FLORIDA STATE UNIVERSITY

COLLEGE OF ARTY SCIENCE

THIS IS MY TITLE:

AND THIS IS ITS SECOND LINE

By

VIKTOR SPOYLES

A Dissertation submitted to the
Department of Furtive Studies
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2015

Viktor Spoyles defended this dissertation on October 31, 2015.
The members of the supervisory committee were:

Faux Causson Yorverk
Professor Directing Thesis

Verda Boizaar
University Representative

Beauxeau D'Claune
Committee Member

Arlip Zarseeld
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with university requirements.

ACKNOWLEDGMENTS

I am greatly indebted to many people during my life to pursue this degree.

First and foremost, I want to express my gratitude to Dr. Jinfeng Zhang,my thesis major professor. Without his patience and insight, it is impossible to complete this thesis. His profound knowledge in statistics and programming, smart vision in research direction and passion for both work and life has benefited me a lot. Dr. Zhang has provided every help I would imagine from a research advisor and it was an enjoyable experience of working with him.

I also deeply appreciate my co advisor Dr. Giray Ökten, he put a lot of effort on this thesis and provided many insightful advice. Besides the solid knowledge of monte carlo methods he taught me, he also helped me so much from the very first day that I joined department, including selecting courses,career preparation and job search. I feel very fortunate to have Dr. Ökten as my co advisor for accomplishing this thesis.

I want to extend my thanks to Dr.Alec N. Kercheval and Dr. as my committee members and every course I had at our department in Florida state university is a wonderful experience and I would like to thank all my instructors for their patience and effort: Dr.Kopriva, Dr. Nichols,Dr. Ewald, Dr. Aldrovandi ,Dr.Fahim, Dr.Gallivan,Dr.Kim, Dr. Sussman and Dr. Zhu. Those courses help me to cultivate solid background of financial math, numerical analysis and programming.

Furthermore, I owe my special thanks to Dr. Penelope Kirby, Dr. Penny LeNoir, Ms.Blackwelder, Mr. Dodaro, Mr. Grigorian and Mr. Wooland, who gave me a lot of instructions and suggestions when I worked as a teaching assistant in our department and mathematics lab. They are all very nice and point out my weak points during work, which gave me a good opportunity to enhance my presentation skills and teaching abilities.

I also want to thank my colleagues and friends at department: Yang Liu, He Huang, Jinghua Yan, Bo Zhao, Jian Geng, Ming Zhu, Wanwan Huang, Yuan Zhang, Wen Huang, Yaning Liu,Qiuping

Xu, Guifang Zhou, Pierre Garreau, Ahmed Islim, Dawna Jones, Linlin Xu, Yuying Tzeng, Szu-Yu Pai, Nguyet Nguyen, Dong Sun, Daozhi Han, Yuanda Chen, Chun-Yuan Chiu, Chenchen Zhou, Yao Dai, Chaoxu Pei, Fangxi Gu, Zailei Cheng, Jian Li, Xinru Yuan, Haixu Wang, Wei Meng and Qi Si for their help, cooperation and meaningful suggestions to my research.

Finally, I want to thank my father, Xinshan Wang, my mother, Hua Yang and my other family members for their supports during my PhD study period. They are the origin of all my motivation, confidence and strength.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

According to rapid development in information technology, limit order books(LOB) mechanism has emerged to prevail in today's financial market. In this paper, we proposes ensemble machine learning architectures for capturing the dynamics of high frequency limit order books such as predicting price spread crossing opportunities in future time interval. The paper is more data-driven oriented, so experiments with 5 real time stock data from NASDAQ, measured by nanosecond, are established. The models are trained and validated by training and validation data sets. Compared with other models,such as logistic regression, support vector machine(SVM),our out-of-sample testing results has shown that ensemble methods had better performance on both statistical measurement and computational efficiency. A simple trading strategy that we devised by our models has shown good profit and loss(P&L) results. Although this paper focuses on limit order books, the similar frameworks and processes can be extended to other classification research area.

Keywords: limit order books, high frequency trading, data analysis, ensemble methods, F1 score.

CHAPTER 1

LATEX EXAMPLES

This is an introductory paragraph. summary all the situation and example in latex file, such as how to write algorithm, how to input programming code in latex file,etc.

Each section contains one aspect of how to use latex, for example, we write the example of input code in latex file in the first section.

