

**Assessment of the Limit Order  
Books Modeling Based on  
Poisson Point Process Under  
Both Buy Side and Sell Side**

Zongrui Liang

CID: 02087003

Supervised by Passeggeri, Riccardo

1 September 2023

Submitted in partial fulfilment of the requirements for the MSc in Statistics of  
Imperial College London

The work contained in this thesis is my own work unless otherwise stated.

Signed: Zongrui Liang

Date: 1 September 2023

## Abstract

This paper examines the modeling of the Limit Order Book (LOB) to uncover the inherent relationships between both the buy side (comprising submission of buying limit orders, cancellation of selling limit orders, and execution of selling limit orders or market buying) and the sell side (entailing submission of selling limit orders, cancellation of buying limit orders, and execution of buying limit orders or market selling). Our objective is to predict market trends by visualizing buying power.

We explored models including the Homogeneous/Inhomogeneous Poisson Point Process, the Cox Process, and the Multivariate Hawkes Process. Evaluation metrics employed encompassed MA, EMA, ACF, and PACF for clustering analysis; Pearson's and other correlation coefficients for dependency assessment; KDE and KS for the non-parametric estimation of the Inhomogeneous Poisson Point Process; Q-Q plots of residuals for each relevant model; the Akaike Information Criterion (AIC) for model selection; OLS regression and RMSE for rolling window prediction; and the Euclidean Distance paired with MSE to evaluate the market trend model's fit.

Our findings indicate that, among the models considered, only the Multivariate Hawkes Process aligns well with the LOB when compared against other models and traditional time series models. Since the properties of clustering and self-exciting of the time series of LOB, the multivariate Hawkes Process presents a better fit of Q-Q plot, better AIC value and better intensity plot.

Nevertheless, the Multivariate Hawkes Process demonstrates a poor fit in rolling window prediction across various time horizons, manifesting in predictions with similar shapes but varying amplitudes. Market trend predictions also reveal an inverse relationship with buying power where the discrepancy may arise due to the Multivariate Hawkes Process's inability to adequately capture large-scale event arrivals.

## Acknowledgements

My deepest gratitude goes out to my supervisor, Riccardo, whose guidance, expertise, and unwavering support have been invaluable throughout my research journey.

Also, my flatmate Yezhang Li, who is studying at Imperial, assisted me with brainstorming research concepts while drinking alcohol and I deeply cherish the moments we shared dancing at Heaven Club, which served as a wonderful reprieve from my stress. This is his [Linkedin](#)

On a personal note, I often find myself missing my cherished dog, RainDrop. He has always been a beacon of love and comfort, offering unwavering support during my melancholic moments.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Background . . . . .	3
1.2	Limit Order Book and Modeling . . . . .	3
1.3	Market Tendency . . . . .	4
1.4	Outline of the thesis . . . . .	5
1.5	Random Measure and Point Process . . . . .	6
1.5.1	Random Measure . . . . .	6
1.5.2	Counting Process . . . . .	6
1.5.3	Poisson Point Process (PPP) . . . . .	6
1.5.4	Homogeneous Poisson Point Process (HPPP) . . . . .	7
1.5.5	Inhomogeneous Poisson Point Process (INHPPP) . . . . .	7
1.6	Data Description . . . . .	8
<b>2</b>	<b>Research Methodology and Modeling Approaches</b>	<b>10</b>
2.1	Data Preparation and Classification . . . . .	10
2.1.1	Data Differentiation . . . . .	10
2.1.2	Event arrivals and Time Interval . . . . .	11
2.1.3	Sizing Effect . . . . .	12
2.2	Basic Point Processes under random measure . . . . .	12
2.2.1	Homogeneous Poisson Point Process . . . . .	12
2.2.2	Inhomogeneous Poisson Point Process . . . . .	13
2.3	Cox Process: Theory and modeling under assumption . . . . .	18
2.3.1	Cox Process Theory . . . . .	18
2.3.2	Log-likelihood of intensity function . . . . .	19
2.4	Hawkes Process: Definition and Implementation . . . . .	19
2.4.1	Univariate and Multivariate Hawkes Process . . . . .	20
2.4.2	MLE Estimation for multivariate Hawkes Process . . . . .	21
2.4.3	Decay Parameter ( $\beta$ ) derivation (MoM and CCF applying) . . . . .	25
2.4.4	Gradient-Based Optimization . . . . .	26
2.4.5	Model Evaluation . . . . .	27
2.5	Rolling Window Prediction . . . . .	27
2.5.1	Regression of Market Impact Model Establishment . . . . .	28
2.5.2	Handle Size Parameter . . . . .	29
2.5.3	Multivariate Hawkes Process Prediction . . . . .	30
2.5.4	RMSE and Correlation Evaluation . . . . .	30

<b>3 Model Performance and Evaluation</b>	<b>32</b>
3.1 Preliminary Data Insights . . . . .	32
3.1.1 Events Arrivals for Each State . . . . .	32
3.1.2 Size Parameter Processing . . . . .	32
3.2 Homogeneous Poisson Point Process Evaluation . . . . .	35
3.2.1 95% Confidence Interval testing . . . . .	35
3.2.2 Basic Time Series . . . . .	37
3.3 Inhomogeneous Poisson Point Process Analysis . . . . .	38
3.3.1 MA and EMA Implementation . . . . .	38
3.3.2 The Volatility Clustering . . . . .	39
3.3.3 States Dependency . . . . .	39
3.3.4 Non-Parametric Estimator of The Intensity Function . . . . .	42
3.3.5 Q-Q Plot of Residuals of INHPPP . . . . .	43
3.4 Cox Process Evaluation . . . . .	45
3.4.1 Intensity Chosen . . . . .	45
3.4.2 Q-Q Plot for Cox Process . . . . .	46
3.5 Multivariate Hawkes Process Performance . . . . .	47
3.5.1 MLE Parameters and Algorithm . . . . .	47
3.5.2 Self-Exciting Intensity . . . . .	50
3.5.3 Model Assessment . . . . .	52
<b>4 Model Application to Rolling Window</b>	<b>54</b>
4.1 Weights Manipulation . . . . .	54
4.2 Rolling Window Fitting under Various Time Horizon . . . . .	55
4.3 Market Trend Display and Evaluation . . . . .	57
<b>5 Conclusion and Further Thought on Current Model</b>	<b>62</b>

# 1 Introduction

The introduction part delves into the significance of understanding the limit order book in high-frequency trading. It introduces modeling techniques for the order book and factors influencing market trends. The section also outlines the structure of the thesis, discussing various modeling methods and their efficacy as well as the source of the dataset.

## 1.1 Background

With the booming progress and growth of the advanced electronic field and sophisticated algorithms, individuals are now better equipped to navigate and participate in the realms of stock and cryptocurrency trading. Nowadays, people are getting used to making transactions online through mobile phones in some exchanges like the New York Stock Exchange and Nasdaq. The studies of exchanges are evoked and fascinated by many participants around the world.

After walking through the dark era of the Over-the-counter (OTC) market which transactions are made so slowly and the information is spread in a very small area, people started paying attention to quantifying trading which the exchanges can process hundreds and thousands of orders at the same time. High-frequency trading has emerged and with an understanding of the order book of the trading, people can gain profit by observing the order queue and order depth in High-frequency trading within a millisecond. With a growing number of population joining the market to compete, figuring out the limit order book can be crucial for gaining power against opponents.

## 1.2 Limit Order Book and Modeling

The limit orders, by definition, are the instructions from traders to buy or sell a security at a specific price or the best possible price, but not executed as soon as possible. It is the core component of liquidity and depth where the submitters of the limit order book no matter on the buy or sell side are called the price makers. The limit order book displays these orders with the best (highest) buy prices known as the bid prices and the best (lowest) sell prices known as the ask or offer prices. The difference between the highest bid price and the lowest ask price is called the bid-ask spread. ([Harris \(2003\)](#))

In contrast, market orders are the orders to buy or sell a stock at the best available price given by the price makers, which are executed immediately after submission. It is the main cause of price changes and also leads to the reduction of the depth of the order book. The submitters of market orders are often called price takers, and they will be charged more transaction fees when the orders are executed compared to price makers. ([O'hara \(1998\)](#))

The cancellation of the limit order book is the action made by the traders who change their minds about the desirability of the stocks or bonds, which might be the price shifts or the news prompting. In High-frequency cases, traders might place and cancel orders at a rapid pace to respond to minute-by-minute changes in the market.

To model the limit order book (LOB), the point process might be the best measure which counts on the event arrivals as a point arrivals to proceed modeling. ([Andoh and Kangro](#)) used the stochastic approach by Laplace transform and conditional probability over continuous time intervals to model the dynamic of the LOB. ([Morariu-Patrichi and Pakkanen \(2022\)](#)) used the Hawkes Process to model the change between each state (Limit buying, Limit selling, and so on) by applying the state-dependent Hawkes processes and cross-exciting counting process. ([Vinkovskaya \(2014\)](#)) used the multivariate Hawkes Process to proceed with the modeling of the switching regime.

They did a marvelous job exploring the limit order book modeling and inspired me a lot. In this paper, the modeling will try the different aspects which is the market trend or say the modeling of buying power to estimate the market trend in a short time interval. This will be beneficial for the traders to quickly determine the side they choose to trade and arbitrage based on the model.

### 1.3 Market Tendency

Basically, market trend, also known as market tendency is one of the key components for analysts to excavate the internal relationship between the bull market and the bear markets. Trying to figure out the change in the points of the bull market and the bear market can be profitable and worth researching. Here the LOB can be one of the important indicators for market trends by the further model construction to show the buying power of the current market.

To measure the market trend, first need to know how the trend would be affected, so the factors need to be considered. The market trend would be affected by many aspects of the limit order book: ([Bowe et al. \(2009\)](#))

- Limit order imbalance, where a significant excess of buy orders over sell orders might suggest an upward trend, while the reverse could indicate a potential downward trend.
- The locked or crossed inside spreads, which are the Hidden and Iceberg Orders.

Some traders use hidden orders (orders not displayed in the public order book) or iceberg orders (only a portion of the order is shown) to conceal their full trading intentions.

- Spread Dynamics. The bid-ask spread where a narrowing spread often suggests increased liquidity and possibly a strong trend, while a widening spread might indicate decreased liquidity and uncertainty.
- Order size effects. Large orders can significantly influence price direction. For example, a sudden large buy order can drive prices upward, especially if sell-side liquidity is not deep.
- The cancellation or revision of existing limit order. This significantly reduces the depth of buy or side limit order books and since it has no cost, the action indicates the potential dynamic of market trend.

So looking at these aspects would essentially affect the market trend as well as the modelling predictions. It will be discussed further within the model.

## 1.4 Outline of the thesis

This paper contains 4 parts which are firstly, the methodology part, where presents the measures and approaches made for modeling. It starts with the differernt type of data processing, and then it discusses the feasibility of the LOB modeling of several point or stochastic processes like HPPP, INHPPP, Cox process, and the Multivariate Hawkes Process using the clustering, dependency, MLE value, Q-Q plot, and AIC value. Then the introduction of rolling window prediction by the Multivariate Hawkes Process is estimated by the RMSE and correlation of the plot.

Then the second part presents the model performance using the method mentioned above. The clustering and dependency show the incapability of modeling LOB by HPPP and INHPPP, and by Q-Q plot and AIC value to get the conclusion of the best model fit of the Multivariate Hawkes Process. However, it also shows some incapability of capturing large values by the Multivariate Hawkes Process-based model.

The last part shows based on the Multivariate Hawkes Process, the constracted model performance. The events of arrivals show the same shape but different amplitude and OLS based market trend prediction model shows inverse relationships. Then proceed with some further thoughts like to change the rolling-window prediction to the normal arrivals prediction rather than using regression model.

## 1.5 Random Measure and Point Process

### 1.5.1 Random Measure

Generally, a random measure is a stochastic process, which assigns measures to subsets of a given set in a random manner. It has definition of:

Given a measurable space  $(\Omega, \mathcal{F}, P)$  where  $P$  is a probability measure, a random measure  $M$  on another measurable space  $E$  is a function  $M : \Omega \times E \rightarrow [0, \infty)$  such that:

1. For every  $\omega \in \Omega$ ,  $M(\omega, \cdot)$  is a measure on  $E$ .
2. For every measurable set  $A \subset E$ , the function  $M(\cdot, A) : \Omega \rightarrow [0, \infty)$  is measurable.

This means that for each outcome  $\omega$  in the underlying probability space, we get a measure  $M(\omega, \cdot)$  on  $E$ . And, for each set  $A$  in  $E$ , the function  $M(\cdot, A)$  is a random variable. ([Kallenberg et al. \(2017\)](#))

### 1.5.2 Counting Process

A counting process is a type of stochastic process that represents the number of events that have occurred up to time  $t$ . Formally, it's a collection of random variables  $\{N(t), t \geq 0\}$  such that:

1.  $N(0) = 0$ .
2.  $N(t) \geq 0$ .
3. For all  $t \geq 0$ ,  $N(t)$  is a non-decreasing, integer-valued random variable.
4. For  $0 \leq s < t$ ,  $N(t) - N(s)$  represents the number of events that have occurred in the interval  $(s, t]$ . ([Chen \(2013\)](#))

### 1.5.3 Poisson Point Process (PPP)

A Poisson point process (often abbreviated as PPP) is a type of random measure that describes the random location of points (events) in a space. The points will always follow a Poisson distribution. The counting process is said to be the Poisson Point Process with rate  $\lambda$  where,

1.  $N(0) = 0$ . This means that no events have occurred at the start.
2. The number of events in non-overlapping time intervals is independent. That is, for any  $0 \leq s < t$ , the increment  $N(t) - N(s)$  (number of events in the interval  $(s, t]$ ) is independent of  $N(s)$  and the past history of the process.
3. The number of events in any interval of length  $t$  has the same distribution. This means that the process is "memoryless" in a sense; the expected number of events in a given time frame doesn't depend on where we start observing.

4. The number of events in an interval of length  $t$  follows a Poisson distribution with parameter  $\lambda t$ . That is, for any  $t > 0$ ,

$$P(N(t) - N(s) = n) = \frac{e^{-\lambda(t-s)}(\lambda(t-s))^n}{n!},$$

for  $n = 0, 1, 2, \dots$  ([Chen \(2013\)](#))

#### 1.5.4 Homogeneous Poisson Point Process (HPPP)

The HPPP is a type of stochastic point process where events occur randomly in time or space with a constant average rate and are independent of each other. Here are its main characteristics: ([Daley et al. \(2003\)](#))

1. The rate at which events occur is constant over time or space.
2. The occurrence of one event does not influence the occurrence of another. This means that events happen independently of each other.
3. Mathematically, the number of events  $N(t)$  that occur up to time  $t$  in an HPPP follows a Poisson distribution:

$$P(N(t) = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}. \quad (1.1)$$

#### 1.5.5 Inhomogeneous Poisson Point Process (INHPPP)

INHPPP has the rate (or intensity) at which events occur and can change over time. It contrasts with the homogeneous Poisson process where the rate is constant. ([Daley et al. \(2003\)](#))

An inhomogeneous Poisson point process on a time interval  $[0, T]$  is characterized by an intensity function  $\lambda(t)$  such that:

1.  $\lambda(t) \geq 0$  for all  $t$  in  $[0, T]$ .
2. The number of events in non-overlapping intervals is independent.
3. The probability of observing exactly one event in a small interval  $[t, t + \delta t]$  is  $\lambda(t)\delta t$  for small  $\delta t$ .
4. The probability of observing more than one event in  $[t, t + \delta t]$  is negligible for small  $\delta t$ .
5. The intensity function  $\lambda(t)$  determines the average rate of events at time  $t$ . It can be any non-negative function, and it can vary over the interval.
6. The probability  $P(N(t) = k)$  of observing exactly  $k$  events in the interval  $[0, t]$  is given by:

$$P(N(t) = k) = \frac{e^{-\Lambda(t)}(\Lambda(t))^k}{k!}, \quad (1.2)$$

where  $N(t)$  is the number of events up to time  $t$ , and  $\Lambda(t)$  is the cumulative intensity function defined as:

$$\Lambda(t) = \int_0^t \lambda(s)ds < \infty. \quad (1.3)$$

## 1.6 Data Description

The data for modelling the Limit Order Book of both the buy and sell side is from the LOBSTER (<https://lobsterdata.com>). Having access to the sampling data from the 'samples.' page, the presented dataset is sourced from level 10, the Amazon (AMZN) limit order book message on June 21st, 2012. The trading time is from 34200 seconds to 57600 seconds in counted from the start of the day of June 21st, 2012, which can be transformed from 9:30 to 16:00.

Here the higher level will reveal more transaction data hiding in same time interval, which make the dataset more accurate. Like for level 1 AMZN LOB, it only shows 57515 data while for level 10 AMZN LOB in my dataset, it collects 269748 data, where more data in the dataset would bring better model development.

As for the raw dataset, the message has indicated several crucial data elements: Time, Type, Order ID, Size, Price, and Direction.

The Time has many decimals which is precisely distributed where several events could happen within a very short or even the same time interval so that the time series can be established based on Time data.

The type has 6 different numbers:  $\{1, 2, 3, 4, 5, 7\}$ <sup>1</sup>. 1 means new limit order arrived. 2 and 3 mean cancellation of partial and total limit order, while 4 and 5 mean execution of existing order or iceberg orders.

Order ID is the unique reference number for tracking. Size is the number of shares and Price is the stock price in dollars.

Direction has two signs which are  $\{-1, 1\}$  where -1 means limit order sell, and 1 means limit order buy. Combining the Type and Direction would reveal the buying and selling of the limit order book and is essential for modeling. Especially after the data cleaning and processing. These will be further discussed in [2.1.1](#).

Also for the Size part, it is not important for the prior modeling but essential for the prediction part, and it will be illustrated in [2.5.2](#). In this thesis, it will only consider

---

<sup>1</sup>7 is rarely happened and not appear in this dataset, so is not considered

Time, Type, Size, and Direction to form the model for different stochastic processes. Other parameters within the raw dataset will be discussed further.

## 2 Research Methodology and Modeling Approaches

The chapter illustrates the methods used for testing if the homogeneous and inhomogeneous Poisson point process, Cox process, and the different types of Hawkes processes can fit the LOB data well. Also, it mentions the construction of the rolling window prediction and how to evaluate the prediction based on the selected model.

