Delft University of Technology
Faculty of EEMCS

Fill Probabilities in a Limit Order Book with Applications to Foreign Exchange Spot Markets

by
F.J.P. Lokin

to obtain the degree of Master of Science in Applied Mathematics - Financial Engineering
at the Delft University of Technology,
to be defended publicly on Friday September 15, 2023 at 9:00 AM.

Student number:	4607775		
Project duration:	January 2, 2023 – September 15, 2023		
Thesis committee:	Dr. F. Yu,	TU Delft,	daily supervisor
	Prof. Dr. Ir. C. Vuik,	TU Delft,	responsible professor
	E. Hazeveld,	MN,	external supervisor

Abstract

Financial markets continue to see an increase in the share of trades executed by algorithmic trading systems. A key component of an efficient algorithmic trading system is its ability to accurately estimate the probability an order will be executed: the fill probability. This thesis aims to determine whether the dynamics of the order book can be captured accurately by a stochastic model, which subsequently leads to the following research question:

Can we accurately compute fill probabilities of limit orders, conditional on the state of the order book?

Our research builds upon the stochastic model proposed by Cont, Stoikov and Talreja, which assumes orders and cancellations have unit size and arrive according to independent Poisson processes [12]. There are two main advantages of the model proposed by Cont et al., which include:

- Straightforward parameter estimation based on observable order book data,
- The possibility to compute various probabilities of interest semi-analytically.

These probabilities include the fill probability of limit orders and the direction of the next mid-price move.

In this model, the number of orders at each price level in the order book is modelled as a birth-death process, in which births represent an increase in the number of orders and conversely, deaths represent a decrease. The first-passage time is the time it takes for such a process to reach a certain state for the first time. Since the order arrivals follow Poisson processes, we can find an expression for the Laplace transforms of the density functions of the first-passage times. By numerically inverting these Laplace transforms we can calculate the probabilities of interest. The parameters of the Poisson processes that determine the arrivals of orders are calibrated using observable order book data of the Euro-US Dollar currency pair of the foreign exchange (FX) spot market.

We extend the model of Cont et al. by integrating spread-dependent arrival rates of orders and cancellations, meaning that the arrival rates change for different sizes of the spread. This extension is based on the order book data, which clearly shows a difference in arrival rates for various spread sizes. Additionally, we provide an expression to compute fill probabilities of orders posted not only at the best bid price, but also for one price level below. This expression can be generalised to obtain an expression for the fill probabilities of orders posted even deeper in the order book. Finally, by employing this stochastic model, we also show that this model has the potential to be applied to asset classes other than equities, which in our case is the FX market.

To evaluate this model, we compare the computed probabilities with the empirical probabilities based on the FX data. Our findings indicate that the model has the ability to effectively capture the dynamics at each price level and the short-term movements of the mid-price. We are able to show that the fill probability for orders posted at the best quote can be estimated with reasonable accuracy when the order has the highest priority of execution. For orders with lower priority, the model is not able to capture all the dynamics, resulting in most cases in an overestimation of the fill probability. Additionally, we evaluate the method for computing the fill probability at one price level lower than the best quote. We also compare the results of this method to empirical probabilities. However, it is challenging to evaluate the performance of the model due to limited data availability and the small magnitude of the computed probabilities, which are typically less than 1.0%, resulting in large relative errors.

Overall, this research contributes to better understanding order book dynamics and provides insights on the precision of the computation of fill probabilities via an extension of [12]. Further improvements and refinement of the model, such as a more realistic order flow or allowing multiple order sizes, could lead to more accurate estimations and a better comprehension of the market dynamics and fill probabilities.

Keywords: Limit Order Books, Fill Probabilities, Stochastic Modelling, Foreign Exchange Market, Laplace Transforms, Continued Fractions, Queueing Systems, Birth-Death Process, Algorithmic Trading

Acknowledgements

This thesis before you is written to obtain the degree Master of Science in Applied Mathematics – Financial Engineering at TU Delft. As a thesis intern I was part of the Academic Excellence Program (AXP), a collaboration between MN and PGGM, which gave me the opportunity to put the knowledge I have gathered as a mathematics student over the past seven years into practice. This thesis also marks the end of my time as a student in Delft, which I have thoroughly enjoyed throughout.

I would like to thank dr. Fenghui Yu for her time and guidance during this project and for the useful discussions we have had regarding the project. I would like to thank Erwin Hazeveld for his guidance and knowledge, which provided useful insights regarding the business side and applicability of this thesis, and most importantly for granting me the opportunity to conduct this research. I would also like to express my gratitude towards dr. Fang Fang for her advise and insights and towards prof. Kees Vuik for his role as chairman of the thesis committee.

I wish you a pleasant read,

Felix Lokin
Delft, September 2023

Contents

Acronyms

AFM Authority for Financial Markets. 4

cdf cumulative distribution function. 20

FIFO First In, First Out. 2

Forex Foreign Exchange. 3

FX Foreign Exchange. 3

HFT High Frequency Trading. 4

LOB limit order book. 1

MAAPE mean arctangent absolute percentage error. 48

MAPE mean absolute percentage error. 46

OTC over-the-counter. 3

pdf probability density function. 12

Glossary

best ask price lowest price for which a sell limit order is submitted. 1

best bid price highest price for which a buy limit order is submitted. 1

best quote most favorable bid or ask price. 2

bid-ask spread difference between the best bid price and best ask price. 1

limit order order to trade a certain quantity of an asset for a pre-specified price. 1

market order order to immediately trade a certain quantity of an asset. 2

mid-price average of the best bid price and best ask price. 1

spread difference between the best bid price and best ask price. 1, *see* bid-ask spread

tick size smallest possible price increment in the order book. 2

List of Figures

List of Tables

1 | Introduction

Driven by recent technological developments, algorithmic trading now accounts for an extensive portion of all trading activity across global financial markets, particularly highly liquid[1] markets. Applications of algorithmic systems range from detecting opportunities that could generate profit to optimising execution and reducing transaction costs. This thesis focuses on the latter and aims to provide an insight on the fill probabilities of orders in the order book, which plays a crucial role in the execution optimisation in this complex environment. The fill probability, or execution probability, refers to the likelihood that a limit order is executed. This probability is affected by both intrinsic characteristics of an order, such as the price and quantity, and by external factors, mainly determined by the market conditions at the time the order is submitted. Estimating these probabilities is a complicated task, due to the nature of the order book which changes at a very high frequency. However, accurately predicting the filling probability is a key component of effective algorithmic trading, as it can significantly reduce the cost of execution. In the following sections we first describe the structure of the order book and its dynamics in more detail and provide some more background on algorithmic and high-frequency trading. We then discuss the existing literature regarding order book modelling and estimating fill probabilities and provide the motivations behind our choice to use a stochastic model to describe the order book dynamics and compute the fill probabilities on that basis.

1.1 Limit Order Books

In an *order driven market*, market participants can place orders to indicate that they want to trade a certain amount of a security. A participant can submit a limit order, indicating he or she wants to trade a certain quantity at a specified price. Limit orders are passive, liquidity providing orders. All incoming limit orders are collected in the limit order book (LOB). A visual representation of the limit order book is given in Figure 1.1. Buy orders are collected on the bid side, while sell orders are collected on the ask side. The highest (resp. lowest) price for which a buy (resp. sell) limit order is submitted is known as the best bid price P^{Bid} (resp. best ask price P^{Ask}). The difference between the best bid price and best ask price is known as the bid-ask spread or spread S, with

$$S \equiv P^{\text{Ask}} - P^{\text{Bid}}. \tag{1.1}$$

The mid-price is the average of the best ask and best bid price, given by

$$P^{\text{Mid}} \equiv \frac{P^{\text{Ask}} + P^{\text{Bid}}}{2}. \tag{1.2}$$

In some cases it can be useful to calculate the micro-price or weighted mid-price, which also incorporates information about the quantities at the best bid and best ask. The micro-price P^{Micro} is given by

$$P^{\text{Micro}} \equiv \frac{q^{\text{Bid}} P^{\text{Ask}} + q^{\text{Ask}} P^{\text{Bid}}}{q^{\text{Ask}} + q^{\text{Bid}}}, \tag{1.3}$$

where q^{Ask} and q^{Bid} denote the quantities at the best ask and best bid, respectively. Orders submitted for prices lower than the best bid price on the bid side and higher than the best ask price on the ask

[1] The term liquidity refers to the extent at which an asset can be traded quickly without significantly affecting its price. Typically, a large number of buyers and sellers is active in liquid markets.

side determine the depth of the order book. The smallest increment between two prices is determined by the tick size. For example, if the tick size is 0.01 then orders can be submitted at 1.05 and 1.06, but not at 1.055.

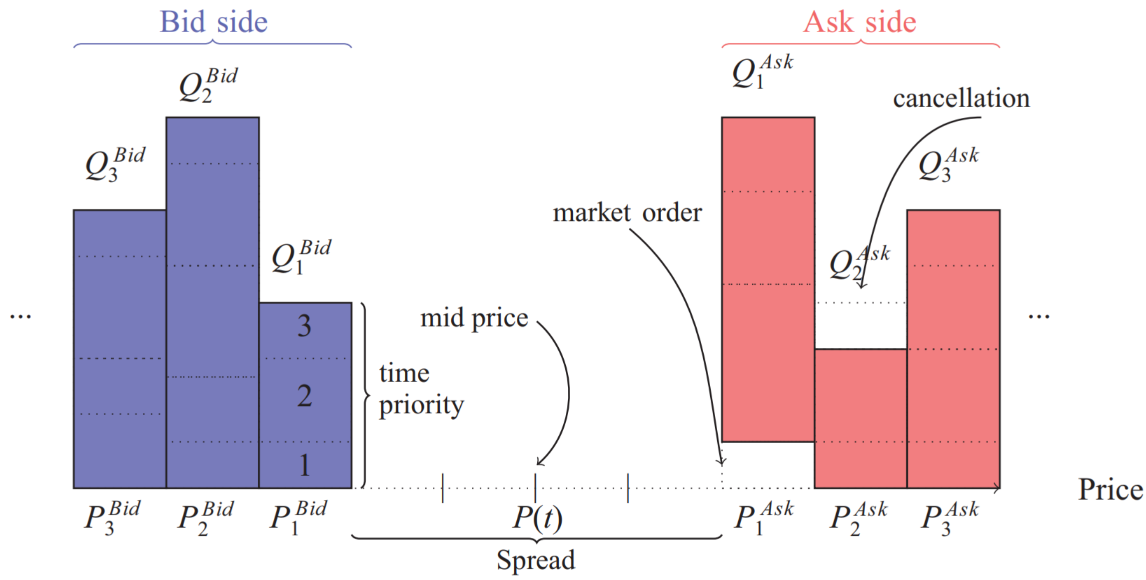


Figure 1.1: Visual representation of the first three levels of a limit order book. P_1^{Bid} and P_1^{Ask} denote the best bid and best ask price, respectively. Prices deeper in the order book (i.e. at price levels P_i, $i > 1$) are decreasing for the bid side and increasing for the ask side. The quantities at each level i of the order book are given by Q_i^{Bid} for the bid side and by Q_i^{Ask} for the ask side. Retrieved from: [18, p. 369].

Limit orders collected in the LOB stay there until they are cancelled, which is allowed at any time, or executed. An execution or transaction occurs when a limit order is matched with an incoming market order. Market orders are aggressive, liquidity taking orders. A trader submits a market order in the case he demands immediate execution and the order is then filled at the best available price. In most markets, limit orders are executed against market orders following the time priority rule for orders submitted at the same price level. The time priority rule states that for orders submitted at the same price, the order that was submitted earlier will be filled first. This is also known as the First In, First Out (FIFO) rule. Upon arrival, limit orders are thus placed at the back of a queue at the price level for which the order is submitted, and only move forward if orders with a higher priority are either filled or cancelled. Other priority rules also exist, such as size-based priority or pro-rata matching. With size-based priority, larger orders are prioritised over smaller orders. A pro-rata matching matches the incoming market order with limit orders proportional to their size [10]. In this thesis we will assume that orders are executed according to the time priority rule, since this is a necessary assumption for our order book model, and furthermore, this is the priority rule on the trading venue from which we have retrieved the data.

Sizes of incoming market orders and outstanding limit orders are generally not equal. Suppose an incoming market order is of size Q_m and a limit order with the highest priority is of size Q_l. If $Q_m > Q_l$, then the limit order is filled completely, and the remainder of the market order $Q_m - Q_l$ is matched with the next available limit order, according to the FIFO rule. In case $Q_l > Q_m$, the limit order is only partially filled, and the remainder $Q_l - Q_m$ remains in the order book until it is filled or cancelled. Table 1.1 shows an example of the change in the order book after a sell market order of 5 million is submitted.

When submitting limit orders, a market participant needs to consider the trade-off between fast execution and a better price for the order. An order submitted for a price near or at the best quote will have a higher chance of getting filled, since market orders are always matched with orders with the best possible price. On the other hand, orders placed deeper in the order book are less likely to be filled, but when filled it will be at a price more favourable to the trader placing the limit order. Traders seek to strike the balance between using market orders, which ensures execution but potentially at a less favourable price, and limit orders, which may lead to a better price but also carry the risk of non-execution. A key factor in this choice is the fill probability of a limit order, and for that reason we aim to obtain a better understanding of this probability in this thesis.

Table 1.1: Example of the change in the order book after the arrival of a sell market order of 5 million. Since the size of limit order with the highest priority (8 million) is larger than the size of the incoming market order (5 million), the difference of 3 million remains in the order book.

	Amount (in millions)	**Price**
	3	1.104
	0	1.103
Ask	8	1.102
	2	1.101
	5	1.100
Spread		
	8	1.099
	1	1.098
Bid	1	1.097
	4	1.096
	6	1.095

(a) Order book before the arrival of a sell market order.

	Amount (in millions)	**Price**
	3	1.104
	0	1.103
Ask	8	1.102
	2	1.101
	5	1.100
Spread		
	3	1.099
	1	1.098
Bid	1	1.097
	4	1.096
	6	1.095

(b) Order book after the arrival of a sell market order.

1.2 Foreign Exchange Market

On the foreign exchange market, also referred to as FX or Forex market, traders can buy, sell and exchange currencies. The price of each currency is determined by the exchange rate, which denotes its value with respect to another currency. For example, if the EURUSD rate is 1.1, then 1 Euro is worth 1.1 US Dollars. The same applies to all other possible currency pairs.

Currencies are traded both via exchanges and over-the-counter (OTC). The latter is significantly larger than the former and operates 24 hours a day and five days a week. In an OTC market, participants can trade securities without the involvement of a central exchange. The FX market is the most liquid and - in terms of trading volume - largest market in the world. In April 2022, the average daily turnover in OTC FX markets was USD 7.5 trillion, of which 28% were FX spot trades [6]. Unlike for example the stock market, the OTC FX market is *decentralised*. This means that there is not a central location where currencies are traded, but trades can be done directly between two participants. In general, decentralised markets have lower transaction costs, but they can also be more complex and less regulated than centralised markets due to a lack of transparency.

Foreign Exchange Spot Market

On the FX spot market currencies are traded ‘immediately’ for the current price, also known as the spot price or spot rate. Most transactions on the FX spot market have a settlement date of $t + 2$, meaning that the day of the actual exchange is two (business) days after the trade date t. An exception to this is the US Dollar and Canadian Dollar (USDCAD) currency pair, which has a settlement date of one business day. The EURUSD, GBPUSD and USDJPY are the most actively traded currency pairs [6].

1.3 Algorithmic and High-Frequency Trading

Using computing algorithms and mathematical models to execute orders is referred to as algorithmic trading. The algorithms make decisions based on input parameters submitted by traders and market conditions such as volatility, time of day etc. These decisions are often made on very short timescales. Algorithmic trading systems have a variety of applications, such as the detection of arbitrage opportunities[2], the optimal execution of orders, detecting and exploiting the strategies of competitors and disguising own intentions. All sorts of market participants, and both retail and institutional investors[3] make use of algorithmic trading. The Dutch Authority for Financial Markets (AFM) distinguishes between two terms: algorithmic trading and algorithmic execution [4]. Both definitions will be discussed shortly.
Algorithmic execution is employed to execute orders based on predetermined investment strategies or decisions. The most common use of algorithmic execution systems is to place the orders in a way that minimises costs and the market impact. The algorithmic execution systems of MN belong to this type of algorithmic trading. First, a decision is made to buy or sell a certain amount of a certain currency. The algorithmic execution system is then used to optimise the trading process.
Algorithmic trading, on the other hand, seeks to automate a specific investment strategy, which includes the execution as a part of the algorithm. In contrast to algorithmic execution, the system makes the investment decision itself based on the input from the market and the constraints in the model.
High-frequency trading (HFT) is an area of algorithmic trading in which the decisions of an algorithm are made on extremely short time scales, often in the order of milliseconds. Using extremely fast signal processing, traders try to outpace other market participants when opportunities arise.

The part of the trading activity by algorithmic systems is growing increasingly. In November 2020, 74% of all transactions for all asset classes[4] on Dutch trading venues was executed by algorithmic trading systems [4, p. 5].

The AFM also identifies some advantages and potential risks associated with algorithmic trading. On the one hand, algorithmic trading has resulted in increased speed and faster transfer of risks by HFT market makers[5]. Together with the growing competition between market participants, this has significantly reduced spreads for a number of financial instruments. As a result, financial instruments can be bought and sold easier and cheaper, which has increased the overall market liquidity. Another advantage is that algorithmic trading systems act as programmed and do not make decisions that are based on emotion. For this reason, errors that result from human emotion can be avoided.

On the other hand, the AFM points out that there are certain risks associated with algorithmic trading. First of all, algorithmic systems can lead to market disruption, which can impact the stability and efficiency of the market significantly. During flash crashes, for example, the price of an asset drops sharply during a short period of time, which can be a result of acts of algorithmic trading systems. Another point of attention is that trading algorithms have the ability to learn to manipulate the market. This concern is fueled by the development in machine learning technologies, especially the subtype Reinforcement Learning, where algorithmic systems could learn to manipulate the market and get rewarded for it, even if it is not the intention of its developers. A last concern is that algorithmic trading systems could affect the quality of the market. This quality is assessed by metrics such as the bid-ask spread, order book depth and market efficiency. Algorithms could impact this quality either intentionally or unintentionally.

[2]An arbitrage opportunity refers to a situation in which market participants are able to profit from - mostly short term - anomalies in the market.

[3]Retail investors or individual investors invest on their own behalf and with their own money. Institutional investors, such as hedge funds or pension funds, invest money on behalf of their clients.

[4]An asset class is a group of financial instruments with similar financial behaviour and characteristics. Examples are equities, bonds, commodities and also currencies.

[5]A market maker is a financial entity that provides liquidity to the market by buying and selling securities and profits from the differences between bid and ask prices.

1.4 Related Literature

Various approaches for estimating filling probabilities can be found in academic literature. We present an overview of the most relevant methodologies in this section. These methodologies include a simplified expression for the filling probabilities, often used in optimisation problems, econometric models based on a statistical method known as survival analysis, machine learning methods, and stochastic models that are used to describe the dynamics of the order book, which can be used to predict the fill probabilities.

First of all, most optimisation problems in which optimal trading strategies are derived, such as the optimal liquidation of a portfolio, assume the filling probability to be exponential and dependent on the distance a limit order is posted from the best price. According to this assumption, the fill probability decreases for orders posted at a larger distance from the best quote. A decay parameter determines the rate of decrease in probability for each tick away from the best quote. The main advantage of this approach is its simplicity, making it a particularly useful method for solving optimisation problems. A simplified fill probability ensures that the problem remains mathematically tractable. However, this approach does not take into account important factors that can potentially influence the fill probability, such as time and the state of the order book. This simplified approach has been used in various studies, including those referenced as [7], [8] and [15].

Another approach used in earlier studies involves econometric models. These models analyse historical data to predict future behaviour of the order book. Relationships between metrics that could potentially influence fill probabilities can be defined using regression models. These models are generally easy to interpret and relationships between variables are clearly defined. On the other hand, they may fail to capture all the complexities of the order book, as well as the interactions between variables. Econometric models often rely on a statistical method known as survival analysis. Survival analysis can be used to estimate the expected duration of time for an event to occur. In this particular application, the 'survival time' of a limit order can be seen as the total time an order is in the order book before it is executed. Cho and Nelling assume that market orders arrive following a non-homogeneous Poisson process, implying that the survival function for time to execution follows a Weibull distribution [9]. The authors treat cancellations as censored and limit orders that were partially filled as if the size of the execution was the initial order size. Furthermore, the weighted average of the filling times was computed for orders that were matched with multiple market orders, where the weight is determined by the size of each execution. Using data from the New York Stock Exchange, Cho and Nelling find that the likelihood of execution decreases as a limit order remains in the book for longer. Moreover, buy orders are less likely to be executed than sell orders and the fill probability is lower for orders posted at large distance from the best price and for larger orders. Conversely, periods of high price volatility and wider spreads tend to increase the execution probability. Lo, MacKinlay and Zhang propose a model based on survival analysis and they are able to compute the distribution of execution times, conditional on various (economic) factors [19]. These factors include the limit price, size of the order, spread and market volatility. The authors treat expired or cancelled orders as censored observations. Using stock data from the S&P 500, Lo et al. compute the distribution for time-to-first-fill and time-to-completion for buy and sell limit orders. They assume the filling time follows a generalised gamma distribution with an accelerated failure time. The accelerated failure time determines the influence of certain explanatory variables on the fill time. The authors find that the probability is influenced considerably by certain variables, such as the limit order price, but less so by others, such as the order size.

Recent developments in the field of artificial intelligence and machine learning have lead to more research regarding its applications to estimating filling probabilities. Unlike for example the econometric models, machine learning models are capable of capturing complex and possibly non-linear relationships between explanatory variables. These models have the ability to make accurate predictions, even under rapidly changing circumstances. Drawbacks of this approach are the need for extensive training data and considerable computational resources. Furthermore, the reasoning behind estimated probabilities can be difficult to interpret due to the black-box nature of these type of models. Maglaras et al. integrate machine learning methods with the concept of survival analysis to estimate filling probabilities and filling times [21]. The authors use a Recurrent Neural Network to estimate the hazard rate parameters of the survival function, which determines the distribution of time duration for a limit order fill. This distribution is conditioned on the current market conditions, such as the imbalance in the order book. The model is trained with historical NASDAQ data and using synthetic limit orders, which are assumed to be of infinitesimal size and to have no market impact. The limit orders are tracked over time and when fill conditions are met, an order is treated as executed. Maglaras et al. find that their model outperforms benchmark models in terms of accuracy of the predictions. A second research that uses machine learning techniques, conducted by Arroyo et al., also builds on survival analysis [5]. The authors base their method on a convolutional-Transformer encoder and a monotonic neural network decoder. The model is trained using proper scoring rules for survival analysis, and they find that their model outperforms benchmark models significantly.

The last type of models used for predicting filling probabilities are stochastic models. The nature of the order book and its dynamics make it a suitable candidate for stochastic modelling. The randomness of arrivals of limit orders, market orders and cancellations can be captured and used for simulating the dynamics of an order book. Since these types of events are modelled directly, it can produce accurate predictions of execution probabilities. Furthermore, analytical formulas for these probabilities can be derived. The disadvantage of stochastic models is that they often rely on complex mathematical theory and on strong, simplifying assumptions on the dynamics of the order book, which are needed to preserve the mathematical tractability. A statistical model for order book dynamics is developed by Smith et. al, under the assumption that the order flow is random and independent and identically distributed [23]. The authors predict observable metrics such as price volatility, spread etc. based on the properties of the order book and order flow, as well as the filling probability. They find that orders placed closer to the best price are more likely to be filled. They also find that the probability of execution for orders placed close to the opposite best quote, i.e. sell orders placed close to the best bid and vice versa, is smaller than 1, meaning that there is a substantial probability that the order is cancelled before it is executed. Smith et al. conclude that although most assumptions are not fully realistic, stochastic models are still able to capture market dynamics reasonably well. Cont, Stoikov and Talreja argue that due to the nature of its dynamics, the limit order book can be modelled as a queueing system [12]. Limit orders arrive at the order book with a certain rate and then remain in the queue until they are either cancelled or executed. They assume the orders are of unit size and arrive according to a Poisson process. The queues at each price level can be modelled as a birth-death process, where incoming limit orders are seen as 'births', while cancellations and market orders result in 'deaths'. Using Laplace transforms, the authors are able to compute several conditional probabilities, including the direction of the next mid-price movement and probability of execution before the next mid-price movement. The main advantages of the approach presented in this paper are the analytical tractability and the simple estimation of parameters using order book data. Cont et al. estimate the parameters on data from the Tokyo stock exchange and conclude that their model is able to make short-term predictions that are reasonably accurate. Huang and Kercheval seek to improve this model by incorporating limit orders with multiple sizes [16]. The authors compute the same conditional probabilities as Cont et al., but do not assess how well their calculations compare to empirical probabilities.

1.5 Model Selection and Main Contributions

For this thesis we have chosen to model the dynamics of the order book using a stochastic model which is an extension of the model proposed by Cont, Stoikov and Talreja [12]. Our choice follows from two main advantages of this model, namely

- the possibility to compute the probabilities of interest semi-analytically,
- the straightforward parameter estimation using observable order book data.

We expect that computing the probabilities on the deepest level, i.e. not only taking into account the price level but also the underlying quantities of orders outstanding at these price levels, will provide accurate results. By employing this stochastic approach, we are able to directly model the randomness and uncertainty of the order book. The proposed model also integrates market conditions such as order flow, imbalance and order book depth, which all influence filling probabilities.

We extend the research of Cont et al. by incorporating state dependent arrival rates of orders and cancellations. This extension follows from our findings that the arrival rates of orders vary substantially for different spread sizes. Furthermore, we provide a general expression to compute the filling probabilities not only at the best bid, as found by Cont et al., but also at one price level below the best bid. This method can be extended to price levels even deeper in the order book, but it is worth noting that the complexity of the analytical formula will increase quite rapidly. This thesis fills a gap in research, since to the best of our knowledge there are no analytical expressions to compute the fill probability at deeper levels of the order book, conditional on the state of the order book. Finally, by using data from the foreign exchange spot market, we also show that the model has the potential to be applied to asset classes other than equities.

With the previous in mind, the main research topic of this thesis is:

Can we accurately compute fill probabilities of limit orders, conditional on the state of the order book?

We will focus on the fill probabilities for orders placed at the best bid price or one price level below the best bid price. By symmetry of the model, the results are the same for orders placed on the ask side of the order book. To tackle this problem, we will first evaluate to what extend the dynamics of the order book can be captured by a simplified stochastic model.

1.6 Outline

This thesis is organised as follows. The stochastic order book model which will be used to capture the dynamics of the order book is described in Chapter 2. Some preliminary mathematical concepts that are used for the computation of the probabilities are provided in Chapter 3. These concepts include Laplace transforms, continued fractions and first-passage times of birth-death processes. Chapter 4 provides the methodology for the computation of the order book probabilities, including the probability of a mid-price increase and the fill probability of limit orders, which are conditional on the state of the order book. A description of the data, together with the analysis on the model assumptions and the model calibration is given in Chapter 5. In Chapter 6, we use the results of Chapter 4 and Chapter 5 to obtain the numerical results for the various probabilities of interest, including the probability of executing an order at the best bid price and one price level below the best bid price. The computed probabilities are compared to the empirical probabilities which are estimated from the data. We present concluding remarks, as well as a discussion on our findings in Chapter 7.

2 | Limit Order Book Model

In this chapter we will describe a stochastic model for the order book dynamics as proposed by Cont, Stoikov and Talreja in [12]. The authors argue that the order book can be modelled as a queuing system due to the nature of its dynamics. Limit orders arrive and stay in the queue until they are either executed or cancelled. The rates at which limit and market orders arrive and the rate at which outstanding limit orders are cancelled can be determined using observable order book data. These rates can then be used to model the arrivals of different events as point processes. Cont et al. use the class of Poisson point processes to model the order book as a continuous-time Markov process which tracks the number of outstanding limit orders at each price level. Section 2.1 describes how the order book can be described as a continuous time Markov-Process. In Section 2.2 we explain how the dynamics of the order book are captured, i.e. how the order book changes as a result of incoming orders and cancellations.

2.1 Limit Order Book as a Stochastic Model

We consider a limit order book as described in Section 1.1, where market participants can place limit orders at price levels, which are multiples of the tick size. These price levels are given by a *price grid* $\{1, ..., n\}$, where the upper boundary n is chosen to be sufficiently large such that the probability that a limit order will be placed at a level $i > n$ is close to 0 within the analysed time-frame. Following the notation in [12], the state of the order book is monitored through a continuous-time process

$$X(t) \equiv (X_1(t), ..., X_n(t))_{t \geq 0}, \tag{2.1}$$

where $|X_p(t)|$ is the number of outstanding limit orders at time t at price level p, with $1 \leq p \leq n$. To make a distinction between the price levels where bid orders are outstanding and price levels where sell orders are outstanding, the bid levels are denoted by negative quantities. That is, if $X_p(t) < 0$, there are $-X_p(t)$ bid orders at price level p and similarly if $X_p(t) > 0$, there are $X_p(t)$ ask orders at price level p. The best ask price $p_A(t)$ at time t is given by

$$p_A(t) \equiv \inf\{p = 1, ..., n : X_p(t) > 0\} \wedge (n + 1), \tag{2.2}$$

i.e. it is the lowest price level for which there is a positive number of orders outstanding. Since we condition the price levels on having a positive quantity, the orders outstanding on this price level are sell orders by definition. Similarly, the best bid price $p_B(t)$ is given by

$$p_B(t) \equiv \sup\{p = 1, ..., n : X_p(t) < 0\} \vee 0, \tag{2.3}$$

i.e. the highest price level for which there is a negative number of orders outstanding, implying the outstanding orders on this level are buy limit orders. Furthermore, the mid-price $p_M(t)$ and spread $p_S(t)$ at time t are defined by

$$p_M(t) \equiv \frac{p_B(t) + p_A(t)}{2} \quad \text{and} \quad p_S(t) \equiv p_A(t) - p_B(t). \tag{2.4}$$

2.2 Dynamics of Stochastic Limit Order Book Model

The dynamics of the model can be fully described by the following events: incoming limit orders, market orders and cancellations. The model is updated for incoming events as follows: let $X(t_j)$ denote the state of the order book at time t_j, and let $X(t_{j+1})$ denote the state of the order book after the next event, which is either the arrival or cancellation of an order. Assuming that orders have unit size, which is taken to be equal to the average size of the limit orders, orders arrive one by one and the order book changes as a result of the possible events in the following way:

1. Since *limit buy orders* are denoted by negative quantities, the state of the order book at a certain level p_i, denoted by X_{p_i}, is negative for all price levels on the buy side of the order book. For this reason, the arrival of a limit buy order at price p_i, with $p_i < p_A$ results in a decrease of one order (unit) of the quantity of the queue at level p_i:

$$X_{p_i}(t_{j+1}) = X_{p_i}(t_j) - 1.$$

   The condition $p_i < p_A$ implies that bid orders can be placed on any price level below the best ask.

