

Spring 5-2017

# Statistical Analysis of Markovian Queueing Models of Limit Order Books

Yiyao Luo

*Washington University in St. Louis*

Follow this and additional works at: [https://openscholarship.wustl.edu/art\\_sci\\_etds](https://openscholarship.wustl.edu/art_sci_etds)



Part of the [Probability Commons](#), and the [Statistical Models Commons](#)

---

## Recommended Citation

Luo, Yiyao, "Statistical Analysis of Markovian Queueing Models of Limit Order Books" (2017). *Arts & Sciences Electronic Theses and Dissertations*. 1076.

[https://openscholarship.wustl.edu/art\\_sci\\_etds/1076](https://openscholarship.wustl.edu/art_sci_etds/1076)

This Thesis is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact [digital@wumail.wustl.edu](mailto:digital@wumail.wustl.edu).

Washington University in St. Louis

Department of Mathematics

Statistical Analysis of Markovian Queueing Models of Limit Order Books

by

Yiyao Luo

A thesis presented to  
The Graduate School  
of Washington University in  
partial fulfillment of the  
requirements for the degree  
of Master of Arts

May 2017  
Saint Louis, Missouri

copyright by

Yiyao Luo

2017

# Contents

- List of Tables . . . . . **iv**
  
- List of Figures . . . . . **v**
  
- Acknowledgments . . . . . **vi**
  
- Abstract . . . . . **vii**
  
- 1 Introduction . . . . . 1**
  
- 2 Model Setup and Comparison of the Two Models . . . . . 5**
  - 2.1 Model Setup . . . . . 5
  - 2.2 Review of Huang’s Model setting . . . . . 8
  - 2.3 Review of Cont’s Model setting . . . . . 11
  - 2.4 Brief Comparison of Two Models . . . . . 14
  
- 3 Numerical Results . . . . . 16**
  - 3.1 Comparison of Estimators . . . . . 16
  - 3.2 Multiple-Size Rates and Other Assumption . . . . . 23

<b>4 Conclusion . . . . .</b>	<b>30</b>
References . . . . .	31

# List of Tables

3.1	Estimates of rate parameters in multiples of $S_m$ per second . . . . .	18
3.2	Estimates of rate parameters in number of 100 shares per second . . . . .	19
3.3	Percentage of new outstanding queues with specific initial volume when spread is one . . . . .	20
3.4	Percentage of time with a given spread . . . . .	23
3.5	Percentage of orders on level I with given volumes (spread=1) . . . . .	24
3.6	Estimates of multiple-size rate parameters . . . . .	25
3.7	The probability of either side to shrink the spread condition on a certain price change . . . . .	26

# List of Figures

3.1	Histogram of volumes of the market orders . . . . .	17
3.2	Histogram of volumes of the existing limit orders on the level I . . . . .	18
3.3	CDF of first-passage time to state 0 with initial volume of 100 shares . . . . .	21
3.4	CDF of first-passage time to state 0 with initial volume of 100 shares . . . . .	22
3.5	CDF of first-passage time to state 0 with initial volume of 100 shares . . . . .	25
3.6	Density of volumes on either side if bid orders insert spread after a price increase	28
3.7	Density of volumes on either side if ask orders insert spread after a price increase	28
3.8	Density of volumes on either side if bid orders insert spread after a price decrease	29
3.9	Density of volumes on either side if ask orders insert spread after a price decrease	29

# Acknowledgments

I am deeply grateful for the guidance of my advisor and committee members throughout the process of writing this thesis. I would like to thank my advisor, Dr. José Figueroa-López, for his willingness to work with me even before I finished all the course requirements for a Master's degree. I greatly appreciate the freedom and support he provided me in searching a thesis topic. And I thank him for the great amount of time he invested in working with me and valuable insight he provided into my work. I am also thankful for his patience and encouragement every time I encountered difficulties and even felt depressed in writing this thesis. I could not have had a better mentor than Dr. Figueroa-López. I also would like to thank my committee members, Dr. Kuffner and Dr. Figueroa-López, for their knowledgeable advice, time and effort.

Last but certainly not least, I want to thank my parents and all my family members. Without their support, I could never imagine the achievements I have made so far.

Yiyao Luo

*Washington University in Saint Louis*  
*May 2017*



## ABSTRACT OF THE THESIS

Statistical Analysis of Markovian Queueing Models of Limit Order Books

by

Yiyao Luo

Master of Art in Statistics

Washington University in St. Louis, May 2017

Research Advisor: Professor José Figueroa-López

The objective of this thesis is to investigate the suitability of some Markovian queueing models in being able to effectively describe the dynamical properties of a limit order book more specifically. We review and compare the assumptions proposed by Huang et al. [Quantitative Finance, **12**, 547-557 (2012)] and Cont et al. [SIAM Journal for Financial Mathematics, **4**, 1-25 (2013)], and estimate the intensity parameters in both ways, based on real data of a stock on the Nasdaq Stock Market. Through comparing by cumulative distribution functions of first-passage time to state 0, we will show that the estimators of Cont's model fit our data better and we put forward the assumption of multiple-size rates as a better alternative to Cont's frame work. At last, we investigate the stationary joint distribution of volumes on either side after each price change.

# Chapter 1

## Introduction

Electronic markets hold an increasing number of financial instruments such as equity, options and futures and, especially, a substantial volume of them are traded in high frequency. In electronic markets, selling and buying participants centralize their orders in a limit order book. A limit order book is divided into two sides called ask and bid. More specifically, participants who want to buy stock shares put orders on the bid side, and those who want to sell their shares put orders on the ask side. In a limit order book based market, there are three types of orders on either the ask or bid side, named limit order, market order, and cancellation.

Limit orders can be placed with a specified price and with any volumes. Generally, the prices for limit orders placed on the ask side should be strictly higher than those on the bid side. In this way, the limit orders on the ask side with the lowest price and the limit orders on the bid side with the highest price are called level I orders or outstanding orders. Also, we define the spread as the price difference between the outstanding orders on the two sides, so the spread should be always no less than 1 tick, where tick is the unit of stock price in a limit order book, typically, 1 cent in US market.

$$spread = P_{\text{level I ask}} - P_{\text{level I bid}} \quad (1.1)$$

Moreover, when the spread is larger than 1 tick, it can be expected that limit orders on either sides will be placed into and then shrink the spread.

The first way for orders in a limit order book to be depleted is called execution. When a market order comes into the book, it will firstly execute the outstanding limit orders on either sides depending on the type (ask or bid) and then continue to execute the level II limit orders if the limit orders of level I are all executed. In that case, the limit orders of level II become to level I. Therefore, on the one hand, participants can only place market orders with specific volumes but without certain prices; on the other hand, only the outstanding limit orders can be depleted by market orders.

The participants who placed limit orders also allowed to cancel them all or partly, which is another way to deplete orders in a limit order book, named cancellation. That implies that cancellations could be set with certain prices and volumes. In other words, limit orders of any level can be depleted by cancellations.