### 2.1 Data Preparation and Classification

#### 2.1.1 Data Differentiation

First of all, to classify the buy side and sell side of the limit order book, it was crucial to dive into the market behavior of the limit order book. In refer to Section 1.6, 'Type' and 'Direction' were helpful for the differentiation of both sides. The table of each state relationship was drawn below:

Table 2.1: Buy side and Sell side based on the Type and Direction Status

	<b>Direction {1}</b>	<b>Direction {-1}</b>
<b>Type {1}</b>	Submission of limit order buying <b>(LB), {1, 1}</b>	Submission of limit order selling <b>(LS), {1, -1}</b>
<b>Type {2}</b>	Cancellation of limit order buying <b>(CB), {2, 1}</b>	Cancellation of limit order selling <b>(CS), {2, -1}</b>
<b>Type {3}</b>	Cancellation of limit order buying <b>(CB), {3, 1}</b>	Cancellation of limit order selling <b>(CS), {3, -1}</b>
<b>Type {4}</b>	execution of buy limit order (Market selling) <b>(MS), {4, 1}</b>	execution of sell limit order (Market buying) <b>(MB), {4, -1}</b>
<b>Type {5}</b>	execution of buy limit order (Market selling) <b>(MS), {5, 1}</b>	execution of sell limit order (Market buying) <b>(MB), {5, -1}</b>

In the table 2.1, based on the combination of 5 different types and 2 different directions of the limit order book message, the entire dataset could be arranged into 6 states where each side occupied 3 states. The buy side in green color consisted of the willing to push the price up or long the stock, so the submission of limit order buying (LB){1, 1}, and execution of limit order selling (MB){{4, -1}, {5, -1}} which is the market buying were considered. Also, the action of cancellation of limit order selling (CS){{2, -1}, {3,

$-1\}$ } indicated the thought of price-up market behavior. The same ideas for sell side in red color,  $(LS)\{1, -1\}$ ,  $(CB)\{\{2, 1\}, \{3, 1\}\}$ ,  $(MS)\{\{4, 1\}, \{5, 1\}\}$  were considered to be the sell side market behavior. The state is listed in this table:

Table 2.2: State 1 to State 6 for Buy Side and Sell Side

<b>Buy side</b>	
State 1	$(LB)\{1, 1\}$
State 2	$(CS)\{\{2, -1\}, \{3, -1\}\}$
State 3	$(MB)\{\{4, -1\}, \{5, -1\}\}$
<b>Sell side</b>	
State 4	$(LS)\{1, -1\}$
State 5	$(CB)\{\{2, 1\}, \{3, 1\}\}$
State 6	$(MS)\{\{4, 1\}, \{5, 1\}\}$

So for each state from 1 to 6, the data would be generated within the state and had an effect on each other for further model research.

### 2.1.2 Event arrivals and Time Interval

All the events happened within the trading time so no need for further data trimming outside the trading time. After the removal of NaN value, we wanted to mention that the time stamps were at least milliseconds and up to nanoseconds, but for the subsequent table and graph, time stamps would be treated to 1 second in accuracy. However, the establishment of time series would include the event arrivals so each time stamp that appeared would express the arrival of certain events. So checking the rate of arrival and drawing the event occurrence for some time interval to visualize the data would be helpful for understanding the sample data.

Since the dataset was huge and up to 269748 data for each time, segmentation of the event arrivals into some discrete time intervals would be essential for subsequent analysis and easy to implement the number of event arrivals and the results loop within the acceptable processing time. The indicated time intervals for this thesis would be 10 seconds, 60 seconds, and 120 seconds, which was 1000 milliseconds, 1 minute, and up to 2 minutes. So the intervals would be fractured into 2340, 390, and 195 time intervals.

In this thesis, the modeling would solely be concentrated on 60-second interval and the other 10-second interval and 120-second interval would be used in the prediction part for different time horizon comparisons and model testing. The reason why chose 60 second to be the main analytical hyperparameter was that 1 minute was easy to implement as well as data segmentation and the 390 segments were the middle ground that captured short-term dynamics without being overly noisy. Also, certain granularities might be more readily available than others, and 1 minute would avoid some overfitting due to the high precision of event arrivals within a certain short time range even within a second.

Then analyzing each interval with respect to each time horizon was helpful for the results validation and robust for the model establishing.

### 2.1.3 Sizing Effect

As for the last parameter, the size parameter processing, we mainly focused on descriptive statistics to find the mean, variance, maximum value, and minimum value of the whole dataset. After that, it would deploy the outlier detections by using the box plot and replace the outliers above the 99th percentile with the value at the 99th percentile as a threshold called capping method.([Tiwari et al. \(2007\)](#))

It was worth saying that the size parameter had not been included in the main model construction but played a role in the prediction part since the established market tendency regression model [2.5](#) had included the size to be the main factor to have an effect on market tendency. So size had to be analyzed comprehensively and find its coherent elements or inherent properties.

After the size processing, it was also important to measure the mean size for both the buy and sell sides under 60 second intervals. Drew the time series graph to compare the similarity and trends. Subsequently, estimated the correlation coefficient of the time series of size between both buy and sell sides to find the relationship between them. Then, it would be well analyzed before the prediction of market tendency [2.5.2](#).

## 2.2 Basic Point Processes under random measure

Under the random measure theory, the point process counts the number of events that appear in a given set, while the random measure is also a point process but assigned the same weight. Therefore, both the homogeneous and inhomogeneous poisson point process can be thought of as a counting process for counting the number of events arrivals.

### 2.2.1 Homogeneous Poisson Point Process

In considering the homogeneous Poisson point process, the formula is derived from [1.4](#). It was good to use the HPPP since The HPPP is one of the simplest point processes. It's defined by a single parameter ( $\lambda$ ), making it easy to understand, implement, and interpret. Also, many complex models, like the Hawkes process and Cox process, were built upon the Poisson process. It is the fundamental model and might be a good start of the paper to establish the base of the model and estimate further.

To proceed with this model, we have to make some assumptions:

1. Events occurred independently and at a constant average rate ( $\lambda$ ) over time or space.
2. Two or more events cannot occur at the exact same point in time or space.
3. At the first time  $t = 0$ , there would be no events happening at this time.

Therefore, resolving the constant rate parameter of HPPP was crucial. However, under 60 second intervals, we should draw the number of event arrivals for each interval and for 6 states individually. It was good for visualization and determining if some properties were violated by the rule. Also, a 95% confidence interval was robust to assess if the model has higher heterogeneity. Following by The time series plot of each state would directly reveal the fluctuation of event arrivals of a certain time range or if it included some spike, tails, or clustering.

After the plot as well as the confidence interval plot, we should proceed to the data analysis by using Descriptive statistics for each state under 60 second intervals. It can also show the data spread of each state and the deviation.

### 2.2.2 Inhomogeneous Poisson Point Process

With respect to the homogeneous Poisson Point process which had a constant rate, the Inhomogeneous Poisson Point process (INHPPP) had a time-varying intensity function  $\lambda(t)$ , which gave the rate of events expected to happen at a time( $t$ ).

The financial data, like the limit order book, was more likely to have a non-constant rate, especially under the condition of high-frequency data trading. For instance, the trading activity might be higher at the start and end of the trading day time and may be weak during the day.  $\int_a^b \lambda_i(t)dt$  represented the number of events that occurred between the time interval  $[a,b]$  and  $i$  mean different states  $i = 1$  to  $6$ .

So to start the INHPPP, we had to determine the time-varying intensity. However this cannot be directly estimated, we should start from its assumption first and then proceed to its model implication before modeling. The assumptions were listed as:

1. Intensity Function  $\lambda(t)$  must be non-negative for all  $t$ , and it is often required to be integrable over the time intervals.
2. The number of events in non-overlapping intervals is independent. For instance, the number of events happening in the time interval  $[a, b]$  is independent of the number of events in  $[c, d]$  as long as  $[a, b]$  and  $[c, d]$  don't overlap.
3. The probability of two or more events happening at the exact same time is zero.
4. The integral of the intensity function over the entire domain should be finite, i.e.,  $\int \lambda(t) dt < \infty$  over the time intervals.
5. The occurrence of an event does not influence the probability of future events.

### Moving Average (MA):

$$MA_n^k = \frac{1}{k} \sum_{i=1}^k x_{n-i}. \quad (2.1)$$

or

$$MA_n^k = \frac{1}{k} \sum_{i=1}^k x_{n-i} w_i. \quad (2.2)$$

Where  $x_1, x_2, \dots, x_n$  here represented the events arrived per interval, and  $k$  was the window width, which was also considered to be the rolling window in Python. Also, increasing the window size would produce smoother results but might hide some minor fluctuations. Decreasing it would make the results closer to the raw data but might introduce more noise. Here I chose the window width = 10 which is  $60 \times 10 = 600$  seconds = 10 minutes under 60-second time intervals. In total, 390 intervals would be considered.

A moving average is used to analyze data points by creating a series of averages of different subsets of the full data set. Then the time series would become smoother in the plot.

### Exponential Moving Average (EMA)

The Exponential Moving Average (EMA) is similar to the simple moving average, but it gives more weight to the recent prices and less weight to the previous prices. This makes it more responsive to recent price changes.

It has the following Formula:

$$EMA_t = \alpha \cdot x_t + (1 - \alpha) \cdot EMA_{t-1}, \quad (2.3)$$

where t is the total number of time intervals, here is 390 intervals.  $\alpha$  is the weighted value where  $\alpha \leq 1$  which followed the basic geometric series of the original deducted formula.  $EMA_t$  is the fitness history value, which contributes to the historical importance. In this case, the most recent value is  $EMA_t * (1 - \alpha) * \alpha^0$ . [Haynes et al. \(2012\)](#)

Here the EWM function is used in Python to compute the EMA.

By using both MA and EMA, it smoothed the time series of empirical data and can be analyzed directly through the MA and EMA line to see extrinsic properties for each state.

After the plotting, volatility clustering was needed to prove if it had some clustering properties which suggested that, a large amount of event arrivals was commonly followed by another large amount of event arrivals. [\(Cont \(2007\)\)](#)

To prove if it had some volatility clustering, the 2 tools in time series called AutoCorrelation Function (ACF) and Partial Autocorrelation Function (PACF) were introduced: [\(Robert et al. \(2006\)\)](#)

### **Autocorrelation Function (ACF)**

it described the correlation between time series and its lags and often helped to identify the repeated patterns in a time series.

$$\rho_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}, \quad (2.4)$$

where  $\rho_k$  is the autocorrelation coefficient at lag  $k$ ,  $y_t$  is the number of arrivals of the time series at time  $t$   $\bar{y}$  is the mean of the event arrivals and  $T$  is the length of the time series, under 60 seconds time intervals, it should be 390.

The partial Autocorrelation Function (PACF) described the contrast with the autocorrelation function, which was the partial correlation of a time series with its own lagged values.

After the evaluation of its time series pattern, it was important to test the assumption of independence, here I would use different approaches to estimate the correlation coefficient to test the dependency of each state from state 1 to state 6, which are:

#### **1. Pearson's Correlation Coefficient**

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (2.5)$$

where  $x_i$  and  $y_i$  were the amount of event arrivals at some time points.  $\bar{x}$  and  $\bar{y}$  were mean value of x and y. n is the number of time points. ([Pearson \(1895\)](#))

## 2. Spearman's Rank Correlation

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \quad (2.6)$$

where  $d_i$  is the rank difference between 2 different states. n is the number of time points. ([Spearman \(1987\)](#))

## 3. Kendall's Tau

$$\tau = \frac{(n_c - n_d)}{n(n - 1)/2}, \quad (2.7)$$

where  $n_c$  is the number of concordant pairs  $n_d$  is the number of discordant pairs n is the number of time points. ([Kendall \(1938\)](#))

## 4. Mutual Information

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right), \quad (2.8)$$

where  $p(x, y)$  is the joint probability distribution function of  $X$  and  $Y$   $p(x)$  and  $p(y)$  are the marginal probability distribution functions of  $X$  and  $Y$  respectively. Sometimes the mutual information would be 0 because of the lack of marginal probability of certain states at some data points. ([Shannon \(1948\)](#))

For each approach, the coefficient value varied from -1 to 1, closer to -1 referred to a strong negative relationship, closer to 1 referred to a strong positive relationship, and closer to 0 implied no such correlation. While Pearson described the linear relationship, Spearman described the rank correlation using a monotonic function, Kendall's Tau found the correlation between two measured quantities and Mutual Information measured the information of two time series.

## Intensity Estimation

Then to proceed with the modelling of LOB, the intensity of INHPPP was estimated by a non-parametric estimator with the empirical data for each state, under the 60 second time interval. There were several approaches to estimating the intensity:

### 1. Kernel Density Estimation (KDE)

$$\Lambda(t) = \frac{1}{nh} \sum_{i=1}^n K \left( \frac{t - t_i}{h} \right), \quad (2.9)$$

where  $\Lambda(t)$  is the estimated density function.  $n$  is the number of time points.  $h$  is the bandwidth, and here was the time interval, which was 60.  $K$  is the kernel function and  $t_i$  are the event of arrival for some time points.

The kernels were defined by the following equations: ([Scott \(2015\)](#))

#### Gaussian (Normal) Kernel:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \quad (2.10)$$

#### Tophat (Uniform) Kernel:

$$K(u) = \begin{cases} 0.5 & \text{if } |u| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.11)$$

#### Epanechnikov Kernel:

$$K(u) = \begin{cases} \frac{3}{4}(1-u^2) & \text{if } |u| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.12)$$

#### Exponential Kernel:

$$K(u) = \frac{1}{2} e^{-|u|} \quad (2.13)$$

#### Cosine Kernel:

$$K(u) = \begin{cases} \frac{\pi}{4} \cos\left(\frac{\pi u}{2}\right) & \text{if } |u| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.14)$$

**2. Kernel Smoothing (KS)** The formula and principle were similar to KDE, for each observed event time, a kernel was centered at that event time. The intensity at any given point in time was the sum of all kernels' values at that point. ([Wand and Jones \(1994\)](#))

The difference between KDE was that the result didn't need to integrate to 1 over its domain since it represents intensity rather than probability. Whilst, KDE was mainly centered on the distribution characteristics and KS was interested in the trend of certain time series. However, combining both of the approaches could indicate a well-rounded explanation and understanding of the LOB model by applying distinct specific intensity functions. KDE could show the distribution of order flow and KS could show the average order flow over time.

**3. Exponential Moving Average function (EMA)** For the last approach, it was directly calculated the EMA by setting a specific decay factor and iterated for each time intervals using the formula 2.3.

#### Performance Assessment:

Q-Q plots of residuals to assess the performance of non-parametric intensity estimator using the 3 approaches listed above. The process were listed

##### 1. Compute Residuals:

- For each time interval  $[t_n, t_{n+1}]$ , estimated the intensity of the point process using different non-parametric estimations. denoted this as  $\lambda(t_n; t_{n+1})$ .
- Count the number of events arrivals in the interval,  $N(t_n; t_{n+1})$ .
- Define the residual for the interval,  $R(t_n; t_{n+1})$ , then normalized by the square root of the expected number.

##### 2. Simulate Residuals:

- Use Monte Carlo simulation to generate several simulated time series based on the non-parametric intensity.
- For each simulated series, compute the model-based residuals.

##### 3. Construct Q-Q Plot:

- For each time interval, plot the quantile of the observed residual against the quantile of the simulated residuals. To test if the points were orderly lie on the line  $y = x$ .

It indicated Q-Q plots of residuals are a valuable diagnostic tool in assessing the fit of point process models.

## 2.3 Cox Process: Theory and modeling under assumption

### 2.3.1 Cox Process Theory

The Cox process, also known as the Doubly Stochastic Poisson Process, under random measure base theory, was another type of point process. Unlike other point processes like HPPP, INHPPP and Hawkes process, which have certain determined intensity functions, no matter time-varying or historical based. The Cox process had the intensity function itself as a random process, given by  $\Lambda(t)$  for the random process, where the intensity function  $\lambda(t) = \Lambda(t)$ . ([Daley and Vere-Jones \(2008\)](#)) The Cox process found applications in various domains such as high-frequency trading in finance, weather prediction in ecology, and in neuroscience for modeling neuron firing patterns.

The probability of observing  $k$  events in an interval  $[t_1, t_2]$  is given by:

$$P(N(t_1, t_2) = k) = \frac{e^{-\int_{t_1}^{t_2} \Lambda(s)ds} \left( \int_{t_1}^{t_2} \Lambda(s)ds \right)^k}{k!}. \quad (2.15)$$

Here, for a better understanding of the thesis, the assumption of the process should be Poisson and the intensity could be chosen as Hawkes process. However, putting the Univariate Hawkes process would not fit the model 2.4.1. The intensity of the Cox process should be chosen as the Multivariate Hawkes Process for better variable control. Even the same intensity would lead to the same AIC value 2.4.5, it would still reveal different Q-Q plot since the inhomogeneous intensity with the assumption of Poisson process of The Cox process would result in different plot.

### 2.3.2 Log-likelihood of intensity function

So based on the definition of the basic poisson process 1.1 the likelihood function  $L$  for observing  $n$  events at times  $t_1, t_2, \dots, t_n$  is:

$$L = \prod_{i=1}^n \lambda(t_i) \exp \left( - \int_0^T \lambda(t) dt \right), \quad (2.16)$$

within the time interval  $[0, T]$

To derive the log-likelihood, take the logarithm of the likelihood:

$$\log L = \sum_{i=1}^n \log \lambda(t_i) - \int_0^T \lambda(t) dt. \quad (2.17)$$

So after importing the empirical data, the MLE can be derived by using the minimize function in Python using the Nelder-Mead downhill simplex algorithm. ([Nelder and Mead \(1965\)](#)).

After that, it was important to test the model fit and evaluation. Therefore the Q-Q plot is used as well as the plotting of the intensity function for each state.

## 2.4 Hawkes Process: Definition and Implementation

The Hawkes process is a self-exciting point process, which means that the occurrence of an event would likely cause the soaring probability of future events occurring in a short period of time afterward. It can be used to predict earthquakes for aftershocks, the sudden increase in spikes in social media activities, and high-frequency market trading including the limit order book modeling.

### 2.4.1 Univariate and Multivariate Hawkes Process

#### Univariate Hawkes Process

Univariate refers to a Hawkes process that only monitors a single state of event over time. The intensity function (or conditional intensity function) of the process, which describes the rate at which events are expected to occur, depends on the entire history of the process up to the current time([Obral \(2016\)](#)). The exponential Kernel is shown as:

$$K(t) = \sum_{m=1}^{d=1} \alpha_m e^{-\beta_m t}, \quad (2.18)$$

where  $d$  was the total order, and in the univariate case, it was 1.  $\alpha$  was the jump size in intensity after each event.  $\beta$  denoted the decay factor at which the intensity would decay back to baseline. ([Daley and Vere-Jones \(2008\)](#))

Therefore, The intensity function for a univariate Hawkes process can be expressed as:

$$\Lambda(t) = \mu + \sum_{t_i < t} \alpha e^{-\beta(t-t_i)}, \quad (2.19)$$

where  $\mu$  is the baseline intensity and the sum is over all events  $t_i$  that have occurred before time  $t$ .