2. Following the same logic, the arrival of a *limit sell order* at price p_i, with $p_i > p_B$ increases the quantity of the queue at price level p_i with one order:

$$X_{p_i}(t_{j+1}) = X_{p_i}(t_j) + 1.$$

   This follows from the fact that the state of the order book at price level p_i is positive for price levels on the ask side of the order book. Again, sell orders can be placed at any price level above the best bid price, hence the condition $p_i > p_B$.

3. Since *market orders* take outstanding limit orders out of the order book and only arrive at the best quotes, buy market orders lead to a decrease of one order of the quantity at the best ask:

$$X_{p_A}(t_{j+1}) = X_{p_A}(t_j) - 1.$$

4. Conversely, sell market orders lead to an increase of the quantity of the queue at the best bid:

$$X_{p_B}(t_{j+1}) = X_{p_B}(t_j) + 1.$$

5. Like market orders, cancellations take outstanding limit orders out of the book. Since the queues at the bid side of the order book are denoted by negative values, a *cancellation of a limit buy order* at price p_i, with $p_i \leq p_B$ increases the quantity of the queue at level p_i by one order:

$$X_{p_i}(t_{j+1}) = X_{p_i}(t_j) + 1.$$

   Note that in this case we have the condition $p_i \leq p_B$ for buy order cancellations, since a cancellation can only arrive at a price level with outstanding orders.

6. Similarly, a *cancellation of a limit sell order* at price p_i, with $p_i \geq p_A$, decreases the quantity of the queue at level p_i by one order:

$$X_{p_i}(t_{j+1}) = X_{p_i}(t_j) - 1.$$

All of the events mentioned above are modelled by independent Poisson processes. The advantage of this is that quantities of interest can be computed analytically using first-passage times of birth-death processes. This will be explained in more detail in Chapter 3. (Homogeneous) Poisson processes can be used to model random points in time, such as the arrival times of orders. We have the following definitions.

Definition 2.1. *A **counting process** is a stochastic process $\{N(t), t \geq 0\}$ with the following properties:*

1. $N(t) \in \mathbb{N}$
2. $N(s) \leq N(t)$ *for* $s \leq t$,

with $\mathbb{N}$ *the set of natural numbers.*

Definition 2.2. *A **homogeneous Poisson process** with rate $\lambda > 0$ is a counting process $N(t)$ with the following properties:*

1. $N(0) = 0$,
2. $N(t)$ *has independent increments,*
3. *the number of events in each interval of length t follows a Poisson distribution with parameter λt.*

The probability that there are n arrivals in $[0, t]$ is then given by:

$$\mathbb{P}(N(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}. \tag{2.5}$$

All order book events are modelled as Poisson processes. As empirical data suggests, the rates of incoming limit orders or cancellations differ for each price level. We will further examine this feature in Chapter 5. The different events arrive at (mutually) independent and exponentially distributed times with the following rates:

- Limit orders arrive with rate $\lambda(\delta)$, with δ the distance in ticks from the *opposite* best price. For example, if the spread equals 2 ticks, then limit orders arrive at the best bid and best ask with rate $\lambda(2)$, and orders are placed inside the spread with rate $2 \cdot \lambda(1)$. Note that we multiply this term by 2, since both buy orders and sell orders can arrive inside the spread.
- Market orders arrive at the best bid and best ask with rate μ.
- Cancellations of limit orders at a distance δ from the opposite best quote arrive at a rate that is *proportional* to the number of orders outstanding at this level k with rate $k\theta(\delta)$. The reasoning behind this is that if there are more orders outstanding at a certain price level, there is also a higher probability of a cancellation. Each order can be cancelled with rate $\theta(\delta)$, so if there are k orders outstanding at a certain price level, the total cancellation rate is $k\theta(\delta)$. Also, there will be no cancellations when there are no outstanding limit orders, since $k = 0$.

The rates are assumed to be the same for the buy and sell side. This assumption will be further elaborated upon in Chapter 5.

We can now describe the dynamics of the order book as a continuous-time Markov process, with state space $\mathbb{Z}^n$ and the following transition rates from time t_j to the time of the next event given by t_{j+1}. For the price levels at which buy limit orders can be submitted, i.e. the price levels p_i below the best ask p_A, we have the transition

$$X_{p_i}(t_{j+1}) = X_{p_i}(t_j) - 1, \text{ with rate } \lambda(p_A(t_j) - p_i), \text{ for } p_i < p_A(t_j). \tag{2.6}$$

As mentioned above, the arrival rate of limit orders at a certain price level depends on the distance δ between this price level and the opposite best quote, which for buy orders is given by $\delta = p_A(t_j) - p_i$. For this reason, the arrival rate is given by $\lambda(\delta) = \lambda(p_A(t_j) - p_i)$. Similarly, ask limit orders can be submitted at price levels larger than the best bid price, $p_i > p_B(t_j)$, and the transitions at each price level p_i are given by

$$X_{p_i}(t_{j+1}) = X_{p_i}(t_j) + 1, \text{ with rate } \lambda(p_i - p_B(t_j)), \text{ for } p_i > p_B(t_j), \tag{2.7}$$

since in this case we have $\delta = p_i - p_B(t_j)$. Market orders arrive only at the best bid and best ask price and therefore we have the following dynamics at the best bid price and best ask price

$$X_{p_B}(t_{j+1}) = X_{p_B}(t_j) + 1, \text{ with rate } \mu, \tag{2.8}$$
$$X_{p_A}(t_{j+1}) = X_{p_A}(t_j) - 1, \text{ with rate } \mu. \tag{2.9}$$

Finally, the cancellation rate $\theta(\delta)$ at each price level p_i depends on both the distance of this price level from the opposite best quote and on the number of orders outstanding at that level. Therefore, the dynamics at each price level are given by

$$X_{p_i}(t_{j+1}) = X_{p_i}(t_j) + 1, \text{ with rate } \theta(p_A(t_j) - p_i)|X_{p_i}(t_j)|, \text{ for } p_i \leq p_B(t_j), \tag{2.10}$$
$$X_{p_i}(t_{j+1}) = X_{p_i}(t_j) - 1, \text{ with rate } \theta(p_i - p_B(t_j))|X_{p_i}(t_j)|, \text{ for } p_i \geq p_A(t_j). \tag{2.11}$$

Note that the absolute value of the quantities is used since quantities on the bid side are denoted by negative values. A schematic representation of the order book dynamics at the best bid and best ask are given in Figure 2.1 and Figure 2.2, respectively.

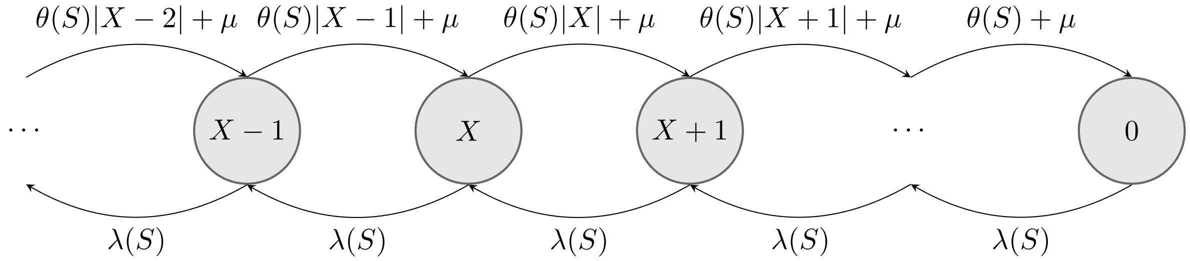


Figure 2.1: Schematic representation of the order book dynamics at the best bid $X = X_{p_B}$.

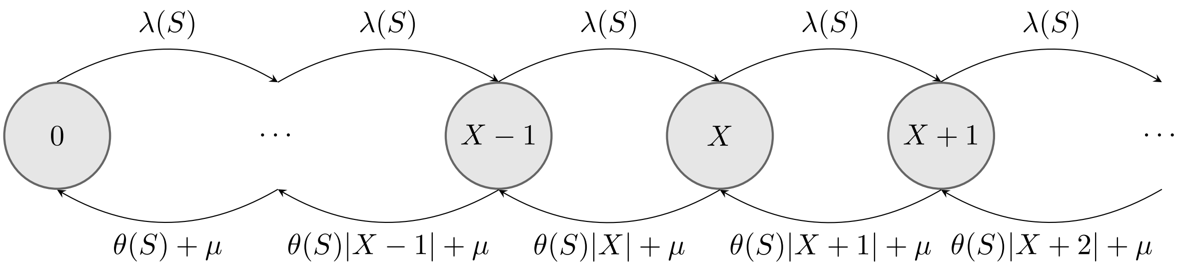


Figure 2.2: Schematic representation of the order book dynamics at the best ask $X = X_{p_A}$.

3 | Preliminary Mathematical Concepts for Computing Probabilities

In this chapter we will introduce some preliminary mathematical concepts that are used to compute the probabilities of interest in the order book. These concepts include Laplace transforms, continued fractions and first-passage times of birth-death processes. Section 3.1 provides a description of Laplace transforms, together with two numerical methods to invert these transforms. We introduce the concept of continued fractions in Section 3.2. Finally, a description of first-passage times of birth-death processes is given in Section 3.3.

3.1 Laplace Transforms

Laplace transforms are a powerful mathematical tool that can be used to transform complex systems. Their main application is to simplify differential equations into a simpler algebraic equation. The (two-sided) Laplace transform $\mathcal{L}[f](s)$ or $\hat{f}(s)$ of a function $f : \mathbb{R} \to \mathbb{R}$ is given by

$$\mathcal{L}[f](s) = \hat{f}(s) = \int_{-\infty}^{\infty} e^{-st} f(t) \mathrm{d}t, \tag{3.1}$$

with $s = \sigma + i\omega$ and i the imaginary unit: $i^2 = -1$. If a random variable X has probability density function (pdf) f, then $\hat{f}$ is the Laplace transform of X. For two independent random variables X and Y whose Laplace transforms are well-defined, we have

$$\hat{f}_{X+Y}(s) = \mathbb{E}[e^{-s(X+Y)}] = \mathbb{E}[e^{-sX}]\mathbb{E}[e^{-sY}] = \hat{f}_X(s)\hat{f}_Y(s). \tag{3.2}$$

The inverse of a Laplace transform $\hat{f}(s)$ is given by the Bromwich contour integral

$$f(t) = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} e^{st} \hat{f}(s) \mathrm{d}s, \tag{3.3}$$

if $f(t)$ is continuous at t and if we have $\int_{-\infty}^{\infty} |\hat{f}(\gamma + i\omega)| \mathrm{d}\omega < \infty$ for some $\gamma \in \mathbb{R}$. Abate and Whitt propose a method for inverting the Laplace transform called the Euler method [2], which is an implementation of the Fourier-series method. The name refers to the Euler summation used in this method. However, due to the nature of the Laplace transform, we propose to use a different method for inverting the Laplace transform. Let $\sigma = 0$, then $s = i\omega$ and

$$\hat{f}(s) = \hat{f}(i\omega) = \int_{-\infty}^{\infty} e^{-i\omega t} f(t) \mathrm{d}t. \tag{3.4}$$

This integral can be approximated using a Fourier-cosine series expansion known as the *COS-method*, which is described below [14]. Like the Euler method, this method is an implementation of the Fourier-series method. The advantage of the COS-method is that in most cases the convergence rate is exponential and the complexity of the computation is linear. This is a result of the close relation between the coefficients of the Fourier-cosine expansion of the density function and the characteristic function, as described in [14]. We extend this by using the relationship between the Laplace transform and the characteristic function to use the COS-method for the computation of the Bromwich contour integral. In the following sections we discuss the Euler method and COS-method in more detail.

Euler Method

In [2], Abate and Whitt show that the Bromwich contour integral can be written as

$$\begin{aligned} f(t) &= \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} e^{st}\hat{f}(s)\mathrm{d}s \\ &= \frac{2e^{\gamma t}}{\pi}\int_0^{\infty}\Re(\hat{f}(\gamma+iu))\cos(ut)\mathrm{d}u, \end{aligned} \tag{3.5}$$

where $\Re(z)$ denotes the real part of a variable z. Using the trapezoidal rule with step size h, the integral can be evaluated by

$$f(t) \approx f_h(t) = \frac{he^{\gamma t}}{\pi}\Re\{\hat{f}(\gamma)\} + \frac{2he^{\gamma t}}{\pi}\sum_{k=1}^{\infty}\Re\{\hat{f}(\gamma+ikh)\}\cos(kht). \tag{3.6}$$

Taking $h = \frac{\pi}{2t}$ and $\gamma = \frac{A}{2t}$ gives us

$$f_h(t) = \frac{e^{A/2}}{2t}\Re\left\{\hat{f}\left(\frac{A}{2t}\right)\right\} + \frac{e^{A/2}}{t}\sum_{k=1}^{\infty}(-1)^k\Re\left\{\hat{f}\left(\frac{A+2k\pi i}{2t}\right)\right\}. \tag{3.7}$$

If $|f(t)| \leq 1$ for all t, which is the case for probability functions, the error is bounded and approximately equal to e^{-A} when e^{-A} is small. To compute Equation 3.7 numerically, Abate and Whitt suggest using Euler summation. The Euler summation is the weighted average of the last m partial sums, where the weights are determined by a Binomial distribution with parameters m and $p = \frac{1}{2}$. The Euler sum approximation of Equation 3.7 is then given by

$$E(m,n,t) = \sum_{k=0}^{m}\binom{m}{k}2^{-m}s_{n+k}(t), \tag{3.8}$$

with

$$s_n(t) = \frac{e^{A/2}}{2t}\Re\left\{\hat{f}\left(\frac{A}{2t}\right)\right\} + \frac{e^{A/2}}{t}\sum_{k=1}^{n}(-1)^k\Re\left\{\hat{f}\left(\frac{A+2k\pi i}{2t}\right)\right\}. \tag{3.9}$$

To achieve a discretisation error of 10^{-8}, Abate and Whitt take $A = 18.4$ and furthermore, for the Euler summation they take $m = 11$ and $n = 15$. For a more concise derivation of this method we refer to [2].

COS-Method

As mentioned previously, the key idea of the COS-method is using the Fourier-cosine expansion to approximate the density function. The Fourier-cosine series expansion of a function $f(t)$ supported on any finite interval $[a,b] \in \mathbb{R}$ is given by

$$f(t) = {\sum_{k=0}^{\infty}}' \bar{A}_k \cdot \cos\left(k\pi\frac{t-a}{b-a}\right), \tag{3.10}$$

with

$$\bar{A}_k = \frac{2}{b-a}\int_a^b f(t)\cos\left(k\pi\frac{t-a}{b-a}\right)\mathrm{d}t. \tag{3.11}$$

Since a real function supported on a finite interval has a cosine expansion, the derivation for the density function begins with truncating the infinite integral in the Bromwich contour integral in Equation 3.3. The integral in this equation has to decay to zero at $\pm\infty$ by the conditions for the existence of a Fourier transform. The integral can therefore be truncated without too much loss of accuracy. If we choose $[a,b] \in \mathbb{R}$ such that the truncated integral approximates the infinite counterpart well, then we have for the approximation of the Laplace transform $\hat{f}^*(s)$

$$\hat{f}^*(s) = \int_a^b e^{-st}f(t)\mathrm{d}t \approx \int_{-\infty}^{\infty} e^{-st}f(t)\mathrm{d}t = \hat{f}(s) \tag{3.12}$$

Furthermore, for a constant $a \in \mathbb{R}$ we have

$$\hat{f}(i\omega)e^{ia} = \mathbb{E}[e^{-i\omega t+ia}] = \int_{-\infty}^{\infty} e^{i(-\omega t+a)} f(t)\mathrm{d}t. \tag{3.13}$$

Substituting the Fourier argument $\omega = -\frac{k\pi}{b-a}$ and multiplying the Fourier transform in Equation 3.12 by $\exp\left(-i\frac{ak\pi}{b-a}\right)$, we obtain

$$\hat{f}^*\left(-i\frac{k\pi}{b-a}\right) \cdot \exp\left(-i\frac{ak\pi}{b-a}\right) = \int_a^b \exp\left(i\frac{tk\pi}{b-a} - i\frac{ak\pi}{b-a}\right) f(t)\mathrm{d}t. \tag{3.14}$$

Taking the real part of both sides and using the Euler formula $e^{iu} = \cos(u) + i\sin(u)$ gives us

$$\Re\left\{\hat{f}^*\left(-i\frac{k\pi}{b-a}\right) \cdot \exp\left(-i\frac{ak\pi}{b-a}\right)\right\} = \int_a^b \cos\left(k\pi\frac{t-a}{b-a}\right) f(t)\mathrm{d}t, \tag{3.15}$$

with $\Re(z)$ the real part of z. We see that $\bar{A}_k$ as in Equation 3.11 can be obtained by multiplying both sides of the equation by $\frac{2}{b-a}$

$$\bar{A}_k = \frac{2}{b-a}\Re\left\{\hat{f}^*\left(-i\frac{k\pi}{b-a}\right) \cdot \exp\left(-i\frac{ak\pi}{b-a}\right)\right\}. \tag{3.16}$$

From Equation 3.12 follows that $\bar{A}_k \approx \bar{F}_k$, with

$$\bar{F}_k := \frac{2}{b-a}\Re\left\{\hat{f}\left(-i\frac{k\pi}{b-a}\right) \cdot \exp\left(-i\frac{ak\pi}{b-a}\right)\right\}. \tag{3.17}$$

For the truncated series summation we obtain

$$f^*(t) \approx {\sum_{k=0}^{N-1}}' \bar{F}_k \cdot \cos\left(k\pi\frac{t-a}{b-a}\right). \tag{3.18}$$

Here, the operator $\sum'$ denotes that the first term in the summation should be multiplied by $\frac{1}{2}$.

An important aspect of the COS-method is to determine the range of integration for $[a,b]$. Fang and Oosterlee propose the following range

$$[a,b] = \left[c_1 - L\cdot\sqrt{c_2+\sqrt{c_4}},\ c_1 + L\cdot\sqrt{c_2+\sqrt{c_4}}\right], \text{ with } L \in [6,12], \tag{3.19}$$

where c_n denotes the n-th cumulant of the underlying stochastic process. These cumulants can be computed using the cumulant-generating function $C_X(t)$, which is given by

$$C_X(t) = \log\mathbb{E}[e^{tX}] = \log\hat{f}(-t), \tag{3.20}$$

with $\hat{f}(s)$ the Laplace transform of a pdf $f(t)$. The n-th cumulant can then be computed by taking the n-th order derivative of $C_X(t)$ and evaluating it at 0, i.e.

$$c_1 = \frac{\mathrm{d}C_X(t)}{\mathrm{d}t}|_{t=0}, \quad c_2 = \frac{\mathrm{d}^2C_X(t)}{\mathrm{d}t^2}|_{t=0}, \quad c_4 = \frac{\mathrm{d}^4C_X(t)}{\mathrm{d}t^4}|_{t=0}. \tag{3.21}$$

For a more detailed description of the COS-method we refer to [14]. A numerical experiment to compare the performances of the Euler method and COS-method is provided in Section 6.2.

3.2 Continued Fractions

We follow the notation of [20] and let $\{f_n\}$ denote the sequence

$$f_n = \cfrac{a_1}{1 + \cfrac{a_2}{1 + \cfrac{a_3}{\ddots + a_n}}}. \tag{3.22}$$

In case we assume $\{a_n\} \neq 0$ for all n and allow $f_n = \infty$, $\{f_n\}$ is well-defined in $\hat{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$, with $\mathbb{C}$ the set of complex numbers. The infinite sequence

$$\mathrm{K}_{n=1}^{\infty} \frac{a_n}{1} = \cfrac{a_1}{1 + \cfrac{a_2}{1 + \cfrac{a_3}{1 + \cdots}}}, \tag{3.23}$$

is called a *continued fraction* and can be written as $f = \mathbf{K}_{n=1}^{\infty}(a_n/1) = \mathbf{K}(a_n/1)$. Similarly, we can construct the following continued fraction

$$\mathrm{K}_{n=1}^{\infty} \frac{a_n}{b_n} = \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{b_3 + \cdots}}}, \tag{3.24}$$

from two sequences $\{a_n\}$ and $\{b_n\}$, with $\{a_n\}$,$\{b_n\} \in \mathbb{C}$ and all $\{a_n\} \neq 0$, and we write $\mathbf{K}(a_n/b_n)$. For simplicity and more efficient notation, we also write

$$\mathrm{K}_{n=1}^{\infty} \frac{a_n}{b_n} = \frac{a_1}{b_1+} \frac{a_2}{b_2+} \frac{a_3}{b_3+} \cdots . \tag{3.25}$$

For the convergence of a continued fraction we have the following results.

Definition 3.1. Lorentzen and Waadeland [20, p. 6] *A continued fraction converges to a value $f \in \hat{\mathbb{C}}$ if* $\lim f^{(k)} = f$*, with $f^{(k)}$ the k-th approximant of the continued fraction f.*

Suppose we have a continued fraction f of the form

$$f = \frac{a_1}{b_1+} \frac{a_2}{b_2+} \frac{a_3}{b_3+} \cdots , \tag{3.26}$$

then the k-th approximant $f^{(k)}$ of f is given by

$$f^{(k)} = \frac{a_1}{b_1+} \frac{a_2}{b_2+} \cdots \frac{a_k}{b_k} = \frac{A_k}{B_k}. \tag{3.27}$$

Here, A_k and B_k satisfy the same recurrence given by

$$\begin{aligned} A_k &= b_k A_{k-1} + a_k A_{k-2}, \text{ and} \\ B_k &= b_k B_{k-1} + a_k B_{k-2}, \end{aligned} \tag{3.28}$$

with initial values $A_0 = 0$, $A_1 = a_1$, $B_0 = 1$, $B_1 = b_1$ [13].

The simplest method to approximate f is to evaluate $f^{(k)}$ by truncating it at a pre-specified number of terms k. However, Crawford and Suchard demonstrate that pre-specifying the number of terms has serious limitations [13]. For this reason, we follow their approach and use the modified Lentz method to evaluate continued fractions of the form of Equation 3.26 [22][25]. The modified Lentz method method stabilises the computation by using the following ratios

$$C_k = \frac{A_k}{A_{k-1}} \text{ and } B_k = \frac{B_{k-1}}{B_k}. \tag{3.29}$$

Rewriting these fractions leads to the following expressions

$$C_k = b_k + \frac{a_k}{C_{k-1}} \tag{3.30}$$

$$D_k = \frac{1}{b_k + a_k D_{k-1}}. \tag{3.31}$$

By iterating, we can evaluate $f^{(k)}$ by

$$f^{(k)} = f^{(k-1)} C_k D_k. \tag{3.32}$$

The algorithm terminates when the difference $|f^{(k)} - f^{(k-1)}|$ is small, i.e. when

$$|C_k D_k - 1| < \epsilon, \tag{3.33}$$

with ϵ a small number.

3.3 First-Passage Times of Birth-Death Processes

In the context of Markov-processes, a birth-death process refers to a process in which the state increases by one unit with the arrival of a 'birth' and decreases by one unit with the arrival of a 'death'. Here, state is referred to as the number of units present in the system at a certain time. We consider such a process where the arrivals of births and deaths are modelled by Poisson processes, and we have a constant birth rate λ and state-dependent death rate μ_i in state $i \geq 1$. The first-passage time of a birth-death process is the total time it takes for the process to move from one state to another for the first time. The first-passage time of this process to zero given that it begins in state b, denoted by σ_b, is given by the sum

$$\sigma_b = \sigma_{b,b-1} + \sigma_{b-1,b-2} + \cdots + \sigma_{1,0}. \tag{3.34}$$

Here, $\sigma_{i,i-1}$ denotes the first-passage time of the birth-death process from state i to state $i-1$. First-passage times can be very useful as they provide information on how quickly the system moves between states, which can help to understand the dynamics of for example a queueing system.

Let $\hat{f}_b$ denote the Laplace transform of σ_b and let $\hat{f}_{i,i-1}$ denote the Laplace transform of $\sigma_{i,i-1}$ for $i = 1, ..., b$, then by Equation 3.2 we have for the Laplace transform of Equation 3.34

$$\hat{f}_b(s) = \prod_{i=1}^{b} \hat{f}_{i,i-1}(s). \tag{3.35}$$

This holds since all the terms on the right-hand side of Equation 3.34, i.e. all separate first passage times $\sigma_{i,i-1}$ for $i = 1, ..., b$, are independent. We now prove that a birth-death process with constant birth rate λ and death rate μ_i in state $i \geq 1$ has the following Laplace transform [1].

Proposition 3.1. *The Laplace transform $\hat{f}_{i,i-1}$ of the density function of the first-passage time $\sigma_{i,i-1}$ of a birth-death process with constant birth rate λ and death rate μ_i in state i is given by*

$$\hat{f}_{i,i-1}(s) = -\frac{1}{\lambda} \mathrm{K}_{k=i}^{\infty} \frac{-\lambda\mu_k}{\lambda + \mu_k + s}, \tag{3.36}$$

Proof. Consider the first-passage time of state i to state $i-1$, denoted by $\sigma_{i,i-1}$. Let λ be the birth rate and μ_i the death rate in state $i \geq 1$. Since the arrivals follow a Poisson process, the time spent in state i is exponentially distributed with parameter $\lambda + \mu_i$ and therefore has density function $(\lambda+\mu_i)e^{-(\lambda+\mu_i)t}$. Furthermore, with probability $\frac{\lambda}{\lambda+\mu_i}$ the next state is $i+1$ and with probability $\frac{\mu_i}{\lambda+\mu_i}$ the next state is $i-1$. Let $f_{i,i-1}(t)$ be the density function of $\sigma_{i,i-1}$, then we have

$$\begin{aligned} f_{i,i-1}(t) &= \frac{\mu_i}{\lambda+\mu_i}(\lambda+\mu_i)e^{-(\lambda+\mu_i)t} + \frac{\lambda}{\lambda+\mu_i}(\lambda+\mu_i)e^{-(\lambda+\mu_i)t} * f_{i+1,i}(t) * f_{i,i-1}(t) \\ &= \mu_i e^{-(\lambda+\mu_i)t} + \lambda e^{-(\lambda+\mu_i)t} * f_{i+1,i}(t) * f_{i,i-1}(t), \end{aligned} \tag{3.37}$$

where the operator $*$ denotes convolution, i.e. $f * g$ is the convolution of two functions f and g. We can now use a useful property of the Laplace transform of a convolution of (probability) functions, since this Laplace transform is given by the product of the separate Laplace transforms. Let $\mathcal{L}[f*g](s)$ denote the Laplace transform of the convolution $f * g$, then

$$\mathcal{L}[f * g](s) = \mathcal{L}[g](s)\mathcal{L}[g](s) = \hat{f}(s)\hat{g}(s). \tag{3.38}$$

Now by taking the (one-sided) Laplace transform on both sides of Equation 3.37 and using

$$\begin{aligned} \int_0^\infty e^{-st}e^{-(\lambda+\mu_i)t}\mathrm{d}t &= \int_0^\infty e^{-(\lambda+\mu_i+s)t}\mathrm{d}t \\ &= \left[-\frac{1}{\lambda+\mu_i+s}e^{-(\lambda+\mu_i+s)t}\right]_0^\infty \\ &= \frac{1}{\lambda+\mu_i+s}, \end{aligned} \tag{3.39}$$

we obtain

$$\hat{f}_{i,i-1}(s) = \frac{\mu_i}{\lambda+\mu_i+s} + \frac{\lambda}{\lambda+\mu_i+s}\hat{f}_{i+1,i}(s)\hat{f}_{i,i-1}(s). \tag{3.40}$$

Rewriting this expression leads to

$$\begin{aligned} \hat{f}_{i,i-1}(s) &= \frac{\mu_i}{\lambda+\mu_i+s-\lambda\hat{f}_{i+1,i}(s)} \\ &= \cfrac{\mu_i}{\lambda+\mu_i+s-\cfrac{\lambda\mu_{i+1}}{\lambda+\mu_{i+1}+s-\cfrac{\lambda\mu_{i+2}}{\lambda+\mu_{i+2}+s-\cdots}}} \\ &= -\frac{1}{\lambda}\mathrm{K}_{k=i}^{\infty}\frac{-\lambda\mu_k}{\lambda+\mu_k+s} \end{aligned} \tag{3.41}$$

□

We combine the result of Proposition 3.1 and Equation 3.35 to find the expression for the Laplace transform of the density function of the first-passage time σ_b

$$\begin{aligned} \hat{f}_b(s) &= \prod_{i=1}^{b}\hat{f}_{i,i-1}(s) \\ &= \prod_{i=1}^{b}\left(-\frac{1}{\lambda}\mathrm{K}_{k=i}^{\infty}\frac{-\lambda\mu_k}{\lambda+\mu_k+s}\right). \end{aligned} \tag{3.42}$$

4 | Order Book Probabilities

Cont et al. use the Laplace transforms of the density functions of the first-passage times of birth-death processes to compute several conditional probabilities of interest. In this chapter we will employ the same techniques and methodology to compute the probability of an increase in mid-price and the probability of executing an order that is placed at the best bid. Additionally, building on this methodology, we provide an expression for the fill probability of orders placed at one level below the best bid price. All these probabilities are *conditional* on the state of the order book at time $t = 0$, which is described by the number of outstanding limit orders at each price level and the spread S.

Before computing the probabilities, we first introduce some notation. Let $X_A = X_{p_A(\cdot)}(\cdot)$ and $X_B = |X_{p_B(\cdot)}(\cdot)|$ denote the quantities at the best ask and best bid, respectively. Furthermore, let $W_B = \{W_B(t), t \geq 0\}$ denote the number of orders remaining at the bid at time t of the initial $X_B(0)$ orders and ϵ_B the first-passage time of W_B to 0. We define W_A and ϵ_A similarly. We will use these quantities to compute the probabilities of interest in the following sections.

