive, while returns of limit order conditional on nonexecution are negative. They also find that picking off risk is not a cost to limit order traders, but that nonexecution is. Thus, it is more appropriate for liquidity trader to transact by the market order strategy, while traders who gain relatively little by trading at current prices (and who are willing to risk not executing) may prefer the limit order strategy.

Another study comparing the profitability of these three choices by Nevmyvaka, Kearns, Papandreou, and Sycara [73] suggests that the limit order strategy performs better than the market order strategy. Although this may seem to contradict [41], they study the limit strategy in finer detail. Specifically, they present a method to estimate return, risk, and risk-return profiles of each strategy from historical data, as well as a method to derive optimal pricing frontiers based on the trade-off between risk and return. Their quantitative method allows traders to optimally price their limit orders to minimise trading costs and control corresponding risks. The importance of a number of microstructure variables (e.g. order size, time window and liquidity) is also highlighted.

Static strategies for informed traders

Unlike liquidity traders, informed traders do not have a responsibility to fill the trade when the deadline is approached. Thus the decision these traders face is simply whether to trade aggressively by submitting a market order or to trade passively by placing a limit order. This static decision problem is usually formalised as an optimisation problem that considers the trade-off between the payoff associated with limit orders and the risk of nonexecution. On one hand, traders would prefer to place their orders very far from the best bid/ask price because this will increase their payoff. On the other hand, the larger the distance from the best price the larger the chance that the order will not be executed. Thus, in this setting, traders have to find the right trade-off between these two opposite choices in order to maximise the expected profit obtained from the trade. This section briefly reviews static order submission strategies for informed traders previously proposed in the literature.

Parlour [76] presents a model of the evolution of the limit order book. The optimal choice between submitting a limit order and a market order is characterised as a single-period optimisation model. The central intuition of her research is that each trader knows that his order will affect the order submission strategies of other traders who follow; thus, he take this effect into account, which in equilibrium, generate systematic patterns in prices and order placement strategies even without asymmetric information. Her study also suggests that both side of the order book are important in determining an agent's order choice.

Foucault [32] describes a game theoretical model of price formation and order submission decisions in a dynamic limit order market where traders arrive sequentially and choose to submit either a market order or a limit order with one-period life. His results indicate that (i) the proportion of limit orders in the order flow is positively related to asset volatility, (ii) the ratio of filled limit orders to total number of limit orders is negatively related to asset volatility, (iii) the proportion of limit orders is positively related to the average size of the spread, (iv) the increase in trading cost at the end of the trading day is negatively related to the level of competition between limit order traders and (v) the size of the sum of trading costs for buy and sell orders is maximised when the ratio of buy to sell orders, is equal to one.

Hollifield, Miller and Sandas [49] present empirical restrictions of a model of optimal order submission in limit order markets. A trader's optimal order submission depends on the trader's valuation of the asset and the trade-offs between order prices, execution probability, and picking off risks. The optimal order submission strategy is a monotonic function of a trader's valuation, characterised in terms of threshold valuations. The threshold valuations are functions of the order prices and the trader's subjective beliefs about the execution probabilities and picking off risks.

Foucault, Kadan and Kandel [33] propose a dynamic model of a limit order market populated by strategic liquidity traders of varying impatience who aim to optimise the trade-off between the cost of delayed execution and the cost of immediacy (the spread). The optimal order submission strategy of each trader is modelled as a single period optimisation problem. Under several simplifying assumptions, they derive the equilibrium order placement strategies. They find that the proportion of patient traders in the population and the order arrival rate are the key determinants of the limit order book dynamics. Traders submit aggressive limit orders, which improve current best quotes by large amounts, when the order arrival rate is low or when the proportion of patient traders is large. As a result, markets with a high proportion of patient traders or a small order arrival rate are more resilient. Also, a reduction in the tick size reduces market resiliency, and, in some case, increases the average spread. Their analysis also yields several testable predictions: (i) a positive relationship between inter-trader durations and market resiliency, (ii) a negative relationship between the order arrival rate and market resiliency, (iii) a joint decline of limit order aggressiveness and market resiliency at the end of the trading session and (iv) limit order traders submit more (less) aggressive orders when the spread is large if patient (impatient) traders dominate the trading population.

Lillo [56] considers the problem of the optimal limit order price for a financial asset in the framework of utility maximisation. The analytical solution of the problem gives insight into the origin of the recently empirically observed power law distribution of limit order prices. In the framework of the model, the most likely proximate cause of this power law is power law heterogeneity of traders' investment time horizons.

7.2.2 Dynamic order submission strategies

Unlike static order submission strategies, dynamic order submission strategies allow traders to monitor the changing market conditions and make changes to their order any time before the order expires or is executed. Empirically, traders change their order submission as market conditions change. They continuously monitor the market and make appropriate changes to their orders whenever necessary. For example, to reduce the execution risk, they may convert their limit orders to market orders when the demand for immediacy increases. They may also reprice or cancel their limit orders when the underlying value of the asset changes to manage the adverse selection cost. Hence, it is more appropriate to model this decision with dynamic strategies than with static strategies. This section presents the existing dynamic strategies previously proposed in the literature.

Cohen, Maier, Schwartz and Whitcomb [19] consider an order submission strategy as a dynamic optimisation problem. Traders in their model may seek to trade via limit order, trade with certainty via a market order, or not to trade at all. Their result demonstrates that transaction costs cause bid-ask spreads to be an equilibrium property of financial markets since, with transaction costs, the execution probability of a limit order does go to unity as the order is placed infinitesimally close to the opposite market quote; thus, with certainty of execution at the opposite market quote, a "gravitational pull" that keeps the opposite quotes from being placed infinitesimally close to each other is generated. They also define an equilibrium market spread and illustrate that it is negatively related to the order arrival rate.

Harris [42] derives optimal dynamic order submission strategies for trading problems faced by three stylised traders: an uninformed liquidity trader, an informed trader and a value-motivated trader. Separate solutions are obtained for quote- and order-driven markets. This results suggest that traders are most aggressive when volatility is high and when their information advantages, if any, are large and decay quickly. Traders are patient when their deadlines are not pressing and when bid/ask spread are wide. The numerical results suggest that most traders should place limit orders close to the market when they trade. Although it may sometimes be optimal for risk neutral traders to place orders far from the market (when deadlines are distant or when private information will not be revealed soon), the expected additional benefits from this strategy are very small. If monitoring costs are high or if the trader is risk averse, distant order placement strategies will not be optimal. The only exception to this rule is for traders who believe that prices are mean-reverting. They may place limit orders far from the market to benefit if prices move far from fundamental values.

Nevmyvaka, Feng and Kearns [72] present the first large-scale empirical application of reinforcement learning to the problem of trade execution. In their problem, the goal is to sell (respectively, buy) V shares of a given stock within a fixed period of time in a manner that maximises the revenue received (respectively, minimises the capital spent). Their results indicate that introducing market variables into the model can greatly improve the execution result and reinforcement learning can indeed result in significant improvement over simpler forms of a single-period optimisation model.

Wang and Zhang [92] present dynamic focus strategies that incorporate a series of market orders of different volume into the limit order strategy and dynamically adjust their volume by monitoring state variables such as inventory and order book imbalance in real-time. The sigmoid function is suggested as the quantitative model to represent the relationship between the state variables and the volume to be adjusted. The empirical results indicate that the dynamic focus strategies can outperform the limit order strategy, which does not adopt dynamic volume adjustment.

Slive [87] derives the optimal dynamic order submission strategies of a trader in a limit order market who has the ability to actively monitor his order and use cancellations and order changes to mitigate the adverse selection and execution risks inherent in limit orders. His results suggest that the ability to implement a dynamic strategy has a large impact on the payoffs to submit limit orders and on limit order submission strategies. After calibrating the parameters to a stock on the Vancouver Stock Exchange, profits from limit order submission are 48% higher when implementing a dynamic strategy compared to a one-shot strategy. Cancellations and order changes are used to avoid adverse selection by moving orders when the underlying value changes. Order changes are used to mitigate execution risk by converting to a market order when the probability of execution declines.

7.3 Framework for an order placement strategy

This section presents a new framework for making order placement decisions which is general enough to solve all three trading problems mentioned above in a mean-variance optimisation framework by extending the model presented in [42]. This framework is based on the trade-off between the profit gained from better execution and the risk of non-execution. As a starting point, we will consider only a static strategy where traders can make decisions only once before the trade begins. After the order is submitted, no additional change can be made to the order until the trading period ends. In particular, we consider an order placement problem of a trader who wants to buy a particular instrument[1] within a time period T in order to maximise his utility function. In this situation, a trader has four possible choices:

1. Do nothing;
2. Execute the order at the beginning of the period at the current market price p_0^M;
3. Execute the order at the end of the period at the closing price p_T^M; or
4. Firstly, submit a limit order to execute a trade at a limit price p^L. If the order is not executed, the trader then either executes the order at the closing price or does nothing depending on his objective.

Although the trader has four choices, these choices can be represented by using only the last strategy. This is because the first and the third choices can be represented by a very low limit price so that the chance that the order will be executed is zero. The second choice can be represented by a limit price higher then the current market price so that the order will be executed immediately. To model these choices, the trader needs to specify his trading objective by defining two payoff functions: a function $U_E(p)$, that defines the payoff he will get when he executes the trade at a price p, and a function $U_{NE}(p)$, that defines the cost he needs to pay if his order is not executed when the asset price at the end of the period is p. Consequently, the payoff the trader will get from submitting a limit buy order at price level p^L is characterised by

$$U(p^L) = \begin{cases} U_E(p^L) & \text{,if the order is executed} \\ U_{NE}(p_T^M) & \text{,if the order is not executed} \end{cases} \tag{7.1}$$

This payoff is a random variable since its value depends on whether the submitted order will be executed or not as well as the price of the asset at the end of the period p_T^M, whose values are not known beforehand. Given the probability that the limit order at price p^L will be executed before the end of the period together with the distribution of the asset price at the end of the period, the expected payoff the trader might get can be computed from

$$E[U(p^L)] = P_E(p^L)U_E(p^L) + [1 - P_E(p^L)] \int_{-\infty}^{\infty} U_{NE}(p) f_{p_M^T|p^L}(p)dp, \tag{7.2}$$

[1] An order placement problem for a trader who want to sell a particular instrument can be formulated in a similar way

where $P_E(p^L)$ is the probability that the limit order at price p^L will be executed before the end of the period, and $f_{p_M|p^L}(.)$ is the probability density function of the asset price at the end of the period given that the limit order at price level p^L is not executed, i.e., the asset price is never lower than or equal to p^L. Similarly, the variance of this utility function can be computed from

$$\begin{aligned} V[U(p^L)] &= P_E(p^L)\big(U_E(p^L) - E[U(p^L)]\big)^2 \\ &\quad + [1 - P_E(p^L)] \int_{-\infty}^{\infty} \big(U_{NE}(p) - E[U(p^L)]\big)^2 f_{p_M^T|p^L}(p)dp, \end{aligned} \tag{7.3}$$

Inserting Equation (7.2) into (7.3), Equation (7.3) can be rewritten as

$$\begin{aligned} V[U(p^L)] &= \big[1 - P_E(p^L)\big] \left[P_E(p^L) \left(U_E(p^L) - \int_{-\infty}^{\infty} U_{NE}(p) f_{p_M^T|p^L}(p)dp \right)^2 \right] \\ &\quad + \big[1 - P_E(p^L)\big] \left[\int_{-\infty}^{\infty} \big[U_{NE}(p)\big]^2 f_{p_M^T|p^L}(p)dp - \left(\int_{-\infty}^{\infty} U_{NE}(p) f_{p_M^T|p^L}(p)dp \right)^2 \right] \end{aligned} \tag{7.4}$$

To perform a mean-variance optimisation, a utility function of a trader who executes his trade by using a limit order at price level p^L could be defined as

$$U_O(p^L) = E[U(p^L)] - \lambda V[U(p^L)], \tag{7.5}$$

where λ is proportional to the trader's risk aversion, or inversely proportional to trader's aggressiveness. As an example, a value of $\lambda = 0$ indicates a trader who is concerned only about the profit gained from better execution while a value of $\lambda = 1$ indicates a trader who equally concerns about the profit gained from better execution and the risk of non-execution. Consequently, the optimal order placement strategy that balances the trade-off between the profit gained from limit orders and the risk of non-execution at the trader's specified level of risk aversion can be determined by maximising the above utility function and can be defined as

$$\begin{aligned} \hat{p} &= \underset{p^L}{\operatorname{argmax}}\ U_O(p^L), \\ &= \underset{p^L}{\operatorname{argmax}}\ E[U(p^L)] - \lambda V[U(p^L)]. \end{aligned} \tag{7.6}$$

By specifying the form of these two utility functions, this framework can be utilised to solve all three trading problems mentioned in the beginning of this chapter. In particular, the problem of liquidity traders who need to transact shares before a deadline can be modelled by setting $U_{NE}(p) = U_E(p)$, so that the cost that a trader needs to pay if his limit order is not executed is equal to the cost of buying at the price at the end of the period. The utility function of informed traders and value motivated traders can be modelled by setting $U_{NE}(p) = \max\{U_E(p), U_{NT}\}$, where U_{NT} is a utility gained from not trading, so that they will trade at the end of the period only when it is profitable to do so. To completely specify this framework, one also needs to specify a model of the probability that the limit order at price level p^L

will be executed before the end of the period, as well as a probability density function of the price at the end of the period given that the limit order submitted at price level p^L is not executed. In the rest of this chapter, we will utilise the price fluctuation model developed in Chapter 5 to model the probability of execution, while utilise the ARMA-GARCH model to model the probability density function of the price at the end of the period. More detail about this will be discussed in the next section where we derive the optimal order placement strategy for liquidity traders.

7.4 Order placement strategy for liquidity traders

This section derives an order placement strategy for liquidity traders who need to transact their orders before some specified deadline from the framework discussed in the previous section. Particularly, we consider a problem of how a trader who want to buy shares should submit his order to execute the trade before the deadline T. In this setting, a trader has three possible choices:

1. Execute the order at the beginning of the trading period, at the current market price p_0^M;
2. Execute the order at the end of the trading period, at the future closing price p_T^M; or
3. Submit a limit order at price level p^L, at the beginning of the trading period. If the order is not executed, the trader then executes the trade at the future closing price p_T^M.

In all these cases, the order is guaranteed to have been executed by the end of the period but the money spent to open this position will be different. The objective of this section is to determine the best way to execute this order to get the most favourable price at a specified risk aversion parameter. To achieve this, Section 7.4.1 firstly derives the model for optimising this decision based on the framework discussed in the previous section. We then describe three different approaches for modelling the execution probability and the probability density of the asset price at the end of the trading period. Particularly, Section 7.4.2 discusses the unconditional model implied from an arithmetic Brownian model. The unconditional empirical model obtained from density estimation is described in Section 7.4.3, while the conditional empirical model using ACD and ARMA-GARCH model is presented in Section 7.4.4. Finally, the results obtained from applying these model to make trading decision are analysed in Section 7.4.5.