Based on the three types of orders mentioned above, a limit order book is defined on a fixed discrete grid of price levels and varying number of shares from limit orders standing on each level Cartea, Jaimungal, and Penalva (2015). Practically, we are interested in the midprice of a limit order book, as an approximation of the true underlying price of the asset, which is the arithmetic average of the outstanding price levels on the ask and bid side.

$$p_{mid} = \frac{P_{\text{level I bid}} + P_{\text{level I ask}}}{2} \quad (1.2)$$

Driven by the interest of midprice, we only investigate the statistical features of orders on the outstanding price level or the level I. We treat the outstanding limit orders as two dynamical queues on the ask and bid side. The limit orders placed on level I increase the queues, and market orders and the cancellations on the level I decrease the queues. In this way, we can treat the two queues as two Markovian queueing models with a state given as the number of shares on the level I. Each time the outstanding ask queue gets depleted, the midprice will move upwards since level II ask queue becomes level I. Hence a main motivation for analyzing high-frequency dynamical properties of a limit order book is to make a short-term prediction of the price changes based on the current state of the book. Therefore, it is of interest to model the outstanding queues with valid and apposite assumptions

Various studies have attempted to explain the process of price changing in limit order book. From a comprehensive perspective, Gould et al. (2013) propose that there are three mainstream directions being discussed. Perfect rationality models presume deterministic orders flows which enable us to obtain optimal strategy among many feasible strategies. The second direction is zero intelligence models, which assume that orders flows are stochastic processes and allow to estimate parameters and measure the predictive power of models. Agent based models is the combination of the other two models and postulate a large number of possibly heterogeneous agents interact in a certain way. Equilibrium models proposed by Parlour (1998), T. Foucault and Kandel (2005) and Rosu (2009) focus on the price formation process, however they are intractable and not easily to use in the application due to the unobservable parameters. Also, some statistical properties of limit order books proposed by empirical studies like J. Bouchaud and Potters (2002), Farmer, Gillemot, Lillo, Mike, and Sen (2004) and Hollified, Miller, and Sandas (2004) are incompatible in a single model.

Cont, Stoikov, and Talreja (2010) firstly propose a continuous-time Markovian queueing model to describe the dynamic properties of a limit order book. The arrival of limit orders, market orders and cancellation of existing orders are viewed as independent Poisson processes. Cont et al. (2010) postulate that the orders arrive in a unit size measured as the average size of limit orders and that the cancellation rate are proportional to the volume of existing orders. Also, according to a power law, where the arrival rates of those orders depend on the distance of their price levels to the outstanding level on the other side.

This thesis follows the model setting of Cont et al. (2010), by viewing the level I orders in a limit order book as two Markovian queues. Based on such a setting, we further investigate the assumptions on orders to better describe the behavior of level I queues. HHuang and Kercheval (2012) and Cont and de Larrard (2013) put forward two distinct perspectives to estimate the parameters of the queues. We aim to compare the two estimation procedures by first computing the cumulative distribution function of first-passage time of the queues to deplete and then to compare with the empirical distribution based on data from the Nasdaq Stock Market. Consequently, we aim to come up with better estimators of the parameters, which has closest cumulative distribution function to the empirical one.

The outline of this thesis is as follows. Chapter 2 introduces the Markovian queueing model on a limit order book and statistical feature to describe the model, and then reviews the assumptions listed in Huang and Kercheval (2012) and Cont and de Larrard (2013), and make a comparison. In Chapter 3, we estimate the rate parameters by estimators proposed by both papers, and compare the cumulative distribution function of the first-passage time to state 0 of the estimates. Then we propose a better assumption of multiple-size rates. Finally, we also investigate the joint distribution of volumes on either side after price changes. Chapter 4 concludes.

# Chapter 2

## Model Setup and Comparison of the Two Models

In this chapter, we first introduce a Markovian queueing model and set up the model in the context of a limit order book, and then list essential parameters in our model and introduce first-passage time of the Markovian queue, an important statistical feature of our model.

To compare the methods of Huang and Kercheval (2012) and Cont and de Larrard (2013) for describing the queues, we review the model setting and assumptions of the two papers. Next, we summarize the similarity and difference of their methods.

### 2.1 Model Setup

The Markovian queueing model we use in this thesis has state  $Z^+ = \{0, 1, 2, \dots\}$ , where state 0 means the queue gets depleted. Since a limit order book can be visualized as a price grid with limit orders standing on and the volumes of those orders increase or decrease because of arrival of the three types of orders, we can model the limit orders on each price level as a Markovian queue.

More specifically, the Markovian queueing model can be modeled as a birth-and-death process, denoted by  $X(t)$  whose value is called the state at time  $t$ . Because of new arrival of limit orders, the limit orders on a specific price level increase, which implies that the queue  $X(t)$  standing for that price reaches a higher state by “births”. Meanwhile, the arrivals of market orders and cancellations, as “deaths”, drive the queue  $X(t)$  to reach a lower state. To describe this process, we need two essential parameters: birth rate and death rate. Under the context of a limit order book, the birth rate dictates the arrival rate of limit orders and the death rate decides the arrival of market orders and cancellations. The arrival rate of orders means the number of one type of orders with specific shares arrive at ask side or bid side per second. Hence we have different rates of orders representing various volume of shares to force the queue reach different states. We denote  $\lambda_n$  the arrival rate of limit orders with size  $n$  units on either side, and  $\mu_n, \theta_n$  for markets orders and cancellations, respectively. Since the outstanding limit orders always get depleted to force the price of stock to change, by the theory of birth-and-death process, the rate of limit orders should be less than the sum of those rates market orders and cancellations. Moreover, if each order arrive independently, then the probabilities for  $X(t)$  moving from one state  $i$  to another state  $i + n$  or state  $i - n$  for  $0 \leq n \leq i$  are given as follows: if  $X(t)$  stands on level  $I$ :

$$\begin{aligned}\mathbb{P}[X(t+1) = i+n | X(t) = i] &= \frac{\lambda_n}{\sum_n (\lambda_n + \mu_n + \theta_n)} \\ \mathbb{P}[X(t+1) = i-n | X(t) = i] &= \frac{\mu_n + \theta_n}{\sum_n (\lambda_n + \mu_n + \theta_n)}\end{aligned}\tag{2.1}$$

if  $X(t)$  doesn't stand on level  $I$ :

$$\begin{aligned}\mathbb{P}[X(t+1) = i+n | X(t) = i] &= \frac{\lambda_n}{\sum_n (\lambda_n + \theta_n)} \\ \mathbb{P}[X(t+1) = i-n | X(t) = i] &= \frac{\theta_n}{\sum_n (\lambda_n + \theta_n)}\end{aligned}\tag{2.2}$$

Another statistical feature we are interested in is the time between two arrivals of orders on the same side, denoted by  $\tau$ . Generally, we denote  $\tau_{i,j}$  be the first-passage time for  $X(t)$  moving from state  $i$  to state  $j$ , which is also called the first-passage time from state  $i$  to state  $j$ :

$$\tau_{i,j} = \inf\{\Delta t : X(t) = i, X(t + \Delta t) = j\} \quad (2.3)$$

Furthermore, we define  $\tau_i$  to be the first-passage time of  $X(t)$  to state 0 with starting at state  $i$ .

$$\tau_i = \inf\{\Delta t : X(t) = i, X(t + \Delta t) = 0\} \quad (2.4)$$

Generally, we assume the times  $T_{n-1,n}$ ,  $n \geq 1$  between any two consecutive arrivals of orders are independent and identically distributed random variables. Additionally, one property that should satisfy is loss of memory property, that is, conditioning on there has been time  $s$  since the last order arrived, the probability of that the next order will arrive at the book after the next time  $t$  is as same as the probability of that without such condition. Mathmatically,

$$\mathbb{P}[T_{n-1,n} \geq s + t | T_{n-1,n} \geq s] = \mathbb{P}[T_{n-1,n} \geq t] \quad (2.5)$$

Based on this property, the only distribution satisfying the property is the exponential distribution. Hence,  $T_{n-1,n}$ , the times between arrivals of orders in a Markovian queueing model, follow exponential distribution with parameter  $\sum_n (\lambda_n + \mu_n + \theta_n)$  for the level I and  $\sum_n (\lambda_n + \theta_n)$  for other levels.



## 2.2 Review of Huang's Model setting

Huang and Kercheval propose limit orders could be placed on a finite price grid  $\Pi = \{1, 2, \dots, n\}$ , where prices are measured as multiples of a tick and the upper boundary  $n$  should be chosen large enough. They measure orders and cancellation as multiples of the average size of market orders, say  $S_m$  shares, so the state of the Markovian queues is set as  $\mathbf{Z}^+ = \{0, 1, 2, 3, \dots\}$ .