4.1 Probability of an Increase in Mid-Price

Let T be the time of the first change in mid-price, i.e.

$$T \equiv \inf\{t \geq 0 : p_M(t) \neq p_M(0)\},$$

where $p_M(t)$ denotes the mid-price at time t. In this section we consider the probability that the first mid-price change is an increase, but since the model is symmetric, the same result holds for the probability that the mid-price decreases at the next change. The probability that the next price move is an increase, *conditional* on the state of the order book, is given by

$$\mathbb{P}[p_M(T) > p_M(0) \mid X_A(0) = q_A,\ X_B(0) = q_B,\ p_S(0) = S], \tag{4.1}$$

with $S > 0$. To compute this probability, Cont et al. use the following lemma.

Lemma 4.1. Cont et. al [12, p. 555]: *Let $p_S(0) = S$, in this case:*

1. *there exist independent birth-death processes $\tilde{X}_A$ and $\tilde{X}_B$ with constant birth rates $\lambda(S)$ and death rates $\mu + i\theta(S)$, $i \geq 1$, such that for all $0 \leq t \leq T$, $\tilde{X}_A = X_A(t)$ and $\tilde{X}_B = X_B(t)$,*

2. *there exist independent pure death processes $\tilde{W}_A$ and $\tilde{W}_B$ with death rate $\mu + i\theta(S)$ in state $i \geq 1$, such that for all $0 \leq t \leq T$, $\tilde{W}_A = W_A(t)$ and $\tilde{W}_B = W_B(t)$. Furthermore, $\tilde{W}_A$ ($\tilde{W}_B$) is independent of $\tilde{X}_A$ ($\tilde{X}_B$) and $\tilde{W}_A \leq \tilde{X}_A$, $\tilde{W}_B \leq \tilde{X}_B$.*

Proof. The proof is provided in [12]. □

We can now compute the probability given by Equation 4.1 using the following proposition.

Proposition 4.1. Cont et. al [12, p. 555]: *Let σ_A and σ_B denote the first-passage times of $\tilde{X}_A$ and $\tilde{X}_B$ to 0, respectively. For $\hat{f}^S_{\sigma_j}(s)$, the Laplace transform of the pdf of σ_j, we have*

$$\hat{f}^S_{\sigma_j}(s) = \prod_{i=1}^{q_j} \left(-\frac{1}{\lambda(S)} \mathrm{K}_{k=i}^{\infty} \frac{-\lambda(S)\big(\mu + k\theta(S)\big)}{\lambda(S) + \mu + k\theta(S) + s} \right), \tag{4.2}$$

with $j \geq 1$. Let $\Lambda_S = \sum_{\delta=1}^{S-1} \lambda(\delta)$, then Probability 4.1 is given by the inverse Laplace transform of

$$\hat{F}^S_{\sigma_A,\sigma_B}(s) = \frac{1}{s}\Big(\hat{f}^S_{\sigma_A}(\Lambda_S+s)+\frac{\Lambda_S}{\Lambda_S+s}(1-\hat{f}^S_{\sigma_A}(\Lambda_S+s))\Big)\cdot\Big(\hat{f}^S_{\sigma_B}(\Lambda_S-s)+\frac{\Lambda_S}{\Lambda_S-s}(1-\hat{f}^S_{\sigma_B}(\Lambda_S-s))\Big), \quad (4.3)$$

evaluated at 0. From 4.3 follows that for $S = 1$ we have

$$\hat{F}^1_{\sigma_A,\sigma_B} = \frac{1}{s}\hat{f}^1_{\sigma_A}(s)\hat{f}^1_{\sigma_B}(-s). \quad (4.4)$$

Proof. Let σ_j be the first-passage time to 0 of a birth-death process with birth rate λ and death rate μ_i in state i, given that it started in state j. In Section 3.3 we showed that the Laplace transform $\hat{f}_{\sigma_j}(s)$ of the pdf of σ_j is given by

$$\hat{f}_{\sigma_j}(s) = \prod_{i=1}^{j}\Big(-\frac{1}{\lambda}\mathrm{K}_{k=i}^{\infty}\frac{-\lambda\mu_k}{\lambda+\mu_k+s}\Big). \quad (4.5)$$

By filling in $\lambda = \lambda(S)$ and $\mu_k = \mu + k\theta(S)$ we obtain the expression for $\hat{f}_{\sigma_j}(s)$ as in Equation 4.2.

Now recall the definition of the mid-price p_M

$$p_M(t) \equiv \frac{p_B(t)+p_A(t)}{2}. \quad (4.6)$$

From this definition follows that a price move occurs after a change in the best bid p_B or the best ask p_A, and the price *increases* if and only if $p_B(T) > p_B(0)$ or $p_A(T) > p_A(0)$. We have $p_B(T) > p_B(0)$ if a buy limit order is posted inside the spread, and $p_A(T) > p_A(0)$ if the number of orders at the best ask reaches 0. For the proof, two situations are considered:

1. For the case $S = 1$, a change in mid-price can only occur when either the queue at the best bid is depleted or the queue at the best ask is depleted, since no orders can be posted inside the spread. If we take $\tilde{X}_A$ and $\tilde{X}_B$ as described in Lemma 4.1, then the probability of an increase in price before a decrease is equal to the probability that $\hat{X}_A$ reaches 0 before $\hat{X}_B$. In this case, there are no orders left at the best ask, so by definition there is a new best ask at a price level that is higher than the previous one, i.e. $p_A(T) > p_A(0)$. Probability 4.1 is then given by

$$\mathbb{P}[\sigma_A < \sigma_B] = \mathbb{P}[\sigma_A - \sigma_B < 0]. \quad (4.7)$$

   Using Equation 3.2, by independence of the first-passage times σ_A and σ_B, the conditional Laplace transform of $\sigma_A - \sigma_B$ is equal to

$$\hat{f}_{\sigma_A-\sigma_B}(s) = \mathbb{E}[e^{-s(\sigma_A-\sigma_B)}] = \mathbb{E}[e^{-s\sigma_A}]\mathbb{E}[e^{s\sigma_B}] = \hat{f}_{\sigma_A}(s)\hat{f}_{\sigma_B}(-s). \quad (4.8)$$

   Combining this with the result of Lemma 4.2, we can show that the conditional Laplace transform of the cumulative distribution function $\hat{F}^1_{\sigma_A,\sigma_B}$ is given by Equation 4.4. If we let $\mathcal{L}^{-1}[\cdot]$ denote the inverse Laplace transform, then from

$$\mathbb{P}[\sigma_A - \sigma_B < 0] = F^1_{\sigma_A,\sigma_B}(0) = \mathcal{L}^{-1}\big[\hat{F}^1_{\sigma_A,\sigma_B}(s)\big](0) = \mathcal{L}^{-1}\big[\frac{1}{s}\hat{f}_{\sigma_A}(s)\hat{f}_{\sigma_B}(-s)\big](0), \quad (4.9)$$

   follows that we can compute Probability 4.1 for $S = 1$ by inverting the Laplace transform of $\hat{F}^1_{\sigma_A,\sigma_B}$ evaluated at 0.

2. For the case $S > 1$, the mid-price changes not only due to passage of the queues at the best quotes to 0, but also due to limit orders posted inside the spread. Let τ^δ_A (τ^δ_B) denote the first time an ask (bid) order arrives δ ticks away from the bid (ask), for $\delta = 1, ..., S-1$. In this case, τ^δ_A and τ^δ_B are mutually independent and have exponential distribution with rate $\lambda(\delta)$. Furthermore, τ^δ_A and τ^δ_B are independent of $\tilde{X}_A$ and $\tilde{X}_B$. For the time of the first change in mid-price we have

$$T = \sigma_A \wedge \sigma_B \wedge \min\{\tau^\delta_A, \tau^\delta_B, \quad \delta = 1, ..., S-1\}.$$

The first price change is an *increase* if either a buy limit order is posted inside the spread, since then we have a new best bid at a higher price level, or if $\tilde{X}_A$ reaches 0 *before* a sell limit order is posted inside the spread or $\tilde{X}_B$ reaches 0. Let τ_A and τ_B denote two exponentially distributed random variables with rate $\Lambda_S = \sum_{\delta=1}^{S-1} \lambda(\delta)$. Then Probability 4.1 is given by

$$\mathbb{P}[\sigma_A \wedge \tau_B < \sigma_B \wedge \tau_A] = \mathbb{P}[\sigma_A \wedge \tau_B - \sigma_B \wedge \tau_A < 0]. \tag{4.10}$$

The conditional Laplace transform of $\sigma_A \wedge \tau_B$ (and similarly $\sigma_B \wedge \tau_A$) is given by Lemma 4.3. From this follows that the Laplace transform of Equation 4.10 is given by Equation 4.3. Probability 4.1 can be computed by inverting this Laplace transform and evaluating it at 0.

□

Lemma 4.2. *Let f and F be the pdf and cdf of a random variable X, respectively. The Laplace transform $\hat{F}$ of the cdf F is given by*

$$\hat{F}(s) = \frac{1}{s}\hat{f}(s), \tag{4.11}$$

where $\hat{f}(s)$ is the Laplace transform of the pdf f.

Proof. See Appendix A □

Lemma 4.3. Cont et. al [12, p. 556]: *Let Z be an exponentially distributed random variable with parameter Λ, then the Laplace transform of $\sigma_B \wedge Z$ is given by*

$$\hat{f}^1_{\sigma_B}(\Lambda + s) + \frac{\Lambda}{\Lambda + s}(1 - \hat{f}^1_{\sigma_B}(\Lambda + s)),$$

where $\hat{f}^1_b$ is given by Equation 4.2.

Proof. See Appendix A □

4.2 Probability of Execution at the Best Bid Price

We now move to the probability of executing an order placed at the best bid before the mid-price moves, given that it is never cancelled. Let NC_b be the event that an order that never gets cancelled is placed at the best bid at time $t = 0$. The conditional probability that an order placed at the best bid is executed before the mid-price moves is given by

$$\mathbb{P}[\epsilon_B < T \mid X_A(0) = q_A,\ X_B(0) = q_B,\ p_S(0) = S, NC_b], \tag{4.12}$$

with ϵ_B the first-passage time of a pure-death process to 0, given that it started in state q_B. We now analyse a pure death process in stead of a birth-death process, since the position in the queue of the order that is placed at the best bid follows a pure death process by the time priority rule in the order book. That is, orders arriving at the same or a lower price level will have lower priority of execution, and therefore have no influence on the filling probability. We use the following proposition to compute the probability of execution.

Proposition 4.2. Cont et. al [12, p. 556]: *Define $\hat{f}^S_{\sigma_A}(s)$ as in Equation 4.2 and let $\hat{g}^S_{\epsilon_B}$ be the Laplace transform of the pdf of ϵ_B given by*

$$\hat{g}^S_{\epsilon_B}(s) = \prod_{i=1}^{q_B} \frac{\mu + (i-1)\theta(S)}{\mu + (i-1)\theta(S) + s}, \tag{4.13}$$

for $q_B \geq 0$ and let $\Lambda_S = \sum_{\delta=1}^{S-1} \lambda(\delta)$. Then Probability 4.12 is given by the inverse Laplace transform of

$$\hat{F}^S_{\epsilon_B,\sigma_A}(s) = \frac{1}{s}\hat{g}^S_{\epsilon_B}(s)\left(\hat{f}^S_{\sigma_A}(2\Lambda_S - s) + \frac{2\Lambda_S}{2\Lambda_S - s}(1 - \hat{f}^S_{\sigma_A}(2\Lambda_S - s))\right), \tag{4.14}$$

evaluated at 0. When $S = 1$, we obtain:

$$\hat{F}^1_{\epsilon_B,\sigma_A}(s) = \frac{1}{s}\hat{g}^1_{\epsilon_B}(s)\hat{f}^1_{\sigma_A}(-s). \tag{4.15}$$

Proof. The proof is similar to the proof of Proposition 4.1. First, to obtain the value for the Laplace transform of the pure-death process $g_{\epsilon_B}(s)$, we consider the first-passage time from state i to state $i-1$ for $1 \leq i \leq q_B$. If $g_{i,i-1}(t)$ is the density function of the first-passage time of the pure-death process which follows an exponential distribution with parameter μ_i, we have

$$g_{i,i-1}(t) = \mu_i e^{-\mu_i t}. \tag{4.16}$$

The Laplace transform $\hat{g}_{i,i-1}(s)$ is then given by

$$\begin{aligned} \hat{g}_{i,i-1}(s) &= \int_0^\infty \mu_i e^{-\mu_i t} e^{-st} \mathrm{d}t \\ &= \int_0^\infty \mu_i e^{-(\mu_i+s)t} \mathrm{d}t \\ &= \frac{\mu_i}{\mu_i + s}. \end{aligned} \tag{4.17}$$

Since the order that is placed at the best bid never gets cancelled, the cancellation rate in state i is given by $(i-1)\theta(S)$ and therefore $\mu_i = \mu + (i-1)\theta(S)$. Now, by independence of the arrival times, we obtain

$$\hat{g}^S_{\epsilon_B}(s) = \prod_{i=1}^{q_B} \frac{\mu + (i-1)\theta(S)}{\mu + (i-1)\theta(S) + s}. \tag{4.18}$$

For the case $S = 1$, the probability of executing a bid order before the mid-price moves comes down to

$$\mathbb{P}[\epsilon_B < \sigma_A] = \mathbb{P}[\epsilon_B - \sigma_A < 0], \tag{4.19}$$

i.e. the probability that the first-passage time of the pure-death process ϵ_B to 0 is smaller than the first-passage time of the birth-death process σ_A to 0. Again, from the fact that

$$\mathbb{P}[\epsilon_B - \sigma_A < 0] = F^1_{\epsilon_B,\sigma_A}(0), \tag{4.20}$$

we can compute Probability 4.12 by evaluating the inverse of the Laplace transform $\hat{F}^1_{\epsilon_B,\sigma_A}(s)$ at 0. For the case $S > 1$, the desired quantity is given by

$$\mathbb{P}[\epsilon_B < \sigma_A \wedge \tau_B \wedge \tau_A], \tag{4.21}$$

where the conditional distribution $\tau_B \wedge \tau_A$ is exponential with parameter $2\Lambda_S$. This is a result of the possibility that the mid-price can also change due to limit orders being posted inside the spread. Here, τ_A and τ_B denote the first time an ask or bid limit order is posted inside the spread, respectively. The Laplace transform of $\sigma_A \wedge \tau_B \wedge \tau_A$ is given by Lemma 4.3 for $Z = \tau_B \wedge \tau_A$, an exponentially distributed random variable with parameter $2\Lambda_S$. Probability 4.12 can in this case be computed by inverting the Laplace transform $\hat{F}^S_{\epsilon_B,\sigma_A}(s)$ given by Equation 4.14 and evaluating it at 0. □

4.3 Probability of Execution at One Price Level below the Best Bid Price

The next step is to compute the fill probabilities of orders posted deeper in the order book, i.e. at one price level below the best bid for bid orders and above the best ask for sell orders. In this section we consider the case of a limit buy order posted at the price level below the best bid: $p_B - 1$. By symmetry of the model, the reasoning is the same for limit sell orders. In reality there are two possibilities for orders posted at deeper levels to be executed. The first possibility is that a large market order arrives which executes multiple limit orders, possibly outstanding at multiple price levels. The second possibility is that the limit order lies at the best bid after a certain time because the bid price has moved down. In this case, we can compute the conditional fill probability as described in Section 4.2. Since all orders are assumed to have unit size, the first scenario of a large market order coming in and executing multiple limit orders is not possible. As we will show in Section 5.1.3, the empirical data will also support this assumption, since more than 90% of the executed orders lie at the best quote at the time of execution. For this reason, for an order to be executed when it is not submitted at the best bid price, the bid price needs to move down. After the price has moved down, we can calculate the probability that the order is executed the same way as before, since now the order lies at the best bid. Before we look at the probability of executing an order at price level $p_B - 1$, we first (re)introduce some notation. Let $X_B(t)$ and $X_A(t)$ again denote the quantities at the best bid and the best ask at time t, respectively. Let $X_{B-}(t)$ denote the quantity at the price level one below the best bid at time t and $W_{B-}(t)$ denotes the number of orders remaining at price level $p_B - 1$ at time t of the initial $X_{B-}(0)$ orders. Furthermore, we define T_{bid} as the first change in mid-price as a result of the best bid moving down,

$$T_{\text{bid}} \equiv \inf\{t \geq 0 : p_B(t) < p_B(0)\}, \tag{4.22}$$

and let T_{other} be the first change in mid-price as a result of a different event, i.e. the bid price moving up or the ask price moving in either direction,

$$T_{\text{other}} \equiv \inf\{t \geq 0 : p_B(t) > p_B(0) \wedge p_A(t) \neq p_A(0)\}. \tag{4.23}$$

The probability that bid price moves down before the mid-price moves due to a different event, is then given by

$$\mathbb{P}[T_{\text{bid}} < T_{\text{other}} \mid X_B(0) = q_B,\ X_A(0) = q_A,\ p_S(0) = S]. \tag{4.24}$$

For ease of notation, we will omit the condition in Probability 4.24, which denotes the state of the order book at time $t = 0$. We have

$$\mathbb{P}[T_{\text{bid}} < T_{\text{other}} \mid X_B(0) = q_B,\ X_A(0) = q_A,\ p_S(0) = S] = \mathbb{P}[T_{\text{bid}} < T_{\text{other}}]. \tag{4.25}$$

Now from Section 4.2 we know how to compute the probability of executing an order placed at the best bid before the mid-price moves, which is given by Equation 4.12. For an order placed at price level $p_B - 1$, this order lies at the new best bid after the previous best bid moved down. For that reason, we can compute the filling probability after the best bid moved down using

$$\mathbb{P}[\epsilon_{B-} < T \mid W_{B-}(T_{\text{bid}}) = q_{B-},\ X_A(T_{\text{bid}}) = q_A,\ p_S(T_{\text{bid}}) = S + 1], \tag{4.26}$$

with ϵ_{B-} the first-passage time of a pure death process (in this case the pure death process at price level $p_B - 1$, T the first time the mid-price moves, $W_{B-}(T_{\text{bid}})$ the remaining number of orders outstanding at the new best bid at $t = T_{\text{bid}}$ of the original number of orders $X_{B-}(0)$, $X_A(T_{\text{bid}})$ the number of orders outstanding at the best ask at $t = T_{\text{bid}}$ and $p_S(T_{\text{bid}})$ the spread size at $t = T_{\text{bid}}$. Since we assume that only the best bid price moves down one price level, we know that the spread size increases by one tick. Due to the dynamics of the order book, both W_{B-} and X_A are unknown at time T_{bid}. By the law of total expectation, we can compute the filling probability by summing over all possible values and combinations for the quantities W_{B-} and X_A at time T_{bid}. The probability of a combination of i orders outstanding at the best bid and j orders outstanding at the best ask after the price moved down is given by

$$\mathbb{P}[W_{B-}(T_{\text{bid}}) = i, X_A(T_{\text{bid}}) = j \mid X_B(0) = q_B,\ X_{B-}(0) = q_{B-},\ X_A(0) = q_A,\ p_S(0) = S], \tag{4.27}$$

which, by independence of the processes W_{B-} and X_A and independence of X_B and X_A can be written as

$$\begin{aligned} \mathbb{P}[W_{B-}(T_{\text{bid}}) = i,\ X_A(T_{\text{bid}}) = j \mid X_B(0) = q_B,\ X_{B-}(0) = q_{B-},\ X_A(0) = q_A,\ p_S(0) = S] = \\ \mathbb{P}[W_{B-}(T_{\text{bid}}) = i \mid X_B(0) = q_B,\ X_{B-}(0) = q_{B-},\ p_S(0) = S] \\ \cdot\mathbb{P}[X_A(T_{\text{bid}}) = j \mid X_A(0) = q_A,\ p_S(0) = S]. \end{aligned} \tag{4.28}$$

Again, to ease the notation, we omit the conditions in Probabilities 4.26 and 4.28 that are known at $t = 0$, which are $X_B(0)$, $X_{B-}(0)$, $X_A(0)$, $p_S(0)$ and $p_S(T_{\text{bid}})$. Note that $p_S(T_{\text{bid}})$ is known at $t = 0$, since the spread increases by one tick if the best bid moves down, so we have $p_S(T_{\text{bid}}) = p_S(0) + 1$. We write Probability 4.26 as

$$\mathbb{P}[\epsilon_{B-} < T \mid W_{B-}(T_{\text{bid}}) = q_{B-},\ X_A(T_{\text{bid}}) = q_A], \tag{4.29}$$

and Probability 4.28 as

$$\mathbb{P}[W_{B-}(T_{\text{bid}}) = i] \cdot \mathbb{P}[X_A(T_{\text{bid}}) = j]. \tag{4.30}$$

Now summing over all possible combinations for the quantities W_{B-} and X_A at time T_{bid} gives us

$$\sum_{j=1}^{A}\sum_{i=1}^{B}\Big(\mathbb{P}[\epsilon_{B-} < T \mid W_{B-}(T_{\text{bid}}) = i,\ X_A(T_{\text{bid}}) = j] \cdot \mathbb{P}[W_{B-}(T_{\text{bid}}) = i] \cdot \mathbb{P}[X_A(T_{\text{bid}}) = j]\Big), \tag{4.31}$$

with A the possible number of orders outstanding at the ask at time T_{bid} and B the number of possible orders in front of and including the submitted bid order at T_{bid}.

Since a bid order was submitted at price level $p_B - 1$ at $t = 0$, we know that W_{B-} follows a pure-death process and therefore has a finite number of possible states i at time T_{bid}. That is, the queue in front of the submitted bid order has only moved down due to cancellations, since limit orders submitted at a later time will have lower priority and therefore will arrive at a place in the queue behind our initial order. For that reason, they do not influence the fill probability. If the number of orders at price level $p_B - 1$ at $t = 0$ was, including our own order, equal to q_{B-}, then the number of orders in front of and including our order at time T_{bid} will be between 1 and q_{B-}. We can calculate these probabilities using Laplace transforms of the first-passage times. Let σ_B be the first-passage time of the best bid to 0. Suppose the quantity at price level $p_B - 1$ is equal to q at $t = 0$, then the probability that the remaining quantity is equal to i at time T_{bid} is given by

$$\mathbb{P}[W_{B-}(T_{\text{bid}}) = i] = \begin{cases} \mathbb{P}[\sigma_B < \epsilon_{q,q-1}], & \text{for } i = q, \\ \mathbb{P}[\epsilon_{q,1} < \sigma_B], & \text{for } i = 1, \\ \mathbb{P}[\epsilon_{q,i} < \sigma_B] - \mathbb{P}[\epsilon_{q,i-1} < \sigma_B], & \text{for } 1 < i < q, \end{cases} \tag{4.32}$$

where $\epsilon_{q,i}$ is the first-passage time of a pure-death process from state q to state i. For simplicity we omit the condition on $X_{B-}(0)$, $X_B(0)$ and $p_S(0)$ in the notation. In other words, it is the probability that the queue at the best bid is depleted before the queue at price level $p_B - 1$ has hit state $i - 1$. This is justified since it is a pure-death process and thus can only move down, and the first-passage time will also be the last-passage time.

The process X_A, on the other hand, follows a birth-death process and therefore could have - in theory - infinitely many values. However, as one might suspect, in reality there will also be only a finite number of possibilities for the quantity at the best ask after the best bid price has moved down. As will become clear in Section 6.5, we can restrict ourselves to a finite number of possibilities. We will fit a (empirical) distribution to the number of orders outstanding at the best ask after a downward move of the bid price to compute $\mathbb{P}[X_A(T_{\text{bid}}) = j]$, following the approach of Cont and De Larrard in [11]. Combining Equations 4.24 and 4.31, we obtain for the probability that an order placed at price level p_{B-1} is filled

$$\begin{aligned}&\mathbb{P}[T_{\text{bid}} < T_{\text{other}}]\cdot \\ \sum_{j=1}^{A}\sum_{i=1}^{B}\Big(&\mathbb{P}[\epsilon_{B-} < T \mid W_{B-}(T_{\text{bid}}) = i,\ X_A(T_{\text{bid}}) = j] \cdot \mathbb{P}[W_{B-}(T_{\text{bid}}) = i] \cdot \mathbb{P}[X_A(T_{\text{bid}}) = j]\Big).\end{aligned} \tag{4.33}$$

We now prove the following propositions to compute the probability given in Equation 4.33.

Proposition 4.3. *Let $\hat{f}^S_{\sigma_j}(s)$ denote the Laplace transform of the density function of first-passage time to 0 of a birth-death process σ_j, given by*

$$\hat{f}^S_{\sigma_j}(s) = \prod_{i=1}^{q_j}\bigg(-\frac{1}{\lambda(S)}\mathop{\mathrm{K}}_{k=i}^{\infty}\frac{-\lambda(S)\big(\mu + k\theta(S)\big)}{\lambda(S) + \mu + k\theta(S) + s}\bigg), \tag{4.34}$$

with $j \geq 0$. Let $\Lambda_S = \sum_{i=1}^{S-1} \lambda(i)$, then Probability 4.24 is given by the inverse Laplace transform of

$$\hat{F}^S_{\sigma_B,\sigma_A}(s) = \frac{1}{s}\hat{f}^S_{\sigma_B}(s)\bigg(\hat{f}^S_{\sigma_A}(2\Lambda_S - s) + \frac{2\Lambda_S}{2\Lambda_S - s}(1 - \hat{f}^S_{\sigma_A}(2\Lambda_S - s))\bigg), \tag{4.35}$$

evaluated at 0. When $S = 1$, we obtain:

$$\hat{F}^1_{\sigma_B,\sigma_A}(s) = \frac{1}{s}\hat{f}^1_{\sigma_B}(s)\hat{f}^1_{\sigma_A}(-s). \tag{4.36}$$

Proof. For the case $S = 1$, the probability that the mid-price moves down due to the best bid moving down before it moves as a result of a different event is given by the probability that the queue at the best bid reaches 0 earlier than the queue at the best ask. Let σ_A denote the first-passage time to 0 of the birth-death process at the ask price, and let σ_B denote the first-passage time to 0 of the birth-death process at the bid price. The probability that the best bid queue reaches 0 before the ask queue is then given by

$$\mathbb{P}[\sigma_B < \sigma_A] = \mathbb{P}[\sigma_B - \sigma_A < 0], \tag{4.37}$$

which by independence of σ_A and σ_B and Equation 3.2 can be computed using the inverse Laplace transform of

$$\hat{F}^1_{\sigma_B,\sigma_A}(s) = \frac{1}{s}\hat{f}^1_{\sigma_B}(s)\hat{f}^1_{\sigma_A}(-s), \tag{4.38}$$

evaluated at 0.
For the case $S > 1$, the mid-price can also move when limit orders are posted within the spread. Probability 4.24 is then given by

$$\mathbb{P}[\sigma_B < \sigma_A \wedge \tau_B \wedge \tau_A], \tag{4.39}$$

where τ_B and τ_A are exponentially distributed random variables with rate Λ_S. These random variables denote the first time either a buy or sell limit order is posted inside the spread. Using Lemma 4.3, the probability can be computed by inverting the Laplace transform

$$\hat{F}^S_{\sigma_B,\sigma_A}(s) = \frac{1}{s}\hat{f}^S_{\sigma_B}(s)\bigg(\hat{f}^S_{\sigma_A}(2\Lambda_S - s) + \frac{2\Lambda_S}{2\Lambda_S - s}(1 - \hat{f}^S_{\sigma_A}(2\Lambda_S - s))\bigg), \tag{4.40}$$

and evaluating it at 0. □

Proposition 4.4. *Define $\hat{f}^S_{\sigma_A}(s)$ as in 4.34 and let $\hat{g}^S_{\epsilon_{B-}}$ be the Laplace transform of the density function of the first-passage time to 0 of a pure-death process ϵ_{B-}, given by*

$$\hat{g}^S_{\epsilon_{B-}}(s) = \prod_{i=1}^{q_{B-}} \frac{\mu + (i-1)\theta(S)}{\mu + (i-1)\theta(S) + s}, \tag{4.41}$$

for $j \geq 0$ and let $\Lambda_S = \sum_{\delta=1}^{S-1} \lambda(\delta)$. Then Probability 4.26 is given by the inverse Laplace transform of

$$\hat{F}^S_{\epsilon_{B-},\sigma_A}(s) = \frac{1}{s}\hat{g}^S_{\epsilon_{B-}}(s)\bigg(\hat{f}^S_{\sigma_A}(2\Lambda_S - s) + \frac{2\Lambda_S}{2\Lambda_S - s}(1 - \hat{f}^S_{\sigma_A}(2\Lambda_S - s))\bigg), \tag{4.42}$$

evaluated at 0.

Proof. The proof is similar to the proof of Proposition 4.1. □

Proposition 4.5. *Define $\hat{f}^S_{\sigma_j}(s)$ as in Equation 4.34 and let $\hat{h}^S_{q,i}(s)$ be the Laplace transform of the density function of the first-passage time of a pure-death process to state i, given that it started in state q, given by*

$$\hat{h}^S_{q,i}(s) = \prod_{j=i}^{q} \frac{(j-1)\theta(S+1)}{(j-1)\theta(S+1) + s}, \tag{4.43}$$

for $q \geq 1$. Then the probabilities in Equation 4.26 are given by the inverse Laplace transforms of

$$\hat{F}^S_{\sigma_B,q,i}(s) = \begin{cases} \frac{1}{s}\hat{f}^S_{\sigma_B}(s)\hat{h}^S_{q,q-1}(-s), & \text{for } i = q, \\ \frac{1}{s}\hat{f}^S_{\sigma_B}(-s)\hat{h}^S_{q,1}(s), & \text{for } i = 1, \\ \frac{1}{s}\hat{f}^S_{\sigma_B}(-s)\hat{h}^S_{q,i}(s) - \hat{f}^S_b(-s)\hat{h}^S_{q,i-1}(s), & \text{for } 1 < i < q, \end{cases} \tag{4.44}$$

evaluated at 0.

