

#### Delft University of Technology Faculty of Electrical Engineering, Mathematics and Computer Science Delft Institute of Applied Mathematics

MSc thesis Applied Mathematics

Stochastic Modeling of Order Book Dynamics

by Clemencia de Jesus Ladeira de Abreu

Delft, The Netherlands August 24, 2010

 $\bigodot$  2010 by Clemencia de J.L. de Abreu. All rights reserved.



#### MSc thesis Applied Mathematics Delft University of Technology

#### "Stochastic Modeling of Order Book Dynamics"

[Abstract] In this project the order book model proposed by Cont et al. [10] is used as a starting point to model order book dynamics. This model nicely combines three desirable properties from earlier studies: it is easy to calibrate, it reproduces statistical properties of the order book and it allows to make analytical computations in the order book. The model is studied, calibrated and tested on real-time data from the London Stock Exchange. Possible improvements to the model are discussed and tested. A method to compute probabilities in the model will be presented: recovering densities by inverting continued fraction representations of Laplace transforms. This is also implemented and evaluated.

#### Supervisor

Dr. J. A. M. van der Weide

#### Other board of assessment members

Prof.dr.ir. M.H. Vellekoop (UVA)

Prof.dr. C.W. Oosterlee

Dr. G. Luculli (All Options)

Prof.dr. F.M. Dekking

August 24, 2010

Delft

© 2010 by Clemencia de J.L. de Abreu. All rights reserved.

#### Preface

This document presents an overview of my work at All Options in Amsterdam, where I carried out my thesis research. All Options is an option trading company established in 1998 and is a leading liquidity provider on all major European derivative markets. All Options' core business is trading options on listed financial instruments. This project was supervised by The Derivatives Technology Foundation (TDTF). TDTF was founded in 2000, and is sponsored by All Options. The foundation has the objective to stimulate co-operation between academics and people who actually work with derivatives. The motivation behind being that there remains a considerable gap between how specialists in the academic and practical worlds think and communicate. Discussions between academics and practitioners result in projects, as for example this thesis, that are of value to all concerned. The eventual result is relevant from an academic perspective, whilst at the same time offering a solution to problems encountered in everyday practice.

After spending an In-house day at the old Amsterdam Stock Exchange where All Options is situated, I became very motivated to do an internship at this company, and see how every day life is working as a quant at the forefront of the financial markets. Therefore I decided to apply for a research project at All Options. This report presents the results of this thesis research, and is the final assessment to complete the Master Programme in Applied Mathematics at the Delft University of Technology. The goal of this project was to develop and test a model that computes probable behavior in the order book. I accomplished to do this in 6 months, and am proud to present my results and conclusions here. This report describes every step that I took in reaching my conclusions, and is written for people who do have a certain basis in Mathematics.

#### Acknowledgment

First of all I would like to express my appreciation to Hans van der Weide, from whom I have had my very first lectures in Financial Mathematics years ago and who has ever since been an important motivation and patient teacher. Next to that I would like to thank Michel Vellekoop who was my supervisor on behalf of All Options and TDTF for his advice, comments and pleasant collaboration. Gabriele Luculli I want to thank for his great efforts to compensate for the absence of the quant force by putting countless hours in reading and understanding my work, and brainstorming new ideas. Sing, Imran, Ariyanto, Sebastiaan, Rob and Geeske, they thought me most of what I know now about order books and trading and gave valuable feedback. Of course not to forget my intern-buddy Michiel for being there to discuss the elections and the world cup while waiting for Matlab to finish. Lastly I would like to thank All Options for giving me the opportunity to experience the excitement of working on a trading floor and facilitating me in carrying out this research.

All of this would not have been possible without TU Delft and the department of Applied Mathematics, where I have followed my education and learned all that I know today about Mathematics. However, I believe that it takes much more than just attending lectures and exams to become a 'Delftse Ingenieur'. And for this I especially want to thank ORAS, CH, EESTEC and TU Delft, who gave me the opportunity to learn so many things and gain so much valuable experience that I would have lacked else. Next to everyone mentioned so far, it is still Sofia, Frank, Igor, Tina and of course Jasper who have been most important in my day-to-day life over the past years. And last but not least, João and Oralia because in the end it is all thanks to them.

Clemencia de Abreu clemencia@ch.tudelft.nl Amsterdam, January 12, 2011

# Contents

	Preface	5						
	Acknowledgment	5						
	An Order Book Model	0						
1		9 11						
1	1.1 What is a Limit Order Book	11						
	1.2 Positioning of this Project	13						
	1.2 Positioning of this Project	14						
2	A Stochastic Order Book Model	15						
-	2.1 Dynamics of the Limit Order Book	15						
	22 Parameter Estimation	17						
	2.3 Direct Simulation of the Order Book	21						
3	Discussion of the Order Book Model	25						
•	3.1 Points of Improvement	25						
	3.2 Test Results	27						
	3.3 Final Order Book Model	30						
П	Computing Probabilities	35						
4	Approximating Laplace Transforms	37						
	4.1 Basic Concepts of Continued Fractions	37						
	4.2 Laplace Transforms via Continued Fractions	39						
	4.3 First Passage Times in Birth-Death Processes	40						
5	Probabilities in the Order Book	43						
	5.1 Probabilities of Influence on Trading Decisions	43						
	5.2 Move of the Mid-Price	45						
	5.3 Probability of Order Execution	47						
	5.4 Probability of Making the Spread	49						
	5.5 Probability of Reaching the Stop Loss	49						
6	Implementation	51						
	6.1 Computing Continued Fractions	51						
	6.2 Numerical Inversion of Laplace Transforms	52						
	6.3 Results	55						
7	Conclusion and Recommendations	59						
	7.1 Summary $\ldots$	59						
	7.2 Recommendations for Further Research	59						
	7.3 Conclusion	61						
Appendix								
	Bibliography	65						

# Part I An Order Book Model

### Chapter 1

# Introduction

Quantitative research on financial markets is not only driven by the human desire to capture the dynamics behind the global markets. Many of the theoretical developments in mathematics have found immediate application in financial markets and these markets also offer an amazing source of detailed data that enable to perform analysis. Here we study the dynamics of the order book, which is the smallest level of description of financial markets. The study of the order book is very interesting both from an academical and a practical point of view. It provides information about price formation dynamics, while for traders who participate in the markets the expected merits of possible trading strategies are computed based on the dynamics of the order book.

#### 1.1 What is a Limit Order Book

Trading can be viewed as a search problem. Buyers and sellers need to find each other to establish a trade. Buyers want to find sellers who are willing to trade the quantities they desire at a price which is as low as possible. Vice versa, sellers want to find buyers who are willing to trade the desired quantities at a price which is as high as possible. Stock exchanges usually organize markets so that everyone who wants to trade gathers at the same place, so that traders can easily find the best prices. Once, markets were exclusively organized on physical trading floors. Nowadays this is done by Electronic Communication Networks (ECN's) that allow participants to trade remotely. There are several types of participants in the markets. Traders who can estimate prices using theoretical models buy when prices better than their estimated price are available in the market. Their buying and selling pushes the prices in the market up and down, causing the market to reflect their estimated values. The whole dynamics of price movements caused by buying and selling constitutes what we know as the financial markets.

Research on market dynamics traditionally focused on quote-driven markets. In a quote-driven market, only buy and sell offers of centralized market makers are displayed. These market makers will post the buy and sell price at which they are willing to trade. Even though individual orders are not seen in a quote-driven market, the market maker will either fill an order from its own inventory or match you with another order. An advantage of this type of market is its liquidity, as the market makers are required to trade at their quoted prices. The major drawback of the quote-driven market is that it does not show transparency in the market.

In recent years, quote-driven markets have lost a significant share of order flow to ECN's, or 'order driven markets'. In such platforms, all orders are submitted to the electronic trading system, where all orders are listed stating their size and price. This is called the limit order book and it is accessible to all market participants. Exchanges operating in this fashion are NYSE, Euronext, Deutsche Börse, Nasdaq, the Tokyo Stock Exchange and the London Stock Exchange. In the limit order book, all outstanding orders both of buyers and sellers are displayed. This provides transparency in the market. The mechanical nature of order execution makes it interesting to stochastically model order-driven markets. We will therefore only focus on this type of markets in this study.

A participant in an order-driven market can place two basic types of orders, a limit and a market order. A limit order is an order to trade a certain quantity of an asset at a given price. A market order is an order to trade a certain quantity at the best price available in the limit order book. The best (lowest) available price at which one can buy an asset is the ask price, the best (highest) price at which one can sell is called the bid price. A market order is always executed, while a limit order either stays in the order book until it is either matched to an incoming market order, canceled or immediately matched to an outstanding order in the book. There are several other order types which we will not take into account here. These order types can theoretically be represented by limit and market orders.

In Figure 1.1 an example is shown of an order book for Vodafone stocks, at three consecutive time steps. This example only shows five levels of the order book, for a liquid stock such as Vodafone up to 30 levels may be available. The two columns in the middle show the available buy prices and sell prices, the two outer columns show the corresponding quantities. In the example we see that at the second time, the ask order at price 1.44500 disappeared so it has been canceled. At the third time the number of limit orders at price 1.45000 has decreased by 170000, which means that this number of contracts has been canceled. Imagine that an order at the best price disappears. In such a case it can be either canceled or executed against an incoming market order. A trader knows whether a cancellation or a trade took place from the 'Last' trade shown in the order book.

```
BOOK: LSE VODE
                         (ID: 25568)
Time: 08:00:00.691357
Last: 1.437000/0
TVol: 0
  BidQty
           Bid
                                  AskQty
                       Ask
    3231
           1.43350
                       1.43750
                                  207672
                       1.43900
                                  232794
     646
           1.43100
  193672
            1.42900
                       1.44200
                                  7689
  187377
            1.42700
                       1.44500
                                  1500
      60
            1.42250
                       1.44900
                                  2000
BOOK: LSE VODE
                         (ID: 25568)
Time: 08:00:00.691556
Last: 1.437000/0
TVol: 0
  BidQty
           Bid
                       Ask
                                  AskQty
    3231
           1.43350
                       1.43750
                                  207672
     646
           1.43100
                       1.43900
                                  232794
  193672
           1.42900
                       1.44200
                                 7689
  187377
            1.42700
                       1.44900
                                  2000
      60
            1.42250
                       1.45000
                                  2433364
BOOK: LSE VODE
                         (ID: 25568)
Time: 08:00:00.691626
Last: 1.437000/0
TVol: 0
  BidQty
           Bid
                       Ask
                                  AskQty
            1.43350
                       1.43750
                                  207672
    3231
     646
            1.43100
                       1.43900
                                  232794
  193672
            1.42900
                       1.44200
                                  7689
  187377
            1.42700
                       1.44900
                                  2000
      60
            1.42250
                       1.45000
                                 2263364
```

Figure 1.1: Example of three consecutive snap shots of the order book for Vodafone stocks, level 5

The above order book is aggregated which means that any displayed order may consist of multiple orders from different traders at the same price. The large volumes and small difference between the highest bid price and lowest ask price show that we are dealing with a liquid stock.

#### 1.2 Positioning of this Project

Various studies have focused on order books so far. Some have focused on reproducing as many of the statistical features observed in order books as possible and others have concentrated on creating a model that is useful for analysis. Some examples on research on order books are the following.

Parlour [22], Foucault et al. [14], Rosu[24] and Biais and Weill [7] all propose equilibrium models in which every trader explicitly takes into account that his order affects all future orders. In such models, the fact that both the current state of the order book and the expected order flow affect the strategies of traders results in a theoretical equilibrium of the order book from which its dynamics can be deduced. Equilibrium models provide insights in the price formation process, but are difficult to estimate, and therefore difficult to use in applications. Examples of empirical studies of the statistical properties of the order book are Bouchaud et al. [8], Hollifield [15], Bovier et al. [9] and Smith et al. [26]. They have focused on average properties and expectations of the order book without taking into account the information provided by the current order book state. However, players in an order market can choose to submit a limit order or a market order, depending on the current state of the limit order book. Every order will affect the placing of subsequent orders and consequently its own execution probability.

Consider for example a trader who badly wants to sell a certain number of contracts. If he knows that it is very likely for the price to go up, he might decide to submit a limit order at a slightly higher price than the current bid, instead of immediately submitting a market order. Therefore, the motivation for modeling order book dynamics is to use all the information provided by the current state of the order book, to predict its short term behavior. Conditional on the current state of the order book, this short term behavior includes for example the probability that a limit order is executed before the mid price moves. This information is of interest for traders because it can be used in designing trading strategies. We will go deeper into this in Chapter 5.



Figure 1.2: Schematic representation of the problem

Figure 1.2 shows a schematic representation of the problem. Studies that are focused on reproducing the statistical features of the order book can be placed in the lower left square and correspond to the arrow emanating from there. Studies such as equilibrium models are focused on the balance between the trading strategy and the order book. The difficulty in modeling an order book is to create a model that both reproduces its empirically observed statistical properties and is suitable for quantitative analysis. In this study we have tried to combine both of these, taking the model proposed by Rama Cont et al. [10] as a starting point. Using this model, we have attempted to go around the circle by establishing an order book model that reproduces as many of the statistical features as possible. Using this simplified representation of the reality, we will compute probabilities in the order book. These probabilities can be used in the trading strategy that interacts with the order book. A chosen trading strategy has again an impact on the order book, which closes the circle.

#### 1.3 Outline

This report is organized as follows. In Chapter 2 the model for the dynamics of the order book introduced by Cont et al. [10] will be explained and calibrated on data taken from the London Stock Exchange. Using direct simulation, the average behavior of a simulated order book will be compared to observed average behavior in the real order book. Based on these results, Chapter 3 proposes possible improvements to the model. These are again used to calibrate the model and using direct simulation average properties are compared to empirical observations. In this chapter, also a final order book model will be proposed.

A technique that can be used to compute probabilities in our model, based on Laplace transforms, will be introduced in Chapter 4. This second approach has the advantage that it can compute full distributions of random behavior in the order book. In Chapter 5 we will start by discussing interesting probabilities in the order book. It will be shown how these probabilities can be computed based on the Laplace technique. Implementation and results will be discussed in Chapter 6. We will conclude with remarks about the chosen approach and recommendations for further research.

### Chapter 2

## A Stochastic Order Book Model

In this chapter a model will be established for the dynamics of the order book. We start by introducing the model proposed by Cont et al. [10]. Reasoning that the dynamics of a limit order book resembles that of a queuing system, it is proposed that the number of limit orders per price level should be modeled as a time-continuous Markov process. Desirable properties of this model are that it can be easily estimated from high-frequency data, it satisfies empirical order book features and it is analytically tractable. Later on it will be shown that the model is in fact simple enough to use Laplace transforms to compute various probabilities, conditional on the state of the order book.

#### 2.1 Dynamics of the Limit Order Book

In the model proposed in [10], it is assumed that orders can be placed on a price grid  $\{1, ..., n\}$  that represents the possible prices known as price ticks. As the time frame of the analysis chosen is in milliseconds or seconds, an upper boundary n can be set so that it is highly unlikely that orders at prices higher than price n will be placed. The number of orders at every price level in the order book is modeled by the time-continuous process  $(X_t^{(1)}, ..., X_t^{(n)})_{t\geq 0}$ , where  $X^{(p)} < 0$  means that there are  $|X^{(p)}|$  bid orders at price p, and  $X^{(p)} > 0$  indicates that there are  $|X^{(p)}|$  ask orders at p.

The ask price, or simply the ask  $p_A(t)$  is defined as

$$p_A(t) = \inf \{ p = 1, ..., n; X^{(p)} > 0 \} \land (n+1),$$
(2.1)

and the bid price or bid  $p_B(t)$  is defined as

$$p_B(t) = \sup \left\{ p = 1, ..., n; X^{(p)} < 0 \right\} \lor 0.$$
(2.2)

Note that if there is no ask price available, n + 1 is taken as ask. If no bid price is available, we take 0 as bid. Furthermore the mid-price  $p_M(t)$  and the bid-ask spread  $p_S(t)$  are defined as

$$p_M(t) = \frac{p_B(t) + p_A(t)}{2}$$
 and  $p_S(t) = p_A(t) - p_B(t)$ . (2.3)

**Example 2.1.** Consider the order book snap shot:

```
BOOK: EXAMPLE
Time: 08:00:00.691626
  BidQty
            Bid
                    Ask
                            AskQty
     100
            1.50
                    1.60
                            20
      25
            1.45
                    1.65
                            400
            1.40
                    1.70
                            150
     350
      88
            1.30
                    1.75
                            200
      60
            1.25
                    1.90
                            90
```

where the tick size is 0.05. Then the price grid is  $\{1.25, ..., 1.90\}$  with steps of 0.05 and the corresponding time-continuous process for t = 0.691626 sec. is given by

$$X_t = (-60, -80, 0, -350, -25, -100, 0, 20, 400, 150, 200, 0, 0, 90),$$
(2.4)

with  $p_A = 1.60$ ,  $p_B = 1.50$ ,  $p_S = 0.10$  and  $p_M = 1.55$ .  $\diamondsuit$ 

Now assume that all incoming and outgoing orders are of size one. We can then describe the change in the order book by inflow of the following possible events:

- A limit buy order at price  $p < p_A(t)$  decreases the quantity at price  $p: X_{t+\Delta t}^{(p)} = X_t^{(p)} 1$
- A limit sell order at price  $p > p_B(t)$  increases the quantity at price  $p: X_{t+\Delta t}^{(p)} = X_t^{(p)} + 1$
- A market buy order decreases the quantity at the ask price:  $X_{t+\Delta t}^{(p_A)} = X_t^{(p_A)} 1$
- A market sell order increases the quantity at the bid price:  $X_{t+\Delta t}^{(p_B)} = X_t^{(p_B)} + 1$
- A cancellation of an outstanding limit buy order at price level  $p < p_A(t)$  increases the quantity at p:  $X_{t+\Delta t}^{(p)} = X_t^{(p)} + 1$
- A cancellation of an outstanding limit sell order at price level  $p > p_B(t)$  decreases the quantity at  $p: X_{t+\Delta t}^{(p)} = X_t^{(p)} 1$

Note that in this model a limit buy order at price  $p > p_A(t)$ , or a limit sell order at  $p < p_B(t)$  do not occur because in such cases it is more convenient to place a market order. Limit buy (sell) orders only exist at price levels lower (higher) than the most favorable opposite order, and can therefore only be canceled there. Furthermore, orders can be changed in practice. In this model such an order change is represented by a cancellation plus the submission of a new order.