7.4.1 Order placement model for liquidity traders

To utilise the proposed framework to make order placement decisions for liquidity traders, one needs to specify a utility function $U_E(p)$ that defines the payoff the trader will get from executing a trade at price p, a utility function $U_{NE}(p)$ that defines the cost the trader will pay if his order is not executed and the price at the end of the period is p, a probability model $P_E(p)$ that describes the probability that the order submitted at price p will be executed before the end of the period, and a probability density function $f_{p_T^M|p^L,p_0^M}(p_T^M)$ that describes the probability that the closing price will be p_T^M, if the limit order submitted at price p^L is not executed. In this setting, we specify the payoff that the trader will get from executing his order at price p^L as the difference between the current best ask price and the execution price, or equivalently, we set

$$U_E(p^L) = p_0^M - p^L, \tag{7.7}$$

where p_0^M is the market price at the beginning of the trading period. Similarly, the payoff that the trader will get if his order is not executed, when the price at the end of the period is p_T^M, is defined as

$$U_{NE}(p_T^M) = p_0^M - p_T^M, \tag{7.8}$$

which is basically the profit gained from executing the order at the end of the trading period. Although the payoff the trader gets from executing his order at limit price p^L is always greater than zero, as the limit price p^L must be lower than p_0^M, the payoff that the trader will get if his order is not executed can be both positive and negative. In fact, this payoff will be negative if the asset price increases at the end of the trading period, while it will be positive if the asset price decreases at the end of the period. Consequently, when the limit order submitted by the trader is not executed, the trader will still gain price improvement if the asset price decreases and he will suffer a loss only when the asset price increases.

The probability that the limit order at each price level will be executed and the distribution of the asset price at the end of the trading period, given that the limit order place at price level p^L is not executed, can be modelled by several approaches. In this section, we will discuss three different approaches for modelling these two distributions, which are the distribution implied by an arithmetic Brownian motion, the unconditional model that utilises the empirical distribution, and the conditional model that uses the ACD model and ARMA-GARCH model.

7.4.2 Unconditional model implied by the arithmetic Brownian motion

The first and only theoretical model considered here is the model implied by an arithmetic Brownian motion. In this situation, the asset price is assumed to follow the arithmetic Brownian motion and a limit order is assumed to be executed when the asset price hits or crosses the limit price. Particularly, the asset price p_t is assumed to follow

$$dp_t = \sigma dW_t + \mu dt, \tag{7.9}$$

where μ is the instantaneous drift, σ is the standard deviation, and W_t is a Wiener process. In the context of a buy problem, p_t is best thought of as the best ask price and the limit order submitted at price level p^L will be executed before the end of the trading period T only when $p_t \leq p^L$ for some time t, $0 \leq t \leq T$. As discussed in Section 5.4.1, the probability that the limit buy order submitted at price level $p^L < p_0$ will be executed before the end of the period T can be obtained from

$$P_E(p^L|p_0) = \Phi\left(\frac{p^L - p_0 - \mu T}{\sigma\sqrt{T}}\right) + \exp\left(\frac{2\mu(p^L - p_0)}{\sigma^2}\right)\Phi\left(\frac{p^L - p_0 + \mu T}{\sigma\sqrt{T}}\right), \tag{7.10}$$

where $\Phi(.)$ is the cumulative distribution function of a standard Normal distribution. When there are no constraints, the distribution of the asset price at the end of the trading period is simply a Normal distribution with mean μT and variance $\sigma^2 T$. However, the condition that the limit order at price level p^L is not executed constrains p_t to be higher than p^L for all $0 < t < T$. This means that the distribution of p_T is not simply a Normal distribution truncated on the left at p^L but is a more complicated expression given by (see [44] page 165)

$$f_{p_T|p^L,p_0}(p_T) = \frac{\phi\left(\frac{p_0 - p_T + \mu T}{\sigma\sqrt{T}}\right) - \exp\left(\frac{2\mu(p^L - p_0)}{\sigma^2}\right)\phi\left(\frac{2p^L - p_0 - p_T + \mu T}{\sigma\sqrt{T}}\right)}{\sigma\sqrt{T}\big(1 - P_E(p^L|p_0)\big)}, \tag{7.11}$$

where $\phi(.)$ is the probability density function of a standard Normal distribution. To gain more insight into the property of this distribution, Figure 7.1 plots this distribution in several parameter settings. The results displayed in the first column indicates that the distribution of asset prices at the end of the trading period given no execution of limit order at price level p^L is generally right-skewed and tends to shift upward to the right when p^L increases. This suggests that this distribution typically has higher mean and lower variance as p^L increases. The results displayed in the second and third columns convey similar outcomes as this distribution tends to shift downward to the right when the drift parameter and the volatility parameter increases. This suggests that this distribution generally has higher mean and higher variance when the drift and volatility parameter of the arithmetic Brownian motion increase.

By inserting Equation (7.10) and (7.11) into Equation (7.2) and (7.4), one can compute the expectation and the variance of the payoff the trader will get from executing his trade using a limit order at each price level under the arithmetic Brownian motion assumption. Consequently, the optimal order placement strategy for a specified level of risk aversion parameter can be derived accordingly by finding the limit order price that maximises the utility function described in Equation (7.5). To gain more insight into the optimal order placement strategy generated from this model, Figure 7.2 displays the expectation and the variance of this payoff function at several parameter settings. The result displayed in the first row of this figure indicates that this expectation is a monotonic decreasing function when the drift parameter is negative, a straight horizontal line when the drift parameter is zero, and a monotonic increasing function when the drift parameter is positive. This indicates that the optimal order placement strategy of a trader whose only concern is the expected profit gained from limit order trading depends heavily on the drift parameter of the arithmetic Brownian motion in the sense that it is always optimal to execute the order immediately at the beginning of the trading period when the drift parameter is positive, while it is

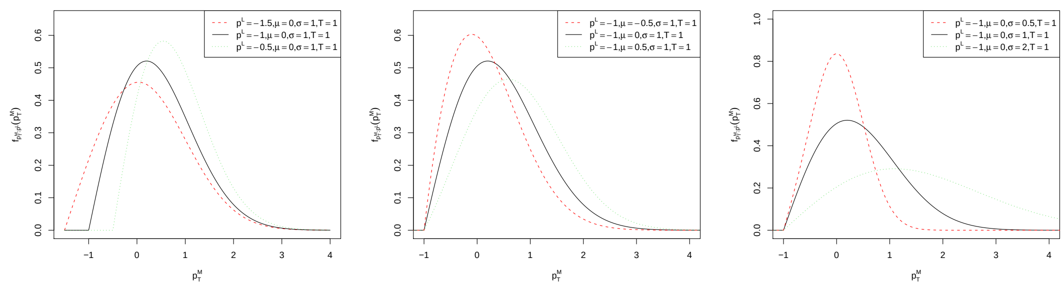


Figure 7.1: The probability density function of the asset price at the end of the trading period T given no execution of the limit order at price level p^L when the asset price is assumed to follow the arithmetic Brownian motion with the drift parameter equal to μ and the volatility parameter equal to σ at several parameter settings.

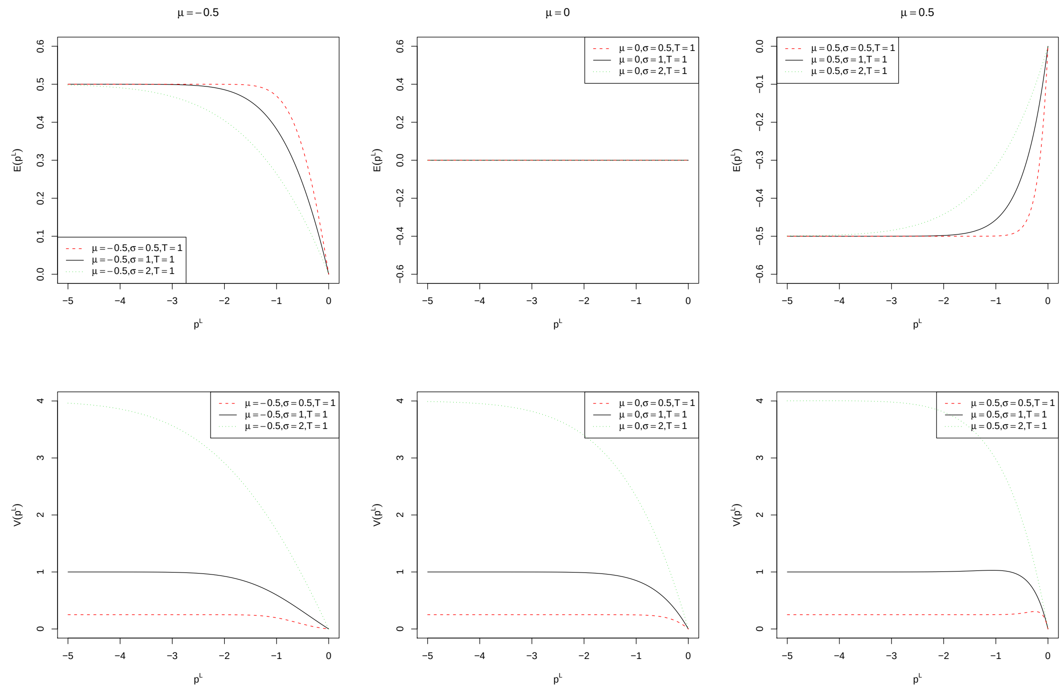


Figure 7.2: The expectation (top row) and the variance (bottom row) of the profit that liquidity traders will get from executing their trade using limit order at price level p^L when the asset price is assumed to follow the arithmetic Brownian motion with the drift parameter equal to μ and the volatility parameter equal to σ at several parameter settings.

always optimal to execute the order at the end of the period when the drift is negative. When the drift parameter is zero, all trading strategies are optimal since the expectation of this payoff is the same for all price levels. Additionally, the variance of this payoff function, displayed in the second row of Figure 7.2 indicates that the variance of this payoff is always a monotonic decreasing function with a minimum at zero. This suggests that it is always optimal to execute the trade immediately at the beginning of the trading period if the trader's objective is to minimise the risk of non-execution.

To analyse the optimal strategy of a trader who is concerned about the trade-off between the profit gained from limit order trading and the risk of non-execution, we plot the combined utility function at three different risk aversion levels when the drift parameter is negative in Figure 7.3[2]. The result indicates that the combined utility function can be monotonic increasing, monotonic decreasing as well as unimodal depending on the parameters of the model. Unlike the optimal strategy of the trader whose only concern is the expected payoff or the risk of non-execution, the optimal strategy in this case also involves the use of limit orders especially when the combined utility function is unimodal. This also illustrates that the optimal strategy depends heavily on the parameters of the model, as the optimal limit price tends to move towards zero, which results in immediate execution, when trader's risk aversion level or the volatility parameter of the arithmetic Brownian motion increases.

[2]We only analyse the case when the drift parameter is negative since it is always optimal to execute the trade immediately at the beginning of the trading period when the drift parameter is positive.

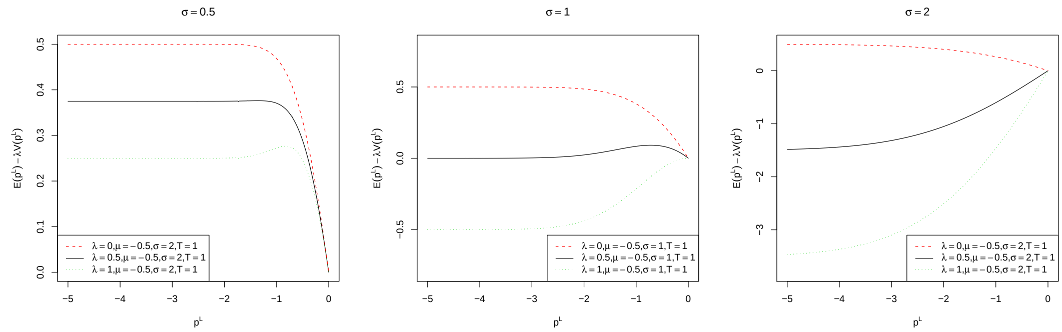


Figure 7.3: The utility function that liquidity traders will get from executing their trade using limit order at price level p^L at three different risk aversion levels, $\lambda = 0.1, 0.5, 1.0$, when the asset price is assumed to follow the arithmetic Brownian motion with drift parameter equal to μ and the volatility parameter equal to σ at several parameter settings.

7.4.3 Empirical unconditional model using density estimation

Since the results analysed in Section 5.4.1 indicates that the probability execution model implied by the arithmetic Brownian motion does not fit very well with the empirical distribution, it might be more appropriate to make order placement decision from the empirical distributions. To achieve this, this section presents a method for estimating the unconditional distribution of the execution probability and the asset price at the end of the trading period given no execution from the history of asset price. In particular, we estimated the probability that the limit order submitted at each price level will be executed before the end of the trading period is estimated from the empirical distribution of price fluctuations as discussed in Section 5.2, while the probability density function of future closing price given no execution will be estimated from the empirical distribution of asset returns at the end of the trading period.

Before discussing a method for modelling these two distributions from asset price history, we will firstly review the relationship between price fluctuations and the probability of execution as discussed in Section 5.2. Define a buy price fluctuation during time T, denoted by M_T^B, as the difference between the initial price level and the lowest price level reached during the trading period T, or equivalently

$$M_T^B = \sup\{p_0 - p_t;\ 0 \le t \le T\}. \tag{7.12}$$

Accordingly, the probability that a limit buy order submitted at price level $p^L < p_0$ will be executed before the end of the period T can be estimated from

$$\begin{aligned} P_E(p^L|p_0) &= \Pr\big\{\sup\{p_0 - p_t;\ 0 \le t \le T\} \ge p_0 - p^L\big\}, \\ &= \Pr\big\{M_T^B \ge p_0 - p^L\big\}, \\ &= 1 - F_{M_T^B}(p_0 - p^L), \end{aligned} \tag{7.13}$$

where $F_{M_T^B}(.)$ is a cumulative distribution function of the buy price fluctuations during time period T. Given a history of buy price fluctuations $(\Delta_1, \ldots, \Delta_2)$, this unconditional cumulative distribution function can be estimated from

$$F_{M_T^B}(\Delta) = \frac{1}{N}\sum_{i=1}^{N} I\{\Delta_i \leq \Delta\}, \tag{7.14}$$

where $I\{.\}$ is an indicator function which is equal to one when the enclosed expression is true, while it will be zero when the enclosed expression is false. Inserting Equation (7.14) into (7.13), the probability that the limit order at price level p^L will be executed before the end of the period can be estimated from

$$P_E(p^L|p_0) = 1 - \frac{1}{N}\sum_{i=1}^{N} I\{\Delta_i < p_0 - p^L\}. \tag{7.15}$$

Let us now consider a method to estimate the probability density function of the asset price at the end of the trading period given no execution of the limit order at price level p^L from the empirical distribution of asset returns at the end of the period. Given a history of asset price dynamics sampling every T unit time $(p_0, \ldots, p_N)$, the unconditional empirical distribution of asset returns at the of the trading period T can be estimated from

$$f_{r_T}(r) = \frac{1}{N}\sum_{i=1}^{N} I\{p_i - p_{i-1} = r\}. \tag{7.16}$$

Consequently, under the assumption that the asset price at the end of the period is independent of the execution of limit order at price level p^L, the distribution of asset prices at the end of the period can be estimated from

$$f_{p_T^M|p_0^M}(p_T^M) = f_{r_T}(p_T^M - p_0^M) = \frac{1}{N}\sum_{i=1}^{N} I\{p_i - p_{i-1} = p_T^M - p_0^M\}. \tag{7.17}$$