1.1 Example of input code

In section 1.1, we show the example of input code:

```java
// Hello.java
import javax.swing.JApplet;
import java.awt.Graphics;

public class Hello extends JApplet {
    public void paintComponent(Graphics g) {
        g.drawString("Hello, world!", 65, 95);
    }
}
```

1.2 Example of algorithm

Example of algorithm is as follows, using the package of algorithm2e:

Algorithm 1.1: My algorithm

Initialize the observation weights ω_i=1/N,i=1,2,...,N;

for *m=1 to M* **do**

  Fit a classifer $G_m(x)$ to the training data using weights ω_i;

  Compute ;

$$err_m = \frac{\sum_{i=1}^{N} \omega_i I(y_i \neq G_m(x_i))}{\sum_{i=1}^{N} \omega_i}$$

  Compute $\alpha_m = log((1 - err_m)/err_m)$;

  Set $\omega_i \leftarrow \omega_i \cdot exp[\alpha_m \cdot I(y_i \neq G_m(x_i))]$, $i = 1, 2, ..., N$;

Output $G(x) = sign[\sum_{m=1}^{M} \alpha_m G_m(x)]$;

1.3 Example of array

$$\begin{cases} a + b = 2 \\ a - b = 4 \end{cases}$$

1.4 cite reference

This is my abstract

CHAPTER 2

INTRODUCTION

2.1 High frequency trading system

2.1.1 Evolution of high frequency trading

Over the last few decades, information technology, including computing speed and memory volume, has made great development. According to this trend, a new class of trading system, which is called high frequency trading (HFT), has appeared to today's financial markets. Generally speaking, HFT represents a program trading platform that uses powerful computers to transact a great number of orders at very fast speeds. It becomes more and more popular because of some key factors:

- **Narrowing Spreads**. In 2001, the unit of quoting prices in U.S. Stock exchanges changed from fractions to decimals. so the minimum spread between the bid and ask prices decreased from 1/6th of a dollar (6.25 cents) to one cent. The change of price unit provides traders better alternatives to seek spread arbitrages, which results in a strong boost in algorithmic trading system.

- **Regulation changes**. In 2005, the Securities and Exchange Commission(SEC) passed the Regulation National Market System(Reg.NMS), which improved transparency and competition among different financial markets. Besides, this regulation also required trade orders to be posted nationally instead of at individual exchanges. So traders can be beneficial of profit from small price difference of a security among different exchanges.

High frequency trading is an extension case of algorithmic trading, which turns over small positions of a security very frequently. The U.S. Securities and Exchanges Commission conclude specific characteristics of HFT:

- Submit some orders and cancel them soon after the submission.

- Maintain very few or no overnight positions.
- Maintain very short time intervals for specific security positions and turn over very frequently of many small positions in one or more financial tools.
- Utilize complicated and high performance computing program to generate, execute or cancel orders.
- Make use of individual data from exchanges and servers that belong to co-location provider in order to minimize network or other types of latencies.

According to recent survey(Agarwal (2012)), high frequency trading has taken a great number of share in U.S. and European equity trading volume. As shown in figure 2.1, in U.S., the percentage of HFT in equity turnover by volume maintained growth trend overall from year 2005 to 2010. For example, in 2010, HFT occupied 56% by volume of the entire equity turnover, increased from 21 % in 2005.