So it is assumed that the order book can be represented as a queuing system, and its dynamics is driven by incoming limit orders, market orders, and cancellations. In the stochastic model proposed by Cont et al.[10], the events described above are modeled with independent exponentially distributed inter-arrival times. The resulting process X is a time-continuous Markov chain, with state space  $\mathbb{Z}^n$  and transition rates

$$\begin{split} X_{t+\Delta t}^{(p)} &= X_t^{(p)} - 1 \quad \text{with transition rate} \quad \alpha(p, x_t) \quad \text{for} \quad p < p_A(t), \\ X_{t+\Delta t}^{(p)} &= X_t^{(p)} + 1 \quad \text{with transition rate} \quad \alpha(p, x_t) \quad \text{for} \quad p > p_B(t), \\ X_{t+\Delta t}^{(p_A)} &= X_t^{(p_A)} - 1 \quad \text{with transition rate} \quad \beta(x_t), \\ X_{t+\Delta t}^{(p_B)} &= X_t^{(p_B)} + 1 \quad \text{with transition rate} \quad \beta(x_t), \\ X_{t+\Delta t}^{(p)} &= X_t^{(p)} + 1 \quad \text{with transition rate} \quad \gamma(p, x_t) \quad \text{for} \quad p < p_A(t), \\ X_{t+\Delta t}^{(p)} &= X_t^{(p)} - 1 \quad \text{with transition rate} \quad \gamma(p, x_t) \quad \text{for} \quad p > p_B(t), \end{split}$$

where  $x_t$  represents the observed order book status at time t. Market orders only arrive at the bid and ask, it is therefore not necessary to make  $\beta$  dependent on the price level p.

To make sure that the Markov chain satisfies the characteristics of a limit order book, we define the set of admissible states x as  $\mathcal{A} \subset \mathbb{Z}^n$  as

$$\mathcal{A} := \{ x \in \mathbb{Z}^n | \exists k \le l \in \mathbb{Z}_+ \text{ s.t. } : x^{(p)} \ge 0 \text{ for } p \ge l; x^{(p)} = 0 \text{ for } k (2.5)$$

The following proposition states that if the order book is initially in an admissible state, it will remain admissible almost surely.

**Proposition 2.2.** If  $X_0 \in A$ , then  $P(X_t \in A, \forall t \ge 0) = 1$ .

Proof. See Cont et al. [10].

Now we can model the limit order book as a Markov chain, where orders leave and arrive in the system after independent exponentially distributed inter arrival times. The parameters  $\beta(x_t)$ ,  $\alpha(p, x_t)$  and  $\gamma(p, x_t)$  may depend on the price level p and the prevailing configuration of the order book  $x_t$ . In the next section it will be shown how to estimate the parameters.

#### 2.2 Parameter Estimation

We will now turn our attention to the parameters in the model, about which Cont notices the following.

- It has been empirically observed by Bouchaud et al.[8] that the frequency of incoming limit orders is higher around the bid and ask prices. Therefore, the arrival rates of limit buy orders and limit sell orders are modeled exponentially with a rate that depends on the distance to the opposite best price (ask or bid). In [8] it is suggested to specify

$$\alpha(p, x_t) := \lambda(i) = \frac{k}{i^{\alpha}}, \tag{2.6}$$

where *i* represents the distance in ticks between p and the opposite best price (so for p on the bid side of the order book, *i* is the distance between p and  $p_A$ , and vice versa).

- Market buy and sell orders are assumed to arrive at independent, exponentially distributed times with rate  $\beta(x_t) = \mu$ .
- Limit orders at a distance of *i* price ticks from the best opposite quote, are canceled at a rate proportional to the number of outstanding orders  $x_t^{(p)}$  at price *p*, so

$$\gamma(p, x_t) := \theta(i) x_t^{(p)}. \tag{2.7}$$

Based on this, the following is suggested in [10] for the parameters.

$$\begin{split} X_{t+\Delta t}^{(p)} &= X_t^{(p)} - 1 \quad \text{with transition rate} \qquad \lambda(p_A(t) - p) & \text{ for } p < p_A(t), \\ X_{t+\Delta t}^{(p)} &= X_t^{(p)} + 1 \quad \text{with transition rate} \qquad \lambda(p - p_B(t)) & \text{ for } p > p_B(t), \\ X_{t+\Delta t}^{(p_A)} &= X_t^{(p_A)} - 1 \quad \text{with transition rate} \qquad \mu, \\ X_{t+\Delta t}^{(p_B)} &= X_t^{(p_B)} + 1 \quad \text{with transition rate} \qquad \mu, \\ X_{t+\Delta t}^{(p)} &= X_t^{(p)} + 1 \quad \text{with transition rate} \qquad \theta(p_A(t) - p) |x_t^{(p)}| & \text{ for } p < p_A(t), \\ X_{t+\Delta t}^{(p)} &= X_t^{(p)} - 1 \quad \text{with transition rate} \qquad \theta(p - p_B(t)) |x_t^{(p)}| & \text{ for } p > p_B(t), \end{split}$$

where  $x_t^{(p)}$  represents the observed number of outstanding orders at price level p.

The next step is to estimate these parameters from high-frequency data. The approach described in [10] is used here. The data we have to our disposal are from Vodafone stocks traded on the London Stock Exchange in January 2010 (16 trading days). The data set contains continuous order book data, which means that a snapshot of the order book is taken every time an event takes place. For Vodafone this corresponds to an average of 4 events per second. This means that per day over 120000 order book snapshots are available. An example of an order book snapshot with level 2 is shown in Figure 2.1. The data set used in this project has level 10. The tick-size of Vodafone stocks is  $\pounds 0.0005$ .

```
Timestamp (in seconds): 08:00:00.691357
          PriceBid1
                     SizeAsk1
                                           SizeBid2
SizeBid1
                                PriceAsk1
                                                      PriceBid2
                                                                 SizeAsk2
                                                                            PriceAsk2
6025
          1.4365
                     26827
                                1.437
                                            73781
                                                      1.436
                                                                  5141
                                                                            1.4375
PriceLast: 1.437
                    SizeLast: 26827
```

Figure 2.1: Order book snap shot of level 2 of Vodafone stocks traded on the London Stock Exchange.

Before trading starts at the London Stock Exchange, there is a Pre-Trading period of 1.5 hour during which participants are allowed to prepare for the opening of trading. Quotes and orders can be entered, changed or deleted. However, information about the ask, bid and their corresponding sizes is not available. There is also a Post Trading period during which orders may still be entered for the next trading day, but exercises

are no longer accepted. Because no actual trades take place, we will ignore the order book snap shots of the Pre-Trading and Post-Trading periods in our analysis.

In order to estimate the parameters in our model, we now proceed by extracting a list of events from the data set. This means that at every time we check which event took place (limit order, market order or cancellation), at which side of the book (bid or ask), and what the corresponding price and volume are. Because orders are assumed to be of unit size in our model, we first compute the average sizes of limit orders  $S_l$ , market orders  $S_m$  and cancellations  $S_c$ . The results for the Vodafone data are displayed in Table 2.2.

	$S_l$	$S_m$	$S_c$
Average	17401	11757	15863

Table 2.1: Average sizes of limit orders, market orders and cancellations for Vodafone stocks.

Next we compute the number of times that a market order is submitted  $N_m$ , a limit order is submitted  $N_l(i)$  or an order is canceled  $N_c(i)$  at distance *i* from the opposite best (see Table 2.2 and Figure 2.2).

i	1	2	3	4	5	6	7	8	9	10
$N_l(i)$	168434	270302	155195	97001	57428	44719	41023	21415	12672	8792
$N_c(i)$	252976	233402	169534	93442	54942	46192	43516	21875	13371	8062
$N_m$	18042									

Table 2.2: Numbers of limit orders, market orders and cancellations at distance i from the opposite best price.



Figure 2.2: left: Numbers of limit orders and cancellations at distance i from the opposite best price. right: Estimated parameters for Vodafone stocks.

Furthermore it is known that there are 8.5 trading hours per day, which corresponds to 30600 seconds. So the total trading time of our data set is T = 489600 seconds. Because the cancellation rate is assumed to be proportional to the number of outstanding orders at the particular price level, we need to know what the average number of orders N(i) is at a distance of i ticks from the opposite best quote. This is what is called the shape of the book. Because orders are assumed to be of unit size, the shape of the order book can be computed by

$$N(i) = \frac{S(i)}{S_l(i)},$$
(2.8)

where S(i) represents the average volume at i ticks from the opposite best quote. This is obtained by considering for every snap-shot what the volume is at distance i. Taking the average for every i over all the

snap-shots of the entire month gives S(i). The parameters can now be estimated as follows.

а

$$\begin{split} \hat{\lambda}(i) &=& \frac{N_l(i)}{T}, \\ \hat{\mu} &=& \frac{N_m}{T}\frac{S_m}{S_l}, \\ \text{nd} \quad \hat{\theta}(i) &=& \frac{N_c(i)}{TN(i)}\frac{S_c}{S_l} \end{split}$$

This approach is analogous to the one presented by Cont. Note that we have to multiply by the ratio between order sizes because we assumed all orders in the model to have the same (unit) size. Results of the parameter estimation for i = 1, ..., 12 are shown at the right hand side of Figure 2.2 and in Table 2.3.

i	1	2	3	4	5	6	7	8	9	10
$\hat{\lambda}(i)$	0.3238	0.5196	0.2983	0.1865	0.1104	0.0860	0.0789	0.0412	0.0244	0.0169
$\hat{\theta}(i)$	0.1108	0.0584	0.0330	0.0164	0.0088	0.0067	0.0069	0.0035	0.0023	0.0014
$\hat{\mu}$	0.0234									

Table 2.3: Estimated parameters for Vodafone stocks.

Instead of estimating the parameters based on the entire data set, we can also compute estimates per day. This shows how the estimates behave from day to day and gives an idea of how sensitive the estimates are. Computing the sample standard deviation of the estimates also allows to compute how many days of data are required in order to reach a certain level of confidence. In Figures 2.3-2.4 the parameters per daily data set are shown. It follows from the figures that the parameters vary more at the end of the month. The most extreme parameters correspond to January  $22^{nd}$ , which was the expiration day that month. This is the day on which options expire and thus market participants update and re-hedge their positions. It would therefore be a good idea to perform a statistical analysis on the influence of the trading day of the month on the parameters. The sample standard deviations corresponding to the daily estimates are displayed in Table 2.4.



Figure 2.3: Estimated  $\mu$  per day.



Figure 2.4: Estimated  $\lambda$  and  $\theta$  per day.

i	1	2	3	4	5	6	7	8	9	10
$\hat{\lambda}(i)$	0.0755	0.1352	0.0800	0.0593	0.0403	0.0320	0.0280	0.0121	0.0064	0.0051
$\hat{ heta}(i)$	0.0334	0.0189	0.0091	0.0059	0.0040	0.0039	0.0043	0.0019	0.0010	0.0006
Â	0.0063									

Table 2.4: Standard deviations corresponding to the estimated parameters.

The graphs in Figure 2.5 show that although the estimates may vary per day, the intensities move around constants and have highs at the end of the month (upper graphs) and the shape of the parameter-curve as a function of i is preserved (lower graphs). An explanation for the observed higher intensities at the end of the month can be found in the expiry of many related contracts at the end of the month. Unwinding positions and re-hedging portfolios results in higher volumes being traded. Note that the shape of the parameter curves are different for  $\lambda$  and  $\theta$ . Suggestions on functions to fit the parameter curves by are made by Zovko [28] and Bouchaud et al. [8]. In this project we will not further go into fitting parameter functions. The order book data available for this project is of sufficient depth to estimate parameters up to 30 ticks from the opposite best bid.



Figure 2.5: Comparison between estimated  $\lambda$  and  $\theta$  per day.

#### 2.3 Direct Simulation of the Order Book

In this section we will simulate an order book as simple as possible, based on the model assumptions introduced at the beginning of this chapter and the parameters estimated in the last section. With this simulation we hope to gain insight into the behavior and the average properties of the order book according to our model. Although this is not of interest to traders, it allows us to compare our model with empirically observed behavior and thus serves as a test of the model.

In the model introduced in Chapter 2, it is assumed that six different types of events – limit/market orders and cancellations on both the bid and the ask side – arrive in the order book. Depending on the type of the event, it may arrive at certain price levels p. The inter arrival times between two consecutive events of the same type at p are taken to be exponentially distributed with a parameter depending on the corresponding price level p. Arrivals of events of different types are independent of each other. Recall the arrival rates discussed in Section 2.2. Remember that the process  $X_t = (X_t^{(1)}, ..., X_t^{(n)})$  denotes the number of orders per price level 0 in the order book at time <math>t, and that it is a time-continuous Markov chain. Based on this information it is easy to simulate the behavior of a given order book over a certain number of events. Using Markov chain Monte Carlo it is also possible to determine the average shape of the order book. This shape was found to be universal by Bouchaud [8] after studying several liquid stocks.

To determine the average shape of a simulated order book consider the following. A Markov chain is said to be ergodic if all states are aperiodic and positive recurrent. Ergodicity (irreducibility in some texts) is a desirable feature for our Markov chain, since it is known that for ergodic chains a unique stationary distribution exists. This allows to compute long-term averages in the model and compare this to observed long-term behavior. From Cont et al. [10] we have the following proposition.

**Proposition 2.3.** If  $\min_{1 \le i \le n} \theta > 0$ , then  $X_t$  is an ergodic Markov process, and has a stationary distribution.

Proof. See Cont et al. [10]

A stationary distribution corresponds to the normalized average shape of the order book. Now let f denote a function on  $\mathcal{A}$  as defined in (2.5). Then for  $X_t \in \mathcal{A}$  the shape of the order book at time t is given by  $f(X_t)$ . Since  $X_t$  is an ergodic Markov chain, we can generate random elements from  $\mathcal{A}$  by simulating  $X_t$  for a long time. In the long run, the average output of the chain should follow the average shape of the book. An average or steady-state shape of the order book therefore corresponds to the expectation of  $f(X_t)$ .

$$E[f(X_{\infty})] \approx \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T} f(X_t).$$
(2.9)

More details about Markov chain Monte Carlo can be found in Madras [20].

In order to determine an average order book shape we now proceed as follows. Generate an admissible order book state with n price levels. By taking random samples from exponential distributions with the parameters given in Table 2.3, we can simulate arrival times. For all prices  $1 \le p < p_A$  generate the arrival times and cancellations of limit buy orders, for prices  $p_B generate the arrival and cancellation times of limit sell orders. For the arrival of market orders only two samples need to be generated because market orders have no price but they always influence the order book at price <math>p_A$  or  $p_B$ . Now choose the smallest arrival time out of the  $2(n + p_A - p_B)$  values generated. The corresponding event is the first event to take place in the order book at corresponding price level p: in the generated order book state the value at p either increases or decreases by one. To simulate another event, repeat the same procedure again, with the new order book as the initial book. If we repeat this simulation a large number of times, (2.9) allows us to approximate the steady-state shape of the book by taking the mean of all obtained X(t)'s.



Figure 2.6: Simulated order book shape based on the estimated parameters, compared to the observed shape.