#### Multivariate Hawkes Process

In the multivariate case, there were multiple states of events existence, and each state had its past events can influence not only its own future occurrences but also the future occurrences of other states of events. The influences were universal and every state would participate in the events arrivals for each state. For instance, a large LB might influence future LS orders, and vice versa. ([Embrechts et al. \(2011\)](#))

Under the same exponential kernel, the intensity function for the  $m^{th}$  type in a multivariate Hawkes process can be expressed as:

$$\Lambda_m(t) = \mu_m + \sum_{n=1}^d \sum_{t_i^n < t} \alpha_{mn} \exp(-\beta_m(t - t_i^n)), \quad (2.20)$$

where

- $d$  is the total number of states, here is 6.
- $\mu_m$  is the baseline intensity for the  $m^{th}$  state.

- $\alpha_{mn}$  represents the influence of an event of state  $n$  on the future rate of events of state  $m$ .
- $\beta_m$  is the decay parameter.
- The outer sum  $\sum_{n=1}^d$  is over all event state, and the inner sum  $\sum_{t_i^n < t}$  is overall events of state  $n$  that have occurred before time  $t$ .

Based on the 2.2, the classification of the AMZN trading data showed the 6 different states and each 3 states were assigned to the buy and sell side. The main purpose of this paper was to find the relationship between the buy side and the sell side to estimate the market trend in the total time horizon.<sup>1</sup> Therefore, due to the nonnegligible interaction between various states like the large arrival of LB might trigger the large arrival of LS, the cross-influence and inter-dependence should be considered and Multivariate Hawkes process was taking more advantages in modeling the limit order book.

As well as the indicated model complexity, it would have 48 parameters including 4 parameters of  $\mu_m$ , 8 parameters of  $\beta_{mn}$ , and 36 parameters of  $\alpha_{mn}$ . To simplify the model without destroying its continuity and completeness, shortening the dimension of  $\beta_{mn}$  was considerable. Therefore, in the subsequent analysis, the  $\beta_m$  which indicated the decaying factors only in relating to itself.

#### 2.4.2 MLE Estimation for multivariate Hawkes Process

After cutting the dimension of decay parameter  $\beta_m$ , the maximum likelihood estimator (MLE) under the multivariate Hawkes Process was needed to compute for further modeling research. For simplicity, we could begin with the univariate Hawkes process, and then extend it to the multivariate case.

##### **Log-likelihood of Univariate Hawkes Process:**

##### **Probability of observing an event in the interval**

in an infinitely small interval  $[t, t + dt]$ . For an inhomogeneous Poisson process, this expectation is  $\Lambda(t)dt$  for intensity function  $\Lambda(t)$

##### **Probability of observing no event in the interval**

In the case of a homogeneous Poisson process,  $P_0$  is the probability of observing no events in an interval of length  $t$ , which has the equation:

$$P_0(t) = e^{-\Lambda t}.$$

For the Inhomogeneous Poisson Process, The probability of no events occurring in the

---

<sup>1</sup>Time interval of 10s, 60s and 120s

infinitely small interval  $[t, t + dt]$  is:

$$1 - \Lambda(t)dt.$$

Then, for no event happens within the whole time interval  $[0, T]$ .  $1 - \lambda(t)dt$  should be timed together for infinite times. Then the product would be:

$$P_0(t) = e^{-\int_0^T \Lambda(s)ds}.$$

The probability of observing an event at time  $t$  and no other events have happened up within the infinitely small interval  $[t, t + dt]$  is:

$$P(t)dt = \Lambda(t)e^{-\int_0^t \Lambda(s)ds}dt,$$

where  $e^{-\int_0^t \Lambda(s)ds}$  is the probability of no other events happening before the  $t$ , and  $\Lambda(t)dt$  is the probability that an event happening right at the time  $dt$ .

Taking the log of both sides:

$$\log(P(t)dt) = \log(\Lambda(t)) - \int_0^t \Lambda(s)ds + \log(dt).$$

Then integrated over the entire observation window, which is the time from 34200 to 57600,  $[0, T] = [34200, 57600]$ , but here would still use the  $[0, T]$  for better understanding of the formula. The log-likelihood of the observed sequence of events is formulated.

$$\log L(\theta) = \sum_i \log(\Lambda(t_i)) - \int_0^T \Lambda(s)ds,$$

where  $\theta = (\alpha, \beta, \mu)$  So in the m-dimension Multivariate case, it has the same equation of log-likelihood, and the integral part is urgent to be solved:

$$\int_0^T \Lambda_m(s)ds,$$

given that the intensity function for dimension  $m$  in a multivariate Hawkes process 2.20 is:

$$\Lambda_m(t) = \mu_m + \sum_{n=1}^d \sum_{t_i^n < t} \alpha_{mn} \exp(-\beta_{mn}(t - t_i^n)),$$

where the baseline intensity:

$$\int_0^T \mu_m ds = \mu_m \int_0^T ds = \mu_m T.$$

The past events influence part:

$$\int_0^T \sum_{t_i^n < t} \alpha_{mn} \exp(-\beta_m(t - t_i^n)) dt.$$

For a fixed event time  $t_i^n$  of dimension  $n$ :

$$\int_0^T \alpha_{mn} \exp(-\beta_m(t - t_i^n)) dt.$$

Break down the integral into two parts, from 0 to  $t_i^n$  and from  $t_i^n$  to  $T$ :

$$\int_{t_i^n}^T \alpha_{mn} \exp(-\beta_m(t - t_i^n)) dt.$$

This is an exponential function integrated over  $t$ . Solving this:

$$= \alpha_{mn} \left[ -\frac{1}{\beta_m} \exp(-\beta_m(t - t_i^n)) \right]_{t_i^n}^T$$

$$= \alpha_{mn} \left( -\frac{1}{\beta_m} \exp(-\beta_m(T - t_i^n)) + \frac{1}{\beta_m} \right)$$

$$= \frac{\alpha_{mn}}{\beta_m} (1 - \exp(-\beta_m(T - t_i^n))).$$

This is for a single event  $t_i^n$  of dimension  $n$ . To account for all past events of dimension  $n$  that occurred before  $T$ , we sum over all such  $t_i^n$ :

$$\sum_{t_i^n < T} \frac{\alpha_{mn}}{\beta_m} (1 - \exp(-\beta_m(T - t_i^n))).$$

Therefore:

$$\int_0^T \Lambda_m(s) ds = \mu_m T + \sum_{n=1}^d \sum_{t_i^n < T} \frac{\alpha_{mn}}{\beta_m} (1 - \exp(-\beta_m(T - t_i^n))).$$

Combining both parts:

$$\begin{aligned} \log L(\theta) &= \sum_i \log(\Lambda(t_i)) - \int_0^T \Lambda(s) ds \\ &= \sum_i \log(\Lambda(t_i)) \\ &\quad - (\mu_m T + \sum_{n=1}^d \sum_{t_i^n < T} \frac{\alpha_{mn}}{\beta_m} (1 - \exp(-\beta_m(T - t_i^n)))) \\ &= \sum_i \log(\mu_m + \sum_{n=1}^d \sum_{t_i^n < t_i} \alpha_{mn} \exp(-\beta_m(t_i - t_i^n))) \\ &\quad - (\mu_m T + \sum_{n=1}^d \sum_{t_i^n < T} \frac{\alpha_{mn}}{\beta_m} (1 - \exp(-\beta_m(T - t_i^n)))). \end{aligned}$$

For the Python part, the equation should be solved by minimizing the algorithm using The Nelder-Mead downhill simplex algorithm, which didn't need gradient information for optimizing a function. So the log-likelihood should take the negative sign for manipulating:

$$f(\theta) = -\log L(\theta). \tag{2.21}$$

Also indicated the property of asymptotic Gaussian for MLEs:

$$\sqrt{N} (\hat{\theta} - \theta_0) \sim \mathcal{N} \left( 0, \frac{1}{N} \mathcal{I}(\theta_0)^{-1} \right).$$

In this context,  $\hat{\theta}$  represents the Maximum Likelihood Estimation (MLE) parameters, while  $\theta_0$  denotes the true empirical parameters. The symbol  $N$  signifies the number of event arrivals, applicable for LB, LS, among others. The Fisher information matrix is denoted by  $\mathcal{I}(\theta_0)$ . ([Vinkovskaya \(2014\)](#))

#### 2.4.3 Decay Parameter ( $\beta$ ) derivation (MoM and CCF applying)

After the simplifying of  $\beta_m$  dimension, the total number of parameters required to be solved is reduced. However, by the direct calculation, the number of parameters is still a huge amount which is 44, and if considering the MLE estimation for each parameter, the processing would take ages to manipulate the loop or iteration. Due to the properties of  $\beta_m$ , which controls the rate at which the excitation caused by past events decays over time. Then the Method of Moments (MoM) approach could be used on the time-decay of the Cross-Correlation Function (CCF) to deduce the  $\beta$  value in advance. [Vinkovskaya \(2014\)](#)

The method of Moments (MoM) is to estimate the parameter of a distribution. It includes equating the sample moment to the empirical moments of the distribution and solving the resulting parameters ([Casella and Berger \(2021\)](#)). In this case,  $\beta_m$  is the time-decay parameter and is following the exponential kernel of Hawkes Process. Therefore,  $\beta_m$  is likely to have an exponential decay distribution, and applying MoM would be efficient to find the parameter while MLE cannot do.

The Cross-Correlation Function (CCF) is the measure of similarity between two signals as a function of the time lag applied. It shows if 2 time series are similar for some time intervals. ([Hellesteth \(1976\)](#)) The formula is:

For two discrete-time signals  $x(t)$  and  $y(t)$ , the cross-correlation  $R_{xy}(\tau)$  is defined as:

$$R_{xy}(\tau) = \sum_t x(t) \cdot y(t + \tau).$$

For continuous signals  $x(t)$  and  $y(t)$ :

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t) \cdot y(t + \tau) dt,$$

where  $\tau$  is the time-lag.

Ideally, the theoretical CCF would be applied to measure the parameter  $\beta_m$ . However, the mathematical formulation is quite challenging and in this case, only empirical CCF would be used to approach the real theoretical value of  $\beta_m$ .

#### 2.4.4 Gradient-Based Optimization

Since the loop and iteration for the MLE optimization using the method of Nelder-Mead downhill simplex algorithm did not consider the gradient of the log-likelihood. It was likely to apply the Gradient-Based Optimization by differentiating the negative log-likelihood for the first order. Then use another Gradient-Based algorithm like L-BFGS-B to speed up the operation.

So we had to derive the first derivative of negative log-likelihood (NLL). The  $\Theta = (\alpha, \mu)$  only had 2 parameters now, where the  $\beta_m$  were deducted by the MoM and CCF by empirical data approach and would be treated as constant. So in this case, we derived the NLL in respect to 2 parameters, start by  $\alpha_{mn}$ :

Derivative with respect to  $\alpha_{mn}$ :  $\frac{\partial L(\Theta)}{\partial \alpha_{mn}}$

From the expression for  $\Lambda_m(t)$  2.20, the derivative with respect to  $\alpha_{mn}$  is:

$$\frac{\partial \Lambda_m(t)}{\partial \alpha_{mn}} = \sum_{t_i^n < t} \exp(-\beta_m(t - t_i^n)).$$

Differentiating the term inside the second sum in the log-likelihood expression:

$$\frac{\partial}{\partial \alpha_{mn}} \left( \frac{\alpha_{mn}}{\beta_m} (1 - \exp(-\beta_m(T - t_i^n))) \right) = \frac{1}{\beta_m} (1 - \exp(-\beta_m(T - t_i^n))).$$

Now, we can plug in the above results to compute the derivative of the log-likelihood:

$$\begin{aligned} \frac{\partial L(\Theta)}{\partial \alpha_{mn}} &= \sum_i \frac{1}{\Lambda(t_i)} \frac{\partial \Lambda(t_i)}{\partial \alpha_{mn}} - \sum_{t_i^n < T} \frac{1}{\beta_m} (1 - \exp(-\beta_m(T - t_i^n))) \\ &= \sum_i \frac{\sum_{t_i^n < t_i} \exp(-\beta_m(t_i - t_i^n))}{\mu_m + \sum_{n=1}^d \sum_{t_i^n < t_i} \alpha_{mn} \exp(-\beta_m(t_i - t_i^n))} \\ &\quad - \sum_{t_i^n < T} \frac{1}{\beta_m} (1 - \exp(-\beta_m(T - t_i^n))). \end{aligned}$$

Derivative with respect to  $\mu_m$ :  $\frac{\partial L(\Theta)}{\partial \mu_m}$

From the expression for  $\Lambda_m(t)$ , the derivative with respect to  $\mu_m$  is:

$$\frac{\partial \Lambda_m(t)}{\partial \mu_m} = 1.$$

Differentiating the term  $\mu_m T$  in the log-likelihood expression:

$$\frac{\partial}{\partial \mu_m} (\mu_m T) = T.$$

Plug to compute the derivative of the log-likelihood:

$$\frac{\partial L(\Theta)}{\partial \mu_m} = \sum_i \frac{1}{\Lambda(t_i)} - T.$$

So implementing both  $\mu_m$  and  $\alpha_{mn}$  to a gradient-based algorithm to find the MLE of them would be efficient and robust.

#### 2.4.5 Model Evaluation

After the essential plotting of the multivariate Hawkes process in respect to different state of events arrivals, the Akaike Information Criterion (AIC) and also Q-Q plot are implemented to test if the model was nested or unfitted.

Akaike Information Criterion (AIC), is a measure to compare the goodness of fit for different models to the time series. It takes into account the complexity of the model, penalizing models that are overfitting. ([Akaike \(1974\)](#)) The formula is listed:

$$AIC = 2k - 2 \log(L),$$

where  $k$  is the number of estimated parameters in the model and  $L$  is the maximum likelihood function under MLEs for the model.

Here I would use the Multivariate Hawkes process to compare with Cox process, Autoregressive model AR(m), Moving Average MA(m), and the combined Autoregressive Integrated Moving Average ARIMA(m). To test if the Multivariate Hawkes process was fitted well or not, and also to test whether the model was overfitted or nested.

## 2.5 Rolling Window Prediction

In LOB modeling, it was crucial to predict upcoming events with historical data. In this case, rolling window prediction was used for time series forecasting where a fixed-size window of historical data is adapted to train a model. Due to the non-stationary nature

of LOB, this rolling or sliding window approach was most likely to be utilized to establish the model basis. In this paper, the window size was chosen to be 10 which meant 10 time intervals of the number of event arrivals which was considered to be the historical data, and use the 10 prior windows to predict the number of events arriving for the next time interval.

### 2.5.1 Regression of Market Impact Model Establishment

To get a deeper understanding of market impact for short-term trading, the buying power could directly reflect the market trend for high-frequency trading. The prediction of the event arrivals for different states on different sides was illustrated as the prediction of the buying power changing related to time intervals. So the arrival of state1,2,3 for the buy-side was treated to be a positive impact on buying power while the arrival of state4,5,6 for the sell-side was treated to be a negative effect on buying power. The simulated graph is like Figure 2.1 for better understanding:

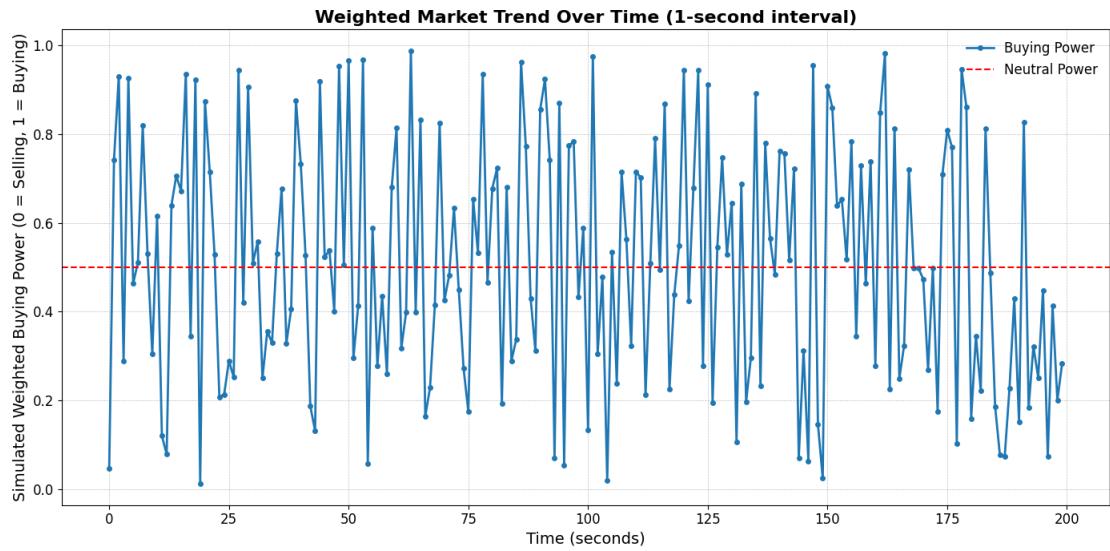


Figure 2.1: 200 simulated points from 0 to 1 in the 1-second time interval. The market trend is not clear and fluctuates a lot. Value closer to 1, the market is buying power dominated. Closer to 0, the market is selling power-dominated. Closer to the red dotted line = 0.5, which means the market stays neutral

Therefore, the buying power was formulated by Equation 2.22

$$\mathcal{B} = \frac{\beta_{BUY} * X_{BUY}}{\beta_{BUY} * X_{BUY} + \beta_{SELL} * X_{SELL}} \quad (2.22)$$

$$= \frac{\beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3}{\beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + \beta_4 * X_4 + \beta_5 * X_5 + \beta_6 * X_6}. \quad (2.23)$$

where  $\mathcal{B}$  is the market buying power.

$\beta_i$  for i from 1 to 6 represented state 1 to state 6 were the estimated coefficient and here indicated the weighted.

$X_i$  for i from 1 to 6 also represented state 1 to state 6, which were the predictor variables.

Also, to turn the coefficients into weights that sum to 1 (for interpretability), normalize them by using:

$$\text{Weight of State}_i = \frac{|\beta_i|}{\sum_{j=1}^N |\beta_j|}.$$

We used the absolute values to ensure all weights are positive, reflecting the magnitude of importance regardless of the direction of the relationship.

In fact, the linear regression for the buy side and sell side were considered to be the predictor variables, where size was contributed to the coefficient estimator to assign the weight to each state.

Here the distribution of test and validation set is the test size of 20% and 80% for the validation size.

Under the 3 different time intervals: 10 seconds, 60 seconds, and 120 seconds. The formulated regression model should be adapted for 3 different time horizons to achieve a better understanding of the model and prediction in the significance of changes in trading frequency.