However, this assumption is generally violated in most situations since the condition that a limit order at price level p^L is not executed constrains the asset price to be higher than p^L for all time. This suggest that it might not be appropriate to estimate the probability density function of the asset price at the end of the trading period given no execution of limit order at price level p^L using the above equation. To improve this estimation, we choose to model this distribution using a left-truncated version of the above distribution which can be computed from

$$\begin{aligned} f_{p_T^M|p^L,p_0^M}(p_T^M) &= f_{r_T}(p_T^M - p_0^M | p_T^M \geq p^L) \\ &= f_{r_T}(p_T^M - p_0^M | p_T^M - p_0^M \geq p^L - p_0^M) \\ &= \frac{\sum_{i=1}^{N} I\{p_i - p_{i-1} = p_T^M - p_0^M\}}{\sum_{i=1}^{N} I\{p_i - p_{i-1} \geq p^L - p_0^M\}}. \end{aligned} \tag{7.18}$$

Since the distribution considered in this section is discrete in nature, the expectation of the payoff function in Equation (7.2) is changed to

$$E[U(p^L)] = P_E(p^L)U_E(p^L) + [1 - P_E(p^L)]\sum_{i=-\infty}^{\infty} U_{NE}\left(p_i^L + i\delta\right) f_{p_T^M|p^L,p_0^M}\left(p^L + i\delta\right), \tag{7.19}$$

where δ is the tick size of the instrument considered. Similarly, the variance of this payoff in Equation (7.4) is changed to

$$\begin{aligned} V[U(p^L)] \quad = \quad & \left[1 - P_E(p^L)\right] \left[P_E(p^L) \left(U_E(p^L) - \sum_{i=-\infty}^{\infty} U_{NE}(p^L + i\delta) f_{p_T^M | p^L, p_0^M}(p^L + i\delta) \right)^2 \right] \\ & + \left[1 - P_E(p^L)\right] \left[\sum_{i=-\infty}^{\infty} [U_{NE}(p^L + i\delta)]^2 f_{p_T^M | p^L, p_0^M}(p^L + i\delta) \right] \\ & - \left[1 - P_E(p^L)\right] \left[\sum_{i=-\infty}^{\infty} U_{NE}(p^L + i\delta) f_{p_T^M | p^L, p_0^M}(p^L + i\delta) \right]^2 . \end{aligned} \tag{7.20}$$

Consequently, one can determine the optimal order placement strategy for a liquidity trader with a specified risk aversion level λ by finding a limit order p^L that maximises the combined utility function

$$\hat{p}^L = \underset{p^L}{\operatorname{argmax}} \left(E[U(p^L)] - \lambda V[U(p^L)] \right). \tag{7.21}$$

7.4.4 Empirical conditional model using ACD and ARMA-GARCH models

Whilst the models considered in the previous two sections are mainly focused on the unconditional distribution of the probability that the limit order will be executed before the deadline and the unconditional distribution of the asset price at the of the trading period, it might be more useful to consider conditional models of these two distribution so that the optimal order placement strategy can adapt according to current market situations rather than fixing at a specified price level. To achieve this, this section presents a method for estimating the conditional distribution of the execution probability using the Autoregressive Condition Duration (ACD) model as described in Section 5.5.3, while modelling the distribution of asset price at the end of the trading period using the Autoregressive Moving Average - Generalised Autoregressive Conditional Heteroskedasticity (ARMA-GARCH) model.

Let $(\Delta_1, \ldots, \Delta_N)$ be a sequence of price fluctuations during the considered time period. The ACD(p, q) model for price fluctuation, as discussed in Section 5.5.3 and 5.6, can be specified by

$$\begin{aligned} \Delta_i &= \hat{\Delta}_i s(t_i), \\ \hat{\Delta}_i &= \psi_i \epsilon_i, \\ \psi_i &= \Psi\left(\hat{\Delta}_{i-1}, \ldots, \hat{\Delta}_{i-p}, \psi_{i-1}, \ldots, \psi_{i-q}\right), \end{aligned}$$

where $\hat{\Delta}_i$ is the seasonally adjusted price fluctuations at the i-th time step, $s(t_i)$ is the seasonality component at the i-th time step, ψ_i is the conditional expectation of price fluctuations at the i-th time step, $\{\epsilon_i\}$ is an independent and identically distributed white noise with unit mean and finite variance, and $\Psi(.)$ is the conditional mean function. If $\{\epsilon_i\}$ has a cumulative distribution function $F_\epsilon(.)$, the conditional distribution of price fluctuations at the i-th time step, when the tick size of the considered instrument is δ, will be given by

$$f_{ACD}(\Delta_i | \mathcal{F}_{i-1}) = F_\epsilon\left(\frac{\Delta_i + \delta}{\psi_i s(t_i)}\right) - F_\epsilon\left(\frac{\Delta_i}{\psi_i s(t_i)}\right), \tag{7.22}$$

and the probability that the limit order at price level p_i^L, submitted at the i-th time step will be executed before the end of the trading period can be computed from

$$\begin{aligned} P_E(p_i^L|p_{i-1}) &= 1 - \Pr\left\{\Delta_i < p_{i-1} - p_i^L\right\}, \\ &= 1 - \sum_{i=-\infty}^{(p_{i-1}-p_i^L)/\delta - 1} f_{ACD}(i\delta), \\ &= 1 - F_\epsilon\left(\frac{p_{i-1} - p_i^L}{\psi_i s(t_i)}\right). \end{aligned} \tag{7.23}$$

In the rest of this section, we will assume that the conditional mean function has the form of the ABAMACD model which can be specified by

$$\begin{aligned} \psi_i^{\delta_1} &= \omega + \sum_{j=1}^{p} \alpha_j \psi_{i-j}^{\delta_1} \left(|\epsilon_{i-j} - b| + c_j(\epsilon_{i-j} - b)\right)^{\delta_2} \\ &\quad + \sum_{j=1}^{p} \lambda_j \left(|\epsilon_{i-j} - b| + c_j(\epsilon_{i-j} - b)\right)^{\delta_2} + \sum_{j=1}^{q} \beta_j \psi_{i-j}^{\delta_1}, \end{aligned} \tag{7.24}$$

where $\theta_\Psi = (\delta_1, \delta_2, b, c_1, \ldots, c_p, \alpha_1, \ldots, \alpha_p, \lambda_1, \ldots, \lambda_p, \beta_1, \ldots, \beta_q)$ are the parameters of the model. The seasonality component at the i-th time step will be specified by a time-of-day equation

$$s(t_i) = \nu_1 I_1(t_i) + \nu_2 I_2(t_i) + \ldots + \nu_s I_s(t_i), \tag{7.25}$$

where $\theta_s = (\nu_1, \ldots, \nu_s)$ are the parameters to be estimated and $I_k(t_i)$ is an indicator function whose value can be either one or zero indicating whether the time t_i is in a particular time interval of the day or not. Additionally, the independent and identically distributed white noise $\{\epsilon_i\}$ will be assumed to follow the unit generalised gamma distribution whose cumulative density function has the form

$$F_{GG}(\epsilon) = \frac{\gamma\left(\lambda, [\epsilon\Gamma(\lambda + 1/\kappa)/\Gamma(\lambda)]^\kappa\right)}{\Gamma(\lambda)}, \tag{7.26}$$

where $\theta_\epsilon = (\kappa, \lambda)$ are parameters of the generalised gamma distribution, $\gamma(\lambda, x) = \int_0^x t^{\lambda-1} e^{-t} dt$ is a lower incomplete gamma function, and $\Gamma(\lambda) = \int_0^\infty t^{\lambda-1} e^{-t} dt$ is a gamma function. Consequently, given a time-series of price fluctuations $(\Delta_1, \ldots, \Delta_2)$ the parameters of this model can be estimated by maximising the log-likelihood of the probability density function described in Equation (7.22), or, equivalently, the maximum likelihood estimator $\hat{\theta} = (\hat{\theta}_\Psi, \hat{\theta}_s, \hat{\theta}_\epsilon)$ of this model can be obtained from

$$\begin{aligned} \hat{\theta} &= \underset{\theta \in (\theta_\Psi, \theta_s, \theta_\epsilon)}{\operatorname{argmax}} \sum_{i=p+1}^{N} \ln f_{ACD}(\Delta_i|\mathcal{F}_{i-1}), \\ &= \underset{\theta \in (\theta_\Psi, \theta_s, \theta_\epsilon)}{\operatorname{argmax}} \sum_{i=p+1}^{N} \ln\left(F_{GG}\left(\frac{\Delta_i + \delta}{\psi_i s(t_i)}\right) - F_{GG}\left(\frac{\Delta_i}{\psi_i s(t_i)}\right)\right), \end{aligned} \tag{7.27}$$

where ψ_i are computed from Equation (7.24) using the observations $\Delta_1, \ldots, \Delta_N$ with $\psi_i = \hat{\Delta}$ for all $i \leq 0$ when $\hat{\Delta}$ is the sample mean of $(\Delta_1, \ldots, \Delta_N)$.

To model the probability density function of the asset price at the end of the i-th time step given no execution of limit order at price level p_i^L, we will utilise the ARMA-GARCH model to model the return at the i-th time step, and, then, estimate the density of the asset price at the end of the i-th time step given no execution of limit order at price level p^L from the density of the return at the i-th time step, as discussed in the previous section. Particularly, let $(p_0, \ldots, p_N)$ be a series of asset price sampling every T unit time. We model the return of the i-th time step, $r_i = p_i - p_{i-1}$, using the ARMA(p,q)-GARCH(r,s) model of the form

$$\begin{aligned}
r_i &= \hat{r}_i + s_r(t_i), \\
\hat{r}_i &= \omega + \sum_{j=1}^{p} \alpha_j r_{i-j} + \sum_{j=1}^{q} \beta_j z_{i-j} + z_i, \\
\sigma_i^2 &= \gamma + \sum_{j=1}^{r} \theta_j z_{i-j}^2 + \sum_{j=1}^{s} \lambda_j \sigma_{i-j}^2, \\
z_i &= \sigma_i \epsilon_i, \qquad (7.28)
\end{aligned}$$

where $\theta_a = (\omega, \gamma, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \theta_1, \ldots, \theta_r, \lambda_1, \ldots, \lambda_s)$ are the parameters of the model, $\{\epsilon_i\}$ is an independent and identically distributed white nose with zero mean and unit variance, $\hat{r}_i$ is the seasonality adjusted return at the i-th time step, and $s_r(t_i)$ is the seasonality component at the i-th time step which has the form

$$s_r(t_i) = \nu_1 I_1(t_i) + \nu_2 I_2(t_i) + \ldots + \nu_s I_s(t_i), \qquad (7.29)$$

where $\theta_s = (\nu_1, \ldots, \nu_s)$ are the parameters to be estimated and $I_k(t_i)$ is an indicator function whose value can be either one or zero indicating whether the time t_i is in a particular time interval of the day or not, similar to the seasonality component of the ACD model. If the white noise $\{\epsilon_i\}$ has a probability density function $f_\epsilon(.)$, the conditional distribution of the asset return at the i-th time step can be computed from

$$\begin{aligned}
f_{r_i}(r_i|\mathcal{F}_{i-1}) &= \frac{1}{\sigma_i} f_\epsilon \left(\frac{z_i}{\sigma_i} \right), \\
&= \frac{1}{\sigma_i} f_\epsilon \left(\frac{r_i - (\omega + \sum_{j=1}^{p} \alpha_j r_{i-j} + \sum_{j=1}^{q} \beta_j z_{i-j} + s(t_i))}{\sigma_i} \right). \qquad (7.30)
\end{aligned}$$

Consequently, under the assumption that the asset price at the end of the period is independent of the execution of limit orders at price level p^L, the distribution of the asset price at the end of the i-th time step given no execution of limit order at price level p_i^L can be estimated from

$$\begin{aligned}
f_{p_i}(p_i|\mathcal{F}_{i-1}) &= f_{r_i}(p_i - p_{i-1}|\mathcal{F}_{i-1}) \\
&= \frac{1}{\sigma_i} f_\epsilon \left(\frac{p_i - p_{i-1} - (\omega + \sum_{j=1}^{p} \alpha_j r_{i-j} + \sum_{j=1}^{q} \beta_j z_{i-j} + s(t_i))}{\sigma_i} \right), \qquad (7.31)
\end{aligned}$$

while the probability density function of the asset price at the end of the i-th time step given no execution of limit order at price level p_i^L, using the left-truncated version of the distribution above can be computed from

$$\begin{aligned} f_{p_i|p^L}(p_i|\mathcal{F}_{i-1}) &= f_{r_i}\left(p_i - p_{i-1}|p_i \geq p^L, \mathcal{F}_{i-1}\right), \\ &= \frac{f_{r_i}\left(p_i - p_{i-1}|\mathcal{F}_{i-1}\right)}{1 - \int_{-\infty}^{p^L} f_{r_i}\left(p - p_{i-1}|\mathcal{F}_{i-1}\right) dp}, \\ &= \frac{f_\epsilon\left(\left(p_i - p_{i-1} - \left[\omega + \sum_{j=1}^{p} \alpha_j r_{i-j} + \sum_{j=1}^{q} \beta_j z_{i-j} + s(t_i)\right]\right)/\sigma_i\right)}{\sigma_i\left[1 - F_\epsilon\left(\left(p_i - p_{i-1} - \left[\omega + \sum_{j=1}^{p} \alpha_j r_{i-j} + \sum_{j=1}^{q} \beta_j z_{i-j} + s(t_i)\right]\right)/\sigma_i\right)\right]}, \end{aligned} \tag{7.32}$$

where $F_\epsilon(.)$ is the cumulative distribution function of the independent and identically distributed white noise $\{\epsilon_i\}$. In the rest of this section, we will assume that distribution of the white noise $\{\epsilon_i\}$ follows a standard Normal distribution whose probability density function and cumulative distribution function are

$$f_{NN}(\epsilon) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \text{ and } F_{NN}(\epsilon) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right], \tag{7.33}$$

where $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ is the Gauss error function. Consequently, given a sequence of asset prices $p_0, \ldots, p_N$, the parameter of this ARMA-GARCH model can be estimated by maximising the log-likelihood implied by the probability density function described in Equation (7.31), or equivalently, the maximum likelihood estimator $\hat{\theta} = (\theta_a, \theta_s)$ of this ARMA(p, q)-GARCH(r, s) model can be computed from

$$\hat{\theta} = \underset{\theta \in (\theta_\Psi, \theta_s, \theta_\epsilon)}{\text{argmax}} \sum_{i=p+1}^{N} \left[\ln f_{NN}\left(\frac{z_i}{\sigma_i}\right) - \ln \sigma_i\right], \tag{7.34}$$

where z_i are computed from Equation (7.28) using the observation $p_0, \ldots, p_N$ with $z_i = \mathbb{E} z_i = 0$ for $i = \min(p - q + 1, p), \ldots, p$. Since the distribution considered in this section is discrete, the expectation of the payoff function for executing the trade using limit orders at price p_i^L at the i-th time step is

$$\begin{aligned} E[U(p_i^L)] = \; & P_E(p_i^L|p_{i-1}) U_E(p_i^L) \\ & + \left[1 - P_E(p_i^L|p_{i-1})\right] \sum_{j=-\infty}^{\infty} U_{NE}\left(p_i^L + j\delta\right) f_{p_i|p_i^L}\left(p_i^L + j\delta\right), \end{aligned} \tag{7.35}$$

where δ is the tick size of the instrument considered. Similarly, the variance of this payoff at the i-th time step is characterised by

$$\begin{aligned} V[U(p_i^L)] \quad = \quad & \left[1 - P_E(p_i^L|p_{i-1})\right] \left[P_E(p_i^L|p_{i-1}) \left(U_E(p_i^L) - \sum_{j=-\infty}^{\infty} U_{NE}(p^L + i\delta) f_{p_i|p_i^L}\left(p_i^L + j\delta\right) \right)^2 \right] \\ & + \left[1 - P_E(p_i^L|p_{i-1})\right] \left[\sum_{j=-\infty}^{\infty} [U_{NE}(p_i^L + j\delta)]^2 f_{p_i|p_i^L}\left(p_i^L + j\delta\right) \right] \\ & - \left[1 - P_E(p_i^L|p_{i-1})\right] \left[\sum_{j=-\infty}^{\infty} U_{NE}(p_i^L + j\delta) f_{p_i|p_i^L}\left(p_i^L + j\delta\right) \right]^2 . \end{aligned} \quad (7.36)$$