If we apply the described procedure to a randomly generated order book for large numbers of events  $(l = 10^4, 10^5, 10^6)$ , we find the average states of the order book displayed in Figure 2.6. From this figure we can conclude that the model used indeed leads to a steady-state shape of the order book corresponding to the one observed from the order book data. The tail of the observed order book shape fluctuates heavily and shows a relatively large discrepancy with the simulated shape. This is not so surprising if we take into account that the arrival rates beyond 15 ticks from the opposite best are almost everywhere equal to zero. In the most important part of the order book (around the ask and bid) however, even up to 15 ticks away from the ask and the bid, the simulated order book approaches the observed shape very well. The hump observed around five ticks from the opposite best bid corresponds to empirical data. It can be explained by the presence of algorithmic traders in the market who, as part of their trading strategy, make small trades with a small gain, at a high frequency. In order to win trades, they are willing to quote at slightly better prices than those

available. Basically, the hump shows the position of the actual market, while the orders in front of it represent the market participants such as high-frequency traders, who are seeking liquidity and are therefore willing to take smaller margins of profit.

Another property that is interesting to compare is the volatility of the mid-price. Define the realized volatility of the mid-price over a day by

$$V(p_M) = \frac{1}{T} \sqrt{\sum_{i=1}^{n} \left( \log\left(\frac{P_M(t_{i+1})}{P_M(t_i)}\right) \right)^2},$$
(2.10)

where n is the number of snap-shots during the day,  $p_M(t_i)$  is the mid-price corresponding to the  $i^{th}$  snap-shot and T is the total trading time in seconds (= 30600 per day). In Table 2.5 you can see the volatility per day in our data set and the corresponding number of events n in the order book observed during that trading day.

After repeatedly simulating a model order book, we find a volatility of  $0.5247 \pm 0.0759$  with probability 95%. The corresponding average number of events per trading day is 208303. This is almost twice as much as observed in the real data. An additional idea would be to compute the total intensity of event arrivals from the parameters, and compare this to an observed order book. It is also interesting to see that in the observed order book an average of around 6000 trades per day is reached, while in our simulation this is around 1000. This leads us to conclude that these parameters do not approximate the dynamics of the order book well enough.

date	4-Jan	5-Jan	6-Jan	8-Jan	11-Jan	12-Jan	13-Jan	14-Jan
$V_{obs}(p_M)$	0.7799	0.3036	0.3805	0.2282	0.3137	0.2203	0.3429	0.4354
n	98135	92911	107119	116697	95332	109248	76197	99944
date	15-Jan	18-Jan	19-Jan	20-Jan	21-Jan	22-Jan	25-Jan	26-Jan
$V_{obs}(p_M)$	0.4237	0.2251	0.2117	0.1544	0.3533	0.1912	0.1605	0.2082
n	114154	70676	113791	141484	144459	179684	135421	150811

Table 2.5: Observed volatility of the mid-price compared to number of order book events per day.

### Chapter 3

### Discussion of the Order Book Model

Although the model introduced in the last chapter gives an almost perfect approximation of the order book shape, several points of doubt were formulated by traders. In this chapter we will discuss these and some possibilities to change the model and the parameters presented so far. We will follow Cont in assuming that the number of orders at any price level follows a birth-death process, because it enables us to analytically compute probabilities in the order book. Other small changes to the model and its parameters will be proposed and tested by direct simulation, similar as we did in last chapter.

#### 3.1 Points of Improvement

First of all let's take a close look at the left hand side of Figure 2.2. It shows the number of limit orders and cancellations versus the distance to the opposite best quote. It seems that the number of cancellations at i = 1 is excessively high. Remember that in Cont's model it was assumed that limit buy orders arrive at price levels  $p < p_A$  and limit sell orders arrive at price levels  $p > p_B$  with rate  $\alpha(p, x_t)$ . In reality however, limit orders may arrive at any price level. This has to do with the fact that different market participants have different speeds of trading. It might therefore happen that the best price in the market moves between the moments of submission and actual arrival in the market of a certain order. Limit orders arriving at better price levels than the opposite best are not displayed in the book but immediately executed against the best available price. Taking this into account, it results that a part of the observed 'cancellations' at i = 1 are in reality limit orders at i = 0. Figure 3.1 shows that this way of interpreting the observed order book is more as one would expect.



Figure 3.1: Numbers of limit orders and cancellations at distance i from the opposite best price.

As a result, define the transition rates in the order book process X as

- $$\begin{split} X^{(p)}_{t+\Delta t} &= X^{(p)}_t 1 \ \text{with transition rate} \quad \alpha(p,x_t) \quad \text{for} \ p \leq p_A(t), \\ X^{(p)}_{t+\Delta t} &= X^{(p)}_t + 1 \ \text{with transition rate} \quad \alpha(p,x_t) \quad \text{for} \ p \geq p_B(t), \end{split}$$
  (3.1)
- (3.2)
- $X_{t+\Delta t}^{(p_A)} = X_t^{(p_A)} 1 \ \, \text{with transition rate} \quad \beta(x_t),$ (3.3)
- $X_{t+\Delta t}^{(p_B)} = X_t^{(p_B)} + 1$  with transition rate (3.4)

$$X_{t+\Delta t}^{(p)} = X_t^{(p)} + 1 \quad \text{with transition rate} \quad \gamma(p, x_t) \quad \text{for} \quad p < p_A(t), \tag{3.5}$$

$$X_{t+\Delta t}^{(p)} = X_t^{(p)} - 1 \quad \text{with transition rate} \quad \gamma(p, x_t) \quad \text{for} \quad p > p_B(t), \tag{3.6}$$

where  $x_t$  represents the observed order book status at time t.

Second, consider the assumption of all orders having size one. This has been one of the points questioned by traders. Obviously, orders have very different sizes. An idea would therefore be to let the order size follow a certain distribution as was done by Bouchaud in [8]. According to Cont et al. however, the distribution of the size of incoming orders plays a minor role except in the far tail of the average order book. They thus claim that direct simulation of an order book with random order sizes gives similar results as taking orders to have constant size in the model. Based on this we keep the assumption of order sizes being one for now.

The third thing we will discuss here is the symmetry of the model. From the arrival rates formulated in Section 2.2 it is clear that the order book in our model is assumed to be symmetric. Except for the cancellation rates that depend on the volume at i, all the rates in the order book just depend on i and not on the side (ask or bid) of the book. Bouchaud [8] and Biais [6] observed that the average shape of the order book is symmetrical. It makes sense to let the parameters be equal for both sides of the book, else a directional trend in the price moves would result in the model. Intuitively however, one would expect the shape of the order book at every time t to be of more influence on the arrivals of new orders. The following examples clarify this statement.

**Example 3.1.** Consider the order book displayed at the left hand side.

BidQty	Bid	Ask	AskQty		BidQty	Bid	Ask	AskQty
5	1.50	1.55	1000		1000	1.50	1.55	1000
1000	1.43	1.56	1000		1000	1.49	1.56	1000
1000	1.42	1.57	1000		1000	1.48	1.57	1000
1000	1.41	1.58	1000		1000	1.47	1.58	1000

The best bid in this example has a relative small size, while there seems to be a 'gap' between the best bid and the second bid. This means that the bid side of the order book has a very irregular shape. According to Conts model, the arrival rates at price levels between  $\pounds 1.01 - \pounds 1.50$  in this example are equal to the arrival rates in the book at the right hand side. In reality however, one would expect the arrival rate of orders just above  $\pounds 1.03$  to be higher in the order book at the left hand side. $\diamondsuit$ 

One could derive from this example that once the order book shows an irregular shape, it has a high probability of returning to its average shape. This mean reverting behavior of the order book is also indicated by traders. To make the intensity of limit order arrivals mean reverting, we can make it depend on the distance between the current order book shape and the average order book shape for every i. We therefore propose to define  $\alpha(p, x_t) = \lambda(p_A(t) - p) \left(1 + \rho(\bar{x_i} - x_t^{(p)})\right)$ , where  $\bar{x_i}$  is the observed average number of quotes at i,  $x_t^{(p)}$  is the observed number of quotes at price level p at time t, and the parameter ho gives the strength of the mean-reversion.

**Example 3.2.** Consider another extreme order book (left hand side).

BidQty	Bid	Ask	AskQty	BidQty	Bid	Ask	AskQty
1000000	5.50	5.55	1000	1000	5.50	5.55	1000
150	5.49	5.56	1000	150	5.49	5.56	1000
1200	5.48	5.57	1000	1200	5.48	5.57	1000
700	5.47	5.58	1000	700	5.47	5.58	1000

Imagine that the huge order at the bid just arrived. According to the model of Cont, the orders at prices below  $\pounds 5.50$  are canceled at equal rates in both the left and the right hand side order book. In reality one would expect the orders below the huge order to be canceled at a higher rate.  $\diamond$ 

This example leads to the idea of making the cancellation rate proportional to not only the number of quotes at the corresponding price level, but also the sum of the numbers of quotes at better price levels. This brings us to the following order book model.

- Limit buy orders arrive at  $p \le p_A(t)$  with rate  $\lambda(p_A(t) p) \left(1 + \rho(\bar{x_i} x_t^{(p)})\right)$ , (3.7)
- $\text{Limit sell orders arrive at} \quad p \ge p_B(t) \quad \text{with rate} \quad \lambda(p p_B(t)) \left(1 + \rho(\bar{x_i} x_t^{(p)})\right), \qquad \textbf{(3.8)}$ 
  - Market buy orders arrive with rate  $\mu$ , (3.9)
  - Market sell orders arrive with rate  $\mu$ , (3.10)
- Limit buy orders are canceled at  $p < p_A(t)$  with rate  $\theta(p_A(t) p) \sum_{k=p}^{p_A} x_t^{(k)}$ , (3.11)

Limit sell orders are canceled at  $p > p_B(t)$  with rate  $\theta(p - p_B(t)) \sum_{k=p_B}^{p} x_t^{(k)}$ , (3.12)

where  $x_t^{(k)}$  represents the observed number of orders at price level k at time t and  $\bar{x_i}$  corresponds to the average number of orders observed at i ticks from the opposite best order. The number  $\bar{x_i}$  must be estimated from data.

#### 3.2 Test Results

In this section it will be tested whether the proposed changes in the model give a better approximation of the average behavior of the order book. Direct simulation will be used to test the influence on the average behavior. Again we start by computing the average sizes of limit orders, market orders and cancellations (Table 3.1), and the number of times that market orders, limit orders and cancellations arrive at distance *i* from the opposite best quote (Table 3.2). This time we take into account that limit orders may also arrive at  $p_B$  and  $p_A$ , in which case the limit order has the same influence on the order book as a market order. Furthermore, the time of the data set still is T = 489600 sec, and also the shape of the order book  $N(i) = \frac{S(i)}{S_l(i)}$  will be used again.

	$S_l$	$S_m$	$S_c$
Average	16822	14363	16463

Table 3.1: Average sizes of limit orders, market orders and cancellations for Vodafone stocks.

i	0	1	2	3	4	5	6	7	8	9
$N_l(i)$	57513	198153	243539	149533	96137	57381	44330	40798	21325	12898
$N_c(i)$	0	188483	225860	168617	93131	54823	46119	43489	21861	13354
$N_m$	39192									

Table 3.2: Numbers of limit orders, market orders and cancellations at distance i from the opposite best price.

Estimating the parameters in the same way as in Section 2.2 but now based on the information in Tables 3.1 and 3.2, we find the results shown in Table 3.3 and Figure 3.2.

Figure 3.3 shows the results for the simulated average shape of the order book using direct simulation. After repeatedly simulating we find 95%-confidence interval for the realized volatility of the mid-price per day of  $0.8076 \pm 0.1085$ . An average number of 217423 events per day were simulated. Althought the shape of the order book is equally good approximated by the model of Cont, this order book change will be adopted. The reason for this being that the assumption that limit orders cannot arrive at the best price is wrong. Next to that, the proposed change does not make the model any more complicated.



Figure 3.2: Estimated parameters corresponding to (3.1-3.6).

i	0	1	2	3	4	5	6	7	8	9
$\hat{\lambda}(i)$	0.1175	0.4047	0.4974	0.3054	0.1964	0.1172	0.0905	0.0833	0.0436	0.0263
$\hat{\theta}(i)$	0	0.0935	0.0586	0.0376	0.0177	0.0095	0.0077	0.0074	0.0039	0.0025
$\hat{\mu}$	0.0683									



Secondly, the cancellation parameter will be re-estimated to satisfy  $\gamma(p, x_t) = \theta(p_A(t) - p) \sum_{k=p}^{p_A} x_k(t)$ . Therefore we now divide by the cumulative sum of the average number of quotes at *i* ticks from the opposite best price. The parameter  $\theta$  can thus be estimated by

$$\hat{\theta}(i) = \frac{N_c(i)}{T\sum_{k=1}^{i} N(k)} \frac{S_c}{S_l}.$$
(3.13)

The results of this estimation are in Table 3.4 and Figure 3.4.

i	0	1	2	3	4	5	6	7	8	9
$\hat{\lambda}(i)$	0.1175	0.4047	0.4974	0.3054	0.1964	0.1172	0.0905	0.0833	0.0436	0.0263
$\hat{ heta}(i)$	0	0.0935	0.0385	0.0163	0.0060	0.0026	0.0017	0.0013	0.0006	0.0003
$\hat{\mu}$	0.0683									



Using direct simulation we find the 95%-confidence interval  $0.8195 \pm 0.0998$  for the realized volatility of the mid-price per day, and an average of 218988 events per trading day The average order book shape is displayed in Figure 3.5. This parameter estimation does not give a better approximation of the order book shape nor the volatility.

Thirdly we will check what the influence of the proposed arrival rate of limit orders is on the model. The parameter  $\lambda$  can again be estimated by

$$\hat{\lambda}(i) = \frac{N_l(i)}{T},\tag{3.14}$$

Note that the term  $(1 + \rho(\bar{x}_i - x_t^{(p)}))$  does not appear in the estimation of  $\lambda$  because its average equals one at every *i*. We can therefore use the parameters from Table 3.3. For the simulation we use  $\rho = 1/\bar{x}_i$ . By direct simulation, the 95%-confidence interval  $0.1110 \pm 0.0189$  is found, and an average of *y* events in the order book per day. The shape of the order book is shown in Figure 3.6. Altought the volatility gives a much better approximation to the observed volatility, the shape of the order book is worse than in the previous simulations.



Figure 3.3: Order book shape by direct simulation, corresponding to (3.1-3.6).



Figure 3.4: Estimated parameters corresponding to (3.13).

The last model to test is the one including all the three changes mentioned in this section. The parameters are equal to those shown in Table 3.4. Simulating gives the average shape shown in Figure 3.7. A 95%-confidence interval for the realized volatility of the mid-price is given by  $0.1370 \pm 0.0288$ , with an average of 312547 simulated events per trading day. The simulated volatility is closer to the observed volatility. However, at the most important points (i = 1, ..., 5), the simulated shape does not give a better approximation of the observed shape than the first two simulations.



Figure 3.5: Order book shape by direct simulation, corresponding to (3.13).



Figure 3.6: Order book shape by direct simulation, using (3.14).

#### 3.3 Final Order Book Model

This section gives an overview of the model that will be used for our further analysis, and study some of its properties. The model is based on the following assumptions about the order book.

- The order book consists of a finite number of price levels,
- All order are of size one,
- The order book can be represented as a queuing system,



Figure 3.7: Order book shape by direct simulation, corresponding to (3.7-3.12).

- The order book dynamics is driven by market orders, limit orders and cancellations, and
- All events have independent exponentially distributed inter arrival times.

These assumptions are also made by Cont et al. [10]. In Section 3.2 we have seen that the proposed changes to the parameter estimation do not improve the performance of the model significantly. We will therefore stick to the parameter estimation proposed by Cont. The assumption about limit orders arriving at the ask and bid will be adopted because it corresponds to reality. The arrival rates of events in the final order book model are therefore modeled by

Limit buy orders arrive at	$p \leq p_A(t)$	with rate	$\lambda(p_A(t) - p)$	),	(3.1	.5)
----------------------------	-----------------	-----------	-----------------------	----	------	-----

Limit sell orders arrive at 
$$p \ge p_B(t)$$
 with rate  $\lambda(p - p_B(t))$ , (3.16)

Market buy orders arrive with rate 
$$\mu$$
, (3.17)

Market sell orders arrive with rate 
$$\mu,$$
 (3.18)

Limit buy orders are canceled at 
$$p < p_A(t)$$
 with rate  $\theta(p_A(t) - p)|x_k(t)|$ , (3.19)

Limit sell orders are canceled at 
$$p > p_B(t)$$
 with rate  $\theta(p - p_B(t))|x_k(t)|,$  (3.20)

where  $x_k(t)$  represents the observed number of orders at price level k at time t.

Note that this way of estimating the parameters does not take trends in the order book into account, such as recurring orders. Recurring orders appear often and are caused by quoting obligations that apply on some market participants or iceberg orders. This will be discussed in Chapter 7.

For the chosen model, the estimated parameters are given by Table 3.2. To get an idea of the sensitivity of this parameter estimation we repeat the estimation per trading day. The resulting parameters are shown in Figures 3.8 - 3.9. As in Section 2.2, an influence of the particular trading day of the month can be observed. As mentioned before, it would therefore be interesting to perform statistical analysis on the influence of a monthly trend on the parameters. A data set over several months is necessary for that. Unfortunately, this is not available for this project. The sample standard deviations in Table 3.5 give a quantification of the sensitivity of the estimation, and show that the sensitivity of the estimates is almost equal to the sensitivity of the estimates in Chapter 2 (see Table 2.4).

