UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2012:11

Irregularly-Spaced Financial High-Frequency Data Simulation Using Multi-Dimensional Hawkes Processes: Estimation, Prediction And Corresponding Trading Strategy

Long Zhou

Examensarbete i matematik, 30 hp
Handledare och examinator: Kaj Nyström
Juni 2012

Department of Mathematics
Uppsala University

Acknowledgments

To thank my amiable supervisor, Mr. Professor Dr. Kaj Nyström comes always to the first position. Without his appreciation, I would not have the opportunities to deal with situations in real financial industries and to talk with professional specialists in field of algorithmic trading before getting out of ivory tower. His guidance always inspires me of incoming problems' solving and leads me going deeper into financial high-frequency trading analysis. He makes me feel so strongly about what I'd like to do after graduation and my potential in software programming.

Being a master in mathematics, my fiancé forwardly gives me plenty of ideas in detection and test (applicable or not). He tolerates (or enjoys) my bad tempers when frequently no results displayed after running codes for over 72 hours or I am stuck by awful problems. He stands with me all the way to help me pass through my most desperate time when the third time my laptop collapsed and everything fell in that battle, even though he does not live in Sweden. His humourous expressions and latest heard gossips make me feel amused and relaxed all the time. Thanks a lot for all what he has done for me, ignoring the teasing part.

Possibly my mom will not read my thesis, but I still feel grateful for her encouraging me forever and supporting me mentally and physically. Maybe she does not understand what I am doing right now, but she is always glad to be my loyal audience, exhorts me not to stay up in the late night.

Thanks to NASDAQ OMX Stockholm for supporting me and providing their precious data; thanks to all co-workers in my team that we discuss a lot and make progresses together after each time of meeting; thanks to all my friends and family members for their unlimited support.

Contents

Chapter 1

Introduction

1.1 Background and Motivation

With the growing popularity of financial high-frequency trading, the availability of modelling high-frequency data has inspired the analyses of quantitative finance. Furthermore, presently simulating high-frequency data is a mainstream area in relevant specialized field owing to its remarkable industrial applications. Investors are informed of trading forms and rules (which differ widely from one exchange to another) in advance, then they strive to optimize their trading strategies according to reports offered by specific electronic trading platforms. As a consequence, the academic research in econometric modelling is of greater interest owing to the expansion of OTC (over-the-counter) trading. Simultaneously, this expansion feeds a greater variety of challenges: unpredictability increases with not only the innovation of trading principles, but also invention of fully or partially unobserved orders (iceberg orders, non-displayed orders, etc). The difficulty and complexity determine that there remains large area of blank in financial point process analysis. Therefore, it is worth researching this field of study with pace of HFT (high-frequency trading) market evolution. Important tasks of high-frequency data analyses are to simulate financial point process, to describe the features of each single process, to improve the model's goodness-of-fit, to predict price dynamics, next event arrival timestamp or the volatility of price movement, etc.

This thesis focuses on a sort of self-exciting intensity process which is called Hawkes process. For instance, such kind of intensity process is particularly interested for its advantages in modelling event arrival processes in continuous time by modelling the dynamics of intensities directly. Corresponding empirical analyses including goodness-of-fit research and stationary conditions of these simulations are given in detail. Also the high-frequency market making strategies under different circumstances according to estimation and forecast results are discussed. Empirical data are supported by NASDAQ OMX Stockholm, therefore, temporally the framework of this article lies on NASDAQ trading forms and rules.

1.2 Basic Concepts

Order and Order Book

An order represents an instruction of a trader who cannot personally negotiate his trades and therefore determines what to trade, when to trade and how much to trade (Hautsch (2011)). Buyers and sellers negotiate and bargain for a stock (or bond, currency, etc) through electronic means, all series of traders' activities (offer, bid or cancel) regarding the stock (or bond, currency, etc) are recorded on a "book". This "book" is the so-called order book. Best bid price (or bid quote) is the price at which the sellers most willingly to accept, whereas best ask price (or ask quote) reflects the situation on the opposite side. The discrepancy of bid and ask quote is known as the bid-ask spread. When best bid equals to the best ask price, it is called that BBO (best bid and offer) is attained. Note that the best ask shall be always no less than the best bid, otherwise everybody would buy orders at the best ask and sell them at the best bid, i.e., there exist arbitrage opportunities, then BBO is almost attained in quite a short time according to the low latency trading frequency.

Market Order and Limit Order

When specifying the order, it is ordinarily taken under discussion: whether it is a market order or limit order. To set his order position at the market order price is the simplest strategy for an investor when he wants to trade. A market order indicates that the trader wants to pay the current ask quote or sell it at the best bid immediately. If the investor's willingness come true, the order is executed immediately, being recorded in the order book as a market trade. The logic of the market trade is that it fills as much as possible at the best bid (ask) level regardless of the remaining volume, no matter how large quantity of order is supplied. However, a limit order seems to be more flexible since the traders can state the price at which he would like to trade (may not be best prices). It stipulates restrictions to best price levels. If the incoming order is not same as the bid (ask) quote, it is recorded on the order book (but not cancelled) and participates in the queue of orders according to time priority. As is summarized in figure 1, all the information mentioned above is illustrated, moreover, it is explicit that the state of limit order book could be depicted by bid (ask) prices and quantities of diverse (first and second) limit bid (ask) orders at the bid (ask) side.

NASDAQ Trading Foundations

NASDAQ stock market is a sort of quote-driven dealer market where their member traders can execute their trades immediately without communicating with dealers in advance. Orders here can be executed flexibly, fully or partially in several steps. If an investor places his order at the market order to sell or buy, he is willing to trade at the currently best bid or ask price, but there is no guarantee for his order arrival in time and execution exactly at the market price. Stock orders with the same price should stand in queue and be executed according to the trading priority principles in corresponding exchange. As is

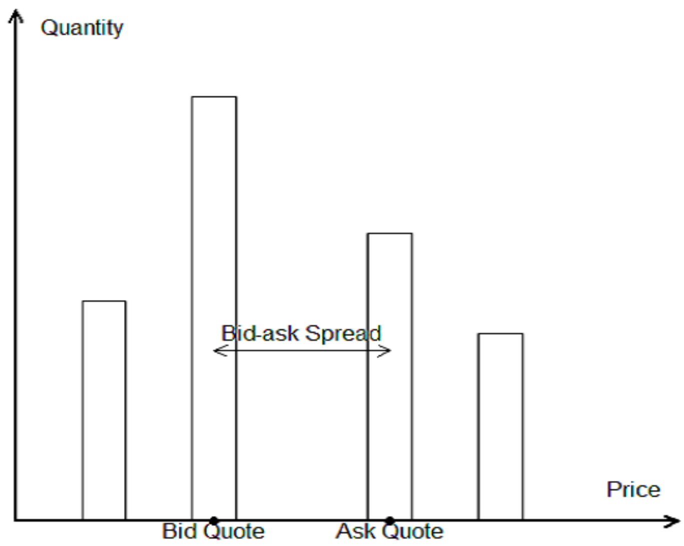

Figure 1.1: example of a simplified limit order book

announced in NASDAQ OMX Nordic market model 2.1 (2011), unlike in an oral auction whose order priorities are price precedence then submitted time, orders in the same order book stand in a queue according to the priority:

$$price > internal > displayed > time.$$

That is to say, when several orders enter an order book with the same price level, firstly the internal priority is concerned, then the displayed orders shall be prior to those hidden ones and iceberg orders, and at last the time precedence is under consideration. Further possible trading features in use are infiltrated and combined with empirical data analysis in chapter 2.

Chapter 2

Data Analysis

2.1 Data Description

As the research is based on a trader's angle of view, only those displayed data are taken and observed in the entire article. At the very beginning of research, all information of the order book for Ericsson traded at NASDAQ on 7th November in 2011 are observed through four tables. All bid, ask, and cancellation events are recorded on two excel files, one for cancellation and another one shows the bid and ask orders. The former table concludes the physical time of cancellation's existence (nanoseconds after midnight) and their quantities, whereas the latter file clearly depicts all bid and ask orders with timestamps, prices and corresponding volumes. A supplementarily table offers the execution data including the quantity, price and liquidity at specific time which capture messages about the market order. Additionally in the last dataset, nineteen queues of bid (ask) limit orders with prices and volumes emerge a portrait of the limit order book. The logic among the tables can be tracked through order sequence number and ID code for each event.

number of orders	115589	number of displayed orders	114274
number of executions	25140	number of displayed executions	11643
number of cancellations	106317	number of displayed cancellations	105598
number of events	247046	number of displayed events	231515

Table 2.1: Glimpse at Empirical Data

2.2 Empirical Properties

All the tasks in the initial stage shall be to grasp empirical properties of data and the implicit logical combination between the datasets. Here observations on elementary market trading factors, i.e, event, time, volume and price are chosen as point of penetration.

09:00-17:25 are normal continuous trading hours in NASDAQ OMX Stockholm followed by a pre-close period which lasts ca. 5 minutes. During continuous trading hours, traders can submit bid/ask orders, execute their orders or cancel them. But from the

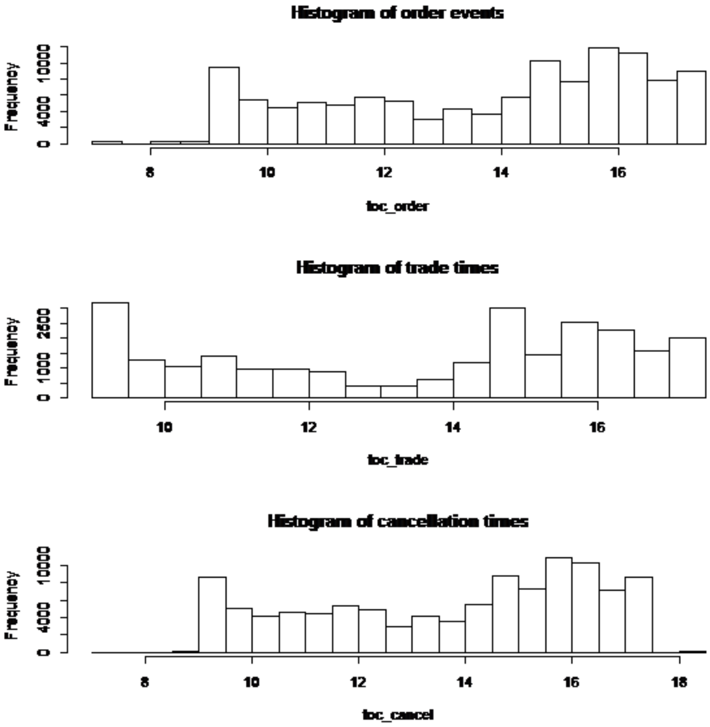

Figure 2.1: Frequencies of intraday orders, trades and cancellation events

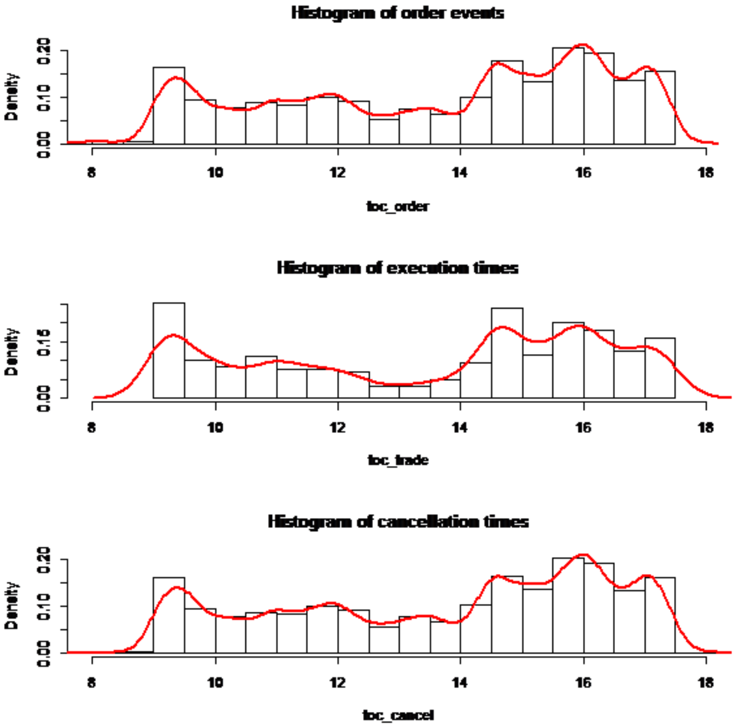

Figure 2.2: Densities of intraday orders, trades and cancellation events

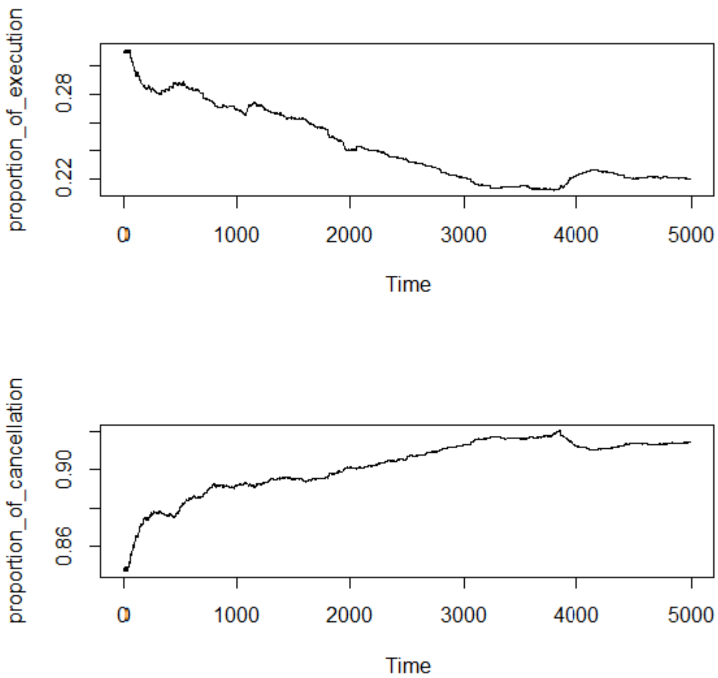

Figure 2.3: Proportion of trade and cancellation at normal trading time(09:30-16:00)

pre-open session (08:00-09:00) to post-trading period (17:30-18:00), new pegged orders are halted when it is outside the continuous trading session whereas order cancellation actions are allowed all the time. Figure 2.1 and Figure 2.2 describe intraday events in terms of frequencies and densities respectively. When focusing on the density curves in Figure 2.1, quantities of orders, trades and cancellations are similarly distributed with the well-known U-shape pattern: in addition of large volumes within circa half an hour after market opening and two hours before closure, there exists a significant dip in the mid-day of trading around 13:00.

Another possible aspect to research the quantities of each sort of events is the quotes/ executions/ cancellations' proportions and ratios between them (see Figure 2.3 and Figure 2.4). Cancellation actions are obviously more widely ranged than trades because of specific trading principles in NASDAQ. Notice that since newly pegged orders are held on the order book from the pre-open session, there exist severe changes of the events' proportions at the initial stage of continuous trading time.