Consequently, one can determine the optimal order placement strategy for a liquidity trader with a specified risk aversion level λ at the i-th time step by finding a limit order p^L that maximises the combined utility function

$$\hat{p}_i^L = \underset{p_i^L}{\operatorname{argmax}}\, E[U(p_i^L)] - \lambda V[U(p_i^L)]. \quad (7.37)$$

7.4.5 Experimentation results

This section investigates the performance of applying the proposed model to make order placement decisions for liquidity traders by utilising the historical data from the Multi Commodity Exchange of India (MCX) and the New York Stock Exchange (NYSE) as described in Section 5.3.1. In particular, we focus here on six instruments, which are the gold, silver, and natural gas futures contracts traded at the MCX and the GE, IBM and Microsoft stock traded at the NYSE. To measure the performance of the proposed order placement model, the trading period, T, for the instruments in the MCX is set to five minutes, while the trading period for the instruments in the NYSE is set to ten minutes. To reveal any bias in our results, we test our model on both a buy and a sell problem. We also vary the risk aversion parameters at several levels to analyse the effect of this parameter on the trading performance.

Since the order placement strategy obtained from unconditional models is static in the sense that they will execute the trade at the same limit order price in all situations, we will not analyse their results directly but will represent them using the static strategy that has the best performance in each dataset. However, we will replace some components of the conditional model discussed in Section 7.4.4 with the corresponding component of the unconditional model discussed in Section 7.4.3 with the aim of identifing the improvement gained from the conditional model. In particular, we will perform the experiment with three order placement models which are: i) the ACD-ARMA model that models the probability of execution and the density of the asset price at the end of the trading period using the ACD model and ARMA-GARCH model, respectively, ii) the ACD-DENSITY model that models the probability of execution using the ACD model, while modeling the density of future closing prices using density estimation, and iii) DENSITY-ARMA that utilises the density estimation to model the probability of execution but utilises the ARMA-GARCH model to model the density of the future closing prices. Additionally, since the probability density of the asset price at the end of the trading period given no execution of limit order can be estimated under both independent and truncated assumptions, for each of the three models, we also estimate this distribution under both assumptions. As a result, this section will investigate six order placement models as summarised in Table 7.2. For each of these models, we firstly estimate its parameters from the first 75% of the dataset using the method described in Section 7.4.3 and 7.4.4. We then perform an experiment to find the risk aversion level that produces the best result in the first 75% of the dataset, and utilise the best risk aversion level to perform trading in the last 25% of the dataset. Finally, we compare the performance of the best trading strategy obtained from each model in order to determine the best model for making order placement decisions.

Trading with no market variables

Let us firstly analyse the results obtained from applying the ACD and ARMA-GARCH model without any market variables. Table 7.3 displays the trading results for the IBM stock in NYSE market using the above models at several risk aversion levels. The results indicate that the risk aversion level that provide the best profit is not zero, but generally higher than that. However, the result clearly confirms the important of the risk aversion parameters in controlling the risk of non-execution as the variance of the profit gained from using the strategy is generally lower when the risk aversion parameter is higher.

Table 7.4 and Table 7.5 report the performance of the best risk aversion level for each of the six models together with the performance of the static strategy that always executes the trade immediately at the beginning of the period, the static strategy that always executes the trade at the end of the trading period and the static strategy that always executes the trade at the best level. The result indicates that all of the six models beat the strategy that executes the trade at the beginning and the end of the trading period in all situations both in the training and testing dataset. Additionally, the best of these six models can also beat the best static strategy in all instruments in the MCX market, while, in the NYSE market, it can beat the static strategy only in the IBM stock. Although our proposed model cannot beat the best static strategy in all cases, the improvement gained from our model when it does beat the best static strategy is significant and ranges from 0.1% to 6.4% in the training dataset and 0.1% to 14.2% in the testing dataset. However, no clear winning models can be identified since the best models vary considerably from case to case. This suggests that it is more appropriate to try all possible models and select the best performing one to make trading decisions.

To compare the performance of these models, we calculate the profit gained (in number of ticks) from these strategies over immediately executing the trade at the beginning of the period averaged over all instruments in the same markets. The result reported in Table 7.6 indicates that the ACD-ARMA model with independent assumption seems to be the best performing model with 26.25% and 15.72%

Table 7.2: List of models for making order placement decisions.

Short name	Model
T-ACD-ARMA	The ACD-ARMA model with truncated assumption
I-ACD-ARMA	The ACD-ARMA model with independent assumption
T-DEN-ARMA	The DENSITY-ARMA model with truncated assumption
I-DEN-ARMA	The DENSITY-ARMA model with independent assumption
T-ACD-DEN	The ACD-DENSITY model with truncated assumption
I-ACD-DEN	The ACD-DENSITY model with independent assumption

Table 7.3: The performance of the proposed order placement strategies at several risk aversion levels when used to make order placement decisions for the IBM stock traded in the New York Stock Exchange. The performance reported includes the probability that the submitted order is executed (P_E), the average profit obtained when the submitted order is executed ($E(U_E)$), the average loss incurred when the submitted order is not executed ($E(U_{NE})$), the average profit obtained ($E(U)$), and the variance of the profit ($V(U)$).

	Buy						Sell					
λ	P_E	$E(U_E)$	$E(U_{NE})$	$E(\Delta)$	$E(U)$	$V(U)$	P_E	$E(U_E)$	$E(U_{NE})$	$E(\Delta)$	$E(U)$	$V(U)$
Truncated ACD-ARMA Model												
0.0	0.49	0.0499	-0.0406	0.37	**0.0035**	0.0381	0.38	0.1604	-0.0927	0.15	0.0041	0.0401
0.5	0.86	0.0230	-0.1196	0.05	0.0025	0.0118	0.41	0.1485	-0.0963	0.14	0.0047	0.0373
1.0	0.93	0.0135	-0.1635	0.02	0.0019	0.0062	0.44	0.1387	-0.0987	0.13	**0.0047**	0.0350
1.5	0.96	0.0083	-0.1863	0.01	0.0013	0.0034	0.46	0.1303	-0.1007	0.12	0.0045	0.0334
2.0	0.98	0.0054	-0.1945	0.01	0.0009	0.0022	0.47	0.1223	-0.1022	0.11	0.0042	0.0312
2.5	0.99	0.0035	-0.2069	0.00	0.0005	0.0015	0.49	0.1151	-0.1031	0.11	0.0044	0.0294
3.0	0.99	0.0023	-0.2111	0.00	0.0001	0.0011	0.51	0.1078	-0.1030	0.10	0.0043	0.0273
3.5	0.99	0.0016	-0.2227	0.00	0.0000	0.0008	0.52	0.1011	-0.1026	0.10	0.0043	0.0255
4.0	1.00	0.0011	-0.2507	0.00	-0.0001	0.0007	0.54	0.0948	-0.1023	0.09	0.0039	0.0240
Independent ACD-ARMA Model												
0.0	0.42	0.1484	-0.0988	0.15	0.0048	0.0389	0.41	0.1505	-0.0959	0.15	0.0044	0.0393
0.5	0.54	0.1045	-0.1107	0.11	0.0046	0.0288	0.53	0.1051	-0.1081	0.10	0.0043	0.0291
1.0	0.64	0.0748	-0.1188	0.08	0.0052	0.0210	0.63	0.0744	-0.1163	0.08	0.0044	0.0215
1.5	0.72	0.0550	-0.1257	0.06	**0.0053**	0.0157	0.72	0.0542	-0.1234	0.06	0.0041	0.0158
2.0	0.79	0.0412	-0.1327	0.04	0.0046	0.0121	0.79	0.0402	-0.1277	0.04	**0.0046**	0.0117
2.5	0.84	0.0312	-0.1391	0.03	0.0045	0.0090	0.84	0.0302	-0.1345	0.03	0.0044	0.0091
3.0	0.88	0.0237	-0.1427	0.03	0.0043	0.0064	0.88	0.0228	-0.1419	0.02	0.0038	0.0071
3.5	0.92	0.0180	-0.1478	0.02	0.0039	0.0048	0.92	0.0172	-0.1455	0.02	0.0034	0.0054
4.0	0.94	0.0136	-0.1499	0.01	0.0032	0.0036	0.94	0.0129	-0.1522	0.01	0.0029	0.0036
Truncated Density-ARMA Model												
0.0	0.04	0.1918	-0.0075	0.62	0.0006	0.0508	0.39	0.1450	-0.0860	0.14	0.0036	0.0322
0.5	0.41	0.0132	-0.0081	0.21	**0.0006**	0.0113	0.42	0.1342	-0.0892	0.13	0.0036	0.0307
1.0	0.67	0.0034	-0.0075	0.09	-0.0002	0.0036	0.44	0.1258	-0.0909	0.12	0.0038	0.0293
1.5	0.88	0.0008	-0.0055	0.03	0.0000	0.0010	0.46	0.1178	-0.0924	0.11	0.0040	0.0277
2.0	0.97	0.0002	-0.0042	0.00	0.0001	0.0002	0.48	0.1104	-0.0923	0.11	0.0042	0.0256
2.5	1.00	0.0000	-0.0315	0.00	-0.0001	0.0001	0.50	0.1021	-0.0922	0.10	0.0043	0.0232
3.0	1.00	0.0000	0.3400	0.00	0.0000	0.0000	0.51	0.0954	-0.0913	0.10	0.0041	0.0215
3.5	1.00	0.0000	0.3400	0.00	0.0000	0.0000	0.52	0.0896	-0.0892	0.09	**0.0044**	0.0192
4.0	1.00	0.0000	0.3400	0.00	0.0000	0.0000	0.54	0.0833	-0.0883	0.09	0.0045	0.0176
Independent Density-ARMA Model												
0.0	0.35	0.1450	-0.0758	0.15	0.0023	0.0291	0.35	0.1460	-0.0743	0.15	0.0031	0.0293
0.5	0.47	0.0883	-0.0732	0.10	0.0027	0.0178	0.46	0.0905	-0.0713	0.11	0.0036	0.0177
1.0	0.56	0.0612	-0.0714	0.08	0.0032	0.0121	0.55	0.0630	-0.0699	0.08	0.0035	0.0126
1.5	0.63	0.0448	-0.0683	0.06	0.0034	0.0087	0.63	0.0460	-0.0678	0.06	0.0039	0.0091
2.0	0.70	0.0333	-0.0644	0.04	**0.0035**	0.0059	0.70	0.0346	-0.0654	0.05	0.0042	0.0065
2.5	0.75	0.0249	-0.0614	0.03	0.0032	0.0042	0.75	0.0262	-0.0616	0.03	**0.0043**	0.0046
3.0	0.80	0.0187	-0.0582	0.03	0.0031	0.0031	0.80	0.0199	-0.0565	0.03	0.0043	0.0032
3.5	0.84	0.0137	-0.0537	0.02	0.0029	0.0021	0.83	0.0147	-0.0517	0.02	0.0037	0.0023
4.0	0.88	0.0096	-0.0472	0.01	0.0026	0.0014	0.87	0.0104	-0.0445	0.01	0.0032	0.0016
Truncated ACD-Density Model												
0.0	0.71	0.0703	-0.1569	0.10	**0.0041**	0.0350	0.37	0.1633	-0.0897	0.15	0.0041	0.0392
0.5	0.81	0.0502	-0.2023	0.07	0.0032	0.0280	0.42	0.1446	-0.0954	0.13	0.0045	0.0358
1.0	0.88	0.0361	-0.2489	0.05	0.0021	0.0219	0.45	0.1296	-0.1000	0.12	0.0042	0.0331
1.5	0.92	0.0261	-0.2892	0.03	0.0007	0.0175	0.49	0.1175	-0.1042	0.11	0.0044	0.0308
2.0	0.94	0.0192	-0.3252	0.02	-0.0001	0.0142	0.52	0.1074	-0.1088	0.10	0.0042	0.0293
2.5	0.96	0.0145	-0.3615	0.02	-0.0002	0.0116	0.55	0.0985	-0.1123	0.09	0.0042	0.0278
3.0	0.97	0.0111	-0.3848	0.01	-0.0003	0.0095	0.61	0.0871	-0.1227	0.09	0.0045	0.0262
3.5	0.98	0.0086	-0.4198	0.01	-0.0002	0.0078	0.66	0.0766	-0.1360	0.08	**0.0049**	0.0249
4.0	0.98	0.0068	-0.4614	0.01	-0.0006	0.0068	0.71	0.0670	-0.1489	0.07	0.0047	0.0234
Independent ACD-Density Model												
0.0	0.43	0.1457	-0.1007	0.14	0.0048	0.0387	0.42	0.1473	-0.0974	0.14	0.0046	0.0391
0.5	0.61	0.0967	-0.1348	0.10	**0.0068**	0.0322	0.60	0.0972	-0.1314	0.10	**0.0052**	0.0328
1.0	0.75	0.0671	-0.1711	0.07	0.0064	0.0262	0.74	0.0666	-0.1668	0.07	0.0050	0.0268
1.5	0.84	0.0473	-0.2102	0.05	0.0050	0.0212	0.83	0.0463	-0.2045	0.05	0.0039	0.0217
2.0	0.89	0.0340	-0.2443	0.04	0.0026	0.0178	0.89	0.0328	-0.2435	0.04	0.0021	0.0181
2.5	0.92	0.0249	-0.2749	0.03	0.0015	0.0146	0.93	0.0237	-0.2787	0.03	0.0010	0.0150
3.0	0.94	0.0186	-0.3037	0.02	0.0004	0.0122	0.95	0.0176	-0.3158	0.02	0.0005	0.0123
3.5	0.96	0.0141	-0.3289	0.02	-0.0004	0.0103	0.96	0.0133	-0.3484	0.02	-0.0004	0.0106
4.0	0.97	0.0109	-0.3582	0.01	-0.0006	0.0087	0.97	0.0103	-0.3749	0.01	-0.0004	0.0088

Table 7.4: The performance of the proposed order placement strategies and the best static strategy when used to make order placement decisions for the instruments in the Multi Commodity Exchange of India. The performance reported includes the probability that the submitted order is executed (P_E), the average profit obtained when the submitted order is executed ($E(U_E)$), the average loss incurred when the submitted order is not executed ($E(U_{NE})$), the average profit obtained ($E(U)$), and the gain/loss in percentage when compared to the best static strategy.