The graphs in Figure 3.10 show the relative behavior of the estimates per day. It is clear that the intensities fluctuates around parallel lines and have higher values at the end of the month. Furthermore, the shape of the parameter curves is similar for every trading day.

Simulating 1000 trading days we find the 95%-confidence interval for the volatility of the mid-price  $0.1386 \pm 0.0056$ . This is not in correspondence with the observed values in Table 2.5. Estimating volatility from high-frequency data is a familiar problem in finance. For frequencies per minute or higher, noise results to be a







Figure 3.9: Estimated  $\lambda$  and  $\theta$  per day.

dominant consideration. Research on this topic is done by Ait-Sahalia [4]. In our simulation, an average of 216095 events took place per simulated trading day, with an average of 5359 trades. Although the simulated number of events is much higher than in the observed order book, the more important number of trades is well approximated by the simulation.

i	0	1	2	3	4	5	6	7	8	9
$\hat{\lambda}(i)$	0.0260	0.0908	0.1264	0.0799	0.0609	0.0437	0.0313	0.0283	0.0122	0.0064
$\hat{ heta}(i)$	_	0.0270	0.0194	0.0096	0.0062	0.0042	0.0041	0.0046	0.0020	0.0011
$\hat{\mu}$	0.0197									



Table 3.5: Standard deviations corresponding to the estimated parameters.

Figure 3.10: Comparison between estimated  $\lambda$  and  $\mu$  per day.

Another interesting idea is to compute a 95%-confidence interval for the shape of the order book as a function of i. To do so, it is necessary to compute the the standard deviation of the number of orders at every i. Computing the average number of orders at distance i from the opposite best bid, and the corresponding 95% upper and lower boundaries per day, we find the results shown in Figure 3.11. It is obvious that the shape of the order book is not stable at all. Since the parameters in the model depend on the shape of the order book, it would make sense to use this confidence interval to e.g. decide when to re-estimate the parameters. The results from Figure 3.11 however show that implementing this idea is not that straightforward. Therefore the possibility of implementing this idea needs to be further explored. It should be noted that the average orderbook shapes on January  $18^{th}$  and January  $20^{th}$  are very uncommon compared to the rest of the month. A close look at the data does not show anything unexpected however. Comparing the average observed standard deviation of the number of orders at i ticks from the opposite best quote to the simulated one, we find the results in Table 3.6. The results are reasonably close to the observed values.

i	0	1	2	3	4	5	6	7	8	9
average observed stdev	3.37	5.54	5.07	6.15	7.10	8.51	8.71	9.12	8.96	9.04
simulated stdev	1.98	3.18	4.22	5.52	6.77	8.41	10.74	13.24	15.59	17.68

Table 3.6: Standard deviations of the number of orders at i from the opposite best, observed vs. simulated.



Figure 3.11: Average order book shape and corresponding 95%-confidence interval per day. The red dotted line represents the upper boundary of the interval, the green dotted line represents the lower boundary of the interval.

# Part II Computing Probabilities

### Chapter 4

# **Approximating Laplace Transforms**

As noted before, the most important motivation for modeling order book dynamics, is to use all the information provided by the current limit order book to predict its short-term behavior, such as the probabilities of the mid-price moving up or down, or execution of an order before the bid and ask prices move, etc. In this chapter we show how the model introduced in Chapter 2 allows to analytically compute these conditional probabilities using Laplace transforms. We will start by introducing the concept of continued fractions, which will be used to construct the Laplace transform.

#### 4.1 Basic Concepts of Continued Fractions

To understand the concept of a continued fraction we draw the analogy with more familiar sequences that have in common with the continued fraction that each element can iteratively be computed from its predecessor. Similarly to the sequence  $\{\sigma_n\}$  of partial sums,

$$\sigma_n := \sum_{k=1}^n a_k = \sigma_{n-1} + a_n,$$
(4.1)

and the sequence  $\{p_n\}$  of partial products,

$$p_n := \prod_{k=1}^n a_k = p_{n-1} \cdot a_n, \tag{4.2}$$

let  $\{f_n\}$  be the sequence

$$f_n := \frac{a_1}{b_1 + \frac{a_2}{b_2 + \frac{a_3}{b_3 + \frac{a_3}{b_$$

If we assume that  $a_n \neq 0$  for all n and allow  $f_n = \infty$ , then  $\{f_n\}$  is well defined in  $\hat{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$ . As is the case for sums and products, an infinite continuation of this process is also defined here, the continued fraction:

$$f := \prod_{n=1}^{\infty} \frac{a_n}{b_n} = \frac{a_1}{b_1 + \frac{a_2}{b_2 + \frac{a_3}{\cdots}}}$$
(4.4)

For simplicity (4.4) it is also often written as

$$f := \prod_{n=1}^{\infty} \frac{a_n}{b_n} = \frac{a_1}{b_1 + b_2 + b_2 + b_3 + \cdots}$$
(4.5)

Now note the common pattern in the first two sequences above: the partial sequence  $\{\Phi_n\}$  can be constructed by function composition. By applying a function on  $\Phi_{n-1}$  we can iteratively find its successive element  $\Phi_n = \Phi_{n-1} \circ \phi_n = \phi_1 \circ \phi_2 \circ \cdots \circ \phi_n$ . For partial sums  $\phi_n$  simply is a summation and for products it obviously is a product. Also for continued fractions such a function exists and is given by

$$s_n(\omega) := \frac{a_k}{b_k + \omega},\tag{4.6}$$

so that

$$S_n(\omega) = s_1 \circ s_2 \circ \dots \circ s_n(\omega) = \frac{a_1}{b_1 + b_2 + \dots + b_n + \omega}$$

$$(4.7)$$

represents  $f_n$  in (4.3). This leads to the following definition.

Definition 4.1. A continued fraction

$$b_0 + \mathop{K}\limits_{n=1}^{\infty} \frac{a_n}{b_n}$$

is an ordered pair  $((\{a_n\}, \{b_n\}), \{S_n\})$ , given by

$$S_n(\omega) = b_0 + \frac{a_1}{b_1 +} \frac{a_2}{b_2 +} \dots \frac{a_n}{b_n + \omega},$$
(4.8)

where  $\{a_n\}$  and  $\{b_n\}$  are sequences of complex numbers with  $a_n \neq 0$  for all n.

 $S_n$  is called the  $n^{th}$ -approximant, and if  $\lim_{n\to\infty} S_n = S$  exists, the Continued Fraction is said to converge and S is called its value. More about convergence of continued fractions can be found in Jacobsen [16] and Lorentzen and Waadeland [19]. An important result is the following lemma, taken from [19]. It shows that there is a recursion to calculate successive approximants of a continued fraction.

**Lemma 4.2.** Let the  $n^{th}$ -approximant  $S_n$  be given by (4.8). Then

$$S_n(\omega) = \frac{A_{n-1}\omega + A_n}{B_{n-1}\omega + B_n} \text{ for } n = 1, 2, ...,$$
(4.9)

where

 $A_n = b_n A_{n-1} + a_n A_{n-2}, \quad B_n = b_n B_{n-1} + a_n B_{n-2},$ 

with initial values  $A_{-1} = 1$ ,  $A_0 = b_0$ ,  $B_{-1} = 0$  and  $B_0 = 1$ .

Proof. see Lorentzen and Waadeland [19]

Using continued fractions, rational approximations of irrational numbers can be constructed. To give an idea of how this works, consider the continuous fraction expansion of  $\pi$  as an example.

**Example 4.3.** We try to find a representation of  $\pi$  of the form  $b_0 + \frac{a_1}{b_1+} \frac{a_2}{b_2+} \frac{a_3}{b_3+}$ ... Since we know that  $\pi > 3$ , we start with  $b_0 = \lfloor \pi \rfloor = 3$  and set  $a_1 = 1$ .

$$\pi := 3 + \frac{1}{b_1 + x}$$

From this it follows that

$$\frac{1}{\pi - 3} := b_1 + x \approx 7.0625.$$

Again we take the largest integer that is smaller than the term on the right-hand side to estimate  $b_1 = \left|\frac{1}{\pi-3}\right| = 7$ , and get

$$\pi := 3 + \frac{1}{7+b_2 + x}.$$

If we continue like this, we find the continued fraction

$$\pi := 3 + \frac{1}{7+} \frac{1}{15+} \frac{1}{1+} \frac{1}{929+} \frac{1}{1+} \frac{1}{1+} \frac{1}{1+} \frac{1}{2+} \frac{1}{2+} \dots,$$

that converges to  $\pi$ .  $\diamond$ 

Obviously, the sequence  $\{b_n\}$  in the  $\pi$ -example does not show any pattern. If we repeated this same example for  $\sqrt{2}$ , we would find two constant sequences  $a_n = 1$  and  $b_n = 2$  for all  $n \ge 0$ . Once we have found sequences  $\{a_n\}$  and  $\{b_n\}$ , we can recursively calculate the approximants  $S_n$  by using Lemma 4.2. An expansion similar to the one for irrational numbers can be used to approximate functions. When doing this, it is always important to verify whether the continued fraction indeed converges to the function it is trying to approximate. Continued fraction function expansions are less known than power series expansions, but have grown in popularity because they are easy to program, they have good convergence properties, and their convergence is often easy to accelerate. Much has been written about the convergence of continued fractions. Therefore we refer to [16], pp. 201-245 in [5] and pp. 59-60 in [19] here.

Many continued fraction expansions of useful functions are of the form

$$b_0 + \prod_{n=1}^{\infty} \frac{a_n z}{1} = b_0 + \frac{a_1 z}{1+} \frac{a_2 z}{1+} \frac{a_3 z}{1+} \dots; b_0 \ge 0, a_n > 0 \text{ for all } n.$$
(4.10)

Continued fractions of this special type are called Stieltjes continued fractions, or in brief S-fractions. They are known for having nice convergence properties. For more about S-fractions and their convergence properties we refer to Lorentzen and Waadeland [19].

#### 4.2 Laplace Transforms via Continued Fractions

Let  $f : \mathbb{R} \to \mathbb{R}$  be a pdf of a non-negative random variable X. Then the (one-sided) Laplace transform of the random variable X is given by

$$\hat{f}(\tau) = \int_0^\infty e^{-\tau t} f(t) dt, \quad \text{for } \tau \in \mathbb{C}.$$
(4.11)

If  $\hat{f}(\tau)$  is available, the pdf f(t) can be recovered from its Laplace transform by using the inverse Laplace transform

$$f(t) = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} e^{t\tau} \hat{f}(\tau) d\tau, \qquad (4.12)$$

for some  $\gamma \in \mathbb{R}$  for which  $\int_{-\infty}^{\infty} |\hat{f}(\gamma + i\omega)| d\omega < \infty$ . So if we can find Laplace transforms for the order book probabilities, theseprobabilities conditional on the current order book state can be computed directly. Unfortunately, Laplace transform values are often unavailable. Abate and Whitt show in [2] that it is possible to find S-fraction representations for Laplace transforms of completely monotone pdf's (see next section). We will first start by showing how Laplace transforms and S-fractions are related.

Recall that the  $n^{th}$  moment of a pdf f is given by

$$m_n(f) = \int_0^\infty t^n f(t) dt \text{ for } n = 0, 1, 2, \dots$$
 (4.13)

Using the Maclaurin expansion of  $e^{-\tau t}$  we can rewrite (4.11)

$$\hat{f}(\tau) = \int_0^\infty e^{-\tau t} f(t) dt = \int_0^\infty \left( \sum_{n=0}^\infty \frac{(-\tau t)^n}{n!} \right) f(t) dt.$$
(4.14)

If we write out the summation and take all the terms that do not depend on t out of the integral, by dominated convergence we find the following expression for the Laplace transform in terms of the moment-generating function (4.13).

$$\hat{f}(\tau) = \sum_{n=0}^{\infty} c_n \tau^n$$
 where  $c_n = (-1)^n \frac{m_n(f)}{n!}$ , (4.15)

which means that  $\hat{f}(\tau) - \sum_{n=0}^{N} c_n \tau^n = O(\tau^{N+1})$  as  $\tau \to 0$ . However, in general we cannot conclude that (4.15) converges.

This is connected to the so-called Stieltjes moment problem: For a given sequence  $\{m_n\}_{n=0}^{\infty}$  of real numbers, find a distribution function f(t) on  $[0,\infty)$  that satisfies (4.13). In his famous paper published in 1894 [27], Stieltjes showed the following.

• A solution f(t) to the moment problem for  $\{m_n\}$  exists if and only if, the series (4.15) at  $\tau = 0$  corresponds to a continued fraction of the form

$$\mathbf{K}_{n=1}^{\infty} \frac{a_n \tau^{-1}}{1}, a_n > 0,$$

which is exactly an S-fraction and is known to converge under minor conditions.

• A solution f(t) to the moment problem for  $\{m_n\}$  is unique if and only if the corresponding S-fraction converges for  $\tau > 0$ .

This tells us that if (4.15) can be written as an S-fraction, a solution f exists and moreover, if this S-fraction converges then the pdf f is unique. A property of the pdf to ensure that the S-fraction actually converges is complete monotonicity. A function f on  $[0, \infty)$  is said to be completely monotone if it satisfies

$$(-1)^n f^{(n)}(t) \ge 0$$
 for all  $t, n \ge 0.$  (4.16)

Keilson [18] shows that the family of completely monotone pdf's is closed under mixtures. So if a solution f exists and is completely monotone, then the corresponding S-fraction converges and according to Stieltjes f is unique.

#### 4.3 First Passage Times in Birth-Death Processes

Now we know that the pdf f can be recovered by inverting its Laplace transform, and we have seen under which conditions a Laplace transform can be represented by an S-fraction. In a birth-death process the passage time to a neighboring state is exponentially distributed, so the probability density function corresponding to the transition from any state to another is completely monotone. Therefore, existence of an S-fraction representation of the Laplace transform implies that a unique density function f exists. Here we will show how to derive an S-fraction representation of the first passage time pdf in a birth-death process.

Consider a birth-death process with  $\lambda_i$  and  $\theta_i$  birth and death rates in state *i*. Let  $T_{\theta}$  and  $T_{\lambda}$  be independent exponentially distributed with parameters  $\theta_i$  and  $\lambda_i$ . Now assume that we are in state *i* and we are interested in finding the Laplace transform of the first passage time to state i - 1,

$$\hat{f}_i(\tau) = E[e^{-\tau t_{i,i-1}}],\tag{4.17}$$

where  $t_{i,i-1}$  represents the first passage time from i to i-1. Given that we are in state i, two events can occur: if  $T_{\lambda} < T_{\theta}$  we go up to state i+1, if  $T_{\theta} < T_{\lambda}$  we go immediately to state i-1. Therefore we have that

$$\left\{ \begin{array}{ll} t_{i,i-1} = T_{\lambda} + t_{i+1,i-1} & \text{ if } \min(T_{\lambda},T_{\theta}) = T_{\lambda}, \text{ and} \\ t_{i,i-1} = T_{\theta} & \text{ if } \min(T_{\lambda},T_{\theta}) = T_{\theta}. \end{array} \right.$$

Rewriting the Laplace transform (4.17) gives

$$\hat{f}_{i}(\tau) = E[e^{-\tau(T_{\lambda} + t_{i+1,i-1})} \mathbb{1}_{\{T_{\lambda} < T_{\theta}\}}] + E[e^{-\tau T_{\theta}} \mathbb{1}_{\{T_{\theta} < T_{\lambda}\}}]$$
(4.18)

First consider the first expectation. Because of the independence between arrival times in a the birth-death process we can write the first expectation as

$$E[e^{-\tau(T_{\lambda}+t_{i+1,i-1})}\mathbb{1}_{\{T_{\lambda}< T_{\theta}\}}] = E[e^{-\tau T_{\lambda}}\mathbb{1}_{\{T_{\lambda}< T_{\theta}\}}]\underbrace{E[e^{-\tau t_{i+1,i}}]}_{=\hat{f}_{i+1}(\tau)}\underbrace{E[e^{-\tau t_{i,i-1}}]}_{=\hat{f}_{i}(\tau)}$$
$$= \int_{0}^{\infty}\int_{0}^{\infty}e^{-\tau t}f_{T_{\lambda},T_{\theta}}(t,u)\mathbb{1}_{\{t< u\}}dtdu\;\hat{f}_{i+1}(\tau)\hat{f}_{i}(\tau).$$

Because of independence we can also write  $f_{T_{\lambda},T_{\theta}}(t,u) = f_{T_{\lambda}}(t)f_{T_{\theta}}(u) = \lambda_i e^{-\lambda_i t} \theta_i e^{-\theta_i u}$ . Using this, the integral term becomes

$$\begin{split} \int_0^\infty \int_0^\infty e^{-\tau t} \lambda_i e^{-\lambda_i t} \theta_i e^{-\theta_i u} \mathbbm{1}_{\{t < u\}} dt du &= \int_0^\infty \lambda_i e^{-(\tau + \lambda_i) t} \left[ \int_t^\infty \theta_i e^{-\theta_i u} du \right] dt \\ &= \int_0^\infty \lambda_i e^{-(\tau + \lambda_i) t} e^{-\theta_i t} dt \\ &= \int_0^\infty \lambda_i e^{-(\tau + \lambda_i + \theta_i) t} dt \\ &= \frac{\lambda_i}{\tau + \lambda_i + \theta_i}. \end{split}$$