### 2.5.2 Handle Size Parameter

As for the raw dataset some redundant data still not be used. However, in referring to 1.3, the market impact or say the market trend would be largely affected by the size of the limit order book, no matter on the buy side or sell side. It plays a huge role in affecting the price and prompting the market to its expected flow. For instance, at this time interval, MB only appeared once whilst MS appeared 10 times. However, MB showed 100 times of shares in the indicated quoted price level, which raised the price and drove the market to more buying power, or say instant bull market. Although MB only emerged once, it still had overwhelming buying power compared to the selling power of 10 times that emerged of MS. Therefore, it should be considered in the rolling window model of market impact.

After estimating the mean size of 60-second time intervals for both buy-side and sell-side, as well as the correlation coefficient, finding its relationship could reveal the weights for different states in the model. Since the regression model needs factors to weigh if some states take the majority of market impact, the size will be considered to be one of the Explanatory Variables.

### 2.5.3 Multivariate Hawkes Process Prediction

For different time horizons: 10 seconds, 60 seconds, and 120 seconds. Each time horizon should be estimated by putting different interval lengths in the model prediction. Maintained the intensity function 2.20 of the Multivariate Hawkes process fixed by MLEs value inserting, then we predicted the intensity for the interval  $[t, t+10]$  for 10 seconds time intervals. Subsequently, we predicted the number of events arriving in the next time interval based on past 10 time intervals.

To ensure the prediction operating, the first 10 steps of time intervals which is 100 seconds under the condition of 10 seconds time horizon were eliminated and could not be shown on the predicted graph after plotting.

### 2.5.4 RMSE and Correlation Evaluation

After the manipulation of prediction value, the model for prediction of the number of events arriving and the subsequent rolling window model were both significant to be analyzed using various approaches. Here 2 approaches were introduced where Root Mean Square Error (RMSE) was used for evaluating the accuracy of the prediction of events arriving. It represented the sample standard deviation of the differences between predicted values and observed values (known as residuals). ([Hyndman and Athanasopoulos \(2018\)](#)) RMSE gave the error between actual data and predicted value. The formula was listed

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (2.24)$$

where  $n$  is the total number of time intervals  $y_i$  is the number of events arrivals for the  $i^{th}$  intervals and  $\hat{y}_i$  is the predicted number of events arrivals for the  $i^{th}$  intervals.

Finally, the rolling window model was also urgent to be evaluated. Another Correlation evaluation would be made after the plotting of both the predicted buying power of the market trend and the real buying power of the market trend. The Euclidean Distance, mean squared error (MSE) and Pearson Correlation coefficient 2.5 would be implemented for analyzing the model performance.

Euclidean Distance was the straight-line distance between two points in Euclidean space,

which indicated the geometric similarity between the predicted buying power graph and the actual buying power graph. The formula is listed:

For two points  $P$  and  $Q$  in  $n$ -dimensional space with coordinates  $P(x_1, y_1, \dots, z_1)$  and  $Q(x_2, y_2, \dots, z_2)$ , the Euclidean distance  $d$  is:

$$d(P, Q) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + \dots + (z_2 - z_1)^2}. \quad (2.25)$$

For Mean Squared Error (MSE), it was similar to the RMSE method where no root was applied.

For  $n$  observed data points  $y_1, y_2, \dots, y_n$  and their corresponding predicted values  $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n$ , the MSE is:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (2.26)$$

## 3 Model Performance and Evaluation

This Chapter analyzes the model by the methods mentioned in Chapter 2. It processes the raw data, estimates the various point processes, and by the end, the model assessment uses AIC and Q-Q plots to get the best model.

### 3.1 Preliminary Data Insights

Following data differentiation, the data was filtered into buy and sell sides across six states in total. Additionally, the size parameter is under processing and will be ready for the prediction part.

#### 3.1.1 Events Arrivals for Each State

Firstly, after computing the arrival rate per second for each state, we get the table 3.1. This shows that averagely, LB, CS, LS and CB appear around 3 times per second while MB and MS appear less frequently by only 0.2 times per second. This is because the market buying and market selling would directly lead to the price change and will cost the real funds to perform MB and MS. So MB and MS are rare events compared to the no-cost events LB, CS, LS, and CB.

Also, after the plotting of the buy side event arrival of the first 10 seconds of the trading day by Figure 3.1. It is obvious that State 3 (MB) appears less frequent compared to State 1 (LB) and State 2 (CS).

To see the direct trend and discrepancy of each state under a 60-second time interval<sup>1</sup>, the scatter plot of different colors is shown by Figure 3.2. State 3 and State 6 show fewer events happening during the time intervals, which mostly fall below 50 times per 60-second time intervals. Whilst, States 1, 2, 4, and 5 show a similar trend, and states 3 and 6 show a similar trend. This plot will indicate some internal relationships between these states.

#### 3.1.2 Size Parameter Processing

After utilizing the Descriptive Statistics of the size parameter, the table 3.2 is shown. It can be seen that the discrepancy between minimum 1 and maximum values (33570),

---

<sup>1</sup>Since under a 10-second time interval, the plot becomes noisy and not smooth.

Table 3.1: The Event Arrival Rate from State 1 to 6 (per second)

State 1 (LB)	State 2 (CS)	State 3 (MB)	State 4 (LS)	State 5 (CB)	State 6 (MS)
2.7095	2.7755	0.2523	2.9295	2.7755	0.2356

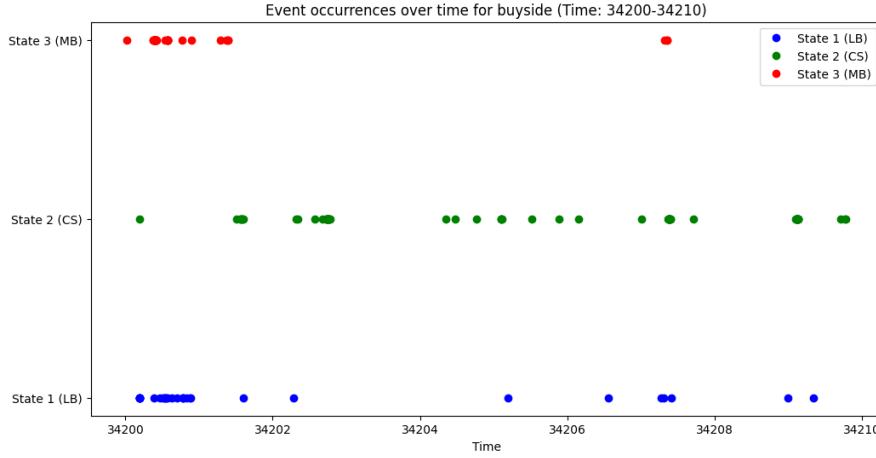


Figure 3.1: The event arrival shown as the point arrival for the buy side as an example. The time interval is from 34200-34210, which is the first 10 seconds of the trading day. Data from Amazon.com (AMZN), June 21st, 2012

indicates the existence of outliers with a mean of 96.87. Also, the box plot has proven the outliers in Figure 3.3.

Table 3.2: Descriptive Statistics of Size

	count	mean	std	min	25%	50%	75%	max
Value	269748.00	96.87	168.84	1.00	21.00	100.00	100.00	33570.00

So we deploy the capping method by replacing the outliers with the threshold value which is the 95th percentile of the confidence interval. Then calculate the mean size for each 60-second time interval for both the buy side and the sell side. The plot is shown in Figure 3.4. It is clear to see a similar trend of size between both sides.

However, since the thesis is analyzing various time horizons, the different time intervals should be applied. To find the relationships between the similarities of the trend of buy and sell side and the time intervals, the plot of the correlation coefficient of buy and sell side for different time intervals from 1s to 120s are shown in Figure 3.5. The greater time interval reveals the larger correlation coefficients. The thesis would mainly focus on the 10s, 60s, and 120s time intervals. Therefore, the correlation coefficient is always above 0.8322 which shows the strong positive relationships of the mean size between the

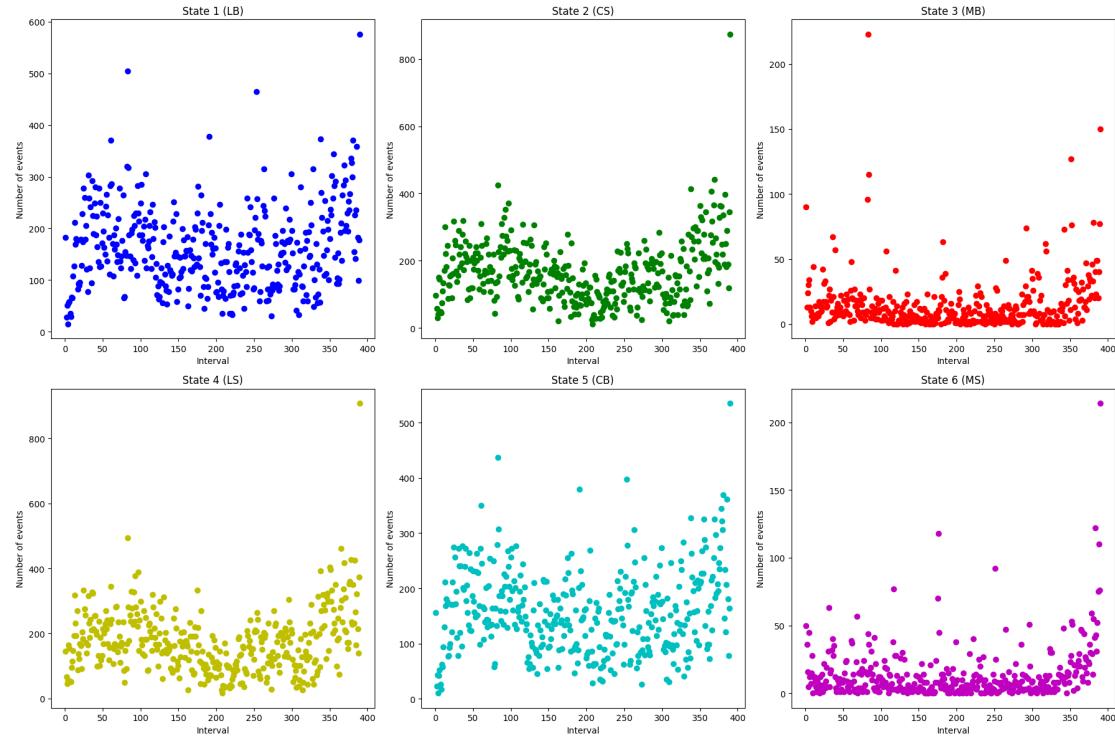


Figure 3.2: The plot of event arrival for each state under 60-second time intervals. Data from Amazon.com (AMZN), June 21st, 2012

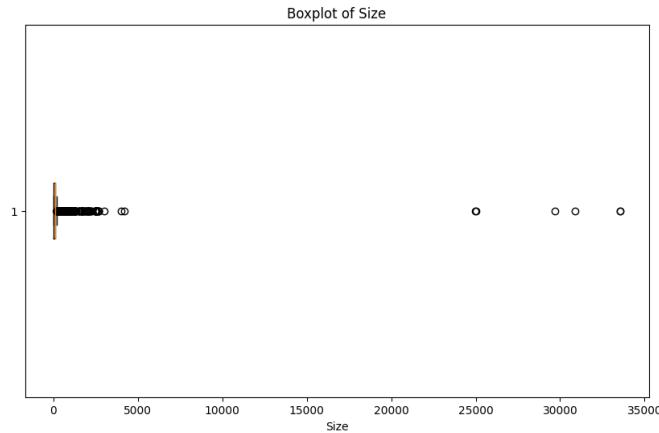


Figure 3.3: The box plot of size parameter of all time

buy and sell side. The result is 0.9681 for the 60s time horizon which can be considered to be the same distribution of the buy and sell side for larger time intervals.

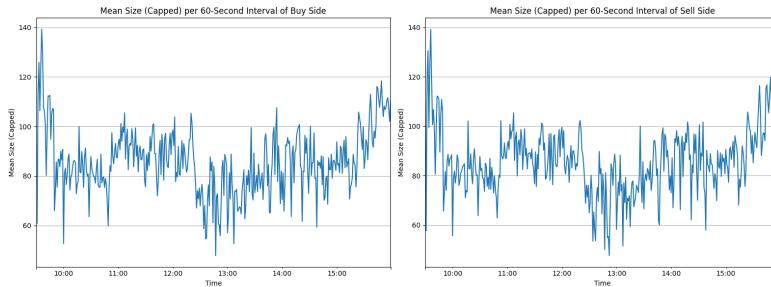


Figure 3.4: The mean size plot after capping method, under the condition of the 60-second time interval for both buy side and sell side

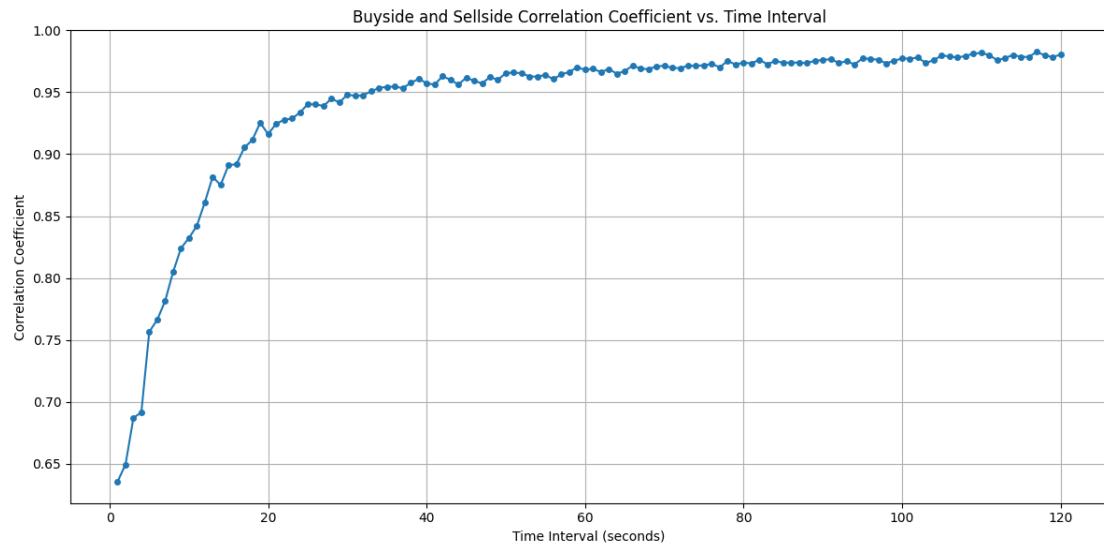


Figure 3.5: The correlation plot between the buy side and the sell side of mean size plot of different time intervals

## 3.2 Homogeneous Poisson Point Process Evaluation

As for the section 2.2.1, the Homogeneous Poisson Point Process (HPPP) has several assumptions which need to be tested. The following section will focus on testing if the model could fit the observed dataset.

### 3.2.1 95% Confidence Interval testing

Considering the 95% confidence interval of the buy side as the sampling data. We have to mention that, When calculating confidence intervals for a sample mean, we typically divide by the square root of the sample size (i.e., we divide by the standard error). This is because we assume that the sample mean will approach the population mean

as the sample size increases. However, in this analysis, we're not calculating confidence intervals for a mean. We're calculating confidence intervals for the number of events in a fixed time interval (e.g., every 10 seconds). In this case, our "sample" is the number of events in each 10-second interval. So, we use the mean and standard deviation of the number of events in each time interval to calculate the confidence intervals, rather than the sample mean and sample standard error.

Figure 3.6 shows the red dotted line as the 95% confidence interval where state 1 (LB), state 2 (CS), and state 3 (MB) are in blue. green and red. The corresponding 95% confidence interval is ([6.078549419139449, 319.0650403244503], [-10.278371031659447, 343.35016590345435], and [-26.429053781529767, 56.70597685845285]).

we need to note that the lower bounds of the confidence interval for state 2 and state 3 are negative, which does not make sense in our case, as the number of events cannot be negative. This may be due to the high heterogeneity of our data, causing the standard deviation to be very large relative to the mean. In this situation, using a Homogeneous Poisson Point process may not be the best modeling choice, as it assumes that events are evenly distributed across all time periods, which may not correspond to our data.

Furthermore, it shows that the number of events in many intervals far exceeds the confidence interval. This again suggests that data may not be suitable for modeling using HPPP.

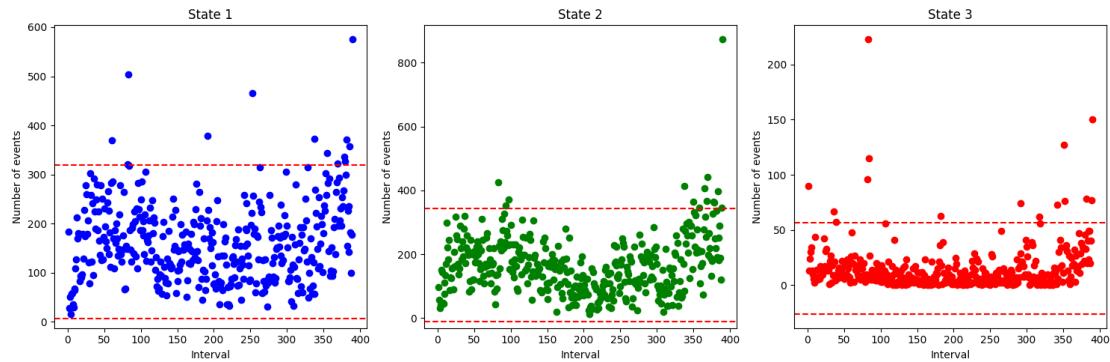


Figure 3.6: The 95% confidence interval of states 1, 2, and 3 under 60-second time interval. The red-dotted line is the confidence interval and each color represents each state differently

For a better understanding of the graphics and statistics, the time series plot is shown in Figure 3.7. It shows each 3 states lie in the buy and sell side and their relationships in time series. The time is starting from 34200s to 57600s of a trading day. Each state is represented in a different color.

As shown by the graph, the number of events varies significantly from interval to interval. So the intensity or the frequency is highly dependent on the time and varied by the time changing. So the assumptions of 'constant average rate' are not made since the obvious

non-average events happened in the time range. The homogeneous Poisson point process may mislead.