		Training dataset					Testing dataset				
Strategy	Δ/λ	P_E	$E(U_E)$	$E(U_{NE})$	$E(U)$	%	P_E	$E(U_E)$	$E(U_{NE})$	$E(U)$	%
Liquidity buy problem for Gold future, MCX											
MARKET	-	1.00	-2.6356	-	-2.6356	-	1.00	-3.3537	-	-3.3537	-
LMARKET	-	0.00	-	-2.7558	-2.7558	-	0.00	-	-3.593	-3.593	-
BEST LIMIT	7	0.55	4.0956	-9.9886	-2.263	-	0.60	3.2913	-12.5233	-2.9682	-
T-ACD-ARMA	0.002	0.58	0.6447	-6.0006	-2.1432	5.30%	0.56	1.2604	-8.286	-2.8942	2.50%
I-ACD-ARMA	0.038	0.73	1.5695	-12.8164	-2.2903	-1.20%	0.73	1.3562	-14.8049	-2.928	1.40%
T-DEN-ARMA	0.007	0.48	-0.8093	-3.5061	-2.2011	2.70%	0.54	-1.5367	-4.9033	-3.0948	-4.30%
I-DEN-ARMA	0.035	0.67	1.2568	-9.5154	-2.2726	-0.40%	0.7	0.3866	-11.5807	-3.1913	-7.50%
T-ACD-DEN	0.000	0.62	3.1335	-11.2481	-2.3453	-3.60%	0.59	4.4427	-13.6369	-2.8997	2.30%
I-ACD-DEN	0.009	0.58	4.3987	-11.5322	-2.2203	1.90%	0.6	4.6666	-14.122	-2.9291	1.30%
Liquidity sell problem for Gold future, MCX											
MARKET	-	1.00	-2.6356	-	-2.6356	-	1.00	-3.3537	-	-3.3537	-
LMARKET	-	0.00	-	-2.5233	-2.5233	-	0.00	-	-3.0558	-3.0558	-
BEST LIMIT	5	0.68	2.1814	-10.937	-2.0635	-	0.72	1.3887	-13.0614	-2.5963	-
T-ACD-ARMA	0.049	0.55	4.9488	-10.5892	-2.0948	-1.50%	0.59	4.577	-12.5053	-2.505	3.50%
I-ACD-ARMA	0.026	0.61	3.6299	-11.1175	-2.1133	-2.40%	0.62	3.4711	-13.2599	-2.917	-12.40%
T-DEN-ARMA	0.058	0.57	3.9071	-10.0436	-2.0708	-0.40%	0.64	3.0458	-12.1469	-2.4361	6.20%
I-DEN-ARMA	0.014	0.45	5.535	-8.3021	-2.055	0.40%	0.51	4.6932	-10.2455	-2.6442	-1.80%
T-ACD-DEN	0.057	0.63	3.6295	-11.6495	-2.071	-0.40%	0.65	3.369	-13.8379	-2.5926	0.10%
I-ACD-DEN	0.006	0.51	5.9787	-10.5162	-2.0823	-0.90%	0.52	6.4013	-12.6331	-2.6534	-2.20%
Liquidity buy problem for Silver future, MCX											
MARKET	-	1.00	-13.9272	-	-13.9272	-	1.00	-8.5941	-	-8.5941	-
LMARKET	-	0.00	-	-14.2661	-14.2661	-	0.00	-	-8.7152	-8.7152	-
BEST LIMIT	20	0.4	2.7885	-22.6995	-12.4615	-	0.48	10.3945	-23.2267	-6.9312	-
T-ACD-ARMA	0.000	0.43	-4.3234	-17.9896	-12.1312	2.70%	0.39	3.5043	-14.3259	-7.3623	-6.20%
I-ACD-ARMA	0.018	0.48	2.8293	-25.2498	-11.8178	5.20%	0.64	3.2918	-24.9985	-6.8902	0.60%
T-DEN-ARMA	0.001	0.23	-1.6784	-15.6905	-12.4859	-0.20%	0.28	3.4333	-11.7572	-7.4675	-7.70%
I-DEN-ARMA	0.014	0.4	4.2479	-22.8506	-12.0092	3.60%	0.59	3.8189	-22.7046	-7.1371	-3.00%
T-ACD-DEN	0.002	0.62	-3.1283	-28.0195	-12.6417	-1.40%	0.55	6.8026	-24.8378	-7.5635	-9.10%
I-ACD-DEN	0.000	0.37	6.7799	-23.5323	-12.3283	1.10%	0.40	16.3573	-22.8207	-7.1582	-3.30%
Liquidity sell problem for Silver future, MCX											
MARKET	-	1.00	-13.9272	-	-13.9272	-	1.00	-8.5941	-	-8.5941	-
LMARKET	-	0.00	-	-13.6492	-13.6492	-	0.00	-	-8.2202	-8.2202	-
BEST LIMIT	27	0.31	8.9971	-20.8138	-11.4747	-	0.37	16.9567	-21.2201	-7.2121	-
T-ACD-ARMA	0.000	0.41	6.8507	-23.8936	-11.3018	1.50%	0.43	14.6263	-23.783	-7.1659	0.60%
I-ACD-ARMA	0.014	0.43	6.6679	-24.0288	-10.8055	5.80%	0.59	5.452	-23.8278	-6.5993	8.50%
T-DEN-ARMA	0.000	0.41	4.4635	-22.6947	-11.5549	-0.70%	0.46	12.3014	-23.0055	-6.7662	6.20%
I-DEN-ARMA	0.012	0.38	6.1091	-21.9234	-11.1369	2.90%	0.57	4.8188	-21.1451	-6.4428	10.70%
T-ACD-DEN	0.000	0.38	7.8089	-23.1627	-11.3751	0.90%	0.43	15.01	-23.4589	-7.0858	1.80%
I-ACD-DEN	0.000	0.37	8.873	-22.9993	-11.3011	1.50%	0.4	16.2643	-22.7785	-7.1268	1.20%
Liquidity buy problem for Natural Gas future, MCX											
MARKET	-	1.00	-0.9031	-	-0.9031	-	1.00	-0.3633	-	-0.3633	-
LMARKET	-	0.00	-	-0.8942	-0.8942	-	0.00	-	-0.3607	-0.3607	-
BEST LIMIT	0.9	0.23	-0.3671	-1.0106	-0.8611	-	0.12	0.2589	-0.43	-0.349	-
T-ACD-ARMA	0.033	0.43	-0.7164	-0.8718	-0.8058	6.40%	0.40	-0.2267	-0.3785	-0.3176	9.00%
I-ACD-ARMA	0.085	0.12	0.5026	-1.0248	-0.8415	2.30%	0.13	0.3754	-0.4558	-0.3468	0.60%
T-DEN-ARMA	0.089	0.14	-0.5275	-0.8815	-0.8335	3.20%	0.07	0.0857	-0.3715	-0.3374	3.30%
I-DEN-ARMA	0.084	0.11	0.5007	-1.0135	-0.845	1.90%	0.09	0.4614	-0.4348	-0.3539	-1.40%
T-ACD-DEN	0.037	0.56	-0.555	-1.2675	-0.8677	-0.80%	0.69	-0.2281	-0.6399	-0.3541	-1.40%
I-ACD-DEN	0.000	0.32	-0.3219	-1.1052	-0.8573	0.40%	0.27	0.0627	-0.4867	-0.3366	3.60%
Liquidity sell problem for Natural Gas future, MCX											
MARKET	-	1.00	-0.9031	-	-0.9031	-	1.00	-0.3633	-	-0.3633	-
LMARKET	-	0.00	-	-0.9105	-0.9105	-	0.00	-	-0.3681	-0.3681	-
BEST LIMIT	1.000	0.23	-0.2868	-0.989	-0.8261	-	0.81	-0.2775	-0.6328	-0.3454	-
T-ACD-ARMA	0.017	0.43	-0.3247	-1.1818	-0.8091	2.10%	0.44	0.027	-0.5473	-0.2965	14.20%
I-ACD-ARMA	0.085	0.14	0.576	-1.0346	-0.8105	1.90%	0.17	0.4076	-0.458	-0.3141	9.10%
T-DEN-ARMA	0.000	0.49	-0.5099	-1.1232	-0.8252	0.10%	0.36	-0.0069	-0.4731	-0.3044	11.90%
I-DEN-ARMA	0.085	0.13	0.5403	-1.0176	-0.8174	1.10%	0.09	0.5206	-0.4232	-0.3383	2.10%
T-ACD-DEN	0.010	0.37	-0.2891	-1.1268	-0.8138	1.50%	0.32	0.0991	-0.4959	-0.3074	11.00%
I-ACD-DEN	0.000	0.37	-0.2485	-1.1365	-0.8109	1.80%	0.35	0.077	-0.5132	-0.3048	11.80%

Table 7.5: The performance of the proposed order placement strategies and the best static strategy when apply to make order placement decision for the instruments in the New York Stock Exchange. The performance reported include the probability that the submitted order is executed (P_E), the average profit obtained when the submitted order is executed ($E(U_E)$), the average loss incurred when the submitted order is not executed ($E(U_{NE})$), the average profit obtained ($E(U)$), and the gain/loss in percentage when comparing to the best static strategy.

		Training dataset					Testing dataset				
Strategy	Δ/λ	P_E	$E(U_E)$	$E(U_{NE})$	$E(U)$	%	P_E	$E(U_E)$	$E(U_{NE})$	$E(U)$	%
Liquidity buy problem for GE stock, NYSE											
MARKET	-	1.00	0.0000	-	0.0000	-	1.00	0.0000	-	0.0000	-
LMARKET	-	0.00	-	0.0002	0.0002	-	0.00	-	0.0005	0.0005	-
BEST LIMIT	0.02	0.80	0.0200	-0.0354	0.0092	-	0.87	0.0200	-0.0683	0.0085	-
T-ACD-ARMA	0.20	0.78	0.0186	-0.0290	0.0081	-11.89%	0.73	0.0329	-0.0620	0.0068	-19.63%
I-ACD-ARMA	5.20	0.80	0.0211	-0.0409	0.0090	-1.66%	0.81	0.0276	-0.0781	0.0073	-14.73%
T-DEN-ARMA	3.00	0.91	0.0057	-0.0088	0.0043	-52.56%	0.96	0.0018	-0.0124	0.0013	-84.77%
I-DEN-ARMA	5.00	0.81	0.0172	-0.0294	0.0082	-10.16%	0.91	0.0110	-0.0449	0.0057	-32.88%
T-ACD-DEN	0.00	0.70	0.0263	-0.0334	0.0084	-8.55%	0.60	0.0538	-0.0646	0.0063	-25.70%
I-ACD-DEN	2.40	0.75	0.0259	-0.0407	0.0091	-0.77%	0.70	0.0436	-0.0762	0.0071	-16.38%
Liquidity sell problem for GE stock, NYSE											
MARKET	-	1.00	0.0000	-	0.0000	-	1.00	0.0000	-	0.0000	-
LMARKET	-	0.00	-	-0.0002	-0.0002	-	0.00	-	-0.0005	-0.0005	-
BEST LIMIT	0.02	0.80	0.0200	-0.0365	0.0090	-	0.86	0.0200	-0.0773	0.0067	-
T-ACD-ARMA	11.00	0.59	0.0346	-0.0321	0.0073	-18.50%	0.65	0.0435	-0.0638	0.0063	-5.30%
I-ACD-ARMA	4.40	0.79	0.0224	-0.0416	0.0087	-2.76%	0.78	0.0299	-0.0765	0.0063	-5.93%
T-DEN-ARMA	8.60	0.61	0.0322	-0.0308	0.0075	-16.23%	0.71	0.0339	-0.0642	0.0057	-14.29%
I-DEN-ARMA	4.80	0.82	0.0168	-0.0305	0.0084	-6.00%	0.91	0.0107	-0.0461	0.0057	-15.19%
T-ACD-DEN	11.00	0.74	0.0258	-0.0399	0.0085	-4.74%	0.72	0.0380	-0.0768	0.0062	-6.82%
I-ACD-DEN	2.60	0.77	0.0240	-0.0431	0.0085	-5.20%	0.70	0.0419	-0.0784	0.0060	-9.68%
Liquidity buy problem for IBM stock, NYSE											
MARKET	-	1.00	0.0000	-	0.0000	-	1.00	0.0000	-	0.0000	-
LMARKET	-	0.00	-	-0.0004	-0.0004	-	0.00	-	-0.0029	-0.0029	-
BEST LIMIT	0.08	0.57	0.0800	-0.0975	0.0041	-	0.60	0.0800	-0.1159	0.0019	-
T-ACD-ARMA	0.00	0.49	0.0499	-0.0406	0.0035	-16.58%	0.49	0.0422	-0.0374	0.0015	-18.71%
I-ACD-ARMA	1.50	0.72	0.0550	-0.1257	0.0053	27.62%	0.72	0.0566	-0.1377	0.0020	6.98%
T-DEN-ARMA	0.20	0.20	0.0365	-0.0074	0.0014	-67.44%	0.16	0.0435	-0.0103	-0.0014	-175.92%
I-DEN-ARMA	1.20	0.59	0.0538	-0.0696	0.0037	-9.63%	0.63	0.0599	-0.1008	0.0011	-41.32%
T-ACD-DEN	0.00	0.71	0.0703	-0.1569	0.0041	-0.68%	0.67	0.0664	-0.1303	0.0015	-19.47%
I-ACD-DEN	0.60	0.64	0.0899	-0.1414	0.0068	63.15%	0.62	0.0860	-0.1337	0.0034	76.13%
Liquidity sell problem for IBM stock, NYSE											
MARKET	-	1.00	0.0000	-	0.0000	-	1.00	0.0000	-	0.0000	-
LMARKET	-	0.00	-	0.0004	0.0004	-	0.00	-	0.0029	0.0029	-
BEST LIMIT	0.07	0.62	0.0700	-0.1015	0.0043	-	0.66	0.0700	-0.1189	0.0049	-
T-ACD-ARMA	0.80	0.43	0.1426	-0.0978	0.0049	11.67%	0.45	0.1314	-0.0964	0.0062	25.98%
I-ACD-ARMA	0.20	0.46	0.1309	-0.1013	0.0048	10.36%	0.47	0.1234	-0.0989	0.0059	20.35%
T-DEN-ARMA	5.20	0.58	0.0708	-0.0850	0.0048	9.77%	0.61	0.0787	-0.1070	0.0057	17.38%
I-DEN-ARMA	2.80	0.78	0.0222	-0.0585	0.0044	2.13%	0.84	0.0207	-0.0879	0.0036	-27.36%
T-ACD-DEN	3.50	0.66	0.0766	-0.1360	0.0049	12.26%	0.67	0.0760	-0.1345	0.0057	17.35%
I-ACD-DEN	0.40	0.56	0.1052	-0.1238	0.0053	21.91%	0.56	0.1008	-0.1172	0.0058	17.88%
Liquidity buy problem for Microsoft stock, NYSE											
MARKET	-	1.00	0.0000	-	0.0000	-	1.00	0.0000	-	0.0000	-
LMARKET	-	0.00	-	0.0007	0.0007	-	0.00	-	-0.0005	-0.0005	-
BEST LIMIT	0.02	0.82	0.0200	-0.0415	0.0092	-	0.83	0.0200	-0.0427	0.0091	-
T-ACD-ARMA	0.00	0.77	0.0225	-0.0401	0.0081	-12.30%	0.79	0.0201	-0.0371	0.0083	-9.07%
I-ACD-ARMA	6.20	0.81	0.0218	-0.0470	0.0089	-3.29%	0.82	0.0207	-0.0444	0.0088	-2.98%
T-DEN-ARMA	2.60	0.94	0.0063	-0.0188	0.0047	-48.88%	0.94	0.0065	-0.0221	0.0048	-47.46%
I-DEN-ARMA	5.00	0.84	0.0159	-0.0316	0.0081	-12.23%	0.84	0.0167	-0.0355	0.0082	-9.80%
T-ACD-DEN	0.80	0.75	0.0268	-0.0496	0.0079	-13.89%	0.77	0.0242	-0.0455	0.0079	-12.61%
I-ACD-DEN	3.00	0.74	0.0286	-0.0475	0.0087	-5.90%	0.74	0.0264	-0.0442	0.0083	-8.46%
Liquidity sell problem for Microsoft stock, NYSE											
MARKET	-	1.00	0.0000	-	0.0000	-	1.00	0.0000	-	0.0000	-
LMARKET	-	0.00	-	-0.0007	-0.0007	-	0.00	-	0.0005	0.0005	-
BEST LIMIT	0.02	0.83	0.0200	-0.0444	0.0088	-	0.84	0.0200	-0.0429	0.0098	-
T-ACD-ARMA	11.00	0.67	0.0235	-0.0314	0.0055	-37.31%	0.68	0.0232	-0.0299	0.0061	-37.13%
I-ACD-ARMA	1.80	0.75	0.0184	-0.0331	0.0057	-36.03%	0.75	0.0177	-0.0259	0.0069	-29.02%
T-DEN-ARMA	10.20	0.56	0.0294	-0.0255	0.0053	-40.35%	0.56	0.0313	-0.0261	0.0063	-35.74%
I-DEN-ARMA	1.00	0.68	0.0153	-0.0205	0.0037	-58.08%	0.68	0.0150	-0.0162	0.0049	-49.97%
T-ACD-DEN	14.80	0.78	0.0249	-0.0515	0.0084	-4.73%	0.79	0.0239	-0.0458	0.0094	-3.31%
I-ACD-DEN	3.40	0.75	0.0273	-0.0500	0.0081	-8.30%	0.77	0.0257	-0.0447	0.0092	-5.84%