The state of the order book is described with two continuous-time Markovian queue with values in  $\mathbf{Z}^+$ , specifically, the ask queue

$$A(t) = (A_1(t), \dots, A_n(t)) \quad (2.6)$$

and the bid queue

$$B(t) = (B_1(t), \dots, B_n(t)) \quad (2.7)$$

where  $t \geq 0$ . Here,  $A_k(t)$  ( $B_k(t)$ ) denotes the number of ask (bid) outstanding limit orders (in multiples of  $S_m$  shares) at price  $k$  and time  $t$ . And  $A_k(t) \wedge B_k(t) = 0$  for all  $k$  and  $t$  because there should be no overlapping ask and bid limit orders, where  $\wedge$  is the symbol for minimum. The level I ask price at time  $t$  is

$$p_A(t) = \inf\{p \in \Pi : A_p(t) > 0\} \wedge (n + 1) \quad (2.8)$$

and similarly the level I bid price at time  $t$  is

$$p_B(t) = \sup\{p \in \Pi : B_p(t) > 0\} \vee 0 \quad (2.9)$$

and we assume that  $p_B(t) < p_A(t)$  always holds.

Hence Huang and Kercheval presume the following assumptions to govern how the two queues evolve.

- Market orders and cancellations are of constant size  $S_m$ , where  $S_m$  is the average size of market orders over the time period we inspect.
- The times between the arrivals of market orders on both sides are independently drawn from an exponential distribution with rate  $\mu$ .
- Limit orders viewed as  $k$  multiples of  $S_m$  shares arrive at a distance of  $j$  ticks from the opposite outstanding price level at independent, exponentially distributed times with rates denoted  $\lambda_j^{(k)}$ , for  $k = 1, 2, \dots, M$  and  $j \geq 1$ . Here,  $M$  is the number of order sizes to be handled by the model.
- Cancellations at a distance of  $j$  ticks from the outstanding price level on the same side arrive at independent, exponentially distributed times at a rate proportional to the number of outstanding shares. For instance, the cancellation rate is  $k\theta_j$  of existing orders with size  $kS_m$  at a distance of  $j$  ticks from the outstanding price level on their side.
- All the events mentioned above are jointly independent.

The rate parameters  $\mu$ ,  $\lambda_j^{(k)}$  are measured in “orders per second” and  $\theta_j$  is measured in “orders per second per existing order” where one order has size  $S_m$  shares.

Now, we can interpret the change behavior on the two Markovian queues with state  $\mathbf{Z}^+$  as

$$A_i(t) \rightarrow A_i(t) + k \text{ at rate } \lambda_{i-p_B(t)}^k \text{ for } i > p_B(t), M \geq k \geq 0 \quad (2.10)$$

$$A_i(t) \rightarrow A_i(t) - 1 \text{ at rate } A_i(t)\theta(i - p_A(t)) \text{ for } i \geq p_A(t), i \geq p_A(t) \quad (2.11)$$

$$A_i(t) \rightarrow A_i(t) - 1 \text{ at rate } \mu \text{ for } i = p_A(t) > 0 \quad (2.12)$$

$$B_i(t) \rightarrow B_i(t) + k \text{ at rate } \lambda_{p_A(t)-i}^k \text{ for } i < p_A(t), M \geq k \geq 0 \quad (2.13)$$

$$B_i(t) \rightarrow B_i(t) - 1 \text{ at rate } B_i(t)\theta(p_B(t) - i) \text{ for } i \leq p_B(t), i \geq p_A(t) \quad (2.14)$$

$$B_i(t) \rightarrow B_i(t) - 1 \text{ at rate } \mu \text{ for } i = p_B(t) < n + 1 \quad (2.15)$$

$$(2.16)$$

Furthermore, we define truncated model by setting the two Markovian queues to have finite states, denoted  $\mathbf{Z}_K = \{0, 1, \dots, K\}$ . To modify the queues defined above, we state that if a transition would cause the state of  $A_i < K$  or  $B_i < k$  to exceed  $K$  then the state is reset to  $K$ ; and that at state  $K$ , if  $j$  is the number of ticks from  $K$  to the opposite outstanding price level, the rate of limit orders  $\lambda_j^{(k)}$  is reset to zero for all  $k$ .

Based on the model built up above, Huang and Kercheval propose the estimator to calibrate the parameters. Let  $T$  be the length of time period of interest,  $N_m$  the number of shares of market orders placed during the time, and  $N_c$  the number of shares cancelled at the level  $i$  during the time.

The rate parameter of market orders (on both sides) is estimated by

$$\hat{\mu} = \frac{N_m}{S_m T} \quad (2.17)$$

and the cancellation rate for the level  $i$  can be estimated by

$$\hat{\theta}(i-1) = \frac{N_c(i)}{L(i)T} \quad (2.18)$$

where  $L(i)$  is the average number of shares in a limit order book at level  $i$  over the time.

For limit orders, we set possible largest size of them to be  $M$  (in multiples of  $S_m$  shares). An order is in the  $k$ th group if its size is closer to  $kS_m$  than  $k'S_m$  for any other  $k'$ ,  $k = 1, \dots, M$ .  $N_l^{(k)}$  denotes the total number of shares in the limit orders of the  $k$ th group. Hence we can estimate the rate parameter of limit orders of size  $k$  by

$$\hat{\lambda}_j^{(k)} = \frac{N_l^{(k)}(j)}{S_m T} \quad (2.19)$$

## 2.3 Review of Cont's Model setting

From Cont and Larrard's perspective, the state of a limit order book can be represented as:

- denote the outstanding ask and bid price level as  $s_t^a$  and  $s_t^b$  respectively;
- $q_t^b$  denotes the size of the level I bid queue which represents the outstanding limit buy orders on the bid;
- $q_t^a$  denotes the size of the level I ask queue which represents the outstanding limit sell orders on the ask.

Cont and Larrard also postulate that the prices of all orders and cancellations are multiples of the tick size  $\delta$ . Meanwhile, since the outstanding price level moves upwards or downwards

when the queue on that is depleted by market orders and cancellations, the price processes  $s_t^b, s_t^a$  are piecewise constant processes, whose transitions correspond to hitting times of the axes  $\{(0, y), y \in \mathbb{N}\} \cup \{(x, 0), x \in \mathbb{N}\}$  by the Markovian queues  $q_t = (q_t^b, q_t^a)$ .

If a limit order book is sparse which means there are gaps in the book, then the price level might move upwards or downwards by more than one tick  $\delta$  when the level I queue is depleted. In Cont and Larrard's assumption, the gaps can be ignored. Moreover, when the level I queue on either side get depleted, the spread widens immediately to more than one tick. And Cont and Larrard suggest that, once the spread widens, we can expect some limit orders quickly fill the gap and force the spread back to one tick again. Since in most of time the spread in a limit order book is one tick, we assume that the spread is one tick all the time. In this case, when we describe the state of the order book after a price change, we focus on the state of the book after the spread has returned to one tick, i.e.  $s_t^b = s_t^a - \delta$ .

Under this assumption, each time one of the two Markovian queues get depleted, both the queues move to a new position and the spread still be one tick after the change. However, the state of the new position is not being tracked because of the complexity. Instead, we treat the queue sizes after a price change as stationary variables from a specific distribution  $f$  on  $\mathbb{N}^2$ . More specifically,  $f(x, y)$  represents the probability of observing  $(q_t^b, q_t^a) = (x, y)$  right after a level I queue on ask side gets depleted. Similarly,  $\tilde{f}(x, y)$  denotes the probability of observing  $(q_t^b, q_t^a) = (x, y)$  right after a level I queue on bid side gets depleted. More precisely, we denote by  $\mathcal{F}_t$  the history of prices and order book events on  $[0, t]$ , and the following independence holds:

- if  $q_{t-}^a = 0$ , then  $(q_t^b, q_t^a)$  is a random variable with distribution  $f$ , independent of  $\mathcal{F}_{t-}$ ;
- if  $q_{t-}^b = 0$ , then  $(q_t^b, q_t^a)$  is a random variable with distribution  $\tilde{f}$ , independent of  $\mathcal{F}_{t-}$ .