Figure 2.5, on the other hand, shows the times series of volumes. It reflects sufficiently the trend of HFT: one would not like to effect the price of stock when entering the market. Therefore, traders divide the trades and orders in several parts such that the lag between events could keep decreasing progressively and volumes remaining at a very low level. All graphs imply that a large amount of operations are submitted within fairly short time intervals, which indicates an analysis of numerous data. Also they reflect a fact that quite a large number of new pegged orders exist everyday, but only less than a quarter of them are executed fully or partially whereas the others are cancelled or kept in the queue

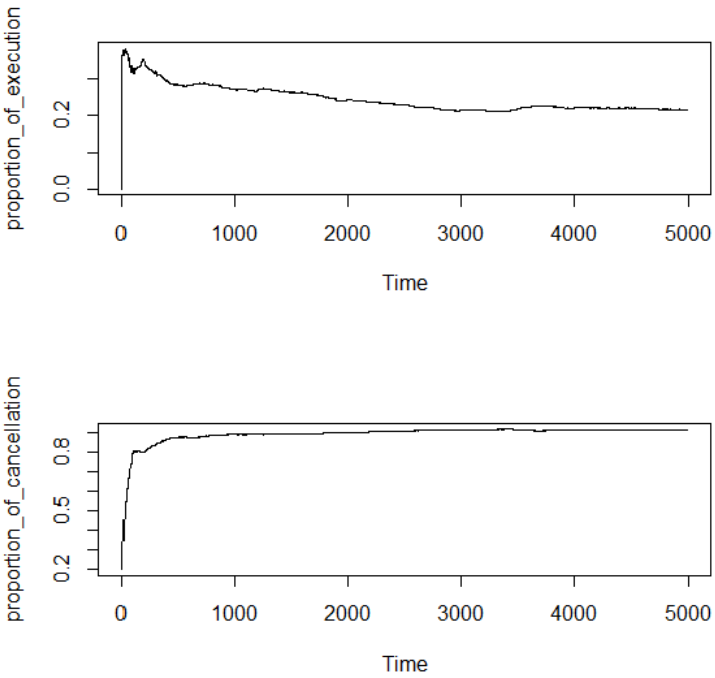

Figure 2.4: Proportion of intraday trade and cancellation (09:00-17:25)

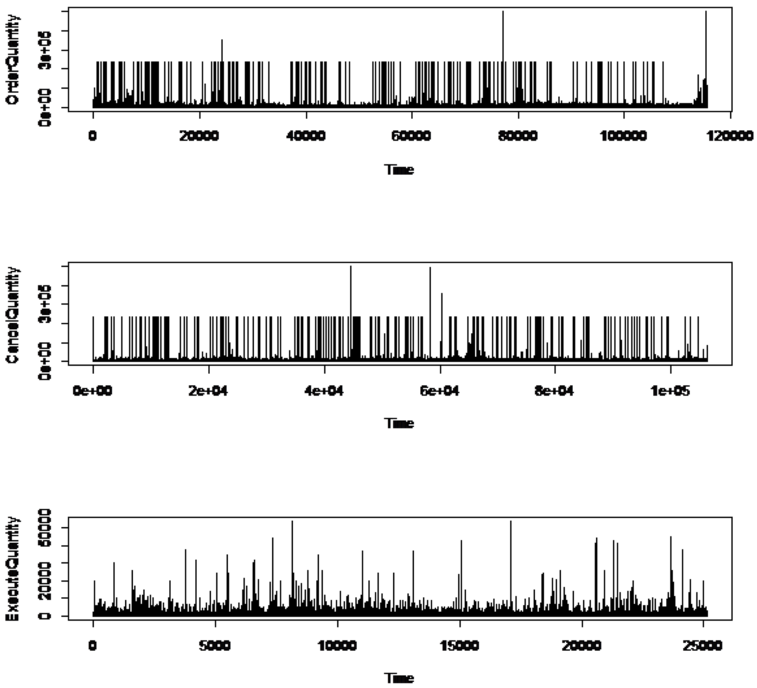

Figure 2.5: Time series of intraday orders, trades and cancellation quantities

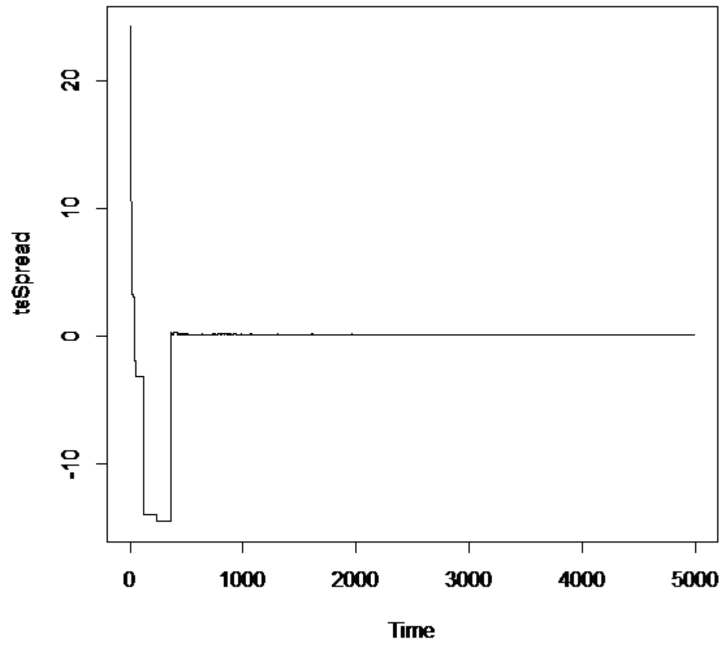

Figure 2.6: Time series of bid-ask spreads (09:00-10:00)

without any actions. The flexibility and inconstancy of market trading make it convenient for investors to arrange their portfolios, whilst they also lead the predictions more difficult to implement accurately: one should consider interactions between orders, executions and cancellations jointly in order to simulate the reality as efficiently as possible.

In aspect of the price, as is shown in Figure 2.6, the bid-ask spread is relatively large in the initial opening time. It means that the bid and ask prices do not closely attain the BBO (best-bid-offer) balance until around 09:30. Whilst analogously in Figure 2.7, BBO approach is also not held from about 17:00 any more. Price is also a significant benchmark for differentiating the market order and limit order: an order enters into the order book with a specific price, whether there exists any price variation before execution determines its property. As the simulation model (more details are specified in Chapter 3) requires market buy/sell activities but not limit ones, the property above is a necessary when attempting to institute market bid/ask information.

2.3 Data Handling

Firstly, as is talked before, some data are unobservable, which leads to the result that 15531 out of 247046 events should be omitted (see Table 2.1). Looking back at Figure 2.7 secondly, the bid-ask spread is not stable (BBO is unattainable) during the primary and final period of continuous trading session. Thus, it is highly suggested to analyze only the data occurred from 09:30 to 17:00, enforcing the number of observations to reduce

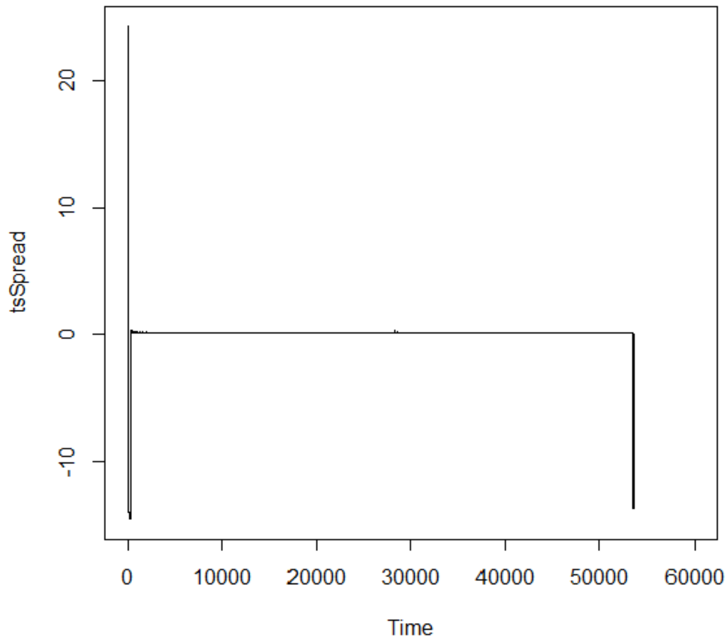

Figure 2.7: Time series of intraday bid-ask spreads

once more. In the end, one can record all timestamps with corresponding event type on new files separately and stare straightly on the newly created tables, only focusing on the event arrival times simulation. When doing software programming, one should exploit the inner logic between original documents here due to unknown of some specific event properties: event type, side, visibility, and so on. After completing these steps, it is enough to implement Hawkes process.

Chapter 3

Model Framework

3.1 Intensity Process

The intensity function plays a fairly important role in the financial point process theory because of its attainability to corresponding theoretical likelihood function. Existence of analytical likelihood function makes the likelihood-based statistical inference applicable to computational programming. Therefore, before directly modelling financial point processes, a concrete mathematical introduction to several fundamental statistical concepts is a necessary. One can look through Hautsch (2004, 2011), Karr (1986) and some introductory readings about point processes, but for instance, only chief and highly required concepts are introduced.

Definition 3.1. (Point Process)(Hautsch) Let t denote physical (calendar) time and let $\{t_i\}$ $i \in \{1, 2, 3, ...\}$ be a random sequence of event arrival times. Then, the sequence $\{t_i\}$ is called a point process on $[0, \infty)$. Specifically, a chronologically ordered time sequence is called a simple point process.

Inference theory tells that the order statistic (see Figure 3.1) is always a sufficient statistic for unknown parameters given the sample X with its corresponding conditional distribution. Intuitively speaking, the simple point process can always keep all important information in the initial point process. In this article, for instance, only simple point processes are taken into consideration because its convenience of chronological observation and support from inference theory.

More generally, a K-dimensional point process further indicates a process that simultaneously observes K different types of event arrival times which are denoted by $\{t_i^k\}$, $i \in \{1, 2, ..., n_k\}$ $k = 1, 2, ..., K$. Here n_k means the total number of observations that belongs to the k-th type.

In probability theory, point processes are defined in a more rigorous way. We start the concepts with intuitive expressions first, then a more rigorous depiction of point process based on probability theory is mentioned in Section 3.2.

Definition 3.2. (Counting Process)(Hautsch) The process $N(t)$ with $N(t) := \sum_{i\geq 1} \mathbf{1}_{\{t_i \leq t\}}$ is a right-continuous counting process associated with $\{t_i\}$.

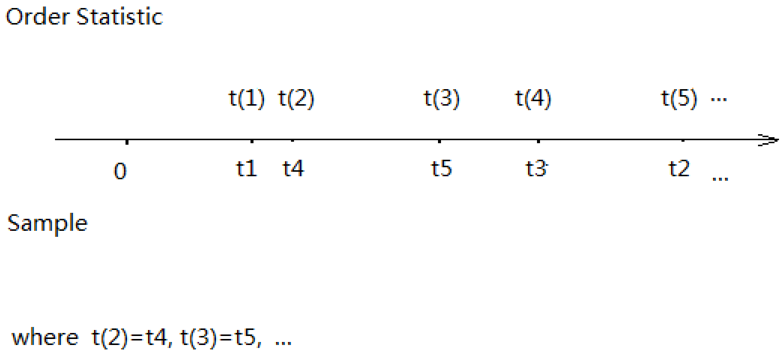

Figure 3.1: Example of order statistic

Definition 3.3. (Regular Point Process (RPP)) A counting process $N(t)$ is called a regular point process (RPP) when satisfying the probability system:

$$\begin{cases} \Pr\{N(t+\Delta) - N(t) = 1 \mid N(s) \ (s \le t)\} & = \Lambda(t) \cdot \Delta + \mathcal{O}(\Delta) \\ \Pr\{N(t+\Delta) - N(t) > 1 \mid N(s) \ (s \le t)\} & = \mathcal{O}(\Delta) \\ \Pr\{N(t) = \infty\} = 0 & \forall t, \end{cases} \tag{3.1}$$

where

$$\lim_{\Delta \downarrow 0} \frac{\mathcal{O}(\Delta)}{\Delta} = 0, \qquad \forall \Delta$$

Definition 3.2 profiles the counting process $N(t)$ as a step function with magnitude one that reports the number of events happened no later than time t. It is also important to mention about the corresponding left-continuous counting process named after $\check{N}(t)$ with $\check{N}(t) := \sum_{i \ge 1} \mathbf{1}_{\{t_i < t\}}$. Clearly it is a step function recording the number of events before time t. On the other hand, the probabilistic expression of Definition 3.3 makes people prefer observing RPPs than Counting processes when applying martingale method into statistical inference (see Section 3.2).

Definition 3.4. (Intensity Process) The $\mathcal{F}_t$-intensity process $\lambda(t; \mathcal{F}_t)$ is a left-continuous, positive-valued process defined as:

$$\lambda(t; \mathcal{F}_t) \approx \lambda(t+) := \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \mathbb{E}[N(t+\Delta) - N(t) | \mathcal{F}_t], \quad \lambda(t+) > 0, \forall t \ge 0 \tag{3.2}$$

where $\lambda(t+) := \lim_{\Delta \downarrow 0} \lambda(t+\Delta; \mathcal{F}_t)$ and $N(t)$ is a regular point process (RPP) based on $\mathcal{F}_t$.

Notice that $\mathcal{F}_t$ is the σ-field generated by N up to time t and can be seemed as all known historical information that determines a simple point process from the initial stage up to time t. More explanation of $\mathcal{F}_t$ is clarified in Section 3.2.