Table 7.6: The profit gained from using the proposed strategy over immediately executing the trade at the beginning of the trading period in number of ticks averaged over all instruments in the same markets together with the improvement over the best static strategy in percentage terms and its ranking.

	MCX Training			MCX Testing			NYSE Training			NYSE Testing		
Model	$E(U)$	%	#	$E(U)$	%	#	$E(U)$	%	#	$E(U)$	%	#
BEST LIMIT	1.0088	0.00%	6	0.7516	0.00%	6	0.7445	0.00%	2	0.6811	0.00%	1
T-ACD-ARMA	1.2281	21.74%	2	0.8491	12.96%	2	0.6218	-16.47%	5	0.5884	-13.61%	5
I-ACD-ARMA	1.2735	26.25%	1	0.8698	15.72%	1	0.7063	-5.13%	3	0.6207	-8.88%	3
T-DEN-ARMA	1.0481	3.90%	5	0.8301	10.43%	3	0.4660	-37.41%	7	0.3732	-45.22%	7
I-DEN-ARMA	1.1818	17.15%	3	0.8042	6.99%	5	0.6106	-17.99%	6	0.4859	-28.67%	6
T-ACD-DEN	0.9899	-1.86%	7	0.7343	-2.31%	7	0.7044	-5.38%	4	0.6205	-8.90%	4
I-ACD-DEN	1.0956	8.61%	4	0.8135	8.23%	4	0.7738	3.93%	1	0.6635	-2.59%	2

improvement over the best static strategy in training and testing dataset, respectively. For the NYSE market, the ACD-DENSITY model with the independent assumption is the best performing model with 3.93% improvement in the training dataset. The result also indicates that the model with independent assumption generally performed better than the model with the truncated assumption (except in the case of the DENSITY-ARMA model in the MCX testing dataset).

Introducing market variables

We now add market variables into our model to investigate whether adding market variables can improve the order placement decision or not. To achieve this, we modify the ACD and the ARMA-GARCH model in Equation (7.24) and (7.28) to include the effect from market variables by adding the market variables into the conditional mean equations resulting in

$$\begin{aligned}\psi_i^{\delta_1} &= \omega + \sum_{j=1}^{p} \alpha_j \psi_{i-j}^{\delta_1} \left(|\epsilon_{i-j} - b| + c_j(\epsilon_{i-j} - b)\right)^{\delta_2} \\ &\quad + \sum_{j=1}^{p} \lambda_j \left(|\epsilon_{i-j} - b| + c_j(\epsilon_{i-j} - b)\right)^{\delta_2} + \sum_{j=1}^{q} \beta_j \psi_{i-j}^{\delta_1} + \sum_{j=1}^{s} \gamma_j x_{ij}, \qquad (7.38)\end{aligned}$$

for the conditional mean function of the ACD model, and

$$\hat{r}_i = \omega + \sum_{j=1}^{p} \alpha_j r_{i-j} + \sum_{j=1}^{q} \beta_j z_{i-j} + z_i + \sum_{j=1}^{s} \gamma_j x_{ij}, \qquad (7.39)$$

for the conditional mean function of the ARMA-GARCH model when $x_i = (x_{i1}, \ldots, x_{is})$ is the market variables at the i-th time step.

Since we only have market variables for the MCX dataset, this section will perform the analysis on only the MCX dataset. The market variables considered here consist of the bid-ask spread, order imbalance (the difference between the number of orders at the best bid and the best ask), and trading volume in the previous trading period, which is reported to improve trading performance in Nevmyvaka et al. [72]. Table 7.7 displays the performance of the best risk aversion level for each of the six models together with the three static strategies mentioned above. Similar to the case of no market variables, the result indicates that all models beat the static strategies that always execute the trade at the beginning and at the end of the trading period in all situations. Additionally, the best of these six models also beat

Table 7.7: The performance of the proposed order placement strategies with market variables and the best static strategy when used to make order placement decision for the instruments in the Multi Commodity Exchange of India. The performance reported includes the probability that the submitted order is executed (P_E), the average profit obtained when the submitted order is executed ($E(U_E)$), the average loss incurred when the submitted order is not executed ($E(U_{NE})$), the average profit obtained ($E(U)$), and the gain/loss in percentage when compared to the best static strategy.

		Training dataset					Testing dataset				
Strategy	Δ/λ	P_E	$E(U_E)$	$E(U_{NE})$	$E(U)$	%	P_E	$E(U_E)$	$E(U_{NE})$	$E(U)$	%
Liquidity buy problem for Gold future, MCX											
MARKET	-	1.00	-2.6356	-	-2.6356	-	1.00	-3.3537	-	-3.3537	-
LMARKET	-	0.00	-	-2.7558	-2.7558	-	0.00	-	-3.5930	-3.5930	-
BEST LIMIT	7	0.55	4.0956	-9.9886	-2.2630	-	0.60	3.2913	-12.5233	-2.9682	-
T-ACD-ARMA	0.008	0.73	0.1681	-8.7285	-2.2025	2.68%	0.70	0.3004	-10.7919	-3.0363	-2.30%
I-ACD-ARMA	0.031	0.67	2.5107	-11.9494	-2.2839	-0.92%	0.69	1.9915	-14.3339	-3.0576	-3.01%
T-DEN-ARMA	0.015	0.76	-1.2868	-4.8906	-2.1687	4.17%	0.75	-1.9302	-6.2305	-2.9847	-0.56%
I-DEN-ARMA	0.028	0.61	1.9138	-9.0821	-2.3455	-3.64%	0.67	0.7044	-11.0116	-3.2038	-7.94%
T-ACD-DEN	0.016	0.83	0.4393	-15.7548	-2.3717	-4.80%	0.81	0.6914	-18.8440	-3.1136	-4.90%
I-ACD-DEN	0.03	0.84	0.5771	-16.5422	-2.2369	1.16%	0.50	6.8490	-12.8132	-2.9387	0.99%
Liquidity sell problem for Gold future, MCX											
MARKET	-	1.00	-2.6356	-	-2.6356	-	1.00	-3.3537	-	-3.3537	-
LMARKET	-	0.00	-	-2.5233	-2.5233	-	0.00	-	-3.0558	-3.0558	-
BEST LIMIT	5	0.68	2.1814	-10.9370	-2.0635	-	0.72	1.3887	-13.0614	-2.5963	-
T-ACD-ARMA	0.064	0.58	4.4270	-10.9370	-2.0640	-0.02%	0.62	4.0155	-13.0351	-2.5072	3.43%
I-ACD-ARMA	0.021	0.57	4.4582	-10.5350	-2.0437	0.96%	0.57	4.5130	-12.5243	-2.8389	-9.35%
T-DEN-ARMA	0.052	0.55	4.1547	-9.7784	-2.0518	0.57%	0.62	3.3729	-12.1278	-2.4542	5.47%
I-DEN-ARMA	0.016	0.49	4.4640	-8.3040	-2.0455	0.88%	0.54	3.6905	-10.2512	-2.6644	-2.62%
T-ACD-DEN	0.068	0.66	3.2422	-12.2083	-2.0573	0.30%	0.69	2.9486	-14.5518	-2.5219	2.86%
I-ACD-DEN	0.009	0.57	5.1129	-11.3937	-2.0596	0.19%	0.58	5.4707	-13.6478	-2.6033	-0.27%
Liquidity buy problem for Silver future, MCX											
MARKET	-	1.00	-13.9272	-	-13.9272	-	1.00	-8.5941	-	-8.5941	-
LMARKET	-	0.00	-	-14.2661	-14.2661	-	0.00	-	-8.7152	-8.7152	-
BEST LIMIT	20	0.40	2.7885	-22.6995	-12.4615	-	0.48	10.3945	-23.2267	-6.9312	-
T-ACD-ARMA	0.005	0.62	-5.3800	-21.5245	-11.4349	8.24%	0.72	-2.5691	-18.1243	-6.9101	0.30%
I-ACD-ARMA	0.013	0.41	5.4125	-24.3563	-12.0391	3.39%	0.60	4.9270	-25.3493	-7.1992	-3.87%
T-DEN-ARMA	0.009	0.55	-6.4325	-18.2981	-11.7399	5.79%	0.69	-4.2680	-14.9368	-7.5683	-9.19%
I-DEN-ARMA	0.007	0.32	6.3043	-21.6261	-12.6626	-1.61%	0.48	7.8295	-22.2245	-7.7134	-11.29%
T-ACD-DEN	0.003	0.65	-2.5233	-29.4288	-11.8876	4.61%	0.67	3.4547	-28.7167	-7.2335	-4.36%
I-ACD-DEN	0.001	0.39	7.8801	-24.5098	-11.8030	5.28%	0.48	13.5446	-25.3791	-6.8440	1.26%
Liquidity sell problem for Silver future, MCX											
MARKET	-	1.00	-13.9272	-	-13.9272	-	1.00	-8.5941	-	-8.5941	-
LMARKET	-	0.00	-	-13.6492	-13.6492	-	0.00	-	-8.2202	-8.2202	-
BEST LIMIT	27	0.31	8.9971	-20.8138	-11.4747	-	0.37	16.9567	-21.2201	-7.2121	-
T-ACD-ARMA	0	0.39	9.8776	-24.1011	-10.9930	4.20%	0.43	15.4360	-24.2853	-7.1006	1.55%
I-ACD-ARMA	0.013	0.42	7.9824	-24.1931	-10.8167	5.73%	0.59	5.9884	-24.3773	-6.5772	8.80%
T-DEN-ARMA	0.000	0.40	6.2994	-23.2220	-11.3821	0.81%	0.46	12.8073	-23.0657	-6.6454	7.86%
I-DEN-ARMA	0.008	0.32	11.1002	-21.6620	-11.0348	3.83%	0.50	8.8558	-20.9699	-6.1176	15.18%
T-ACD-DEN	0.003	0.41	7.8032	-24.0072	-11.1120	3.16%	0.47	13.1287	-24.8965	-7.0700	1.97%
I-ACD-DEN	0.003	0.43	6.8151	-24.9905	-11.2268	2.16%	0.49	12.0850	-25.7921	-7.2800	-0.94%
Liquidity buy problem for Natural Gas future, MCX											
MARKET	-	1.00	-0.9031	-	-0.9031	-	1.00	-0.3633	-	-0.3633	-
LMARKET	-	0.00	-	-0.8942	-0.8942	-	0.00	-	-0.3607	-0.3607	-
BEST LIMIT	0.9	0.23	-0.3671	-1.0106	-0.8611	-	0.12	0.2589	-0.4300	-0.3490	-
T-ACD-ARMA	0.060	0.47	-0.5156	-1.0321	-0.7883	8.46%	0.44	-0.1711	-0.4289	-0.3155	9.61%
I-ACD-ARMA	0.089	0.14	0.4654	-1.0485	-0.8292	3.72%	0.15	0.3827	-0.4574	-0.3355	3.89%
T-DEN-ARMA	0.089	0.11	-0.2569	-0.9115	-0.8391	2.56%	0.07	0.1522	-0.3761	-0.3382	3.10%
I-DEN-ARMA	0.060	0.10	0.5543	-1.0061	-0.8451	1.87%	0.07	0.6389	-0.4248	-0.3540	-1.43%
T-ACD-DEN	0.013	0.59	-0.3882	-1.4561	-0.8217	4.58%	0.69	-0.1953	-0.6540	-0.3360	3.74%
I-ACD-DEN	0.000	0.34	-0.1294	-1.1801	-0.8231	4.41%	0.28	0.1146	-0.5031	-0.3280	6.02%
Liquidity sell problem for Natural Gas future, MCX											
MARKET	-	1.00	-0.9031	-	-0.9031	-	1.00	-0.363332	-	-0.363332	-
LMARKET	-	0.00	-	-0.9105	-0.9105	-	0.00	-	-0.36812	-0.36812	-
BEST LIMIT	1	0.23	-0.2868	-0.9890	-0.8261	-	0.81	-0.277526	-0.63279	-0.345426	-
T-ACD-ARMA	0.003	0.41	-0.1576	-1.2298	-0.7938	3.92%	0.41	0.0691839	-0.552794	-0.294706	14.68%
I-ACD-ARMA	0.088	0.14	0.5725	-1.0417	-0.8127	1.63%	0.16	0.403118	-0.451998	-0.313196	9.33%
T-DEN-ARMA	0.037	0.48	-0.4544	-1.1477	-0.8166	1.16%	0.39	0.0041958	-0.49375	-0.299728	13.23%
I-DEN-ARMA	0.074	0.12	0.6390	-1.0152	-0.8181	0.98%	0.10	0.490945	-0.425961	-0.335306	2.93%
T-ACD-DEN	0.012	0.36	-0.1664	-1.1599	-0.8028	2.82%	0.28	0.14509	-0.487378	-0.309381	10.43%
I-ACD-DEN	0.009	0.34	-0.1339	-1.1535	-0.8038	2.71%	0.29	0.135054	-0.49147	-0.309537	10.39%

Table 7.8: Comparison between the performance of the best strategy with market variables and the one without market variables.