Combining these results we find for the first expectation of (4.18) that

$$E[e^{-\tau(T_{\lambda}+t_{i+1,i-1})}\mathbb{1}_{\{T_{\lambda}< T_{\theta}\}}] = \frac{\lambda_i}{\tau+\lambda_i+\theta_i} \ \hat{f}_{i+1}(\tau)\hat{f}_i(\tau).$$
(4.19)

The second expectation in (4.18) can be computed in a similar way:

$$E[e^{-\tau T_{\theta}} \mathbb{1}_{\{T_{\theta} < T_{\lambda}\}}] = \int_{0}^{\infty} \int_{0}^{\infty} e^{-\tau u} f_{T_{\lambda}, T_{\theta}}(t, u) \mathbb{1}_{\{u < t\}} du dt$$
$$= \int_{0}^{\infty} \int_{0}^{\infty} e^{-\tau u} \lambda_{i} e^{-\lambda_{i} t} \theta_{i} e^{-\theta_{i} u} \mathbb{1}_{\{u < t\}} du dt$$
$$= \int_{0}^{\infty} \theta_{i} e^{-(\tau + \theta_{i})u} \left[ \int_{u}^{\infty} \lambda_{i} e^{-\lambda_{i} t} dt \right] du$$
$$= \int_{0}^{\infty} \theta_{i} e^{-(\tau + \theta_{i} + \lambda_{i})u} du$$
$$= \frac{\theta_{i}}{\tau + \theta_{i} + \lambda_{i}}$$

Inserting all these results into equation (4.18) we find the following expression for the Laplace transform of the first passage time pdf,

$$\hat{f}_i(\tau) = \frac{\lambda_i}{\tau + \lambda_i + \theta_i} \ \hat{f}_{i+1}(\tau) \hat{f}_i(\tau) + \frac{\theta_i}{\tau + \theta_i + \lambda_i}$$

This can be written as

$$\hat{f}_i(\tau)\left(1 - \frac{\lambda_i \hat{f}_{i+1}(\tau)}{\tau + \lambda_i + \theta_i}\right) = \frac{\theta_i}{\tau + \theta_i + \lambda_i},$$

from which we obtain

$$\hat{f}_i(\tau) = \frac{\theta_i}{\tau + \lambda_i + \theta_i - \lambda_i \hat{f}_{i+1}(\tau)}.$$
(4.20)

Writing out the iteration in (4.20) we find the following continued fraction representation for the Laplace transform.

$$\hat{f}_i(\tau) = \frac{\theta_i}{\tau + \lambda_i + \theta_i - \frac{\lambda_i \theta_{i+1}}{\tau + \lambda_{i+1} + \theta_{i+1} - \frac{\lambda_{i+1} \theta_{i+2}}{\tau + \lambda_{i+2} + \theta_{i+2} - \ddots}}$$

Multiplying by  $\frac{-\lambda_{i-1}}{-\lambda_{i-1}}$  and using Definition 4.8, this continued fraction can be written as

$$\hat{f}_i(\tau) = \frac{-1}{\lambda_{i-1}} \prod_{k=i}^{\infty} \frac{-\lambda_{k-1}\theta_k}{\lambda_k + \theta_k + \tau},$$
(4.21)

which has the form of a so called real Jacobi fraction or J-fraction. On pp. 166-167 of Baker and Graves-Morris [5] it is shown that J-fractions of the form

$$J(\tau) = \bigwedge_{n=1}^{\infty} \frac{-k_n}{l_n + \tau}$$
(4.22)

have an equivalent S-fraction representation. The S-fraction equivalent to (4.22) is

$$\frac{1}{\tau} \frac{a_1}{1+1} \prod_{n=2}^{\infty} \frac{a_n \tau^{-1}}{1} = \prod_{n=1}^{\infty} \frac{a_n \tau^{-1}}{1},$$
(4.23)

where  $\{a_n\}$  satisfies

$$a_1 = k_1, \ a_2 = l_1, \ a_{2m+1} = \frac{k_{m+1}}{a_{2m}}, \ \text{and} \ a_{2m+2} = \frac{l_{m+1}}{a_{2m+1}},$$
 (4.24)

for m = 1, 2, 3, ... It can therefore be concluded that the Laplace transform of a first passage-time pdf in a birth-death process (4.21) has an S-fraction representation

$$\hat{f}_{i}(\tau) = \frac{-1}{\lambda_{i-1}} \bigvee_{k=i}^{\infty} \frac{a_{k}\tau^{-1}}{1},$$
(4.25)

with

$$a_1 = \lambda_{i-1}\theta_i, \quad a_2 = \lambda_i + \theta_i, \quad a_{2m+1} = \frac{\lambda_m \theta_{m+1}}{a_{2m}}, \quad \text{and} \quad a_{2m+2} = \frac{\lambda_{m+1} + \theta_{m+1}}{a_{2m+1}},$$
(4.26)

for m = 1, 2, 3, ... From Stieltjes' results it now follows that a unique density function  $f_i$  for the first-passage time from state i to i - 1 can be recovered from  $\hat{f}_i$ , where  $\hat{f}_i$  is iteratively given by the birth and death rates of the process. From (4.25) it follows that the first-passage time from some state b of the process to state a is given by

$$\hat{f}_{b,a}(\tau) = \hat{f}_{a+1}(\tau) \cdots \hat{f}_{b}(\tau) = \frac{-1}{\lambda_{a}} \prod_{k=a+1}^{\infty} \frac{a_{k}\tau^{-1}}{1} \cdots \frac{-1}{\lambda_{b-1}} \prod_{k=b}^{\infty} \frac{a_{k}\tau^{-1}}{1}$$
$$= \prod_{i=a+1}^{b} \left( \frac{-1}{\lambda_{i-1}} \prod_{k=i}^{\infty} \frac{a_{k}\tau^{-1}}{1} \right),$$
(4.27)

with  $a_k$  given by (4.26). This result will be used in the next two chapters. Note that the Markov chain assumption also enables us to compute probabilities of interest using a transition matrix as for example in Section 4.4 of Ross [23]. However, the approach presented in this chapter has the advantage that it can compute full distributions of random variables fast.

### Chapter 5

### Probabilities in the Order Book

In this chapter we will apply the results from Chapter 4 to compute probabilities in our order book model that are of interest for traders. We will start by discussing some probabilities and how knowledge about them can influence a traders' strategy. The information presented in Section 5.1 is gained from discussions with traders and other people involved in trading. The choices are based on what practitioners actually are interested in.

#### 5.1 Probabilities of Influence on Trading Decisions

As mentioned in the introduction, the motivation for modeling order book dynamics is to use all the information provided by the current state of the order book, to say something about its short-time behavior. Knowing that every change in the order book will affect the future events in the order book, a trader can decide to submit or cancel orders conditioned on the current order book. The current state of the order book allows to compute how probable following states are to occur. We would like to emphasize once more that we are considering statistical probabilities on a very short-time scale only, without predicting general market movements.

Imagine a trader who in order to hedge his position needs to buy a certain number of contracts. The contracts are 'valued' by the market at the mid-price  $p_M$ . The trader is willing to buy the contracts at a relative high price by submitting a market order. By submitting a market order a trader basically 'pays' for the certainty of having the order executed. The 'price' he pays is what we call the margin, which is the difference between the price at which the trade is executed and the mid-price. So the margin payed by our trader if he submits a market buy order is equal to  $p_A - p_M$ , his counterparty wins this margin. Now suppose that the trader knows that the mid-price is very likely to decrease. It would then be wise to, instead of a market buy order, submit a limit buy order at a slightly higher price than the current bid. Because the mid-price is likely to decrease, it is likely that the limit buy order is hit (by an incoming market sell order). The trader than pays less for the contracts and even 'wins margin'. Therefore it is useful to know the probabilities of the mid-price increasing or decreasing. In Section 5.2 it is shown how to compute an expression for these probabilities.

Relevant in the situation above is also the probability that a limit order that is submitted at the bid (ask) is executed before the mid-price moves. Suppose that the trader from our example, in order to get a better price, submits an order at the current bid. Then his bid has to line up and thus lies last in queue at  $p_B$ . It is relevant to know the execution probability before the mid-price moves because it gives an indication of the expected gain of the order and helps quantifying the choice between placing a limit order and placing a market order. In Section 5.3 it is shown how the results from Chapter 4 enable us to compute this probability.

The trader can also decide to place his buy order at a price level above the current bid. Then this price level is the new bid, and the order of our trader is the first in queue at the bid. Note that the place in queue is considered because orders are executed according to a price/time priority rule: the first order in queue at a certain price is the first one to go out. Proposition 5.4 presented in Section 5.3 allows us to compute the execution probability of hypothetically submitted orders at any price level  $p_B , conditional on the current configuration of the order book. This enables us to compute the expected gain of orders at any of these price levels: submitting a bid order at a higher price will give less margin but has a higher probability of being executed. Choosing the price level with the highest expected gain gives the optimal price to submit an order.$ 

Another probability interesting to traders that we can compute is the probability that both the bid and ask are executed before the mid-price moves. This is referred to as 'making the spread'. If the probability of making the spread is high, than a trader can submit an ask and a bid simultaneously. If both are executed before the mid-price moves, than the strategy has paid of the spread  $p_S = p_A - p_B$ . This is called a statistical arbitrage opportunity.

A different way of making the spread is to submit orders just above the bid and under the ask, such that you are the first order to be executed when a market order arrives. Note that the smaller the spread between the submitted bid and ask, the higher the probability of success, but smaller the gain of the trade would be. Also combinations of these two approaches are possible. In Section 5.4 only the first two approaches are treated. Expressions to calculate success probabilities of other approaches can be computed in a similar way.

Note that computing the aforementioned probabilities does not mean that it is possible to make free money. Strategies based on these probabilities may often result in losses, especially if we take the costs of trading into account. The idea of this research however is, to optimize the 'win'-rate such that the gains compensate for the costs of the lost bets. Take for example the strategy of making the spread. Because successful implementation of this strategy pays out the spread, it makes little sense to apply it on liquid stocks, where the spread is very small (smaller than the costs of the trade). Therefore, this strategy is only relevant for stocks with a relatively large bid-ask spread. But even if the spread is large, implementation of this strategy might result in losses: while successful implementation pays out the spread, a move of the market in a certain direction may result in a loss n times the spread. In order to break even, one loss then has to be countervailed by n successes. Therefore, to make such a strategy a winning strategy, it is important to perform order book analysis, such as in this project.

Note that liquidity is a very relative concept which is influenced by the bid-ask spread, the available bid and ask sizes and the traded volume. For the probability of making the spread it was mentioned that the stock should be 'relatively illiquid' by which we mean that the spread has to be reasonably large. However, in order to make the spread making strategy successful the stock in question needs to be actively traded. A better formulation would therefore be 'a liquid stock (in terms of turnover) with a high volatility (large spread)'. For now we won't quantify such expressions, this will become relevant when implementing the theory presented here in certain trading strategies.

A last probability that will be computed in this chapter is the probability of reaching the stop loss. When a trader goes long or short stocks, a 'stop loss order' can be submitted simultaneously. This is an order that is not visible in the market, but is converted into a market order if it is triggered by the market. The stop loss order is triggered by the market when the mid-price moves such that the (long or short) position of the trader hits a certain level of loss. This level of loss is specified by the trader depending on his strategy and the composition of his book. Knowing the probability that the stop loss is reached helps in designing strategies.

RV	corresponding event
$X_0$	configuration of the order book at the initial time
T	first time at which the mid-price moves
$T^A$	first-passage time to zero of number of orders at the ask
$T^B$	first-passage time to zero of number of orders at the bid
$M^A$	first arrival time of an ask order in the spread
$M^B$	first arrival time of a bid order in the spread
$E^A$	execution time of the first order in line at the ask
$E^B$	execution time of the first order in line at the bid
$I^A$	first-passage time to zero of number of orders that were initially at the ask
$I^B$	first-passage time to zero of number of orders that were initially at the bid
$NC^A$	one order at the ask will never be canceled
$NC^B$	one order at the bid will never be canceled

Table 5.1: Overview defined Random Variables

#### 5.2 Move of the Mid-Price

Recall from section 4.3 that the Laplace transform of the first-passage time from state b to state a in a birth-death process with birth rates  $\lambda_i$  and death rates  $\theta_i$  is given by

$$\hat{f}_{a,b}(\tau) = \prod_{i=a+1}^{b} \left( \frac{-1}{\lambda_{i-1}} \prod_{k=i}^{\infty} \frac{-\lambda_{k-1}\theta_k}{\lambda_k + \theta_k + \tau} \right).$$
(5.1)

We will use this to compute the probabilities discussed in Section 5.1, conditional on the current order book. The first probability we will consider is the probability that the mid-price increases at its next move.

There are two different events that can make the mid-price move:

- The number of orders at  $p_A$  or  $p_B$  passes to zero, or
- a limit order arrives between  $p_A$  and  $p_B$ , given that the spread  $p_S > 1$ .

If we let the initial state of the order book be  $X_0 = (X_0^{(1)}, X_0^{(2)}, ..., X_0^{(n)})$ , and define the first time at which the mid-price moves as

$$T := \inf\{t \ge 0, p_M(t) \neq p_M(0)\}.$$
(5.2)

Then the probability that the mid-price goes up at its next move can be written as

$$P(p_M(T) > p_M(0) | X_0^{(p_A)} = a, X_0^{(p_B)} = b, p_S(0) = S),$$
(5.3)

with a and b the numbers of orders at the best bid and best ask. Recall that in our model the number of orders at the best ask and best bid follow a birth-death process with rates as in Figure 5.1.



Figure 5.1: Transition rates for  $p_A$  and  $p_B$ .

Note that these transition rates are valid because i does not change! Using these rates, it follows from (5.1) that the Laplace transform of the first passage time of the number of orders at  $p = p_A, p_B$  from j to zero is given by

$$\hat{f}_j^S(\tau) = \prod_{i=1}^j \left( \frac{-1}{\lambda(S)} \bigvee_{k=i}^\infty \frac{-\lambda(S)(\mu + \theta(S)|k|)}{\lambda(S) + \mu + \theta(S)|k| + \tau} \right).$$
(5.4)

For the density of arrival times of orders in the spread consider the following. We assumed in our model that, at any price level, the inter-arrival times between limit orders is exponentially distributed with parameter  $\lambda(i)$ . It is easy to see that the inter-arrival time of orders at any price level p in the spread  $p_S = S$  is exponentially distributed with parameter

$$\Lambda(S) = \sum_{i=1}^{S-1} \lambda(i).$$
(5.5)

Now let  $T^A$  and  $T^B$  denote the respective passage times to zero of the number of orders at the ask and the bid, and let  $M^A$  and  $M^B$  denote the first times that an ask order resp. bid order arrives in the spread. Then  $T^A$  and  $T^B$  are independent and are distributed corresponding to the inverse Laplace transforms of (5.4),  $T^A \sim f_a^S(t)$ ,  $T^B \sim f_b^S(t)$ , and  $M^A$  and  $M^B$  are independent and exponentially distributed with parameter

 $\Lambda(S)$  as in (5.5). Remember that in our model, all possible events  $Y_i$  are modeled with independent interarrival times. It is known that for a sequence of mutually independent random variables  $(Y_1, Y_2, Y_3, ...)$ , also the sequence  $(Y_1 + Y_2, Y_3, ...)$  is mutually independent. The first passage time of the number of orders at the ask or at the bid is the sum of independent inter-arrival times. Therefore, we can conclude that also  $T^A, T^B, M^A$  and  $M^B$  are independent. An overview of all random variables used in this chapter can be found in Table 5.1.

First consider the situation where the spread  $p_S := S = 1$ . Then probability (5.3) can be written as

$$P(p_M(T) > p_M(0)|X_0, S = 1) = P(T^A < T^B|X_0, S = 1) = P(T^A - T^B < 0|X_0, S = 1).$$
(5.6)

We can obtain the Laplace transform of the conditional probability density as follows (leaving the condition out of the notation).

$$\hat{f}_{T^{A}-T^{B}}(\tau) = \int_{0}^{\infty} e^{-\tau t} f_{T^{A}-T^{B}}(t) dt = \int_{0}^{\infty} e^{-\tau t} \int_{0}^{\infty} f_{a}^{S}(u) f_{b}^{S}(u-t) du dt$$

where we have used the independence of  $T^A$  and  $T^B$  to write the density of the difference between variables as the cross-correlation of their densitities. Now use Fubini's theorem and the substitution v = u - t to write

$$\begin{split} \hat{f}_{T^A-T^B}(\tau) &= \int_0^\infty f_a^S(u) \left( \int_0^\infty e^{-\tau t} f_b^S(u-t) dt \right) du \\ &= \int_0^\infty e^{-\tau u} f_a^S(u) \left( \int_0^\infty e^{\tau v} f_b^S(v) dv \right) du \\ &= \hat{f}_b^S(-\tau) \int_0^\infty e^{-\tau u} f_a^S(u) du = \hat{f}_a^S(\tau) \hat{f}_b^S(-\tau). \end{split}$$

The Laplace transform of the cdf  $F_{T^A-T^B}(t)$  then satisfies

$$\hat{F}_{T^{A}-T^{B}}(\tau) = \frac{1}{\tau}\hat{f}_{T^{A}-T^{B}}(\tau) = \frac{1}{\tau}\hat{f}_{a}^{S}(\tau)\hat{f}_{b}^{S}(-\tau).$$
(5.7)

We can now find the probability that the mid-price increases when it moves, given that S = 1 by evaluating the inverse Laplace transform of (5.7) at zero, with  $\hat{f}_j^S(\tau)$  given by (5.4) with S = 1.