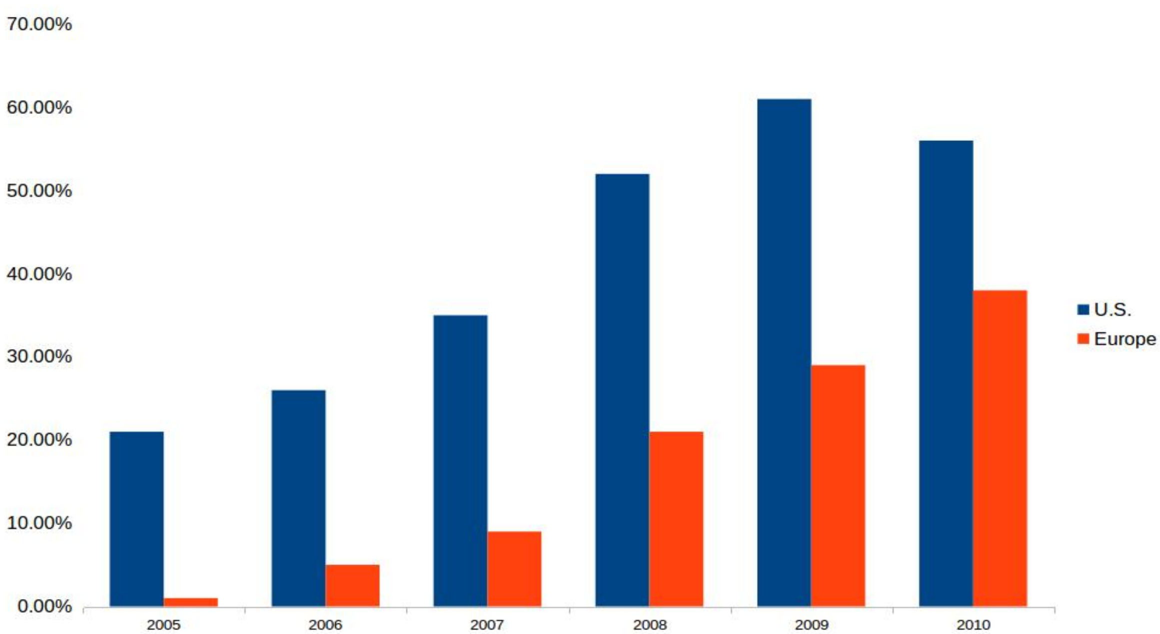


Figure 2.1: High frequency trading as a % of equity turnover by volume, U.S. and by value, Europe 2005-2010

Similar situation happened in Europe. HFT only accounted for 1% of equity turnover by value in 2005. The percentage surged to 38 % in 2010, increased from 29 % just a year ago.

2.1.2 High frequency trading strategies

Generally speaking, there are two main high frequency trading strategies: passive and aggressive trading strategies. Passive strategies uses limit orders which let the brokerage to buy or sell the stocks at a specified level price, while aggressive strategies utilizes market orders to buy or sell the stocks immediately. In the following, we give the description of main types of these two trading strategies.

For passive HFT strategies, the first widely used method is passive market making. This method allow the market maker to purchase a company's securities and,at the same time, the market maker is also acted as an underwriter of the securities in a secondary public offering. Since the market maker can place bids of a security before its publication, the real buyers would have to place their buy orders higher than the market maker's bid price. Therefore the market maker can benefit from this higher opening. The second method for passive strategies is arbitrage trading. As described by its name, this method makes profit by the price difference of the same or related securities. A simple example is as following: the stock of one company that we called X is trading at $10 on the New York Stock Exchange(NYSE), while at the same moment it is trading at the price of $10.05 on the London Stock Exchange(LSE). Given that a trader can trade the stock on both market with no time lag, he can buy the stock on the NYSE and immediately sell the same shares on the LSE. Obviously, a profit of 5 cents without risk can be earned by him. Those arbitrage opportunity will persist until the specialists on the NYSE or LSE adjust their price to eliminate the price difference.

The main type of aggressive HFT strategies consists of the following two types: Momentum ignition and order anticipate. The former one is a strategy that a proprietary trading firm buy or sell volumes of orders that will cause the price of underlying security significantly going up or down in the near future. Such quick submission and cancellation of many orders of a stock will trigger other traders' algorithm to buy or sell the same stock more aggressively. After the trend of price movement is made in the market, the momentum maker will benefit from selling the stock at a higher price or buying the stock at a lower price. However it is very difficult to distinguish between momentum ignition and *spoof*, which was defined as illegal according to Dodd Frank Act. For the order anticipate method, it can be described as a liquidity detection trading which confirms the existence of large institutional buys or sellers in the marketplace and then trade ahead of these buyers or cellers in anticipation that their large orders will move market prices(Securities and

Exchange Commission,2014,p.8). The line between this method and another illegal action called *front-running* can be nuanced. The *front-running* is the unethical practice that a stockbroker trades securities in his personal account based on the knowledge of advance knowledge of pending orders from its customers.

2.2 Limit order book dynamics

According to the above section, we know that under the aggressive strategies situation, it is very likely that the strategy will be deemed as illegal. Therefore choosing passive strategies in high frequency trading is a good idea. Since most transactions in the passive trading strategies are buying and selling the securities at a specific price, limit order books play an indispensable role in this strategy. Actually, in today's financial market, more than half of stock exchanges now use a limit order book(LOB) mechanism to facilitate trade(Rosu (2009)). Some exchanges, such as Helsinki, Hong Kong, Shenzhen, Swiss, Tokyo, Toronto, and Vancouver Stock Exchanges, now use pure LOBs(Luckock (2001)). Some exchanges use hybrid of hybrid LOBs, which include the New York Stock Ex-change (NYSE), NASDAQ, and the London Stock Exchange (LSE) (Cont et al., 2010).