According to Equation 3.1, only a finite number of recurrences are allowed in a specific time interval and

$$\begin{aligned}\Pr\{N(t+\Delta)-N(t)=0\,|\,\mathcal{F}_t\}+\Pr\{N(t+\Delta)-N(t)=1\,|\,\mathcal{F}_t\}\\+\Pr\{N(t+\Delta)-N(t)>1\,|\,\mathcal{F}_t\}=1\end{aligned}\tag{3.3}$$

$$\Rightarrow \Pr\{N(t+\Delta)-N(t)=1\,|\,\mathcal{F}_t\}=1-\mathcal{O}\{t\}-\Pr\{N(t+\Delta)-N(t)=0\,|\,\mathcal{F}_t\}\tag{3.4}$$

Therefore, the intensity function can be transformed into a probabilistic function by exploiting the definition of expectation:

$$\begin{aligned}\lambda(t;\mathcal{F}_t)&=\lim_{\Delta\downarrow 0}\frac{1}{\Delta}\mathbb{E}\left[N(t+\Delta)-N(t)\,|\mathcal{F}_t\right]\\&=1\cdot\lim_{\Delta\downarrow 0}\frac{1}{\Delta}\Pr\{N(t+\Delta)-N(t)=1|\mathcal{F}_t\}\\&\quad+0\cdot\lim_{\Delta\downarrow 0}\frac{1}{\Delta}\Pr\{N(t+\Delta)-N(t)=0|\mathcal{F}_t\}\\&\quad+\cdots+(N(t+\Delta)-N(t)=i)\cdot\lim_{\Delta\downarrow 0}\frac{1}{\Delta}\Pr\{N(t+\Delta)-N(t)=i|\mathcal{F}_t\}+\cdots\\&=\lim_{\Delta\downarrow 0}\frac{1}{\Delta}\Pr\{N(t+\Delta)-N(t)=1|\mathcal{F}_t\}+(N(t+\Delta)-N(t))\\&\quad\cdot\lim_{\Delta\downarrow 0}\frac{1}{\Delta}\mathcal{O}(\Delta)\quad\left(N(t+\Delta)-N(t)<\infty;\quad\lim_{\Delta\downarrow 0}\frac{\mathcal{O}(\Delta)}{\Delta}=0,\forall\Delta\right)\\&=\lim_{\Delta\downarrow 0}\frac{1}{\Delta}\Pr\{N(t+\Delta)-N(t)=1|\mathcal{F}_t\}\end{aligned}\tag{3.5}$$

Equation 3.5 indicates that the intensity process function reflects the rate at which the event occurs instantaneous at marginal moment of t according to all the information summarized until t. By classical transition analysis, one always investigates the duration data instead of the initial point process. It is not only because of features of duration process easier to catch, but also since it is possible to generate point process with relevant duration process. Here the duration process $\{x_i\}$ $i\in\{1,2,...,n\}$ is defined as

$$x_i:=\begin{cases}t_i-t_{i-1}, & i=2,3,...,n\\ t_1, & i=1\end{cases}.$$

On the other hand, by Equation 3.2, 3.2 and 3.5, an intensity process is able to result in the duration process, and furthermore, the point process. The dynamic models based such as on intensity process are continuously distributed whilst some other duration approach techniques are discrete, advantaging the intensity process because of its smoother and more fluid characteristics. Another reason why greater amounts of researchers concentrate on intensity functions is that its likelihood function can be mathematically captured.

In this way, the cooperation of statistical inference theory, computational modelling and empirical data could be associated well. That is to say, it could be more convenient to implement likelihood-based inference numerically. Further general mathematical properties of intensity process are manifested in the following two sections.

For one-variate RPP, Definition 3.4 and Equation 3.5 describes the intensity process. An interesting and important study is to look through the combination between mono- and multi-dimensional RPP.

Let $\lambda_{N_t}(t;\mathcal{F}_t)$ be intensity of a K-dimensional ($K \geq 2$) RPP which is superposed by K independent one-dimensional RPPs. Denote that each one-dimensional RPP has intensity process $\lambda^i_{N_t}(t;\mathcal{F}^i_t)$, $i = 1, 2, ..., K$. Here the field $\left(\mathcal{F}^1_t, \mathcal{F}^2_t, ..., \mathcal{F}^K_t\right)$ is Borel (see Section 3.2), and $\mathcal{F}_t \subseteq \left(\mathcal{F}^1_t, \mathcal{F}^2_t, ..., \mathcal{F}^K_t\right)$ because $\lambda_{N_t}(t;\mathcal{F}_t)$ is generated by $\lambda^i_{N_t}(t;\mathcal{F}^i_t)$, i.e., each event in $\lambda_{N_t}(t;\mathcal{F}_t)$ is decomposed and noted on specific mono-variate RPPs which is a subset of $\left(\mathcal{F}^1_t, \mathcal{F}^2_t, ..., \mathcal{F}^K_t\right)$. Hence, by Equation 3.2, 3.5 and interchange between expectation and probability

$$\begin{aligned}
\lambda_{N_t}(t;\mathcal{F}_t) &= \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\mathbb{E}\left[N(t+\Delta) - N(t)\,|\mathcal{F}_t\right] \\
&= \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\Pr\left\{N(t+\Delta) - N(t) = 1|\mathcal{F}_t\right\} \\
&= \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\left(1 - \Pr\left\{N(t+\Delta) = N(t)\,|\mathcal{F}_t\right\} - \mathcal{O}(\Delta)\right) \\
&= \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\left(1 - \Pr\left\{N(t+\Delta) = N(t)\,|\mathcal{F}_t\right\}\right)
\end{aligned} \tag{3.6}$$

New task emerges: to calculate the probability of $\Pr\{N(t+\Delta) = N(t)\,|\mathcal{F}_t\}$. Notice that for indicator function

$$\mathbf{1}\left\{N(t+\Delta) = N(t)\right\} = \begin{cases} 1 & N(t+\Delta) = N(t) \\ 0 & otherwise \end{cases},$$

we have

$$\begin{aligned}
\mathbb{E}^{\mathcal{F}_t}\left[\mathbf{1}\left\{N(t+\Delta) = N(t)\right\}\right] &= 1\cdot\Pr\left\{N(t+\Delta) = N(t)\,|\mathcal{F}_t\right\} + 0\cdot\Pr\left\{N(t+\Delta) \neq N(t)\,|\mathcal{F}_t\right\} \\
&= \Pr\left\{N(t+\Delta) = N(t)\,|\mathcal{F}_t\right\}.
\end{aligned} \tag{3.7}$$

Hence, using $\mathcal{F}_t \subseteq \left(\mathcal{F}^1_t, \mathcal{F}^2_t, ..., \mathcal{F}^K_t\right)$, the independence between mono-variate RPPs and the properties

$$\mathbb{E}^{\mathcal{F}_u}[x] = \mathbb{E}^{\mathcal{F}_u}\left[\mathbb{E}^{\mathcal{F}_s}[x]\right], \qquad \mathcal{F}_u \subseteq \mathcal{F}_s;$$

$$c + \mathbb{E}^{\mathcal{F}_u}[x] = \mathbb{E}^{\mathcal{F}_u}[c + x], \qquad c\ is\ const.,$$

one gets aware of

$$\begin{aligned}
\Pr\{N(t+\Delta)=N(t)\,|\mathcal{F}_t\} &= \mathbb{E}^{\mathcal{F}_t}\left[\mathbf{1}\{N(t+\Delta)=N(t)\}\right] \\
&= \mathbb{E}^{\mathcal{F}_t}\left[\mathbb{E}^{\mathcal{F}_t^1,\mathcal{F}_t^2,\ldots\mathcal{F}_t^K}\left[\mathbf{1}\{N(t+\Delta)=N(t)\}\right]\right] \\
&= \mathbb{E}^{\mathcal{F}_t}\left[\mathbb{E}^{\mathcal{F}_t^1,\mathcal{F}_t^2,\ldots\mathcal{F}_t^K}\left[\mathbf{1}\left\{N^1(t+\Delta)=N^1(t)\right\}\right.\right. \\
&\quad \left.\left.\cdot\mathbf{1}\left\{N^2(t+\Delta)=N^2(t)\right\}\cdots\mathbf{1}\left\{N^K(t+\Delta)=N^K(t)\right\}\right]\right] \\
&= \mathbb{E}^{\mathcal{F}_t}\left[\mathbb{E}^{\mathcal{F}_t^1}\left[\mathbf{1}\left\{N^1(t+\Delta)=N^1(t)\right\}\right]\right. \\
&\quad \left.\cdot\mathbb{E}^{\mathcal{F}_t^2}\left[\mathbf{1}\left\{N^2(t+\Delta)=N^2(t)\right\}\right]\cdots\mathbb{E}^{\mathcal{F}_t^K}\left[\mathbf{1}\left\{N^K(t+\Delta)=N^K(t)\right\}\right]\right] \\
&= \mathbb{E}^{\mathcal{F}_t}\left[\Pr\left\{N^1(t+\Delta)=N^1(t)\,|\mathcal{F}_t^1\right\}\right. \\
&\quad \left.\cdot\Pr\left\{N^2(t+\Delta)=N^2(t)\,|\mathcal{F}_t^2\right\}\cdots\Pr\left\{N^K(t+\Delta)=N^K(t)\,|\mathcal{F}_t^K\right\}\right]
\end{aligned} \tag{3.8}$$

Therefore, multi-variate intensity function is

$$\begin{aligned}
\lambda_{N_t}(t;\mathcal{F}_t) &= \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\left(1-\Pr\{N(t+\Delta)=N(t)\,|\mathcal{F}_t\}\right) \\
&= \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\left(1-\mathbb{E}^{\mathcal{F}_t}\left[\Pr\left\{N^1(t+\Delta)=N^1(t)\,|\mathcal{F}_t^1\right\}\right.\right. \\
&\quad \left.\left.\cdot\Pr\left\{N^2(t+\Delta)=N^2(t)\,|\mathcal{F}_t^2\right\}\cdots\Pr\left\{N^K(t+\Delta)=N^K(t)\,|\mathcal{F}_t^K\right\}\right]\right) \\
&= \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\left(\mathbb{E}^{\mathcal{F}_t}\left[1-\Pr\left\{N^1(t+\Delta)=N^1(t)\,|\mathcal{F}_t^1\right\}\right.\right. \\
&\quad \left.\left.\cdot\Pr\left\{N^2(t+\Delta)=N^2(t)\,|\mathcal{F}_t^2\right\}\cdots\Pr\left\{N^K(t+\Delta)=N^K(t)\,|\mathcal{F}_t^K\right\}\right]\right)
\end{aligned} \tag{3.9}$$

When trying to solve this equation, firstly one can guess:

$$\begin{aligned}
(1-x_1)\cdot(1-x_2)\cdots(1-x_n) = (-1)^n\cdot&\left(\left(\prod_{i=1}^{n}x_i\right)-1\right) \\
&+(-1)^n\cdot\left(\sum_{i=1}^{n}(1-x_i)\right) \\
&+R_n(1-x_1,1-x_2,...,1-x_n),
\end{aligned} \tag{3.10}$$

where $R_n(1-x_1,1-x_2,...,1-x_n)$ is a polynomial of $1-x_1,1-x_2,...,1-x_n$ with at least order two if it is not null. Employ induction to prove the guess

· For $n=2$:

$$\begin{aligned}
(1-x_1)\cdot(1-x_2) &= 1-x_1-x_2+x_1x_2 \\
&= (-1)^2\cdot(x_1x_2-1)+(1-x_1)+(1-x_2)+0 \quad \textit{satisfied}
\end{aligned} \tag{3.11}$$

· Assume that for $n=k,\ k>2$ the conclusion is true, i.e.,

$$\begin{aligned}(1-x_1)\cdot(1-x_2)\cdots(1-x_k) &= (-1)^k\cdot(x_1x_2\cdots x_k-1)\\ &+(-1)^k\cdot\left(\sum_{i=1}^{k}(1-x_i)\right)\\ &+R_k\left(1-x_1,1-x_2,...,1-x_k\right)\end{aligned} \tag{3.12}$$

· Thus, for $n=k+1$:

$$\begin{aligned}(1-x_1)\cdot(1-x_2)\cdots(1-x_{k+1}) &= ((1-x_1)\cdot(1-x_2)\cdots(1-x_k))\cdot(1-x_{k+1})\\ &= \Big[(-1)^k\cdot(x_1x_2\cdots x_k-1)+(-1)^k\cdot\Big(\sum_{i=1}^{k}(1-x_i)\,\Big)\\ &\quad +R_k\left(1-x_1,1-x_2,...,1-x_k\right)\Big]\cdot(1-x_{k+1})\\ &= (-1)^k\cdot(x_1x_2\cdots x_k-1)\cdot(1-x_{k+1})\\ &\quad +(-1)^k\cdot\left(\sum_{i=1}^{k}(1-x_i)\right)\cdot(1-x_{k+1})\\ &\quad +R_k\left(1-x_1,1-x_2,...,1-x_k\right)\cdot(1-x_{k+1})\end{aligned} \tag{3.13}$$

The sum of second and third term is can be found as also a polynomial of at least order two. Hence, it is part of $R_{k+1}\left(1-x_1,1-x_2,...,1-x_{k+1}\right)$. One shall ignore the calculation for last two terms, note the sum of them as $H_{k+1}\left(1-x_1,1-x_2,...,1-x_{k+1}\right)$ and just focus on the first term

$$
\begin{aligned}
(1-x_1)\cdot(1-x_2)\cdots(1-x_{k+1}) &= (-1)^{k+1} + (-1)^{k+1} x_1x_2\cdots x_{k+1} + (-1)^k x_1x_2\cdots x_k \\
&\quad + (-1)^k x_{k+1} + H_k(1-x_1, 1-x_2, ..., 1-x_{k+1}) \\
&= (-1)^{k+1} + (-1)^{k+1} x_1x_2\cdots x_{k+1} \\
&\quad + \Big(-(1-x_1)\cdot(1-x_2)\cdots(1-x_k) + (-1)^k \\
&\quad - (-1)^k \cdot \left(\sum_{i=1}^{k}(1-x_i)\right) \\
&\quad + R_k(1-x_1, 1-x_2, ..., 1-x_k)\Big) + (-1)^k x_{k+1} \\
&\quad + H_k(1-x_1, 1-x_2, ..., 1-x_{k+1}) \\
&= (-1)^{k+1} + (-1)^{k+1} x_1x_2\cdots x_{k+1} + (-1)^k \\
&\quad - (-1)^k \cdot \left(\sum_{i=1}^{k}(1-x_i)\right) + (-1)^{k+1} x_{k+1} \\
&\quad + R_{k+1}(1-x_1, 1-x_2, ..., 1-x_{k+1}) \\
&= (-1)^{k+1} + (-1)^{k+1} x_1x_2\cdots x_{k+1} + (-1)^k \\
&\quad + (-1)^{k+1}\left(\sum_{i=1}^{k}(1-x_i)\right) + (-1)^k (x_{k+1}-1) + (-1)^k \\
&\quad + R_{k+1}(1-x_1, 1-x_2, ..., 1-x_{k+1}) \\
&= (-1)^{k+1}(x_1x_2\cdots x_{k+1} - 1) + (-1)^{k+1}\left(\sum_{i=1}^{k+1}(1-x_i)\right) \\
&\quad + R_{k+1}(1-x_1, 1-x_2, ..., 1-x_{k+1})
\end{aligned} \tag{3.14}
$$

Q.E.D

Martingale method requires the assumptions that

$$\mathbb{E}\left[\lambda^i\left(t;\mathcal{F}_t^i\right)\right] < \infty, \tag{3.15}$$

and that there exists a function $B^i(t;\mathcal{F}_t^i)$ with $\mathbb{E}[B^i(t;\mathcal{F}_t^i)] < \infty$ such that

$$\frac{1}{\Delta}\Pr\left\{N^i(t+\Delta) - N^i(t) = 1|\mathcal{F}_t^i\right\} \leq B^i\left(t,\mathcal{F}_t^i\right). \tag{3.16}$$