Proof. We know that the Laplace transform of the density function of the first-passage time from state i to state $i-1$ of a pure-death process with death-rate μ_i in state i is given by

$$\hat{h}_{i,i-1}(s) = \frac{\mu_i}{\mu_i + s}, \tag{4.45}$$

see also the computation in Equation 4.17. At the price level below the best bid, the death-rate is determined by the cancellations only, since market orders only arrive at the best quotes. The death-rate μ_i in state i is therefore given by

$$\mu_i = (i-1)\theta(S+1), \tag{4.46}$$

which gives us the following Laplace transform

$$\hat{h}_{i,i-1}(s) = \frac{(i-1)\theta(S+1)}{(i-1)\theta(S+1)+s}. \tag{4.47}$$

By independence of the arrival times, the first-passage time to a state i of a pure-death process, given that it started in state q is given by

$$\hat{h}^S_{q,i}(s) = \prod_{j=i}^{q} \frac{(j-1)\theta(S+1)}{(j-1)\theta(S+1)+s}. \tag{4.48}$$

For the proof we consider three cases for the probabilities that the quantity at price level $p_B - 1$ after the bid price moved down, given that the quantity at $t=0$ was equal to q.

1. If the first-passage time to 0 of the birth-death process σ_B at the best-bid is smaller than the first-passage time to state $q-1$ of the pure-death process $\epsilon_{q,q-1}$, then the bid price moved down before the queue at level $p_B - 1$ moved to state $q-1$. This would mean that the pure-death process would still be in state q, meaning that there would still be q orders remaining. The probability

$$\mathbb{P}[W_{B-}(T_{\text{bid}}) = q] = \mathbb{P}[\sigma_B < \epsilon_{q,q-1}] = \mathbb{P}[\sigma_B - \epsilon_{q,q-1} < 0], \tag{4.49}$$

can therefore be computed by the inverse Laplace transform

$$\hat{F}^S_{\sigma_B,q,q}(s) = \frac{1}{s}\hat{f}^S_{\sigma_B}(s)\hat{h}^S_{q,q-1}(-s), \tag{4.50}$$

evaluated at 0. This follows from Equation 3.2, by independence of σ_B and $\epsilon_{q,q-1}$.

2. Similarly, if the first-passage time of the pure-death process to 1, $\epsilon_{q,1}$, is smaller than the first-passage time of the birth-death process at the best bid to zero, σ_B, then there would be no orders in front of the posted order when the price moves down. The probability

$$\mathbb{P}[W_{B-}(T_{\text{bid}}) = 1] = \mathbb{P}[\epsilon_{q,0} < \sigma_B] = \mathbb{P}[\epsilon_{q,0} - \sigma_B < 0], \tag{4.51}$$

can be computed by the inverse Laplace transform

$$\hat{F}^S_{\sigma_B,q,1}(s) = \frac{1}{s}\hat{f}^S_{\sigma_B}(-s)\hat{h}^S_{q,1}(s), \tag{4.52}$$

evaluated at 0, again by independence of σ_B and $\epsilon_{q,0}$ and using Equation 3.2.

3. For the cases $1 < i < q-1$, we combine the reasoning from the previous two instances. Suppose the first-passage time of the pure-death process to state i, denoted by $\epsilon_{q,i}$ is smaller than the first-passage time of the birth-death process σ_B to 0, then the pure-death process has moved to *at least* state i. This probability is given by

$$\mathbb{P}[W_{B-}(T_{\text{bid}}) \leq i] = \mathbb{P}[\epsilon_{q,i} < \sigma_B] = \mathbb{P}[\epsilon_{q,i} - \sigma_B < 0]. \tag{4.53}$$

To compute the probability that the pure-death process is in state i exactly, we need to subtract the probability that the process is in state $i-1$ or lower, which is given by

$$\mathbb{P}[W_{B-}(T_{\text{bid}}) \leq i-1] = \mathbb{P}[\epsilon_{q,i-1} < \sigma_B] = \mathbb{P}[\epsilon_{q,i-1} - \sigma_B < 0]. \tag{4.54}$$

Combining these probabilities gives

$$\mathbb{P}[W_{B-}(T_{\text{bid}}) = i] = \mathbb{P}[\epsilon_{q,i} - \sigma_B < 0] - \mathbb{P}[\epsilon_{q,i-1} - \sigma_B < 0], \tag{4.55}$$

which is the probability that the process is in state i exactly. This can be computed by the inverse Laplace transform

$$\hat{F}^S_{\sigma_B,q,i}(s) = \frac{1}{s}\hat{f}^S_{\sigma_B}(-s)\hat{h}^S_{q,i}(s) - \frac{1}{s}\hat{f}^S_{\sigma_B}(-s)\hat{h}^S_{q,i-1}(s), \tag{4.56}$$

evaluated at 0.

□

Using Propositions 4.3, 4.4 and 4.5, we ware able to compute all components of Probability 4.33 to obtain the probability that an order posted at price level $p_B - 1$ is filled.

The intuition that we presented in this section to compute the probability of executing an order placed at one price level below the best bid can be extended to price levels even deeper in the order book. It should be noted that the complexity of the expression to compute the fill probability analytically would increase quite rapidly. For example, an expression for the fill probability at two price levels below the best bid price, $p_B - 2$, would need to incorporate all uncertainty of the number of orders outstanding at each of the price levels in between. That is, the bid price needs to move down two levels, and after each price move, the possible number of orders outstanding at the best ask p_A, the price level below the best bid $p_B - 1$ and the price level $p_B - 2$ needs to be incorporated. For this reason, it remains to be seen if computing the fill probability analytically for price levels deeper in the order book is feasible. In Section 5.1.3 we show that the majority of executed limit orders (±85%) was posted at a distance of at most one tick from the best quote. Therefore, we decide to focus only on those price levels.

5 | Verification of Model Assumptions and Model Calibration

This chapter provides a description of the available limit order book data, an analysis of this data and an estimation of the model parameters. The available data is collected from a trading venue called LMAX from November 2nd 2020 until October 29th 2021. This data contains all information of trading activity for several currency pairs, including EURUSD, EURGBP and GBPUSD. We focus on the EURUSD currency pair, since its daily turnover amounts to 22.7% of the total daily turnover on the FX market, the largest share of all currency pairs [6]. Furthermore, most trades executed by MN involve this currency pair.

The LMAX data is formatted according to the LMAX Exchange multicast market data service or Itch service. This is a format in which the data is collected in a binary format and it is commonly used by high-frequency traders, since the data can be stored and distributed efficiently. The data files belonging to LMAX contain both event and trade data, for all currency pairs. Every time an event takes place, i.e. a trade, cancellation or the arrival of a new order, a message is sent out in this binary format. A parser is needed to decode the messages to use it in the data analysis. We can use the parser to transform the data into a format that we can use to create an order book, i.e. for event messages a timestamp, order ID, event type, direction, price and quantity and for trade messages a timestamp, aggressive side[1], price and quantity. We create and simulate the order book using the incoming event messages, adding orders to it when a limit order arrives and deleting orders when trades or cancellations arrive. We are then also able to compute metrics of interest for each new event, including the spread, best bid price, best ask price, quantities at each price level etc. Since the simulation of the order book using the event data is quite an expensive task, we use data from four weeks in the data set from 7–6–2021 to 2–7–2021 to perform the analyses and to calibrate the parameters. For this period, the data was complete for all the days and furthermore, there were no special events or disturbances on the FX market during this period.

In Section 5.1 we conduct an analysis of the order book data of the FX spot market. This includes the time distribution of the spread size, the symmetry of the order flow and the characteristics of orders and cancellations. The model parameters $\lambda(\delta)$, μ and $\theta(\delta)$ are calibrated in Section 5.2.

5.1 Data Analysis on the Foreign Exchange Spot Market

5.1.1 Spread Distribution

First we examine the distribution of the size of the bid-ask spread. For stocks, the tick size typically equals one cent. In contrast, the tick size for currency pairs can be as small as 0.1 cents or in our case 0.001 cents. As a result, there are more possible price levels at which an order can be placed and typically for the EURUSD currency pair, the bid-ask spread is larger than one tick. Table 5.1 shows the distribution of the duration for which the spread is equal to a certain value, for four weeks in the data set from 7–6–2021 to 2–7–2021. We observe that for all weeks, the spread predominantly lies between one and five ticks and for approximately 80% of the time, the spread equals three or four ticks. As will become clear in the sections that follow, the spread size also influences the arrival rate

[1] Aggressive side refers to the side of the trader that submitted the market order.

of orders and cancellations, and thus also the order book probabilities. For that reason we will take this into account when computing the desired conditional probabilities.

Table 5.1: Distribution of the duration the spread is equal to S in ticks.

S	Percentage of duration
1	0.45%
2	5.38%
3	39.31%
4	47.84%
5	6.32%
>5	0.70%

(a) Distribution between 7–6–2021 and 11–6–2021.

S	Percentage of duration
1	0.13%
2	1.66%
3	17.07%
4	62.04%
5	18.54%
>5	0.56%

(b) Distribution between 14–6–2021 and 18–6–2021.

S	Percentage of duration
1	0.75%
2	2.72%
3	19.43%
4	53.96%
5	22.07%
>5	1.07%

(c) Distribution between 21–6–2021 and 25–6–2021.

S	Percentage of duration
1	0.48%
2	18.81%
3	36.17%
4	31.93%
5	7.60%
>5	0.70%

(d) Distribution between 28–6–2021 and 2–7–2021.

5.1.2 Order Flow Symmetry

In our model we assume that the order flow is symmetric, i.e. the rate of incoming sell orders is equal to the rate of incoming buy orders. To see that this is a reasonable assumption, we can take a look at the empirical rates of incoming orders and cancellations. Figure 5.1 shows the arrival rate per second of buy and sell market orders, as well as the total arrival rate of market orders per second. The rates are calculated using

$$\frac{N_m}{T_*}, \tag{5.1}$$

with N_m the number of arrivals of market orders during the time sample and T_* the total time in seconds within the time sample. As we can see, the arrival rates are similar for both the buy and sell side for each week.

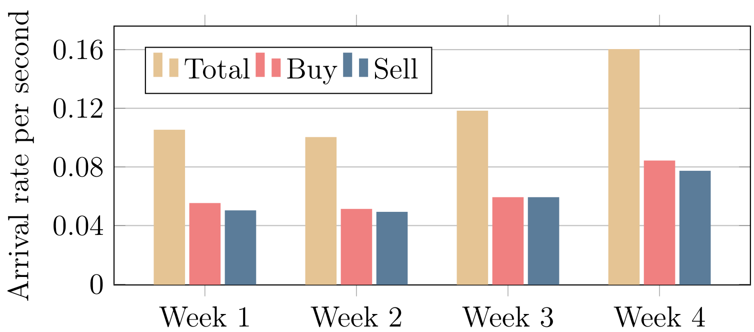
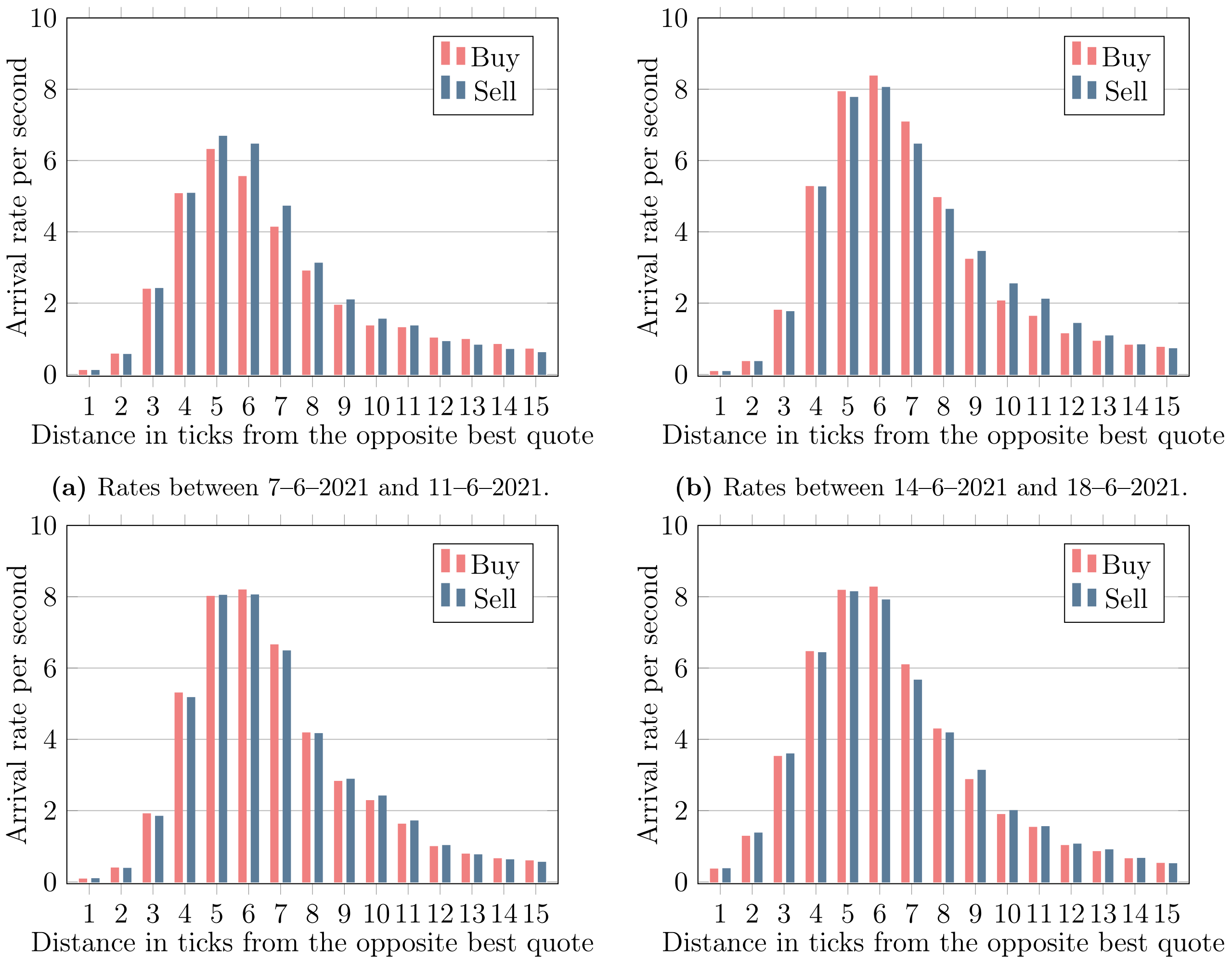


Figure 5.1: Arrival rate per second for sell and buy market orders for each week in the data set.

Unlike market orders, which only arrive at the best quotes, limit orders can be posted at any price level higher than the best bid for sell orders and lower than the best ask for buy orders. For this reason, we determine the arrival rates of both buy and sell orders at each distance δ in ticks from the opposite best quote. For example, for $\delta = 1$ we look at bid orders posted at one price level below the best ask and ask orders posted one price level above the best bid. The rates are calculated by

$$\frac{N_l(\delta)}{T_*}, \tag{5.2}$$

with $N_l(\delta)$ the number of arrivals of limit orders at a distance δ from the opposite best quote during the time sample and T_* the total time in seconds within the time sample. Figure 5.2 shows that for limit orders, the arrival rates for buy and sell orders are similar for each distance from the opposite best quote and for each week in the data set. We only look at the first 15 ticks, since most trading action takes place near the best quotes.



(a) Rates between 7–6–2021 and 11–6–2021.

(b) Rates between 14–6–2021 and 18–6–2021.

(c) Rates between 21–6–2021 and 24–6–2021.

(d) Rates between 28–6–2021 and 2–7–2021.

Figure 5.2: Arrival rates per second for sell and buy limit orders for each distance δ in ticks from the opposite best quote for $\delta = 1, ..., 15$.

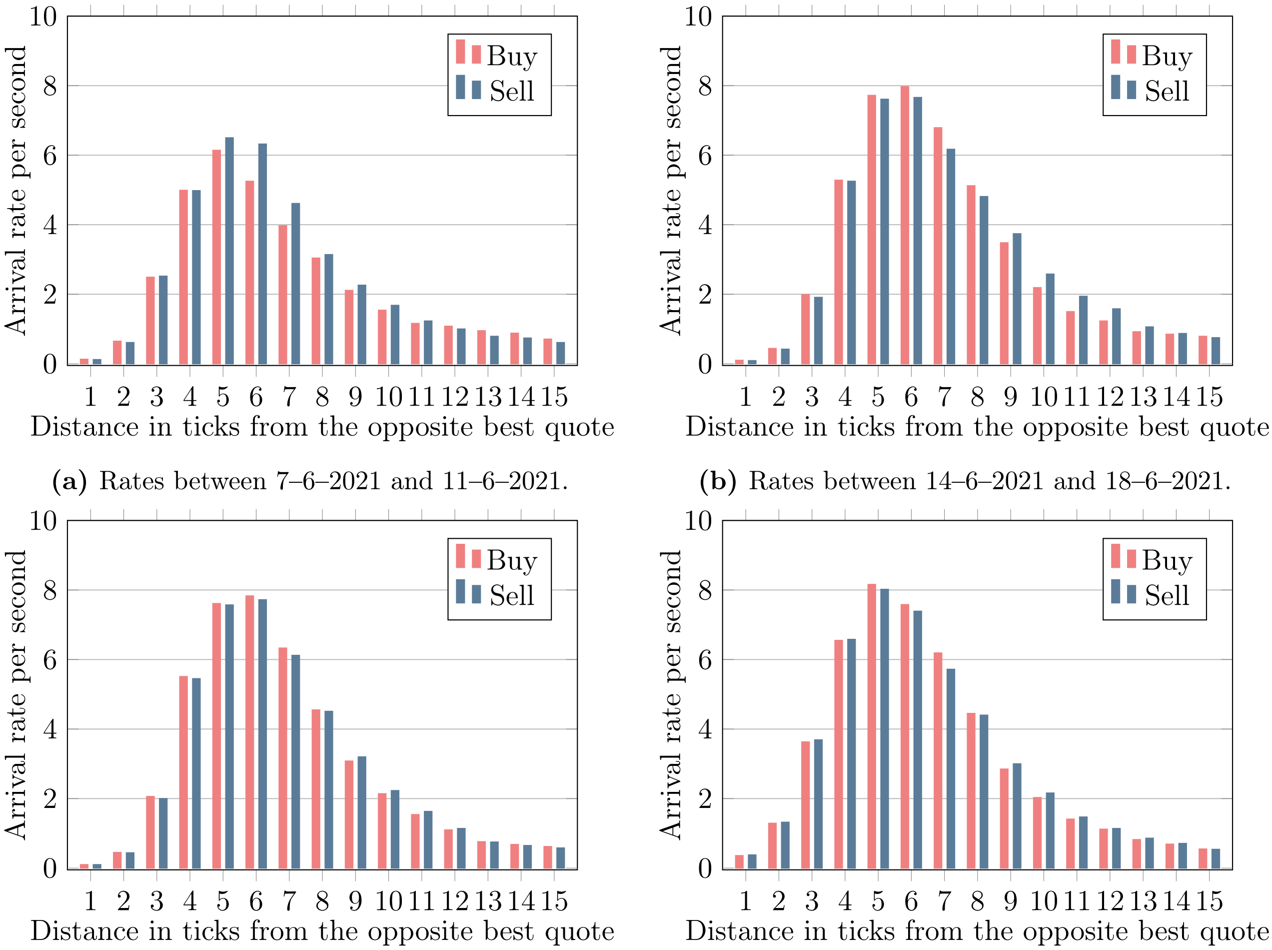
Finally, the arrival rates of cancellations on both buy and sell side of the order book are shown in Figure 5.3. Similar to the limit order arrival rate, the rates are determined by

$$\frac{N_c(\delta)}{T_*}, \tag{5.3}$$

with $N_c(\delta)$ the number of cancellations at a distance δ from the opposite best quote during the time sample and T_* the total time in seconds within the time sample. We notice that the cancellation rates for each distance from the opposite best quote are very similar to the arrival rates of limit orders as depicted in Figure 5.2. This can be explained by the fact that around 99.9% of all limit orders are cancelled, see also Table 5.2.

Table 5.2: Percentage of limit orders cancelled or (partially) filled.

	Week 1	Week 2	Week 3	Week 4
Cancelled	99.91%	99.93%	99.91%	99.87%
(Partially) filled	0.09%	0.07%	0.09%	0.13%



(a) Rates between 7–6–2021 and 11–6–2021.

(b) Rates between 14–6–2021 and 18–6–2021.

(c) Rates between 21–6–2021 and 24–6–2021.

(d) Rates between 28–6–2021 and 2–7–2021.

Figure 5.3: Arrival rates per second for cancellations on sell and buy side for each distance δ in ticks from the opposite best quote for $\delta = 1, \ldots, 15$.

In addition to the visual evidence for order flow symmetry, we can also perform statistical tests to see if the symmetry assumption is reasonable. Since the arrival rates are not normally distributed, we will perform a Wilcoxon signed rank test to test this assumption [26]. This test is used to determine whether the rate of one type of order is likely to be larger than the rate of the other type. Specifically, the method tests if this probability is higher than 50%, indicating that the rates are not symmetrical. The p-values of the Wilcoxon test can be found in Table 5.3. For the limit orders and cancellations we have taken the arrival rates of sell and limit orders at each distance from the opposite best quote for each week. For the market orders we used the difference in rates between buy and sell orders for each day in the week. For p-values larger than 0.05 we can assume that the data is symmetrical. As becomes clear from Table 5.3, this is the case for all orders for each week, except for the market orders in the first week.

Although the arrival rates for the buy and sell side are - naturally - not exactly equal, it is reasonable to assume symmetry from a modelling perspective as well. An asymmetric order flow could lead to a directional price trend, since the price would tend to move into the direction with higher cancellation and market order rates or lower limit order rates.

Table 5.3: Wilcoxon signed-rank test to test the assumption of order flow symmetry. The assumption of symmetry is satisfied for p-values larger than 0.05.

	Week 1	Week 2	Week 3	Week 4
Market orders	0.04	0.50	0.89	0.28
Limit orders	0.16	0.95	0.43	0.65
Cancellations	0.36	0.91	0.31	0.65

5.1.3 Empirical Limit Order Execution

In Section 4.3 we mentioned that due to the unit size assumption of the model, it is not possible to model the situation that multiple limit orders are executed against one large market order. A result of this is also that in the model, limit orders can only be executed when they have the highest priority and therefore by definition lie at the best bid or best ask price. Figure 5.4 shows the distribution of the distance from the best quote for limit orders at the time of execution. For buy orders this is the distance from the best bid price and, conversely, for sell orders it is the distance from the best ask price. A distance of 0 indicates that the order was executed at the best quote, while a distance of more than 0 indicates that a large market order executed limit orders over multiple price levels. We observe that for all weeks in the data set, more than 90% of all executions occurred at the best quotes. This observation also shows that it is reasonable from a modelling perspective to assume that most orders are executed at the best quotes only.

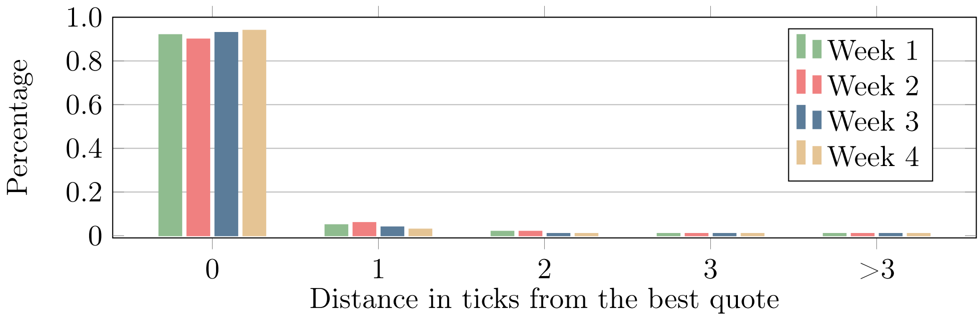


Figure 5.4: Distribution of executed limit orders based on their distance in ticks from the best quote at the time of execution. A distance of 0 indicates the order was executed at the best quote.

Now we examine the distribution of the distance from the best quote at the time of submission. Figure 5.5 shows the percentage of executed limit orders that was submitted at a certain distance from the best quote. A negative distance indicates that the order was posted inside the spread, which in practice would result in a new best quote. We see that around 85% of the executed limit orders was submitted at a distance of at most one tick from the opposite best quote.

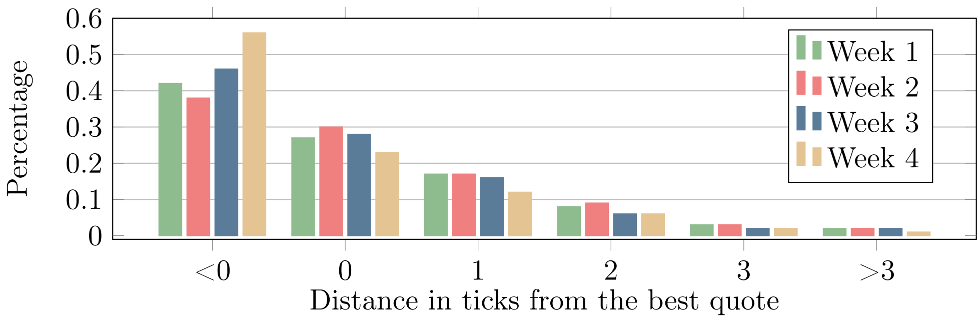


Figure 5.5: Distribution of executed limit orders based on their distance in ticks from the best quote at the time of submission. A distance of 0 indicates the order was submitted at the best quote, while a distance of <0 indicates that the order was posted inside the spread, resulting in a new best quote.

5.1.4 Limit Orders

For the estimation of the parameters used in the model, we begin by examining the limit orders. The model assumes that all orders have unit size, equal to the average limit order size. In Figure 5.6, the distribution of limit order sizes is shown for each week. The x-axis denotes the order size in millions and the y-axis shows the percentage of limit orders of a particular size during each week in the data set. We observe that the distribution of is very similar for each week in the data set. Around 85% of the limit orders are of size 1 million or 500 thousand. Other regularly occurring order sizes are 2 million, 4 million and 525 thousand. Other sizes occur less than 1% of the time. The average limit order size used in the model to determine the unit size is approximately 850 thousand for each week. Although Figure 5.6 shows us that the assumption that orders have unit size is not a very realistic one, it is a necessary assumption in order to have an analytical expression for the fill probability.

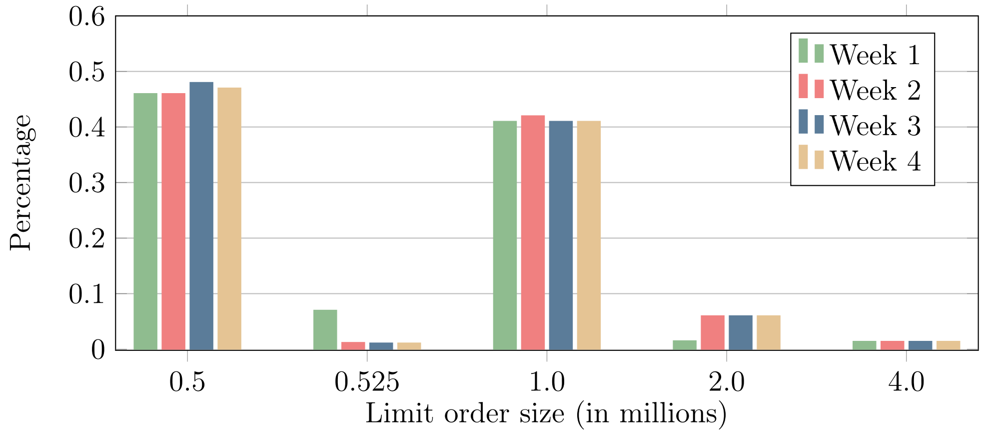
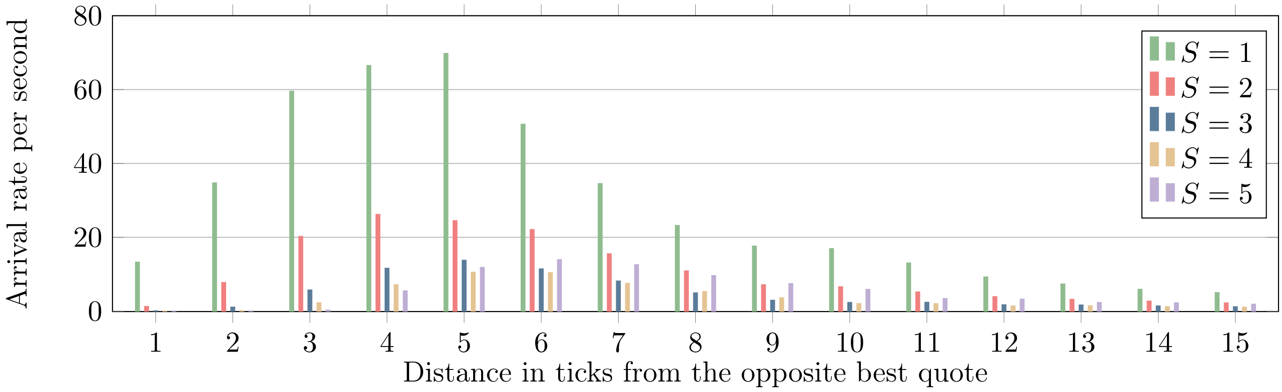


Figure 5.6: Distribution of limit order sizes, for sizes occurring more than 1% of the time.