For three types of orders, we denote by  $T_i^a, T_i^b$  the durations between two consecutive orders arriving at the ask and bid side, and by  $V_i^a, V_i^b$  the size of the associated change in queue size, for  $i \geq 1$ . Intuitively, limit orders come with size  $V_i^a > 0$  or  $V_i^b > 0$ , while market orders and cancellations are corresponding to  $V_i^a < 0$  or  $V_i^b < 0$ . To represent the dynamics of a limit order book, Cont and Larrard make more assumptions to describe the distribution of the variables  $T_i^a, T_i^b, V_i^a, V_i^b$ :

- Market orders arrive at independent, exponential times with rate  $\mu$ .
- Limit orders on either side arrive at independent, exponential times with rate  $\lambda$ .
- Cancellations occur at independent, exponential times with rate  $\theta$ .
- Orders are of a unique size.

The assumptions above implicates the following statistical properties of variables  $T_i^a, T_i^b, V_i^a, V_i^b$ :

- $(T_i^a)_{i \geq 0}$  and  $(T_i^b)_{i \geq 0}$  are independent random variables from exponential distribution with parameter  $\lambda + \mu + \theta$ .
- $(V_i^a)_{i \geq 0}$  and  $(V_i^b)_{i \geq 0}$  are independent random variables with probabilities

$$\mathbb{P}(V_i^a = 1) = \frac{\lambda}{\lambda + \mu + \theta} \quad \mathbb{P}(V_i^a = -1) = \frac{\mu + \theta}{\lambda + \mu + \theta} \quad (2.20)$$

$$\mathbb{P}(V_i^b = 1) = \frac{\lambda}{\lambda + \mu + \theta} \quad \mathbb{P}(V_i^b = -1) = \frac{\mu + \theta}{\lambda + \mu + \theta} \quad (2.21)$$

- All the events are independent.

Since it is assumed that all the orders are of the same size, Cont and Larrard set the unit size to be 100 shares, and estimate the rate parameters based on that, which represent the number of certain orders per second and per 100 shares.

$$\hat{\lambda} = \frac{N_l(1)}{100T} \quad \hat{\mu} = \frac{N_m}{100T} \quad \hat{\theta} = \frac{N_c(0)}{100T} \quad (2.22)$$

where  $T$  is the time period of interest,  $N_l(1)$  denotes the total number of shares of limit orders placed on level I on either side,  $N_m$  is total number of shares of market orders on either side and  $N_c(0)$  is the total number of shares of cancellations on the outstanding price level on either side.

## 2.4 Brief Comparison of Two Models

The similarity of Huang and Kercheval's and Cont and Larrard's model setting is the independence of arrival time for any orders. Under both of their assumptions, the arrival times of all three types of orders are independent exponentially distributed but with different rate parameters. Also, all the events, the arrival of orders, are independent. Another assumption they both agree on is that the Markovian queues can only decrease by one unit-move down by one step in the state-each time an event comes.

However, there are also significant differences between the two model settings. First of all, Huang and Kercheval postulate the limit orders only can be placed on a finite price grid,  $\Pi = \{1, 2, \dots, n\}$ , and they record the queues on all price levels. But Cont and Larrard have no such restriction on price the limit orders can be placed and they are only interested in the outstanding queues on two sides and hence record and model the level I queues only.

In Huang and Kercheval (2012), the rate parameters are measured in a unit size given by the average size of market orders, i.e.  $S_m$ . Then, the rate of market orders is actually the average number of market orders per second. The rate of limit orders are the average number of orders as multiples of unit  $S_m$  per second. Even though the cancellation rate here is measured as a ratio of existing volume of limit orders, the real rate is estimated corresponding to multiples of unit  $S_m$ . On the contrary, Cont and Larrard use other size as a unit of orders. The rate parameters are the number of orders with size 100 shares that arrive at the book per second.

Besides, Huang and Kercheval allow limit orders to arrive with multiple sizes. Specifically, they estimate the rate parameters of limit orders with size  $kS_m$  respectively. That also implies the Markovian queues can move upwards by more than one step in the state. Meanwhile, they set a restriction on the volume of limit orders on all price levels, which indicates the states of the Markovian queues are finite. But Cont and Larrard suppose the limit orders arrive at the book only with unique size-100 shares. In this way, the Markovian queues can move upwards by one step each time. Cont and Larrard also believe that the queues can move up to any state since the state is infinite.

Nevertheless, as viewing  $\theta$  as a ratio of existing limit orders, Huang and Kercheval actually suppose there are more than one rate parameters for cancellations. Cont and Larrard view cancellation rate as a constant value by setting the size of each cancellation as 100 shares.



# Chapter 3

## Numerical Results

In this chapter, we firstly check the assumptions reviewed in Chapter 2 with the empirical data from the Nasdaq Stock Market. We estimate the rate parameters of the limit order book for the stock Microsoft in one day, and then use the parameters to simulate the distribution of first-passage time of the level I queue to reach state 0, which provides us with a tool to compare the assumptions in the two papers. Comparing the two empirical cumulative distribution functions with that from empirical data, we can extract more comparison between the two model settings. Based on the data, we also propose new assumptions on estimation of parameters, which we believe they fit the data better.

### 3.1 Comparison of Estimators

In Table 3.1, we show the estimates of the rate parameters under Huang and Kercheval's model assumptions and the average size of market orders and of existing level I limit orders.

$$S_m = \frac{N_m}{\text{number of market orders}} \quad L(1) = \frac{\sum_{i=1}^N L_{t_i} t_i}{T} \quad (3.1)$$

where  $N_m$  is total shares of the market orders,  $L_{t_i}$  is the existing volume of the limit orders on the level I during time period  $t_i$ , and  $T = \sum_{i=1}^N t_i$  is the whole time period of interest. The Figure 3.1 and 3.2 show the histograms of volumes of the market orders and of the existing limit orders on the level I.

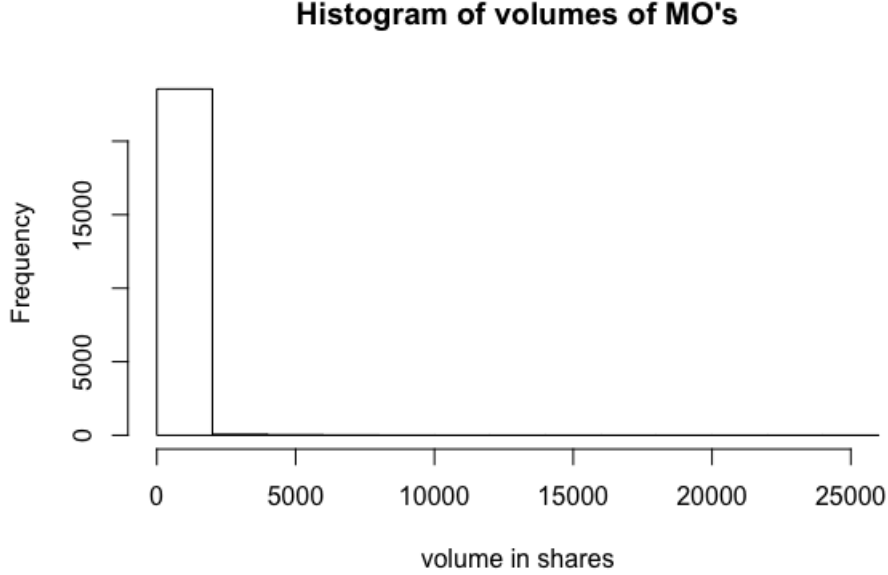


Figure 3.1: Histogram of volumes of the market orders

The parameters of limit orders and cancellations are all estimated on the outstanding levels. More specifically, to estimate  $\hat{\lambda}_1^{(i)}$ , we firstly set the limit orders into groups corresponding to the closest multiples of  $S_m$ , that is, if a limit order has a volume closer to  $kS_m$  than any other  $k'S_m$  then it is set into group  $k$ . Then the parameter is estimated as the average number of limit orders of volume near  $kS_m$  per second in each group.  $L(1)$  represents the average size of existing level I limit orders, and then the cancellation ratio  $\hat{\theta}(i-1)$  with  $i=1$  is estimated as below.