As described above, we can see that LOBs play an important role in financial trading architectures, it is beneficial for both scholars and practitioners to understand dynamics of LOB. The advantages of capturing dynamics of LOBs include: finding optimal opportunity to execute orders(Obizhaeva and Wang, 2013); improving performance of electronic trading algorithms(Engle et al., 2006); Obtaining a better understanding of market micro structure for Practitioners(Harris, 2003); Getting a clearer insight into market volatility(Kirilenko et al., 2015).

In this thesis, we first introduce basic definitions of LOBs, including math definition of bid-ask spread, mid price, bid and ask side depth, spread crossing opportunities and so on. Then we use statistical and data driven methods,especially in the area of machine learning architectures, to model the future arbitrage opportunities of LOBs. We focus on ensemble machine learning algorithms to build our prediction models. As far as we know, there are no evidence shows that these methods were used in LOBs research area. Introduction of ensemble machine learning algorithms will be given in the following section of this chapter and details can be found in chapter ???.

2.3 Ensemble method for classifiers

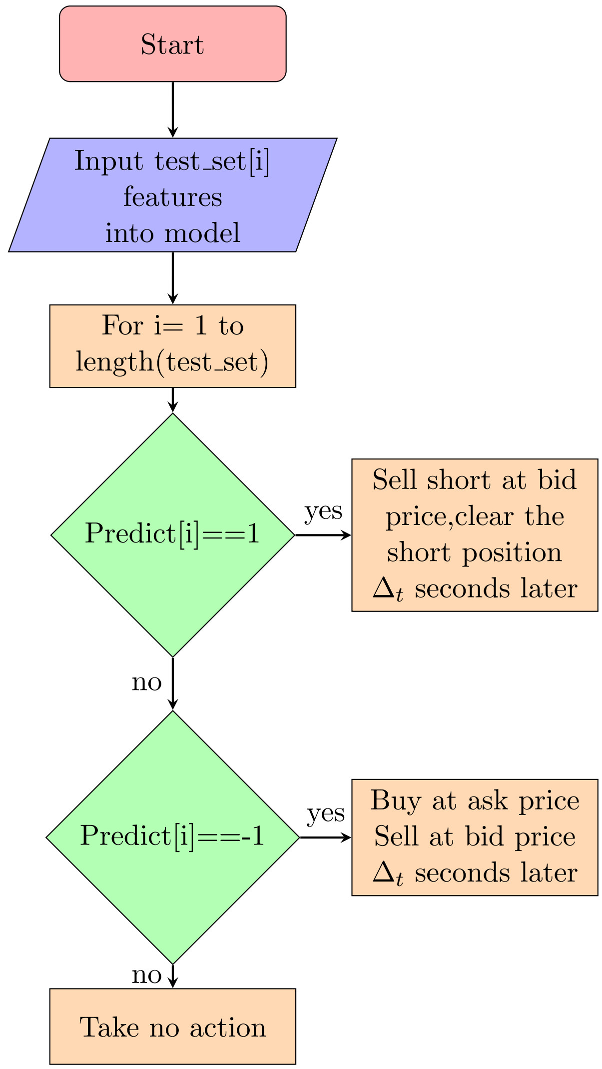
In spite of a great amount of research on limit order books, there is only a few literature which utilize machine learning methods for capturing the limit order books. Furthermore, based on our knowledge, there is little evidence that ensemble methods were allied to this topic. In our research, we try to use AdaBoost(Adaptive boosting) method, which was deemed as the best classifier off the shelf(Kégl, 2013), and random forest method to predict spread crossing over opportunities of LOBs.

Generally speaking, an ensemble classifier contains a set of individually trained classifiers(such as neural networks or decision trees) to get better predictive performance by combining the predicting results of each individual classifier. Some past research show that ensemble classifier is generally more accurate than any of the individual classifier which constitutes the ensemble. For example, Hansen and Salamon (1990) and Hashem (1997) have conducted both theoretical and empirical research which demonstrated that a good ensemble of neural networks are more accurate than its basic classifier.

Like other machine learning classifiers, ensemble methods are double edged swords. The main advantage of ensembles is that there is little probability that all classifiers will make the same mistake. Actually, if each error is made by a minority of the classifiers, you can obtain an optimal classification. Besides,ensembles are very likely to reduce the variance of classifiers. Therefore, if the classification algorithms are sensitivity to small changes in the training data, ensemble methods tend to be very helpful. The disadvantages of ensemble methods are also significant, the biggest might be the lack of interpretation(Bühlmann, 2012). A linear combination of individual classifier is much harder to interpret than a single classifier.

