Spring 5-2017

Statistical Analysis of the Price Jumps of Financial Assets Based on LOB Data

Ying Zhuang
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Mathematics

Statistical Analysis of the Price Jumps of Financial Assets Based on LOB Data

by

Ying Zhuang

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

May 2017
Saint Louis, Missouri

Contents

List of Tables

List of Figures

Acknowledgments

I would like to sincerely thank my thesis advisor Professor José E. Figueroa-López. He is such a friendly and patient professor and always provides me with lots of valuable ideas. Without his guidance and help, it would be impossible for me to accomplish this thesis. I would also like to thank my classmate Qi Wang, who taught me lots of useful skills of using R. Besides, I want to express my gratitude to my committee members Professor José E. Figueroa-López and Professor Todd Kuffner for their time and advices.

In addition, I would like to truly thank my dear parents. As I am growing up, I realize just how much they did for me. Their continuous encouragement and support make me fearless no matter what kind of difficulties I encounter.

Ying Zhuang

Washington University in Saint Louis
May 2017

Dedicated to my parents.

ABSTRACT OF THE THESIS

Statistical Analysis of the Price Jumps of Financial Assets Based on LOB Data

by

Ying Zhuang

Master of Arts in Statistics

Washington University in St. Louis, May 2017

Advisor: Professor José E. Figueroa-Lópes

The price process in electronic markets is one prototypical example of a stochastic process, and it has historically be fitted and analyzed using different stochastic models such as Lévy processes, diffusions, and SDEs (stochastic differential equations). In this thesis, we analyze Microsoft stock data in 2014-11-03 with the goal of studying the presence of jumps based on Limit Order Book (LOB) data. To this end, we divide the whole day's data into many consecutive intervals and proceed to apply a jump detection method to identify the intervals that could potentially have jumps. After obtaining the intervals with potential jumps, we zoom in these intervals and compare them in order to characterize their features. More specifically, we analyze the price LOB data from both the traditional side and the statistical side, and our aim is that try to identify statistical differences between the intervals with jumps and without any jump, and then give evidence to support this jump detection method and conjecture reasons for the appearance of sharp price changes in small intervals.

Key words: Jump Detection, Limit Order Book, Bi-power Variation, Price Process

Chapter 1

Introduction

In this chapter, we introduce different orders in electronic markets at the first section. After that, we describe the price process and introduce several LOB models and the jump detection method. At the end of this chapter, we give details of the LOB data.

1.1 Market Order and Limit Order

When it comes to electronic trading market, a company's stock is a general term which represents the ownership of the company. Companies use stocks to raise money and investors gain profits by buying or selling stocks. Traders will place orders as they want to execute transactions in the stock market. An order contains information about the number of shares and the price at which the investors wants to trade them.

There are two sides in the electronic markets, namely, bid and ask. Traders in the bid (ask) side are willing to buy (sell) certain number of shares at a specific price. Correspondingly, bid (ask) price is the traders' desired transaction price. The highest bid price is called the best bid price (q_1) and the lowest ask price is called the best ask price (p_1), or we call them both level one price. The average of them $S_t = \frac{q_1+p_1}{2}$ is defined as the mid-price. When traders want to trade in the market, there are two options to place their orders: Market Order or Limit Order.

Market orders (MOs) are considered more fundamental and of immediate execute. That is, a trader who wants to immediately buy or sell a certain quantity of shares at the best available price usually place a MO. Since, in principle, the market's affordability of trading is connected with the timing and size of MOs, the transactions may not be actually processed if the size of MOs is too large.

Compared to MOs, Limit Orders (LOs) are relative more passive. A trader places a LO indicates her desire to trade a certain quantity of shares at a specific price. LOs are usually not trade immediately and have to wait until either they are matched with new orders (execution) or they are withdrawn (cancellation). In general, the current market price is usually higher than the price of buying LOs and lower than the price of selling LOs.

All of the non-executed LOs are queued in a type of accounting book called limit order book (LOB) based on their price and the time they are placed. The execution occurs only if a new market order come in and match the LO order. Orders that are stay far away from the best quote are less likely to be executed. Hence traders may choose to select price close to the best bid or ask order in order to increase the probability of execution during some specified time range.

Every time a trader places a LO, she must choose price and amount for purchasing or selling at first. The designated price and quantity make LOs more controllable. The other convenient side of LOs is that traders can place LOs even if the market is closed. The LOs placed when the market is closed are set into a queue and they will be processed as soon as the market open. In one word, when traders place LOs they can select the prices they want and, thus, LOs are better than MOs if the specific price is a priority.

1.2 Price Process

A càdlàg (right-continuous with left limits) stochastic process $(X_t)_{t\geq 0}$ on a probability space $(\Omega, F, \mathbb{P})$ with values in R^d and $X_0 = 0$ is a Lévy process if it satisfies *independent increments*, *stationary increments*, and *stochastic continuity*. Due to these characters, a Lévy process could be treated as the continuous-time analog of a random walk. The Brownian motion process is one of the most well known examples of Lévy process, and usually be used for fitting a price process.

When it comes to price process, we ought to consider both diffusions and jumps and it is very important to consider their influences separately in applications. Suppose X is a stochastic process with initial value $x_0 \in \mathbb{R}$ at time $t = 0$, then X can be used to represent asset price if

$$dX_t = a_t dt + \sigma_t dW_t + dJ_t, \quad t \in [0, T]$$

where a and σ are measurable processes, W is a standard Brownian motion and J is a jump process.

In general, a jump process (J) is the sum of a *finite activity* (J_1) and an *infinite activity* component (J_2), i.e. $J = J_1 + J_2$, where *finite activity* indicates in each finite time interval a jump process owns a.s. a finite number of jumps. Otherwise we consider the jump process have *infinite activity*.

1.3 LOB Models

The simple uncorrelated random walk is similar to the Bernoulli trial process or the discrete time Brownian motion process is some respects. Suppose that $Z = (Z_1, Z_2, ...)$ is a sequence of independent random variables, and each of them take values 1 or -1 with probability $p \in [0, 1]$ and $1 - p$ respectively. Let $X = (X_0, X_1, X_2, ...)$ with

$$X_n = \sum_{i=1}^{n} Z_i, \quad n \in \mathbb{N}$$

Then the sequence X is the simple random walk with parameter p.

Suppose the parameter $p = \frac{1}{2}$. Now $X = (X_0, X_1, X_2, ...)$ is called the uncorrelated symmetric random walk. In an uncorrelated symmetric random walk, the price process assumed to go upward or downward one tick with the same probability $\frac{1}{2}$. The "tick" here is a measure of the minimum upward or downward movement in the price of a stock.

In order to analyze the pattern of two consecutive time intervals mid-price changes and compare the patterns among different type of intervals, in this thesis we try to use a correlated random walk moder to fit the mid-price change.

Under optimal conditions, we will first assume that $X = (X_1, X_2, ...X_t)$ $(1 \leq t \leq T)$ is a sequence of independent variables, each taking value 1 or -1 with $P(X_i = 1) = 1 - P(X_i = -1) = p$ and $p \in [0, 1]$. Let S_t be the mid-price at time t, measured as ticks. Assume that the mid-price increases or decreases at each time $t = 0, 1, ..., T$, and the movement range is always 1 tick, then

$$S_t = s_0 + \sum_{i=1}^{t} X_i, \quad S_0 = 0$$

In this thesis, we apply correlated random walk model to fit the mid-price change and proceed to use statistical tests to analyze the probability of consecutive mid-price changes.

1.4 Jump Detection Method

Power Variations is first defined by Berman (1965) as estimator of the sum of given powers of the jumps. Barndorff-Nielsen & Shephard (2004, 2006) introduced the Multi-power Variation (MPV) and the Bi-Power Variation (BPV) originally and use them to estimate the integrated volatility $\int_0^T \sigma_t^p \, dt$ for a given p. The BPV (p=2) is a precursor of the general MPV.