### 3.2.2 Basic Time Series

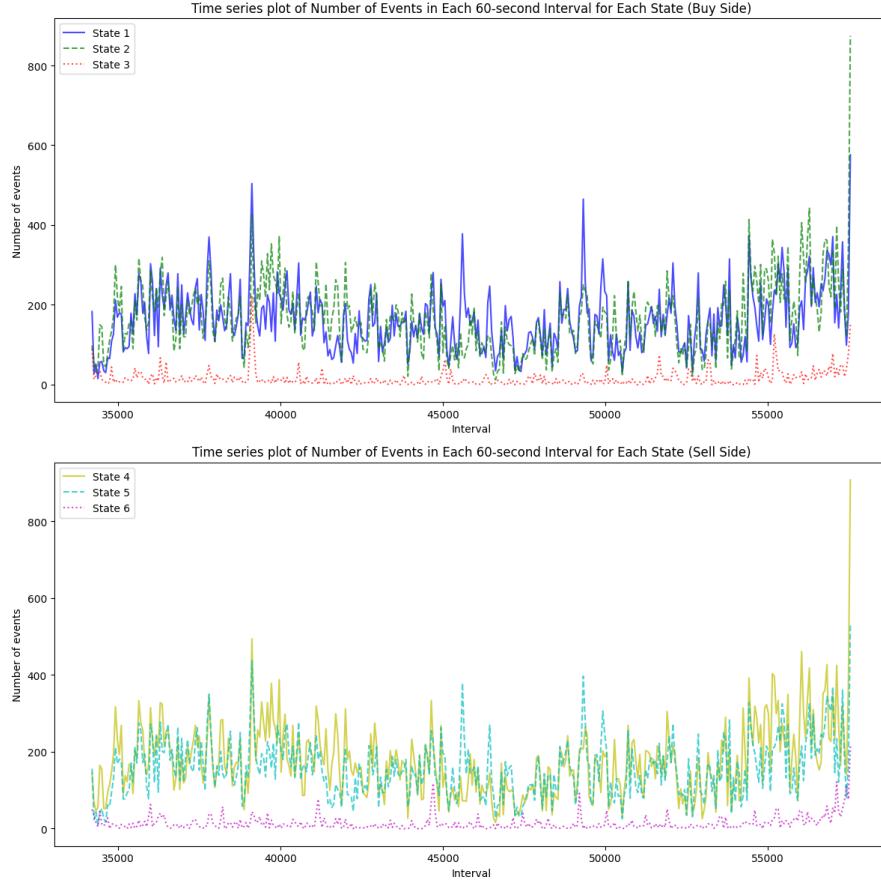


Figure 3.7: The time series plot of number of events in each 60-second intervals for each state from state 1 to state 6 on buy side and sell side. Data from Amazon.com (AMZN), June 21st, 2012

The homogeneous Poisson point process assumes that events are uniformly distributed across all time periods. However, our data show significant differences in the number of events across different time intervals, which contradicts the assumption of a homogeneous Poisson point process. Furthermore, a homogeneous Poisson point process also assumes that events occur independently, but in limit order book data from AMZN, the occurrence of events may be influenced by past events, which also contradicts the assumption of a homogeneous Poisson point process. Therefore, we may need to look for other modeling methods that are more suitable for our data, which will be discussed in the next few sections.

### 3.3 Inhomogeneous Poisson Point Process Analysis

After rejecting the HPPP assumptions, we get the knowledge that to adapt to the conditions in which the number of events happens non-uniformly. The varying intensity function is used.

#### 3.3.1 MA and EMA Implementation

Firstly, to visualize the time series plot. Here we implement the Moving Average (MA) and Exponentially Moving Average (EMA).

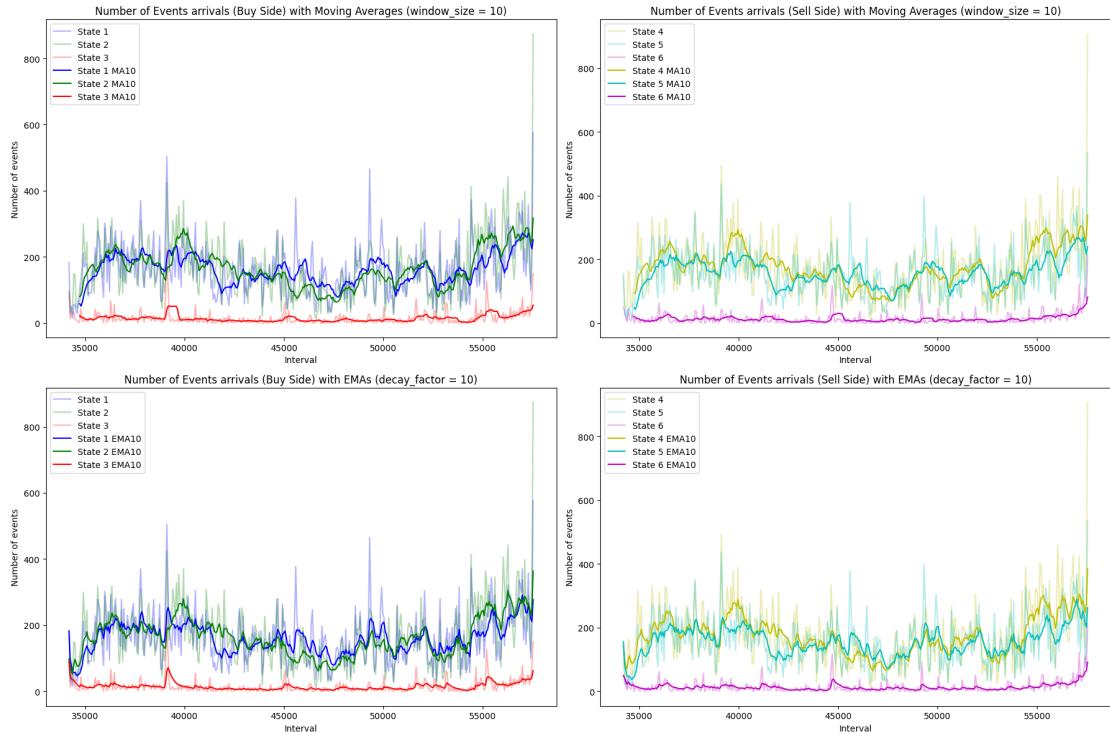


Figure 3.8: The MA, EMA, and time series plot of number of events in each 60-second interval for each state from state 1 to state 6 on buy side and sell side. Data from Amazon.com (AMZN), June 21st, 2012

The plot above shows the MA and EMA with the raw time series plot as the background. The raw time series for each state is represented by semi-transparent lines, while the moving average is shown by opaque lines. It is clearly to see that the moving average line is smoother than the raw data line, revealing the long-term trend of event counts for each state more effectively. This allows us to see changes in event counts over time without being distracted by short-term fluctuations.

It's worth noting that since EMA assigns more weight to recent data, the EMA line

responds quicker and is more sensitive to changes in event counts. Especially on the time interval of 39000 seconds. the soaring of State 1 is represented by EMA more quickly than MA.

The window size of the moving average is chosen as 10 in this case, which means that the rolling window would be 10 intervals for the 60-second interval. Therefore, the next MA curve is drawn by the past 10-minute time series data.

Also, increasing the window size will produce smoother results but might hide some minor fluctuations. Decreasing it will make the results closer to the raw data but might introduce more noise.

### 3.3.2 The Volatility Clustering

Also, Figure 3.8 of MA and EMA clearly shows the extreme varying under the condition of the 10-minute windows for 3 states. The fluctuations of a number of events arrival are significant and the spikes often appear for 3 states, especially for time around 39000 seconds.

Event flow shows similarities and the emergence of simultaneous spikes in the 3 different states of time series indicate inter-dependence. Meanwhile, a large amount of events arrival is commonly followed by another large event arrival, which suggests volatility clustering.

So to prove the properties of volatility clustering, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are induced to test.

According to Figure 3.9, it shows the high significance of each time series for each state from state 1 to state 6. The ACF and PACF for each state both remain high with a value of 0.13 for the first 10 to 15 lags. Especially for ACF of state 2 and state 4, the value decays extremely slow compared to other states and shows the increase of 95% confidence band. Even after 30 lags, the value of ACF is still exceeding 0.1.

The evidence above shows a great clustering, suggesting that large changes are followed by large changes.

### 3.3.3 States Dependency

The proven clustering has indicated that INHPPP may not be the best model selected. However, it is still worth finding the correlation of each state by drawing its dependency graph of the influence by time series of each state, which can bring a deeper understanding of the relationship between each data for further model establishment.

The scatter plot Figure 3.10 has indicated despite the state x vs. state x where shows strong positive relationships, state 1 with state 5, and state 2 with state 4 also show incredibly strong positive relationships.

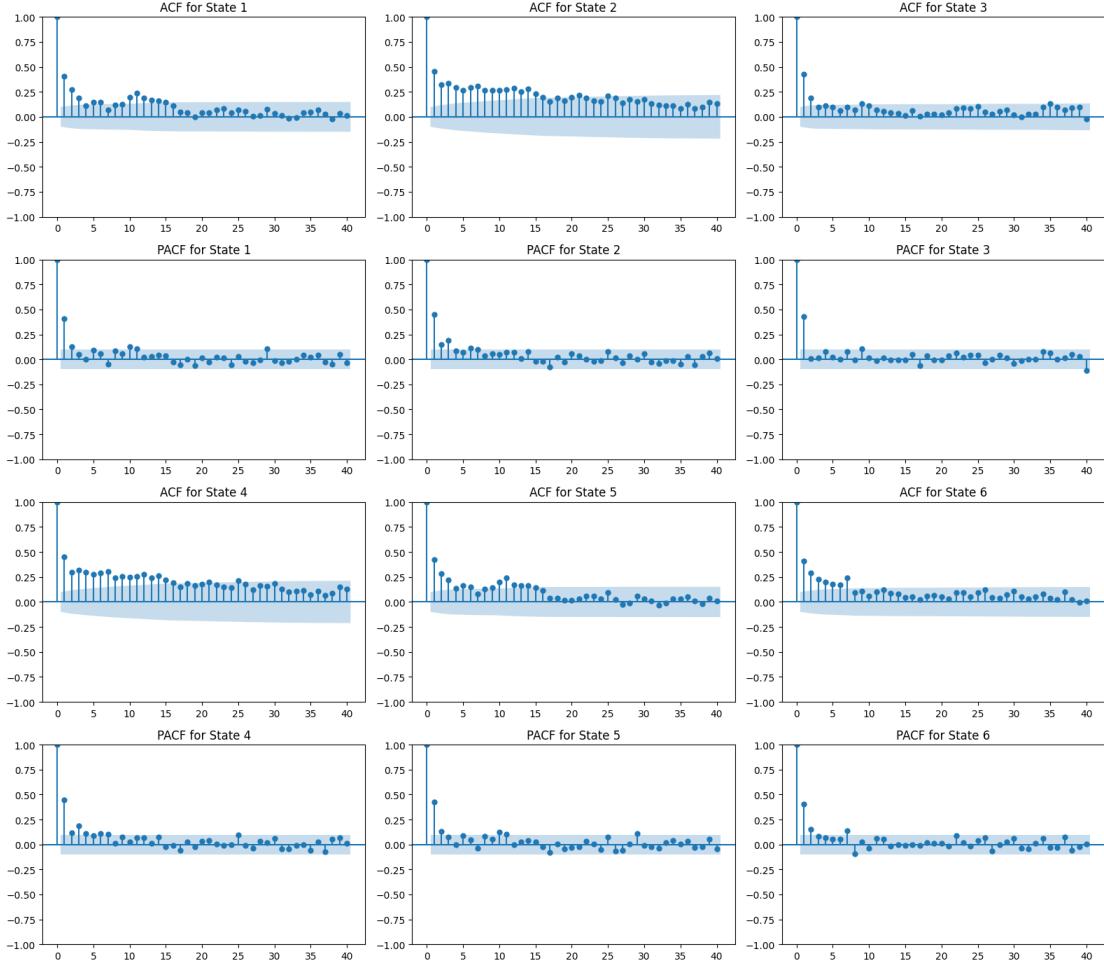


Figure 3.9: The ACF and PACF plot for each state from state 1 to state 6, The blue segment represents the 95 % confidence bands.

The table 3.3 using 4 different correlation coefficient approaches is also listed. The red values represent the highest 2 correlation coefficients among all pairs of states, and the green value represents the lowest 2 correlation coefficients among all pairs of states.

So for the pairs of limit order book buying (LB) related to the cancellation of the limit order book buying (CB), they show really strong positive relationships where the measures of correlation coefficient are closer to 1. The result can be explained explicitly by market behavior, where to maintain the buy side and sell side balanced and the increase in LB, would not obviously make the one side too strong to frighten the opponents investing. Also for the increase in CB, fewer order books for buy orders would trigger more investors to buy the dip, where more LB arrives. Also, The pairs of limit order book selling (LS) and cancellation of the limit order book selling (CS) show the same results.

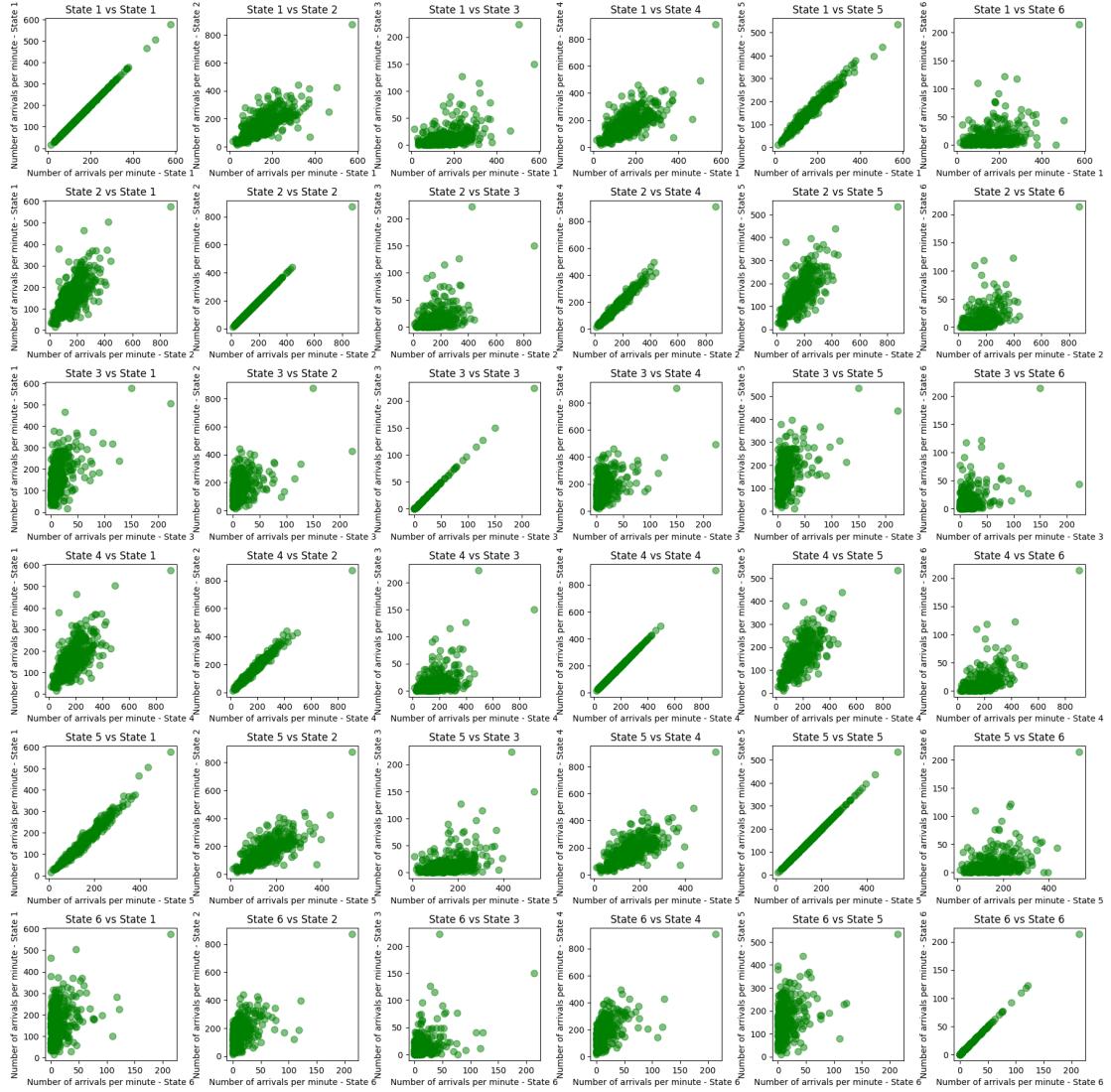


Figure 3.10: The scatter plot of the number of events arrivals for each state from state 1 to state 6 under 60-second time interval, and the graph shows the relationship between each pair of states respectively

Despite the largest correlation coefficient value, the pairs of MS and LB as well as the Paris of MS and CB show the weakest positive relationships among all pairs. Since LB and CB have almost the same distribution, they have a similar correlation value to MS. Intuitively, the submission of limit order book buying and market buying are 2 types of investors who are price taker and price maker. The price taker has no influence on the stock price and the price maker has the ability to influence the price by executing the limit order selling. Consequently, they have less correlations with others indeed.

Table 3.3: Correlation Coefficients between each state under 60-second time interval.  
 Correlations are shown in 4 different approaches: Pearson's, Spearman's, Kendall's Tau and Mutual Information

		State					
		LB	CS	MB	LS	CB	MS
State 1 LB	Pearson's		0.7198	0.5301	0.6970	<b>0.9827</b>	<b>0.3575</b>
	Spearman's		0.7185	0.5147	0.6972	<b>0.9810</b>	<b>0.2688</b>
	Kendall's Tau	1.00	0.5358	0.3705	0.5126	<b>0.8880</b>	<b>0.1832</b>
	Mutual Information		0.3905	0.1258	0.3522	<b>1.7484</b>	<b>0.0727</b>
State 2 CS	Pearson's	-		0.4457	<b>0.9853</b>	0.7307	0.5402
	Spearman's	-		0.4079	<b>0.9832</b>	0.7298	0.4481
	Kendall's Tau	-	1.00	0.2878	<b>0.8935</b>	0.5465	0.3111
	Mutual Information	-		0.0983	<b>1.7676</b>	0.4478	0.1114
State 3 MB	Pearson's	-	-		0.4832	0.4799	0.4222
	Spearman's	-	-		0.4071	0.4575	0.2986
	Kendall's Tau	-	-	1.00	0.2851	0.3264	0.2088
	Mutual Information	-	-		0.1256	0.1211	0.0000
State 4 LS	Pearson's	-	-	-		0.7216	0.5890
	Spearman's	-	-	-		0.7242	0.5220
	Kendall's Tau	-	-	-	1.00	0.5398	0.3668
	Mutual Information	-	-	-		0.4260	0.1708
State 5 CB	Pearson's	-	-	-	-		<b>0.3539</b>
	Spearman's	-	-	-	-		<b>0.2911</b>
	Kendall's Tau	-	-	-	-	1.00	<b>0.2002</b>
	Mutual Information	-	-	-	-		<b>0.0426</b>
State 6 MS	Pearson's	-	-	-	-	-	
	Spearman's	-	-	-	-	-	
	Kendall's Tau	-	-	-	-	-	1.00
	Mutual Information	-	-	-	-	-	

### 3.3.4 Non-Parametric Estimator of The Intensity Function

To further evaluate the INHPPP, we can derive some approaches by using the Non-Parametric Estimator to fit the process, which are Kernel Density Estimation (KDE), Kernel Smoothing (KS), and Exponential Moving Average function (EMA). The time interval is still 60 seconds.