$$\hat{\theta}(0) = \frac{N_c(1)}{L(1)T} \quad (3.2)$$

**Histogram of volumes of existing LO's on the level I**

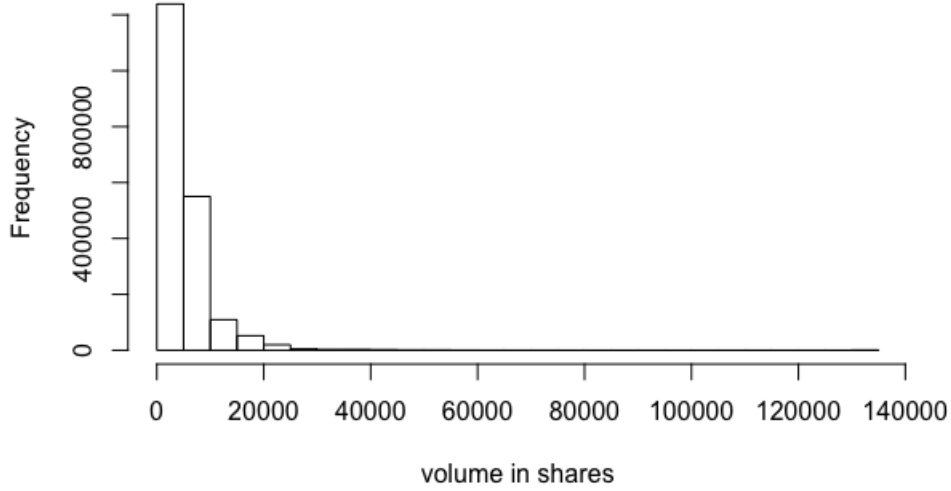


Figure 3.2: Histogram of volumes of the existing limit orders on the level I

where  $N_c(1)$  denotes the shares of the cancellations on the level I. The true cancellation rates are  $k\hat{\theta}(0)$  if there are  $kS_m$  shares of the outstanding limit orders on either side.

Table 3.1: Estimates of rate parameters in multiples of  $S_m$  per second

$\hat{\mu}$	$\hat{\theta}(0)$	$\hat{\lambda}_1^{(1)}$	$\hat{\lambda}_1^{(2)}$	$S_m$	$L(1)$
0.506	0.403	2.397	1.955	201	3862

However, since  $\hat{\lambda}_1^{(2)}$  represents the average number of the limit orders with size  $2S_m$  arriving at the book per second, so we can also roughly consider that as, in each second,  $2\hat{\lambda}_1^{(2)}$  limit orders with size  $S_m$  arrive at the book in average. When the volume of the existing level I limit orders becomes near to 0, the death rate (sum of rates of market orders and cancellations) is significantly less than the birth rate (rates of limit orders), which implies it would be less possible for the queues on the level I to reach state 0.

In Cont and Larrard’s paper framework, we estimate the rate parameters of the three types of orders as shown in Table 3.2. Since Cont and Larrard ignore the case that the spread is larger than one, we only use the book data with spread one to estimate the arrival rates on the level I. Under Cont and Larrard’s estimations, the birth rate  $\hat{\lambda}$  is slightly less than the rate of death  $\hat{\mu} + \hat{\theta}$ . Hence, the parameter estimates satisfy the property that the outstanding Markovian queues should always be able to be depleted.

Table 3.2: Estimates of rate parameters in number of 100 shares per second

$\hat{\lambda}$	$\hat{\mu} + \hat{\theta}$	$\frac{\hat{\mu} + \hat{\theta} - \hat{\lambda}}{\hat{\lambda}}$
9.555	9.587	0.003

Now we further compare the two methods of estimation in terms of their cumulative distribution function of the first-passage time to reach state 0. The reason we care about the first-passage time to state 0 is that we are always interested in the behavior of price change and price gets upwards if the level I queue on ask side gets depleted before the one on bid side, which also implies the first-passage time to state 0 of the outstanding queue is shorter than the other one, and vice versa. Hence, based on the statistical features of first-passage time to state 0 of outstanding queues, we can derive some properties of the price behavior and make prediction on that. Because the cumulative distribution function of first-passage time only depends on the parameters  $\lambda$ ,  $\mu$ , and  $\theta$ , it is a good way to compare different bunches of parameters. We also derive the empirical cumulative distribution function of first-passage time to state 0, and we can decide which model fits the data better depending on whether its cumulative distribution function is closer to the empirical distribution.

Note that we need to observe a whole process for an outstanding queue: starting from becoming level I and ending up with getting depleted, then we can only choose the outstanding queues when the spread is one in case new limit orders insert into the spread and become

level I. Moreover, as shown in Table 3.3, we have more observations with initial volume of 100 shares on both sides, so we should use the empirical cumulative distribution function of first-passage time to state 0 with initial volume 100 as the indicator function in order to obtain more accurate estimation.

Table 3.3: Percentage of new outstanding queues with specific initial volume when spread is one

side	100 shares	200 shares	other possible number of shares
ask	0.196	0.065	< 0.01
bid	0.191	0.027	< 0.02

Since it is hard to derive analytical form of cumulative distribution functions of first-passage time to state 0 with estimates above, we use Monte Carlo simulation to obtain the estimation of that for both Huang and Kercheval's and Cont and Larrard's estimates. The detailed steps of simulation is listed below:

- Step 1. we generate N queues with initial volume 1 unit and time record 0;
- Step 2. we obtain the time for next order to come by generating a random number from exponential distribution with parameter  $\lambda + \mu + \theta$ , and we add the generated times to the time records of every queue;
- Step 3. we generate the volumes of each incoming events with probability  $\frac{\lambda}{\lambda + \mu + \theta}$  and  $\frac{\mu + \theta}{\lambda + \mu + \theta}$  for each queue;
- Step 4. for the queues have non zero volumes, we repeat the step 2 and 3, and we keep the rest queues at state 0;
- Step 5. we repeat the above steps until all the queues get 0 all the maximum of the time records reach some values.

After the simulation, we can get the empirical cumulative distribution function of first-passage time to state 0 provided different estimates. As shown in Figure 3.3, we use the observations on the ask side to generate the empirical function. From the figure, we find out that the cumulative distribution function of Cont and Larrard's estimates is significantly closer to the empirical cumulative distribution function. Therefore, we draw a conclusion that the estimators of Cont and de Larrard (2013) can fit our data better compared with Huang's. Additionally, near time 0, the function of Huang and Kercheval is greatly lower than that of Cont and Larrard, which shows the probability that the queue with initial volume of 100 shares gets depleted early on is relatively small under Huang and Kercheval's assumptions.