Thus, the intensity process is bounded. At the same time, according to the dominated convergence theorem, interchanges between integrations and limits are available. Further studies about these two assumptions are discussed in Section 3.2. Now apply Equation 3.14 to continue calculating the result of Equation 3.9

$$
\begin{aligned}
\lambda_{N_t}(t;\mathcal{F}_t) &= \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\left(\mathbb{E}^{\mathcal{F}_t}\left[1-\Pr\left\{N^1(t+\Delta)=N^1(t)\,|\mathcal{F}_t^1\right\}\right.\right.\\
&\quad\left.\left.\cdot\Pr\left\{N^2(t+\Delta)=N^2(t)\,|\mathcal{F}_t^2\right\}\cdots\Pr\left\{N^K(t+\Delta)=N^K(t)\,|\mathcal{F}_t^K\right\}\right]\right)\\
&= \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\left(\mathbb{E}^{\mathcal{F}_t}\left[(-1)^K\cdot(-1)\cdot\left(1-\Pr\left\{N^1(t+\Delta)=N^1(t)\,|\mathcal{F}_t^1\right\}\right)\right.\right.\\
&\quad\left(1-\Pr\left\{N^2(t+\Delta)=N^2(t)\,|\mathcal{F}_t^2\right\}\right)\cdots\left(1-\Pr\left\{N^K(t+\Delta)=N^K(t)\,|\mathcal{F}_t^K\right\}\right)\\
&\quad+(-1)^{2K}\left(\sum_{i=1}^{K}\left(1-\Pr\left\{N^i(t+\Delta)=N^i(t)\,|\mathcal{F}_t^i\right\}\right)\right)\\
&\quad+(-1)^K\cdot R_K\left(1-\Pr\left\{N^1(t+\Delta)=N^1(t)\,|\mathcal{F}_t^1\right\},...,\right.\\
&\quad\left.\left.\left.1-\Pr\left\{N^K(t+\Delta)=N^K(t)\,|\mathcal{F}_t^K\right\}\right)\right]\right)\\
&= \mathbb{E}^{\mathcal{F}_t}\left[\lim_{\Delta\downarrow 0}\Delta^{K-1}\cdot(-1)^{K+1}\cdot\prod_{i=1}^{K}\lambda^i\left(t;\mathcal{F}_t^i\right)+\sum_{i=1}^{K}\lambda^i\left(t;\mathcal{F}_t^i\right)\right.\\
&\quad\left.+\lim_{\Delta\downarrow 0}\Delta\cdot\tilde{R}_K\left(\Delta,\lambda^1\left(t;\mathcal{F}_t^1\right),\lambda^2\left(t;\mathcal{F}_t^2\right),...,\lambda^K\left(t;\mathcal{F}_t^K\right)\right)\right]\\
&= \mathbb{E}^{\mathcal{F}_t}\left[\sum_{i=1}^{K}\lambda^i\left(t;\mathcal{F}_t^i\right)\right]
\end{aligned}
\tag{3.17}
$$

The whole procedure of deriving repeatedly combines expectational/ probabilistic expressions of intensity functions with martingale specific calculation principles and assumptions. It requires a clear and flexible mind in the exchange between counting processes and two forms of intensity functions which is also a major method of treating corresponding statistical inference problematic.

Equation 3.17 indicates that the intensity function of a K-dimensional RPP can be expressed as the expectation of summation of several superposed independent mono-variate RPPs' intensity. Another conclusion from Equation 3.17 is that the class of RPPs is closed regarding superposition, i.e., each superposed RPP can be interpreted by several observed intensity functions with being conditioned on previously observed past event recurrences. This is an important conclusion since it simplifies a multi-variate intensity process research into detecting corresponding superposed one-dimensional intensity processes. Further studies about RPPs' intensity-based inference exploits this conclusion sufficiently.

3.2 Intensity-based Inference of RPP

Likelihood-based statistical inference is tackled in favour of the martingale method owing to its outstanding advantages in effective methodological calculation of expectation, variance, coefficient and potent technique for constructing asymptotic normality of the estimator. In conclusion, martingale method has a succession of rules for parametric detection and prediction analysis which makes studies of intensity-based RPP models significative in practical applications. Besides, other methods rarely can create the estimators in need resulting in martingale method the dominant tool.

Martingale method assumes that the intensity process is stochastic. In consequence, from the sample space $(\Omega, \mathcal{G})$ one can generate a probability space $(\Omega,\mathcal{G},P)$ on which the history information $\mathcal{F}_t$, K-dimensional intensity process λ_t and corresponding RPP $N(t)$ can be expressed as

- $\mathcal{F}_t = \left(\mathcal{F}_t^1, \mathcal{F}_t^2, ..., \mathcal{F}_t^K\right),\ \ \mathcal{F}_t \subseteq \mathcal{G};$
- $\lambda_t = \left(\lambda_t^1, \lambda_t^2, ..., \lambda_t^K\right)$ is bounded, non-negative and predictable;
- $N(t) = \left(\sum_n \mathbf{1}(t_n \le t, type\, 1), \sum_n \mathbf{1}(t_n \le t, type\, 2), ... \sum_n \mathbf{1}(t_n \le t, type\, K)\right).$

By martingale method, if

$$\sum_{k=1}^{K} \int_0^{t_n} \lambda(u; \mathcal{F}_u)\, du < \infty \qquad a.s \tag{3.18}$$

or moreover

$$\mathbb{E}\left[\sum_{k=1}^{K} \int_0^{t_n} \lambda(u; \mathcal{F}_u)\, du\right] < \infty \qquad a.s,$$

then the probability space above is dominated. As a conclusion, all likelihood-based statistical inference can be applied only if Equation 3.18 is satisfied. Karr (1986) has derived this conclusion in a more rigorous manner, and has proved the uniqueness and existence of the probability measure Π. Readers could browse section 5.1-5.2 for more detail.

Construction of the probability space $(\Omega,\mathcal{G},P)$ and existence of the probability measure Π ensure that point processes can be defined in this probability space.

Definition 3.5. (Point Process in Probability Theory) Let S be the smallest σ-field on probability measure Π depicting all counting points. Then a point process is a measurable map from the probability space $(\Omega,\mathcal{G},P)$ to measurable space (Π, S).

With the assumption that Equation 3.18 is established, the log-likelihood function and prediction analysis for mono-variate and multi-variate RPPs are given in the forthcoming couple of sections.

Likelihood Function

Theorem 3.6. *(Joint* pdf *of RPP) (Rubin (1972)) The joint occurrence probability density function $f(x|\Theta)$ of a regular point process $N(t)$ satisfies the formula*

$$\ln f(t|\Theta) = -\int_0^T \lambda(t|\Theta)\, dt + \int_0^T \ln \lambda(t|\Theta)\, dN(t) \tag{3.19}$$

whenever $\lambda(t_i|\mathcal{F}_{t-1}) > 0, \quad \forall i \ge 1,$ and vanishes otherwise.

Proof: By definition of differentiation, probability conditions of RPP and probabilistic form of intensity function, assum that $\Pr\{N(t+\Delta) = n|\mathcal{F}_s\}$, $\forall t > s \ge 0$, then

$$\begin{aligned}\frac{\partial \Pr\{N(t+\Delta)-N(t)=0|\mathcal{F}_s\}}{\partial t} &= \frac{\partial \Pr\{N(t+\Delta)=N(t)=n|\mathcal{F}_s\}}{\partial t} \\ &= \frac{\partial \Pr\{N(t+\Delta)=n|N(t)=n,\mathcal{F}_s\}}{\partial t}\end{aligned}$$

Thus

$$\begin{aligned}\frac{\partial \Pr\{N(t)=n|\mathcal{F}_t\}}{\partial t} &= \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\left(\Pr\{N(t+\Delta)=n|N(t)=n,\mathcal{F}_s\}\cdot\Pr\{N(t)=n|\mathcal{F}_s\}\right.\\ &\quad+\Pr\{N(t+\Delta)=n|N(t)=n-1,\mathcal{F}_s\}\cdot\Pr\{N(t)=n-1|\mathcal{F}_s\}\\ &\quad\left.+\mathcal{O}(\Delta)\right)-\lim_{\Delta\downarrow 0}\frac{1}{\Delta}\Pr\{N(t)=n|\mathcal{F}_s\}\\ &= \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\left(\left(\Pr\{N(t+\Delta)=n|N(t)=n,\mathcal{F}_s\}-1\right)\cdot\Pr\{N(t)=n|\mathcal{F}_s\}\right.\\ &\quad\left.+\Pr\{N(t+\Delta)=n|N(t)=n-1,\mathcal{F}_s\}\cdot\Pr\{N(t)=n-1|\mathcal{F}_s\}\right)\\ &= \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\left(\left(-\Pr\{N(t+\Delta)=n|N(t)=n-1,\mathcal{F}_s\}-\mathcal{O}(\Delta)\right)\right.\\ &\quad\left.\cdot\Pr\{N(t)=n|\mathcal{F}_s\}\right)+\lambda_{n-1}(t;\mathcal{F}_s)\cdot\Pr\{N(t)=n-1|\mathcal{F}_s\}\\ &= -\Pr\{N(t)=n|\mathcal{F}_s\}\cdot\lambda_n(t;\mathcal{F}_s)+\Pr\{N(t)=n-1|\mathcal{F}_s\}\cdot\lambda_{n-1}(t;\mathcal{F}_s)\end{aligned} \tag{3.20}$$

Substitute $N(s)=n$ into Equation 3.20, then the second term

$$\Pr\{N(t)=n-1|\mathcal{F}_s\}=\Pr\{N(t)=N(s)-1|\mathcal{F}_s\}=0$$

since $N(t)\geq N(s),\forall t>s$. The equation turns out to be

$$\frac{\partial \Pr\{N(t)=N(s)|\mathcal{F}_s\}}{\partial t}=-\Pr\{N(t)=N(s)|\mathcal{F}_s\}\cdot\lambda_{N(s)}(t;\mathcal{F}_s)$$

Solution of this ODE is simply given by

$$\Pr\{N(t)=N(s)|\mathcal{F}_s\}=\exp\left\{-\int_s^t\lambda_{N(s)}(u;\mathcal{F}_s)\,du\right\}. \tag{3.21}$$

On the other hand, analogous to Equation 3.35, by employing the definition of RPP, it is simple to obtain

$$\begin{aligned}\frac{\partial \Pr\{N(t)=n\}}{\partial t} &= \lim_{\Delta\downarrow 0}\frac{\Pr\{N(t+\Delta)=n\}-\Pr\{N(t)=n\}}{\Delta}\\ &= \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\Pr\{N(t+\Delta)=n|N(t)=n\}\cdot\Pr\{N(t)=n\}\\ &\quad+\lim_{\Delta\downarrow 0}\frac{1}{\Delta}\Pr\{N(t+\Delta)=n|N(t)=n-1\}\cdot\Pr\{N(t)=n-1\}\\ &\quad+\lim_{\Delta\downarrow 0}\frac{1}{\Delta}\mathcal{O}(\Delta)\cdot\left(\sum_{i=0}^{n-2}\Pr\{N(t)=i\}\right)-\lim_{\Delta\downarrow 0}\frac{1}{\Delta}\Pr\{N(t)=n\}\\ &= \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\left((\Pr\{N(t+\Delta)=n|N(t)=n\}-1)\cdot\Pr\{N(t)=n\}\right.\\ &\quad\left.+\Pr\{N(t+\Delta)=n|N(t)=n-1\}\cdot\Pr\{N(t)=n-1\}\right)\\ &= \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\left((-\Pr\{N(t+\Delta)=n|N(t)=n-1\}-\mathcal{O}(\Delta))\right.\\ &\quad\left.\cdot\Pr\{N(t)=n\}\right)+\lambda_{n-1}(t)\cdot\Pr\{N(t)=n-1\}\\ &= -\Pr\{N(t)=n\}\cdot\lambda_n(t)+\Pr\{N(t)=n-1\}\cdot\lambda_{n-1}(t),\qquad n\geq 1\end{aligned} \tag{3.22}$$

and

$$\begin{aligned}\frac{\partial \Pr\{N(t)=0\}}{\partial t} &= \lim_{\Delta\downarrow 0}\frac{\Pr\{N(t+\Delta)=0\}-\Pr\{N(t)=0\}}{\Delta}\\ &= \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\Pr\{N(t+\Delta)=0|N(t)=0\}\cdot\Pr\{N(t)=0\}-\Pr\{N(t)=0\}\\ &= \lim_{\Delta\downarrow 0}\frac{1}{\Delta}(\Pr\{N(t+\Delta)=0|N(t)=0\}-1)\cdot\Pr\{N(t)=0\}\\ &= -\lambda_0(t)\cdot\Pr\{N(t)=0\},\qquad n=0\end{aligned} \tag{3.23}$$

For instance, consider the situation with times from null until the n-th event arrival time $s=t_n$, i.e., $0=t_0<t_1<...<t_{n-1}<t_n=s<t$. Letting X_i marks the i-th event, by applying Equation 3.21, then for $n\geq 1$ the *cdf* can be expressed as

$$\begin{aligned}\mathbf{F}_{X_{n+1}}(t|t_1,t_2,...,t_n) &= 1-\Pr\{X_{n+1}>t|X_1=t_1,X_2=t_2,...,X_n=t_n\}\\ &= 1-\Pr\{N(t)=n|X_1=t_1,X_2=t_2,...,X_n=t_n\}\\ &= 1-\exp\left\{-\int_{t_n}^{t}\lambda_n(u;\mathcal{F}_{t_n})\,du\right\}\end{aligned} \tag{3.24}$$

and when $n=0$

$$\begin{aligned} \mathbf{F}_{X_1}(t) &= 1 - \Pr\{X_1 > t\} \\ &= 1 - \Pr\{N(t) = 0\} \\ &= 1 - \exp\left\{-\int_0^t \lambda_0(u)\,du\right\} \end{aligned} \tag{3.25}$$

Take derivatives separately and obtain the conditional *pdf*

$$\begin{aligned} f_{X_{n+1}}(t|t_1, t_2, ..., t_n) &= \frac{\partial}{\partial t}\mathbf{F}_{X_{n+1}}(t|t_1, t_2, ..., t_n) \\ &= \exp\left\{-\int_{t_n}^t \lambda_n(u; \mathcal{F}_{t_n})\,du\right\} \cdot \lambda_n(t; \mathcal{F}_{t_n}) \end{aligned} \tag{3.26}$$

and

$$\begin{aligned} f_{X_1}(t) &= \frac{\partial}{\partial t}\mathbf{F}_{X_1}(t) \\ &= \exp\left\{-\int_0^t \lambda_1(u)\,du\right\} \cdot \lambda_1(t) \end{aligned} \tag{3.27}$$