Now allow the spread to be  $p_S := S > 1$ . Then the conditional probability that the mid-price increases can be interpreted as: the probability that either the number of orders at the ask goes to zero or a bid order arrives in the spread, before the number of orders at the bid goes to zero or an ask order arrives in the spread. Thus probability (5.3) can now be written as

$$P(p_M(T) > p_M(0)|X_0, S > 1) = P(T^A \land M^B - T^B \land M^A < 0|X_0, S > 1).$$
(5.8)

We now need to find the Laplace transform of the joint density of the first passage times to zero and arrival of limit orders in the spread. A useful lemma is the following.

**Lemma 5.1.** Let the random variable  $X \sim f_{a,b}(t)$  denote the first passage time from b to a and let Y be an exponentially distributed random variable with parameter  $\Lambda$  that is independent of X. Then the Laplace transform of the density of  $X \wedge Y$  is given by

$$\frac{\tau}{\Lambda + \tau} \hat{f}_{a,b}(\tau + \Lambda) + \frac{\Lambda}{\Lambda + \tau},$$
(5.9)

with  $\hat{f}_{a,b}$  as in (5.1).

*Proof.* Because of the independence between X and Y and because Y is exponentially distributed, we can write the cdf of  $X \wedge Y$  as

$$P(X \land Y < t) = 1 - P(X \land Y > t) = 1 - P(X > t)P(Y > t) = 1 - (1 - F_{a,b}(t))e^{-\Lambda t},$$

where  $F_{a,b}(t)$  denotes the cdf of X. Taking the derivative we find the pdf

$$f_{X \wedge Y}(t) = f_{a,b}(t)e^{-\Lambda t} + \Lambda(1 - F_{a,b}(t))e^{-\Lambda t}, \text{ for } t \ge 0.$$

The Laplace transform can now be calculated as follows

$$\hat{f}_{X\wedge Y}(\tau) = \int_{0}^{\infty} e^{-\tau t} f_{X\wedge Y}(t) dt = \int_{0}^{\infty} e^{-\tau t} f_{a,b}(t) e^{-\Lambda t} dt + \int_{0}^{\infty} \Lambda(1 - F_{a,b}(t)) e^{-\Lambda t} dt$$

$$= \hat{f}_{a,b}(\Lambda + \tau) + \Lambda \left\{ \frac{-1}{\Lambda + \tau} (1 - F_{a,b}(t)) e^{-t(\Lambda + \tau)} \Big|_{t=0}^{\infty} - \int_{0}^{\infty} \frac{1}{\Lambda + \tau} f_{a,b}(t) e^{-t(\Lambda + \tau)} dt \right\}$$

$$= \hat{f}_{a,b}(\Lambda + \tau) + \Lambda \left\{ \frac{1}{\Lambda + \tau} - \frac{1}{\Lambda + \tau} \int_{0}^{\infty} f_{a,b}(t) e^{-t(\Lambda + \tau)} dt \right\}$$

$$= \hat{f}_{a,b}(\Lambda + \tau) + \frac{\Lambda}{\Lambda + \tau} \left( 1 - \hat{f}_{a,b}(\Lambda + \tau) \right) = \frac{\tau}{\Lambda + \tau} \hat{f}_{a,b}(\Lambda + \tau) + \frac{\Lambda}{\Lambda + \tau}. \quad (5.10)$$

Following the same procedure as we did for S = 1 and using Lemma 5.1, we can find an expression for the Laplace transform of the conditional distribution of  $T^A \wedge M^B - T^B \wedge M^A$ .

**Proposition 5.2.** The conditional probability that the mid-price increases at its next move (5.8), is given by the inverse Laplace transform of

$$\hat{F}_{a,b}^{S}(\tau) = \frac{1}{\tau} \left( \frac{\tau}{\Lambda(S) + \tau} \hat{f}_{a}^{S}(\Lambda(S) + \tau) + \frac{\Lambda(S)}{\Lambda(S) + \tau} \right) \left( \frac{-\tau}{\Lambda(S) - \tau} \hat{f}_{b}^{S}(\Lambda(S) - \tau) + \frac{\Lambda(S)}{\Lambda(S) - \tau} \right), \quad (5.11)$$

evaluated at 0, where  $\Lambda(S) = \sum_{i=1}^{S-1} \lambda(i)$ , a is the number of orders at the ask, b the number of orders at the bid and  $\hat{f}_j^S(\tau)$  is given by (5.4) for j = a, b.

*Proof.* Similar to the case where S = 1, we can express the Laplace transform of the distribution of  $T^A \wedge M^B - T^B \wedge M^A$  as

$$\hat{f}_{T^A \wedge M^B - T^B \wedge M^A}(\tau) = \hat{f}_{T^A \wedge M^B}(\tau) \hat{f}_{T^B \wedge M^A}(-\tau).$$
(5.12)

If we now apply Lemma 5.1 on both Laplace transforms, we find that (5.12) equals

$$\left(\frac{\tau}{\Lambda+\tau}\hat{f}_{a}^{S}(\Lambda+\tau)+\frac{\Lambda}{\Lambda+\tau}\right)\left(\frac{-\tau}{\Lambda-\tau}\hat{f}_{b}^{S}(\Lambda-\tau)+\frac{\Lambda}{\Lambda-\tau}\right).$$
(5.13)

According to (5.7) the Laplace transform  $\hat{F}_{a,b}^{S}(\tau)$  for S > 1 then satisfies (5.11). If we now evaluate (5.11) for S = 1, we find the same result as in (5.7).

Obviously, the probability that the mid-price decreases at its next move, is equal to one minus the probability from Proposition 5.2. The results in this section correspond to Cont et al.'s findings.

#### 5.3 Probability of Order Execution

We will derive the Laplace transform of the conditional probability that the order placed last in queue at the bid or ask is executed before the mid-price moves. Because we assumed the parameters for the bid and ask to be symmetric, the Laplace transform is similar for both sides. We will perform our analysis based on the ask side, a similar result holds for the bid side.

As in Section 5.2, let  $X_0$  be the known initial order book state with  $X_0^{(p_A)} = a$ ,  $X_0^{(p_B)} = b$  and  $p_S(0) = S$ , and let T be the first time at which the mid-price moves (5.2). Let  $I^A$  be the time at which all the orders that were initially at the ask are executed, given that the last order in queue at the ask is not canceled. Let  $NC^A$  denote the event that the last order initially in line at the ask is never canceled (see Table 5.1). Note that we are only interested in the orders that are initially at the ask because all orders that are submitted at the ask after the initial time stand in queue behind our order. The last order in the queue is the order submitted by our trader at time 0. We can therefore condition on  $NC^A$ . The probability of our interest can be written as

$$P(I^{A} < T|X_{0}, NC^{A}) = P(I^{A} - T < 0|X_{0}, NC^{A}).$$
(5.14)

Given that the mid-price does not move and the last order in the initial queue is not canceled, the number of orders initially at the ask follows a pure death process with death-rate  $\mu + \theta(S)|x^{(p_A)} - 1|$ . Recall that the cancellation rate is proportional to the number of orders at the particular price level. In this case, the condition implies that there is one order that will never be canceled. For this reason, the cancellation rate is taken proportional to the number of orders minus one. The Laplace transform of  $I^A$  then is

$$\hat{g}_{a}^{S}(\tau) = \prod_{j=1}^{a} \frac{\mu + \theta(S)|j-1|}{\mu + \theta(S)|j-1| + \tau}.$$
(5.15)

This leads to the following proposition.

**Proposition 5.3.** The conditional probability that the last order initially in queue at the ask is executed before the mid-price moves, given that it is not canceled (5.14), is given by the inverse Laplace transform of

$$\hat{F}_a^S(\tau) = \frac{1}{\tau} \hat{g}_a^S(\tau) \left( \frac{-\tau}{2\Lambda(S) - \tau} \hat{f}_b^S(2\Lambda(S) - \tau) + \frac{2\Lambda(S)}{2\Lambda(S) - \tau} \right)$$
(5.16)

evaluated at 0, where  $\Lambda(S) = \sum_{i=1}^{S-1} \lambda(i)$ ,  $\hat{g}_a^S(\tau)$  is given by (5.15) and  $\hat{f}_b^S(\tau)$  is given by (5.4).

*Proof.* The first time at which the mid-price changes corresponds to the minimum of  $T^B$ ,  $M^A$  and  $M^B$  defined in last section. Because  $I^A$  and  $T^B \wedge M^A \wedge M^B$  are independent, we may write

$$\hat{f}_{I^A - T^B \wedge M^A \wedge M^B}(\tau) = \hat{f}_{I^A}(\tau) \hat{f}_{T^B \wedge M^A \wedge M^B}(-\tau).$$
(5.17)

Since  $M^A \wedge M^B \sim \exp(2\Lambda(S))$ , and  $T_B \sim f_b^S(t)$ , we can use Lemma 5.1 to find the Laplace transform of the conditional pdf:

$$\hat{f}_{I^A}\hat{f}_{T^B\wedge M^A\wedge M^B}(-\tau) = \hat{g}_a^S(\tau) \left(\frac{-\tau}{2\Lambda(S)-\tau}\hat{f}_b^S(2\Lambda(S)-\tau) + \frac{2\Lambda(S)}{2\Lambda(S)-\tau}\right).$$
(5.18)

If we take S = 1 equation (5.16) becomes  $\frac{1}{\tau} \hat{g}_a^S(\tau) \hat{f}_b^S(-\tau)$  which indeed corresponds with  $\hat{f}_{I^A-T^B}(\tau)$ .  $\Box$ 

A similar expression can be found for the equivalent probability at the bid side. The results presented so far correspond to the ones presented by Cont.

In addition to Cont's work, we will consider the probability that the first limit order in queue at the ask is executed before the mid-price moves. This applies to the situation as described in Section 5.1, where a trader decides to submit a limit order at a better price than the current ask or bid. Again, let  $X_0$  be the known initial order book state with  $X_0^{(p_A)} = a$ ,  $X_0^{(p_B)} = b$  and  $p_S(0) = S$ , and let T be the first time at which the mid-price moves (5.2). Let  $E^A$  be the time at which the first order in queue at the ask is executed. We can write the probability of our interest as

$$P(E^{A} < T|X_{0}, NC^{A}) = P(E^{A} - T < 0|X_{0}, NC^{A}).$$
(5.19)

Given that the first order in queue at the ask is not canceled, it follows a pure death process with death-rate  $\mu$ . Therefore the Laplace transform of  $E^A$  is

$$\hat{g}_a(\tau) = \frac{\mu}{\mu + \tau}.$$
(5.20)

This leads to the following proposition.

**Proposition 5.4.** The conditional probability that the first order at the ask is executed before the mid-price moves, given that it is not canceled (5.19), is given by the inverse Laplace transform of

$$\hat{F}_{a,b}^{S}(\tau) = \frac{1}{\tau} \hat{g}_{a}(\tau) \left( \frac{-\tau}{2\Lambda(S) - \tau} \hat{f}_{b}^{S}(2\Lambda(S) - \tau) + \frac{2\Lambda(S)}{2\Lambda(S) - \tau} \right)$$
(5.21)

evaluated at 0, where  $\Lambda(S) = \sum_{i=1}^{S-1} \lambda(i)$ ,  $\hat{g}_a(\tau)$  is given by (5.20) and  $\hat{f}_j^S(\tau)$  is given by (5.4)

Proof. The proof is similar to the proof of Proposition 5.3

Again a similar expression can be derived for the bid side.

#### 5.4 Probability of Making the Spread

We will now compute the probability of 'making the spread' by submitting a buy order and a sell order at the current ask and bid, given that these orders are not canceled. Recall that  $M^A$  and  $M^B$  are the times at which an ask resp. bid arrives in the spread and  $I^A$  is the time at which all orders that were initially at the ask are executed. In addition, let  $I^B$  be the time at which all orders that were initially at the bid are executed, given that the last in queue is not canceled, and let  $NC^B$  denote the event that there is an order initially at the bid the bid the bid the bid the bid the bid the is never canceled.  $I^A$  and  $I^B$  are independent. The conditional probability of our interest then is

$$P(\max\{I^{A}, I^{B}\} < T|X_{0}, NC^{A}, NC^{B}).$$
(5.22)

One might think that because we assumed that there are orders at the bid and ask that are never canceled, the first time T at which the mid-price moves now corresponds to the first time at which an order arrives in the spread:

$$T = \min\{M^A, M^B\}.$$

In that case, inserting T and separating the cases that either  $M^A < M^B$  or  $M^B < M^A$ , we could write Probability (5.22) as

$$P(\max\{I^{A}, I^{B}\} < \min\{M^{A}, M^{B}\} | X_{0}, NC^{A}, NC^{B})$$
  
=  $P(\max\{I^{A}, I^{B}\} < M^{A}, M^{A} < M^{B}) + P(\max\{I^{A}, I^{B}\} < M^{B}, M^{B} < M^{A}),$   
=  $P(I^{A} < M^{A}, I^{B} < M^{A}, M^{A} < M^{B}) + P(I^{A} < M^{B}, I^{B} < M^{B}, M^{B} < M^{A}).$  (5.23)

Following the same approach as in the preceeding sections, this would lead to an expression for the Laplace transform of the probability that is much simpler than the one presented by Cont et al. [10]. The difficulty here however is that the mid-price does move. We are considering the probability that the number of orders at both the ask and the bid go to zero, before the mid-price moves. However, at the time that the first of both queues reaches zero the mid-price actually changes. It seems therefore impossible to find an easier result than the one presented by Cont et al. for which we refer to Proposition 6 of [10].

#### 5.5 Probability of Reaching the Stop Loss

The probability of reaching the stop loss corresponds to the probability of the mid-price moving away a certain number of ticks. A move of n ticks from the mid-price can be reached by different combinations of events. In Section 5.2 we have already computed the conditional probability that the mid-price increases or decreases one tick at its next move. First consider the probability that the mid-price decreases one tick at its first move, conditional on the initial order book state  $X_0$ :

$$p = P(p_M(t_1) < p_M(t_0)|X_0), \tag{5.24}$$

where  $t_1$  is the first time that the mid-price moves and  $t_0$  the initial time. This probability can be computed using Proposition 5.2. The probability that the mid-price decreases by two at its next two moves, can be written as

$$P(p_M(t_1) < p_M(t_0)|X_0)P(p_M(t_2) < p_M(t_1)|X_1),$$
(5.25)

where  $t_n$  with n = 1, 2, 3, ... represents the  $n^{th}$  time that the mid-price moves and  $X_n$  represents the state of the order book after the  $n^{th}$  mid-price move. Both probabilities in (5.25) can be computed using Proposition 5.2. However, the parameters in the second probability depend on a future (unknown) order book state. Conditional on the initial order book state  $X_0$  and assuming that the first mid-price move was the first event that took place in the order book, this unknown order book state has at least two possible configurations: cancellation of the bid or submission of a cheaper ask with respect to  $X_0$ . The number of possible configurations can be much bigger because e.g. an ask can be submitted on any price level between the bid and ask. Therefore, considering the probability of decreasing n ticks within n mid-price moves already gives over  $2^{n-1}$  possible order books that we have to account for. If we also include the possibilities of the mid-price decreasing nticks in any number of price moves, we get a huge number of possible combinations. We therefore choose to calculate the parameters based on the initial book, and use the same parameters after every step. This means that we assume the initial order book configuration to be a reasonable estimate for the expected configuration of the order book and that

$$p = P(p_M(t_n) < p_M(t_{n-1})|X_{n-1})$$
(5.26)

for any n = 1, 2, 3, ... Using this approach, the probability that the mid-price decreases k ticks can be approximated by

$$\sum_{n=k+2j}^{\infty} \binom{n}{(n-k)/2} (1-p)^{(n+k)/2} p^{(n-k)/2}.$$
(5.27)

where p is given by one minus the Laplace transform of (5.11) evaluated at zero.