Since the plot would be too big to fit in the paper, so take only state 1 as an example. Figure 3.11 and Figure 3.12 show the KDE and KS intensity function using 5 different kernels which are: Gaussian (Normal) Kernel, Tophat (Uniform) Kernel, Epanechnikov

Kernel, Exponential Kernel, and Cosine Kernel in different colors. Both KDE and KS show almost the same distribution.

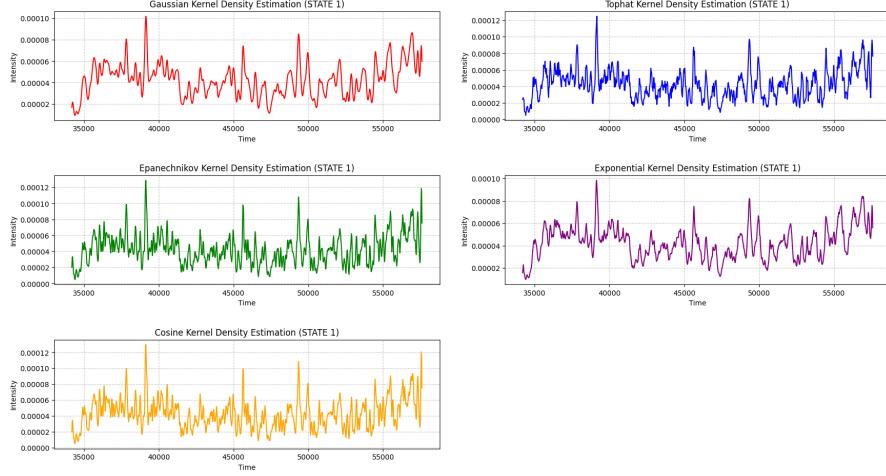


Figure 3.11: The intensity function using KDE for state 1 (LB) under 5 kernels

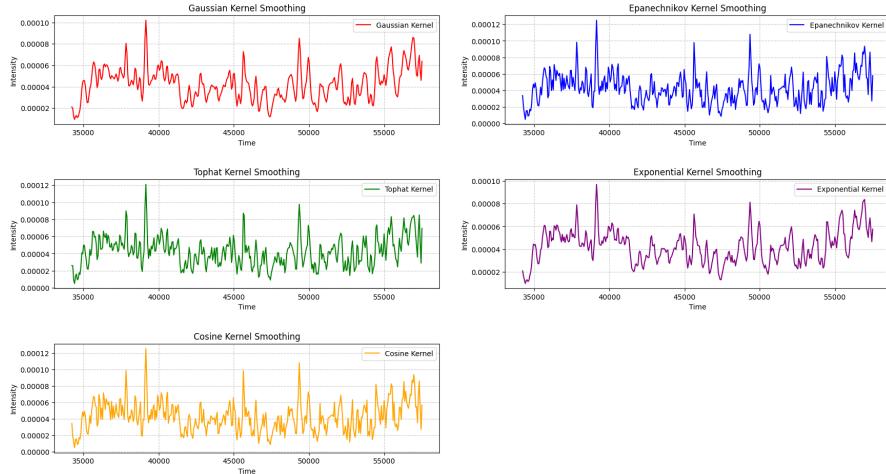


Figure 3.12: The intensity function using KS for state 1 (LB) under 5 kernels

Figure 3.13 displays the intensity function using EMA function. The grey part is the raw number of event arrivals, and the green line with the light green shaded area is the EMA line with the 95% confidence band for state 1 (LB).

### 3.3.5 Q-Q Plot of Residuals of INHPPP

Here we use Q-Q residual plots to perform the model assessment for 3 approaches of Non-Parametric Estimator of The Intensity Function.

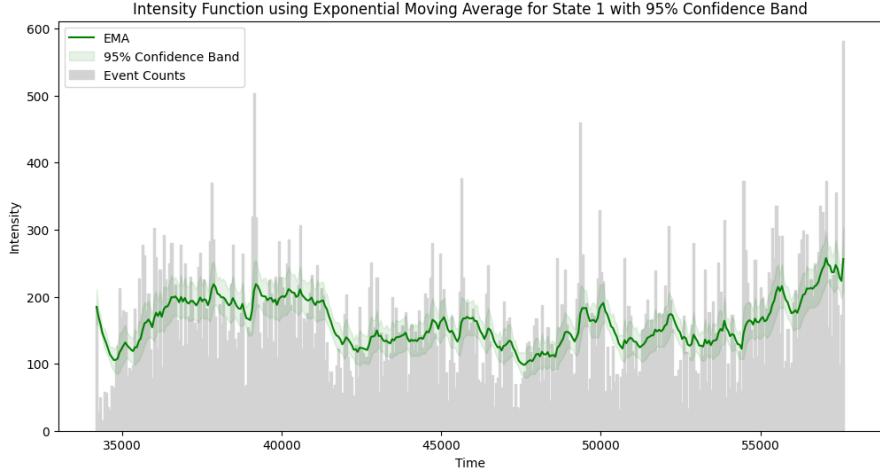


Figure 3.13: The intensity function using EMA for state 1 (LB) with 95% confidence band

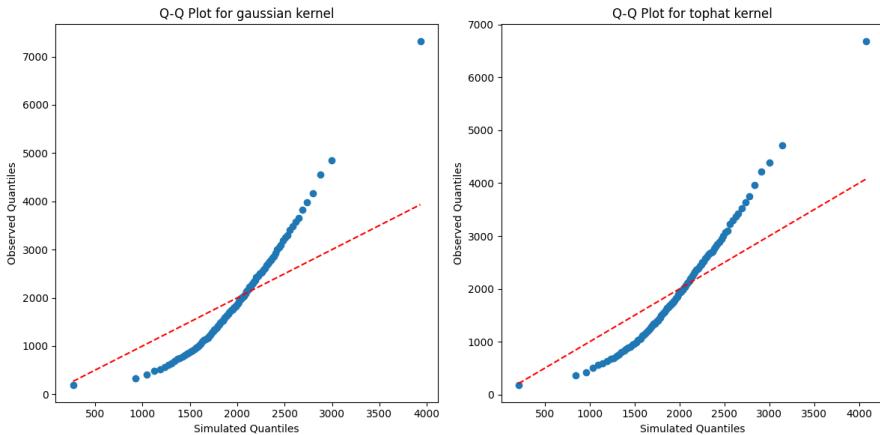


Figure 3.14: Q-Q plot for the residuals with the intensity function driven by KDE and KS. The kernel is chosen to be Gaussian and Tophat. The data is from state 1 (LB) of Amazon.com (AMZN), June 21st, 2012

Figure 3.14 reveals the Q-Q plot of residuals for intensity function under Gaussian and Tophat kernels with the KDE and KS approaches.

Since the Q-Q plots for KDE and KS are almost the same, to save more space, only the Q-Q plot for KDE is displayed. The reason why we chose only 2 kernels rather than 4 is because the remaining kernels are not supported by the libraries. Only 2 kernels are estimated accordingly. Also to mention the number of Monte Carlo simulations, only 100 times of simulation is implemented rather than 1000 times since too waste time and no such improvements of Q-Q plots under more simulation case.

Figure 3.14 also shows the non-linear relationship between the constructed non-parametric intensity INHPPP model and actual data residuals. The model may lack information to explain the residual data also the arrival of events for state 1 (LB).

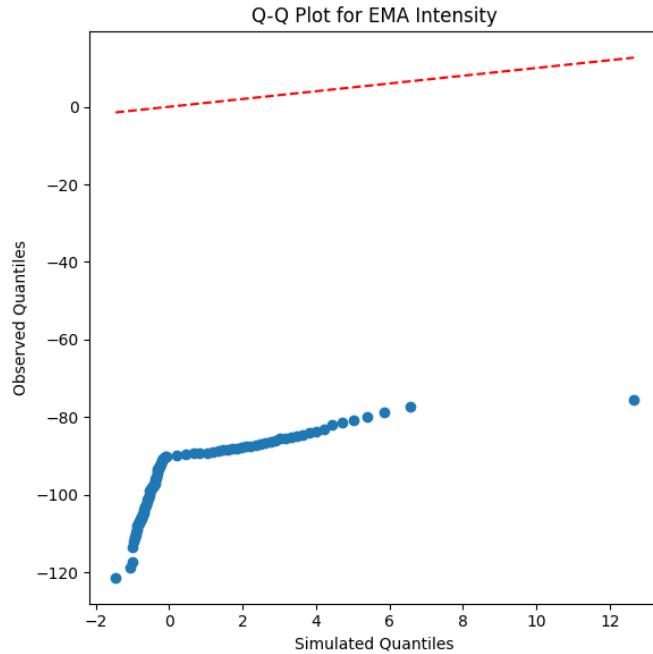


Figure 3.15: Q-Q plot for the residuals with the intensity function driven by EMA. The data is from state 1 (LB) of Amazon.com (AMZN), June 21st, 2012

Figure 3.15 also reveals the Q-Q plot of residuals for the intensity function under the EMA function. However, the model is more deviated than KDE and KS. In my perspective, it may be the cause of volatility clustering, as well as the self-exciting of data itself. Therefore, using the non-parametric measures to establish the model may not be appropriate. A better model should be applied.

## 3.4 Cox Process Evaluation

Under the Poisson assumption of the Cox process, we need first to determine the intensity function  $\Lambda_{COX}$  of the Cox process for further analysis.

### 3.4.1 Intensity Chosen

For better variable control, The intensity  $\Lambda_{COX}$  is determined to be the same intensity function as the Multivariate Hawkes Process. So the  $\Lambda_{COX}(t)$  is the same as equation 2.20, Where

$$\Lambda_m^{COX}(t) = \mu_m + \sum_{n=1}^d \sum_{t_i^n < t} \alpha_{mn} \exp(-\beta_m(t - t_i^n)), \quad (3.1)$$

the log-likelihood function of the Cox process was derived by equation 2.17. So the final MLE would be the same as the Multivariate Hawkes process which is in table 3.4. The intensity plot is the same as Figure 3.21

### 3.4.2 Q-Q Plot for Cox Process

So due to the same intensity but different assumption of the Poisson point process, the Cox process is still different from the 'Pure' Multivariate Hawkes Process. The Q-Q plot of residuals for state 1 given by the Cox process is shown below as Figure 3.16:

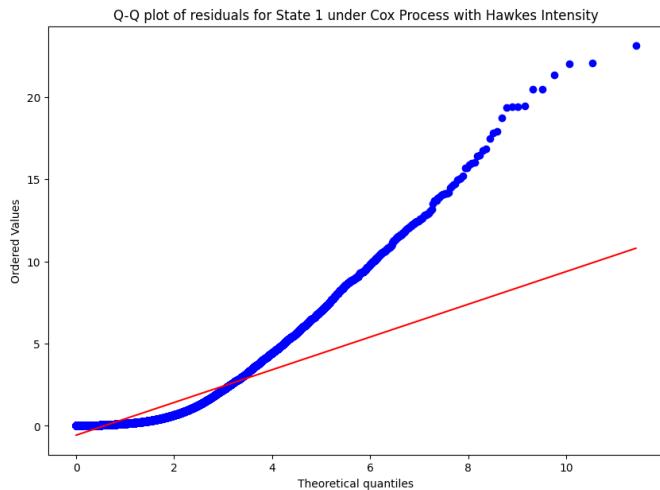


Figure 3.16: Q-Q plot for the residuals of the Cox process under the assumed intensity as the same as the Multivariate Hawkes Process. The data is from state 1 (LB) of Amazon.com (AMZN), June 21st, 2012

The Q-Q plot for residuals only fits a bit well under lower quantiles but does not capture the quantile over 3. The plot shows the non-linear relationship between the Cox process model and actual data.

The result is not satisfied and it is because of the assumption of the Poisson process given a constant rate of events arrivals. So in this case, the observed events arise from a Poisson process, but the rate of events is influenced by past events in a way described by the Multivariate Hawkes process by several dimensions.

### 3.5 Multivariate Hawkes Process Performance

After testing the volatility clustering and self-exciting properties, it would be beneficial to adapt the Multivariate Hawkes Process from state 1 to state 6. However, due to the high model complexity, we will reduce the dimension to 4 states which are states 1, and 3,4,6 for LB, MB, LS and MS respectively. We eliminate state 2 and state 5 which are CS and CB, and it is due to the high correlation between state 1 and state 5, and state 2 and state 4 by Table 3.3. Also, state 2 and state 4 which represent the cancellation of the limit order book, can be treated as the same behavior of reducing the buy queue and sell queue just like the MS and MB. So we retain the submission part of both buy and sell sides which are state 1 and state 4 while eliminating the state 2 and 5 since they represent the same property of 3 parameters  $\alpha$ ,  $\beta$ , and  $\mu$

And for the prediction part Chapter 4, when state 1 arrives, it will always accompanied by the arrival of state 5, and when state 4 arrives, it will always accompanied by the arrival of state 2.

#### 3.5.1 MLE Parameters and Algorithm

After the dimension reduction, m is now equal to 4, which is applied to each log-likelihood formula. Then, using the Method of Moments (MoM) and Cross-Correlation Function (CCF) 2.4.3 to compute the decay parameter  $\beta$  for each state. Then by equation 2.21, MLE of  $\beta$  has already been proceeded and considered to be fixed, so only 20 parameters are in the MLE estimation. which are 4  $\mu$  and 20  $\alpha$

During the time of minimizing function using Nelder-mead downhill simplex, it takes more than 40 hours to estimate the result. So we adapt the following measures to help speed up:

- Reduce the maximum Iteration of the minimizing process. After a maximum iteration of 50, the results become extremely large due to the algorithm of Nelder-mead downhill simplex which will excavate the global maximum rather than the given intervals. Also, the time of processing MLE for higher iteration would exponentially increase the time spent. After several tests, the ideal iteration is around 20 to 25 which saves time and is not dispersed to the global maximum.
- Reduce the sample size estimated, do it by segmentation. Since the sample size for each state is too large to process. Even when cutting 2 states, there are still 200000 data waiting for padding to the time series. So cut it to 1/4 and analyze each part to bring a better approach and save time.<sup>2</sup>
- Shrink the Observation Windows. Due to the reduction of sample size, the observation windows are shrunk to 1/4 by the sample size. However, the data still follows the time series by sequence.

---

<sup>2</sup>Still more than 12 hours for MLE processing

- Better Initial Guess of  $\mu$  and  $\alpha$ . Which by considering the  $\mu$  as the rate of arrival [3.1](#) on the time interval of the observation windows. So  $\mu_{guess} = [2.7095, 0.2523, 2.9295, 0.2356]$ . While alpha is the correlated intensity, like  $\alpha_{m,1LB} == [3, 1, 1, 1]$  means the state 1 happens mostly trigger next state 1 happens indeed
- Gradient-Based method. After computing the first derivative of log-likelihood, using the L-BFGS-B and Powell's method to quickly find the MLE.
- Parallel Optimization. It activates all the cores of the computer to utilize the parallel computation to improve the exercise speed. using Parallel(njobs=-1)
- Monte Carlo approach, using simulation of 1000 to compute. However, no MLE results are given in this paper.

Table [3.4](#) shows the computed MLE using the above-detailed measures:

Table 3.4: Maximum Likelihood Estimators for the time interval of 60-second

States	$\beta$	$\mu$	$\alpha_{m,1LB}$	$\alpha_{m,3MB}$	$\alpha_{m,4LS}$	$\alpha_{m,6MS}$
State 1 LB	5.3074	0.3395	3.9409	0.9199	1.4089	1.2278
State 3 MB	6.3483	0.0214	1.2025	4.0593	1.1671	1.2458
State 4 LS	7.0619	0.5987	1.2532	0.8013	5.0085	1.0113
State 6 MS	4.9644	0.0200	1.2025	1.2458	1.1671	4.0593

$\beta$  represents the decay rates of the influence of past events from different states on the current state's intensity function. A higher beta indicates that the influence of past events decays more rapidly, meaning that past events have a shorter-lasting impact on the current intensity. Conversely, a lower beta means that past events have a longer-lasting impact.

State 4 has the highest decay rate ( $\beta = 7.062$ ), meaning past events in State 4 have the shortest-lasting impact on the intensity of other states. State 6 has the lowest decay rate ( $\beta = 4.964$ ), meaning past events in State 6 have a longer-lasting impact compared to those in the other states.

$\mu$  represents the baseline intensities for each state in the absence of any past events. It's the rate at which events would occur if there were no self-excitation or mutual excitation with other states.

State 4 has the highest baseline intensity ( $\mu = 0.599$ ), suggesting that, all else being equal, events in State 4 are the most frequent. State 6 and State 3 have the lowest baseline intensities, suggesting that events in these states are relatively rare when not influenced by other states.

$\alpha_{mn}$  represents the strength and direction of influence between states. An entry  $\alpha_{ij}$  represents how much state  $j$  influences state  $i$ .

The diagonal elements represent self-excitation. The value at  $\alpha_{4,4}$  is 5.009, suggesting a strong self-excitation for State 4 and Off-diagonal elements represent mutual excitation between different states

The matrix appears to be not entirely symmetric, suggesting that the influence between states is not always reciprocal. For example, while State 3 has some influence on State 1 (with a value of 0.9199), State 1 has a slightly larger influence on State 3 (with a value of 1.2025).

Also, Figure 3.17, represented the abse line intensity  $\mu$  for 4 states in the histogram. Figure 3.18, represented the kernel influence of paired state 1 and state 2. Figure 3.19, represented the  $\alpha_{mn}$  as a diagram chart.

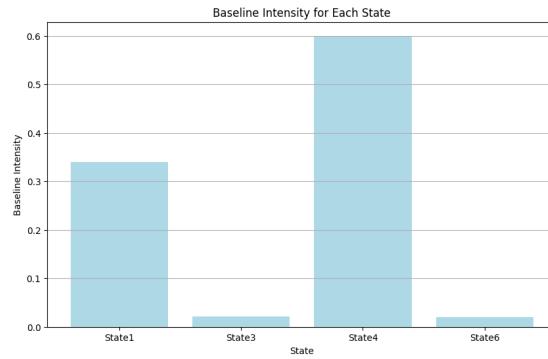


Figure 3.17: Histogram of Baseline intensity of state 1, state 3, state 4, and state 6

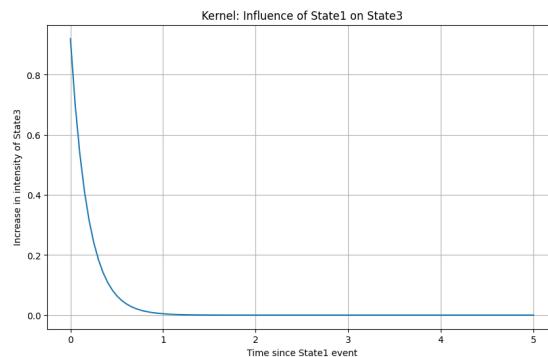


Figure 3.18: kernel influence of state 1 on state 3

In summary, State 4 has the highest baseline intensity  $\mu$  and strong self-excitation  $\alpha_{4,4}$ . State 6 has a longer-lasting impact on other states.