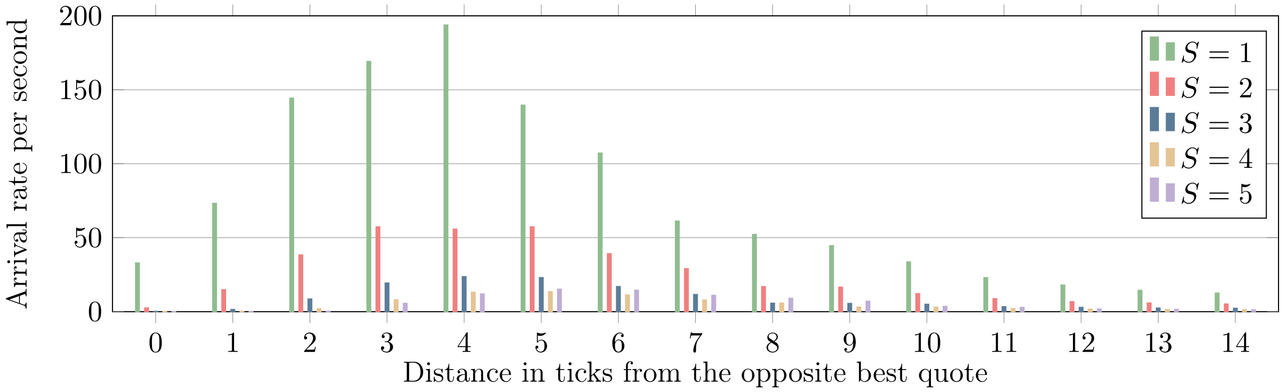
As mentioned in Section 5.1.1, the size of the spread varies quite substantially. In Figure 5.7, the arrival rates of limit orders are shown for each of the most common spread sizes ranging from one to five ticks. The arrival rates are computed using

$$\frac{N_l^S(\delta)}{T_*^S}, \tag{5.4}$$

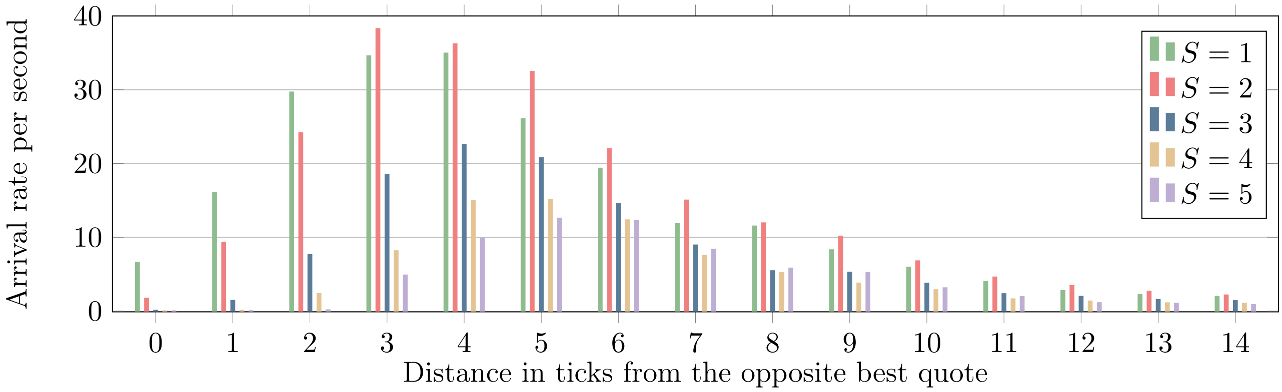
with $N_l^S(\delta)$ the number of arrivals of limit orders at a distance δ in ticks from the opposite best quote during the time sample when the spread was equal to S and T_*^S denotes the total time in seconds the spread was equal to S. We observe a significant difference in the arrival rates w.r.t. the different spread sizes. The arrival rate of limit orders is considerably higher for smaller spread sizes, especially in the first two weeks. Another thing that stands out is that if limit orders are posted within the spread, the price level at which they are posted is in most cases at most one tick better than the current best bid or ask. For example, when the spread equals two ticks, limit orders are posted at a distance of one tick or more from the opposite best quote. Similarly, when the spread equals three ticks, the orders are posted at a distance of two ticks or more from the opposite best quote. The exact arrival rates of limit orders for each week in the data set can be found in Table B.1 in Appendix B.



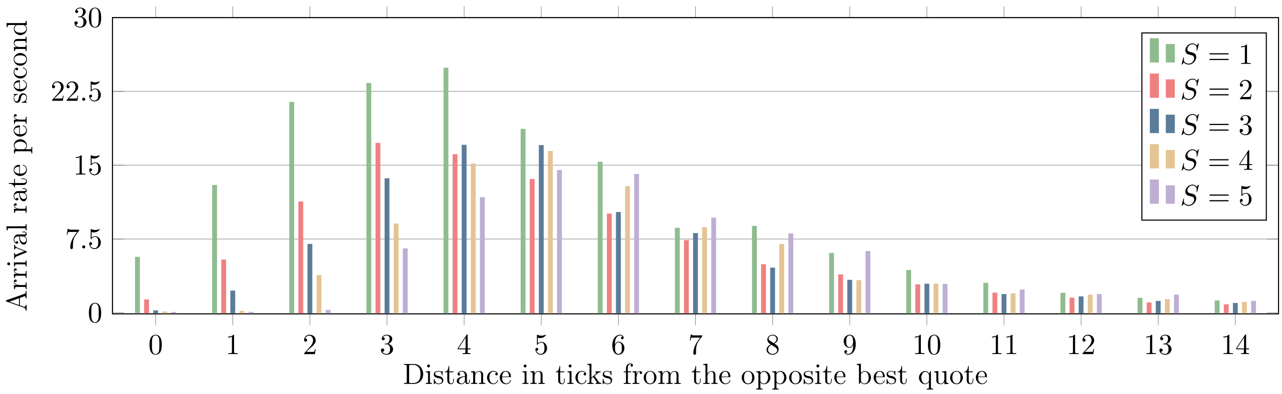
(a) Rates between 7–6–2021 and 11–6–2021



(b) Rates between 14–6–2021 and 18–6–2021



(c) Rates between 21–6–2021 and 25–6–2021

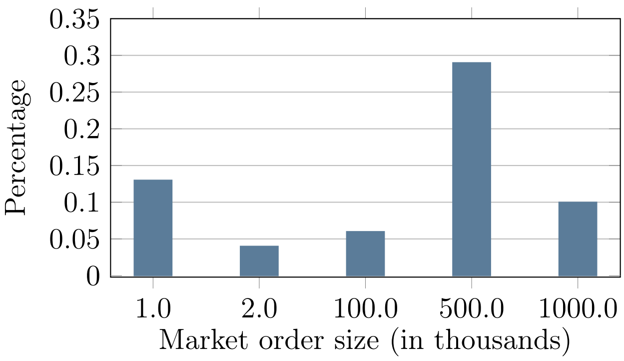


(d) Rates between 28–6–2021 and 2–7–2021

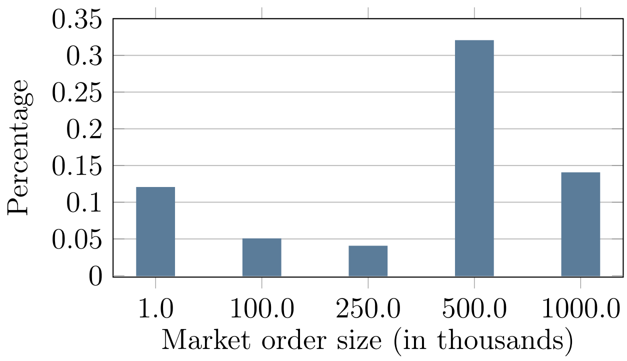
Figure 5.7: Arrival rates of limit orders as a function of the distance δ in ticks from the opposite best quote for each spread size S in ticks.

5.1.5 Market Orders

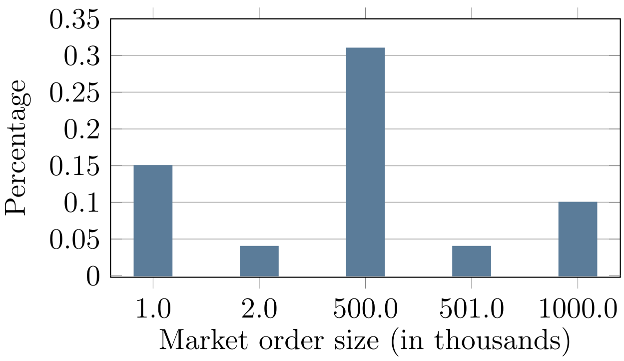
Figure 5.8 displays the distribution of the five most common market order sizes for each week in the data set. The x-axis denotes the order sizes in thousands and the y-axis shows the percentage of market orders with a certain size during each week in the data set. The most common market order sizes are 1 million, 500 thousand and 1 thousand. We see that relatively small orders of a few thousand arrive quite regularly. From the data it is hard to find a reason for the arrival of such small orders, but an explanation could be that algorithmic trading systems post small orders to test the market dynamics and find an optimal price before submitting a larger order. Similar to the limit order sizes, we see that the assumption of unit order size is not a very realistic one, as a wide range of order sizes is observed in the data. However, the set of order sizes that occur most frequently is relatively small. Other order sizes occur around or below 1% of the time.



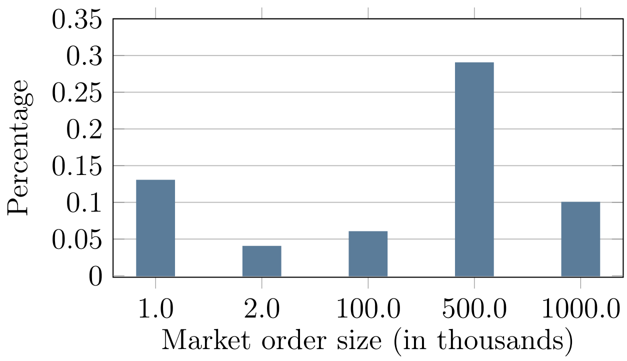
(a) Size distribution between 7–6–2021 and 11–6–2021. The average market order size is 339901.64.



(b) Size distribution between 14–6–2021 and 18–6–2021. The average market order size is 400128.26.



(c) Size distribution between 21–6–2021 and 24–6–2021. The average market order size is 320173.47.



(d) Size distribution between 28–6–2021 and 2–7–2021. The average market order size is 361490.31.

Figure 5.8: Distribution of market order sizes, for the five most commons market order sizes in each week in the data set.

The subfigures in Figure 5.9 show the rates of both buy and sell orders market, for the most common values of the spread size, which range from one to five ticks. The rates are computed using

$$\frac{N_m^S}{T_*^S}, \tag{5.5}$$

with N_m^S the number of arrivals of market orders during the time sample when the spread was equal to S and T_*^S the total time in seconds the spread was equal to S. We observe the arrival rate is the higher for smaller spread sizes and seems to decay exponentially. This makes sense since a small spread size indicates a higher trading activity. The exact rates can be found Table B.4 in Appendix B.

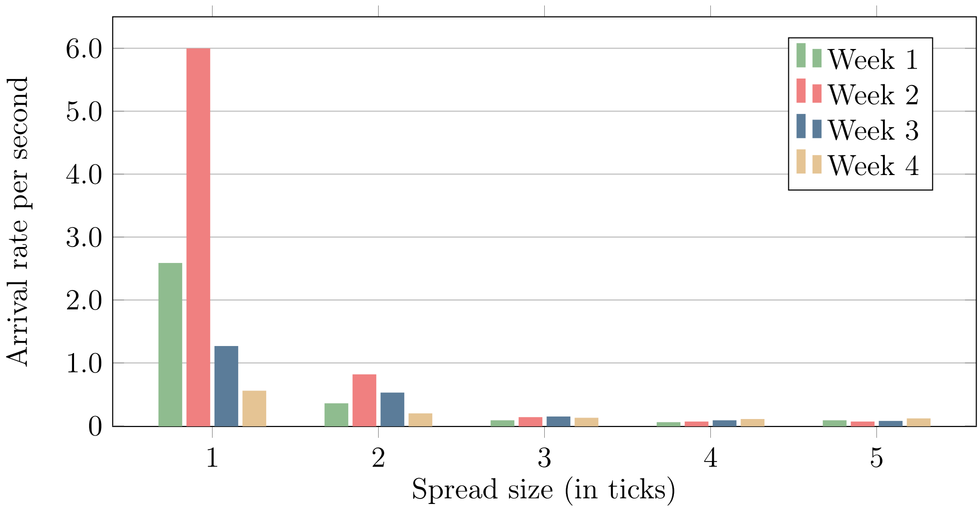


Figure 5.9: Arrival rates per second for market orders for each spread size between one and five ticks.

5.1.6 Cancellations

The last type of event for which the arrival rate needs to be determined is the cancellation of limit orders. As shown in Table 5.2, around 99.9% of all outstanding limit orders is cancelled. For this reason, the distribution of the cancellation sizes as shown in Figure 5.10 is almost exactly the same as the size distribution for limit orders (Figure 5.6). This leads to the conclusion that for cancellations, the assumption of unit size is also not a realistic one. However, since the cancellation sizes are very similar to the limit order sizes, assuming that limit orders have unit size can justify that cancellations have unit size too.

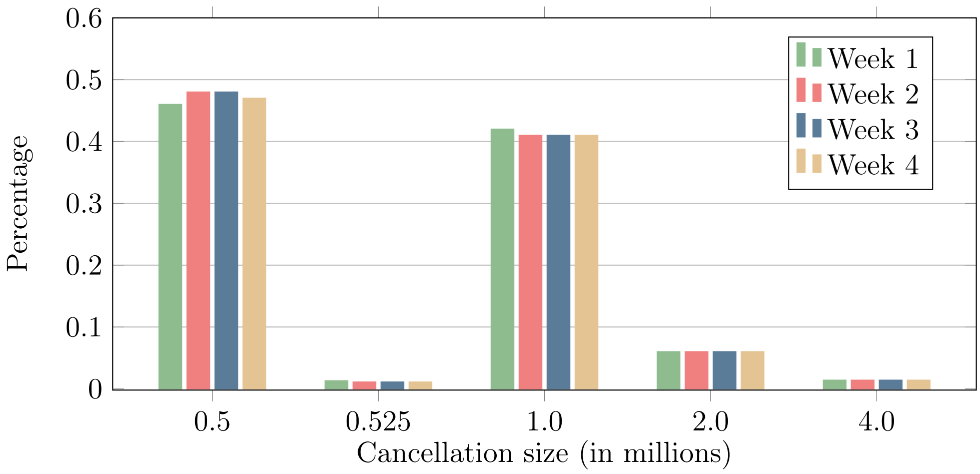
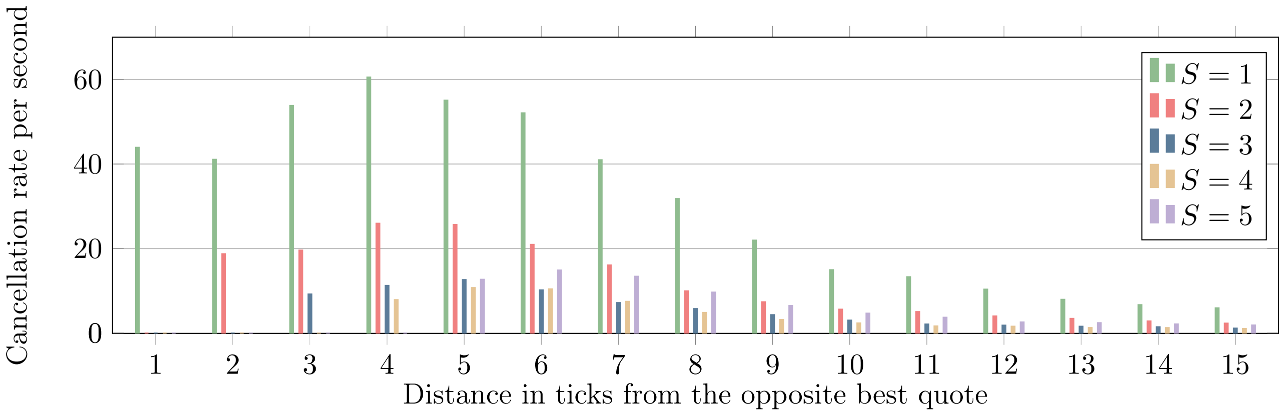


Figure 5.10: Distribution of cancellation sizes, for sizes occurring more than 1% of the time.

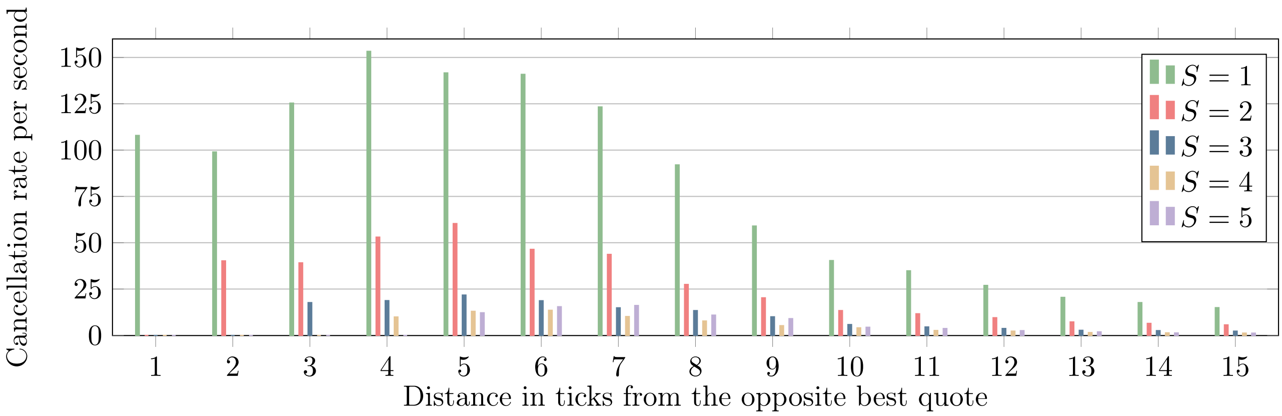
Unlike the size distribution, the arrival rate of cancellations does differ from the arrival rate of limit orders. This makes sense since cancellations, unlike limit orders, can only arrive at price levels with outstanding orders. The arrival rates are again computed by

$$\frac{N_c^S(\delta)}{T_*^S}, \tag{5.6}$$

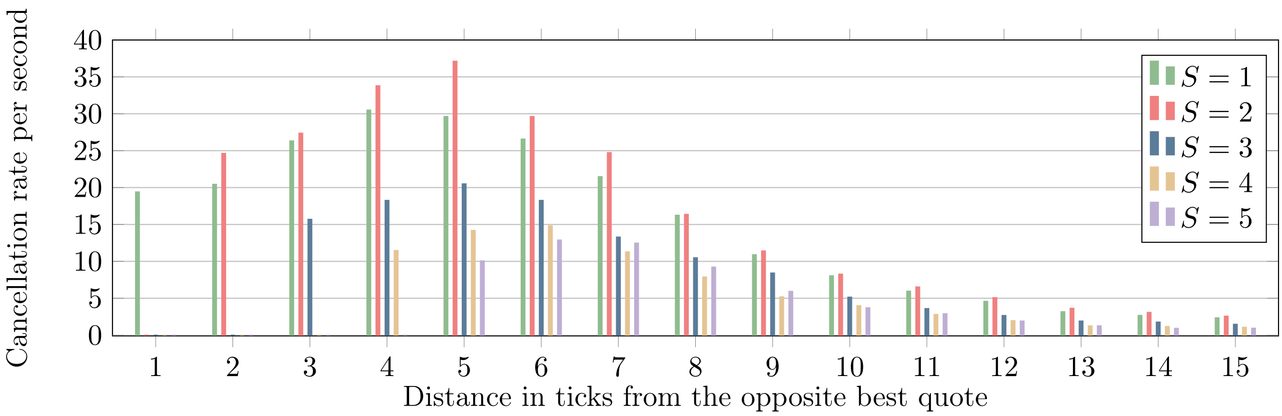
with $N_c^S(\delta)$ the number of arrivals of cancellations at distance δ in ticks from the opposite best quote during the time sample when the spread was equal to S and T_*^S the total time in seconds the spread equalled S. Figure 5.3 confirms that the cancellations arrive at the best quote or deeper in the book, for each value of the spread. Again we observe that the rates differ significantly for different spread sizes. The exact cancellation rates can be found in Table B.5 in Appendix B.



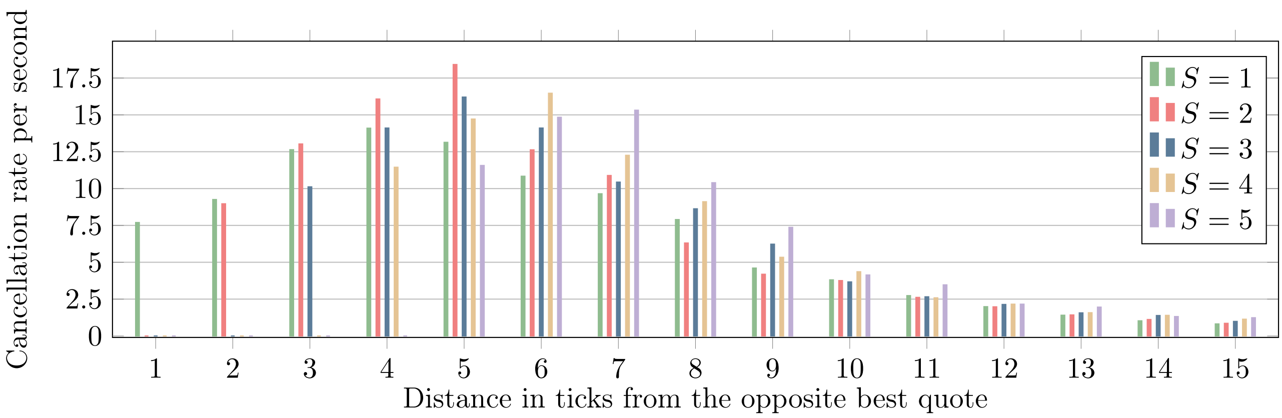
(a) Rates between 7–6–2021 and 11–6–2021.



(b) Rates between 14–6–2021 and 18–6–2021.



(c) Rates between 21–6–2021 and 25–6–2021.



(d) Rates between 28–6–2021 and 2–7–2021.

Figure 5.11: Cancellation rates as a function of the distance δ in ticks from the opposite best quote for each spread size S in ticks.

5.2 Parameter Estimation

5.2.1 Arrival Rates

The arrival rates $\lambda(\delta), \mu$ and $\theta(\delta)$ for incoming orders and cancellations, as described in Section 2.2, can be estimated relatively easy using order book data. Cont et al. provide the following functions to estimate the arrival rates of different events [12]. The function for the arrival of limit orders is estimated by

$$\hat{\lambda}(\delta) = \frac{N_l(\delta)}{T_*}, \tag{5.7}$$

where $N_l(\delta)$ denotes the total number of limit orders that arrived at a distance δ in ticks from the opposite best quote within a certain sample time T_*. Similarly, the arrival rate of market orders is estimated by

$$\hat{\mu} = \frac{N_m}{T_*}\frac{S_m}{S_l}, \tag{5.8}$$

with N_m the number of market orders, T_* the total sample time, S_m the average size of market orders and S_l average size of limit orders. Lastly, the estimator for the cancellation rate function is given by

$$\hat{\theta}(\delta) = \frac{N_c(\delta)}{T_* Q_\delta}\frac{S_c}{S_l}, \tag{5.9}$$

with S_c the average size of cancelled orders and S_l the average size of limit orders. $N_c(\delta)$ denotes the number of cancellations at price level a distance δ in ticks from the opposite best quote within the time frame T_*. Here, Q_δ is the average of Q_δ^{Ask} and Q_δ^{Bid}, with

$$Q_\delta^{\text{Bid}} = \frac{1}{S_l}\frac{1}{M}\sum_{j=1}^{M} S_\delta^{\text{Bid}}(j), \quad Q_\delta^{\text{Ask}} = \frac{1}{S_l}\frac{1}{M}\sum_{j=1}^{M} S_\delta^{\text{Ask}}(j) \tag{5.10}$$

where M denotes the number of event rows in the data and $S_\delta^{\text{Bid}}(j)$ ($S_\delta^{\text{Ask}}(j)$) the number of outstanding bid (ask) orders at a distance of δ ticks from the ask (bid) on the jth row of the data. That means that Q_δ can be seen as the average quantity outstanding at each price level.

For the rates $\hat{\mu}$ and $\hat{\theta}(\delta)$ we multiply them by their average size relative to the average size of limit orders. This is because of the assumption that orders have unit size, equal to the average limit order size.

While these rates provide a comprehensive picture of the dynamics of the order book, they might be further refined to capture the dynamics even better. In the previous sections we have demonstrated that the arrival and cancellations rates are different for different sizes of the spread S. For this reason, we propose to add a dependency to the rates based on the spread size to improve the model. To integrate the spread size into the model, the arrival and cancellation rates are re-estimated using the data specific to each spread size. This involves slightly modified versions of the formulas described above.

For limit orders, the arrival rate of orders arriving at a distance δ from the opposite best quote given that the spread is equal to S is now estimated by

$$\hat{\lambda}^S(\delta) = \frac{N_l^S(\delta)}{T_*^S}, \tag{5.11}$$

where $N_l^S(\delta)$ denotes the total number of limit orders that arrived at a distance δ from the opposite best quote when the spread was equal to S within a certain sample time T_*^S. The arrival rate per second for each week and each spread size is given in Table B.1. However, this is the rate for both sell and buy orders combined. To obtain the arrival rate for both sell and buy limit orders separately we need to multiply the rates in Table B.1 by $\frac{1}{2}$. The arrival rate is then symmetric and equal to the average rate of buy and sell orders.

Similarly, the arrival rate of market orders is estimated by

$$\hat{\mu}^S = \frac{N_m^S}{T_*^S}\frac{S_m}{S_l}, \tag{5.12}$$

with N_m^S the number of market orders that arrived when the spread was equal to S, T_*^S the sample time, S_m the average size of market orders and S_l average size of limit orders. The ratios $\frac{S_m}{S_l}$ for the four separate weeks in the data set are given in Table 5.4. Similar to the limit order arrival rate, we need to multiply the rates by both $\frac{1}{2}$ to average the rate of buy and sell orders, as well as by the ratio between the average order sizes.

Table 5.4: Ratio between average market order size and average limit order size.

	Week 1	Week 2	Week 3	Week 4
S_m/S_l	0.40	0.47	0.38	0.43

Finally, the estimator for the cancellation rate function is given by

$$\hat{\theta}^S(\delta) = \frac{N_c^S(\delta)}{T_*^S Q_\delta^S} \frac{S_c}{S_l}, \tag{5.13}$$

with S_c the average size of cancelled orders and S_l the average size of limit orders. $N_c^S(\delta)$ denotes the number of cancellations at price level a distance δ in ticks from the opposite best quote during spread size S within the time frame T_*^S. Q_δ^S is the average of $Q_\delta^{S,\text{Ask}}$ and $Q_\delta^{S,\text{Bid}}$, which denote the average quantities outstanding at each price level at the ask and bid side, with

$$Q_\delta^{S,\text{Bid}} = \frac{1}{S_l} \frac{1}{M} \sum_{j=1}^{M} S_\delta^{S,\text{Bid}}(j), \quad Q_\delta^{S,\text{Ask}} = \frac{1}{S_l} \frac{1}{M} \sum_{j=1}^{M} S_\delta^{S,\text{Ask}}(j) \tag{5.14}$$

Here, M denotes the number of event rows in the data and $S_j^{S,\text{Bid}}(\delta)$ ($S_j^{S,\text{Ask}}(\delta)$) the number of outstanding bid (ask) orders at a distance of δ ticks from the ask (bid) on the jth row in the event data when the spread was equal to S. The ratio between the average cancellation size and average order size for each week in the data set is given in Table 5.5.

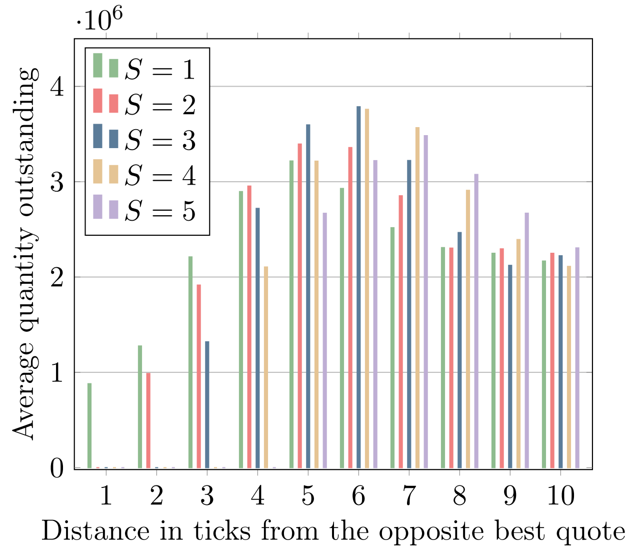
Table 5.5: Ratio between average cancellation size and average limit order size.

	Week 1	Week 2	Week 3	Week 4
S_c/S_l	1.00	1.00	1.00	1.00

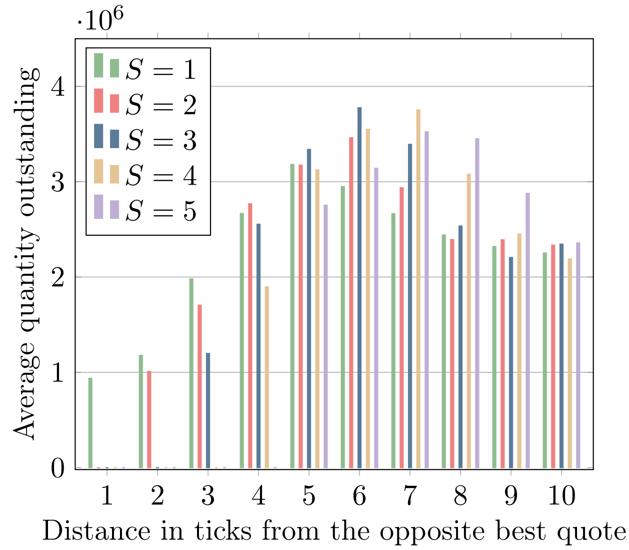
5.2.2 Average Number of Outstanding Orders

As described in Section 5.2.1, the arrival rate of cancellations depends on the average number of outstanding orders at a certain distance from the opposite best quote. To estimate the arrival parameter $\theta^S(\delta)$, the average number of outstanding orders at each price level and for each spread size between one and five ticks needs to be determined. Figure 5.12 displays the average outstanding quantities for each price level at a distance of one to ten ticks from the opposite best quote. In this context, outstanding quantity means the sum of quantities of all limit orders at a price level with a particular distance from the opposite best quote for each timestamp.