To show this, let  $t_n$  for n = k + 2j and with j = 1, 2, 3, ... denote the times at which the mid-price moves, and let  $t_0$  be the initial time. Note that a decrease of k ticks here is only possible if n - k is even. At every time  $t_n$ , the probability p that the mid-price decreases at its next move, is assumed to be given by one minus the inverse Laplace transform of (5.11) in Proposition 5.2, with  $S = p_S(t_0)$ . Using this probability p allows us to derive a density function for the probability that the mid-price decreases k ticks in n mid-price moves. The n price moves follow a Binomial distribution with parameters (n, p). Therefore the probability of the mid-price decreasing k ticks in n price-moves can be written as

$$P(p_M(t_n) = p_M(t_0) - k) = \binom{n}{(n-k)/2} (1-p)^{(n+k)/2} p^{(n-k)/2}.$$
(5.28)

Summing (5.28) over all possible numbers of price moves n gives the desired result.

### Chapter 6

## Implementation

In Chapter 5 we have proposed expressions to compute several probabilities conditional on the current state of the order book. All these expressions are in the form of a Laplace transform whose inverse has to be evaluated at zero to find the corresponding probability. These Laplace transforms are again given as a function of the Laplace transform of the first passage time of the number of orders from *a* to zero. In Chapter 4 we have seen that this second Laplace transform can be written as a continued fraction. In order to compute the probabilities of our interest we therefore need to compute continued fractions and invert Laplace transforms. Implementation and results will be discussed in this chapter.

#### 6.1 Computing Continued Fractions

Recall the Laplace transform of the first probability discussed in Chapter 5: the conditional probability that the mid-price increases at its next move. This probability will be used here to illustrate how the theory presented in Chapters 4 and 5 can be used to numerically compute probabilities in the order book.

Probability	Corresponding Laplace transform
mid-price increase:	$\hat{F}^{S}_{a,b}(\tau) = \frac{1}{\tau} \left( \frac{\tau}{\Lambda(S) + \tau} \hat{f}^{S}_{a}(\Lambda(S) + \tau) + \frac{\Lambda(S)}{\Lambda(S) + \tau} \right) \left( \frac{-\tau}{\Lambda(S) - \tau} \hat{f}^{S}_{b}(\Lambda(S) - \tau) + \frac{\Lambda(S)}{\Lambda(S) - \tau} \right)$

Table 6.1: Probability and corresponding Laplace transform

In the Laplace transform in Table 6.1,  $\tau$  is the variable and  $\Lambda(S)$  is a constant that depends on the order book configuration and thus is given in the condition. In order to compute the probability of the mid-price increasing, the Laplace transform  $\hat{f}_j^S(\tau)$  needs to be computed. The inverse Laplace transform of  $\hat{f}_j^S(\tau)$ corresponds to the first passage time of the number of orders at  $p_A$  or at  $p_B$  from j to zero. In Section 4.3 this was found to satisfy

$$\hat{f}_j^S(\tau) = \prod_{i=1}^j \left( \frac{-1}{\lambda(S)} \prod_{k=i}^\infty \frac{-\lambda(S)(\mu + k\theta(S))}{\lambda(S) + \mu + k\theta(S) + \tau} \right).$$
(6.1)

The continuous fraction in the product is a Jacobi fraction of the form (4.22). Following Section 4.3 this continued fraction is equivalent to the S-fraction

$$\overset{\sim}{\underset{k=i}{\overset{\sim}{\prod}}} \frac{a_k s^{-1}}{1} = \frac{a_i}{\tau + 1} \frac{a_{i+1}}{1 + 1} \frac{a_{i+2}}{\tau + 1} \frac{a_{i+3}}{1 + 1} \cdots,$$
(6.2)

where  $a_k$  satisfies

$$\begin{split} a_i &= \lambda(S)(\mu + i\theta(S)), \\ a_{i+2n} &= \frac{\lambda(S)(\mu + (n+i)\theta(S))}{a_{i+2n-1}}, \end{split} \quad \text{and} \quad \begin{aligned} a_{i+1} &= \lambda(S) + \mu + i\theta(S), \\ a_{2n+i+1} &= \frac{\lambda(S) + \mu + (n+i)\theta(S)}{a_{2n+i}}, \end{split}$$

for n = 1, 2, 3, ... Direct expressions for  $a_{i+2n}$  and  $a_{2n+i+1}$  are easily derived and satisfy (6.5)-(6.6). The S-fraction in (6.1) can then be written as

$$\overset{\sim}{\underset{k=i}{\overset{\scriptstyle}{\overset{\scriptstyle}}}} \frac{a_k}{b_k},$$
(6.3)

with

$$a_i = \lambda(S)(\mu + i\theta(S)), \qquad a_{i+1} = \lambda(S) + \mu + i\theta(S), \qquad (6.4)$$

$$a_{i+2n} = \prod_{k=1}^{n} \frac{\lambda(S)(\mu + (i+k)\theta(S))}{\lambda(S) + \mu + (i+k-1)\theta(S)},$$
(6.5)

$$a_{2n+i+1} = (\lambda(S) + \mu + i\theta(S)) \prod_{k=1}^{n} \frac{\lambda(S) + \mu + (i+k)\theta(S)}{\lambda(S)(\mu + (i+k)\theta(S))},$$

$$b_{i+2(n-1)} = \tau, \quad \text{and} \quad b_{i+2n-1} = 1.$$
(6.6)

This result enables us to compute approximants of the continued fraction in (6.1), and thus compute  $\hat{f}_j^S(\tau)$ , by using Lemma 4.2. A Matlab implementation of this algorithm is included in the appendix. The program output is tested to be equal to the iterative results of computing (6.1). Only 7 approximants are needed in this algorithm to compute the continued fraction in (6.1) up to four digits accuracy. Note that the other Laplace transforms that we have introduced in Chapter 5, such as  $\hat{g}_a^S(\tau)$  given by (5.15), can be computed directly from the order book state on which we condition.

#### 6.2 Numerical Inversion of Laplace Transforms

The next step in computing the conditional probabilities from Chapter 5, is recovering the cdf from the corresponding Laplace transform by computing its inverse Laplace transform

$$f(t) = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} e^{t\tau} \hat{f}(\tau) d\tau.$$
(6.7)

Different numerical methods are available to compute the complicated Bromwich contour integral in (6.7). In [3] Abate and Whitt an inversion method is presented that is based on the idea of using two very different methods so that they can serve as a check for each other. This idea was first introduced by Davies and Martin [11]. The two methods used are easy to understand and shortly described below.

#### Post-Widder method:

According to the Post-Widder theorem [13], f(t) can be expressed as the point-wise limit of

$$f_n(t) = \frac{(-1)^n}{n!} \left(\frac{n+1}{t}\right)^{n+1} \hat{f}^{(n)} \left(\frac{n+1}{t}\right)$$

as  $n \to \infty$ , where  $\hat{f}^{(n)}$  is the  $n^{th}$  derivative of the Laplace transform of f. Following the steps in Jagerman [17] using the Cauchy contour integral, we obtain the integral

$$f_n(t) = \frac{n+1}{t} \frac{1}{2\pi r^n} \int_0^{2\pi} \hat{f}\left(\frac{n+1}{t}(1-re^{tu})\right) e^{-tnu} du.$$

The integral is finite, and using the trapezoidal rule with step size  $\pi/n$  gives the approximation

$$\begin{split} f_n(t) &= \frac{n+1}{2tnr^n} \left\{ \hat{f}\left(\frac{(n+1)(1-r)}{t}\right) + (-1)^n \hat{f}\left(\frac{(n+1)(1+r)}{t}\right) \\ &+ 2\sum_{k=1}^{n-1} (-1)^k \Re\{\hat{f}\}\left(\frac{n+1}{t}(1-re^{\pi ik/n})\right) \right\}. \end{split}$$

The error bound can be shown to satisfy  $|f(t)| \leq r^{2n}$ . For exact derivations see [3] and [17].

Euler method:

The Bromwich integral (6.7) can be rewritten as

$$f(t) = \frac{2e^{at}}{\pi} \int_0^\infty \Re\{\hat{f}(a+iu)\}\cos ut \, du,$$

where  $i = \sqrt{-1}$ . Using the trapezoidal rule with step size h gives

$$f(t) \approx f_h(t) = \frac{he^{at}}{\pi} \Re\{\hat{f}(a)\} + \frac{2he^{at}}{\pi} \sum_{k=1}^{\infty} \Re\{\hat{f}(a+ikh)\} \cos(kht),$$

and with  $h = \pi/2t$  and a = A/2t we get

$$f_h(t) = \frac{e^{A/2}}{2t} \Re\{\hat{f}\left(\frac{A}{2t}\right)\} + \frac{e^{A/2}}{t} \sum_{k=1}^{\infty} (-1)^k \Re\{\hat{f}\left(\frac{A+2k\pi i}{2t}\right)\}.$$
 (6.8)

It can be shown that when f(t) is a cumulative distribution function, for small  $e^{-A}$  the error is approximately bounded by  $e^{-A}$ . To accelerate the numerical calculation of (6.8) with the infinite sum, it is suggested in [25] to use Euler summation. This method can be described as the average of the last m partial sums, weighted by a Bin(1/2, m) distribution. The numerical approximation to (6.8) then becomes

$$E(m, n, t) = \sum_{k=0}^{m} \binom{m}{k} 2^{-m} s_{n+k}(t),$$
(6.9)

where  $s_n$  is the  $n^{th}$  partial sum

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

Typically, m = 11 and n = 15 are taken, increasing n as necessary. For the precise derivation of this expression we refer to Dubner and Abate [12].

In both methods there is a different step for which there is no error bound available. Therefore, the given error bounds of both methods serve as a check for each other. This method is specially useful in the case where f(t) is a cdf, because the fact that  $|f(t)| \leq 1$  for all t can be used in the error analysis. Another feature of this algorithm is that it is intended for computing f(t) at single values of t. Since the probabilities presented in Chapter 5 involve evaluating the inverse of a Laplace transform of a cdf at zero, this method is suitable for the problem. In Cont et al. this method was used to compute probabilities in the order book. Further discussion of the method can be found in [1] and [3].

For this project we choose to not implement the method used by Cont et al., but instead use another interesting method. This method is also based on the relation between a Laplace transform and its Fourier-cosine expansion as presented by Abate and Dubner [12], but is an analogue of the COS method for inverting Fourier transforms presented by Fang and Oosterlee in [21]. They apply the COS method on option pricing, and show that its convergence rate is exponential. A requirement for this method to perform well for Laplace transforms is that the pdf that is to be recovered is sufficiently smooth. A description of how to use this method to find a cdf from its Laplace transform is given below.

To apply this method, it is necessary to have an expression for the Laplace transform of the pdf. In the case of our problem, all the probabilities are given in terms of the Laplace transform of the cdf. Thanks to (5.7) the corresponding pdf expression is easily obtained. The COS method was implemented in order to obtain the results in the next section. The corresponding Matlab code can be found in the appendix. The Symbolic Mathematics Toolbox from Matlab was used for this. It should be noted that this method has not been tested on precision and convergence speed so far. A first step in a follow up study would be to test the performance of inversion methods and explore available alternatives to numerically invert Laplace transforms.

COS method:

As seen in the Euler method, rewriting the Bromwich integral in (6.7) by use of the trapezoidal rule gives

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

A requirement for the COS method to be successful is that the infinite integral in (6.7) can be truncated such that

$$\hat{f}_1(\tau) = \int_a^b e^{i\tau t} f(t) dt \approx \int_0^\infty e^{i\tau t} f(t) dt = \hat{f}(\tau).$$
(6.10)

If we now choose a and b such that (6.10) is satisfied and substitute  $h=\pi/(b-a)$  and  $\gamma=0,$  we find

$$f_h(t) = \frac{1}{b-a} \Re\{\hat{f}(0)\} + \frac{2}{b-a} \sum_{k=1}^{\infty} \Re\{\hat{f}\left(\frac{ik\pi}{b-a}\right)\} \cos(k\pi \frac{t-a}{b-a}).$$
(6.11)

Truncating the infinite sum as well, we find the following numerical approximation of f at t.

$$f_h(t) = \frac{1}{b-a} \Re\{\hat{f}(0)\} + \frac{2}{b-a} \sum_{k=1}^{N-1} \Re\{\hat{f}\left(\frac{ik\pi}{b-a}\right)\} \cos(k\pi \frac{t-a}{b-a}).$$

This corresponds to the method presented in [21]. In order to use this method to recover a cdf from its inverse, we compute the value of  $f_h(t)$  for a range of t, and take the cumulative sum. The approximation for the cdf then becomes

$$F_{h}(t) = \sum_{j=0}^{M-1} \left( \frac{1}{b-a} \Re\{\hat{f}(0)\} + \frac{2}{b-a} \sum_{k=1}^{N-1} \Re\{\hat{f}\left(\frac{ik\pi}{b-a}\right)\} \cos(k\pi \frac{j\,dt}{b-a}) \right),$$
(6.12)

where dt = (b - a)/M and M is the chosen number of discretization steps. The resulting error consists of the truncation of the infinite sum in (6.11), the truncation error in the infinite integral (6.10), and the truncation of the integral over the pdf that is used to obtain the approximation of the cdf (6.12). For error analysis we would like to refer to [12] and [21].

#### 6.3 Results

In Chapter 3 we have compared the average behavior of a simulated order book to the average behavior of an observed order book. Although this is not of interest in practice, this showed that the model gives a reasonable approximation of the average behavior. In order to asses the performance of its short-term prediction, we will compare the probability that the number of orders at distance i from the opposite best quote increases to probabilities that are empirically observed in the Vodafone data.

Let  $Q_i(t_m)$  be the number of orders at a particular price level at distance *i* from the opposite best quote at time  $t_m$ , and define the probability that this number increases by one at its next change as

$$P(Q_i(t_{m+1}) = n+1 | Q_i(t_m) = n) = \begin{cases} \frac{\lambda(S)}{\lambda(S) + \mu + n\theta(S) + \lambda(0)}, & i = S, \\ \frac{\lambda(i)}{\lambda(i) + n\theta(i)}, & i > S, \end{cases}$$
(6.13)

where  $t_{m+1}$  represents the next time at which the number of orders at distance *i* changes. The expressions on the right hand side can be understood intuitively as follows. For a price level *p* at distance *i* > *S* two types of events can occur: arrivals of limit orders and cancellations of existing orders. An increase of the number of orders at *p* takes place with rate  $\lambda(i)$ , and a decrease with rate  $n\theta(i)$ . The probability that a limit order arrives first therefore is  $\lambda(i)/\lambda(i) + n\theta(i)$ .



Figure 6.1: Probability of increase of the number of orders at distance i from the opposite best quote, as a function of the number of orders n, with spread S = 1.

Figures 6.1 - 6.3 show the results for spreads S = 1, 2, 3, at different distances *i* from the opposite best quote. The results are obtained by, for every spread S = 1, 2, 3 and for price levels at *i* ticks from the bid, counting the percentage of times that the number of orders in queue increased, depending on the queue sizes. This same approach was used in [10], our results however are less good than the ones presented by Cont et al. [10]. Although the shapes of the observed probabilities are similar to the ones obtained from the model, there seems to be a parallel shift. An interesting observation is that the observed probabilities at queue size 0 is less than one everywhere. Given that there are 0 orders in queue, the cancellation rate should be zero because it is proportional to the number of orders. The probability of the number of orders increasing then equals one. The observed probabilities however are different. An explanation for this can be found in the assumption of all orders being unit size. Because of this assumption, a small number of orders in the order book is interpreted as zero. However, cancellations can still be observed at the corresponding price level. This result indicates that the parameter estimation based on the assumption of all orders having size one does not give a good description of the short-term dynamics of the order book.



Figure 6.2: Probability of increase of the number of orders at distance i from the opposite best quote, as a function of the number of orders n, with spread S = 2.



Figure 6.3: Probability of increase of the number of orders at distance i from the opposite best quote, as a function of the number of orders n, with spread S = 3.

Next we will compute the probability that the mid-price increases at its next move, conditional on the state of the order book. For this we use Proposition 5.2 and implementation of the COS-method explained in the last section. Note that it is not possible to back-test the order book model proposed here, because every action suggested by it has an influence on the order book, and thus changes the order book. Table 6.2 shows the results for the probability of the mid-price increasing. The initial order book configuration on which we condition is given by the numbers of orders at the ask a, the number of orders at the bid b, and the spread S = 1. The upper table shows the probabilities that are empirically observed in the Vodafone data. Using direct simulation, the same probabilities in the model and their corresponding 95%-confidence interval were computed. The results are shown in the table in the middle. Finally, the same probabilities were computed using the Laplace method. The corresponding results are shown in the bottom table.

	a = 1	a=2	a = 3	a = 4	a = 5
b=1	0.503	0.377	0.273	0.215	0.167
b=2	0.625	0.482	0.388	0.334	0.270
b=3	0.700	0.584	0.486	0.422	0.345
b=4	0.761	0.647	0.561	0.491	0.402
b=5	0.780	0.694	0.607	0.546	0.443

	a = 1	a=2	a = 3	a = 4	a = 5
b = 1	$0.501 \pm 0.003$	$0.338 \pm 0.003$	$0.270\pm0.003$	$0.231 \pm 0.003$	$0.203 \pm 0.003$
b=2	$0.661 \pm 0.003$	$0.500\pm0.003$	$0.417 \pm 0.003$	$0.359 \pm 0.003$	$0.325\pm0.003$
b=3	$0.732 \pm 0.003$	$0.588 \pm 0.003$	$0.502\pm0.003$	$0.442\pm0.003$	$0.403\pm0.003$
b=4	$0.769 \pm 0.003$	$0.638 \pm 0.003$	$0.558 \pm 0.003$	$0.499 \pm 0.003$	$0.459 \pm 0.003$
b=5	$0.796 \pm 0.003$	$0.676 \pm 0.003$	$0.597 \pm 0.003$	$0.541 \pm 0.003$	$0.501 \pm 0.003$

	a = 1	a=2	a = 3	a = 4	a = 5
b = 1	0.516	0.386	0.297	0.225	0.235
b=2	0.616	0.458	0.441	0.365	0.307
b=3	0.679	0.550	0.506	0.415	0.370
b=4	0.704	0.665	0.623	0.455	0.403
b=5	0.652	0.509	0.670	0.631	0.061

Table 6.2: Probability of the mid-price increasing at its next move conditional on a orders at the ask, b orders at the bid and spread S = 1. top: empirical frequencies observed in data, middle: direct simulation results, bottom: Laplace transform results.