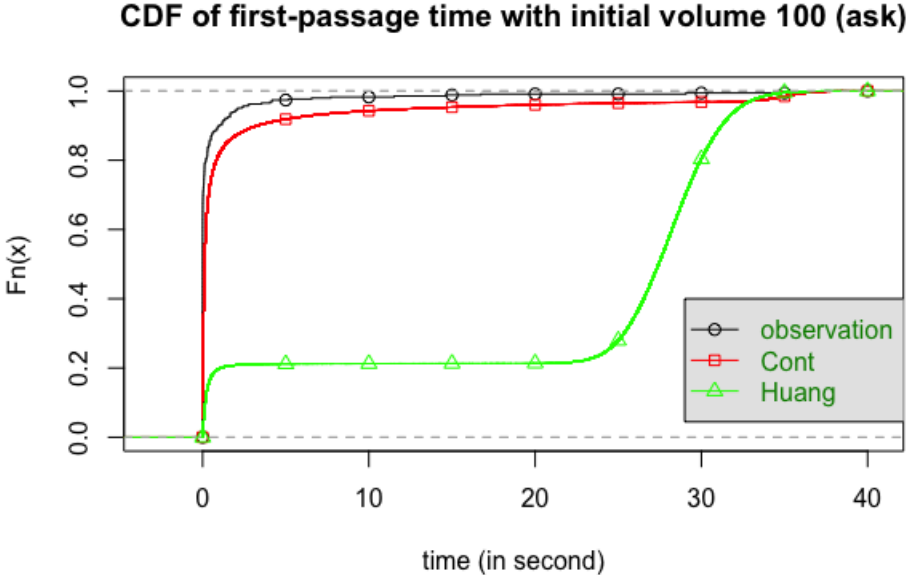


Figure 3.3: CDF of first-passage time to state 0 with initial volume of 100 shares

We try to figure out a reason why Huang and Kercheval's estimates fit our data not that well, so we simulate the cumulative distribution functions for the following two bunches of

estimators:

$$\hat{\mu} = \frac{N_m}{S_m T} \quad \hat{\theta} = \frac{N_c(0)}{S_m T} \quad \hat{\lambda}_j^{(k)} = \frac{N_l^{(k)}(j)}{S_m T} \quad (3.3)$$

and

$$\hat{\lambda} = \frac{N_l(1)}{100T} \quad \hat{\mu} = \frac{N_m}{100T} \quad \hat{\theta}(0) = \frac{N_c(0)}{L(1)T} \quad (3.4)$$

We name the first one “constant theta” and the second one “inconstant theta”. Clearly, for now we just exchange Huang’s and Cont and Larrard’s assumption on cancellation rate.

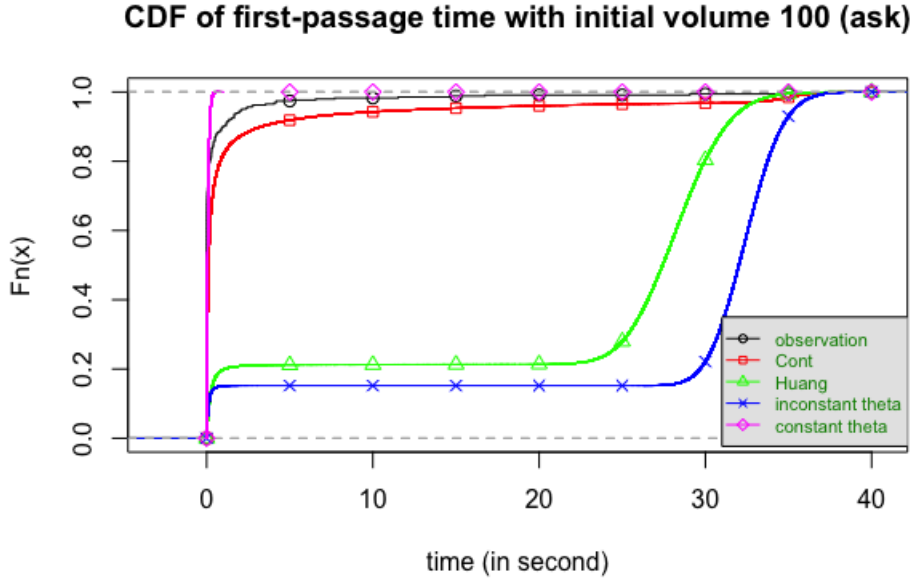


Figure 3.4: CDF of first-passage time to state 0 with initial volume of 100 shares

When we use constant cancellation rate in Huang and Kercheval’s estimators as shown in above equation, we have much larger death rate than birth rate, and hence the queues get depleted as soon as become level I, as shown in Figure 3.4. Meanwhile, if we revise the Cont and Larrard’s estimators with inconstant cancellation rate we have a cumulative distribution function pretty like the one of Huang and Kercheval’s original estimators. Therefore, we conclude that it is the inconstant cancellation rate that causes the estimated cumulative

distribution function of first-passage time to state 0 far away from the true one, and Cont and Larrard’s estimators are better than Huang and Kercheval’s is due to the constant cancellation rate to most degree.

## 3.2 Multiple-Size Rates and Other Assumption

First of all, we agree on some basic assumptions that the events arrive at the book independently and the time between consecutive arrivals is independent, identically drawn from an exponential distribution with a certain rate parameter of orders.

We check the percentage of duration time for the different ticks, and find that most of the time the spread is one tick, which validates the assumption of Cont and Larrard, as shown in Table 3.4. In this way, we also need to focus on the orders of the book when the spread is one tick.

Table 3.4: Percentage of time with a given spread

1 tick	2 tick	$\geq 3$ tick
98.47	1.441	0.09

In Figure 3.3 and 3.4, we notice that the cumulative distribution function of Cont and Larrard is lower than that of observations, which indicates that, given initial volume of 100 shares, the queues with rate estimates of Cont and Larrard have less probability to get depleted before time  $t$  than actual queues, for  $t < 40$ . We believe, as shown in Table 3.2, that  $\frac{\hat{\mu} + \hat{\theta} - \hat{\lambda}}{\hat{\lambda}}$  is diminutive shows the reason for queues getting depleted more slowly. Hence we intend to come up with better estimators of rate parameters to expand the difference between birth rate and death rate slightly.



Based on Table 3.5, we find that about 50% of orders arrive with volume of 100 shares no matter what type they are, so we should consider the rate parameters with unit size as 100 shares, just as in Cont and de Larrard (2013). Meanwhile, if we divide orders into groups according to the closest multiples of unit size to their volumes, and then estimate the rates within each group, we will get smaller estimates than those by Cont and Larrard's estimators.

Table 3.5: Percentage of orders on level I with given volumes (spread=1)

order type	100 shares	200 shares	other possible volume
limit order	49.935%	22.980%	< 9%
market order	47.580%	12.614%	< 7%
cancellation	47.546%	23.728%	< 9%

Therefore we suggest to have different rate parameters standing for multiple size orders and we call them multiple-size rates. The estimators for multiple-size rates are as below:

$$\hat{\lambda}^{(k)} = \frac{N_l^{(k)}(1)}{100T} \quad \hat{\mu}^{(k)} = \frac{N_m^{(k)}}{100T} \quad \hat{\theta}^{(k)} = \frac{N_c^{(k)}(0)}{100T} \quad (3.5)$$

where  $N_l^{(k)}$ ,  $N_m^{(k)}$  and  $N_c^{(k)}(0)$  denote the total number of shares of limit orders, market orders and cancellations on level I with volumes closest to  $100k$  when the spread is one. Here, we only consider the case  $k \leq 2$  since most of orders are in the group 1 and 2. The estimates based on our data are given in Table 3.6 with  $k = 1, 2$ .