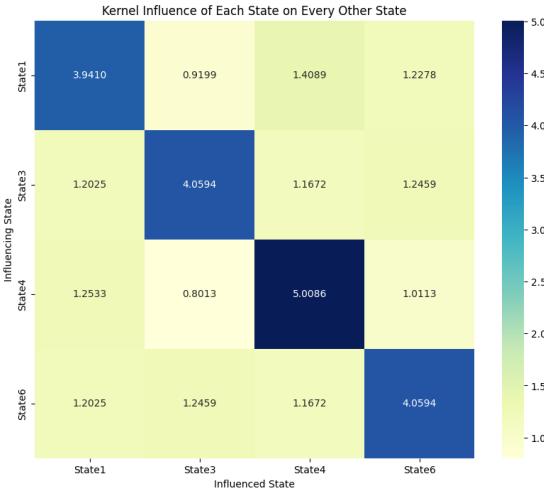


Figure 3.19: strength and direction of influence between each state. The darker the color, the larger the value

### 3.5.2 Self-Exciting Intensity

Firstly, Figure 3.20, reveals the simple 'No Influence' case where only state 1 triggers itself as a self-exciting process with red points as events arrivals. The intensity would be only reacted to the state 1 events arrivals.

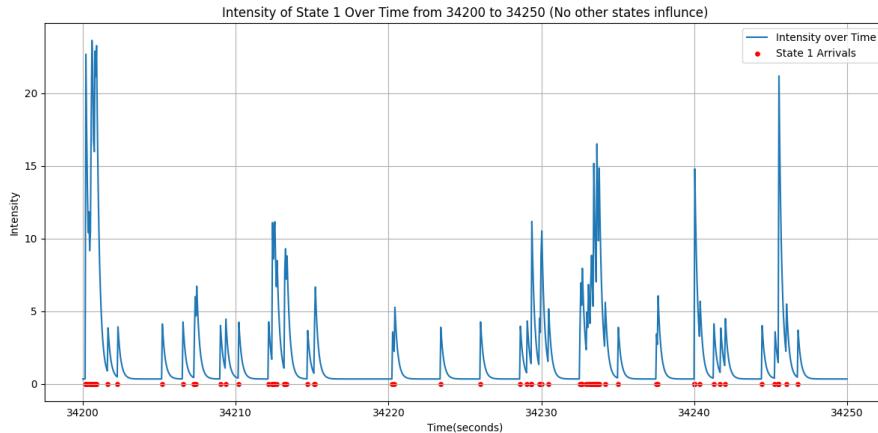


Figure 3.20: The intensity function of Multivariate Hawkes Process of state 1 from time interval of 34200 to 34250 where no other states influenced.

Figure 3.21 shows the intensity function of submission of limit order buying for the first 50 seconds, where the MLE are proceeded by Table 3.4. The plot shows the intensity of state 1 and the change made by other states, also shows some jumps due to the mutual effects of other events arrivals. Events arrivals for all 4 states are shown on the bottom

of the plot, which is below the x-axis. it can be directly seen from the plot that the event occurs then immediately accompanied by the exponential decay by  $\beta_1$ .

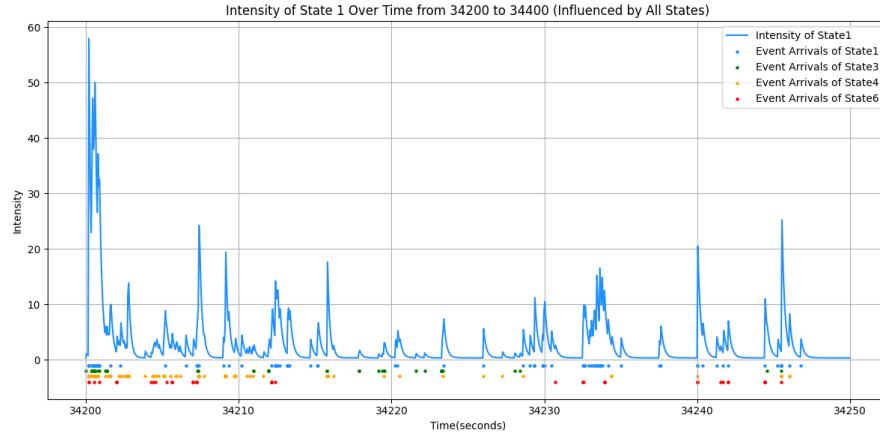


Figure 3.21: The intensity function of Multivariate Hawkes Process of state 1 from time interval of 34200 to 34250 where all other states influenced.

Figure 3.22 is like the 3.21, but add the intensity functions of the remaining 3 states together to see the correlation. Blue = state 1, Green = state 3, Orange = state 4, red = state 6 with all event arrivals. It can be seen that state 4 as the orange color would have a larger intensity value for each short time interval. It is because of the more events of states 4 itself coming and highest baseline intensity  $\mu$  and strong self-excitation  $\alpha_{4,4}$ .

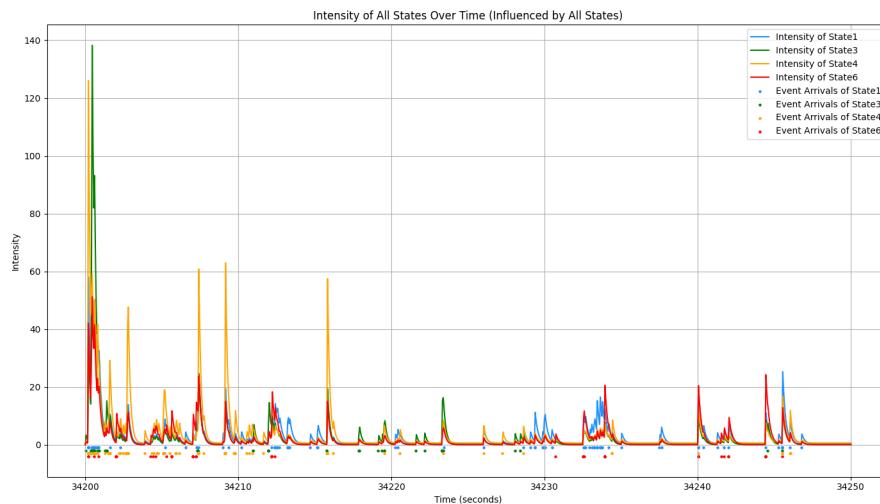


Figure 3.22: The intensities of all states of state 1, state 3, state 4, and state 6 of Multivariate Hawkes Process

### 3.5.3 Model Assessment

Here is the Q-Q plot of residuals for state 1 and the Akaike Information Criterion (AIC) value of Multivariate Hawkes Process, Cox process, AR(4), MA(4), and ARIMA(4,0,4).

For Figure 3.23, the Q-Q plot of the Multivariate Hawkes Process, reveals the goodness-fit of the model for the first 99% quantile, which is overwhelming compared to the previous models. However, the dispersed tail of the last 1% quantile indicates the probability of incapability of capturing the large value events. So might be insufficient to predict the sudden rise of a large number of event arrivals within a short time interval.

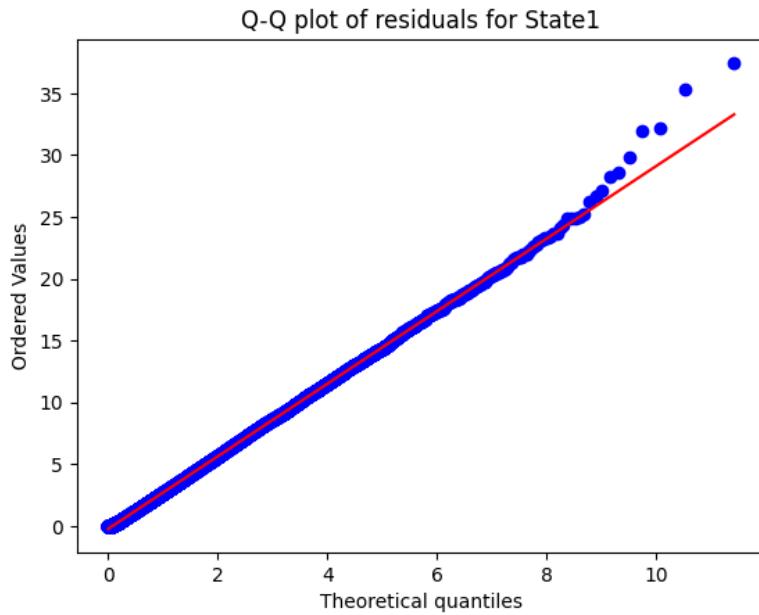


Figure 3.23: The Q-Q plot for residuals of Multivariate Hawkes Process for state 1

Table 3.5: AIC Results for Different States and Approaches

State/Approach	State1	State3	State4	State6
Multivariate Hawkes	<b>-256797.47</b>	<b>-20487.54</b>	<b>-304283.08</b>	<b>-17020.46</b>
Cox Process ( $\Lambda$ is Multivariate Hawkes)	<b>-256797.47</b>	<b>-20487.54</b>	<b>-304283.08</b>	<b>-17020.46</b>
AR(4)	-8.64	-6.08	-8.75	-5.99
MA(4)	-8.65	-6.08	-8.76	-5.99
ARIMA(4,0,4)	-0.63	1.92	-0.75	2.00

Table 3.5 indicates the smallest value in red, which is the Multivariate Hawkes process and Cox process with  $\Lambda$  as a Multivariate Hawkes process. This suggests that the

Multivariate Hawkes process provides the best fit to the data among all the models considered, for every state.

However, The AIC values for the Cox process with  $\Lambda$  modeled as a Multivariate Hawkes process are identical to the AIC values of the Multivariate Hawkes process itself. This indicates that, in this modeling approach, the Cox process is essentially equivalent to the Multivariate Hawkes process in terms of model fit.

The ARIMA model has the highest AIC values among all the models, which indicates the poorest fit to the data

Also, these AIC results imply that under the Multivariate Hawkes process, the model is not nested and is sufficient and significant for subsequent predictions.

## 4 Model Application to Rolling Window

To compute the rolling window prediction using the Multivariate Hawkes Process, it is essential to find the buying power using the formula 2.22. Therefore, the steps of building the Ordinary linear regression to find the  $\beta_{BUY}$  and  $\beta_{SELL}$  which are the weights of each state respectively, are made by sm.OLS and the .fit() in Python. The test size is set by 20% of the whole time series and the rest 80% of the time series is treated as the validation set.

### 4.1 Weights Manipulation

Based on the basic weights manipulation cut the coefficient to make the weights added equal to 1. Table 4.1 shows the weights results:

Table 4.1: Weights with MSE and  $R^2$  for Different States and Time Intervals

Time Interval/State	Buy-side			Sell-side		
	State1	State2	State3	State4	State5	State6
10s Weights	0.1875	0.1576	0.6549	0.0647	0.1294	0.8058
60s Weights	0.1774	0.1749	0.6477	0.0231	0.1492	0.8277
120s Weights	0.2190	0.2365	0.5446	0.0573	0.1716	0.7711
10s MSE	0.004984			0.005233		
60s MSE	0.0006704			0.0006355		
120s MSE	0.0003530			0.0003788		
10s $R^2$	0.0786			0.0326		
60s $R^2$	0.0778			0.1259		
120s $R^2$	0.1423			0.0796		

For the buy-side, State3 has the highest weight across the whole time interval of 0.6549, 0.6477, and 0.5466. For the sell-side, State3 has the highest weight across the whole time interval of 0.8058, 0.9277, and 0.7711. The value suggests state 3 and state 6 have the most influence in the model for Buy-side and Sell-side.

From Table 4.1, the MSE indicates the average squared difference between the observed actual value and the values predicted by the model. A smaller MSE is preferable as it indicates a better fit of the model to the data. For instance, under a 120s time interval,

it represents the lowest MSEs which are 0.00035 for the buy-side and 0.00037 for the sell-side, implying good predictability. For the buy-side, the MSE decreases as the interval increases, suggesting the model's predictions are more accurate at longer intervals.

The  $R^2$  indicates the proportion of the variance in the dependent variable that's predicted by independent variables. For instance, the largest  $R^2$  is the 120s time interval for the buy-side, which is 0.1423, implying that 14.23% of the variability in the dependent variable can be explained by the model. For both buy-side and sell-side, the  $R^2$  values aren't very high, suggesting the model doesn't explain a large portion of the variance in the data.

## 4.2 Rolling Window Fitting under Various Time Horizon

Since the dataset is too large to compute the plot like 2.1, it only has 200 points while The AMZN has 200000 more data to compute, which would cause the noise and unsmooth of the plot of Buying Power to insufficiently express the market trend. Therefore, only the top 10% of the data from AMZN is selected for a 10-second time interval, the top 50% of the data from AMZN is selected for a 60-second time interval and no observation data changes for a 120-second time interval since only 196 intervals in this case.

So after the manipulation of the observed data, the observed intervals for the different states are similar and around a number of 200, but in different horizons where the total plot of the market trend under 10-second is the scaling of the first 10% plot of the market trend under 120-second time interval.

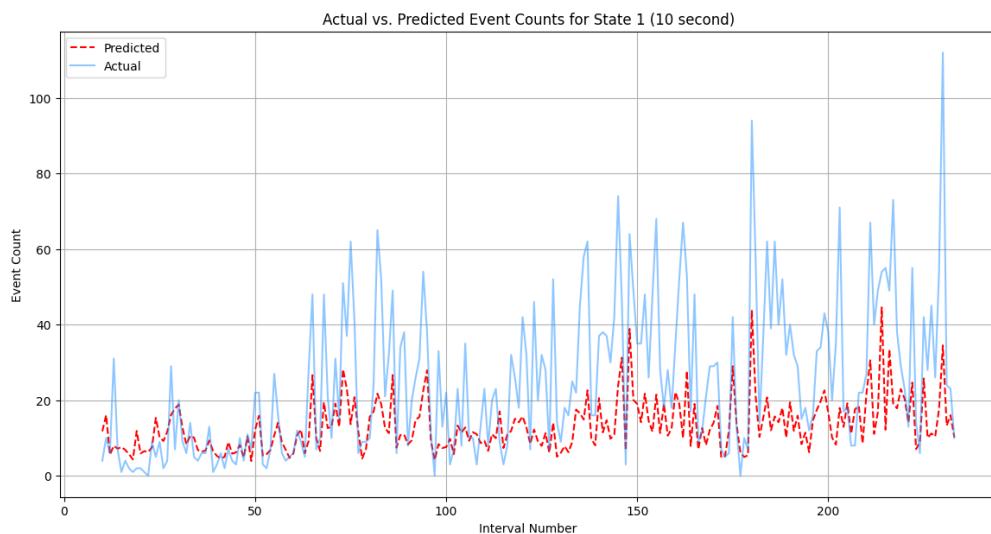


Figure 4.1: The Actual plot of events arrivals of state 1 under top 10% of 10-second time intervals with respect to the predicted plot of events arrivals. Data from Amazon.com (AMZN), June 21st, 2012

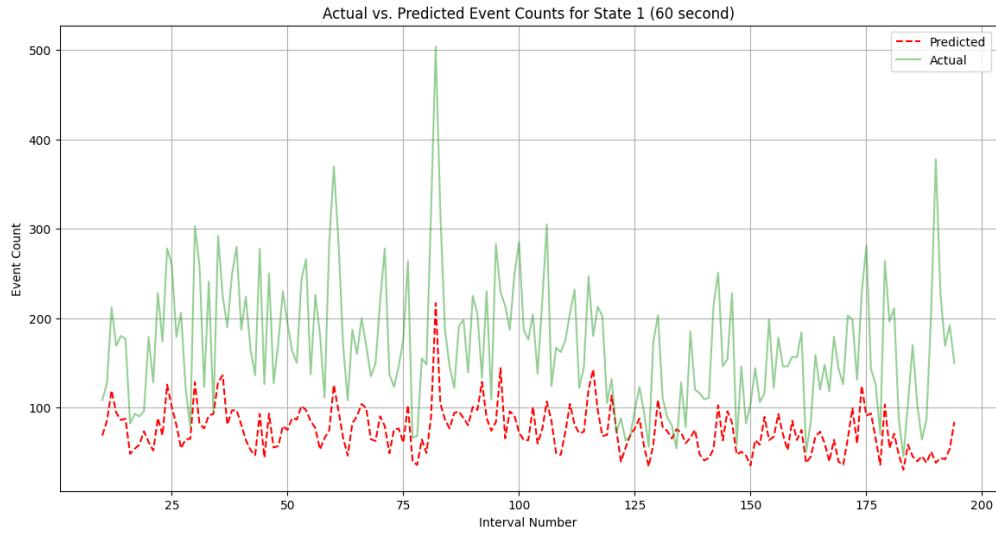


Figure 4.2: The Actual plot of events arrivals of state 1 under top 50% of 60-second time intervals with respect to the predicted plot of events arrivals. Data from Amazon.com (AMZN), June 21st, 2012

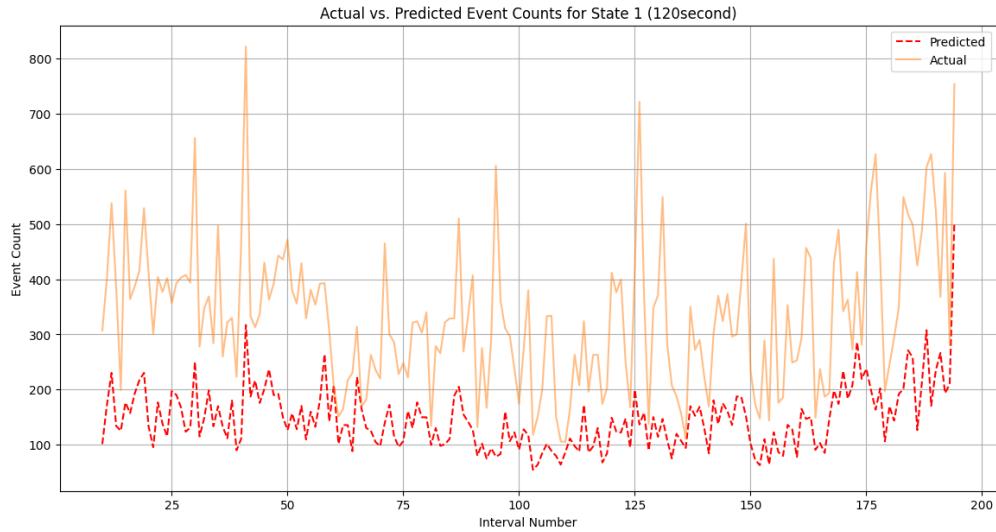


Figure 4.3: The Actual plot of events arrivals of state 1 under 120-second time intervals with respect to the predicted plot of events arrivals. Data from Amazon.com (AMZN), June 21st, 2012

Here, Figures 4.1, 4.2 and 4.3 represent the actual and prediction of event arrivals under time horizons of 10s, 60s, and 120s for state 1. The predicted line is shown as a red dotted line and the background line is the actual events that happen across the given

time intervals.