The electronic market opens at 9:30 and closes at 16:00, during the 6.5 hours we can divide the time into many consecutive and length-equal intervals. For both ends of each interval we can obtain a mid-price $S_t = \frac{q_1+p_1}{2}$, where q_1 is the best bid price and p_1 is the best ask price. It is clear that two consecutive intervals are going to share one same mid-price. For the n consecutive intervals the consecutive increments $\{\Delta X_i, \ 1 \leq i \leq n\}$ are defined as

$$\Delta X_i = log(S_i) - log(S_{i-1}) = log(\frac{S_i}{S_{i-1}})$$

For the purpose of recovering the integrated volatility $\int_0^T \sigma_t^2 \, dt$ of a continuous process without jump, we can make use of the traditional realized quadratic variation

$$RQV = \sum_{i=1}^{n} |\Delta X_i|^2 = \sum_{i=1}^{n} |log(\frac{S_i}{S_{i-1}})|^2$$

However, this purpose can not be achieved if jump occurs. Hence, On the basis of RQV, the estimator BPV is proposed

$$BPV = \frac{\pi}{2}\sum_{i=1}^{n-1}|\Delta X_i||\Delta X_{i+1}| = \frac{\pi}{2}\sum_{i=1}^{n-1}|log(\frac{S_i}{S_{i-1}})||log(\frac{S_{i+1}}{S_i})|$$

The heuristic of BPV comes from the fact that jumps are unlikely to occurs in both of two adjacent intervals, in which case $|\Delta X_i||\Delta X_{i+1}| \approx |\Delta X_i|^2$. In other cases, even if one interval have jumps in $[t_{i-1}, t_{i+1}]$, we still have $|\Delta X_i||\Delta X_{i+1}| \approx 0$. The constant part $\frac{\pi}{2}$ of BPV is needed for the consistency of the estimator.

The most popular form of a power shreshold is Ch^α with $C > 0$ and $\alpha \in (0, 0.5)$. In this thesis we use a shreshold that has been typically done in existing work

$$Ch^\alpha = 4\sqrt{\frac{BPV}{T}} \times h^{0.49}$$

where T=6.5 (9:30-16:00), and h is the length of each interval measured in hour.

We think the i^{th} interval may contain at least one jump if $|\Delta X_i| > Ch^\alpha$. Otherwise, it is reasonable to assume that there is no jump in i^{th} interval. More specifically, $\Delta X_i > Ch^\alpha$ indicates the likely presence of a up jump while if $\Delta X_i < Ch^\alpha$ there is possible a down jump.

1.5 Introduction of LOB data

The data we use is Microsoft stock during the day 2014-11-03. The original data contains seven variables *Time*, *ID tag for initial posted message*, *Message Type*, *Number of shares*, *Price*, *Exchange*, *BidSide Flag*.

Table 1.1: The first three rows of modified data

Seconds	Nanoseconds	Best Bid Price	Volume at Best Bid	Best Ask Price	Volume at Best Ask
34200	8078523	467300	100	467500	2695
34200	8087812	467300	100	467500	1495
34200	8222545	467300	100	467500	1395

$Time$ measures the number of milliseconds from midnight, so the market open at 9:30:00.000 with $Time$=34200000 and close at 16:00:00.000 with $Time$=57600000. *ID tag for initial posted message* is an ID number for every posted order in the limit order book that can be used to keep track of the order. *Message Type* is encoded by numbers and there are 8 different messages in total:

66: 'B' – Add buy limit order on the bid side; 83: 'S' – Add sell limit order on the ask side; 69: 'E' – Execute outstanding market order in part; 70: 'F' – Execute outstanding market order in full; 67: 'C' – Cancel outstanding LOB order in part; 68: 'D' – Delete outstanding LOB order in full; 88: 'X' – Bulk volume for the cross event; 84: 'T' – Execute non-displayed order.

Number of shares and $Price$ are measured in units and 0.01 cents respectively. *Exchange* always equal to 1 indicates that the stock is in NASDAQ. *BidSide Flag* has only two different values 0 or 1. Zero represents the messages was posted on the ask side and one indicates the bid side.

By python, we cleaned the original data and extract a modified data which is an $n \times 6$ matrix form where the six columns are: *Seconds*, $Nanoseconds$, *Best Bid Price*, *Volume at Best Bid*, *Best Ask Price*, *Volume at Best Ask*. *Seconds* is the same as Time, $Nanoseconds$ are since the last seconds. *Best Bid Price* and *Best Ask Price* are the level one prices at a given time and measured in 0.01 cents. *Volume at Best buy* and *Volume at Best Ask* are volumes under the best prices. For example, table 1.1 is the first three rows of the data matrix, each row represent a state of LOs at the best ask price and the best bid price at a specific time.

Chapter 2

Detection of Jumps

This chapter aims to apply the jump detection method to LOB data to obtain intervals with potential jumps. We separate the whole LOB data into many consecutive intervals with time length equal. There are 4 different time length of the intervals: 5-minute interval, 1-minute interval, 0.5-minute interval, and 10-second interval. For each time length we will apply the jump detection method once and find out which intervals have potential jumps.

2.1 5-minute Intervals

The range of our data is from 9:30 to 16:00, so T=6.5 hour. At first, we separate the data into 78 consecutive 5-minute intervals. We show in Figure 2.1 the cumulative sum of increments $\Delta X_i = log(S_i) - log(S_{i-1}) = log(\frac{S_i}{S_{i-1}})$ where $S_i = \frac{q_1+p_1}{2}$, q_1 is the best bid price and p_1 is the best ask price at time $t = i$.

After calculation we can obtain

$$BPV = \frac{\pi}{2} \sum_{i=1}^{77} |log(\frac{S_i}{S_{i-1}})||log(\frac{S_{i+1}}{S_i})| = 6.239 \times 10^{-5}$$

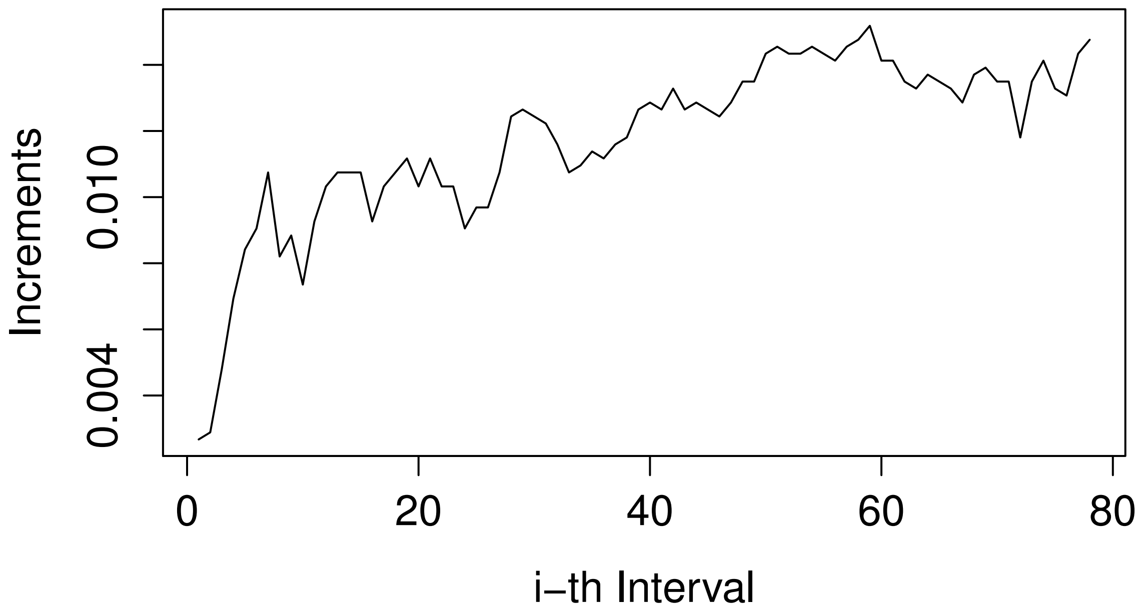

Figure 2.1: Trend of increments of 5-minute intervals

This corresponds to an initialized volatility of the process. Hence we can calculate the shreshold

$$Ch^{\alpha} = 4\sqrt{\frac{BPV}{6.5}} \times (\frac{5}{60})^{0.49} = 0.00367$$

which is about three "ticks".