Therefore, the joint *pdf* for the event recurrence is

$$\begin{aligned} f(t_1, t_2, ..., t_n) &= f_{X_n}(t_n|t_1, t_2, ..., t_{n-1}) \cdot f(t_1, t_2, ..., t_{n-1}) \\ &= f_{X_n}(t_n|t_1, t_2, ..., t_{n-1}) \cdot f_{X_{n-1}}(t_{n-1}|t_1, t_2, ..., t_{n-2}) \cdot f(t_1, t_2, ..., t_{n-2}) \\ &= ... \\ &= f_{X_n}(t_n|t_1, t_2, ..., t_{n-1}) \cdot f_{X_{n-1}}(t_{n-1}|t_1, t_2, ..., t_{n-2}) \cdots f_{X_1}(t_1) \\ &= \exp\left\{-\int_0^{t_1} \lambda_1(u)\,du\right\} \cdot \lambda_1(t_1) \cdots \exp\left\{-\int_{t_{n-2}}^{t_{n-1}} \lambda_{n-2}\left(u; \mathcal{F}_{t_{n-2}}\right) du\right\} \\ &\quad \cdot \lambda_{n-2}\left(t_{n-1}; \mathcal{F}_{t_{n-2}}\right) \cdot \exp\left\{-\int_{t_{n-1}}^{t_n} \lambda_{n-1}\left(u; \mathcal{F}_{t_{n-1}}\right) du\right\} \cdot \lambda_{n-1}\left(t_n; \mathcal{F}_{t_{n-1}}\right) \\ &= \prod_{i=1}^{n} \lambda_{i-1}(t_i; \mathcal{F}_{i-1}) \cdot \exp\left\{-\sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \lambda_{j-1}\left(u; \mathcal{F}_{t_{j-1}}\right) du\right\} \end{aligned} \tag{3.28}$$

From Equation 3.21, 3.23, 3.28 and definition of conditional probability, furthermore, one can calculate the joint *pdf* of an RPP regarding the interval [0,T] by institute N_T into n and $s = t_{N_T}$

$$
\begin{aligned}
f\left(N_T, t_1, t_2, ..., t_{N_T}\right) &= f\left(N_T | t_1, t_2, ..., t_{N_T}\right) \cdot f\left(t_1, t_2, ..., t_{N_T}\right) \\
&= f\left(N\left(T\right) = N_T | t_1, t_2, ..., t_{N_T}\right) \cdot f\left(t_1, t_2, ..., t_{N_T}\right) \\
&= \exp\left\{-\int_{t_{N_T}}^{T} \lambda_{N_T}\left(u; \mathcal{F}_{N_T}\right) du\right\} \cdot f\left(t_1, t_2, ..., t_{N_T}\right) \\
&= \exp\left\{-\int_{t_{N_T}}^{T} \lambda_{N_T}\left(u; \mathcal{F}_{N_T}\right) du\right\} \cdot \prod_{i=1}^{N_T} \lambda_{i-1}\left(t_i; \mathcal{F}_{i-1}\right) \\
&\quad \cdot \exp\left\{-\sum_{j=1}^{N_T} \int_{t_{j-1}}^{t_j} \lambda_{j-1}\left(u; \mathcal{F}_{t_{j-1}}\right) du\right\} \\
&= \prod_{i=1}^{N_T} \lambda_{i-1}\left(t_i; \mathcal{F}_{i-1}\right) \cdot \exp\left\{-\int_{t_{N_T}}^{T} \lambda_{N_T}\left(u; \mathcal{F}_{N_T}\right) du \right. \\
&\quad \left. -\sum_{j=1}^{N_T} \int_{t_{j-1}}^{t_j} \lambda_{j-1}\left(u; \mathcal{F}_{t_{j-1}}\right) du\right\}, \qquad n \geq 1
\end{aligned}
\tag{3.29}
$$

and

$$
f\left(N_T\right) = f\left(N\left(T\right) = 0\right) = \exp\left\{-\int_0^T \lambda_0\left(u\right) du\right\}, \qquad n = 0 \tag{3.30}
$$

Take logarithm on both sides and get

$$
\begin{aligned}
\ln f\left(N_T, t_1, t_2, ..., t_{N_T}\right) = \sum_{i=1}^{N_T} \ln \lambda_{i-1}\left(t_i; \mathcal{F}_{i-1}\right) - \left\{\int_{t_{N_T}}^{T} \lambda_{N_T}\left(u; \mathcal{F}_{N_T}\right) du \right. \\
\left. + \sum_{j=1}^{N_T} \int_{t_{j-1}}^{t_j} \lambda_{j-1}\left(u; \mathcal{F}_{t_{j-1}}\right) du\right\}, \qquad n \geq 1
\end{aligned}
\tag{3.31}
$$

and

$$
\ln f\left(N_T\right) = -\int_0^T \lambda_0\left(u\right) du, \qquad n = 0 \tag{3.32}
$$

Presently mathematical calculations are enough in order to retrieve the result of 3.6. Martingale method guarantees that the intensity process $\lambda_{N_T}\left(t; \mathcal{F}_{t_{N_T}}\right)$ is $\mathcal{F}_t$-measurable over $t \in [0, T]$ and enables it be treated as stochastic in form of

$$
\lambda_{N_T}\left(t; \mathcal{F}_{t_{N_T}}\right) = \begin{cases} \lambda_k\left(t, t_1, t_2, ..., t_k\right) & N\left(t\right) = k, \ i.e., \ t_k \leq t < t_{k+1}; n = 1, 2, ..., N_T \\ \lambda_0\left(t\right) & 0 \leq t < t_1 \end{cases}
$$

Through transforming the sum function into integration and exploiting the definition of counting process, Equation 3.29 and 3.30 can be jointly simplified into the equation in Theorem 3.6.

Q.E.D

Definition of likelihood function announces the joint *pdf* of the distribution in discussion as the likelihood function. The difference between them is just the likelihood function ranges as parameters change with observed data known, but the joint *pdf* is inverse:

$$\mathcal{L}(\theta|x) = f(x|\theta)$$

Therefore, conclusion in Theorem 3.6 supplies also the log-likelihood function of RPP which is the RHS of Equation 3.19.

A mono-variate RPP uncontroverted has a likelihood function as

$$\ln \mathcal{L}(\Theta|t_1, t_2, ..., t_n) = -\int_0^T \lambda(\Theta|t_1, t_2, ..., t_n)\, dt + \int_0^T \ln \lambda(\Theta|t_1, t_2, ..., t_n)\, dN(t), \quad (3.33)$$

where $t_1, t_2, ..., t_n$ are observations, i.e., all event arrival times. Θ is a set of parameters that depicts corresponding distribution of RPP under specific intensity model.

For a multi-variate RPP, for example, a K-dimensional RPP, its likelihood function could be rewritten into a simplified mode. The function is composed of K one-dimensional intensities, applying the simplification of the multi-variate intensity function shown in Section 3.1. By Equation 3.17, with all assumptions satisfied, the likelihood function of a K-dimensional RPP is

$$\begin{aligned}
\ln \mathcal{L}(\Theta|t_1, t_2, ..., t_n) &= -\int_0^T \lambda(\Theta|t_1, t_2, ..., t_n)\, dt + \int_0^T \ln \lambda(\Theta|t_1, t_2, ..., t_n)\, dN(t) \\
&= -\int_0^T \left(\mathbb{E}^{\mathcal{F}_t}\left[\sum_{i=1}^K \lambda^i\left(\Theta|t_1^i, t_2^i, ..., t_{n_i}^i\right)\right]\right) dt \\
&\quad + \int_0^T \ln\left(\mathbb{E}^{\mathcal{F}_t}\left[\sum_{i=1}^K \lambda^i\left(\Theta|t_1^i, t_2^i, ..., t_{n_i}^i\right)\right]\right) dN(t) \\
&= -\int_0^T \left(\sum_{i=1}^K \lambda^i\left(\Theta|t_1^i, t_2^i, ..., t_{n_i}^i\right)\right) dt \\
&\quad + \int_0^T \ln\left(\sum_{i=1}^K \lambda^i\left(\Theta|t_1^i, t_2^i, ..., t_{n_i}^i\right)\right) dN(t) \\
&= -\sum_{i=1}^K \int_0^T \lambda^i\left(\Theta|t_1^i, t_2^i, ..., t_{n_i}^i\right) dt + \sum_{i=1}^K \int_0^T \ln \lambda^i\left(\Theta|t_1^i, t_2^i, ..., t_{n_i}^i\right) dN(t),
\end{aligned} \quad (3.34)$$

where all event arrival times until $\left(t_{n_1}^1, t_{n_2}^2, ..., t_{n_K}^K\right)$ generate exactly the Borel field $\mathcal{F}_t$. With all assumptions for conclusion 3.17 satisfied, the superposed intensities are bounded on the specific probability space according to martingale method. Therefore, the expectation symbols are omitted in the third step.

By the general form of likelihood function for RPP based on its inherent intensity process, one can derive concrete likelihood functions regarding different intensity-based models. Furthermore, maximum likelihood estimator (MLE) is possible to implement numerically. More details and results could be found in Chapter 4.

Prediction Formula

Predictability

Consider the assumptions

$$\mathbb{E}\left[\lambda\left(t;\mathcal{F}_t\right)\right]<\infty, \tag{3.35}$$

and that there exists a function $B\left(t;\mathcal{F}_t\right)$ with $\mathbb{E}\left[B\left(t;\mathcal{F}_t\right)\right]<\infty$ such that

$$\frac{1}{\Delta}\Pr\left\{N\left(t+\Delta\right)-N\left(t\right)=1|\mathcal{F}_t\right\}\leq B\left(t,\mathcal{F}_t\right), \tag{3.36}$$

That is to say, $\frac{1}{\Delta}\Pr\left\{N\left(t+\Delta\right)-N\left(t\right)=1|\mathcal{F}_t\right\}$ is bounded. Thus, by dominated convergence theorem and apply the expectational expression of intensity function, when $s\leq t$

$$\begin{aligned}
\lambda_{N(t)}\left(t;\mathcal{F}_s\right)&=\lim_{\Delta\downarrow 0}\frac{1}{\Delta}\mathbb{E}\left[N\left(t+\Delta\right)-N\left(t\right)|N\left(t\right),\mathcal{F}_s\right]\\
&=\lim_{\Delta\downarrow 0}\frac{1}{\Delta}\mathbb{E}\left[\mathbb{E}\left[N\left(t+\Delta\right)-N\left(t\right)|\mathcal{F}_t\right]|N\left(t\right),\mathcal{F}_s\right]\\
&=\mathbb{E}\left[\lim_{\Delta\downarrow 0}\frac{1}{\Delta}\mathbb{E}\left[N\left(t+\Delta\right)-N\left(t\right)|\mathcal{F}_t\right]|N\left(t\right),\mathcal{F}_s\right]\\
&=\mathbb{E}^{\mathcal{F}_s,N(t)}\left[\lambda_{N(t)}\left(t;\mathcal{F}_t\right)\right]
\end{aligned} \tag{3.37}$$

Consequently, $\lambda_{N(t)}\left(t;\mathcal{F}_s\right)$ tends to $\lambda_{N(t)}\left(t;\mathcal{F}_t\right)$ with probability one when s enlarges towards t.

Now one turns to the RPP. Martingale adapted analysis describes an RPP's property with the definition of martingale, supermartingale and submartingale as

Definition 3.7. A stochastic process N_s, $s\geq 0$ is called a martingale with respect to the σ-field $\{\mathcal{F}_s,\ s\geq 0\}$ if:

· $\forall\ s\geq 0$, N_s is $\mathcal{F}_s$-measurable;
· $\forall\ s\geq 0$, $\mathbb{E}\left[|N\left(s\right)|\right]<\infty$;
· $\forall\ 0\leq s\leq t<\infty$,

$$\mathbb{E}\left[N\left(t\right)|\mathcal{F}_s\right]=N\left(s\right). \tag{3.38}$$

Furthermore, the stochastic process is called a supermartingale or submartingale if Equation 3.38 changes into inequality $\mathbb{E}\left[N\left(t\right)|\mathcal{F}_s\right]\geq N\left(s\right)$ or $\mathbb{E}\left[N\left(t\right)|\mathcal{F}_s\right]\leq N\left(s\right)$.

According to Assumptions 3.35 and 3.36, the first two conditions are fulfilled. Thus, it is clearly that RPP is a supermartingale since for all $0 < s \leq t < \infty$

$$\mathbb{E}\left[N\left(t\right)|\mathcal{F}_s\right] \geq N\left(s\right) \quad a.s., \tag{3.39}$$

which is based on property of a counting process: continuously non-decreasing. Also it is known that the RPP N_t is càdlàg (continue à droite, limitée à gauche) which means the RPP is continuous to direction of right and limited at the left hand side. Once these two conditions are fulfilled, the famous Doob-Meyer Decomposition theorem is applicable.

Theorem 3.8. *(Doob-Meyer Decomposition Theorem) Let N_t be a càdlàg supermartingale with respect to $\mathcal{F}_t$ with $N_0 = 0$. Then there exists a unique, increasing, predictable process A with $A_0 = 0$ such that $N_t = Z_t + A_t$, where Z_t is a martingale.*

According to Theorem 3.8, N_t can be decomposed as sum of a martingale process and a predictable process. Theoretically speaking, the predictability is capable as long as one detects the predictable process A throughly which is defined in next section.

Prediction

To implement the prediction, one could link the relationship between duration and intensity process, knowledge of Markov chains and its physical meaning. Conclusively follows a theorem in concrete

Theorem 3.9. *(Hautsch) Let $\tau\left(t\right)$ be the stopping time of a (transformed) point process $N\left(t\right)$ satisfying*

$$\int_0^{\tau(t)} \lambda\left(t;\mathcal{F}_s\right) ds = t, \quad \forall t, \tag{3.40}$$

then, the correponding durations $t_i - t_{i-1}$ are given by

$$t_i - t_{i-1} = \Lambda\left(t_{i-1}, t_i\right) = \int_{t_{i-1}}^{t_i} \lambda\left(s;\mathcal{F}_s\right) ds. \tag{3.41}$$

On the other hand, the durations are i.i.d. standard exponentially distributed random variates. Thus,

$$\Lambda\left(t_{i-1}, t_i\right) \sim i.i.d. \quad Exp\left(1\right) \tag{3.42}$$

Proof: In this case, t_i and t_{i-1} are exactly successive stopping-times with mathematical expressions

$$\begin{cases} N\left(t_i\right) - N\left(t_{i-1}\right) & = 1 \\ \tau\left(t_i\right) = t_i \\ \tau\left(t_{i-1}\right) = t_{i-1} \end{cases}.$$

t_i and t_{i-1} by definition are

$$\int_0^{t_i} \lambda\left(t_i;\mathcal{F}_s\right)ds = t_i$$

and

$$\int_0^{t_{i-1}} \lambda\left(t_{i-1};\mathcal{F}_s\right)ds = t_{i-1}.$$

Obviously in the time interval $t \in (t_{i-1}, t_i)$, there exists no new event. Notice that during this period of time, history information updates automatically with time t

$$t_i - t_{i-1} = \int_0^{t_i} \lambda\left(t;\mathcal{F}_s\right)ds - \int_0^{t_{i-1}} \lambda\left(t;\mathcal{F}_s\right)ds = \int_{t_{i-1}}^{t_i} \lambda\left(s;\mathcal{F}_s\right)ds \tag{3.43}$$

Now think about the second part of the theorem straightforwardly: the waiting time from t_{i-1} to t_i could be considered to follow the standard exponential distribution because during this period, no event arrives which follows a birth process with parameter one.