		Training dataset			Testing dataset		
Instrument	Side	$E(U)$ with X	$E(U)$ no X	Gain(%)	$E(U)$ with X	$E(U)$ no X	Gain(%)
Gold futures	Buy	-2.1687	-2.1432	-1.19%	-2.9847	-2.8942	-3.13%
Gold futures	Sell	-2.0437	-2.0550	0.55%	-2.8389	-2.6442	-7.36%
Silver futures	Buy	-11.4349	-11.8178	3.24%	-6.9101	-6.8902	-0.29%
Silver futures	Sell	-10.8167	-10.8055	-0.10%	-6.5772	-6.5993	0.33%
Natural gas futures	Buy	-0.7883	-0.8058	2.17%	-0.3155	-0.3176	0.66%
Natural gas futures	Sell	-0.7938	-0.8091	1.89%	-0.2947	-0.2965	0.61%
Average ticks gained		-7.0475	-7.1618	1.60%	-4.2355	-4.1948	-0.97%

the best static strategy in all situations as well.

To compare the improvement gained from including market variables into the model, Table 7.8 compares the performance of the best performing models that includes market variables and the one that does not include market variables. The result indicates that including market variables into the model does not necessarily improve trading performance as we gain performance improvement in the training dataset in only four out of the six cases considered. Additionally, we gain improvement both in the training dataset and testing dataset only in two cases for the Natural gas futures. When considering the average profit gained from including the market variables into the model, averaged over all situations, we find that the model with market variables has better performance in the training dataset while it performs worse in the testing dataset.

Consequently, the result studied in this section indicates that including market variables in our model does not always result in better trading performance, as suggested in [72], and it is more appropriate to try the model with and without market variables to select the best model for making the order placement decision. A general guideline to select the best model for a particular instrument is to adopt the model that has the best combined performance on the buy and the sell problem. By using this criterion, we will select the model without market variables for Gold futures, while we will select the model with market variables for Silver and Natural gas futures which is exactly the best performing model for each instrument.

7.5 Summary

This chapter proposed a new framework for making order placement decisions based on the trade-off between the profit gained from better execution price and the risk of non-execution in a mean-variance optimisation framework. This framework is general enough to solve all trading problems mentioned in Harris [42] as traders can define their trading objective by specifying two payoff functions: a function $U_E(p)$ that defines the payoff that traders will get when they execute the trade at price level p, and a function $U_{NE}(p)$ that defines the cost that traders need to pay when their order is not executed and the price of the asset at the end of the trading period is p. In particular, the order placement problem of liquidity traders who need to transact their order before some specified deadline can be modelled by setting $U_{NE}(p) = U_E(p)$ so that the cost that a trader needs to pay if his order is not executed is equal to the cost of executing the trade at the end of the trading period. The utility function of informed traders and value motivated traders can be modelled by setting $U_{NE}(p) = \max\{U_E(p), U_{NT}\}$, where U_{NT} is a utility gained from not trading, so that they will trade at the end of the period only when it is profitable to do so. After specifying the objective function, traders also need to specify a model for the probability that the limit order at price level p^L will be executed before the end of the period and a probability density function of the asset price at the end of the trading period given no execution of limit order at price level p^L. Accordingly, we discuss three different approaches to model these two probabilities, which are i) the unconditional model implied by an arithmetic Brownian motion, ii) the empirical unconditional model using density estimation and iii) the empirical conditional distribution using ACD and ARMA-GARCH models.

To measure the performance of the proposed framework in making order placement decisions, we performed an experiment to apply the proposed framework to make order placement decision for liquidity traders using the historical data from the Multi Commodity Exchange of India (MCX) and the New York Stock Exchange (NYSE). The results indicated that the proposed framework beats the simple static strategies that always execute the trade at the beginning of the period and at the end of the period in all cases studied. The proposed framework also beated the best static strategy in all cases in the MCX dataset, while it beated the best static strategy only in two out of six cases considered in the NYSE dataset. Although the proposed framework could not beat the best static strategy in all cases, the improvement gained from the proposed framework when it can beat the best static strategy is very significant. Additionally, the result obtained from the model with market variables indicated that adding market variables into the model does not necessary improve the trading performance of the model. This suggests that it is more appropriate to try both models and select the one that performs best in the training dataset to make the trading decision.

Chapter 8

Conclusion

This chapter brings the thesis to a conclusion. We begin by summarising the key points of this work, what guided us in this direction and what can be learned from our models and experiments. We the review our contributions and academic achievements, and suggest some direct applications for practitioners.

8.1 On the origins of this thesis

This thesis had been mainly concerned with the model of execution probability and its application in an order placement strategy.

We started the adventure in these subjects by acknowledging that most of the publications in algorithmic trading are mainly focusing on the trade scheduling problem but the order choice problem, which was equally important, has largely dismissed by researchers in this field. Although the order choice problem was an active research topic during the last few decades in the market microstructure community, these studies were mostly theoretical and mainly focused on analysing the rational for, and the profitability of, limit order trading as well as the dynamic characteristics of limit order markets implied by those findings rather than focusing on developing a profitable algorithmic trading system. Consequently, to fill the gap in this research area, the main objective of this research was set to develop a framework for optimising the order placement strategy in an algorithmic trading system.

After reviewing previous works from the market microstructure community, we found that these theoretical works generally viewed the order type problem as a trade-off between the payoffs associated with limit orders and the risk of non execution. On one hand, traders would prefer to place their orders very far from the best price to increase their payoff. On the other hand, the larger the distance from the best price, the larger the chance that the order will not be executed. Accordingly, in this setting, traders have to find the right trade-off between these two opposite choices, in order to maximise the expected profit obtained from the trade, and one of the most important factors in valuing such a trade-off is a model of execution probability which can be utilised to compute the probability that limit orders at a specific price level will be executed. Unfortunately, the research into how to model such probability is still very limited and mainly focused on identifying the determinants of the execution probability and the effect that each determinants have on the execution probability with the aim of explaining traders decisions, rather than focus on the accuracy of the prediction result, which is important for developing a profitable algorithmic trading system. These limitations led us to develop a model of execution probability with high accuracy and utilise the developed model to implement an order placement strategy for an algorithmic trading system.

Our journey in developing a model of execution probability starts from an in-depth experiment to compare the performance of previously proposed models in a controlled environment by utilising the data generated from simulation models with the aim to understand the advantage, disadvantage and the limitation of each method. The result from this experiment indicated that among previously proposed models, the models that utilise survival analysis to handle cancelled orders seemed to be the best performing method both from theoretical and empirical point of view. However, the result also indicate that traditional survival analysis models (i.e. the proportional hazards model and the accelerated failure time model) utilised in previous works were not flexible enough to model the effect from explanatory variables even in the simplest simulation model since the assumptions made by these models are violated in most of the situations. Although it is tempting to develop a new method by relaxing the assumptions made by these models, we decide not to take this part as the preliminary result obtained from applying Bayesian neural networks to model this probability proved dissatisfactory. Accordingly, we decided to take two alternative approaches inspired by the disadvantage and limitation we experienced during these experimentations.

The first approach was inspired by the disadvantage and limitation of previous models when we want to model the execution probability from first-passage time information at several price levels simultaneously. Particularly, to model the execution probability at p different price levels from n realisations of a particular asset price process, all previous models required us to prepare np data records which could be very large especially when we want to model the execution probability at all possible price levels. To amend the problem, we proposed a new approach to model the execution probability at a specified time period from the price fluctuation during that period so that the data required for estimating the model was reduced from np to n. Although this model did not allow us to compute the execution probability at several trading horizons simultaneously, the ability to compute the execution probability at several price levels is far more important in our trading application.

The second approach was inspired by the lack of theoretical models for computing the execution probability, even in the simplest simulation model. To fill the gap in this area, we tried to find a way to compute this probability in such a model without resorting to simulation, and this led us to derive the model for describing the dynamics of asset prices in the simulation model and estimate the execution probability from the derived asset price dynamics model. Although the estimation we obtained from the model has high accuracy only for a small subset of the parameter space, the developed framework sheds some light on the interaction between order arrival/cancellation processes and the asset price dynamics.

8.2 Contributions and achievements

In this dissertation, we had presented a general framework and computer implementation for simulating an order flow in order-driven markets based on a continuous-time aggregated order flow model, which is able to be used to simulate several previous simulation models such as those proposed by Smith et al. [88] and Cont et al. [20]. The dynamics of the order book in this framework is assumed to be driven by two different types of agents: i) impatient agents who place market orders randomly according to some predefined stochastic process, and ii) patient agents who place limit orders randomly both in time and in price. Additionally, unexecuted limit orders are also assumed to be cancelled according to some predefined stochastic processes. By controlling the properties of these orders submission and cancellation processes, several realisations of the order book dynamics that have similar stochastic properties can be generated. This enabled us to evaluate the developed model in a controlled environment before applying them to the data generated from real markets.

The main contributions of this thesis are probably related to the evaluation and development of execution probability models. Particularly, the evaluation of previous models in a controlled environment using data generated from simulation models reveal several interesting facts about these models. These include the fact that censored observations, in form of unexecuted orders, are the main obstacle that prevents us from applying traditional methods to model the execution probability and survival analysis is an appropriate method for modelling this probability. The analysis of the relationship between the execution probability and possible explanatory variables indicates that the execution probability of a limit buy order is positively correlated with bid-ask spread, number of sell orders in the order book, market order arrival rate and order cancellation rate while it is negatively correlated with the distance from the opposite best price, the number of buy orders in the order book and the limit order arrival rate. Additionally, the results also indicate that standard survival analysis techniques (i.e. the proportional hazards model and the accelerated failure time model) generally used to model this probability in previous works is not flexible enough to model the effect of these variables even in the simplest simulation model.

We then shifted our attention to the relation between price fluctuation and execution probability. Particularly, we derived the equation relating the distribution of price fluctuation and execution probability which led us to a new method for modelling the execution probability at a specified time period from the fluctuation of the price during the interested period. The advantage of this approach over traditional technique is that it requires less data to model the execution probability at all price levels since it requires only one record per sample while survival analysis requires n records per sample to model the execution probability for n price levels. Additionally, this provides a natural way to apply traditional time series analysis techniques to model the execution probability. The statistical analysis of price fluctuation dataset obtained from the Multi Commodity Exchange of India and the New York stock exchange indicates that the form of the market seem to have strong impact on the dynamics of price fluctuations as the strength and persistence of serial dependency in price fluctuations mainly differ between the individual exchanges and less between the different assets traded in the same exchange. The analysis also suggests that the price fluctuation process seem to have a long range dependency with clear intraday seasonality pattern similar to the one observed from volatility processes. The analysis of the dependency between price fluctuation, return and volatility indicates that price fluctuation is heavily dependent on the direction of returns during the same period in the sense that buy price fluctuation is negatively correlated with return, while sell price fluctuation is positively correlated with return. However, the correlation between price fluctuation and previous return is typically weak and might not be useful for predicting future price fluctuations. Additionally, the results indicated that price fluctuations are also strongly correlated with volatility, as estimated by the range between the highest and lowest price. In a search for the best model for modelling the price fluctuation process, we proposed a new method for fitting the distributions of price fluctuations that do not suffer from the problem experienced in traditional maximum likelihood estimation by maximising the likelihood of the implied discrete distribution rather than directly maximising the likelihood of the model. The results obtained from fitting the price fluctuation process with several time series analysis models indicate that the ACD model is the most appropriate model for modelling this process, and the ABAMACD model, which extends the basic ACD parameterisation by adding an additive stochastic component and an asymmetric news response, as well as applying the Box-Cox transform to the conditional mean function, is the best model for modelling the price fluctuations, and, thus, the best model for modelling the execution probability.

We then derived a stochastic model of asset price dynamics in a simulated model of order-driven markets whose dynamics are described by the incoming flow of market orders, limit orders and order cancellation processes. Particularly, we introduced a framework to model the dynamics of asset prices giving the statistical properties of those processes, thus, establishing the relationship between the microscopic dynamics of the limit order book and the long-term dynamics of the asset price process. Unlike traditional methods that model asset price dynamics using a one-dimensional stochastic process, the proposed framework models this dynamic using a two-dimensional stochastic process where the additional dimension represents information about the latest price change. Using dynamic programming methods, we were able to efficiently compute several interesting properties of the asset price dynamics (i.e. volatility, occupation probability and first-passage probability), conditioning on the trading horizon, without resorting to simulation.

Finally, we proposed a new framework for making order placement decision based on the trade-off between the profit gained from better execution price and the risk of non-execution in a mean-variance optimisation setting. We then applied the developed execution probability model based on the relationship between price fluctuation and the execution probability to implement order placement strategy for liquidity traders who need to transact their order before some specified deadline. The results obtained from applying the proposed framework to make order placements in the historical dataset obtained from the Multi Commodity of India and the New York Stock Exchange indicated that the proposed framework has better performance than the best static order placement strategy for all instruments in the Multi Commodity of India, while it beat the best static strategy only in two out of six cases studied in the New York Stock Exchange. Although the proposed framework cannot beat the best static strategy in all cases, the improvement gained when it did beat the best static strategy was very significant.

8.3 Possible extensions

The analysis and experiments in this dissertation suggested several possible extensions that future work can be carried out to answer some open questions.

To begin with, the simulation results, reported in Chapter 3, indicated that previous simulation models based on the concept of aggregated order flow with constant Poisson rate are not able to generate all the stylised facts observed in real markets. This was somehow inline with empirical findings that Poisson processes are not flexible enough to model the order flows in real markets since the order flows in real markets display clustering. Although several alternative models (e.g. Hawkes process and Autoregressive Conditional Duration) have been proposed to model this order flows, no simulation models had incorporated this finding into the model before. Hence, it is interesting to investigate whether a more complicated order flow model is enough to generate all the stylised facts about the real markets or not.

Secondly, the inability to model the effect of explanatory variables on the execution probability of traditional survival analysis techniques since the assumptions made by these models are generally violated, as reported in Chapter 4, suggested that improvement could be made if we could relax these assumptions. Consequently, possible extensions could be achieved by developing a new survival analysis technique that is flexible enough to model these effects by relaxing these assumptions.

Thirdly, although the result reported in Chapter 5 indicated that the basic ACD model is able to capture the dynamic of the conditional mean function in most of situations, there was still a situation when this model cannot capture all the dynamics especially in the price fluctuation time series of the IBM stock traded in the New York Stock Exchange. Accordingly, this suggested the need to develop a new parameterisation of the ACD model or a new model that is able to capture the dynamics of price fluctuations in this case.

Finally, it would be interesting to see the performance of the proposed order placement framework when used to make order placement decisions of other types of traders rather than the liquidity traders. Consequently, one possible future work would apply the framework to make order placement decisions for informed and value-motivated traders. Additionally, since the single period model considered in this study is somewhat limited, future work could also focus on extending the model into a multi-period setting.