The maximal and minimal numbers of ΔX_i are 0.00267 and -0.00254 respectively. Figure 2.2 shows the increment ΔX_i of each 5-minute interval. According to the jump-detection algorithm we mentioned in chapter one, since both the absolute values of 0.00267 and -0.00254 are less than 0.00367, there is no interval detected with a jump of all the 78 intervals.

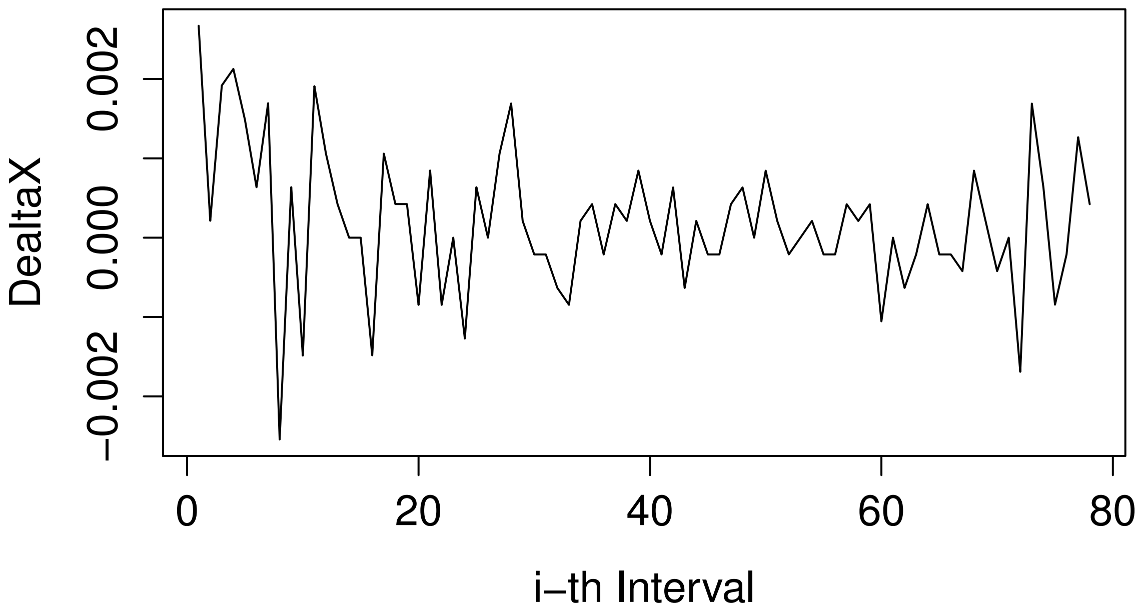

Figure 2.2: Increments of 5-minute intervals

2.2 1-minute Intervals

When we set 1 minute as the length of interval, we can get 390 intervals in total. In this case

$$BPV = \frac{\pi}{2} \sum_{i=1}^{389} |log(\frac{S_i}{S_{i-1}})||log(\frac{S_{i+1}}{S_i})| = 5.506 \times 10^{-5}$$

$$Ch^{\alpha} = 4\sqrt{\frac{BPV}{6.5}} \times (\frac{1}{60})^{0.49} = 0.00157$$

Figure 2.3 shows the cumulative sum of increments of these 390 intervals.

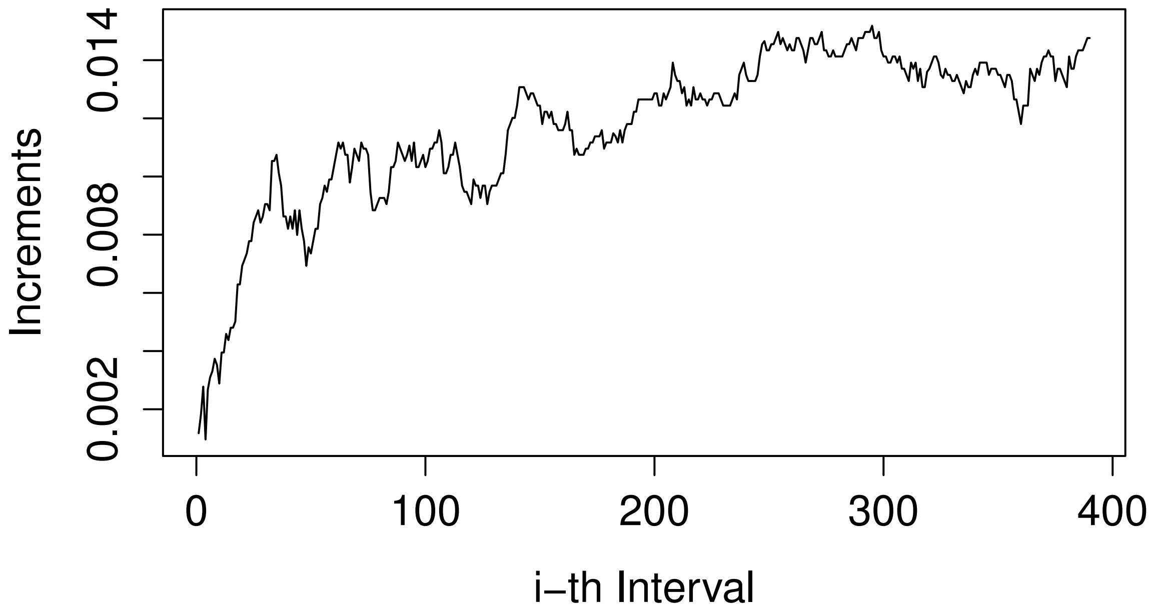

Figure 2.3: Trend of increments of 1-minute intervals

There are three intervals detected with potential jumps: 4^{th} interval (9:33-9:34), 5^{th} interval (9:34-9:35), and 33^{th} interval (10:02-10:03).

$$\Delta X_4 = log(\frac{S_4}{S_3}) = -0.00182 < -Ch^{\alpha}$$

$$\Delta X_5 = log(\frac{S_5}{S_4}) = 0.00171 > Ch^{\alpha}$$

$$\Delta X_{33} = log(\frac{S_{33}}{S_{32}}) = 0.00170 > Ch^{\alpha}$$

which indicate that the 4^{th} interval is detected with jump down and the other two intervals are detected with jump up. Figure 2.4 shows ΔX_is of all these intervals together with shresholds (the horizontal straight lines).

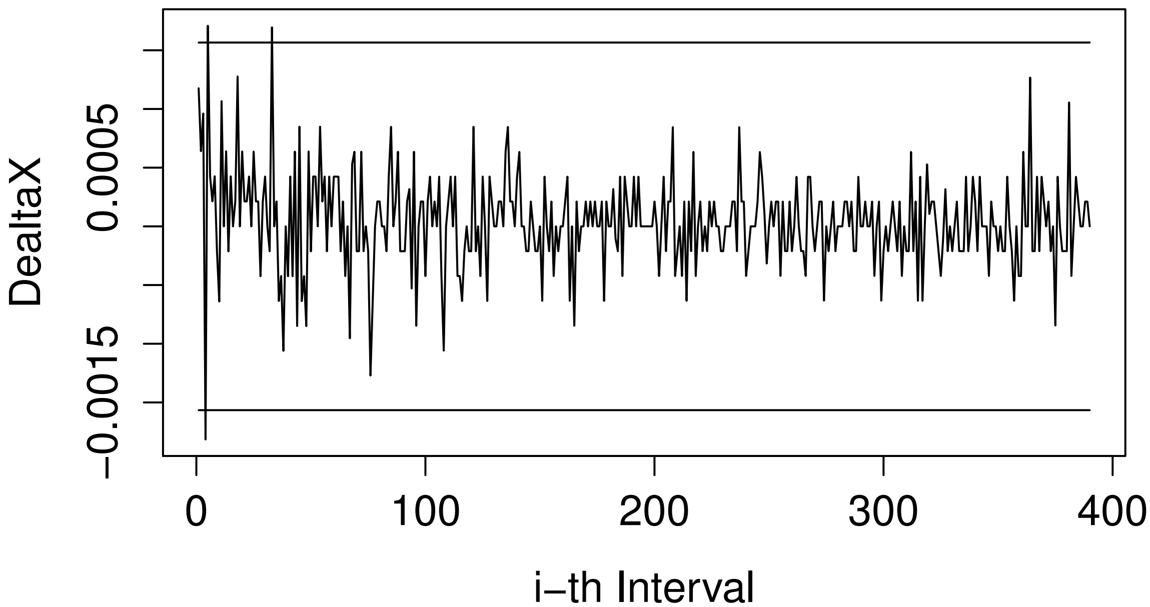

Figure 2.4: Increments of 1-minute intervals together with shresholds

2.3 0.5-minute Intervals

For 0.5-minute interval, we have 780 intervals in total. Then, we have

$$BPV = \frac{\pi}{2}\sum_{i=1}^{779} |log(\frac{S_i}{S_{i-1}})||log(\frac{S_{i+1}}{S_i})| = 5.746 \times 10^{-5}$$

$$Ch^{\alpha} = 4\sqrt{\frac{BPV}{6.5}} \times (\frac{0.5}{60})^{0.49} = 0.00114$$