Also, we can calculate the ratio of birth rate and the difference between birth and death rate.

$$\frac{\hat{\mu}^{(1)} + \hat{\mu}^{(2)} \times 2 + \hat{\theta}^{(1)} + \hat{\theta}^{(2)} \times 2 - \hat{\lambda}^{(1)} - \hat{\lambda}^{(2)} \times 2}{\hat{\lambda}^{(1)} + \hat{\lambda}^{(2)} \times 2} \approx 0.346 \quad (3.6)$$

Table 3.6: Estimates of multiple-size rate parameters

rate parameters	$k = 1$	$k = 2$
$\hat{\lambda}^{(k)}$	2.086	1.954
$\hat{\mu}^{(k)}$	0.263	0.151
$\hat{\theta}^{(k)}$	4.814	1.345

That is larger than the one calculated by Cont and Larrard's estimates. Then we simulate the cumulative distribution function of first-passage time to state 0 with multiple-size rates based on our estimates and the result is shown in Figure 3.5. From Figure 3.5, the cumulative

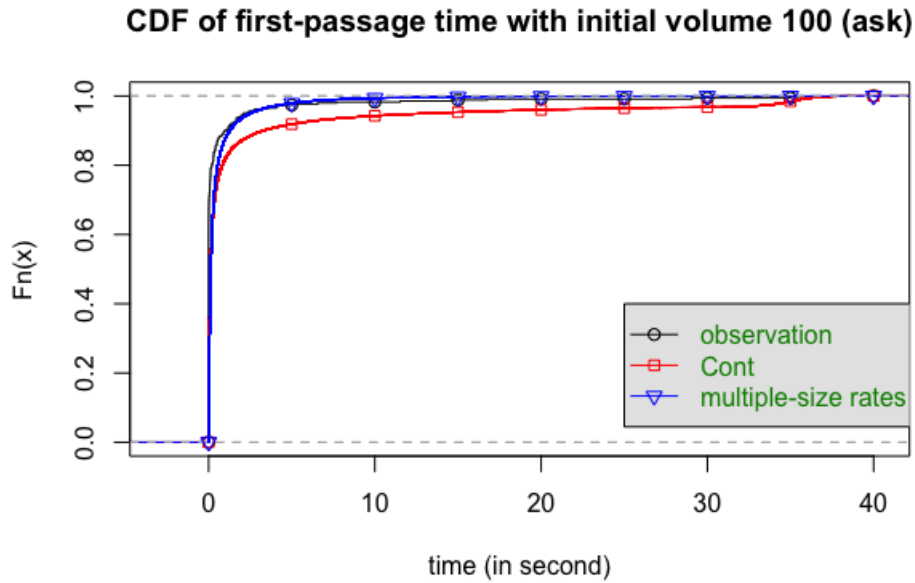


Figure 3.5: CDF of first-passage time to state 0 with initial volume of 100 shares

distribution function of first-passage time to state 0 with multiple-size rates is much closer to the empirical one, and then we can conclude that the multiple-size rates assumption fits our data better.

Table 3.7: The probability of either side to shrink the spread condition on a certain price change

	bid	ask
after a price change	0.448	0.552
after a price increase	0.639	0.361
after a price decrease	0.191	0.809

Because we only consider the outstanding queues on the book when the spread is one, once one of queues gets depleted, we have no prior information about the volumes of the new outstanding queues. So, we further investigate the distribution of volumes of new queues after a price increase and decrease as postulated in Cont and de Larrard (2013). After a price change, there is a limit order on either side arrive at the book to shrink the spread. In table 3.7, we estimate the probabilities of either side to shrink the spread, and the conditional probabilities of either side to shrink the spread given that the price change is increase and decrease. From Table 3.7, it is clearly to observe that limit orders on bid side are more possible to arrive and force the spread to be one again after an outstanding queues on ask side gets depleted, and conversely, limit orders on ask side are more possible to insert the spread after an outstanding queue on bid side that gets depleted.

We can write the joint distribution density functions  $f(x, y)$ ,  $\tilde{f}(x, y)$  of volumes as

$$f(x, y) = P_b^{in} f_b(x, y) + P_a^{in} f_a(x, y) \quad (3.7)$$

and

$$\tilde{f}(x, y) = P_b^{de} \tilde{f}_b(x, y) + P_a^{de} \tilde{f}_a(x, y) \quad (3.8)$$

Here,  $P_b^{in}$  and  $P_a^{in}$  are the probabilities of bid and ask limit orders shrinking the spread after a price increase, and similarly  $P_b^{de}$  and  $P_a^{de}$  are that after a price decrease. Those are estimated in Table 3.7. Also,  $f_b(x, y)$  and  $f_a(x, y)$  are the density of volumes on either side at the time a bid or ask limit order shrinks the spread after a price increase, and  $\tilde{f}_b(x, y)$ ,  $\tilde{f}_a(x, y)$  are the same things for the case after a price decrease. Furthermore, we assume that the volumes of limit orders on the ask side is independent of those on the bid side, then we can derive the following formula:

$$\begin{aligned} f_b(x, y) &= f_b^b(x) f_b^a(y) & f_a(x, y) &= f_a^b(x) f_a^a(y) \\ \tilde{f}_b(x, y) &= \tilde{f}_b^b(x) \tilde{f}_b^a(y) & \tilde{f}_a(x, y) &= \tilde{f}_a^b(x) \tilde{f}_a^a(y) \end{aligned} \tag{3.9}$$

where  $f_b^b(x)$  and  $f_b^a(x)$  are the density functions of volumes on the bid and ask side at the time a bid limit order shrinks the spread after a price increase, and  $f_a^b(x)$ ,  $f_a^a(x)$  are for the case where an ask limit order shrinks the spread after a price increase.  $\tilde{f}_b^b$ ,  $\tilde{f}_b^a$ ,  $\tilde{f}_a^b$ , and  $\tilde{f}_a^a$  are defined in a similar way but after a price decrease. We plot the density functions mentioned above in Figure 3.6 to 3.9.

Moreover, in Figure 3.6 to 3.9, the densities drawn by dash lines are all significantly smaller than the densities on the other side, so we can view those as density of uniform distribution. Also, we can observe that, after a price increase, if we have an ask limit order to shrink the spread then its density shows bimodality as well as if we have a bid limit order to shrink the spread after a price decrease. However, both the densities of a bid limit order which shrinks the spread after a price increase, and of a ask limit order which shrinks the spread after a price decrease, are unimode.

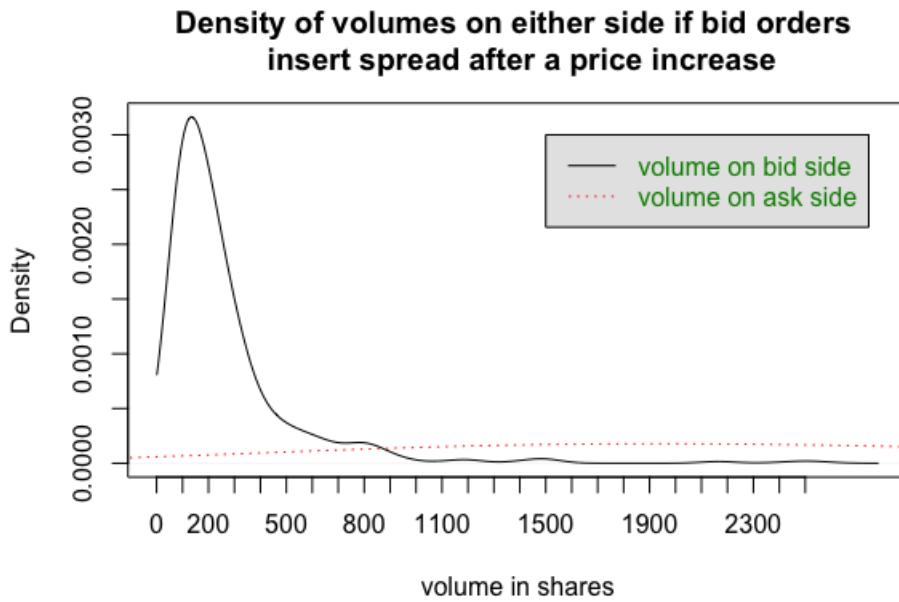


Figure 3.6: Density of volumes on either side if bid orders insert spread after a price increase