In our research, we utilize two typical ensemble algorithms, AdaBoost and random forest, to train and predict the labeled spread crossing over in a fixed time interval of LOBs. The price spread crossing opportunities can be labeled as ask price lower, bid spread higher and no crossing over in a fixed time horizon. Therefore our problem is a multi-class classification case, the one against one and one against all methods will be introduced to solve the multi-class classification problem. The real time data is divided into two part with the percentage of 9:1, to make a training dataset and testing data set respectively. Besides, features with price, volume and order book arrival intensity of each price level are created, ,so every data sample in training and testing dataset is represented as a vector of features. Moreover, Precision, recall and F1 score are used to measure the performance of the models, since the existence of arbitrage in a relatively long time interval is rare and our dataset can be treated as an imbalanced data.

Experiments with real time data from NASDAQ show that the ensemble models built in our paper can not only predict the arbitrage opportunities with high accuracy, but also can improve the prediction performance compared with basic classifiers, such as logistic regression, support vector machine and decision trees. We also design a naive trading strategy in the testing time interval and demonstrate the Profit and Loss (PnL) of our models. The cumulative Pnl curve show that the traders can obtain positive return with zero investment, which indicates that the statistical arbitrages can be found in our models.



2.4 Purpose of the dissertation

CHAPTER 3

LITERATURE REVIEW

3.1 Limit order book dynamics and modeling

3.2 Machine learning methods on capturing limit order book dynamics

CHAPTER 4

MATHEMATICAL DESCRIPTION AND STATISTICAL PROPERTIES OF LIMIT ORDER BOOKS

In this chapter, we give precise mathematical descriptions of LOBs' trading and exchanging principles. Besides,some basic statistical properties of limit order books, which have been proposed by past research, are also studied and described. Those properties include:size of orders, shape of order books,time of arrival of orders, placement of orders and so on. Study for those properties will help us to get a clearer insight of how to capture the dynamics of LOBs. For example, by knowing the distribution of arrival of orders will help us to build more meaningful features in our prepdiction models.

4.1 Mathematical descriptions of LOBs

An LOB can be represent as a three dimensional vector by the following definition:

Definition 4.1.1 *An order* $x = (p_x, \omega_x, t - x)$ *submitted at time* t_x *which price* p_x *and size* $\omega_x >$ 0*(respectively,* $\omega_x < 0$*) is a commitment to sell (respectively, buy) up to* $|\omega_x|$ *units of the traded asset at a price no less than (respectively, no greater than)* P_x*.*

For a given LOB, the units of order size and price are defined as follows:

Definition 4.1.2 *The lot size of an LOB is the smallest amount of the asset that can be traded within it. All orders must arrive with a size* $\omega \in \{\pm k\sigma | k = 1, 2, ...\}$

Definition 4.1.3 *The tick size* π *of an LOB is the smallest permissible price interval between different orders within it. All orders must arrive with a price that is specified to the accuracy of* π

For example, if $\pi = 0.01$, then the largest permissible order price that is strictly less than \$1 is \$ 0.99, and all orders must be submitted aat a price with exactly two decimal places.

Definition 4.1.4 *The lot size σ and tick size π of an LOB are collectively called its resolution parameters.*

Definition 4.1.5 *When a buy (respectively, sell) order x is submitted, an LOB's trade-matching algorithm checks whether it is possible to match x to some other previously submitted sell(respectively, buy) order. If so, the matching occurs immediately. If not, x becomes active*

Active orders in a market make up an LOB:

Definition 4.1.6 *An LOB $\mathcal{L}(t)$ is the set of all active orders in a market at time t.*

An LOB can be treated as a set of queues, each of which contains active bid or ask orders at a specified price.

Definition 4.1.7 *The bid-side depth available at price p and at time t is:*

$$n^b(p,t) := \sum_{\{x \in \mathcal{B}(t) | p_x = p\}} \omega_x$$

The ask-side depth available at price p and at time t, denoted $n^a(p,t)$, is defined similarly using $\mathcal{A}(t)$

The depth available is often demonstrated as multiples of the lot size. Since $\omega_x < 0$ for bid orders and $\omega_x > 0$ for ask orders, it implies that $n^b(p,t) \leq 0$ and $n^a(p,t) \geq 0$ for all prices p.