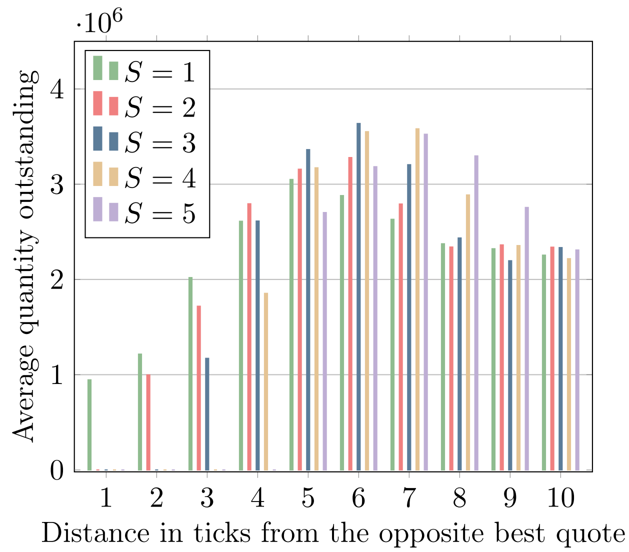
We observe that average number of outstanding quantities is similar for each week in the data. We also see that on average, the best quote has the lowest average outstanding quantity for each spread size and the quantity increases for levels deeper in the order book until approximately four to five ticks from the best quote. The average quantity decreases again and stays more or less constant for levels deeper in the order book. The exact values of the average outstanding quantities can be found in Table B.7 in Appendix B.



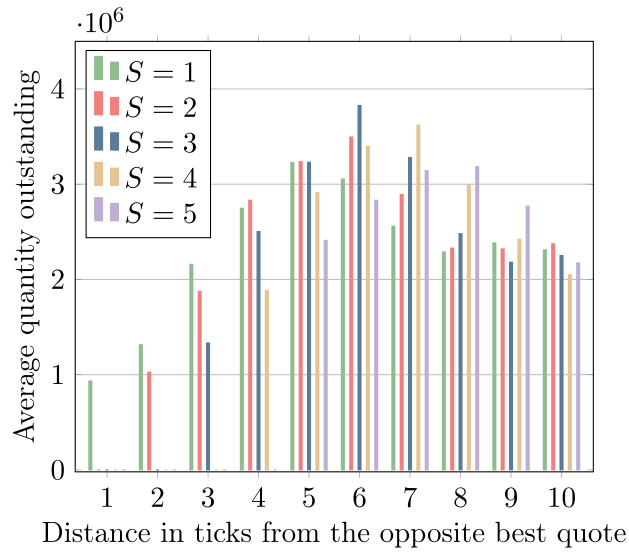
(a) Average quantities between 7–6–2021 and 11–6–2021.



(b) Average quantities between 14–6–2021 and 18–6–2021.



(c) Average quantities between 21–6–2021 and 25–6–2021.



(d) Average quantities between 28–6–2021 and 2–7–2021.

Figure 5.12: Average quantity outstanding at each distance δ in ticks from the opposite best quote for each spread size S in ticks for $\delta = 1, ..., 10$.

5.2.3 Average Order Sizes

The functions in Equations 5.11, 5.12 and 5.13 determine the arrival rates of incoming orders and cancellations for each size of the spread. As previously mentioned, due to the unit size assumption the arrival rates for market orders and cancellations are multiplied by the ratios between their average sizes and that of limit orders. These average sizes, unlike the arrival rates, do not depend on the spread size. Table 5.6 shows the average size of limit orders S_l, market orders S_m and cancellations S_c for each spread size, together with the ratios. We notice that the spread size does not significantly influence the ratios and there is no obvious or significant relationship between the spread size and the ratios. For that reason we assume the ratios to be equal for each spread size.

Table 5.6: Average sizes of limit orders S_l, market orders S_m and cancellations S_c and the ratios between the average sizes for various spread sizes S in ticks.

S	S_l	S_m	S_c	S_m/S_l	S_c/S_l
1	838449.26	331034.87	820172.66	0.40	0.98
2	844179.94	380353.36	832940.77	0.45	0.99
3	854789.86	347039.02	850841.29	0.41	1.00
4	856000.71	354098.51	863128.17	0.41	1.01
5	850569.99	400764.76	861386.95	0.47	1.01

(a) Average sizes between 7–6–2021 and 11–6–2021.

S	S_l	S_m	S_c	S_m/S_l	S_c/S_l
1	839281.24	399852.32	821391.64	0.48	0.98
2	842035.52	394185.99	829351.65	0.47	0.98
3	850608.77	417543.58	841085.84	0.49	0.99
4	857147.21	412903.30	859889.27	0.48	1.00
5	847244.42	458038.04	859902.23	0.54	1.01

(b) Average sizes between 14–6–2021 and 18–6–2021.

S	S_l	S_m	S_c	S_m/S_l	S_c/S_l
1	837536.92	371935.19	823018.62	0.44	0.98
2	838805.46	378048.34	824690.14	0.45	0.98
3	844005.13	338163.91	835460.29	0.40	0.99
4	843282.36	300020.34	845812.33	0.36	1.00
5	832494.86	346079.78	846318.47	0.42	1.01

(c) Average sizes between 21–6–2021 and 25–6–2021.

S	S_l	S_m	S_c	S_m/S_l	S_c/S_l
1	842967.89	393082.81	833429.96	0.47	0.99
2	840465.41	389173.14	832937.51	0.46	0.99
3	841592.27	368210.23	840983.28	0.44	1.00
4	836619.40	341209.80	843416.34	0.41	1.01
5	823680.00	411997.99	839521.97	0.50	1.02

(d) Average sizes between 28–6–2021 and 2–7–2021.

6 | Numerical Results

In this chapter we conduct some numerical experiments to assess the performance of the model. In Section 6.1 we conduct an experiment to test the models ability to capture the short-term behaviour of the order book. We do this by comparing the theoretical probability that the quantity at a certain price level will increase with the empirical frequency. In Section 6.2 we compare the performance of the two methods for inverting the Laplace transform that were introduced in Section 3.1, the Euler method and the COS-method. Furthermore, we compute the conditional probabilities of interest, including the probability of an increase in mid-price (Section 6.3), the probability of executing an order placed at the best bid (Section 6.4) and the probability of executing an order at one price level below the best bid (Section 6.5). The expressions used to compute these probabilities are given in Sections 4.1, 4.2 and 4.3. We compare the probability forecasts to the empirical probabilities, and assess the performance of the model using a relative error measure: the mean absolute percentage error.

We implement the model in Python. The code used to compute the Laplace transforms, continued fractions and density functions can be found in Appendix C.

6.1 Verification of One-Step Transition Probabilities

To assess if the estimated parameters are a good fit for the real data and for predictions about short-term behaviour of the order book, Cont et al. compare the one-step probabilities of the model with the observed empirical frequencies [12]. The one-step probability refers to the likelihood of an *increase* in quantity at a certain price level at the next event, given the current number of outstanding orders. Incoming limit orders are events that increase the quantity, while cancellations result in a quantity decrease for price levels smaller than the best bid or larger than the best ask. At the best quotes, a decrease can also be the result of an incoming market order.

Let T_m be the time of the m-th event in the order book, and similarly, define T_{m+1} as the first event after the m-th event. We have

$$T_0 = 0, \quad T_{m+1} \equiv \inf\{t > T_m \mid X(t) \neq X(T_m)\}. \tag{6.1}$$

Events in the order book at a distance δ from the opposite best quote occur at the best quotes, i.e. for $\delta = S$, with rate

$$\lambda(\delta) + \mu + q\theta(\delta), \tag{6.2}$$

with S the spread size in ticks and q the number of outstanding orders. At other price levels, i.e. for $\delta > S$, events occur with rate

$$\lambda(\delta) + q\theta(\delta), \tag{6.3}$$

since market orders arrive only at the best quote. The probability that the next change in quantity from q to $q+1$ at a price level a distance δ from the opposite best quote is an increase, denoted by $P_\delta(q)$, is therefore given by

$$P_\delta(q) \equiv \mathbb{P}[Q_\delta(T_{m+1}) = q+1 \mid Q_\delta(T_m) = q, Q_\delta(T_{m+1}) \neq q] = \begin{cases} \frac{\lambda(S)}{\lambda(S)+\mu+q\theta(S)}, & \delta = S, \\ \frac{\lambda(\delta)}{\lambda(\delta)+q\theta(\delta)}, & \delta > S, \end{cases} \tag{6.4}$$

where $Q_\delta(t)$ denotes the quantity outstanding at a distance δ from the opposite best quote at time t. Now let Q^A_δ denote the ask quantities outstanding at a distance δ from the best bid and similarly, let Q^B denote the bid quantities outstanding at a distance δ from the best ask. We can compare the theoretical probability $P_\delta(q)$ to the empirical frequencies of an increase $\hat{P}_\delta(q)$, estimated by

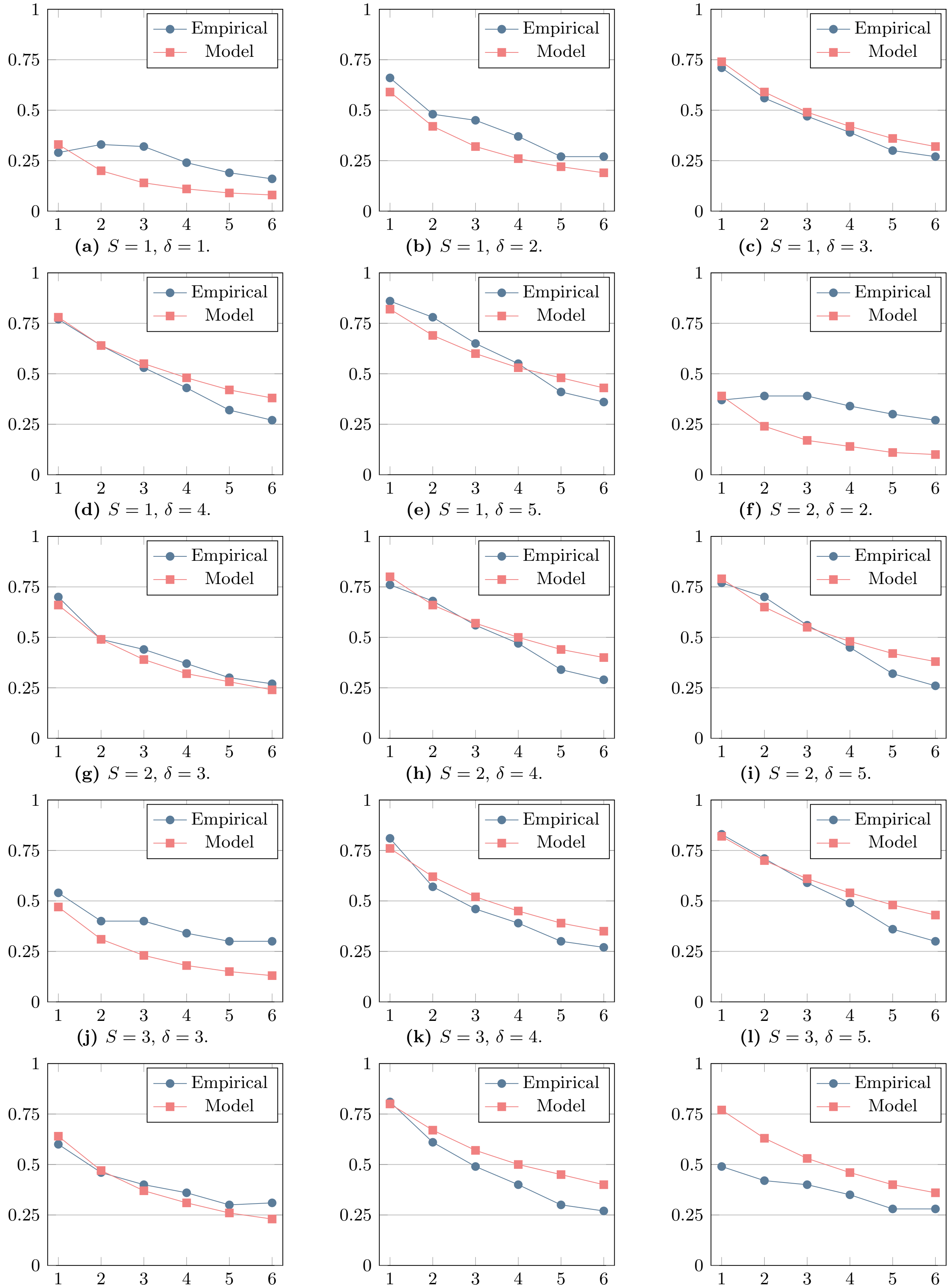
$$\hat{P}_\delta(q) \equiv \frac{\hat{B}_{\text{up}} + \hat{A}_{\text{up}}}{\hat{B}_{\text{up}} + \hat{A}_{\text{up}} + \hat{B}_{\text{down}} + \hat{A}_{\text{down}}}, \tag{6.5}$$

with

$$\begin{aligned} \hat{B}_{\text{up}} &\equiv \#\{\hat{Q}^B_\delta(\hat{T}_m) = q, \hat{Q}^B_\delta(\hat{T}_{m+1}) = q+1\}, \\ \hat{A}_{\text{up}} &\equiv \#\{\hat{Q}^A_\delta(\hat{T}_m) = q, \hat{Q}^A_\delta(\hat{T}_{m+1}) = q+1\}, \\ \hat{B}_{\text{down}} &\equiv \#\{\hat{Q}^B_\delta(\hat{T}_m) = q, \hat{Q}^B_\delta(\hat{T}_{m+1}) = q-1\}, \\ \hat{B}_{\text{down}} &\equiv \#\{\hat{Q}^A_\delta(\hat{T}_m) = q, \hat{Q}^A_\delta(\hat{T}_{m+1}) = q-1\}, \end{aligned} \tag{6.6}$$

where $\hat{Q}^B_\delta$ and $\hat{Q}^A_\delta$ are the empirical quantities outstanding at each level on the bid and ask side of the order book, respectively. Since we assume the orders to be of unit size, $\hat{Q}^B_\delta$ and $\hat{Q}^A_\delta$ are computed by dividing the quantities outstanding at each new event by the average limit order size S_l, rounded off to the nearest positive integer.

Figure 6.1 illustrates the results for $P_\delta(q)$ and $\hat{P}_\delta(q)$ for the prediction of one-step probabilities on July 2nd 2021 for various spread sizes S and distances δ from the opposite best quote. Particularly the results for price levels different to the best quotes, i.e. for $S \neq \delta$, show significant similarity. For the probabilities at the best quotes, the model tends to slightly underestimate the probabilities for $S = 1, 2, 3$ (See Figures 6.1a, 6.1f, 6.1j). On the contrary, for $S = 5$ the model seems to overestimate the probabilities of a price increase, as can be seen in Figure 6.1o. What also stands out from Figure 6.1a and Figure 6.1f is that the probability of an increase in quantity increases for $q = 1$ to $q = 3$ at the best quote when the spread is equal to one or two ticks. This is not in accordance with the model, since the probability of an increase should decrease when the outstanding quantity is larger. This follows from the fact that more orders would result in a higher probability that at least one of them is cancelled, which would result in a quantity decrease. Overall, although it is obvious that not all dynamics are captured by the model, in most cases it shows a reasonable ability to capture the one-step probabilities.



(a) $S = 1$, $\delta = 1$.
(b) $S = 1$, $\delta = 2$.
(c) $S = 1$, $\delta = 3$.
(d) $S = 1$, $\delta = 4$.
(e) $S = 1$, $\delta = 5$.
(f) $S = 2$, $\delta = 2$.
(g) $S = 2$, $\delta = 3$.
(h) $S = 2$, $\delta = 4$.
(i) $S = 2$, $\delta = 5$.
(j) $S = 3$, $\delta = 3$.
(k) $S = 3$, $\delta = 4$.
(l) $S = 3$, $\delta = 5$.
(m) $S = 4$, $\delta = 4$.
(n) $S = 4$, $\delta = 5$.
(o) $S = 5$, $\delta = 5$.

Figure 6.1: Probabilities of an increase in queue size for different values of the spread S and different price levels at a distance δ from the opposite best quote. The y-axis denotes the probability that the quantity at a certain level increases at the next quantity change. The x-axis denotes the quantity outstanding q before the next change.

6.2 Comparison of Numerical Methods for Inverting Laplace Transforms

In Section 3.1 we provided two methods to invert Laplace transforms: the Euler method and the COS-method. To compare both methods, we compute the density functions of the first-passage times using the arrival rate parameters given in the paper by Cont et al. for the case the spread size is equal to one tick, which can be found in Table 6.1. We compute and plot the pdf's for several values of the queue size q and measure the computation time.

Table 6.1: Arrival rate parameters from [12] for a spread size S equal to one tick.

$\hat{\lambda}(\delta)$	1.85
$\hat{\theta}(\delta)$	0.71
$\hat{\mu}$	0.94

The pdf's are computed by inverting the Laplace transforms of the density functions of the first-passage times to 0, conditional on the number of orders outstanding being equal to q. We recall that this Laplace transform is given by

$$\hat{f}_q^S(s) = \prod_{i=1}^{q} \left(-\frac{1}{\lambda(S)} \mathrm{K}_{k=i}^{\infty} \frac{-\lambda(S)\big(\mu + k\theta(S)\big)}{\lambda(S) + \mu + k\theta(S) + s} \right). \tag{6.7}$$

The continued fractions in Equation 6.7 are the expressions of the Laplace transforms of the density functions for the separate first-passage times from state i to state $i-1$, given by

$$\hat{f}_{i,i-1}^S(s) = -\frac{1}{\lambda(S)} \mathrm{K}_{k=i}^{\infty} \frac{-\lambda(S)\big(\mu + k\theta(S)\big)}{\lambda(S) + \mu + k\theta(S) + s}. \tag{6.8}$$

To evaluate these expressions, first we notice that $\hat{f}_{i,i-1}(s)$ is a continued fraction of the form

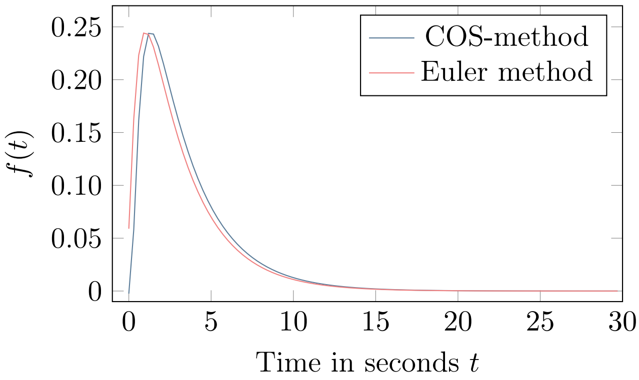
$$\hat{f}_{i,i-1}^S(s) = \mathrm{K}_{k=i}^{\infty} \frac{a_k}{b_k} = \frac{a_k}{b_k+} \frac{a_{k+1}}{b_{k+1}+} \frac{a_{k+2}}{b_{k+2}+} \cdots, \tag{6.9}$$

with

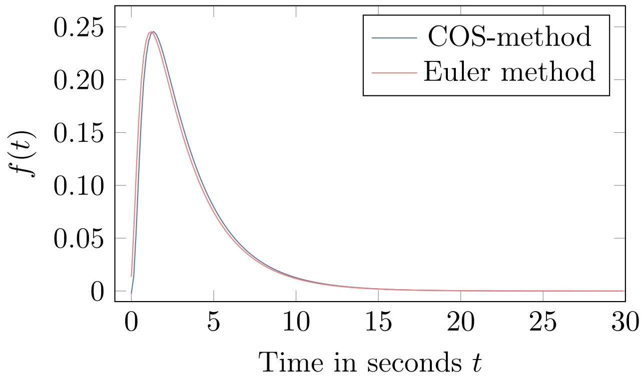
$$\begin{aligned} a_k &= \mu + k\theta(S), \\ a_{k+l} &= -\lambda(S)\big(\mu + (k+l)\theta(S)\big), \quad \text{for } l \geq 1, \\ b_{k+l} &= \lambda(S) + \mu + (k+l)\theta(S) + s, \text{ for } l \geq 0. \end{aligned} \tag{6.10}$$

We can compute the value of this infinite continued fraction using the modified Lentz method as described in Section 3.2 with the terms in the continued fraction a_n and b_n as defined in Equation 6.10.

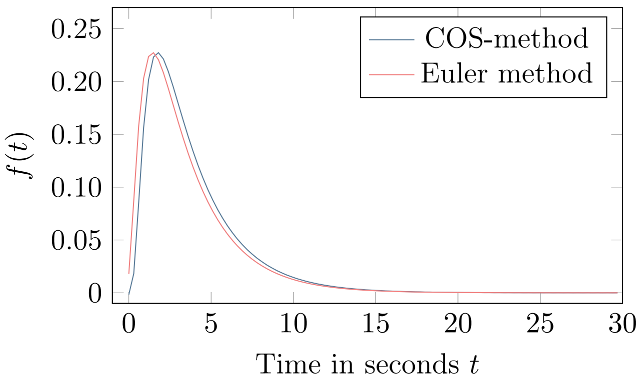
To compare both methods, we take for the Euler method $A = 18.4$, $m = 11$ and $n = 15$, as proposed by Abate and Whitt in [2]. For the COS-method we take 50 for the number of terms in the summation N. We compute the pdf's for the case $S = 1$ for t between 0 and 30 seconds. The number of time steps is varied between 100 and 200. The computation times for both methods are given in Table 6.2 and the plots of the pdf's of the first-passage times can be found in Figure 6.2. Figure 6.2 shows that the COS-method provides similar results to the Euler method. However, from Table 6.2 follows that the computation time of the COS-method is significantly smaller than the Euler method. For this reason we choose to compute the inverse Laplace transforms using the COS-method.



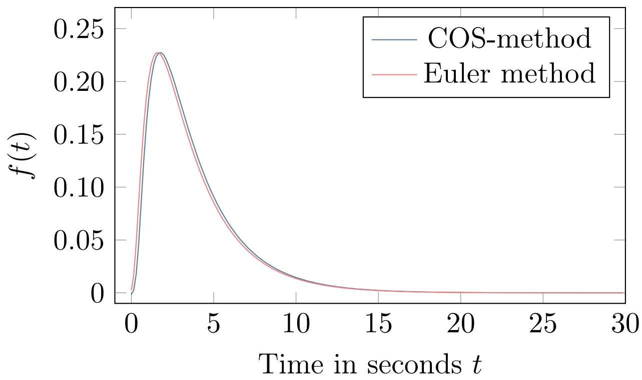
(a) $q = 4,\ M = 100$



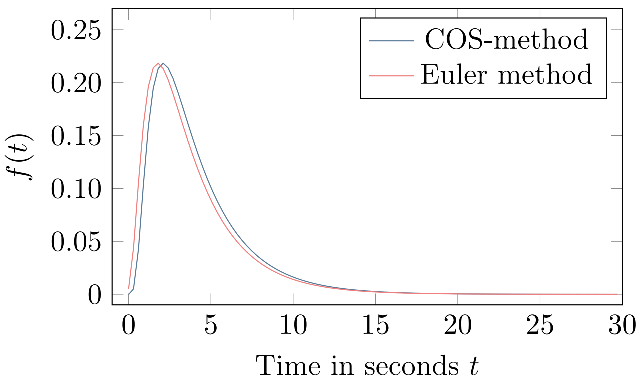
(b) $q = 4,\ M = 200$



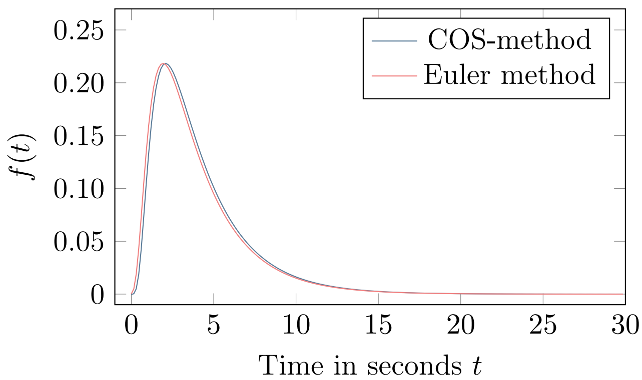
(c) $q = 5,\ M = 100$



(d) $q = 5,\ M = 200$



(e) $q = 6,\ M = 100$



(f) $q = 6,\ M = 200$

Figure 6.2: Probability density functions $f(t)$ of first-passage times for several values of the queue size q and number of time steps M, computed using the Euler method and the COS-method.

Table 6.2: Comparison of computation times of the density functions $f(t)$ of the first-passage times using the Euler method and COS-method for several values of the number of orders q and the number of time steps M.

	q					
	4		5		6	
M	Euler	COS	Euler	COS	Euler	COS
100	7.0s	2.2s	9.5s	2.5s	8.9s	3.0s
200	14.0s	4.4s	15.6s	5.1s	17.6s	5.2s

6.3 Computation of Probability of Mid-Price Increase

In this section we compute the conditional probability of an increase in mid-price using Laplace transforms and compare them to the empirical distribution based on the frequencies of mid-price changes. We will compute the probabilities based on different parameter sets and evaluate which set provides the best results. The probability of a mid-price increase can be computed by inverting the Laplace transforms given in Proposition 4.1 using the COS-method, which is described in Section 3.1. The probabilities are computed for a five day period from 29–6–2021 to 5–7–2021.

First, we compute the parameters using the data from the previous four instances of the same weekday, i.e. we use the data from the four previous Mondays to predict the rates for the next Monday, and the same holds for the other days in the week. We then compute the price move probabilities using rates that are independent of the spread, i.e. the parameters are the same for all spread sizes S, and rates that do depend on the spread size. The probabilities are computed for several quantities q_A and q_B outstanding at $t = 0$ at the best ask and best bid, respectively, ranging from 1 to 5. Since the model assumes orders to have unit size, we have computed the empirical probabilities by dividing the quantities at the best ask and best bid price by the average limit order size, and rounding them off to the nearest positive integer. The empirical probability of an increase in mid-price can then be computed by

$$P^S_{\text{increase}} = \frac{\#\{p_M(t_M) > p_M(t),\ X_A(t) = q_A,\ X_B(t) = q_B,\ p_S(t) = S\}}{\#\{p_M(t_M) \neq p_M(t),\ X_A(t) = q_A,\ X_B(t) = q_B,\ p_S(t) = S\}}, \tag{6.11}$$

with $p_M(t)$ the mid-price at time t, $p_M(t_M)$ the mid-price after the first change in mid-price, $X_A(t)$ and $X_B(t)$ the number of orders outstanding at the best ask and bid price, respectively and $p_S(t)$ the spread size at time t. We also compute the probability of a mid-price increase based on rates calculated on the preceding five days, i.e., data from Monday to Friday is used to predict the rates for the following Monday.

We compute the probabilities using the COS-method with the parameters given in Table B.8. To compare the results for the four different parameter sets, we compute the mean absolute percentage error (MAPE) for all the estimated probabilities, which are shown in Table 6.3. The MAPE is given by

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{P_i - \hat{P}_i}{P_i}\right|, \tag{6.12}$$

with P_i the empirical probability, $\hat{P}_i$ the estimated probability and n the number of predictions. For this error measure holds that the closer its value is to zero, the better the prediction. If the MAPE is less than 5%, the forecast is assumed to be acceptably accurate, while the accuracy of a prediction is considered to be low but acceptable when the MAPE lies between 5% and 25% [24].

Table 6.3: Mean absolute percentage error between empirical probabilities of a mid-price increase and computed probabilities using spread dependent and independent rates and estimated on different parameter sets.

	MAPE			
	Preceding days		Previous instances of the same day	
S	Spread independent rates	Spread dependent rates	Spread independent rates	Spread dependent rates
1	11.3%	**10.4%**	11.4%	10.6%
2	12.0%	11.4%	13.0%	**10.8%**
3	**12.7%**	13.3%	14.7%	14.6%
4	6.4%	**4.7%**	6.6%	5.4%
5	10.8%	8.8%	10.8%	**8.1%**
Average	10.6%	**9.7%**	11.3%	9.9%

As can be observed from Table 6.3, the accuracy of the computed probabilities can therefore be considered acceptably accurate for all parameter sets and all spread sizes. As we would suspect, the probabilities based on spread dependent rates are in general slightly better than using rates that are independent of the spread size. However, no specific set of parameters clearly outperforms the others. More research could be done to examine which time period could be used to estimate the parameters that have the most predictive power. To compute other conditional probabilities in the coming sections, we choose to use the parameter set that is estimated on five preceding days with spread dependent rates, since it has the lowest MAPE over all days and all spread sizes.

Table 6.4 shows the empirical frequencies of a mid-price increase and the model probabilities on 5–7-2021, computed using spread dependent rates based on the five preceding days. For each spread size and each combination of quantities q_B and q_A at the best bid price and best ask price, respectively, the empirical probability of a mid-price increase is computed using Equation 6.11. Not all combinations of q_A and q_B have a value for the empirical probability. This is because these combinations of quantities at a certain spread occurred less than 100 times. We have chosen to disregard these instances to ensure we can assume the empirical probability is reliable. We observe that in most cases, the predictions are reasonably accurate. In the model, the probability of a mid-price increase is decreasing for larger values of q_A, and conversely is increasing for larger values of q_B. This behaviour can be found in the majority of the empirical probabilities as well. This, in combination with the values of the MAPE indicating an acceptable accuracy, suggest that the model is able to capture the short-term mid-price dynamics quite effectively.

Table 6.4: Empirical frequency and computed probability of an increase in mid-price for several sizes of the spread S and initial values of the quantities at the best bid q_B and the best ask q_A on July 5th 2021. The model probability is computed using rates computed on the five preceding days. Only instances where a combination of q_A and q_B occurred more than 100 times are taken into account to compute a reasonable empirical probability.