Q.E.D

Theorem 3.9 depicts strongly not only the relationship between duration and intensity process, but also the method of accomplishing prediction: assume people are informed of all events until t_n, consequently, the next event arrival time shall be solution to the function which has only one unknown t_{n+1}

$$\epsilon_{n+1} = \Lambda\left(t_n, t_{n+1}\right) = \int_{t_n}^{t_{n+1}} \lambda\left(s;\mathcal{F}_s\right)ds, \tag{3.44}$$

where ϵ_{n+1} represents an i.i.d. exponentially distributed random variable.

3.3 Point Process Modelling

Glance at Various Methods

There exist plenty methods of point process modelling whose mainstream ideas are generally devided into four types:

· Intensity-based modelling
· Hazard-based modelling
· Counting-based modelling
· Duration-based modelling.

All these methods can be employed into a dynamic framework and each of them foucuses on a specification of RPP. Each specified process has strong combinations with one another, in the meanwhile, it has distinctive advantage over the others.

Intensity-based RPP models, for example, their largest superiority is that intensity process is denoted as continuously distributed. They are constructed based on an almost surely fluid expanding field $\mathcal{F}_t$ which is also known as the information set and leads an RPP modelling in continuous time.

Similar to intensity process, hazard functions are defined as rate of waiting times' probabilities, but the difference is obvious: the hazard rate is only evaluated at the end of each duration and depends on historical process conclusions.

Duration models and counting model seem to be more intuitive and understandable which separately depicts processes which are generated by lags between each two consecutive event arrivals or total amount of events with respect to a fixed time interval.

Presently it is hard to judge which one is the best when being exploited into empirical studies. The reason is that there exists no convincing theoretical system yet that can persuade people about an absolute dominant of a method mentioned above. To another aspect, when simulating empirical data, it is not capable to dig up information with distinct signal from large quantities of data such that we can incorporate a best specification in order to judge the best fit directly before simulation. If different sorts of empirical model's extent of approach is of interest, one has to implement and evaluate each model with the same data, then to compare and choose the most satisfactory one with significantly financial interpretation.

Hawkes Process

This thesis introduces and discusses only one intensity-based RPP model throughly, namely the Hawkes process. Alan G. Hawkes (1971) instructs this self-exciting intensity process which has a characteristic that all previous backward recurrence timestamps fuel the intensity process to evolve automatically all the way.

Definition 3.10. (Hawkes Process) Under the circumstance of a point process satisfying

$$\begin{cases} \Pr\left\{\Delta N(t) = 1 | N(s), (s \le t)\right\} = \Lambda(t) \cdot \Delta t + \mathcal{O}(\Delta t) \\ \Pr\left\{\Delta N(t) > 1 | N(s), (s \le t)\right\} = \mathcal{O}(\Delta t). \end{cases} \tag{3.45}$$

Assume that $\Lambda(t)$ is a stationary stochastic process and the full set of $N(s), (s \le t)$ here in Equation 3.45 can be interpreted as $\{N(s), (s \le t) : \Lambda(s), s \in (-\infty, \infty)\}$. With all these assumptions being fulfilled, a Hawkes process is a generalized point process model whose intensity function is

$$\Lambda(t) = \mu + \int_{-\infty}^{t} g(t-u)\, dN(u) \tag{3.46}$$

with $\mu > 0$ and

$$\int_{0}^{\infty} g(v)\, dv < 1 \tag{3.47}$$

such that negative-valued probabilities can be restricted to a relative low level.

The assumption regarding a probability system (Equation 3.45) is exactly the same as that in Definition 3.4 which restricts the analysis only to the field of RPP. At the meantime, another assumption $\{N(s), (s \le t) : \Lambda(s), s \in (-\infty, \infty)\}$ manifests the intensity function $\Lambda(s)$ adequately as a pre-defined process at the initial session of retrieving the formula 3.46.

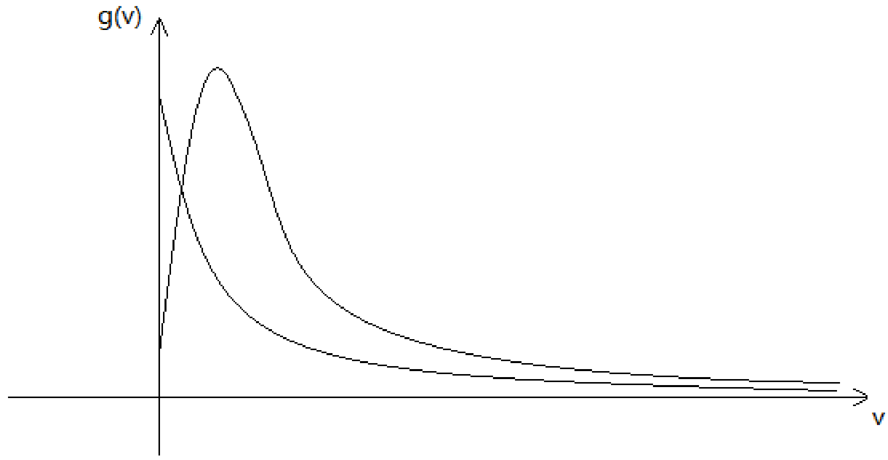

Figure 3.2: Samples of g(v)

Now fix eyes on the intensity function 3.46 whose restrictions guarantee a positively ranged function $g(v)$ starting at value of zero when $v < 0$. Intuitively observe through Figure 3.2, when t moves along the time string from 0, the more severe $g(v)$ declines, the more larger the integral term in 3.46 in accompany with the intensity will be. Distinctively if there exists a hump in $g(v)$, the more height and thickness a hump has, the larger proportion of integral value will cluster on the hump part for integration over a relative longer interval (i.e., for larger t). It testifies that a hump may effect the intensities of longer time period hereafter than a $g(v)$ without hump does, and the extent of "longer" is determined by concrete properties of the hump.

Hawkes process is firstly applied in geographical and geological empirical modelling, such as for earthquake (1970s) and rainfall (1980s) analyses. Both those objects of study possess strongly a physical feature that the situation of present or even the future has a tight correlation with previous circumstances. Also, by the intensity function 3.46, notice that the more the past time approaches to present, the more mightily it will link to the situation for present.

During this decade, econometricians begin bringing this extraordinary model into field of finance. In a financial market, or more precisely, under a high-frequency trading environment, event arrival times have a feature of clustering: when being informed of some news which would influence traders' expectation on the company's operation, for example, merger, scandal or rumor, even because of a tiny fluctuation of the stock price, the investors tend to crowd into the market and take the same trading side (bid or ask) simultaneously. Investors' willings with the same side "stand" in a queue and execute orderly according to local trading priority principles (see also Section 1.2) in specific exchange platform, which indicates an intense relevance between clustering event times. Therefore, the time when a new event arrives is related powerfully to the timestamps which occur previously. This conclusion economically interpretes the feasibility and significance of modelling financial high-frequency data with Hawkes process.

3.4 Intensity-based Inference of Hawkes(P) Process

As is known to all from the definition, the Hawkes process supplies a special form of intensity function for general RPP. That means, all former derivative intensity-based inferences with respect to RPP, including likelihood function and forecast related equations can be applied mechanically into Hawkes process.

An exponentially decaying impulse response function $g(v)$, i.e.,

$$g(v) = \sum_{j=1}^{P} \alpha_j \cdot \exp(-\beta_j \cdot v), \tag{3.48}$$

with $v > 0$ and

$$\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} < 1 \tag{3.49}$$

is employed in entire later research. It is because on one side it is proper for describing empirical financial high-frequency data from the economic point of view; on the other side analytical results for statistical inference is attainable when making use of this function. Here α and β denote the scale and location parameters respectively which determine the "portrait" of intensity function. P is the order of Hawkes process which is picked up exogenously and means the number of superposed different response functions with divers parameters but in the same form as 3.48.

Likelihood Function

As is derived in Section 3.2, the likelihood function for an RPP can be expressed as

$$\ln \mathcal{L}(\Theta | t_1, t_2, ..., t_n) = -\int_0^T \lambda(\Theta|t)\, dt + \int_0^T \ln \lambda(\Theta|t)\, dN(t). \tag{3.50}$$

Thus, for a one-variate Hawkes(P) process which has the intensity function

$$\begin{aligned} \lambda(t) &= \mu + \int_{-\infty}^{t} g(t-u)\, dN(u) \\ &= \mu + \int_{-\infty}^{t} \left\{ \sum_{j=1}^{P} \alpha_j \cdot \exp(-\beta_j \cdot (t-u)) \right\} dN(u) \\ &= \mu + \sum_{j=1}^{P} \sum_{m=1}^{\check{N}(t)} \alpha_j \cdot \exp(-\beta_j \cdot (t - t_m)), \end{aligned} \tag{3.51}$$

where t_m means some observed event arrival time that occurs before t. Thereafter, the likelihood function can be obtained by simply substituting the Equation 3.51 into 3.50 with $\Theta = \{\mu, \alpha_1, \alpha_2, ..., \alpha_P, \beta_1, \beta_2, ..., \beta_P\}$. Note that Equation 3.50 states a general formula for likelihood function until T given all information until time t_n, $T \geq t_n$. Following analyses on likelihood function and prediction are basement of empirical calibrations in next chapter. Therefore, in the further research, likelihood function is simplified with $T = t_n$, and the integrals are rewritten into form of summations. Then, the likelihood function for a mono-variate Hawkes(P) process is

$$
\begin{aligned}
\ln \mathcal{L}(\Theta|t_1, t_2, ..., t_n) = & -\int_0^{t_n} \left\{ \mu + \sum_{j=1}^{P} \sum_{m=1}^{\check{N}(t)} \alpha_j \cdot \exp\left(-\beta_j \cdot (t - t_m)\right) \right\} dt \\
& + \int_0^{t_n} \ln \left\{ \mu + \sum_{j=1}^{P} \sum_{m=1}^{\check{N}(t)} \alpha_j \cdot \exp\left(-\beta_j \cdot (t - t_m)\right) \right\} dN(t) \\
= & -\int_0^{t_n} \mu dt - \sum_{j=1}^{P} \int_0^{t_n} \sum_{m=1}^{\check{N}(t)} \alpha_j \cdot \exp\left(-\beta_j \cdot (t - t_m)\right) dt \\
& + \int_0^{t_n} \ln \left\{ \mu + \sum_{j=1}^{P} \sum_{m=1}^{\check{N}(t)} \alpha_j \cdot \exp\left(-\beta_j \cdot (t - t_m)\right) \right\} dN(t) \\
= & -\int_0^{t_n} \mu dt - \sum_{j=1}^{P} \left[-\sum_{m=1}^{\check{N}(t)} \frac{\alpha_j}{\beta_j} \cdot \exp\left(-\beta_j \cdot (t - t_m)\right) \right]_0^{t_n} \\
& + \int_0^{t_n} \ln \left\{ \mu + \sum_{j=1}^{P} \sum_{m=1}^{\check{N}(t)} \alpha_j \cdot \exp\left(-\beta_j \cdot (t - t_m)\right) \right\} dN(t) \\
= & -\sum_{i=1}^{n} \mu \cdot (t_i - t_{i-1}) \\
& + \sum_{j=1}^{P} \sum_{m=1}^{\check{N}(t_i)} \sum_{i=1}^{n} \left\{ \frac{\alpha_j}{\beta_j} \cdot \exp\left(-\beta_j \cdot (t_i - t_m)\right) - \frac{\alpha_j}{\beta_j} \cdot \exp\left(-\beta_j \cdot (t_{i-1} - t_m)\right) \right\} \\
& + \sum_{i=1}^{n} \ln \left\{ \mu + \sum_{j=1}^{P} \sum_{m=1}^{\check{N}(t_i)} \alpha_j \cdot \exp\left(-\beta_j \cdot (t_i - t_m)\right) \right\} \cdot 1 \\
= & \sum_{i=1}^{n} \{ -\mu \cdot (t_i - t_{i-1}) \\
& + \sum_{j=1}^{P} \sum_{m=1}^{i-1} \left\{ \frac{\alpha_j}{\beta_j} \cdot \exp\left(-\beta_j \cdot (t_i - t_m)\right) - \frac{\alpha_j}{\beta_j} \cdot \exp\left(-\beta_j \cdot (t_{i-1} - t_m)\right) \right\} \\
& + \ln \left\{ \mu + \sum_{j=1}^{P} \sum_{m=1}^{i-1} \alpha_j \cdot \exp\left(-\beta_j \cdot (t_i - t_m)\right) \right\} \}
\end{aligned}
\tag{3.52}
$$

where $t_0 = 0$. Use the technique of recursion and obtain

$$
\begin{aligned}
A_i^j := & \sum_{m=1}^{i-1} \exp\left(-\beta_j \cdot (t_i - t_m)\right) \\
&= \exp\left(-\beta_j \cdot (t_i - t_{i-1})\right) + \exp\left(-\beta_j \cdot (t_i - t_{i-2})\right) + \cdots + \exp\left(-\beta_j \cdot (t_i - t_1)\right) \\
&= \exp\left(-\beta_j \cdot (t_i - t_{i-1})\right) \cdot \{1 + \exp\left(-\beta_j \cdot (t_{i-1} - t_{i-2})\right) + \cdots + \exp\left(-\beta_j \cdot (t_{i-1} - t_1)\right)\} \\
&= \exp\left(-\beta_j \cdot (t_i - t_{i-1})\right) \cdot \left\{1 + \sum_{m=1}^{i-2} \exp\left(-\beta_j \cdot (t_{i-1} - t_m)\right)\right\} \\
&= \exp\left(-\beta_j \cdot (t_i - t_{i-1})\right) \cdot \left(1 + A_{i-1}^j\right),
\end{aligned}
\tag{3.53}
$$

where $A_0^j = 0$. Then, the likelihood function of one-dimensional Hawkes(P) process is simplified

$$
\begin{aligned}
\ln \mathcal{L}\left(\Theta | t_1, t_2, ..., t_n\right) = & \sum_{i=1}^{n} \{-\mu \cdot (t_i - t_{i-1}) \\
& + \sum_{j=1}^{P} \sum_{m=1}^{i-1} \left\{ \frac{\alpha_j}{\beta_j} \cdot \exp\left(-\beta_j \cdot (t_i - t_m)\right) - \frac{\alpha_j}{\beta_j} \cdot \exp\left(-\beta_j \cdot (t_{i-1} - t_m)\right) \right\} \\
& + \ln\left(\mu + \sum_{j=1}^{P} \alpha_j \cdot A_i^j\right) \Bigg\},
\end{aligned}
\tag{3.54}
$$

When dealing with a multi-dimensional Hawkes process, for instance, a K-dimensional Hawkes(P) process, the intensity function regarding the k-th process in favour of conclusion 3.17 is

$$
\begin{aligned}
\lambda^k(t) = & \mathbb{E}^{\mathcal{F}_t^k} \Bigg[\left(\mu^{k,1} + \cdots + \mu^{k,K}\right) + \sum_{j=1}^{P} \Bigg\{ \sum_{m=1}^{\check{N}^l(t)} \alpha_j^{k,1} \cdot \exp\left(-\beta_j^{k,1} \cdot \left(t - t_m^1\right)\right) \\
& + \sum_{m=1}^{\check{N}^l(t)} \alpha_j^{k,2} \cdot \exp\left(-\beta_j^{k,2} \cdot \left(t - t_m^2\right)\right) + \cdots + \sum_{m=1}^{\check{N}^l(t)} \alpha_j^{k,K} \cdot \exp\left(-\beta_j^{k,K} \cdot \left(t - t_m^K\right)\right) \Bigg\} \Bigg] \\
& = \mu^k + \sum_{l=1}^{K} \sum_{j=1}^{P} \sum_{m=1}^{\check{N}^l(t)} \alpha_j^{k,l} \cdot \exp\left(-\beta_j^{k,l} \cdot \left(t - t_m^l\right)\right),
\end{aligned}
\tag{3.55}
$$

where all parameters are non-negative and impact of l-type backward recurrences on k-th time string is under consideration. The sum of $\mu^{k,l}$ can be instituted by mu^k since all μ's are unknown. Also, it is clear that all l-type event arrivals are restricted by t on k-th series which leads an omission of the expectation symbol.