Among the 780 intervals, there are still three intervals detected with jumps: 8^{th} interval (9:33:30-9:34:00), 9^{th} interval (9:34:00-9:34:30), 86^{th} interval (10:12:30-10:13:00).

$$\Delta X_8 = log(\frac{S_8}{S_7}) = -0.00171 < -Ch^{\alpha}$$

$$\Delta X_9 = log(\frac{S_9}{S_8}) = 0.00160 > Ch^{\alpha}$$

$$\Delta X_{86} = log(\frac{S_{86}}{S_{85}}) = 0.00127 > Ch^{\alpha}$$

which indicate the 8^{th} interval is jump down, 9^{th} and 86^{th} intervals are jump up. We show the cumulative sum of increments of all the 780 intervals in Figure 2.5.

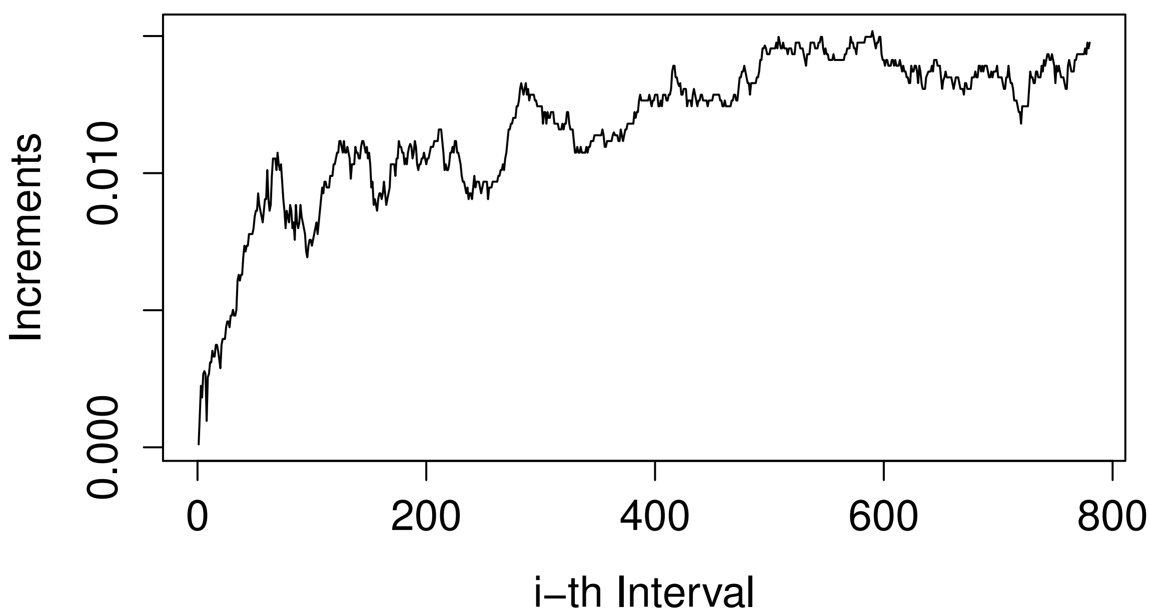

Figure 2.5: Trend of increments of 0.5-minute intervals

Figure 2.6 gives information about ΔX_i of all these intervals together with shresholds. Since the three intervals detected with potential jumps are all belong to the first 100 intervals, we give a magnification, Figure 2.7, of Figure 2.6 that only contains the first 100 ΔX_is of 0.5-minute intervals together with shresholds

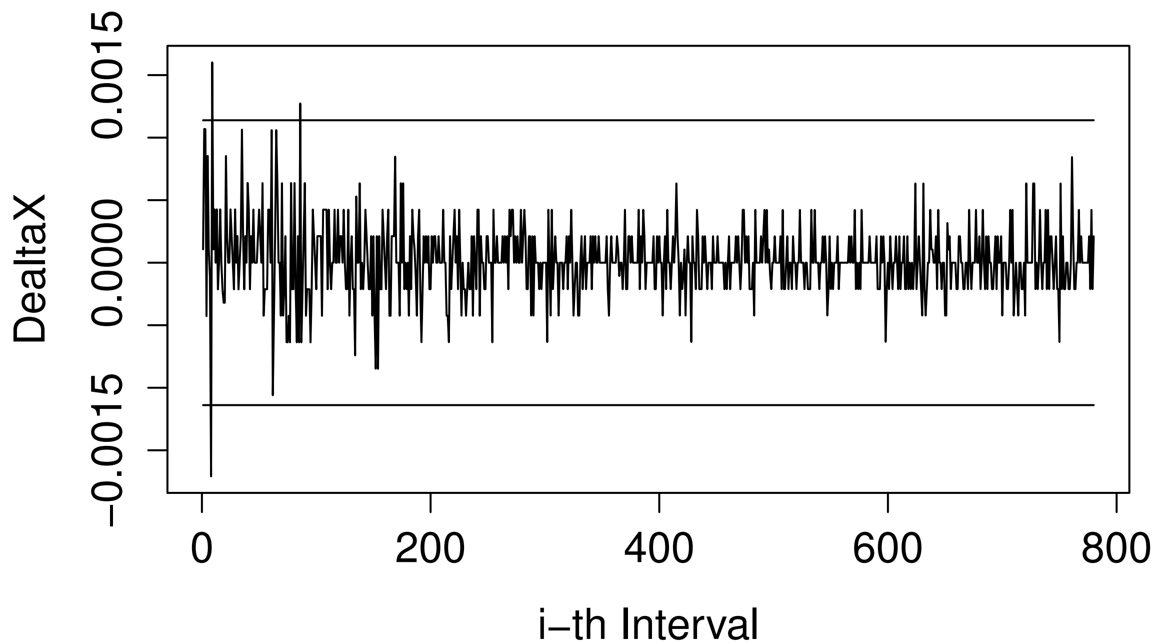

Figure 2.6: Increments of 0.5-minute intervals together with shresholds

We find out three intervals with potential jumps in both 1-minute intervals and 0.5-minute intervals respectively. In order to analyze the connection and difference between them, we put the two group of intervals in one plot. Figure 2.8 shows the cumulative sum of increments of 1-minute intervals (solid line) and 0.5-minute intervals (dashed line). Figure 2.9 is a magnification of figure 2.8 contains intervals with potential jumps.