		Empirical Probability					Model Probability				
		q_A					q_A				
	q_B	1	2	3	4	5	1	2	3	4	5
	1	50.3%	33.0%	22.9%	27.1%	22.4%	50.0%	34.7%	27.1%	22.6%	19.6%
	2	70.5%	56.6%	-	-	-	65.3%	50.0%	41.1%	35.3%	31.1%
$S = 1$	3	78.7%	-	-	-	-	72.9%	58.9%	50.0%	43.8%	39.2%
	4	78.2%	-	-	-	-	77.4%	64.7%	56.2%	50.0%	45.3%
	5	81.4%	-	-	-	-	80.5%	68.9%	60.8%	54.8%	50.0%
	1	49.6%	38.5%	32.2%	25.2%	22.0%	50.0%	36.8%	30.9%	27.6%	25.6%
	2	58.1%	48.5%	52.3%	45.5%	27.5%	63.3%	50.0%	43.2%	39.1%	36.4%
$S = 2$	3	70.1%	49.9%	-	-	-	69.2%	56.8%	50.0%	45.7%	42.7%
	4	75.1%	43.9%	-	-	-	72.4%	60.9%	54.4%	50.0%	47.0%
	5	81.6%	-	-	-	-	74.5%	63.7%	57.3%	53.1%	50.0%
	1	49.9%	42.8%	39.7%	36.8%	31.8%	50.0%	38.9%	34.4%	32.2%	30.8%
	2	54.7%	48.1%	52.0%	47.5%	52.6%	61.1%	50.0%	45.0%	42.2%	40.5%
$S = 3$	3	56.3%	46.0%	45.3%	49.9%	72.2%	65.6%	55.0%	50.0%	47.12%	45.3%
	4	62.5%	49.4%	59.7%	47.0%	-	67.8%	57.8%	52.9%	50.0%	48.1%
	5	65.7%	51.9%	-	-	-	69.2%	59.5%	54.7%	51.9%	50.0%
	1	51.7%	44.0%	41.5%	38.6%	35.3%	50.0%	41.5%	38.8%	37.6%	37.0%
	2	56.5%	47.0%	45.1%	43.5%	44.1%	58.5%	50.0%	47.0%	45.6%	44.9%
$S = 4$	3	58.3%	52.8%	50.5%	52.8%	40.1%	61.2%	53.0%	50.0%	48.6%	47.8%
	4	61.5%	54.8%	55.6%	51.7%	48.3%	62.4%	54.4%	51.4%	50.0%	49.2%
	5	67.7%	58.7%	60.2%	54.0%	44.1%	63.0%	55.1%	52.2%	50.8%	50.0%
	1	39.1%	43.3%	49.3%	42.3%	51.0%	50.0%	44.4%	43.1%	42.7%	42.6%
	2	56.8%	45.5%	66.8%	52.1%	40.0%	55.6%	50.0%	48.6%	48.2%	48.0%
$S = 5$	3	63.2%	48.1%	53.9%	59.7%	53.7%	56.9%	51.4%	50.0%	49.5%	49.4%
	4	65.0%	54.1%	39.4%	53.6%	50.7%	57.3%	51.8%	50.5%	50.0%	49.8%
	5	65.7%	64.2%	42.5%	53.3%	45.2%	57.4%	52.0%	50.6%	50.2%	50.0%

6.4 Computation of Probability of Execution at the Best Bid Price

In this section we compute the conditional probability of executing an order placed at the best bid before the mid-price moves. For the computation we use the parameter set computed on the preceding five days with spread dependent rates. The probability of execution is calculated using the result of Proposition 4.2. We compare the computed probabilities with the empirical frequencies to assess the performance of the model. In reality most orders are cancelled, and the probability we wish to compute is conditional on an order not being cancelled. Therefore, we estimate the empirical probability in the following way:

$$P_{\text{fill}}^{S} = \frac{\#\{\text{Fill},\ p_M(t_F) = p_M(t_0)\}}{\#\{\text{Fill},\ p_M(t_F) = p_M(t_0)\} + \#\{\text{Cancel},\ p_M(t_C) \neq p_M(t_0)\}}. \tag{6.13}$$

Here, p_{fill}^{S} denotes the empirical fill probability, $p_M(t_0)$ the mid-price at the time of order submission t_0, $p_M(t_F)$ the mid-price at the time of execution t_F and $p_M(t_C)$ the mid-price at the time of cancellation t_C. Since our focus is on computing the execution probability before a mid-price move for orders that are 'never' cancelled, we calculate the ratio of the filled orders with respect to orders that are cancelled *after* a mid-price move. The reason for this is that these orders were 'never' cancelled before the mid-price moved, but they were also not executed. For ease of notation, we have omitted the dependency on X_A, X_B and p_S for the sets in Equation 6.13, for which we have

$$\{\text{Fill},\ p_M(t_F) = p_M(t_0)\} = \{\text{Fill},\ p_M(t_F) = p_M(t_0),\ X_A(t_0) = q_A,\ X_B(t_0) = q_B,\ p_S(t_0) = S\},$$

and

$$\{\text{Cancel}, p_M(t_C) \neq p_M(t_0)\} = \{\text{Cancel}, p_M(t_C) \neq p_M(t_0), X_A(t_0) = q_A, X_B(t_0) = q_B, p_S(t_0) = S\}.$$

We calculate all fill probabilities again using the COS-method, with the parameters as provided in Table B.9. Table 6.5 displays the empirical probabilities together with the calculated probabilities for July 5th 2021 for spread sizes S ranging from one to five ticks. Again, we decided only to include cases where a combination of q_A and q_B occurred more than 100 times to ensure a reliable empirical probability. For certain combinations, the model seems to perform quite well, particularly for the cases that $q_B = 1$. However, the decay in probability when q_B is larger than 1 seems to be greater for the empirical probability than for the model probability. Conversely, the empirical probability seems to increase more rapidly for larger values of q_A when $q_B = 1$. These observations suggest that the model may not be able to effectively capture all the dynamics in the order book. Additionally, we observe that for $S > 2$, the predicted fill probability does not change significantly for different values of q_A. An explanation for this could be that the probability of an order posted within spread, and thus leading to a change in mid-price, is larger for bigger spread sizes. The case where the mid-price changes due to the quantity at the best ask reaching 0 might be less significant for larger spread sizes, resulting in a relatively constant fill probability for different values of q_A. We see that the empirical values exhibit different behaviour, indicating that there are some dynamics that are not captured well by the model.

To asses the accuracy of the computed fill probabilities, we propose not to use the mean absolute percentage error. The reason for this is that the probabilities are typically quite small, and a significant amount even less than 1.0%, which will result in a large relative error. This error will even tend to infinity for empirical probabilities close to 0. Instead, we propose to use an error measure described in [17], which is called mean arctangent absolute percentage error (MAAPE). The MAAPE is given by

$$\text{MAAPE} = \frac{1}{n}\sum_{i=1}^{n} \arctan\left(\left|\frac{P_i - \hat{P}_i}{P_i}\right|\right), \tag{6.14}$$

with P_i the empirical probability, $\hat{P}_i$ the estimated probability and n the number of predictions. The main advantage of this method is that the errors are now bounded, so that they do not tend to infinity for small values of P_i. When the empirical value P_i is close or equal to 0, the MAAPE is bounded above by 1.57.

Table 6.5: Empirical frequency and calculated probability of execution of a bid order for several sizes of the spread S and initial values of the quantities at the best bid q_B and the best ask q_A on July 5th 2021. The model probability is computed using rates calculated on the five preceding days. Only instances where a combination of q_A and q_B occurred more than 100 times are taken into account to compute a reasonable empirical probability.

		Empirical Probability					Model Probability				
		q_A					q_A				
	q_B	1	2	3	4	5	1	2	3	4	5
$S=1$	1	2.0%	5.3%	7.1%	13.4%	-	3.0%	3.9%	4.6%	5.1%	5.5%
	2	0.6%	0.0%	-	-	-	1.9%	2.7%	3.2%	3.7%	4.0%
	3	0.3%	-	-	-	-	1.5%	2.2%	2.6%	3.0%	3.3%
	4	0.0%	-	-	-	-	1.2%	1.8%	2.2%	2.6%	2.9%
	5	0.0%	-	-	-	-	0.9%	1.5%	1.9%	2.2%	2.5%
$S=2$	1	1.3%	2.5%	4.1%	5.9%	5.9%	1.5%	1.8%	2.0%	2.1%	2.2%
	2	0.3%	0.1%	0.0%	-	-	1.0%	1.2%	1.3%	1.4%	1.5%
	3	0.3%	1.6%	-	-	-	0.7%	0.9%	1.0%	1.1%	1.2%
	4	0.2%	-	-	-	-	0.6%	0.8%	0.9%	1.0%	1.0%
	5	0.0%	0.0%	0.0%	-	-	0.5%	0.7%	0.8%	0.8%	0.9%
$S=3$	1	0.5%	0.9%	1.4%	1.1%	0.8%	1.2%	1.3%	1.4%	1.4%	1.4%
	2	0.1%	0.0%	0.3%	0.0%	0.0%	0.8%	0.9%	1.0%	1.0%	1.0%
	3	0.1%	0.1%	0.0%	0.0%	-	0.7%	0.7%	0.8%	0.8%	0.8%
	4	0.1%	0.0%	0.0%	-	-	0.5%	0.5%	0.5%	0.6%	0.7%
	5	0.1%	0.0%	-	-	-	0.5%	0.5%	0.6%	0.6%	0.6%
$S=4$	1	1.5%	0.5%	0.3%	0.5%	0.3%	1.1%	1.2%	1.2%	1.2%	1.2%
	2	0.0%	0.0%	0.2%	0.1%	0.0%	0.8%	0.8%	0.8%	0.8%	0.8%
	3	0.0%	0.1%	0.0%	0.0%	0.0%	0.6%	0.7%	0.7%	0.7%	0.7%
	4	0.1%	0.0%	0.0%	0.0%	0.0%	0.5%	0.6%	0.6%	0.6%	0.6%
	5	0.0%	0.0%	0.0%	0.0%	-	0.5%	0.5%	0.5%	0.5%	0.5%
$S=5$	1	0.0%	-	-	-	-	1.1%	1.1%	1.1%	1.1%	1.1%
	2	0.0%	-	-	-	-	0.6%	0.7%	0.7%	0.7%	0.7%
	3	-	-	-	-	-	0.5%	0.5%	0.5%	0.7%	0.5%
	4	-	-	-	-	-	0.3%	0.3%	0.4%	0.4%	0.4%
	5	-	-	-	-	-	0.3%	0.3%	0.3%	0.3%	0.3%

The MAAPE for each spread size S and combination of quantities at the best ask q_A and best bid q_B, along with the average empirical probability is given in Table 6.6 for the days between 29–6–2021 and 5–7–2021. We see that for each spread size, the forecast for the first row, i.e. for $q_B = 1$, appears to be reasonably accurate. For $q_B > 1$, the model consistently overestimates the fill probabilities for all days and spread sizes. We observe that for $q_B > 1$, the empirical probability of executing an order before the mid-price moves is very low, typically less than 0.5%. This also shows the challenge of accurately predicting fill probabilities for such cases. Regarding the cases when $q_B = 1$, the model seems to perform better for smaller spread sizes. Since the empirical probabilities are quite small, a relative error measure will always tend to increase quite rapidly. Considering the strong assumptions on the orders and order flow, we can conclude that for the case $q_B = 1$, the model seems to be able to capture the dynamics of the order book quite accurately. For the case $q_B > 1$, the average empirical fill probability is in most cases negligibly larger than 0%, and the model seems to consistently overestimate these values.

Table 6.6: Average empirical probability and mean arctangent absolute percentage error for computed fill probabilities for each spread size S from 29–6–2021 to 5–7–2021.

		Average Empirical Probability					MAAPE				
		q_A					q_A				
	q_B	1	2	3	4	5	1	2	3	4	5
$S=1$	1	2.1%	5.7%	7.5%	10.3%	9.2%	67%	15%	23%	36%	27%
	2	0.6%	0.5%	-	-	-	126%	139%	-	-	-
	3	0.3%	-	-	-	-	137%	-	-	-	-
	4	0.5%	-	-	-	-	103%	-	-	-	-
	5	0.3%	-	-	-	-	117%	-	-	-	-
$S=2$	1	1.4%	2.5%	4.8%	5.8%	8.1%	32%	21%	26%	52%	60%
	2	0.4%	0.4%	0.4%	0.3%	-	115%	121%	123%	135%	-
	3	1.0%	0.6%	-	-	-	101%	109%	-	-	-
	4	0.1%	0.0%	-	-	-	135%	157%	-	-	-
	5	0.1%	-	-	-	-	139%	-	-	-	-
$S=3$	1	1.0%	1.2%	1.7%	1.9%	2.1%	31%	28%	24%	24%	42%
	2	0.3%	0.2%	0.4%	0.7%	0.3%	112%	123%	99%	105%	105%
	3	0.2%	0.2%	0.2%	0.0%	-	128%	132%	119%	157%	-
	4	0.1%	0.1%	0.1%	0.0%	-	129%	132%	119%	157%	-
	5	0.1%	0.1%	-	-	-	139%	124%	-	-	-
$S=4$	1	0.9%	0.9%	0.8%	0.8%	0.7%	30%	55%	47%	69%	73%
	2	0.1%	0.2%	0.1%	0.2%	0.1%	135%	121%	132%	113%	128%
	3	0.1%	0.1%	0.1%	0.1%	0.1%	122%	134%	139%	145%	139%
	4	0.0%	0.0%	0.1%	0.0%	0.0%	150%	152%	150%	156%	157%
	5	0.0%	0.0%	0.0%	0.0%	0.0%	151%	152%	157%	157%	157%
$S=5$	1	1.0%	0.7%	0.6%	1.3%	0.1%	34%	47%	41%	46%	145%
	2	0.2%	0.1%	0.1%	0.0%	0.0%	114%	137%	139%	157%	157%
	3	0.0%	0.0%	0.1%	0.0%	0.0%	145%	138%	122%	153%	157%
	4	0.0%	0.0%	0.0%	0.0%	0.0%	157%	155%	157%	157%	157%
	5	0.1%	0.0%	0.0%	0.0%	0.0%	132%	154%	157%	157%	157%

6.5 Computation of Probability of Execution at One Price Level below the Best Bid Price

We recall the probability of executing an order at price level $p_B - 1$ from Section 4.3, given by

$$\mathbb{P}[T_{\text{bid}} < T_{\text{other}}] \cdot \sum_{j=1}^{A} \sum_{i=1}^{B} \Big(\mathbb{P}[\epsilon_{B-} < T \mid W_{B-}(T_{\text{bid}}) = i,\ X_A(T_{\text{bid}}) = j] \cdot \mathbb{P}[W_{B-}(T_{\text{bid}}) = i] \cdot \mathbb{P}[X_A(T_{\text{bid}}) = j] \Big). \tag{6.15}$$

Here, A denotes the possible number of orders outstanding at the best ask price after the best bid moved down, and B denotes the possible number of orders at price level $p_B - 1$. As mentioned in Section 4.3, B can only take values between 1 and $X_{B-}(0)$, where $X_{B-}(0)$ denotes the number of orders outstanding at price level $p_B - 1$ after submitting the order. We have $B = 1$ if all the orders in front of the submitted bid order were cancelled before the bid price moved down, and $B = X_{B-}$ if no order was cancelled before the bid price moved down. Recall that W_{B-} denotes the number of orders remaining at price level $p_B - 1$ of the original number of orders $X_{B-}(0)$. In Section 6.4 we saw that the empirical fill probability is negligible for $W_{B-} > 2$ for all days and spread sizes. Therefore, we restrict ourselves to the cases where W_{B-} is either 1 or 2 after the bid moved down, so we assume

$$\mathbb{P}[\epsilon_B < T \mid W_{B-}(T_{\text{bid}}) = i, X_A(T_{\text{bid}}) = j] = 0 \text{ for } i > 2. \tag{6.16}$$

We are also interested in finding an upper bound for A to ensure the tractability of the summation in Equation 6.15. The quantity at the best ask not only moves down, but can also move up due to the arrival of new limit orders. As a consequence, the quantity outstanding after a price move, given by A, could theoretically have infinitely many values. However, as mentioned in Section 4.3, A will have a finite number of possible values in reality. In this section we will focus on the cases where the spread equals one or two ticks when the order is submitted, since the model had the best performance for these spread sizes for estimating the fill probability in the previous section. Table 6.7 shows the empirical distribution for the orders outstanding at the ask after a downward move of the best bid price for these spread sizes. It becomes clear that in both cases, the probability that there are only one or two orders outstanding at the best ask is around 98%. For this reason, we will neglect the other possible values and restrict ourselves to $A = 2$, i.e. we assume $\mathbb{P}[X_A(T_{\text{bid}}) = j] = 0$ for $j > 2$. We will use the distribution as provided in Table 6.7 for the probability $\mathbb{P}[X_A(T_{\text{bid}}) = j]$ for $j = 1, 2$. This follows the approach of Cont and De Larrard, in which they draw the quantities at both the best bid price and the best ask price after a price move from a joint (empirical) distribution [11].

Table 6.7: Empirical distributions of the number of orders at the ask q_A after a downward move of the best bid price for spread size $S = 1, 2$ from 29–6–2021 to 5–7–2021.

	q_A					
S	1	2	3	4	5	>5
1	76.2%	22.0%	1.6%	0.0%	0.0%	0.2%
2	82.2%	15.2%	1.4%	0.0%	0.0%	0.6%

By the aforementioned assumptions, probability 6.15 now becomes

$$\mathbb{P}[T_{\text{bid}} < T_{\text{other}}]\cdot \sum_{j=1}^{2}\sum_{i=1}^{2}\Big(\mathbb{P}[\epsilon_{B-} < T \mid W_{B-}(T_{\text{bid}}) = i, X_A(T_{\text{bid}}) = j]\cdot\mathbb{P}[W_{B-}(T_{\text{bid}}) = i]\cdot\mathbb{P}[X_A(T_{\text{bid}}) = j]\Big). \tag{6.17}$$

Since the fill probability now also depends on $X_{B-}(0)$, the number of orders outstanding at price level $p_B - 1$ at $t = 0$, the dimension of the problem becomes quite large. We need to calculate the probability for each combination of q_B, q_A and q_{B-}. For simplicity, we therefore restrict ourselves to the case where $S = 1$ when the order is submitted, and we let q_B range from one to four orders, and q_A and q_{B-} from one to five orders. We will compare the computed probabilities to the empirical frequencies. Since each combination of quantities is quite specific, we will compute the empirical frequencies based on a longer time frame to obtain a sufficient amount of empirical values. We estimate the probabilities based on order book data from June 21 2021 to July 2 2021. Table 6.8 indicates that even for a data set based on two weeks of order book data, not all combinations of quantities occur more than 100 times. Therefore, we cannot compare all forecasts to an empirical probability, which makes it more difficult to assess the performance of the model.

We observe that in most instances the model seems to overestimate the fill probabilities. This could be a result of the magnitude of the probabilities that we are trying to compute, together with an accumulated error from the results of the fill probability at the best bid. In addition, there might be some order book dynamics that are not effectively described by the model. We can conclude that although the model shows some promising results, it needs to be further refined to be able to capture more dynamics, which could lead to a better estimation of the fill probability.

Table 6.8: Empirical probability and computed theoretical probability for orders placed at price level $p_B - 1$ for $S = 1$. The empirical frequencies are based on order book data from June 21 2021 to July 2 2021. Only combinations of q_B, q_A and q_{B-} that occurred more than 100 times are taken into account. The computed probabilities are calculated using the parameter set based on data from June 28 to July 2.

		Empirical Probability					Model Probability				
		q_A					q_A				
	q_B	1	2	3	4	5	1	2	3	4	5
$q_{B-} = 1$	1	0.19%	0.27%	0.68%	2.16%	-	0.75%	0.97%	1.09%	1.16%	1.21%
	2	0.23%	0.25%	0.39%	0.67%	-	0.53%	0.75%	0.88%	0.97%	1.03%
	3	0.11%	0.69%	-	-	-	0.42%	0.62%	0.75%	0.84%	0.91%
	4	0.12%	-	-	-	-	0.35%	0.54%	0.66%	0.75%	0.82%
$q_{B-} = 2$	1	0.06%	0.44%	-	-	-	0.63%	0.81%	0.91%	0.96%	1.00%
	2	0.12%	0.37%	0.00%	-	-	0.47%	0.67%	0.78%	0.86%	0.91%
	3	0.12%	0.00%	-	-	-	0.38%	0.57%	0.69%	0.77%	0.83%
	4	0.00%	-	-	-	-	0.32%	0.50%	0.61%	0.70%	0.76%
$q_{B-} = 3$	1	0.28%	1.12%	-	-	-	0.42%	0.54%	0.60%	0.64%	0.67%
	2	0.09%	0.22%	-	-	-	0.38%	0.53%	0.62%	0.68%	0.72%
	3	0.04%	-	-	-	-	0.33%	0.49%	0.59%	0.66%	0.71%
	4	0.00%	-	-	-	-	0.29%	0.44%	0.55%	0.62%	0.68%
$q_{B-} = 4$	1	0.06%	0.71%	-	-	-	0.33%	0.42%	0.47%	0.50%	0.53%
	2	0.08%	0.62%	-	-	-	0.32%	0.45%	0.52%	0.58%	0.61%
	3	0.00%	-	-	-	-	0.29%	0.43%	0.52%	0.58%	0.63%
	4	0.03%	-	-	-	-	0.26%	0.40%	0.49%	0.56%	0.61%
$q_{B-} = 5$	1	0.07%	0.00%	-	-	-	0.27%	0.35%	0.39%	0.42%	0.43%
	2	0.08%	0.89%	-	-	-	0.28%	0.39%	0.46%	0.50%	0.53%
	3	0.00%	-	-	-	-	0.26%	0.38%	0.46%	0.52%	0.56%
	4	0.03%	-	-	-	-	0.24%	0.36%	0.45%	0.51%	0.56%

7 | Conclusion and Discussion

The aim of this research was to evaluate the feasibility of estimating fill probabilities of limit orders on the foreign exchange spot market using a simplified stochastic model to model the order book dynamics. Accurately predicting fill probabilities is a key factor in the field of algorithmic trading, since it can lead to more effective trading strategies and significantly reduce transactions cost. The model we used in this thesis describes the order book dynamics by modelling the queues at different price levels as birth-death processes. Here, the arrivals of limit orders lead to an increase in quantity at a certain price level, which can be seen as 'births'. Conversely, the 'deaths' are a result of incoming cancellations and market orders, as they cause a decrease in quantity. The density functions of the first-passage times of these queues to 0 can then be computed using Laplace transforms. By using the density functions, we derive a semi-analytical expression for several probabilities of interest, namely the probability of a mid-price increase, the probability of an order posted at the best bid being executed, and the probability of an order posted at one price level below the best bid being executed. These probabilities are all conditional on the state of the order book, i.e. conditional on the number of orders outstanding and the spread size. The model used in this thesis was based on a stochastic model as proposed by Cont et al., and we extended this model by incorporating order arrival rates that depend on the spread size [12]. This means that the rates at each price level change for different spread sizes, which leads to a more realistic depiction of the order book dynamics. Furthermore, we provided a general expression to analytically compute not only the fill probabilities at the best quotes, but also for one price level below the best bid price and best ask price.

In Chapter 5, we conducted several tests to assess the validity of the model assumptions with respect to our data. We showed that symmetry of the order flow is a reasonable assumption, as the rates did not differ significantly for orders placed on the bid and ask side of the order book. We confirmed these results by employing the Wilcoxon-signed rank test to the arrival rates. On the other hand, we found that the assumption of unit-size orders is not supported by the data. Nevertheless, this assumption does allow for a semi-analytical computation of the probabilities of interest. In addition, we discovered that the arrival rates of incoming orders and cancellations differ substantially for different spread sizes. For this reason, we decided to increase the dimensionality of the arrival rate parameters to incorporate the spread size.

To see how well the model captures the one-step transition probabilities (i.e. the probability that the quantity of the queue increases at the next event at a certain price level) we compared the theoretical probabilities with the empirical frequencies in Chapter 6. We observed that in most cases, the model is able to capture the dynamics reasonably well. For the one-step transition probabilities at the best quotes there exists some discrepancy between the empirical and theoretical values, indicating the presence of dynamics that the model does not effectively replicate.

Additionally, we presented the numerical results of the probabilities of interest, namely the probability that the mid-price increases at the next change and the probability of executing an order. First, we computed the probability of a mid-price increase using four different parameter sets. These parameter sets were calculated on data of five preceding days and data of the previous four instances of the same day to predict the rates for the day of interest. We used both spread-dependent rates and spread-independent rates and compared their performance. Our analysis showed some success, particularly for predictions that incorporated spread-dependent rates, which tended to slightly outperform predictions using rates independent of the spread size. It is important to note, however, that none of the parameter sets was found to clearly outperform the others. More research could be done to determine which time frame would yield the best prediction, for example by incorporating a decay parameter, which would give higher weight to the rates of more recent days. We computed the probabilities of an increase for five days for spread sizes ranging from 1 to 5, and for different combinations of quantities outstanding at the best bid price and best ask price. We compared the probability forecasts to the empirical frequencies of a mid-price increase, and we found that the model exhibits reasonable accuracy for most cases. This suggests its ability to capture short-term mid-price dynamics adequately.

To asses the performance of the prediction of execution probability at the best bid price, again we computed the probability for multiple days, considering various spread sizes and combinations of quantities at the best bid price and best ask price. We compared the predicted values to the empirical ones and we observed certain limitations in the model's capacity to fully capture all dynamics in the order book. For instance, while the model performed relatively well for cases when the submitted order has the highest priority, discrepancies arose for orders with lower priority, which highlights a potential area for further study. This was mainly the result of the magnitude of the empirical probabilities, which were often around 0.5% or lower. The model consistently overestimated the fill probabilities for orders that do not have the highest priority of execution. Additionally, for larger spread sizes, the predicted fill probability did not vary significantly for different values of the quantity at the best ask price, indicating possible shortcomings of the model in replicating more complex market dynamics. This observation offers another promising topic for future research to improve the predictive capacity of the model.

Finally, we computed the probabilities of orders placed at one price level below the best bid price. Since we found that the empirical probabilities of order execution at the best bid were close to 0 for orders that do not have a high priority, we decided to disregard these in the computation of orders posted at the price level below the best bid. We computed the probabilities for the case that the spread size is equal to one tick, for different combinations of quantities at the best bid price, best ask price and one level below the best bid price. Since these combinations are very specific, we were not able to compute an empirical probability for each instance, even when using two weeks of order book data. For the empirical values we were able to compute, the model again seemed to overestimate the probabilities. This could be a result of the model's inability to capture all market dynamics, but it could also be an effect of the accumulation of errors for the fill probability at the best bid price. Similar to the execution probability at the best price level, it is difficult to asses the performance of the model due to the small magnitude of the probabilities of interest.

An obvious limitation of this model is that it makes strong assumptions on the characteristics of the orders and the order flow, such as unit order size and independence of event arrivals. The model could be further refined to capture the subtleties of market dynamics more effectively. Future work could also focus on investigating more sophisticated parameter sets for the arrival rates or integrating additional market factors that could possible improve the predictive accuracy, such as seasonality effects. The assumptions of unit order sizes could be relaxed by incorporating multiple order sizes. We showed in Chapter 5 that the assumption that orders have only one size is not supported by the data, but the set of regularly occurring order sizes is relatively small. For this reason, incorporating a small set of possible order sizes could result in significantly more realistic model. Additionally, a more realistic order flow could potentially lead to better results regarding the order book probabilities. This could, for example, be done by modelling the order flow using a Hawkes-process, in which the arrival of an event leads to a temporary increase of the same (self exciting) or different (mutually exciting) events. Another limitation of this model is that the fill probability is conditional on the mid-price not moving. This condition is needed to ensure that the fill probability can be expressed analytically. In reality, however, a large part of orders will be filled after the mid-price has moved, and is therefore disregarded by the model. In order to incorporate these orders, the whole dynamics of the order book would have to be modelled. It remains to be seen whether this is possible without affecting the analytical tractability of the model. Finally, it would be interesting to see how the model would perform for different currency pairs or asset classes.

In summary, the model shows some promising results that could be used as a foundation in the ongoing process of improving and developing comprehensive models to forecast financial markets, and in particular the fill probability of limit orders.

Bibliography

[1] Joseph Abate and Ward Whitt. "Computing Laplace Transforms for Numerical Inversion Via Continued Fractions." In: *INFORMS Journal on Computing* 11.4 (1999), pp. 394–405. DOI: https://doi.org/10.1287/ijoc.11.4.394.

[2] Joseph Abate and Ward Whitt. "Numerical inversion of Laplace transforms of probability distributions." In: *ORSA Journal on Computing* 7.1 (1995), pp. 36–43. DOI: https://doi.org/10.1287/ijoc.7.1.36.

[3] Joseph Abate and Ward Whitt. "The Fourier-series method for inverting transforms of probability distributions." In: *Queueing systems* 10 (1992), pp. 5–87. DOI: https://doi.org/10.1007/BF01158520.

[4] *AFM Market Watch Edition 8 - Algorithmic Trading.* Tech. rep. The Dutch Authority for the Financial Markets (AFM), (2023). URL: https://www.afm.nl/en/sector/themas/beurzen-en-effecten/afm-market-watch.

[5] Alvaro Arroyo, Alvaro Cartea, Fernando Moreno-Pino, and Stefan Zohren. "Deep Attentive Survival Analysis in Limit Order Books: Estimating Fill Probabilities with Convolutional-Transformers". (2023). DOI: https://doi.org/10.48550/arXiv.2306.05479.

[6] Bank for International Settlements. *OTC foreign exchange turnover in April 2022.* Tech. rep. (2022). URL: https://www.bis.org/statistics/rpfx22_fx.pdf.

[7] Alvaro Cartea, Sebastian Jaimungal, and José Penalva. *Algorithmic and High-Frequency Trading.* Cambridge University Press, (2015). ISBN: 978-1-107-09114-6.

[8] Álvaro Cartea and Sebastian Jaimungal. "Optimal execution with limit and market orders". In: *Quantitative Finance* 15.8 (2015), pp. 1279–1291. ISSN: 1469-7688. DOI: 10.1080/14697688.2015.1032543.

[9] Jin-Wan Cho and Edward Nelling. "The Probability of Limit-Order Execution". In: *Financial Analyst Journal* 56.5 (2000), pp. 28–33. DOI: https://doi.org/10.2469/faj.v56.n5.2387.

[10] Rama Cont. "Statistical modeling of high-frequency financial data". In: *IEEE Signal Processing Magazine* 28.5 (2011), pp. 16–25. DOI: https://dx.doi.org/10.2139/ssrn.1748022.

[11] Rama Cont and Adrien de Larrard. "Price Dynamics in a Markovian Limit Order Market". In: *SIAM Journal on Financial Mathematics* 4.1 (2013), pp. 1–25. DOI: 10.1137/110856605.

[12] Rama Cont, Sasha Stoikov, and Rishi Talreja. "A Stochastic Model for Order Book Dynamics". In: *Operations Research* 58.3 (2010), pp. 549–563. ISSN: 0030-364X. DOI: 10.1287/opre.1090.0780.

[13] Forrest W. Crawford and Marc A. Suchard. "Transition probabilities for general birth–death processes with applications in ecology, genetics, and evolution". In: *Journal of Mathematical Biology* 65.3 (2012), pp. 553–580. ISSN: 0303-6812. DOI: 10.1007/s00285-011-0471-z.