Similar to the technique applied in one-variate Hawkes process analysis, likelihood function of a multi-variate Hawkes(P) process can be calculated. Moreover, the likelihood function of complete model can be derived as sum of that of each single multi-variate process. Thus, likelihood function of the model is

$$\begin{aligned}
\ln \mathcal{L}\left(\Theta | t_1, t_2, ..., t_n\right) = & \sum_{k=1}^{K} \Bigg\{ -\int_0^{t_n} \left\{ \mu^k + \sum_{l=1}^{K} \sum_{j=1}^{P} \sum_{m=1}^{\check{N}^l(t)} \alpha_j^{k,l} \cdot \exp\left(-\beta_j^{k,l} \cdot \left(t - t_m^l\right)\right) \right\} dt \\
& + \int_0^{t_n} \ln \left\{ \mu^k + \sum_{l=1}^{K} \sum_{j=1}^{P} \sum_{m=1}^{\check{N}^l(t)} \alpha_j^{k,l} \cdot \exp\left(-\beta_j^{k,l} \cdot \left(t - t_m^l\right)\right) \right\} dN\left(t\right) \Bigg\} \\
= & \sum_{k=1}^{K} \Bigg\{ -\int_0^{t_n} \mu^k dt - \sum_{l=1}^{K} \sum_{j=1}^{P} \int_0^{t_n} \sum_{m=1}^{\check{N}^l(t)} \alpha_j^{k,l} \cdot \exp\left(-\beta_j^{k,l} \cdot \left(t - t_m^l\right)\right) dt \\
& + \sum_{i=1}^{n} \ln \left\{ \mu^k + \sum_{l=1}^{K} \sum_{j=1}^{P} \sum_{m=1}^{\check{N}^l(t_i)} \alpha_j^{k,l} \cdot \exp\left(-\beta_j^{k,l} \cdot \left(t_i - t_m^l\right)\right) \right\} \cdot 1 \Bigg\} \\
= & \sum_{k=1}^{K} \sum_{i=1}^{n} \Bigg\{ -\mu^k \cdot \left(t_i - t_{i-1}\right) \\
& + \sum_{l=1}^{K} \sum_{j=1}^{P} \sum_{m=1}^{\check{N}^l(t_i)} \left\{ \frac{\alpha_j^{k,l}}{\beta_j^{k,l}} \cdot \exp\left(-\beta_j^{k,l} \cdot \left(t_i - t_m^l\right)\right) \right. \\
& \left. - \frac{\alpha_j^{k,l}}{\beta_j^{k,l}} \cdot \exp\left(-\beta_j^{k,l} \cdot \left(t_{i-1} - t_m^l\right)\right) \right\} \\
& + \ln \left\{ \mu^k + \sum_{l=1}^{K} \sum_{j=1}^{P} \sum_{m=1}^{\check{N}^l(t_i)} \alpha_j^{k,l} \cdot \exp\left(-\beta_j^{k,l} \cdot \left(t_i - t_m^l\right)\right) \right\} \Bigg\} \\
= & \sum_{k=1}^{K} \sum_{i=1}^{n} \Bigg\{ -\int_{t_{i-1}}^{t_i} \mu^k du \\
& + \sum_{l=1}^{K} \sum_{j=1}^{P} \sum_{m=1}^{\check{N}^l(t_i)} \frac{\alpha_j^{k,l}}{\beta_j^{k,l}} \cdot \left\{ \exp\left(-\beta_j^{k,l} \cdot \left(t_i - t_m^l\right)\right) \right. \\
& \left. - \exp\left(-\beta_j^{k,l} \cdot \left(t_{i-1} - t_m^l\right)\right) \right\} \\
& + \ln \left\{ \mu^k + \sum_{l=1}^{K} \sum_{j=1}^{P} \alpha_j^{k,l} \cdot \left(\sum_{m=1}^{\check{N}^l(t_i)} \exp\left(-\beta_j^{k,l} \cdot \left(t_i - t_m^l\right)\right) \right) \right\} \Bigg\}
\end{aligned} \tag{3.56}$$

where

$$
\begin{aligned}
A_i^{j,kl} :&= \sum_{m=1}^{\check{N}^l(t_i)} \exp\left(-\beta_j^{k,l} \cdot \left(t_i - t_m^l\right)\right) \\
&= \exp\left(-\beta_j^{k,l} \cdot \left(t_i - t^l_{\check{N}^l(t_i)}\right)\right) \\
&\quad + \exp\left(-\beta_j^{k,l} \cdot \left(t_i - t^l_{\check{N}^l(t_i)-1}\right)\right) + \cdots + \exp\left(-\beta_j^{k,l} \cdot \left(t_i - t_1^l\right)\right) \\
&= \exp\left(-\beta_j^{k,l} \cdot (t_i - t_{i-1})\right) \cdot \Big\{\exp\left(-\beta_j^{k,l} \cdot \left(t_{i-1} - t^l_{\check{N}^l(t_i)}\right)\right) \\
&\quad + \exp\left(-\beta_j^{k,l} \cdot \left(t_{i-1} - t^l_{\check{N}^l(t_i)-1}\right)\right) + \cdots + \exp\left(-\beta_j^{k,l} \cdot \left(t_{i-1} - t_1^l\right)\right)\Big\} \\
&= \exp\left(-\beta_j^{k,l} \cdot (t_i - t_{i-1})\right) \cdot \Big\{\exp\left(-\beta_j^{k,l} \cdot \left(t_{i-1} - t^l_{\check{N}^l(t_i)}\right)\right) \\
&\quad + \exp\left(-\beta_j^{k,l} \cdot \left(t_{i-1} - t^l_{\check{N}^l(t_{i}-1)}\right)\right) + \cdots + \exp\left(-\beta_j^{k,l} \cdot \left(t_{i-1} - t_1^l\right)\right)\Big\} \\
&= \exp\left(-\beta_j^{k,l} \cdot (t_i - t_{i-1})\right) \cdot \Big\{\exp\left(-\beta_j^{k,l} \cdot \left(t_{i-1} - t^l_{\check{N}^l(t_i)}\right)\right) + A_{i-1}^{j,kl}\Big\} \\
&= \begin{cases} \exp\left(-\beta_j^{k,l} \cdot (t_i - t_{i-1})\right) \cdot \Big\{0 + A_{i-1}^{j,kl}\Big\} & t_{i-1} = \check{N}^l(t_i) \\ \exp\left(-\beta_j^{k,l} \cdot (t_i - t_{i-1})\right) \cdot \Big\{1 + A_{i-1}^{j,kl}\Big\} & t_{i-1} \neq \check{N}^l(t_i) \end{cases} \\
&= \exp\left(-\beta_j^{k,l} \cdot (t_i - t_{i-1})\right) \cdot \left(y_{i-1}^l + A_{i-1}^{j,kl}\right)
\end{aligned}
\tag{3.57}
$$

with $A_0^{j,kl} = 0$. Here y_{i-1}^l is an indicator function taking value one when $(i-1)$-th event is of type l with $y_o = 0$.

With these discrete forms of analytical likelihood functions, empirical models are possible to implement. Conclusion 3.17 associated with this likelihood function makes the relatively complicated multi-dimensional Hawkes process-based modelling systems popular which in an aspect is because of its property of additivity in each single multi-variate process. Further studies, such as conditions and parametric calibrations about maximum likelihood estimator (MLE) are demonstrated in Chapter 4.

Prediction Formula

With all conditions and assumptions for the model, martingale-based calculating principles and stationarity accomplished, to forecast the next event arrival time $t_{n+1}$with given information until t_n, mathematically speaking, is to calculate the conditional expectation $\mathbb{E}\left[t_{n+1}|\mathcal{F}_{t_n}\right]$. As is discussed before, one obtains the forecasting formula for mono- or multi-variate Hawkes process without difficulty by inserting concrete intensity functions into Equation 3.44. Thus the prediction formula for one- and k-th high-variate Hawkes(P) processes separately are presented as

$$
\begin{aligned}
\epsilon_{n+1} &= \Lambda\left(t_n, t_{n+1}\right) \\
&= \int_{t_n}^{t_{n+1}} \lambda\left(s; \mathcal{F}_s\right) ds \\
&= \int_{t_n}^{t_{n+1}} \mu + \sum_{j=1}^{P} \sum_{m=1}^{\check{N}(s)} \alpha_j \cdot \exp\left(-\beta_j \cdot (s - t_m)\right) ds \\
&= \mu \cdot (t_{n+1} - t_n) + \sum_{j=1}^{P} \int_{t_n}^{t_{n+1}} \sum_{m=1}^{\check{N}(s)} \alpha_j \cdot \exp\left(-\beta_j \cdot (s - t_m)\right) ds \\
&= \mu \cdot (t_{n+1} - t_n) - \sum_{j=1}^{P} \left[\sum_{m=1}^{\check{N}(s)} \frac{\alpha_j}{\beta_j} \cdot \exp\left(-\beta_j \cdot (s - t_m)\right) \right]_{t_n}^{t_{n+1}} \\
&= \mu \cdot (t_{n+1} - t_n) - \sum_{j=1}^{P} \sum_{m=1}^{n} \frac{\alpha_j}{\beta_j} \cdot \left[\exp\left(-\beta_j \cdot (t_{n+1} - t_m)\right) - \exp\left(-\beta_j \cdot (t_n - t_m)\right)\right],
\end{aligned}
\tag{3.58}
$$

and

$$
\begin{aligned}
\epsilon^k_{n+1} &= \Lambda^k\left(t_n, t_{n+1}\right) \\
&= \int_{t_n}^{t_{n+1}} \mu^k + \sum_{l=1}^{K} \sum_{j=1}^{P} \sum_{m=1}^{\check{N}^l(s)} \alpha_j^{k,l} \cdot \exp\left(-\beta_j^{k,l} \cdot \left(s - t^l_m\right)\right) ds \\
&= \mu^k \cdot (t_{n+1} - t_n) + \sum_{l=1}^{K} \sum_{j=1}^{P} \int_{t_n}^{t_{n+1}} \sum_{m=1}^{\check{N}^l(s)} \alpha_j^{k,l} \cdot \exp\left(-\beta_j^{k,l} \cdot \left(s - t^l_m\right)\right) ds \\
&= \mu^k \cdot (t_{n+1} - t_n) - \sum_{l=1}^{K} \sum_{j=1}^{P} \left[\sum_{m=1}^{\check{N}^l(s)} \frac{\alpha_j^{k,l}}{\beta_j^{k,l}} \cdot \exp\left(-\beta_j^{k,l} \cdot \left(s - t^l_m\right)\right) \right]_{t_n}^{t_{n+1}} \\
&= \mu^k \cdot (t_{n+1} - t_n) \\
&\quad - \sum_{l=1}^{K} \sum_{j=1}^{P} \sum_{m=1}^{\check{N}^l(t_{n+1})} \frac{\alpha_j^{k,l}}{\beta_j^{k,l}} \cdot \left[\exp\left(-\beta_j^{k,l} \cdot \left(t_{n+1} - t^l_m\right)\right) - \exp\left(-\beta_j^{k,l} \cdot \left(t_n - t^l_m\right)\right)\right],
\end{aligned}
\tag{3.59}
$$

where both ϵ_{n+1} and ϵ^k_{n+1} are random variates that follow the standard exponential distribution Exp(1). In Equation 3.58 and 3.59, with all event arrivals until t_n are given, when parameters are determined, there remains only one unknown factor t_{n+1}. Although unfortunately, a general solution to this nonlinear equation is not reachable, several numerical means can be employed in order to approach the true value. How to generate i.i.d. exponentially distributed random variables, to analyze the forecasting results, etcetera are clarified in next chapter.

Chapter 4

Calibration

Application of models always plays a fairly important role in judging whether a model is proper or not to simulate given data, because a positively cognized model is always considered to retain mighty latent financial values for the industry. Economically speaking, modelling irregularly-spaced financial high-frequency data using Hawkes processes makes sense. And to the aspect of theoretical part, closed discrete forms of log-likelihood function and prediction formula are delightful since one is able to numerically realize empirical simulation, calibrate the parameters in Equation 3.52 or 3.56 using MLE and omen next event arrival time with help of 3.58 or 3.59. After numerical calibrations, one can tell how good empirically the Hawkes process fits after proper results' contrasts and studies regarding the data which are employed in Chapter 2.

In later of this thesis, all simulations and analyses are based on mono- or multi-dimensional Hawkes processes with the order being exogenously chosen as one, i.e., $P = 1$. In this chapter, conditions for existence of MLE and tips or tricks for computational programming, etc, which are not mentioned in Chapter 3 are exhibited here. Some intuitive figures and analyses of results are also jointly settled in this chapter as this session directly reflects the effect of fitting and is manifested as bridge combining Chapter 2 and 3.