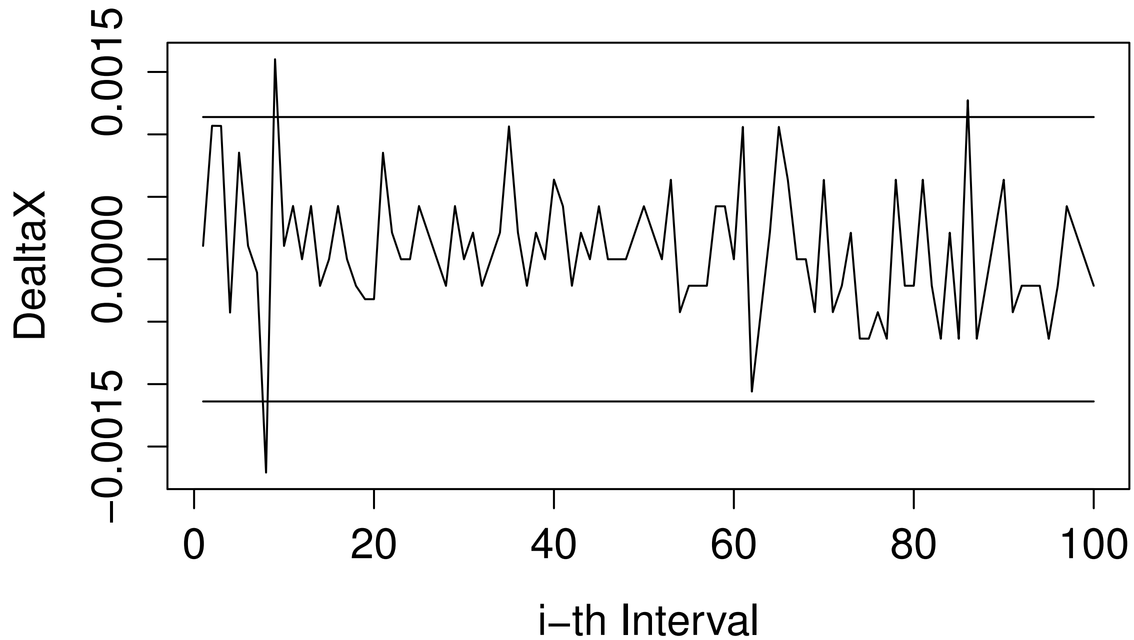

Figure 2.7: The first 100 increments of 0.5-minute intervals together with shresholds

It is clear that the 8^{th} 0.5-minute interval and 9^{th} 0.5-minute interval belong to 4^{th} 1-minute interval and 5^{th} 1minute-interval. From the first plot in Figure 2.9 we can see that the mid-price change of 7^{th} 0.5-minute interval is very small, which can explain why we detect a jump down in the 4^{th} 1-minute interval and also detect a jump down in the 8^{th} 0.5-minute interval. Similarly, it makes sense that we detect the 5th 1minute-interval jump up and then detect the 9th 0.5-minute interval jump up.

As shown before, we detect a potential jump up in the 33^{th} 1-minute interval, but neither the 65^{th} 0.5-minute interval nor the 66^{th} 0.5-minute interval were flagged as having any jump. Take a look at the second plot of Figure 2.9, mid-price changes of the 65^{th} 0.5-minute interval and the 66^{th} 0.5-minute interval are relative equal, which means that $\Delta X_{65}^{0.5-min}$ and

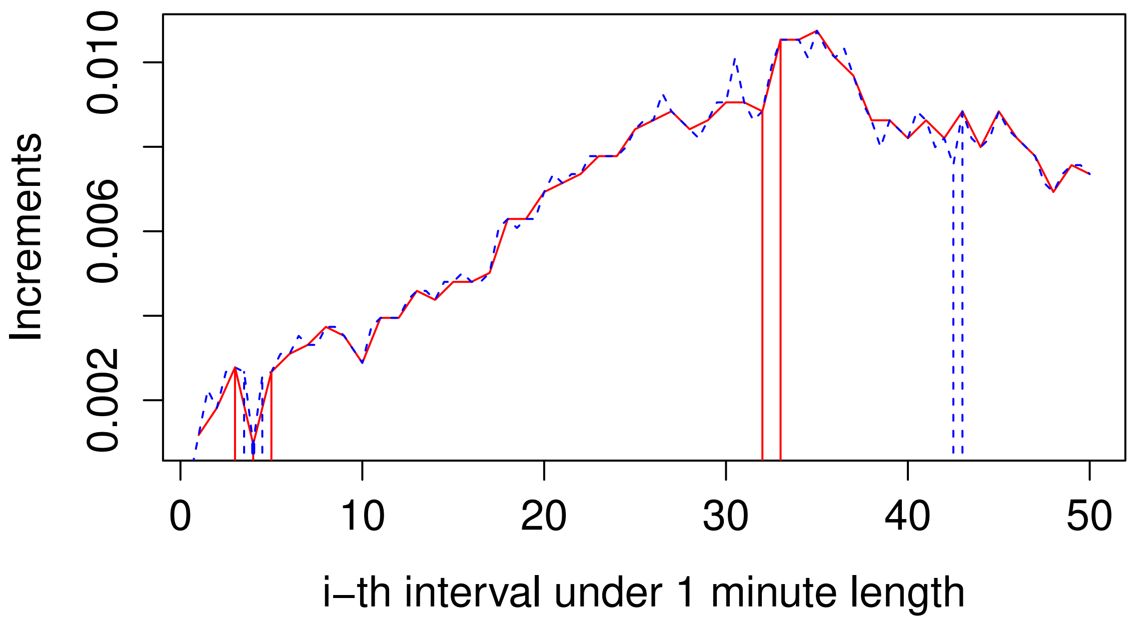

Figure 2.8: Trend of increments of 1-minute intervals and 0.5-minute intervals

$\Delta X_{66}^{0.5-min}$ have similar weight on influencing ΔX_{33}^{1-min}. With all of $\Delta X_{65}^{0.5-min}$,$\Delta X_{66}^{0.5-min}$, and ΔX_{33}^{1-min} are positive and $Ch_{0.5-min}^{\alpha}$ and Ch_{1-min}^{α} are similar, thus although the 33^{th} 1-minute interval jump up ($\Delta X_{33}^{1-min} > Ch_{1-min}^{\alpha}$), it is not unreasonable that the 65^{th} 0.5-minute interval and the 66^{th} 0.5-minute interval have no jumps ($|\Delta X_{65}^{0.5-min}| < Ch_{0.5-min}^{\alpha}$ and $|\Delta X_{66}^{0.5-min}| < Ch_{0.5-min}^{\alpha}$).

According to the third plot of Figure 2.9, we can see that the trend of mid-price from 10:12 (beginning of the 43^{th} 1-minute interval) to 10:13 (beginning of the 44^{th} 1-minute interval) is up. But actually, when we observe the trend based on 0.5-minute interval, it is clear that the mid-price decreases at the first half of the 43^{th} 1-minute interval and then goes up at the second half. So $|\Delta X_{86}^{0.5-min}| = |\Delta X_{85}^{0.5-min}| + |\Delta X_{43}^{1-min}|$ with $|\Delta X_{43}^{1-min}| < Ch_{1-min}^{\alpha}$ and

$|\Delta X_{86}^{0.5-min}| > Ch_{0.5-min}^{\alpha}$. Therefore, we detect the 86^{th} 0.5-minute interval jump up but 43^{th} 1-minute interval does not have jump.

2.4 10-second Intervals

Although we found 3 intervals having potential jumps when we separate the whole day data into 780 0.5-minute intervals, the sample is relatively small to do any statistical analysis. In order to have a larger sample of jump intervals, we set 10 seconds as the length of each interval. Figure 2.10 shows the trend of increments of all 10-second intervals. Figure 2.11 shows ΔX_is of all intervals together with shresholds.

For time frequency 10 seconds, $BPV = 4.889 \times 10^{-5}$ and $Ch^{\alpha} = 0.000613$. Out of all the 2340 intervals, we detect 32 intervals as having potential jumps. Among these 32 intervals, 17 intervals jump up and 15 intervals jump down. Table 2.1 indicates which intervals with potential jumps.