[14] Fang Fang and Cees W. Oosterlee. "A Novel Pricing Method for European Options Based on Fourier-Cosine Series Expansions". In: *SIAM Journal on Scientific Computing* 31.2 (2009), pp. 826–848. ISSN: 1064-8275. DOI: 10.1137/080718061.

[15] Olivier Guéant, Charles-Albert Lehalle, and Joaquin Fernandez-Tapia. "Optimal Portfolio Liquidation with Limit Orders". In: *SIAM Journal on Financial Mathematics* 3.1 (2012), pp. 740–764. ISSN: 1945-497X. DOI: 10.1137/110850475.

[16] He Huang and Alec N. Kercheval. "A generalized birth–death stochastic model for high-frequency order book dynamics". In: *Quantitative Finance* 12.4 (2012), pp. 547–557. ISSN: 14697688. DOI: 10.1080/14697688.2012.664926.

[17] Sungil Kim and Heeyoung Kim. "A new metric of absolute percentage error for intermittent demand forecasts". In: *International Journal of Forecasting* 32.3 (2016), pp. 669–679. ISSN: 0169-2070. DOI: 10.1016/J.IJFORECAST.2015.12.003.

[18] Charles-Albert Lehalle, Othmane Mounjid, and Mathieu Rosenbaum. "Optimal liquidity-based trading tactics". In: *Stochastic Systems* 11.4 (2018), pp. 368–390. DOI: https://doi.org/10.1287/stsy.2021.0078.

[19] Andrew W. Lo, A. Craig MacKinlay, and June Zhang. "Econometric models of limit-order executions". In: *Journal of Financial Economics* 65.1 (2002), pp. 31–71. ISSN: 0304-405X. DOI: 10.1016/S0304-405X(02)00134-4.

[20] Lisa Lorentzen and Haakon Waadeland. *Continued Fractions*. 2nd ed. Vol. 1. Atlantis Press, (2008). ISBN: 978-90-78677-07-9.

[21] Costis Maglaras, Ciamac C. Moallemi, and Muye Wang. "A deep learning approach to estimating fill probabilities in a limit order book". In: *Quantitative Finance* 22.11 (2022), pp. 1989–2003. ISSN: 14697696. DOI: 10.1080/14697688.2022.2124189.

[22] William H. Press and Saul A. Teukolsky. "Evaluating Continued Fractions and Computing Exponential Integrals". In: *Computers in Physics* 2.5 (1988), pp. 88–89. ISSN: 08941866. DOI: 10.1063/1.4822777.

[23] Eric Smith, J. Doyne Farmer, László Gillemot, and Supriya Krishnamurthy. "Statistical theory of the continuous double auction Statistical theory of the continuous double auction". In: *Quantitative Finance* 3.6 (2003), pp. 1–36. ISSN: 1469-7696. DOI: 10.1088/1469-7688/3/6/307.

[24] David A. Swanson. "On the Relationship among Values of the same Summary Measure of Error when used across Multiple Characteristics at the same point in time: An Examination of MALPE and MAPE". In: *Review of Economics and Finance* 5.1 (2015), pp. 1–14. URL: https://escholarship.org/uc/item/1f71t3x9.

[25] Ian J. Thompson and A.R. Barnett. "Coulomb and Bessel functions of complex arguments and order". In: *Journal of Computational Physics* 64.2 (1986), pp. 490–509. ISSN: 00219991. DOI: 10.1016/0021-9991(86)90046-X.

[26] Frank Wilcoxon. "Individual Comparisons by Ranking Methods". In: *Biometrics Bulletin* 1.6 (1945), p. 80. ISSN: 00994987. DOI: 10.2307/3001968.

A | Proofs

Lemma 4.2

Let f and F be the pdf and cdf of a random variable X, respectively. The Laplace transform $\hat{F}$ of the cdf F is given by:

$$\hat{F}(s) = \frac{1}{s}\hat{f}(s), \tag{A.1}$$

where $\hat{f}(s)$ is the Laplace transform of the pdf f.

Proof. We have $F' = f$ (a.s.),

$$\hat{F}(s) = \int_{-\infty}^{\infty} e^{-st}F(t)\mathrm{d}t \quad \text{and} \quad \hat{F}'(s) = \hat{f}(s) = \int_{-\infty}^{\infty} e^{-st}f(t)\mathrm{d}t.$$

By integration by parts, we have:

$$\hat{f}(s) = \hat{F}'(s) = \int_{-\infty}^{\infty} e^{-st}F'(t)\mathrm{d}t \tag{A.2}$$

$$= \left[e^{-st}F(t)\right]_{-\infty}^{\infty} + s\int_{-\infty}^{\infty} e^{-st}F(t)\mathrm{d}t \tag{A.3}$$

$$= s\int_{-\infty}^{\infty} e^{-st}F(t)\mathrm{d}t = s\hat{F}(s). \tag{A.4}$$

□

Lemma 4.3

Let Z be an exponentially distributed random variable with parameter Λ, then the Laplace transform of $\sigma_B \wedge Z$ is given by:

$$\hat{f}_b^1(\Lambda + s) + \frac{\Lambda}{\Lambda + s}(1 - \hat{f}_b^1(\Lambda + s)),$$

where $\hat{f}_b^1$ is given in 4.2.

Proof. Z is exponentially distributed with rate Λ, so for all $t \geq 0$, we have:

$$\begin{aligned}\mathbb{P}[\sigma_B \wedge Z < t] &= 1 - \mathbb{P}[\sigma_B > t]\mathbb{P}[Z > t] \\ &= 1 - (1 - F_{\sigma_B}(t))e^{-\Lambda t}.\end{aligned}$$

From this follows that $f_{\sigma_B \wedge Z}(t)$ is given by:

$$\begin{aligned}f_{\sigma_B \wedge Z}(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\Big(1 - (1 - F_{\sigma_B}^1(t))e^{-\Lambda t}\Big) \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\Big(1 - e^{-\Lambda t} + e^{-\Lambda t}F_{\sigma_B}^1(t)\Big) \\ &= \Lambda e^{-\Lambda t} + e^{-\Lambda t}f_{\sigma_B}^1(t) - \Lambda F_{\sigma_B}^1(t)e^{-\Lambda t} \\ &= e^{-\Lambda t}\Big(f_{\sigma_B}^1(t) + \Lambda(1 - F_{\sigma_B}^1(t))\Big),\end{aligned} \tag{A.5}$$

for $t \geq 0$. Here, $f^1_{\sigma_B}$ and $F^1_{\sigma_B}$ are the pdf and cdf of σ_B, respectively. The Laplace transform is then given by:

$$\begin{aligned}
\hat{f}^1_{\sigma_B \wedge Z}(s) &= \int_{-\infty}^{\infty} e^{-st} f_{\sigma_B \wedge Z}(t)\mathrm{d}t \\
&= \int_0^{\infty} e^{-st} e^{-\Lambda t}\Big(f^1_b(t) + \Lambda(1 - F^1_b(t))\Big)\mathrm{d}t \\
&= \int_0^{\infty} e^{-(s+\Lambda)t} f^1_b(t)\mathrm{d}t + \Lambda \int_0^{\infty} e^{-(s+\Lambda)t}(1 - F^1_b(t))\mathrm{d}t \\
&= \hat{f}^1_b(s+\Lambda) + \Lambda \int_0^{\infty} e^{-(s+\Lambda)t}(1 - F^1_b(t))\mathrm{d}t.
\end{aligned}$$

By integration by parts we have for the second part of the last equality:

$$\begin{aligned}
\Lambda \int_0^{\infty} e^{-(s+\Lambda)t}(1 - F^1_b(t))\mathrm{d}t &= \Lambda\Bigg(\Big[(1 - F^1_b(t)) \cdot -\frac{1}{s+\Lambda} e^{-(s+\Lambda)t}\Big]_0^{\infty} - \frac{1}{s+\Lambda}\int_0^{\infty} e^{-(s+\Lambda)t} f^1_b(t)\mathrm{d}t\Bigg) \\
&= \Lambda\Bigg(\frac{1}{s+\Lambda} - \frac{1}{s+\Lambda}\hat{f}^1_b(s+\Lambda)\Bigg) \\
&= \frac{\Lambda}{s+\Lambda}\big(1 - \hat{f}^1_b(s+\Lambda)\big).
\end{aligned}$$

Combining these, we obtain:

$$\hat{f}^1_{\sigma_B \wedge Z}(s) = \hat{f}^1_b(s+\Lambda) + \frac{\Lambda}{s+\Lambda}\big(1 - \hat{f}^1_b(s+\Lambda)\big). \tag{A.6}$$

□

B | Parameters

B.1 Limit Orders

Table B.1: Arrival rate per secondof limit orders for each spread size S in ticks for each distance δ in ticks from the opposite best quote.

	Distance in ticks to the opposite best quote														
S	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	13.27	34.72	59.54	66.5	69.78	50.61	34.53	23.19	17.61	16.93	13.04	9.25	7.36	5.91	5.02
2	1.27	7.75	20.22	26.18	24.48	22.09	15.53	10.90	7.14	6.6	5.19	3.92	3.19	2.72	2.23
3	0.08	1.09	5.73	11.60	13.77	11.44	8.15	4.94	2.94	2.35	2.4	1.72	1.66	1.40	1.18
4	0.02	0.11	2.27	7.17	10.54	10.43	7.52	5.30	3.61	2.04	2.03	1.40	1.43	1.22	1.09
5	0.02	0.05	0.30	5.48	11.84	13.96	12.57	9.62	7.47	5.89	3.42	3.24	2.34	2.26	1.87

(a) 7-6-2021 - 11-6-2021.

	Distance in ticks to the opposite best quote														
S	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	32.88	73.20	144.32	169.15	196.80	139.58	107.18	61.20	52.17	44.57	33.57	22.95	17.92	14.34	12.54
2	2.45	14.79	38.31	57.27	55.74	57.31	39.16	29.01	16.88	16.51	12.14	8.73	6.71	5.82	5.13
3	0.12	1.43	8.51	19.31	23.6	22.97	16.94	11.53	5.67	5.52	4.95	3.29	2.83	2.38	2.24
4	0.02	0.10	1.88	7.97	13.06	13.47	11.28	7.82	5.75	3.08	3.06	2.02	1.66	1.33	1.24
5	0.01	0.04	0.21	5.56	11.93	15.14	14.41	10.96	8.99	6.97	3.49	2.85	1.74	1.50	1.13

(b) 14-6-2021 - 18-6-2021.

	Distance in ticks to the opposite best quote														
S	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	6.60	16.07	29.68	34.59	34.96	26.07	19.36	11.87	11.52	8.30	5.95	3.98	2.77	2.22	1.98
2	1.74	9.32	24.18	38.28	36.22	32.48	22.00	15.04	11.95	10.15	6.80	4.61	3.47	2.68	2.18
3	0.10	1.44	7.64	18.51	22.60	20.79	14.60	8.95	5.46	5.27	3.78	2.35	1.99	1.57	1.41
4	0.02	0.10	2.37	8.17	15.00	15.14	12.36	7.57	5.23	3.79	2.89	1.65	1.36	1.11	1.04
5	0.01	0.03	0.16	4.88	9.88	12.59	12.26	8.36	5.82	5.23	3.15	1.96	1.13	1.04	0.87

(c) 21-6-2021 - 25-6-2021.

	Distance in ticks to the opposite best quote														
S	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	5.64	12.94	21.38	23.3	24.85	18.64	15.3	8.59	8.78	6.04	4.3	3.00	1.98	1.47	1.21
2	1.31	5.36	11.26	17.21	16.07	13.55	10.03	7.34	4.89	3.86	2.84	2.00	1.49	1.01	0.81
3	0.20	2.21	6.95	13.62	17.02	16.99	10.19	8.05	4.54	3.30	2.91	1.86	1.62	1.16	0.95
4	0.08	0.14	3.79	9.02	15.12	16.39	12.82	8.66	6.94	3.27	2.91	1.93	1.79	1.34	1.05
5	0.04	0.05	0.25	6.50	11.7	14.46	14.05	9.62	8.02	6.23	2.89	2.32	1.85	1.80	1.17

(d) 28-6-2021 - 2-7-2021.

Table B.2: Arrival rate per second of limit orders for each distance δ in ticks from the opposite best quote.

	Distance in ticks to the opposite best quote														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Week 1	0.23	1.14	4.81	10.15	12.99	12.01	8.84	6.02	4.04	2.92	2.67	1.94	1.80	1.54	1.32
Week 2	0.17	0.72	3.57	10.54	15.70	16.42	13.54	9.59	6.68	4.61	3.74	2.58	2.01	1.66	1.49
Week 3	0.19	0.78	3.76	10.47	16.05	16.24	13.13	8.34	5.71	4.69	3.33	2.01	1.54	1.27	1.13
Week 4	0.73	2.65	7.11	12.89	16.31	16.18	11.74	8.47	6.00	3.89	3.08	2.08	1.75	1.31	1.03

B.2 Market Orders

Table B.3: Arrival rate per second of market orders.

Week 1	0.10
Week 2	0.10
Week 3	0.12
Week 4	0.16

Table B.4: Arrival rate per second of market orders for each spread size S in ticks.

	Side		
S (in ticks)	Sell	Buy	Total
1	0.21	0.27	0.48
2	0.09	0.08	0.17
3	0.05	0.04	0.09
4	0.04	0.03	0.07
5	0.04	0.03	0.07

(a) 7-6-2021 - 11-6-2021.

	Side		
S (in ticks)	Sell	Buy	Total
1	0.24	0.30	0.54
2	0.11	0.11	0.22
3	0.05	0.05	0.09
4	0.04	0.04	0.07
5	0.04	0.03	0.07

(b) 14-6-2021 - 18-6-2021.

	Side		
S (in ticks)	Sell	Buy	Total
1	0.27	0.27	0.54
2	0.11	0.10	0.22
3	0.05	0.05	0.11
4	0.05	0.05	0.09
5	0.04	0.05	0.09

(c) 21-6-2021 - 25-6-2021.

	Side		
S (in ticks)	Sell	Buy	Total
1	0.16	0.17	0.34
2	0.10	0.09	0.19
3	0.07	0.06	0.13
4	0.06	0.05	0.11
5	0.08	0.06	0.13

(d) 28-6-2021 - 2-7-2021.

B.3 Cancellations

Table B.5: Arrival rate per second of cancellations for each spread size S in ticks for each distance δ in ticks from the opposite best quote.

	Distance in ticks to the opposite best quote														
S	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	43.95	41.11	53.86	60.57	55.10	52.11	41.01	31.82	21.99	14.99	13.33	10.39	7.98	6.72	5.97
2	-	18.77	19.63	25.98	25.68	20.98	16.12	9.99	7.40	5.65	5.08	4.05	3.47	2.87	2.35
3	-	-	9.25	11.27	12.65	10.22	7.21	5.79	4.35	3.06	2.14	1.87	1.61	1.49	1.17
4	-	-	-	7.91	10.76	10.46	7.50	4.88	3.20	2.40	1.69	1.62	1.30	1.29	1.08
5	-	-	-	-	12.72	14.92	13.44	9.69	6.51	4.71	3.74	2.63	2.46	2.16	1.91

(a) 7-6-2021 - 11-6-2021.

	Distance in ticks to the opposite best quote														
S	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	107.95	99.02	125.40	153.33	141.69	140.91	123.33	92.04	59.04	40.40	34.88	26.99	20.54	17.72	14.99
2	-	40.24	39.18	53.06	60.38	46.46	43.74	27.51	20.30	13.41	11.69	9.56	7.29	6.50	5.70
3	-	-	17.72	18.78	21.79	18.72	14.91	13.40	10.06	5.87	4.58	3.75	2.76	2.59	2.31
4	-	-	-	9.96	13.01	13.57	10.20	7.78	5.26	4.10	2.67	2.32	1.55	1.39	1.25
5	-	-	-	-	12.23	15.50	16.15	10.96	9.06	4.41	3.73	2.58	1.92	1.33	1.23

(b) 14-6-2021 - 18-6-2021.

	Distance in ticks to the opposite best quote														
S	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	19.41	20.44	26.32	30.5	29.63	26.57	21.48	16.25	10.89	8.04	5.96	4.58	3.17	2.67	2.35
2	-	24.63	27.37	33.8	37.12	29.62	24.74	16.36	11.41	8.27	6.53	5.08	3.64	3.08	2.57
3	-	-	15.70	18.26	20.50	18.26	13.29	10.48	8.42	5.15	3.60	2.66	1.91	1.77	1.47
4	-	-	-	11.47	14.19	14.81	11.28	7.88	5.19	3.98	2.81	1.96	1.26	1.17	1.07
5	-	-	-	-	10.06	12.89	12.47	9.22	5.93	3.71	2.90	1.91	1.26	0.92	0.94

(c) 21-6-2021 - 25-6-2021.

	Distance in ticks to the opposite best quote														
S	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	7.70	9.26	12.64	14.10	13.14	10.84	9.65	7.90	4.61	3.81	2.74	1.99	1.41	1.03	0.82
2	-	8.97	13.03	16.08	18.42	12.63	10.89	6.31	4.19	3.76	2.62	1.98	1.43	1.12	0.86
3	-	-	10.12	14.11	16.21	14.11	10.44	8.63	6.23	3.67	2.66	2.14	1.57	1.39	0.99
4	-	-	-	11.45	14.72	16.47	12.26	9.11	5.34	4.36	2.59	2.16	1.58	1.40	1.14
5	-	-	-	-	11.57	14.84	15.32	10.40	7.37	4.14	3.47	2.16	1.96	1.32	1.24

(d) 28-6-2021 - 2-7-2021.

Table B.6: Arrival rate per second of cancellations for each distance δ in ticks from the opposite best quote.

	Distance in ticks to the opposite best quote														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Week 1	0.28	1.27	5.02	9.97	12.64	11.57	8.59	6.18	4.37	3.22	2.38	2.08	1.74	1.61	1.32
Week 2	0.21	0.86	3.91	10.54	15.33	15.63	12.96	9.93	7.22	4.77	3.44	2.81	1.98	1.72	1.54
Week 3	0.22	0.89	4.06	10.96	15.17	15.56	12.45	9.07	6.28	4.37	3.17	2.24	1.51	1.33	1.20
Week 4	0.74	2.61	7.32	13.13	16.18	14.98	11.91	8.85	5.85	4.19	2.88	2.26	1.68	1.40	1.10

B.4 Average Outstanding Quantity

Table B.7: Average quantity outstanding for each spread size S in ticks at each distance δ in ticks from the opposite best quote.

	Distance in ticks to the opposite best quote									
S	1	2	3	4	5	6	7	8	9	10
1	881009	1276706	2211293	2896346	3217561	2929171	2517662	2308628	2249335	2167686
2	-	987704	1915617	2954285	3395169	3358261	2852408	2303484	2295585	2248940
3	-	-	1320495	2720061	3596108	3786409	3222420	2466837	2122430	2223011
4	-	-	-	2105311	3215664	3759231	3567034	2909708	2393342	2111617
5	-	-	-	-	2668738	3221010	3483012	3076086	2669657	2305588

(a) 7-6-2021 - 11-6-2021.

	Distance in ticks to the opposite best quote									
S	1	2	3	4	5	6	7	8	9	10
1	937201	1178273	1980725	2666660	3180881	2948543	2664931	2441223	2320286	2252619
2	-	1009250	1704020	2768598	3173786	3461039	2937258	2393703	2391245	2334668
3	-	-	1197928	2554048	3338846	3776215	3391919	2535485	2204979	2345263
4	-	-	-	1896755	3125428	3548816	3753704	3079049	2450451	2188514
5	-	-	-	-	2753656	3140941	3523565	3450564	2877668	2358425

(b) 14-6-2021 - 18-6-2021.

	Distance in ticks to the opposite best quote									
S	1	2	3	4	5	6	7	8	9	10
1	946203	1216197	2019834	2611302	3050296	2880843	2631826	2374942	2322966	2255550
2	-	997756	1719178	2794963	3157827	3280082	2792516	2341025	2363071	2339150
3	-	-	1171872	2613741	3363508	3637893	3205178	2435920	2196744	2334744
4	-	-	-	1854415	3172715	3551835	3582193	2887714	2356096	2218379
5	-	-	-	-	2702744	3184181	3524908	3297543	2756338	2309784

(c) 21-6-2021 - 25-6-2021.

	Distance in ticks to the opposite best quote									
S	1	2	3	4	5	6	7	8	9	10
1	934078	1314196	2159477	2748123	3227522	3055098	2559423	2289031	2384276	2310791
2	-	1026808	1875125	2831906	3238382	3495244	2892744	2329283	2320882	2374705
3	-	-	1333256	2503048	3230359	3827122	3280678	2481327	2180672	2251154
4	-	-	-	1884665	2911031	3398257	3621024	2989333	2423047	2050902
5	-	-	-	-	2410589	2832350	3143166	3185080	2769294	2174400

(d) 28-6-2021 - 2-7-2021.

B.5 COS-Method

Table B.8: Integration range $[a, b]$ and value of N used in the COS-method to compute the probability of a mid-price increase for both the arrival rate parameter set computed on the five preceding days and four previous instances of the same day.

	Preceding days		Previous instances of the same day	
S	N	$[a, b]$	N	$[a, b]$
1	20	[-1.5,1.5]	20	[-1.0,1.0]
2	20	[-1.5,1.5]	20	[-1.0,1.0]
3	20	[-2.5,2.5]	20	[-2.5,2.5]
4	20	[-3.0,3.0]	20	[-3.5,3.5]
5	20	[-2.0,2.0]	20	[-2.0,2.0]

Table B.9: Integration range $[a, b]$ and and value of N used in the COS-method to compute the fill probability.

S	N	$[a, b]$
1	75	[-35,35]
2	100	[-100,100]
3	150	[-160,160]
4	150	[-280,280]
5	150	[-320,320]

C | Python Codes

```python
import time
import numpy as np
import matplotlib.pyplot as plt
import math
from scipy.special import comb

def COSDensity(f_hat, t, N, a, b):
    ##COS method to compute the pdf of first-passage times
    i = complex(0.0, 1.0)

    # F_k coefficients
    k_values = np.arange(N)
    u_values = k_values * np.pi / (b - a)

    f_hat_values = np.array([f_hat(-i * u) for u in u_values])
    F_k = (2.0 / (b - a)) * np.real(f_hat_values * np.exp(-i * u_values * a)
          * np.cos(u_values * (t - a)))
    F_k[0] *= 0.5  # adjustment for the first term

    return F_k.sum()

def pdf_cosmethod(LT, N, a, b,step_size, plot: bool):
    ##Compute values of the pdf using COS-method
    start = time.time()
    t_vals = np.arange(a, b, step_size)
    pdf_vals = np.vectorize(COSDensity)(LT, t_vals, N, a, b)
    end = time.time()
    elapsed = end - start
    print("Elapsed time:", elapsed, "seconds")
    if plot:
        plt.plot(t_vals, pdf_vals)
        plt.show()
    return pdf_vals

def cdf(values, a, b,step_size, plot: bool):
    ##Compute cdf values
    t_vals = np.arange(a, b, step_size)
    cdf_vals = np.cumsum(values) * step_size
    if plot:
        plt.plot(t_vals, cdf_vals)
        plt.show()
    return cdf_vals
```

```
def pdf_euler(LT, t, m, n, A, num_steps, plot=False):
    ##Compute values of the pdf using Euler method
    step_size = t / (num_steps - 1)
    t_vals = np.linspace(0, t+step_size, num_steps)
    values = []
    start = time.time()
    for i in range(1,num_steps+1):
        values.append(Euler(LT, step_size*i, A, m, n))
    end = time.time()
    elapsed= end - start
    print("Elapsed time:", elapsed, "seconds")
    if plot:
        plt.plot(t_vals, values)
        plt.show()
    return values

def sn(f_hat,t,N,A):
    ##S_n terms in Euler method
    i = complex(0.0,1.0)
    F_k = []

     # F_k coefficients
    for k in range(0,N):
        u = (A+2*k*np.pi*i) / (2*t)
        F_k.append(np.exp(A/2) / t * (-1)**k * np.real(f_hat(u)))

    F_k[0] *= 0.5 # adjustment for the first term

    return sum(F_k)

def Euler(f_hat,t,A,m,n):
    ##Computation of the pdf using Euler method
    som = 0
    for k in range(0,m+1):
        s = sn(f_hat,t,n+k,A)
        som += comb(m,k)/2**m*s
    return som

def a_term(n,x,S,params):
    ## a-terms are the numerators in the continued fraction
    lamda, mu, theta = params
    if n - x == 0:
        a = mu+x*theta[S-1]
    else:
        a = -lamda[S-1]*(mu+n*theta[S-1])
    return a

def b_term(x,s,S,params):
    ## b-terms are the denominators in the continued fraction
    lamda, mu, theta = params
    b = lamda[S-1]+mu+x*theta[S-1]+s
    return b
```

```
def Lentz_approximant(x,s,S,params,eps):
    #modified lentz algorithm to determine the value of the continued fraction
    f = 1e-30
    C = f
    D = 0
    i = 0
    error = 1
    while error>eps:
        D = b_term(x+i,s,S,params)+a_term(x+i,x,S,params)*D
        if D == 0:
            D = 1e-30
        C = b_term(x+i,s,S,params)+a_term(x+i,x,S,params)/C
        if C ==0:
            C = 1e-30
        D = 1/D
        f *= C*D
        error = abs(D*C-1)
        i=i+1
    return f

def laplace_transform_PM(s,S,a,b,params,eps):
    ##Compute Laplace transform for the price moves
    lamda, mu, theta = params
    if S==1:
        CF_a = [Lentz_approximant(j, s, S, params, eps) for j in range(1, int(a +
                                                        1))]
        f_hat_a = np.prod(CF_a)
        CF_b = [Lentz_approximant(j, -s, S, params, eps) for j in range(1, int(b
                                                        + 1))]
        f_hat_b = np.prod(CF_b)
        f_hat = f_hat_a*f_hat_b
    else:
        L = np.sum(lamda[0:S-1])
        CF_a = [Lentz_approximant(j, L+s, S, params, eps) for j in range(1, int(a
                                                        + 1))]
        f_hat_a = np.prod(CF_a)
        CF_b = [Lentz_approximant(j, L-s, S, params, eps) for j in range(1, int(b
                                                        + 1))]
        f_hat_b = np.prod(CF_b)
        f_hat = (f_hat_a+(L/(L+s))*(1-f_hat_a))*(f_hat_b+(L/(L-s))*(1-f_hat_b))
    return f_hat

def laplace_transform_execution(s,S,a,b,params,eps):
    ##Compute Laplace transforms for the execution probabilities
    lamda, mu, theta = params
    if S==1:
        CF_a = [Lentz_approximant(j, -s, S, params, eps) for j in range(1, int(a
                                                        + 1))]
        f_hat_a = np.prod(CF_a)
        g = [(mu + (j-1)*theta[S-1])/(mu + (j-1)*theta[S-1] + s) for j in range(1
                                                        , int(b + 1))]
        g_hat = np.prod(g)
        f_hat = f_hat_a*g_hat
    else:
        L = np.sum(lamda[0:S-1])
        CF_a = [Lentz_approximant(j,2*L-s,S,params,eps) for j in range(1,int(a +
                                                        1))]
        f_hat_a = np.prod(CF_a)
        g = [(mu + (j-1)*theta[S-1])/(mu + (j-1)*theta[S-1] + s) for j in range(1
                                                        , int(b + 1))]
        g_hat = np.prod(g)
        f_hat = (f_hat_a+(2*L/(2*L-s))*(1-f_hat_a))*g_hat
    return f_hat
```

```
def laplace_transform_order_outstanding(s,S,b,q,i,params,eps):
    ##Compute Laplace transforms for the probability of number of orders
                                          outstanding at the new best bid/
                                          ask after a price move down
    lamda, mu, theta = params
    if i!=q:
        g = [((j-1)*theta[S])/((j-1)*theta[S] + s) for j in range(i+1, int(q+1))]
        g_hat = np.prod(g)
        CF = [Lentz_approximant(j, -s, S, params, eps) for j in range(1, int(b +
                                          1))]
        f_hat = np.prod(CF)
    else:
        g_hat = ((i-1)*theta[S])/((i-1)*theta[S] - s)
        CF = [Lentz_approximant(j, s, S, params, eps) for j in range(1, int(b + 1
                                          ))]
        f_hat = np.prod(CF)
    return g_hat*f_hat

def LT_bid_down(s,S,a,b,params,eps):
    ##Compute Laplace transform for the midprice moves as a result of the bid
                                          moving down
    lamda, mu, theta = params
    if S==1:
        CF_a = [Lentz_approximant(j, -s, S, params, eps) for j in range(1, int(a
                                          + 1))]
        f_hat_a = np.prod(CF_a)
        CF_b = [Lentz_approximant(j, s, S, params, eps) for j in range(1, int(b +
                                          1))]
        f_hat_b = np.prod(CF_b)
        f_hat = f_hat_a*f_hat_b
    else:
        L = np.sum(lamda[0:S-1])
        CF_a = [Lentz_approximant(j, 2*L-s, S, params, eps) for j in range(1, int
                                          (a + 1))]
        f_hat_a = np.prod(CF_a)
        CF_b = [Lentz_approximant(j, s, S, params, eps) for j in range(1, int(b +
                                          1))]
        f_hat_b = np.prod(CF_b)
        f_hat = (f_hat_a+(2*L/(2*L-s))*(1-f_hat_a))*f_hat_b
    return f_hat

##Compute cumulants
def mgf(LT, t,S,a,b,params,eps):
    return LT(-t, S, a, b, params, eps)

def cgf(LT, t, S, a, b, params, eps):
    return np.log(mgf(LT, t, S, a, b, params, eps))

def nth_derivative(n, LT, func, t_val, S, a, b, params, eps, h=1e-5):
    if n == 0:
        return func(LT, t_val, S, a, b, params, eps)
    else:
        return (nth_derivative(n-1, LT, func, t_val + h, S, a, b, params, eps) -
                                          nth_derivative(n-1, LT, func,
                                          t_val - h, S, a, b, params,
                                          eps)) / (2 * h)

def compute_cumulant(order, LT, S, a, b, params, eps):
    derivative = nth_derivative(order, LT, cgf, 0, S, a, b, params, eps)
    return derivative / np.math.factorial(order)
```