Definition 4.1.8 *The bid-side depth profile at time t is the set of all ordered pairs(p,$n^b(p,t)$). The ask-side depth profile at time t is the set of all ordered pairs(p,$n^a(p,t)$).*

Definition 4.1.9 *The mean bid-side depth available at price p between times t_1 and t_2 is*

$$\bar{n}^b(p,t_1,t_2) := \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} n^b(p,t)dt$$

The mean ask-side depth available at price p between times t_1 and t_2, denoted as $\bar{n}^a(p,t_1,t_2)$, is defined similarly using the ask-side depth available.

In the following, we define the terms of *bid price, ask price, mid price, and bid-ask spread*

Definition 4.1.10 *The bid price at time t is the highest stated price among active buy orders at time t,*

$$b(t) := \max_{x \in \mathcal{B}(t)} p_x$$

The ask price at time t is the lowest stated price among active sell orders at time t,

$$a(t) := \min_{x \in \mathcal{A}(t)} p_x$$

Definition 4.1.11 *The bid-ask spread at time t is $s(t) := a(t) - b(t)$*

Definition 4.1.12 *The mid price at time t is $m(t) := [a(t) + b(t)]/2$*

More clearly, in an LOB, $b(t)$ is the highest price at which it is immediately possible to sell at least the lot size of the traded asset at time t, and $a(t)$ is the lowest price at which it is immediately possible to buy at least the lot size of the traded asset at time t. Considering prices relative to $b(t)$ and $a(t)$ is helpful in some cases.(Gould et al., 2013)

Definition 4.1.13 *For a given price p, the bid-relative price is $\delta^b(p) := b(t) - p$ and the ask-relative price is $\delta^a(p) := p - a(t)$*

Note that here is difference in signs between the definition of δ for the two sides:$\delta^b(p)$ defines how much smaller that p is less than $b(t)$ and $\delta^a(p)$ illustrates how much larger that p is greater than $a(t)$. It is useful that we compare the properties of orders on both the bid side and the ask side of an LOB.

Definition 4.1.14 *For a given order $x = (p_x, \omega_x, t_x)$, the relative price of the order is :*

$$\delta^x := \begin{cases} \delta^b(p_x) \; if \; the \; order \; is \; a \; buy \; order, \\ \delta^a(p_x) \; if \; the \; order \; is \; a \; sell \; order, \end{cases}$$

Definition 4.1.15 *The bid-side depth available at relative price p and at time t is:*

$$N^b(p,t) = \sum_{\{x \in \mathcal{B}(t) | \delta^x = p\}}$$

The ask-side depth available at relative price p and at time t, denoted $N^a(p,t)$, is defined similarly using $\mathcal{A}(t)$

Definition 4.1.16 *The bid-side relative depth profile at time t is the set of all ordered pairs*$(p, N^b(p,t))$. *The ask-side relative depth profile at time t is the set of all ordered pairs*$(p, N^a(p,t))$.

Definition 4.1.17 *The mean bid-side depth available at relative price p between times t_1 and t_2 is:*

$$\bar{N}^b(p, t_1, t_2) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} N^b(p,t)$$

The mean ask-side depth available at relative price p between times t_1 and t_2 denoted $\bar{N}^a(p, t_1, t_2)$ is defined similarly using the ask-side relative depth available.

Definition 4.1.18 *The mean bid-side relative depth profile between times t_1 and t_2 is the set of all ordered pairs*$(p, \bar{N}^b(p, t_1, t_2))$. *The mean ask-side relative depth profile between times t_1 and t_2 is the set of all ordered pairs $(p, \bar{N}^a(p, t_1, t_2))$*

while most traders use the relative depth profile to measure the phenomenon of LOBs, some research has claimed that the order arrival rates relies on relative prices rather than actual prices(Biais et al. (1995),Bouchaud et al. (2002),Potters and Bouchaud (2003),Zovko et al. (2002)), relative depth profiles do not contain information about the absolute prices at which trades occur. Moreover, relative depth profiles provide little information about the bid-ask spread and mid price. Therefore, it is better that we combine the relative depth profiles, bid price($b(t)$) and ask price($a(t)$) together when we consider the problem of LOBs. A completed view of the evolution of an LOB can be obtained if we consider all the information simultaneously.

4.2 Properties of LOBs

During the last two decades, financial market has become more and more computerized. Therefore, it is easier for us to access extensive data on order books. To our knowledge, Biais et al. (1995) is the pioneer to study the computerized data flows of Paris Bourse. After that, lot of related papers provide more empirical findings,statistical properties and modeling views.(See,e.g.**?**,**?**,**?**,Bouchaud et al. (2002),Potters and Bouchaud (2003)), In this section, we give a briefly introduction of some basic empirical study results. Those fundamental statistical properties can be found through observing the real time data. Some results of observations, such as time of arrival of orders, placement of orders, size of orders, shape of orders books and etc, are essential for calibrating the models of order flows and capturing the dynamics of order books.