Since it is the rolling window prediction, and the rolling window is chosen as 10, the first 10 intervals are erased and start from the 11th interval. It is obvious to see that from the start, only a 10-second time interval performs well for the fitted line, while a 60-second and 120-second time interval indicate the same shape but different amplitude. However, after several time intervals under a 10-second time horizon, the predicted start to perform the same behavior as the 60-second and 120-second time horizon.

In the case of a 10-second time horizon, the predicted events arrivals become stable and accurate for the interval 100 where the number of actual events arrivals reduces a lot from 40 to around 10. For the other 2 time horizons, a large value of event arrivals emerges continuously across each interval. Also based on the plot of Figure 3.23, which shows the incapability of capturing the large events arrivals for over 35 per interval. Therefore, the Multivariate Variate Hawkes Process shows the same shape but different amplitude referring to the capability of capturing the trend across the entire time horizon and the lack of ability to capture the large events arrivals, where the plot always shows the fewer events arrivals. So the differentiation of the time intervals to smaller segments like 10ms or 100ms will be better for the model fit and prediction.

### 4.3 Market Trend Display and Evaluation

Firstly, we need to compute the actual market trend by showing the buying power using the formula 2.22. Then to draw the plot of the predicted market trend compared to the actual plot. Then use RMSE and some correlation between the predicted plot and the actual plot to evaluate the Multivariate Hawkes model performance.

The plot of actual market trend Figure 4.4 and the plot of predicted market trend Figure 4.5 are shown. Directly from the plot, the actual market is in a bull market trend where the buying power has a value always above 0.5 across all time horizons, indicating the buying power dominated the market for AMZN on June 21st, 2012.

However, the predicted plot has shown the market behaved as a bear market among all time horizons. For the top 10% of the 10-second intervals, only the short range of intervals around the 50s are above 0.5 where most of them are under 0.5. 60-second and 120-second time intervals have also shown the same bear market trend where strong selling power dominates the market.

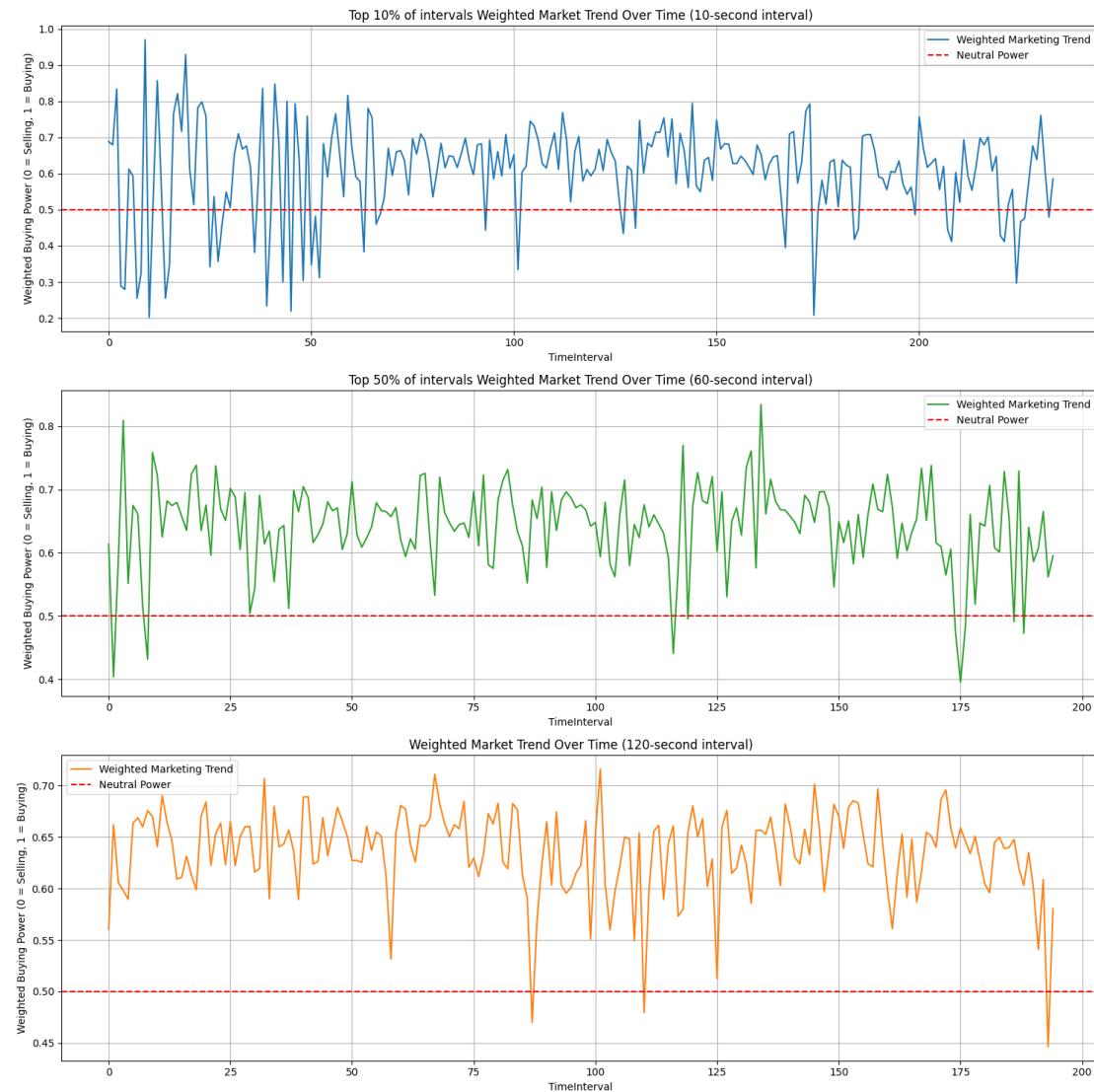


Figure 4.4: The actual plot of the market trend by showing the Buying power. Closer to 1 refers to a long market, closer to 0 refers to a short market, and closer to 0.5 refers to a neutral market. Data from Amazon.com (AMZN), June 21st, 2012

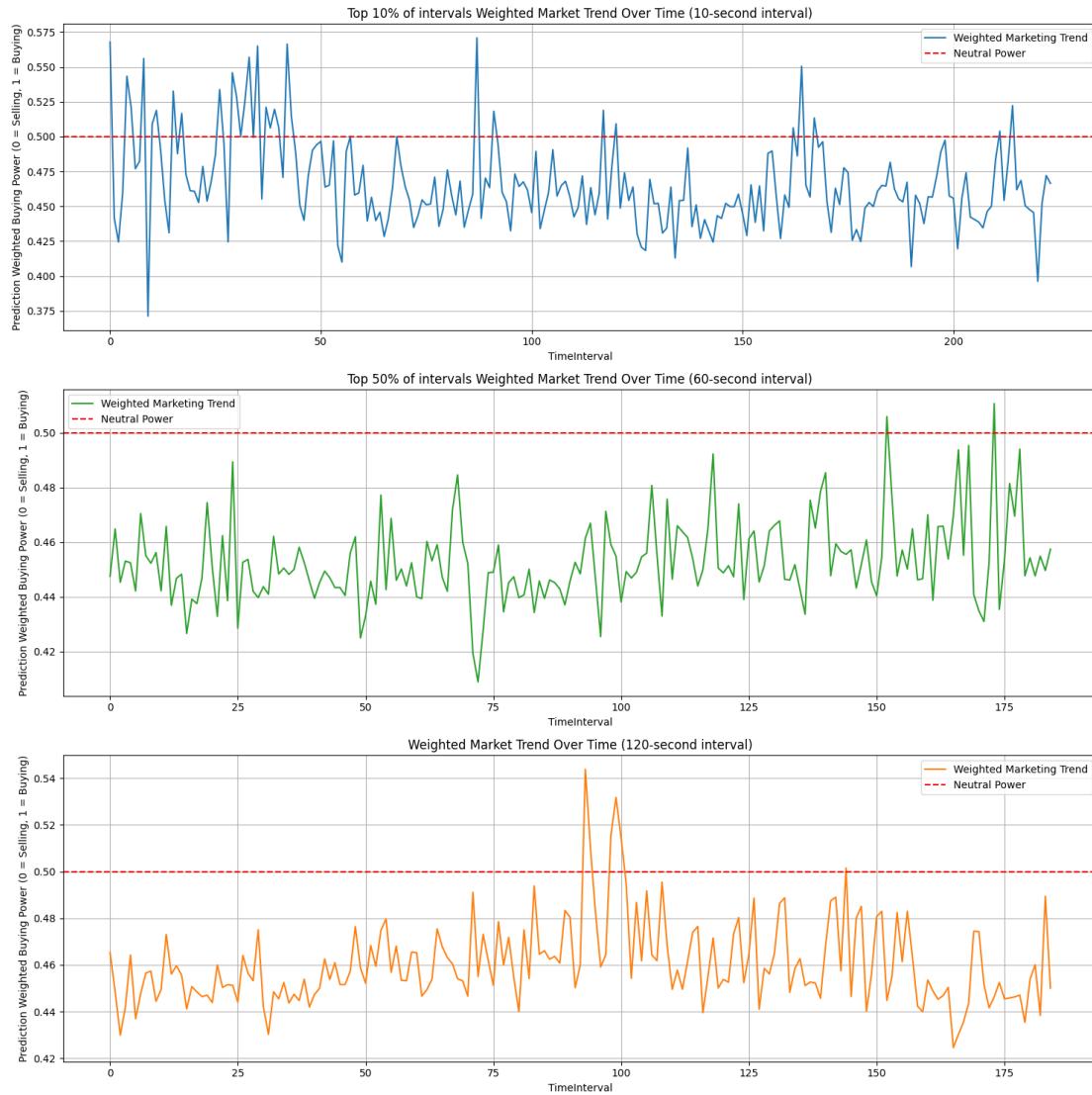


Figure 4.5: The predicted plot of the market trend by showing the Buying power. Closer to 1 refers to a long market, closer to 0 refers to a short market, and closer to 0.5 refers to a neutral market. Data from Amazon.com (AMZN), June 21st, 2012

So for a better understanding of the plot, here RMSE Table 4.2 of Multivariate Hawkes Process-based model compared to models: AR(4), MA(4), and ARIMA(4,0,4) is shown. Sadly to see that only 10-second time intervals under 10 rolling windows for State 1 show better RMSE using the Multivariate Hawkes Process-based model compared to the rest of the models. Interestingly, with the increase of time interval, the RMSE becomes larger than expected for each state. It might be caused by the lack of ability to capture large values shown by the Q-Q plot of the Multivariate Hawkes Process since large intervals indicate larger values within the single interval.

Table 4.2: RMSE for Various Predictions Across States and Time Horizons

Intervals	States	Multivariate Hawkes Process	AR(4)	MA(4)	ARIMA(4,0,4)
10s	State 1	<b>19.24</b>	20.76	20.28	22.68
10s	State 3	11.04	5.63	5.54	5.91
10s	State 4	27.03	23.06	22.09	30.68
10s	State 6	15.36	4.27	4.08	11.97
60s	State 1	111.80	72.88	73.18	74.21
60s	State 3	61.00	22.13	21.86	32.31
60s	State 4	133.44	85.32	81.15	102.03
60s	State 6	83.53	26.63	18.03	50.33
120s	State 1	94.83	72.88	73.18	74.21
120s	State 3	118.74	37.87	17.67	40.75
120s	State 4	260.57	170.94	166.70	189.71
120s	State 6	118.94	37.87	37.67	40.75

Table 4.3: Predicted Market Trend Evaluation

Time Interval	Euclidean Distance	Mean Squared Error	Pearson's
10s	2.995	0.04003	-0.1481
60s	2.769	0.04145	-0.03878
120s	2.485	0.03338	-0.1175

Table 4.3 shows the comparison between the prediction of the market trend and the actual market trend measured by the buying power.

The Euclidean distance for each state is not quite high with 2.995, 2.769, and 2.485. It decreases as the time interval increases, suggesting that predictions for longer intervals (120s) are closer to the actual values compared to shorter intervals (10s).

The MSE values are almost the same across all time intervals which reveals a relatively small value, indicating that the actual and predicted results are fairly close to each other.

Pearson's correlation coefficient values are all negative, indicating an inverse linear rela-

tionship between the predicted and actual market trends. The value is closest to 0 for the 60s interval, suggesting the weakest inverse linear relationship for this interval.

## 5 Conclusion and Further Thought on Current Model

After conducting several experiments on various Poisson Point Processes, including the Homogeneous Poisson Point Process (HPPP), the Inhomogeneous Poisson Point Process (INHPPP), and the Cox Process (under the assumption of the Poisson Process), as well as the Multivariate Hawkes Process, it was found that the Limit Order Books (LOB) for both the buy and sell sides were best modeled by the Multivariate Hawkes Process.

During testing, this paper delved into basic point processes. The findings indicated that neither HPPP nor INHPPP were suitable for the model, as evidenced by the inconsistent rate of arrivals with volatility clustering, the observed inter-dependence between each state, the self-exciting phenomenon, and the strong deviations observed in the Q-Q plot.

Under the assumption that the Cox process has the same intensity as the Multivariate Hawkes Process, the Q-Q plot revealed a poor fit for the LOB time series.

Among all the Point Processes evaluated, the Multivariate Hawkes Process outperformed the rest. It also provided a better fit than traditional time series models like AR, MA, and ARIMA, as indicated by the AIC value.

Subsequently, an OLS regression model was established to gauge market trends by assessing buying power using the number of arrivals for each state across various time intervals. However, using the Multivariate Hawkes Process for the rolling window prediction of LOB resulted in unsatisfactory predictability, as measured by the RMSE value. The predictions exhibited consistent shapes but varied in amplitude. This might stem from the Multivariate Hawkes Process's inability to capture significant events.

To enhance the current model, initiating with the MLE estimation of the Multivariate Hawkes Process is recommended. For simplicity, during the Python coding phase, fewer states were defined and minimal data was displayed. However, the unpredictability of the MLE remains a significant factor affecting the model. Thus, it's advisable to consider full dimensions for states 1 through 6 and the complete time series from 34200 to 57600. A deeper analysis of the raw data could provide a more informed initial estimate.

The MLE estimation coding requires significant refinement. During the minimization process, I experimented with maximum iterations ranging from 5 to 200. While increasing iterations can substantially lengthen computing time, it can yield improved results. The iteration count should be judiciously selected based on the available data. During

the Monte Carlo simulation, the process was hampered by the vast amount of data computed. Implementing a more efficient algorithm could accelerate this. In my intensity and integral intensity functions, I employed a loop for the time range. Utilizing NumPy array operations instead of internal loops might enhance speed and is worth exploring in future iterations.

Regarding the rolling window prediction of market trends using the Multivariate Hawkes Process across different time intervals, several modifications could be beneficial. The application of OLS regression might be substituted with a superior regression model incorporating diverse factors. Although I omitted the size parameter due to the high correlation between the buy and sell sides, reintroducing them as predictors and creating a separate model to forecast the size trend over time might be advantageous.

The prediction evaluation indicated a suboptimal fit. I suspect the rolling window concept might be contributing to this inadequacy. I opted for a window size of 10, but exploring alternative sizes and time intervals could be fruitful. The rolling window model necessitates a swift response to event changes, and the sudden surge in order arrivals typifies the clustering nature of the LOB, where parameters exhibit pronounced instability. The Multivariate Hawkes Process has inherent lag issues, suggesting that transitioning to a different model, such as one focusing on state dynamics, might yield superior results.

## Bibliography

- Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.
- Dominic Isaac Andoh and Raul Kangro. Limit order book modelling.
- Mike Bowe, Stuart Hyde, and Ike Johnson. Emg working paper series. 2009.
- George Casella and Roger L Berger. *Statistical inference*. Cengage Learning, 2021.
- Yuanda Chen. Modelling limit order book dynamics using poisson and hawkes processes. 2013.
- Rama Cont. Volatility clustering in financial markets: empirical facts and agent-based models. In *Long memory in economics*, pages 289–309. Springer, 2007.
- Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume II: general theory and structure*. Springer, 2008.
- Daryl J Daley, David Vere-Jones, et al. *An introduction to the theory of point processes: volume I: elementary theory and methods*. Springer, 2003.
- Paul Embrechts, Thomas Liniger, and Lu Lin. Multivariate hawkes processes: an application to financial data. *Journal of Applied Probability*, 48(A):367–378, 2011.
- Larry Harris. *Trading and exchanges: Market microstructure for practitioners*. OUP USA, 2003.
- David Haynes, Steven Corns, and Ganesh Kumar Venayagamoorthy. An exponential moving average algorithm. In *2012 IEEE Congress on Evolutionary Computation*, pages 1–8, 2012. doi: 10.1109/CEC.2012.6252962.
- Tor Helleseth. Some results about the cross-correlation function between two maximal linear sequences. *Discrete Mathematics*, 16(3):209–232, 1976. ISSN 0012-365X. doi: [https://doi.org/10.1016/0012-365X\(76\)90100-X](https://doi.org/10.1016/0012-365X(76)90100-X). URL <https://www.sciencedirect.com/science/article/pii/0012365X7690100X>.
- Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.
- Olav Kallenberg et al. *Random measures, theory and applications*, volume 1. Springer, 2017.

- Maurice G Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.
- Maxime Morariu-Patrichi and Mikko S Pakkanen. State-dependent hawkes processes and their application to limit order book modelling. *Quantitative Finance*, 22(3):563–583, 2022.
- John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.
- Katarzyna Obral. Simulation, estimation and applications of hawkes processes. *Doctoral dissertation, Master's thesis*, 2016.
- Maureen O'hara. *Market microstructure theory*. John Wiley & Sons, 1998.
- Karl Pearson. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58:240–242, 1895.
- H Robert et al. Time series analysis and its applications with r examples second edition, 2006.
- David W Scott. *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 2015.
- Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.
- Charles Spearman. The proof and measurement of association between two things. *The American journal of psychology*, 100(3/4):441–471, 1987.
- Kunal Tiwari, Krishna Mehta, Nitin Jain, Ramandeep Tiwari, and Gaurav Kanda. Selecting the appropriate outlier treatment for common industry applications. In *NESUG Conference Proceedings on Statistics and Data Analysis, Baltimore, Maryland, USA*, pages 1–5, 2007.
- Ekaterina Vinkovskaya. *A point process model for the dynamics of limit order books*. Columbia University, 2014.
- Matt P Wand and M Chris Jones. *Kernel smoothing*. CRC press, 1994.