Table 2.1: 10-second intervals with potential jumps

Jump up intervals (i^{th})	1, 13, 26, 105, 119, 181, 186, 210, 221, 234, 258, 318, 521, 811, 1893, 1954, 2281
Jump down intervals (i^{th})	23, 24, 183, 184, 185, 187, 220, 224, 247, 259, 280, 284, 454, 461, 642

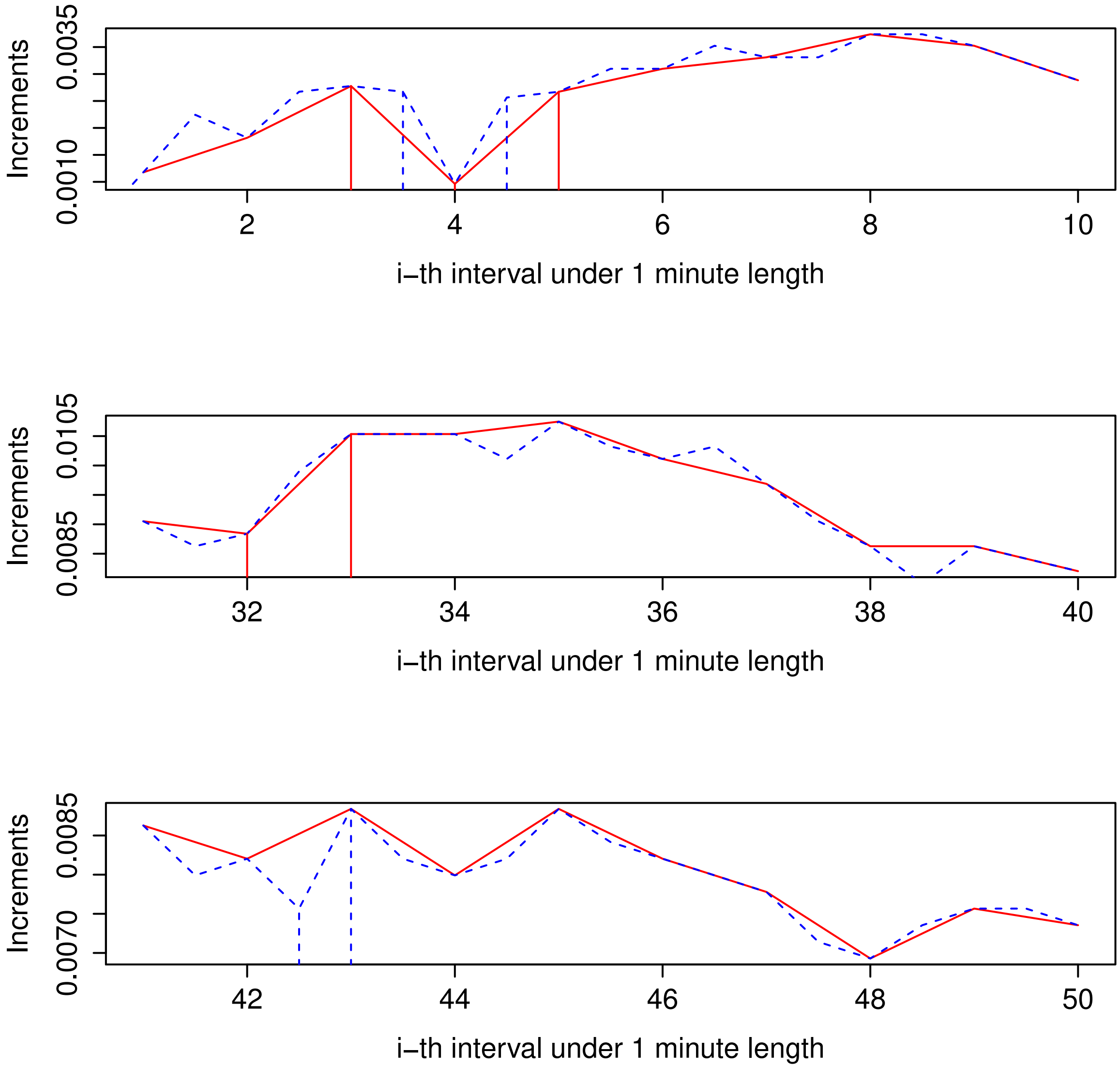

Figure 2.9: Magnification of Figure 2.8

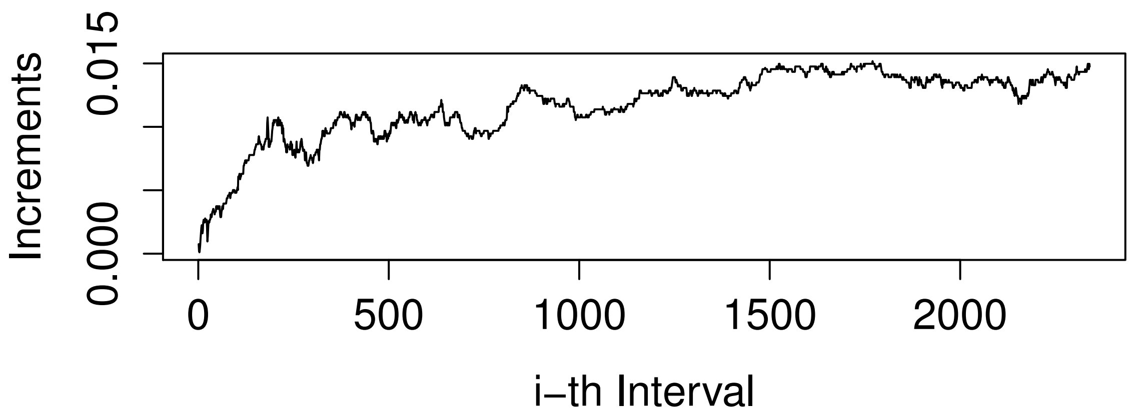

Figure 2.10: Trend of increments of 10-second intervals

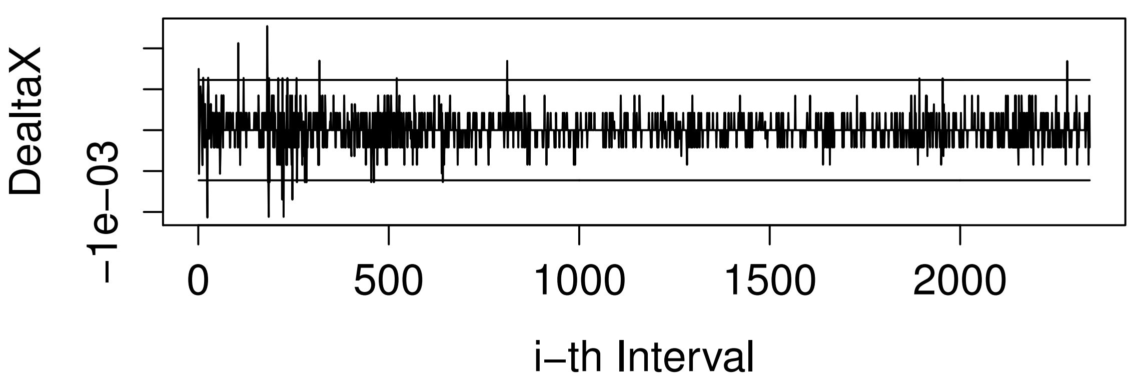

Figure 2.11: Increments of 10-second intervals

Chapter 3

Analysis of jumps

This chapter contains two sections. The first section compares the numbers of level one prices changes of intervals detected with potential jumps and without any jump. The second section makes use of proportion test and binomial test in order to find out the differences of probability of mid-price change trend in different type of intervals.

3.1 Analysis of Level One Prices Changes

In this section, we count the numbers of changes of the best bid price and the best ask price for each interval. By comparing the numbers of times these level one prices changes of intervals having potential jumps and intervals without any jump, we are going to find out that whether the number of level one prices changes is a significant feature of different type of intervals.

As mentioned above, there are three 1-minute intervals detected with potential jumps: 4^{th}, 5^{th}, and 33^{th}. Table 3.1 shows the number of times level one prices changes of these three intervals, and Table 3.2 gives the frequencies of changes of level one prices of the other 387 1-minute intervals without jump. We show the histogram of frequencies in Tables 3.1 in

Figure 3.1. Note: In Table 3.2 and Figure 3.1, BBP meas the number of intervals that the best bid price change *Freq.* times, BAP means the number of intervals that the best ask price change *Freq.* times.