4.2.1 Time of arrivals of orders

We compute the cumulative distribution for inter-arrival times of market orders for four stocks: Amazon,Google, Intel and Microsoft. The results are plotted in figure 4.1. From the figure, it is obvious that the Lognormal and Exponential distribution are not good fits. For the Weibull distribution, it has been suggested in Ivanov et al. (2004).From our results, we can see that weibull distribution(red line in our figure) is relatively a good fit to the original data. However, in our case, the Gamma distribution is the best fit for each stock.

In many past literatures, the non-poisson arrival times models have been introduced to deal with the "irregular" financial data. For example,Engle and Russell (1997) and Engle (2000) have proposed autoregressive condition intensity models, which are beneficial to capturing the processes of orders' submission. Another research area that deal with the non-exponential arrival times relies on the branches of stochastic processes(see,e.g. Clark (1973);Silva and Yakovenko (2007),Huth and Abergel (2012)).

4.2.2 Volume of orders

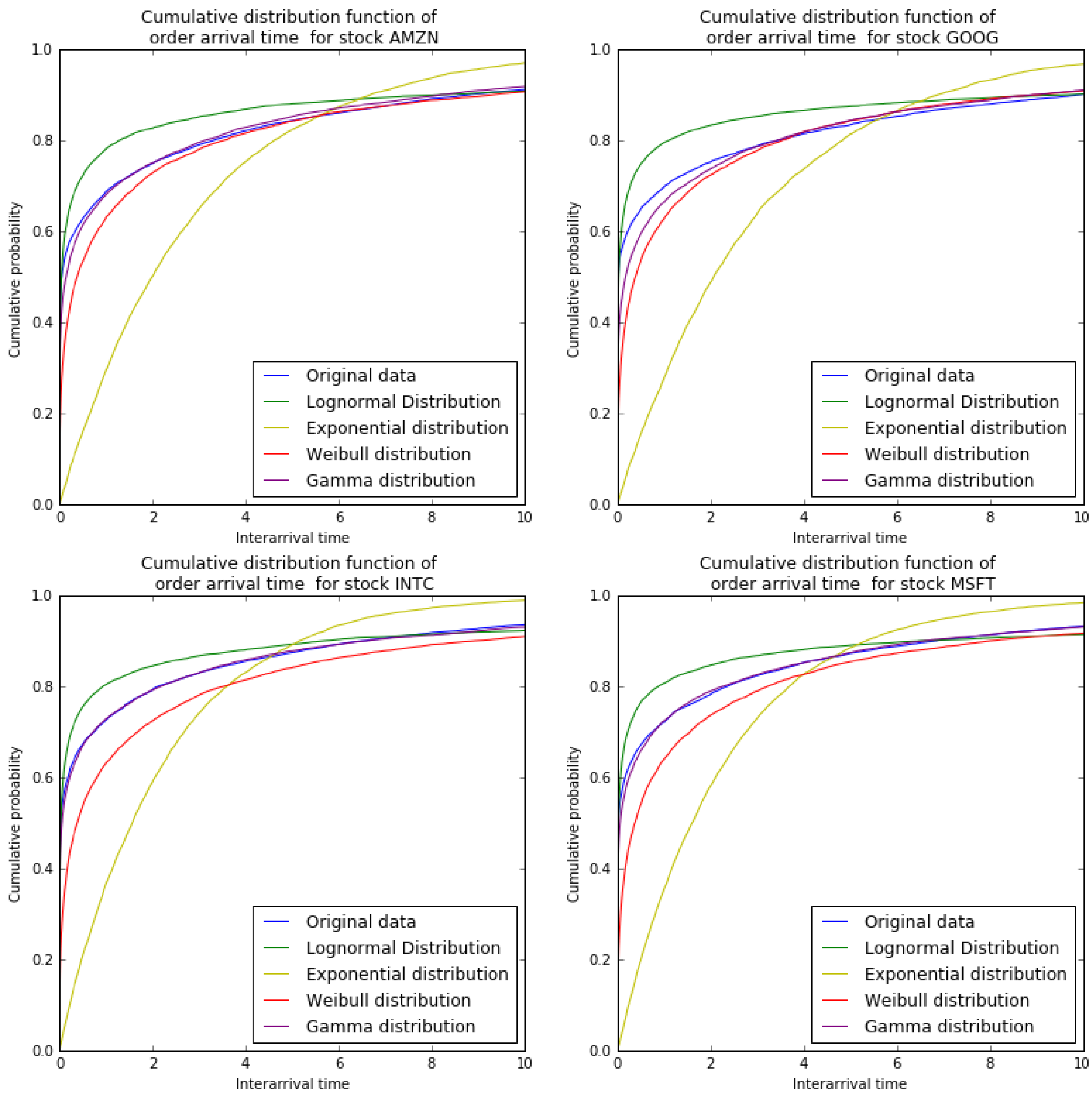


Figure 4.1: Cumulative distribution of inter-arrival time for stock: AMZN,GOOG,INTC and MSFT. In each panel,four distribution, Lognormal, Exponential Weibull and Gamma, are compared with the original dataset. x axis represents the inter-arrival time of market orders and y axis shows the cumulative probability

BIBLIOGRAPHY

Anuj Agarwal. High frequency trading: Evolution and the future. 2012.

Bruno Biais, Pierre Hillion, and Chester Spatt. An empirical analysis of the limit order book and the order flow in the paris bourse. *the Journal of Finance*, 50(5):1655–1689, 1995.

Jean-Philippe Bouchaud, Marc Mézard, Marc Potters, et al. Statistical properties of stock order books: empirical results and models. *Quantitative finance*, 2(4):251–256, 2002.

Peter Bühlmann. Bagging, boosting and ensemble methods. In *Handbook of Computational Statistics*, pages 985–1022. Springer, 2012.

Peter K Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: journal of the Econometric Society*, pages 135–155, 1973.

Rama Cont, Sasha Stoikov, and Rishi Talreja. A stochastic model for order book dynamics. *Operations research*, 58(3):549–563, 2010.

Robert F Engle. The econometrics of ultra-high-frequency data. *Econometrica*, 68(1):1–22, 2000.