Unlike the results presented in [10], the results in the upper two tables show that only a few of the empirically observed probabilities lay in the 95%-confidence interval simulated by the model. This means that the probability of the mid-price increasing is generally not captured well enough by the model. Therefore, the model as proposed here cannot be used in trading yet. It should first be adapted to give good approximations of short-term probabilities. In the next chapter, ideas for possible improvements of the model will be treated.

Given that the results from both the direct simulation and the Laplace method are based on the same model, we use the results in the middle table as a test for the numerical inversion method. The results at the bottom of Table 6.2 however do not really correspond to the ones obtained by direct simulation. Since the approximation of the S-fraction is found to be accurate, this result shows that also the numerical method to invert Laplace transforms needs to be improved. Further study of numerical inversion of Laplace transforms is out of the scope of this project. The Matlab code used for the results in the last table can be found in the appendix.

### Chapter 7

# **Conclusion and Recommendations**

#### 7.1 Summary

In the first part of this project we researched the world of trading and limit order books. The behavior of limit order books was studied based on real market data, information from traders and observations from the trading floor. Next to that, the stochastic model for the dynamics of the limit order book proposed by Cont et al. [10] was studied. This model was calibrated to high-frequency data taken from the London Stock Exchange. We have compared the average behavior of a simulated order book to empirically observed average behavior using direct simulation. Several changes to the model and its parameter estimation were proposed and tested thoroughly. The different estimation approaches were used to calibrate the model and the resulting simulated order books were again compared to average properties of the order book data. Finally, a slightly adapted version of Conts' model was chosen. This model reasonably approximates the average shape of the order book and the variance of this shape.

In the second part of this project, the analytical tools were presented that allow to compute short-term probabilities in the model order book. Although these tools were also used by Cont et al., they were not presented as such. Therefore, a literature study on Continuous Fractions and their use in Laplace transforms was done. Using the Laplace transform methods that were briefly explained by Cont, we showed how to compute short-term probabilities in the order book. We expanded the set of probabilities that were computed by Cont and added extended proofs. Finally, we explained how the Laplace method introduced in Chapter 5 can be implemented in order to compute the probabilities. Therefore it is necessary to compute a continuous fraction as well as to invert a Laplace transform. For this second task, a new method was proposed and implemented. Finally, some results were computed and discussed.

#### 7.2 Recommendations for Further Research

In this section, possible improvements to the model will be discussed, as well as issues that need to be explored before the model can be used. We will start by pointing out how the performance of the model could be improved based on the assumptions and the parameter estimation.

As mentioned before, the assumption of all orders having size one was one of the points questioned by traders. This assumption was taken from [10] because it was claimed that the order-size distribution plays a minor role in the average order book. In Section 6.3 however we found that this assumption leads to a discrepancy between the simulated and the observed order book when it comes to short term probabilities. We therefore recommend to include a size-distribution in the model. For this, the work of Bouchaud [8] can serve as an example.

The parameter estimation offers several possibilities to improve the model. First of all, in Chapter 3 we already mentioned that the proposed parameter estimation does not take trends like e.g recurring orders into account. In reality however, recurring orders such as ice bergs form a significant part of the incoming order flow. Ice bergs are huge orders that are divided into smaller orders of equal size, and are subsequently submitted to the order book. It is therefore recommended to explore how the parameters can be adapted in order to account for such orders. One could for example think about increasing the arrival rate at a certain

price level after observing subsequent limit order arrivals at a certain level. What sort of recurring behavior needs to be included in the order book should be discussed with traders.

Next to that, we have seen that although the parameter estimations can be quite different from day-to-day, the parameter curves have a similar shape. At the end of the month the parameters seem to have higher values. Mondays are e.g. generally more quiet than Fridays, and the highest turn over is almost always reached on the expiry day, which is the third Friday of the month. It is therefore recommended to perform a statistical analysis on the influence of the particular trading day of the month on the parameters. A data set over several months is necessary. Another interesting idea is to estimate the parameters based on only data at certain frequencies. Because many players in the markets can only submit orders at a certain frequency due to limitations of their machines, estimating the parameters based on certain frequencies might give interesting results.

Once a satisfactory estimation method for the parameters is found, the stability of the parameters should be explored. What is important to determine is what the required length of the data set is in order to obtain stable estimates. Also, a measure indicating that the parameters should be re-estimated should be determined. Furthermore, Figure 3.11 showed that the shape of the order book can vary heavily from day to day, not to mention from event to event. Therefore, the mean reverting behavior of the order book shape that was pointed out by traders is not included in this model. This could however be a consideration for future research.

The next recommendations apply to the second part of this report. First of all it is clear that a the Laplace inversion method should be studied more carefully. The method presented in Section 6.2 could be improved, however more recent developments might offer better solutions. This was outside the scope of this project and can be an important part of a follow up research. Also, the computation of Stieltjes-fractions and should be further explored.

After completing this, the loop introduced in Figure 1.2 should be closed. In order to do so, it should be studied how knowledge about short-term price moves would exactly influence trader's strategies, and which precise information is valuable for traders. For this, close collaboration with trading is required. This very much depends on the strategy and its interplay with the computed probabilities. An example of this is to compute probabilities in two order books simultaneously, based on the correlation between both. The move of the mid-price in one order book then influences the probability of the mid-price increasing in the other.

The last step in closing the loop is to include the own impact in the model. Although the arrival rates of events in the model order book already include the current order book state, and thus the impact of the last event, this is still a difficulty. The model assumes that the statistical short-term behavior of the order book is not affected by a market participant who starts implementing a certain trading strategy based on this model. If the short-term behavior were to change, the parameters should be re-estimated. Because of the own impact on the order book, it is not possible to back-test this model and find out what the own impact is without risking losses. It is therefore recommended to carefully develop a method to test the model and apply it for a small portfolio to determine its own impact.

Research on order books so far has mainly focused on algorithmic trading or algo trading. Algo trading means trading by computer programs that follow a certain trading strategy and react upon events in the market. Algo trading is very fast and therefore probabilities on very-short terms are mainly of use in this type of trading. However, also for market participants who trade by manually adapting their parameters to follow the market this model can be useful. In order to decide when the model can be applied, it is necessary to determine how volatile a stock should be for the model to give good results, and for which daily turn over. So a quantification of the required liquidity is needed before this model can be apply.

Algo trading is now estimated to cover over 50% of the transactions in the US and over 30% of the transactions in Europe for stocks. As the global financial markets continue to set record volumes and average trade sizes continue to decline significantly, algorithmic trading is expected to grow even more in the nearby future. It is therefore of great importance for trading firms and other institutions participating in the financial markets to understand the driving dynamics of order books.

#### 7.3 Conclusion

From the results presented in Chapter 3 it is clear that the proposed order book model does not exactly approximate the long-term behavior of the order book in terms of numbers of events and trades per day. Also the short-term probabilities were not accurately approximated as shown in Section 6.3. It is not the purpose of Applied Mathematicians to make a 100% fitting model of reality. Next to the fact that it is almost always impossible to track all the parameters that influence a certain system, an increasing number of parameters also increases the complexity of the model. Therefore, mathematical modeling is a trade-off between reality and functionality. What we mean by this is that an inaccurate approximation of some properties by a model is not necessarily a problem, as soon as the model does appropriately capture the properties of interest.

Obviously, the power of the model presented here lies in the fact that it is both easy to calibrate and analytically tractable. Its analytical tractability is solely based on the assumption that the number of orders at any price level follows a birth-death process. This offers room for improvement of the performance of the model without losing its nice analytical properties, by changing the other assumptions or the parameter estimation. Some ideas for this were explained in the last section. Therefore, although it may not be true, this assumption is very functional. The purpose of modeling order book dynamics in this case is to compute short-term probabilities, conditional on the current order book. If it would however result from future research that it is not possible to improve the approximation of short-term probabilities that we actually want to use this model for, it would be unwise to use this model for trading.

## Appendix

#### Matlab Code: Computing Continued Fractions

```
function K = cont_fraction(x,S,s)
%%% compute 10th approximant
\%\% passage time from x to x-1 orders
            AO = 1;
            A1 = 0;
            BO = O;
            B1 = 1;
            for n=1:10
                A = vpa(b(x+n-1,x,s))*A1 + a(x+n-1,x,S)*A0;
                B = vpa(b(x+n-1,x,s))*B1 + a(x+n-1,x,S)*B0;
                approximant = (A1*(s)+A)/(B1*(s)+B);
                AO = A1; A1 = A; BO = B1; B1 = B; n = n+1;
            end;
            K = approximant;
function a = a(n, x, S)
            n = n - x;
            if n==0
                prod = lambda(S)*(mu(1)+x*theta(S));
            elseif n==1
                prod = lambda(S)+mu(1)+x*theta(S);
            elseif mod(n,2) == 0\% even
                n = n/2;
                prod = 1;
                for k=1:n
                    prod = prod*(lambda(S)*(mu(1)+(x+k)*theta(S)))/...
                      (lambda(S)+mu(1)+(x+k-1)*theta(S));
                end;
            else
                             % odd
                n = (n-1)/2;
                prod = lambda(S)+mu(1)+x*theta(S);
                for k=1:n
                    prod = prod * (lambda(S)+mu(1)+(x+k)*theta(S))/...
                      (lambda(S)*(mu(1)+(x+k)*theta(S)));
                end;
            end:
            a = prod;
function b = b(n, x, s)
            n = n-x;
            if mod(n,2) == 0
                               % even
                prod = (s);
```

```
else % odd
  prod = 1;
end;
b = prod;
```

#### Matlab Code: Inverting Laplace Transforms

```
%%% Compute probability of mid-price increasing
%%% by inverting Laplace transforms
%%% for a,b=1:5
for a = 1:5
   for b=1:5
syms tau
S = 1;
L = sum(lambda(1:S));
N = 100;
                        \% # truncation steps infinite sum
tb = 5;
                        % truncation bound integral
ta = -5;
                        % truncation bound integral
f = [];
dt = 0.01;
        fa = 1;
                        \% compute first passage time to zero at ask
        for i=1:a
            K1 = cont_fraction(i,S,tau+L);
            fa = fa*-1/lambda(S)*K1;
        end:
        fb = 1;
        for j=1:b
                        % compute first pasage time to zero at bid
            K2 = cont_fraction(j,S,L-tau);
            fb = fb*-1/lambda(S)*K2;
         end;
f_hat = ((tau/(tau+L))*fa + (L/(L+tau)))*(-tau/(L-tau)*fb + (L/(L-tau))); % LT of pdf
tau = 0;
f_hat0 = subs(f_hat);
k = [1:N-1];
tau = sqrt(-1)*k*pi/(tb-ta);
t = [ta:dt:tb];
for i=1:(tb-ta)/dt+1
    f(i) = 1*dt/(tb-ta)*real(f_hat0) + 2*dt/(tb-ta)*sum(real(subs(f_hat)).*cos(k*pi*(t(i))));
end;
F = cumsum(f);
F((tb-ta)/(2*dt)+1)
    end;
```

end;

# Bibliography

- J. ABATE AND W. WHITT, The Fourier-series method for inverting transforms of probability distributions, Queueing Systems, 10 (1992), pp. 5–88.
- [2] —, Computing Laplace transforms for numerical inversion via continued fractions, INFORMS Journal on Computing, Vol.11 No.4 (1999), pp. 394–405.
- [3] —, Numerical inversion of Laplace transforms of probability distributions, INFORMS Journal on Computing, Vol.11 No.4 (1999), pp. 394–405.
- [4] Y. AIT-SAHALIA, P. A. MYKLAND, AND L. ZHANG, Ultra high frequency volatility estimation with dependent micro-structure noise, Deutsche Bundesbank, Research Centre - Discussion Paper Series 1: Economic Studies, Vol.30 (2005).
- [5] G. BAKER AND P. GRAVES-MORRIS, Encyclopedia of Mathematics and its Applications: Padé Approximants, 2nd Edition, Cambridge University Press, 1996.
- [6] B. BIAIS, P. HILTON, AND C. SPATT, An empirical analysis of the limit order book and the order flow in the Paris Bourse, Journal of Finance, Vol.50 (1995), p. 1655.
- [7] B. BIAIS AND P.-O. WEILL, Liquidity shocks and order book dynamics, Toulouse School of Economics/University of California Los Angeles and NBER, (2009).
- [8] J.-P. BOUCHAUD, M. MÉZARD, AND M. POTTERS, Statistical properties of stock order books: Empirical results and models, Quantitative Finance, Vol.2 (2002), pp. 251–256.
- [9] A. BOVIER, J. ČERNÝ, AND O. HRYNIV, The opinion game: Stock price evolution from microscopic market modeling, International Journal of Theoretical and Applied Finance, Vol.9 (2006), pp. 91–111.
- [10] R. CONT, S. STOIKOV, AND R. TALREJA, A stochastic model for order book dynamics, IEOR Dept, Columbia University, New York, (2008).
- [11] B. DAVIES AND B. L. MARTIN, Numerical inversion of Laplace transforms: A critical evaluation and review of methods, Journal of Computational Physics, 33 (1970), pp. 1–32.
- [12] H. DUBNER AND J. ABATE, Numerical inversion of Laplace transforms by relating them to the finite Fourier cosine transform, JACM, 15 (1968), pp. 115–123.
- [13] W. FELLER, An Introduction to Probability Theory and its Applications, Vol.II, 2nd edition, Wiley, New York, 1971.
- [14] T. FOUCAULT, O. KADAN, AND E. KANDEL, Limit order book as a market for liquidity, Review of Financial Studies, Vol.18 No.4 (2001), pp. 1171–1217.
- [15] B. HOLLIFIELD, R. A. MILLER, AND P. SANDAS, An empirical analysis of limit order markets, Review of Economic Studies, Vol.71 No.4 (2004), pp. 1027–1063.
- [16] L. JACOBSEN, General convergence of continued fractions, Transactions of the American Mathematical Society, Vol.294 No.2 (1986), pp. 477–485.
- [17] D. L. JAGERMAN, An inversion technique for the Laplace transform, Bell System Technical Journal, 61 (1982), pp. 1995–2002.

- [18] J. KEILSON, Markov Chain Models Rarity and Exponentiality, Springer-Verlag, 1979.
- [19] L. LORENTZEN AND H. WAADELAND, Continued Fractions, second edition, Vol 1: Convergence Theory, Atlantis Studies in Mathematics for Engineering and Science, 2008.
- [20] N. MADRAS, Fields Institute Monographs: Lectures on Monte Carlo Methods, American Mathematical Society, Providence, Rhode Island, 2002.
- [21] C. OOSTERLEE AND F. FANG, A novel pricing method for European options based on Fourier-cosine series expansions, Munich Personal RePEc Archive, 9319 (2008).
- [22] C. A. PARLOUR, Price dynamics in limit order markets, Review of Financial Studies, Vol.11 No.4 (1998), pp. 789–816.
- [23] S. M. Ross, Stochastic Processes, second edition, Wiley, 1996.
- [24] I. ROSU, A dynamic model of the limit order book, Review of Financial Studies, Vol.22 No.11 (2009), pp. 4601-4641.
- [25] R. M. SIMON, M. STROOT, AND G. H. WEISS, Numerical inversion of Laplace transforms with application to percentage labeled experiments, Journal of Biomedical Informatics, 6 (1972), pp. 596–607.
- [26] E. SMITH, J. D. FARMER, L. GILLEMOT, AND S. KRISHNAMURTHY, Statistical theory of of the continuous double action, Quantitative Finance, Vol.3 No.6 (2003), pp. 481–514.
- [27] T. STIELTJES, Recherches sur les fractions continues, Annales de la Faculté des Sciences de Toulouse, (1894).
- [28] I. ZOVKO AND J. FARMER, The power of patience; a behavioral regularity in limit order placement, Quantitative Finance, 2 (2002), pp. 387–392.