Table 3.1: Frequencies of level one prices changes of intervals without jump

Freq.	0	1	2	3	4	5	6	7	8	9	10	11
BBP	30	47	65	52	31	25	22	16	16	10	10	7
BAP	25	32	55	43	33	28	22	20	18	14	9	10
Freq.	12	13	14	15	16	17	18	19	20	21	22	23
BBP	3	8	4	4	7	3	1	3	1	1	2	1
BAP	5	5	8	7	4	2	6	5	1	1	5	2
Freq.	24	25	26	27	28	31	32	33	34	35	37	38
BBP	1	2	1	2	1	3	0	1	2	0	0	0
BAP	1	3	2	1	1	1	3	1	1	1	2	3
Freq.	40	43	56	70	80	120	134	168	174	181	230	
BBP	1	0	0	0	1	1	1	0	0	1	0	
BAP	0	1	1	1	0	0	0	1	1	0	1	

From table 3.1 we can see that the number of price changes of these intervals are relative small. Since limit orders can be placed in a queue during the market is closed and they will be processed as quick as the market open, it is rational that there will be more executions of limit orders when the market is just open and thus level one prices are more likely to change at the beginning of market is open. In fact, the intervals without jump which have a relative large number of times the level one prices changes are the intervals represent the market just open, say, the first three intervals. The best bid prices in the first three intervals (9:30-9:33) change 181 times, 134 times, and 120 times, the best ask prices change 230 times, 168 times, and 174 times.

For 4^{th}, 5^{th}, and 33^{th} intervals, the average number of the best bid price change is $\frac{96+22+32}{3} = 50$, the average number of the best ask price change is $\frac{162+33+30}{3} = 75$. For the other 387 intervals, the average number of the best bid price change is 7.116, the average number of

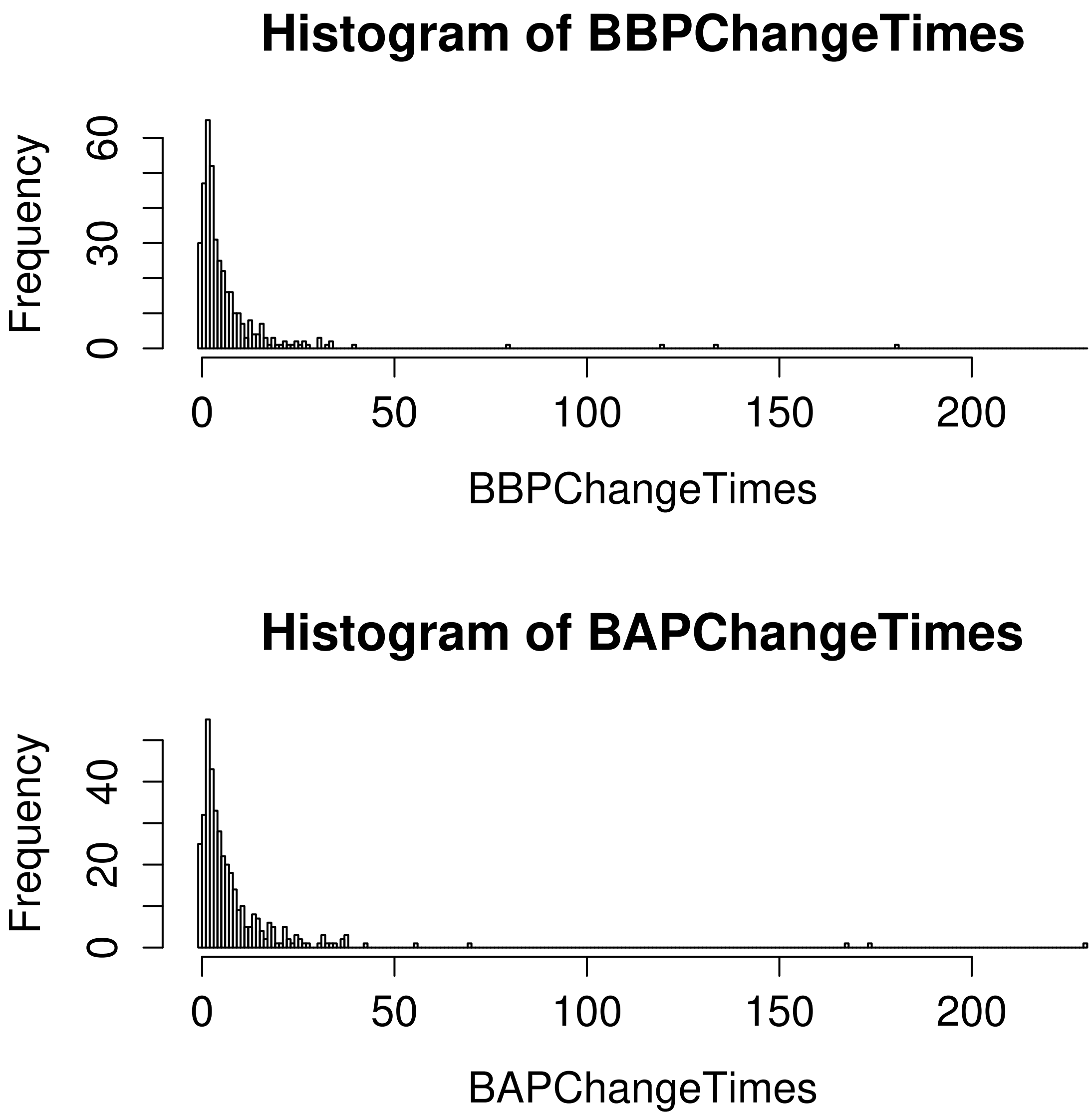

Figure 3.1: Histogram of frequencies in Tables 3.1

the best ask price change is 9.119. If we do not consider the first three intervals, then for the other 384 intervals, the average number of the best bid price change is 6.039, the average number of the best ask price change is 7.701.

It is obvious that there are much more price changes within the intervals detected with jumps. In order to check this argument, let us do the same analysis of 0.5-minute intervals and 10-second intervals.

There are also three 0.5-minute intervals detected with potential jumps: 8^{th}, 9^{th}, and 86^{th}. The average number of the best bid price change of intervals with potential jumps is $\frac{58+17+7}{3} = 27.333$, of the best ask price change is $\frac{105+18+9}{3} = 44$. For the other 777 intervals without jump, the average number of the best bid price change is 3.687, of the best ask price change is 4.758.

For 10-second intervals, there are 32 intervals have jumps. The average number of the best bid price change of these intervals is 11.375, of best ask price change is 13.063. For the other 2308 intervals without jump, the average number of the best bid price change is 1.195, of the best ask price change is 1.564.

According to the results, the intervals with jumps have more level one price changes than the intervals without jump. The number of changes of 86^{th} is small (the magnitude of the change is large and we have discussed about it in chapter two, and compare to the average number of changes of intervals without jump, 7 and 9 are still large enough), which means that we cannot categorize one interval arbitrarily by the number of level one price change. However, the number of level one price change have a certain connection with the jump. In general, one interval with large number of level one price change is more likely to be detected with potential jumps.

3.2 Statistical Tests for the Probability of Mid-price Change

In the section, we do proportion test and binomial test on the probability of mid-price change in order to find out the difference between the change probability and 0.5.

For a mid-price at time t, it will have three different states at next time $t+1$: remain the same, go upward, go downward. In order to be able to better analysis, we only consider the mid-price go upward or go downward. That is to say if mid-price at time t, $t+1$, and $t+2$ are S_0, S_0, and S_1, then we treat t as the first state t', and $t+2$ as the second state $t'+1$, S_0 and S_1 are mid-prices at the two states.