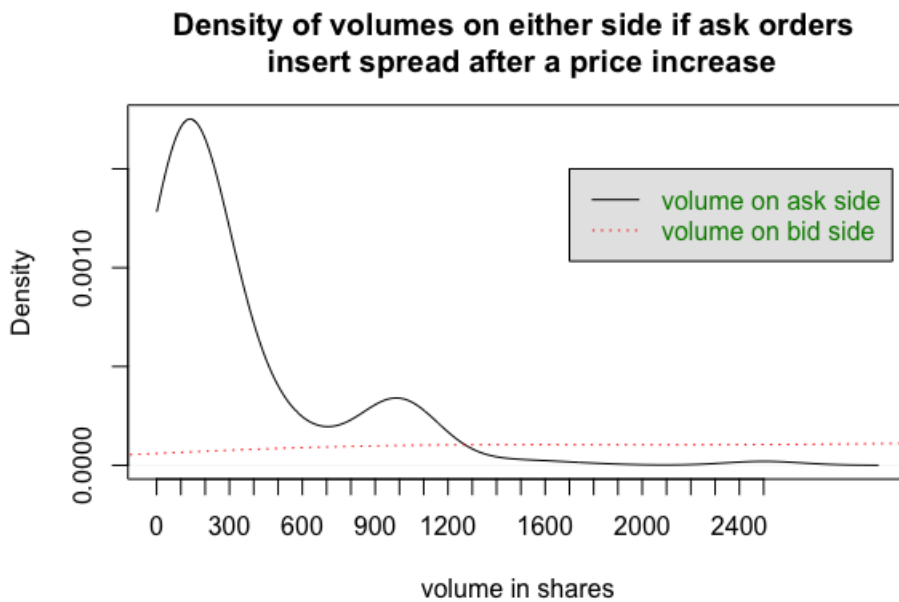


Figure 3.7: Density of volumes on either side if ask orders insert spread after a price increase

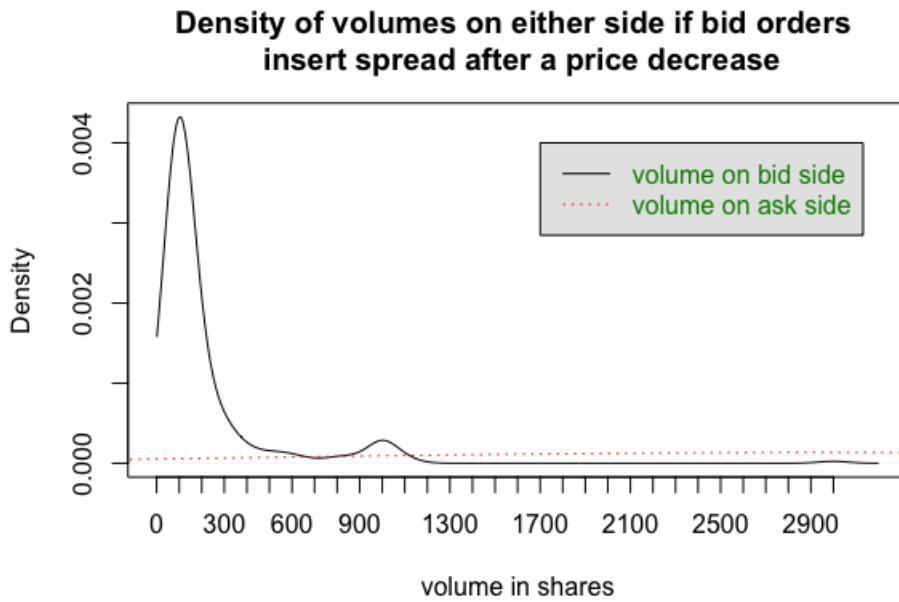


Figure 3.8: Density of volumes on either side if bid orders insert spread after a price decrease

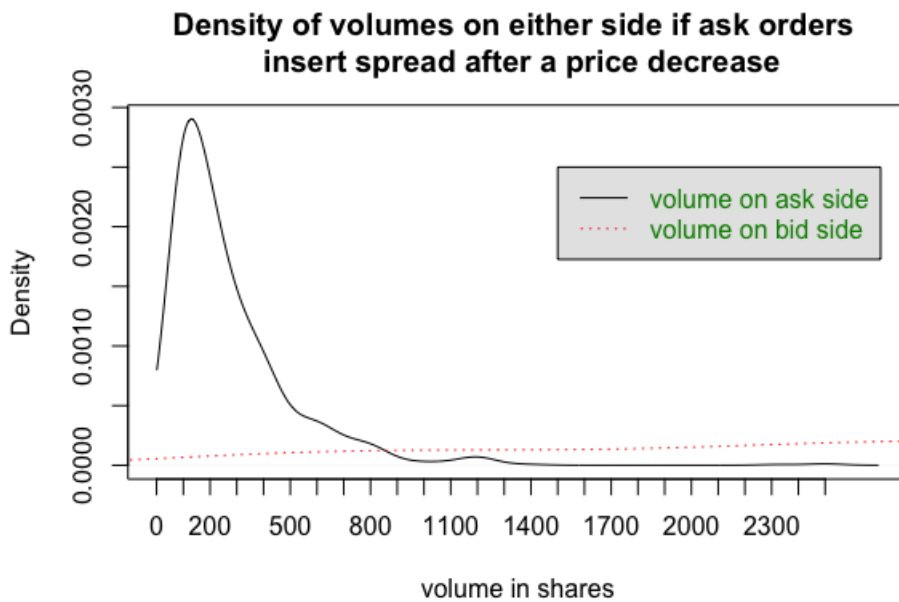


Figure 3.9: Density of volumes on either side if ask orders insert spread after a price decrease

# Chapter 4

## Conclusion

We firstly introduce the basic ideas for modeling a limit order book as a Markovian queueing system and summarize the similarities and differences between the assumptions in Huang and Kercheval (2012) and Cont and de Larrard (2013) models respectively. Specifically, Huang assumes limit orders have various rates corresponding to the volume and rate of cancellations depends on volumes of existing limit orders proportionally.

Based on data from the Nasdaq Stock Market and comparison by cumulative distribution functions of first-passage time to state 0, we conclude that the estimators of Cont's model fit the data better than that of Huang's model. Moreover, under assumption the spread keeps one tick for most of time, we put forward assumption multiple-size rates that all types of orders have various rates depending on their volume, which practically fits our data better in terms of closeness to the empirical cumulative distribution function of first-passage time to state 0. At last, we further investigate the joint distribution of volumes on either side when the spread becomes one again after price increases and decreases, under the assumption that the side of orders inserting the spread is independent of the volumes of them and existing outstanding limit orders on the other side.

## References

- Cartea, A., Jaimungal, S., & Penalva, J. (2015). *Algorithm and high-frequency trading*. University Printing House, Cambridge CB2 8BS, United Kingdom: Cambridge University Press.
- Cont, R., & de Larrard, A. (2013). Price dynamics in a markovian limit order market. *SIAM Journal for Financial Mathematics*, 4(1), 1-25.
- Cont, R., Stoikov, S., & Talreja, R. (2010). A stochastic model for order book dynamics. *Operations Research*, 58(3), 549-563.
- Farmer, J. D., Gillemot, L., Lillo, F., Mike, S., & Sen, A. (2004). What really causes large price changes? *Quantitative Finance*, 4, 383-397.
- Gould, M. D., Porter, M. A., Williams, S., McDonald, M., Fenn, D. J., & Howison, S. D. (2013). Limit order books. *Quantitative Finance*, 13(11), 1709-1742.
- Hollified, B., Miller, R. A., & Sandas, P. (2004). Empirical analysis of limit order markets. *The Review of Economic Studies*, 71(4), 1027-1063.
- Huang, H., & Kercheval, A. N. (2012). A generalized birth-death stochastic model for high-frequency order book dynamics. *Quantitative Finance*, 12(4), 547-557.
- J. Bouchaud, M. M., & Potters, M. (2002). Statistical properties of stock order books: Empirical results and models. *Quantitative Finance*, 2, 251-256.
- Parlour, C. A. (1998). Price dynamics in limit order markets. *Review of Financial Studies*, 11(4), 789-816.
- Rosu, I. (2009). A dynamic model of the limit order book. *Review of Financial Studies*, 22, 4601-4641.
- T. Foucault, O. K., & Kandel, E. (2005). Limit order book as a market for liquidity. *Review of Financial Studies*, 18(4), 1171-1217.