Robert F Engle and Jeffrey R Russell. Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model. *Journal of empirical finance*, 4(2):187–212, 1997.

Robert F Engle, Robert Ferstenberg, and Jeffrey R Russell. Measuring and modeling execution cost and risk. 2006.

Martin D Gould, Mason A Porter, Stacy Williams, Mark McDonald, Daniel J Fenn, and Sam D Howison. Limit order books. *Quantitative Finance*, 13(11):1709–1742, 2013.

Lars Kai Hansen and Peter Salamon. Neural network ensembles. *IEEE transactions on pattern analysis and machine intelligence*, 12:993–1001, 1990.

Larry Harris. *Trading and exchanges: Market microstructure for practitioners*. Oxford University Press, USA, 2003.

Sherif Hashem. Optimal linear combinations of neural networks. *Neural networks*, 10(4):599–614, 1997.

Nicolas Huth and Frédéric Abergel. The times change: Multivariate subordination. empirical facts. *Quantitative Finance*, 12(1):1–10, 2012.

Plamen Ch Ivanov, Ainslie Yuen, Boris Podobnik, and Youngki Lee. Common scaling patterns in intertrade times of us stocks. *Physical Review E*, 69(5):056107, 2004.

Balázs Kégl. The return of adaboost. mh: multi-class hamming trees. *arXiv preprint arXiv:1312.6086*, 2013.

Andrei A Kirilenko, Albert S Kyle, Mehrdad Samadi, and Tugkan Tuzun. The flash crash: The impact of high frequency trading on an electronic market. *Available at SSRN 1686004*, 2015.

Hugh Luckock. A statistical model of a limit order market. *Sidney University preprint (September 2001)*, 2001.

Anna A Obizhaeva and Jiang Wang. Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1):1–32, 2013.

Marc Potters and Jean-Philippe Bouchaud. More statistical properties of order books and price impact. *Physica A: Statistical Mechanics and its Applications*, 324(1):133–140, 2003.

Ioanid Rosu. A dynamic model of the limit order book. *Forthcoming in The Review of Financial Studies*, 2009.

A Christian Silva and Victor M Yakovenko. Stochastic volatility of financial markets as the fluctuating rate of trading: An empirical study. *Physica A: Statistical Mechanics and its Applications*, 382(1):278–285, 2007.

Ilija Zovko, J Doyne Farmer, et al. The power of patience: a behavioural regularity in limit-order placement. *Quantitative finance*, 2(5):387–392, 2002.

BIOGRAPHICAL SKETCH

This is my biography.