Let us define

$$P_{up,up} = Prob(next\ mid-price\ change\ up \mid mid-price\ change\ up)$$

$$P_{up,down} = Prob(next\ mid-price\ change\ down \mid mid-price\ change\ up)$$

$$P_{down,up} = Prob(next\ mid-price\ change\ up \mid mid-price\ change\ down)$$

$$P_{down,down} = Prob(next\ mid-price\ change\ down \mid mid-price\ change\ down)$$

Let N_1 and N_2 be the numbers of mid-price change up or down during $[t_i, t_{i+1}]$ respectively. And n_{11}, n_{12} are the numbers of next mid-price change up or down conditioned on a change up mid-price during $[t_i, t_{i+1}]$. Similarly, n_{21} and n_{22} are the numbers of next mid-price change up or down conditioned on a change down mid-price during $[t_i, t_{i+1}]$. Hence $N_1 = n_{11} + n_{12}$

and $N_2 = n_{21} + n_{22}$, and the estimated probabilities are

$$\hat{P}_{up,up} = \frac{n_{11}}{N_1}, \quad \hat{P}_{up,down} = \frac{n_{12}}{N_1}$$

$$\hat{P}_{down,up} = \frac{n_{21}}{N_2}, \quad \hat{P}_{down,down} = \frac{n_{22}}{N_2}$$

We use the 10-second intervals as our sample since there are more intervals (32 intervals) detected with jumps. At the first we categorize the sample into three sub-samples. Sample one contains all of the 17 intervals detected with up jumps, sample two contains all of the 15 intervals detected with down jumps, sample three contains all the 2308 intervals without jump. Then for each interval we can obtain a $\hat{P}_{up,up}$ and a $\hat{P}_{down,down}$. Our aim is to compare $\hat{P}_{up,up}$ and $\hat{P}_{down,down}$ of these three samples.

For each sample, we do three proportion tests and three binomial tests. Binomial test is a test of a simple null hypothesis about the probability of success in a Bernoulli experiment. Proportion test is a test that can be used for testing the null that the proportions (probabilities of success) in several groups are the same, or that they equal certain given values. Considering the sample size of intervals with potential jumps is not very large, we do both of the two tests on each sample. We want to use correlated random walk model to fit the mid-price change and the next step can only be walk upward or walk downward, then in order to find out the tendency of next "walk", we choose probability 0.5 as target. The null hypothesis of three tests are the same H_0: The probability is equal to 0.5. The alternative hypothesis of three tests are ① H_1: The probability is not equal to 0.5; ② H_1: The probability is greater than 0.5; ③ H_1: The probability is less than 0.5. And we use P-value=0.05 as significant level. Table 3.2, Table 3.3, and Table 3.4 give the results of proportion test and binomial test on different type of intervals.

According to table 3.2, 3.3, and 3.4, we can conclude that the intervals with potential upward jumps have $\hat{P}_{up,up} > 0.5$ and $\hat{P}_{down,down} = 0.5$, the intervals with downward jumps have $\hat{P}_{up,up} = 0.5$ and $\hat{P}_{down,down} > 0.5$, the intervals without jump have$\hat{P}_{up,up} = 0.5$ and $\hat{P}_{down,down} = 0.5$.

Table 3.2: Results of proportion test and binomial test on intervals with up jumps

H_0: The probability is equal to 0.5		P-value of ①	P-value of ②	P-value of ③
Proportion test	$P_{up,up}$	1.602e-06	8.012e-07	1
	$P_{down,down}$	0.3737	0.8131	0.1869
Binomial test	$P_{up,up}$	1.848e-06	9.239e-07	1
	$P_{down,down}$	0.4149	0.8324	0.2075
Estimated value		$\hat{P}_{up,up}$=0.642, $\hat{P}_{down,down}$=0.467		

Table 3.3: Results of proportion test and binomial test on intervals with down jumps

H_0: The probability is equal to 0.5		P-value of ①	P-value of ②	P-value of ③
Proportion test	$P_{up,up}$	0.1776	0.9112	0.08882
	$P_{down,down}$	1.101e-09	5.506e-10	1
Binomial test	$P_{up,up}$	0.2132	0.9269	0.1066
	$P_{down,down}$	9.259e-10	4.629e-10	1
Estimated value		$\hat{P}_{up,up}$=0.430, $\hat{P}_{down,down}$=0.721		

Table 3.4: Results of proportion test and binomial test on intervals without jumps

H_0: The probability is equal to 0.5		P-value of ①	P-value of ②	P-value of ③
Proportion test	$P_{up,up}$	0.06054	0.3027	0.6973
	$P_{down,down}$	0.5233	0.7383	0.2617
Binomial test	$P_{up,up}$	0.6188	0.3094	0.7039
	$P_{down,down}$	0.5364	0.7448	0.2682
Estimated value		$\hat{P}_{up,up}$=0.505, $\hat{P}_{down,down}$=0.494		

Chapter 4

Conclusions

In this thesis, we aim to analyze LOB data to explain the presence of potential "jumps" in the dynamics of asset prices. To this end, first we apply a jump detection procedure commonly used for macro financial models based on Brownian Motion driven SDE. Second, we compare the numbers of level one prices changes of intervals with potential jumps and intervals without jump. Third, we make use of proportion test and binomial test to analyze the probabilities of mid-price change.

As we divide the intervals shorter, the shreshold Ch^{α} becomes smaller (0.00367 for 5-minute interval, 0.00157 for 1-minute interval, 0.00114 for 0.5-minute interval, 0.000613 for 10-second intervals), and we detect more intervals with jumps. Hence the length of interval is one of the factors that have much influence on the detection method.

Another feature of the intervals is that the best bid price and the best ask price of intervals with jumps usually change more frequently than the prices of intervals without jump. The price change may due to big orders come into the limit order book, or there is sparsity between limit orders.

In addition, according to the statistical tests, intervals without jump have $\hat{P}_{up,up} = 0.5$ and $\hat{P}_{down,down} = 0.5$. It is just like a simple random walk model, the price goes upward or downward is random, so a significant price change is less likely to occur and this is the reason that we do not detect any jumps in these intervals. For intervals with up jumps, $\hat{P}_{up,up} > 0.5$ and $\hat{P}_{down,down} = 0.5$. When mid-price goes downward, at the next state the mid-price goes upward or goes downward is random with same probability. However, if the mid-price goes upward, then the mid-price is more likely to goes upward at the next state. Therefore, the mid-price has an overall upward trend, and this can be used to explain why we detect up jumps in these intervals. Similarly, intervals with down jumps have $\hat{P}_{up,up} = 0.5$ and $\hat{P}_{down,down} > 0.5$. The statistical tests results of intervals with down jumps also support our detection.

Overall, this thesis may still have some insufficient place that can be improved, but all of these evidences above could be used to support that this jump detection method is effective and there exists obvious differences between intervals with potential jumps and intervals without any jump.

References

[1] Rama Cont, Peter Tankov. *Financial Modelling With Jump Processes CRC Press LLC*. London, UK, 2004.

[2] Nikolaus Hautsch. *Econometrics of Financial High-Frequency Data Springer*. New York, US, 2012.

[3] Álvaro Cartea, Sebastian Jaimungal, José Penalva. *Algorithmic and HighFrequency Trading Cambridge University Press*. Cambridge, UK, 2015.

[4] Cecilia Mancini. *Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps*. Scandinavian Journal of Statistics, 36: 270-296, 2009.

[5] J.E. Figueroa-López, Jeffrey Nisen. *Optimally thresholded realized power variations for Lévy jump diffusion models*. Stochastic Processes and their Applications (2013), http://dx.doi.org/10.1016/j.spa/2013.04.006.

[6] Rama Cont. *Statistical modeling of high frequency financial data: facts, models and challenges*. (March 1, 2011), Available at SSRN: https://ssrn.com/abstract=1748022.

[7] J.E. Figueroa-López, Jeffrey Nisen. *Second-order properties of thresholded realized power variations of FJA additive processes*. 25 pages. Submitted, 2016.

[8] Xin Guo, Adrien de Larrard, Zhao Ruan. *Optimal placement in a limit order book: an analytical approach*. Math Finan Econ, 26 July 2016.

[9] Jean Jacod, Viktor Todorov. *Efficient estimation of integrated volatility in presence of infinite variation jumps*. The Annals of Statistics, 42: 1029-1069, 2014.

[10] Investopedia: Introduction to order types: Limit Orders, `http://www.investopedia.com/university/intro-to-order-types/limit-orders.asp`

[11] Investopedia: What's the difference between a market order and a limit order, `http://www.investopedia.com/ask/answers/100314/whats-difference-between-market-order-and-limit-order.asp`

[12] PennState: Test of proportion, `https://onlinecourses.science.psu.edu/statprogram/node/164`