Bibliography

[1] Joseph Abate and Ward Whitt. Computing laplace transforms for numerical inversion via continued fractions. *INFORMS Journal on Computing*, 11:394–405, 1998.

[2] Hee-Joon Ahn, Kee-Hong Bae, and Kalok Chan. Limit orders, depth, and volatility: Evidence from the stock exchange of hong kong. *The Journal of Finance*, 56(2):767–788, 2001.

[3] Robert Almgren and Neil Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–39, 2001.

[4] Robert Almgren and Julian Lorenz. Bayesian adaptive trading with a daily cycle. *Journal of Trading*, 2006.

[5] Robert Almgren and Julian Lorenz. Adaptive arrival price. *Algorithmic Trading III: Precision, Control, Execution*, 2007.

[6] James Angle. Limit versus market orders. Working Paper No. FINC-1377-01-293, School of Business Administration, Georgetown University, Washington, DC, 1994.

[7] Luc Bauwens and Pierre Giot. The logarithmic acd model: An application to the bid-ask quote process of three nyse stocks. *Annales d'conomie et de Statistique*, (60):117–149, 2000.

[8] Helena Beltran-Lopez, Pierre Giot, and Joachim Gramming. Commonalities in the order book. Working Paper 2005014, Universit catholique de Louvain, Dpartement des Sciences Economiques, January 2005.

[9] Bruno Biais, Pierre Hillion, and Chester Spatt. An empirical analysis of the limit order book and the order flow in the paris bourse. *Journal of Finance*, 50(5):1655–89, December 1995.

[10] Adam Blazejewski. *Computational Models for Stock Market Order Submissions*. PhD thesis, The University of Sydney, 2005.

[11] Katalin Boer, Uzay Kaymak, and Jaap Spiering Spiering. From discrete-time models to continuous-time, asynchronous modeling of financial markets. *Computational Intelligence*, 23(2):142–161, 2007.

[12] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307 – 327, 1986.

[13] Jean-Philippe Bouchaud, Yuval Gefen, Marc Potters, and Matthieu Wyart. Fluctuations and response in financial markets: the subtle nature of 'random' price changes. *Quantitative Finance*, 4:176–190, 2004.

[14] Jean-Philippe Bouchaud, Marc Mezard, and Marc Potters. Statistical properties of stock order books: empirical results and models. *Quantitative Finance*, 2:251, 2002.

[15] Glenn W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950.

[16] Sergiy Butenko, Alexander Golodnikov, and Stanislav Uryasev. Optimal security liquidation algorithms. *Comput. Optim. Appl.*, 32(1-2):9–27, 2005.

[17] Charles Cao, Oliver Hansch, and Xiaoxin Wang. Order placement strategies in a pure limit order book market. *The Journal of Financial Research*, 31(2):113–140, 2008.

[18] Carl Chiarella and Giulia Iori. A simulation analysis of the microstructure of double auction markets. *Quantitative Finance*, 2(5):346–353, October 2002.

[19] Kalman J. Cohen, Steven F. Maier, Robert A. Schwartz, and David K. Whitcomb. Transaction costs, order placement strategy, and existence of the bid-ask spread. *The Journal of Political Economy*, 89(2):287–305, 1981.

[20] Rama Cont, Sasha Stoikov, and Rishi Talreja. A stochastic model for order book dynamics. *Social Science Research Network Working Paper Series*, September 2008.

[21] Thomas Copeland and Dan Galai. Information effects and the bid-ask spread. *Journal of Finance*, 38:1457–1469, 1983.

[22] D. R. Cox and D. Oakes. *Analysis of Survival Data*. Chapman & Hall, London, first edition, 1984.

[23] Ian Domowitz and Jianxin Wang. Auctions as algorithms : Computerized trade execution and price discovery. *Journal of Economic Dynamics and Control*, 18(1):29 – 60, 1994.

[24] Alfonso Dufour and Robert F Engle. The acd model: Predictability of the time between concecutive trades. ICMA Centre Discussion Papers in Finance icma-dp2000-05, Henley Business School, Reading University, May 2000.

[25] Nicholas Economides and Robert A Schwartz. Electronic call market trading. *The Journal of Portfolio Management*, 21, 1995.

[26] Zoltan Eisler, Janos Kertesz, Fabrizio Lillo, and Rosario N. Mantegna. Diffusive behavior and the modeling of characteristic times in limit order executions, 2007.

[27] R. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of u.k. inflation. In *Econometrica*, volume 50, pages 987–1008, 1982.

[28] Robert F. Engle and Jeffrey R. Russell. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, 66(5):1127–1162, 1998.

[29] E. S. Epstein. A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, 8:985–987, 1969.

[30] J. Doyne Farmer, Paolo Patelli, and Ilija I. Zovko. The predictive power of zero intelligence in financial markets. *Proceedings of the National Academy of Sciences of the United States of America*, 102(6):2254–2259, February 2005.

[31] Marcelo Fernandes and Joachim Grammig. A family of autoregressive conditional duration models. *Journal of Econometrics*, 130(1):1–23, January 2006.

[32] Thierry Foucault. Order flow composition and trading costs in a dynamic limit order market1. *Journal of Financial Markets*, 2(2):99–134, May 1999.

[33] Thierry Foucault, Ohad Kadan, and Eugene Kandel. Limit order book as a market for liquidity. *Review of Financial Studies*, 18(4):1171–1217, 2005.

[34] Francois Ghoulmie, Rama Cont, and Jean-Pierre Nadal. Heterogeneity and feedback in an agent-based market model. *Journal of Physics: Condensed Matter*, 17(14):S1259–S1268, 2005.

[35] Lawrence R Glosten. Is the electronic open limit order book inevitable? *Journal of Finance*, 49(4):1127–61, September 1994.

[36] Patricia M. Grambsch and Terry M. Therneau. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3):515–526, September 1994.

[37] Joachin Grammig, Andreas Heinen, and Erick Rengifo. Trading activity and liquidity supply in a pure limit order book market: An empirical analysis using a multivariate count data model. In *the 14th Annual Conference of the Deutsche Gesellschaft fr Finanzwirtschaft*, Dresden, Germany, September 2007.

[38] Mark D. Griffiths, Brian F. Smith, D. Alasdair S. Turnbull, and Robert W. White. The costs and determinants of order aggressiveness. *Journal of Financial Economics*, 56(1):65–88, April 2000.

[39] Anthony D. Hall and Nikolaus Hautsch. Order aggressiveness and order book dynamics. *Emprical Economics*, 30(4):973–1005, 2006.

[40] Puneet Handa, Robert Schwartz, and Ashish Tiwari. Quote setting and price formation in an order driven market. *Journal of Financial Markets*, 6(4):461–489, August 2003.

[41] Puneet Handa and Robert A Schwartz. Limit order trading. *Journal of Finance*, 51(5):1835–61, December 1996.

[42] Lawrence Harris. Optimal dynamic order submission strategies in some stylized trading problems. *Financial Markets, Institutions and Instruments*, 7(2):1–76, 1998.

[43] J. Michael Harrison. *Brownian Motion and Stochastic Flow Systems*. Krieger Publishing Company, 1990.

[44] Joel Hasbrouck. *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading*, chapter 2. Oxford University Press, New York, Dec 2007.

[45] Nikolaus Hautsch. *Modelling Irregularly Spaced Financial Data*. Springer-Verlag Berlin Heidelberg, 2004.

[46] Tristen Hayfield and Jeffrey S. Racine. Nonparametric econometrics: The np package. *Journal of Statistical Software*, 27(5):1–32, 7 2008.

[47] Ludger Hentschel. All in the family nesting symmetric and asymmetric garch models. *Journal of Financial Economics*, 39(1):71–104, September 1995.

[48] Burton Hollifield, Robert Miller, Patrik Sandas, and Joshua Slive. Liquidity supply and demand: Empirical evidence from the vancouver stock exchange. GSIA Working Papers 1999-E19, Carnegie Mellon University, Tepper School of Business, March 2001.

[49] Burton Hollifield, Robert A. Miller, and Patrik Sandas. Empirical analysis of limit order markets. *Review of Economic Studies*, 71(4):1027–1063, October 2004.

[50] Patrik Idvall and Conny Jonsson. Algorithmic trading: Hidden markov models on foreign exchange data. Master's thesis, Department of Mathematics, Linkopings Universitet, Linkopings, Sweden, 2008.

[51] Jonathan E. Ingersoll. *Theory of financial decision making*. Rowman and Littlefield publishers, 1987.

[52] J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. Wiley, New York, 2 edition, 2002.

[53] E. L. Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.

[54] Robert Kissell and Roberto Malamut. Algorithmic decision making framework. *Journal of Trading*, 2006.

[55] T. Lancaster. *The Econometric Analysis of Transition Data*. Cambridge University Press, Cambridge, 1990.

[56] Fabrizio Lillo. Limit order placement as an utility maximization problem and the origin of power law distribution of limit order prices. *The European Physical Journal B*, 55(4):453–459, February 2007.

[57] Fabrizio Lillo and J. Doyne Farmer. The long memory of the efficient market. *Studies in Nonlinear Dynamics & Econometrics*, 8(3):1226–1226, 2004.

[58] Fabrizio Lillo, Szabolcs Mike, and J. Doyne Farmer. Theory for long memory in supply and demand. *Physical Review E*, 71(6):066122, Jun 2005.

[59] G. M. Ljung and G. E. P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

[60] Andrew W. Lo. Long-term memory in stock market prices. *Econometrica*, 59(5):pp. 1279–1313, 1991.

[61] Andrew W. Lo, A. Craig MacKinlay, and June Zhang. Econometric models of limit-order executions. *Journal of Financial Economics*, 65(1):31–71, July 2002.

[62] Ingrid Lo and Stephen Sapp. Price aggressiveness and quantity: how are they determined in a limit order market? In *The Microstructure of Equity and Currency Markets*, Oslo, Norway, September 2005. Norges Bank/BI (Norwegian School of Management).

[63] Hugh Luckock. A steady-state model of the continuous double auction. *Quantitative Finance*, 3:385–404, Oct 2003.

[64] A. Lunde. A generalized gamma autoregressive conditional duration model. Working paper, Department of Economics, Politics and Public Administration, Aalborg University, Denmark, 1999.

[65] N. MANTEL and W. HAENSZEL. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22(4):719–748, April 1959.

[66] Sergei Maslov and Mark Mills. Price fluctuations from the order book perspective - empirical facts and a simple model. *Physica A*, 229:234–246, 2001.

[67] Lukas Menkhoff, Carol L. Osler, and Maik Schmeling. Order-choice dynamics under asymmetric information: An empirical analysis. In *12th Conference of the Swiss Society for Financial Market Research*, SWX Swiss Exchange, Zrich, April 2008.

[68] Szabolcs Mike and J. Doyne Farmer. An empirical behavioral model of price formation. Quantitative Finance Papers physics/0509194, arXiv.org, September 2005.

[69] Szabolcs Mike and J. Doyne Farmer. An empirical behavioral model of liquidity and volatility. *Journal of Economic Dynamics and Control*, 32(1):200–234, January 2008.

[70] Daniel B Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–70, March 1991.

[71] W. Nelson. Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14:945–965, 1972.

[72] Yuriy Nevmyvaka, Yi Feng, and Michael Kearns. Reinforcement learning for optimized trade execution. In *Proceedings of the 23rd international conference on Machine learning (ICML06)*, pages 673–680, New York, NY, USA, 2006. ACM.

[73] Yuriy Nevmyvaka, Michael Kearns, Amy Papandreou, and Katia Sycara. Electronic trading in order-driven markets: efficient execution. In *Proceedings of the 7th IEEE International Conference on E-Commerce Technology (CEC05)*, pages 190–197, 2005.

[74] Keiichi Omura, Yasuhiko Tanigawa, and Jun Uno. Execution Probability of Limit Orders on the Tokyo Stock Exchange. *SSRN eLibrary*, 2000.

[75] Alan Pankratz. *Forecasting with univariate Box-Jenkins models : concepts and cases / Alan Pankratz*. Wiley, New York :, 1983.

[76] Christine A Parlour. Price dynamics in limit order markets. *Review of Financial Studies*, 11(4):789–816, 1998.

[77] Marc Potters and Jean-Philippe Bouchaud. More statistical properties of order books and price impact. *Physica A*, 324:113–140, 2003.

[78] Jens T. Præstgaard. Permutation and bootstrap Kolmogorov-Smirnov tests for the equality of two distributions. *Scandinavian Journal of Statistics*, 22(3):305–322, 1995.

[79] Tobias Preis, Sebastian Golke, Wolfgang Paul, and Johannes J. Schneider. Statistical analysis of financial returns for a multiagent order book model of asset trading. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 76(1):016108, 2007.

[80] R. L. Prentice. Discrimination among some parametric models. *Biometrika*, 62(3):607–614, December 1975.

[81] Marco Raberto, Enrico Scalas, and Francesco Mainardi. Waiting-times and returns in high-frequency financial data: an empirical study. *Physica A*, 314:749–755, 2002.

[82] Angelo Ranaldo. Order aggressiveness in limit order book markets. *Journal of Financial Markets*, 7(1):53–74, January 2004.

[83] Ioanid Rosu. A dynamic model of the limit order book. *Review of Financial Studies*, 22(11):4601–4641, November 2009.

[84] Enrico Scalas, Taisei Kaizoji, Michael Kirchler, Jrgen Huber, and Alessandra Tedeschi. Waiting times between orders and trades in double-auction markets. *Physica A*, 366:463–471, 2006.

[85] David Schoenfeld. Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1):239–241, April 1982.

[86] Duane J. Seppi. Liquidity provision with limit orders and strategic specialist. *The Review of Financial Studies*, 10(1):103–150, 1997.

[87] Joshua Slive. Dynamic strategies in a limit order market. In *Annual Meetings of The Eastern Finance Association 2008*, St. Pete Beach, Florida, USA, April 2008.

[88] Eric Smith, J. Doyne Farmer, LÁszlÓ Gillemot, and Supriya Krishnamurthy. Statistical theory of the continuous double auction. *Quantitative Finance*, 3(6):481–514, 2003.

[89] M. A. Stephens. Edf statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347):pp. 730–737, 1974.

[90] Alexandre A. Trindade, Yun Zhu, and Beth Andrews. Time series models with asymmetric laplace innovations. *Journal of Statistical Computation and Simulation*, 2009.

[91] Peter Verhoeven, Simon Ching, and Hock Guan Ng. Determinants of the decision to submit market or limit orders on the asx. *Pacific-Basin Finance Journal*, 12(1):1–18, January 2004.

[92] Jiaqi Wang and Chengqi Zhang. Dynamic focus strategies for electronic trade execution in limit order markets. In *CEC-EEE '06: Proceedings of the The 8th IEEE International Conference on E-Commerce Technology and The 3rd IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services*, page 26, Washington, DC, USA, 2006. IEEE Computer Society.

[93] Paul Wilmott. *Paul Wilmott on Quantitative Finance 3 Volume Set (2nd Edition)*. Wiley, 2006.

[94] Song Yang. Some scale estimators and lack-of-fit tests for the censored two-sample accelerated life model. *Biometrics*, 54(3):1040–1052, 1998.

[95] I. Zovko and J. Doyne Farmer. The power of patience; A behavioral regularity in limit order placement. *Quantitative Finance*, 2:387–392, 2002.