4.1 Model Setting

For instance, a model that consists of intensities for bid and ask orders, market buy and sell and bid and ask cancellations is detected. The complete 6-dimensional Hawkes(1) model possesses an intensity system as

$$\begin{cases}
\lambda^{S}\left(t;\mathcal{F}_{t}\right)=\mu^{S}+\sum_{m=1}^{\check{N}^{S}(t)}\alpha^{S,S}\cdot\exp\left(-\beta^{S,S}\cdot\left(t-t_{m}^{S}\right)\right)\\
\qquad+\sum_{m=1}^{\check{N}^{B}(t)}\alpha^{S,B}\cdot\exp\left(-\beta^{S,B}\cdot\left(t-t_{m}^{B}\right)\right)\\
\qquad+\cdots+\sum_{m=1}^{\check{N}^{CB}(t)}\alpha^{S,CB}\cdot\exp\left(-\beta^{S,CB}\cdot\left(t-t_{m}^{CB}\right)\right)\\
\lambda^{B}\left(t;\mathcal{F}_{t}\right)=\mu^{B}+\sum_{m=1}^{\check{N}^{S}(t)}\alpha^{B,S}\cdot\exp\left(-\beta^{B,S}\cdot\left(t-t_{m}^{S}\right)\right)\\
\qquad+\sum_{m=1}^{\check{N}^{B}(t)}\alpha^{B,B}\cdot\exp\left(-\beta^{B,B}\cdot\left(t-t_{m}^{B}\right)\right)\\
\qquad+\cdots+\sum_{m=1}^{\check{N}^{CB}(t)}\alpha^{B,CB}\cdot\exp\left(-\beta^{B,CB}\cdot\left(t-t_{m}^{CB}\right)\right)\\
\lambda^{MS}\left(t;\mathcal{F}_{t}\right)=\mu^{MS}+\sum_{m=1}^{\check{N}^{S}(t)}\alpha^{MS,S}\cdot\exp\left(-\beta^{MS,S}\cdot\left(t-t_{m}^{S}\right)\right)\\
\qquad+\sum_{m=1}^{\check{N}^{B}(t)}\alpha^{MS,B}\cdot\exp\left(-\beta^{MS,B}\cdot\left(t-t_{m}^{B}\right)\right)\\
\qquad+\cdots+\sum_{m=1}^{\check{N}^{CB}(t)}\alpha^{MS,CB}\cdot\exp\left(-\beta^{MS,CB}\cdot\left(t-t_{m}^{CB}\right)\right)\\
\lambda^{MB}\left(t;\mathcal{F}_{t}\right)=\mu^{MB}+\sum_{m=1}^{\check{N}^{S}(t)}\alpha^{MB,S}\cdot\exp\left(-\beta^{MB,S}\cdot\left(t-t_{m}^{S}\right)\right)\\
\qquad+\sum_{m=1}^{\check{N}^{B}(t)}\alpha^{MB,B}\cdot\exp\left(-\beta^{MS,B}\cdot\left(t-t_{m}^{B}\right)\right)\\
\qquad+\cdots+\sum_{m=1}^{\check{N}^{CB}(t)}\alpha^{MB,CB}\cdot\exp\left(-\beta^{MB,CB}\cdot\left(t-t_{m}^{CB}\right)\right)\\
\lambda^{CS}\left(t;\mathcal{F}_{t}\right)=\mu^{CS}+\sum_{m=1}^{\check{N}^{S}(t)}\alpha^{CS,S}\cdot\exp\left(-\beta^{CS,S}\cdot\left(t-t_{m}^{S}\right)\right)\\
\qquad+\sum_{m=1}^{\check{N}^{B}(t)}\alpha^{CS,B}\cdot\exp\left(-\beta^{CS,B}\cdot\left(t-t_{m}^{B}\right)\right)\\
\qquad+\cdots+\sum_{m=1}^{\check{N}^{CB}(t)}\alpha^{CS,CB}\cdot\exp\left(-\beta^{CS,CB}\cdot\left(t-t_{m}^{CB}\right)\right)\\
\lambda^{CB}\left(t;\mathcal{F}_{t}\right)=\mu^{CB}+\sum_{m=1}^{\check{N}^{S}(t)}\alpha^{CB,S}\cdot\exp\left(-\beta^{CB,S}\cdot\left(t-t_{m}^{S}\right)\right)\\
\qquad+\sum_{m=1}^{\check{N}^{B}(t)}\alpha^{CB,B}\cdot\exp\left(-\beta^{CS,B}\cdot\left(t-t_{m}^{B}\right)\right)\\
\qquad+\cdots+\sum_{m=1}^{\check{N}^{CB}(t)}\alpha^{CB,CB}\cdot\exp\left(-\beta^{CB,CB}\cdot\left(t-t_{m}^{CB}\right)\right),
\end{cases} \tag{4.1}$$

where the set $\{S, B, MS, MB, CS, CB\}$ denotes {*order sell, order buy, market sell, market buy, cancellation sell, cancellation buy*}.

Thus, each intensity is a 6-variate process and correlated with the others. The μ's in Hawkes processes can be treated as functions as well as constants, but in this thesis μ's are not treated as dynamics for convenience of MLE approaches. Thus, number of unknown parameters that should be estimated is $1 + 2 \cdot 6 = 13$ for each single process, which leads the model to own up to $13 \cdot 6 = 78$ parameters to approach. Now look back to Equation 3.56 and 3.57, there are at least five for-loops when having accomplished them on software and each one for-loop spends plenty of time in calculations.

As a consequence, it is naturally and appropriate that one begins the studies by implementing the simplest 6-dimensional model, and then take efforts in improving the existing one. Being restricted by the capacity of personal laptop, empirical studies about Hawkes(1) processes in this thesis focus only on this highly simplified model which is

$$\begin{cases}
\lambda^{S}\left(t;\mathcal{F}_{t}\right)=\mu^{S}+\sum_{m=1}^{\check{N}^{S}(t)}\alpha^{S,S}\cdot\exp\left(-\beta^{S,S}\cdot\left(t-t_{m}^{S}\right)\right)\\
\lambda^{B}\left(t;\mathcal{F}_{t}\right)=\mu^{B}+\sum_{m=1}^{\check{N}^{B}(t)}\alpha^{B,B}\cdot\exp\left(-\beta^{B,B}\cdot\left(t-t_{m}^{B}\right)\right)\\
\lambda^{MS}\left(t;\mathcal{F}_{t}\right)=\mu^{MS}+\sum_{m=1}^{\check{N}^{MS}(t)}\alpha^{MS,MS}\cdot\exp\left(-\beta^{MS,MS}\cdot\left(t-t_{m}^{MS}\right)\right)\\
\lambda^{MB}\left(t;\mathcal{F}_{t}\right)=\mu^{MB}+\sum_{m=1}^{\check{N}^{MB}(t)}\alpha^{MB,MB}\cdot\exp\left(-\beta^{MB,MB}\cdot\left(t-t_{m}^{MB}\right)\right)\\
\lambda^{CS}\left(t;\mathcal{F}_{t}\right)=\mu^{CS}+\sum_{m=1}^{\check{N}^{CS}(t)}\alpha^{CS,CS}\cdot\exp\left(-\beta^{CS,CS}\cdot\left(t-t_{m}^{CS}\right)\right)\\
\lambda^{CB}\left(t;\mathcal{F}_{t}\right)=\mu^{CB}+\sum_{m=1}^{\check{N}^{CB}(t)}\alpha^{CB,CB}\cdot\exp\left(-\beta^{CB,CB}\cdot\left(t-t_{m}^{CB}\right)\right).
\end{cases} \tag{4.2}$$

Apparently in the simplified model, each intensity for a specific type of event is only affected by all past occurrences of the same type. That is to say, by this case, it is assumed that any different sorts of intensities are independent to each other. Therefore, in Model 4.2, there exist only $3 \cdot 6 = 18$ unknown parameters, which reduces the computational effort largely whilst the accuracy of simulation may decline to some extent.

4.2 Parameter Estimation

The maximum likelihood estimator (MLE) is always firstly considered as a powerful tool in estimating parameters once a closed form of joint probability density function for each intensity in Model 4.2 is achieved.

Definition 4.1. (Maximum Likelihood Estimator) $\mathcal{L}(\Theta|\boldsymbol{x})$ is log-likelihood function with respect to a specific distribution for each sample point $\boldsymbol{x}$. Let $\hat{\Theta}(\boldsymbol{x})$ be a set of paramater values at which $\mathcal{L}(\Theta|\boldsymbol{x})$ attains its maximum with $\boldsymbol{x}$ fixed. Thus, a maximum likelihood estimator (MLE) of parameter set Θ based on sample $\boldsymbol{X}$ is $\hat{\theta}(\boldsymbol{X})$.

If $\mathcal{L}(\Theta|\boldsymbol{x})$ is differentiable in all elements of Θ, $\Theta = \{\theta_1, \theta_2, ..., \theta_n\}$. Therefore, one is able to solve the first derivative function system

$$\frac{\partial}{\partial \theta_i} \mathcal{L}(\Theta|\boldsymbol{x}) = 0, \quad i = 1, 2, ..., n$$

and obtain candidates for MLE. After satisfying specific conditions for second derivative functions which are specified later, the candidates can be recognized as ultimate MLE.

Since log-likelihood function is a monotone, one-to-one function of likelihood function, thus the log-likelihood function can also exploited in finding MLE without having any compact on parameters' choice. For log-likelihood function $\ln \mathcal{L}(\Theta|t_1, t_2, ..., t_n)$ employed in this thesis

$$\begin{aligned} \ln \mathcal{L}(\Theta|t_1, t_2, ..., t_n) = \sum_{i=1}^{n} & \{-\mu \cdot (t_i - t_{i-1}) \\ & \sum_{m=1}^{i-1} \left\{ \frac{\alpha}{\beta} \cdot \exp(-\beta \cdot (t_i - t_m)) - \frac{\alpha}{\beta} \cdot \exp(-\beta \cdot (t_{i-1} - t_m)) \right\} \\ & + \ln \left\{ \mu + \sum_{m=1}^{i-1} \alpha \cdot \exp(-\beta \cdot (t_i - t_m)) \right\} \Bigg\}, \end{aligned} \tag{4.3}$$

one searches for a proper $\{\mu, \alpha, \beta\}$ such that

$$\begin{cases} \frac{\partial}{\partial \mu} \ln \mathcal{L}(\mu, \alpha, \beta|t_1, t_2, ..., t_n) = 0 \\ \frac{\partial}{\partial \alpha} \ln \mathcal{L}(\mu, \alpha, \beta|t_1, t_2, ..., t_n) = 0 \\ \frac{\partial}{\partial \beta} \ln \mathcal{L}(\mu, \alpha, \beta|t_1, t_2, ..., t_n) = 0, \end{cases}$$

which can be materially expressed as

$$\begin{cases} \frac{\partial}{\partial\mu}\ln\mathcal{L}(\Theta|t) = \sum_{i=1}^{n}\left\{-(t_i - t_{i-1}) + \frac{1}{\ln\left\{\mu+\sum_{m=1}^{i-1}\alpha\cdot\exp(-\beta\cdot(t_i-t_m))\right\}}\right\} = 0 \\ \frac{\partial}{\partial\alpha}\ln\mathcal{L}(\Theta|t) = \sum_{i=1}^{n}\sum_{m=1}^{i-1}\left\{\frac{1}{\beta}\left(\exp\left(-\beta\cdot(t_i-t_m)\right) - \exp\left(-\beta\cdot(t_{i-1}-t_m)\right)\right)\right\} \\ \qquad + \sum_{i=1}^{n}\frac{\sum_{m=1}^{i-1}\exp(-\beta\cdot(t_i-t_m))}{\ln\left\{\mu+\sum_{m=1}^{i-1}\alpha\cdot\exp(-\beta\cdot(t_i-t_m))\right\}} = 0 \\ \frac{\partial}{\partial\beta}\ln\mathcal{L}(\Theta|t) = \sum_{i=1}^{n}\sum_{m=1}^{i-1}\left\{\left(-\frac{\alpha}{\beta^2}-\alpha\right)\cdot\left(\exp\left(-\beta\cdot(t_i-t_m)\right) - \exp\left(-\beta\cdot(t_{i-1}-t_m)\right)\right)\right\} \\ \qquad + \sum_{i=1}^{n}\frac{\sum_{m=1}^{i-1}-\alpha\beta\cdot\exp(-\beta\cdot(t_i-t_m))}{\ln\left\{\mu+\sum_{m=1}^{i-1}\alpha\cdot\exp(-\beta\cdot(t_i-t_m))\right\}} = 0. \end{cases} \tag{4.4}$$

The solutions $\left\{\hat{\mu}, \hat{\alpha}, \hat{\beta}\right\}$ to 4.4 is a MLE when they satisfy at least the conditions for second partial derivatives

$$\frac{\partial^2}{\partial\mu^2}\ln\mathcal{L}(\Theta|t) < 0 \quad or \quad \frac{\partial^2}{\partial\alpha^2}\ln\mathcal{L}(\Theta|t) < 0 \quad or \quad \frac{\partial^2}{\partial\beta^2}\ln\mathcal{L}(\Theta|t) < 0 \tag{4.5}$$

and the Jacobian be

$$\left|\begin{matrix} \frac{\partial^2}{\partial\mu^2}\ln\mathcal{L}(\Theta|t) & \frac{\partial^2}{\partial\mu\partial\alpha}\ln\mathcal{L}(\Theta|t) & \frac{\partial^2}{\partial\mu\partial\beta}\ln\mathcal{L}(\Theta|t) \\ \frac{\partial^2}{\partial\alpha\partial\mu}\ln\mathcal{L}(\Theta|t) & \frac{\partial^2}{\partial\alpha^2}\ln\mathcal{L}(\Theta|t) & \frac{\partial^2}{\partial\alpha\partial\beta}\ln\mathcal{L}(\Theta|t) \\ \frac{\partial^2}{\partial\beta\partial\mu}\ln\mathcal{L}(\Theta|t) & \frac{\partial^2}{\partial\beta\partial\alpha}\ln\mathcal{L}(\Theta|t) & \frac{\partial^2}{\partial\beta^2}\ln\mathcal{L}(\Theta|t) \end{matrix}\right|_{\mu=\hat{\mu},\alpha=\hat{\alpha},\beta=\hat{\beta}} > 0 \tag{4.6}$$

It seems difficult to retrieve a MLE by means of solving all non-linear equations above. Therefore, one prefers exploiting numerical approaches instead, i.e., one could optimize the log-likelihood function with scientific computing. Notice that the likelihood function is in the same form as joint *pdf* of specific distribution, which is restricted in the interval $[0, 1]$. Consequently, value of log-likelihood function is always non-positive with non-negative parameters. With all these conditions added, numerical approaches to MLE are able to figure out as concrete numbers instead of formulas.

For this thesis, we separately take experiments over three time intervals chosen from morning, mid-day and afternoon, 09:30-11:00, 12:00-13:30 and 14:00-15:30. For convenience of contrast, all eight tests are observed over series of different time intervals: equally divided three half hours in the morning session, the last half hour in both second and third time durations and whole periods of time intervals. And there are eighteen parameters to estimate for each single experiment. The results of MLE are listed:

```
> length(dat$time[9.5*3600*1E9<dat$time & dat$time<10*3600*1E9])
[1] 10968
> par1
           [,1]       [,2]         [,3]         [,4]       [,5]       [,6]
[1,] 0.02858798 0.04196397  0.001993641  0.003181868 0.02004016 0.05867768
[2,] 1.36006386 1.36052085 10.999943939 15.412955117 1.36310356 1.32021660
[3,] 1.51158459 1.51794953 16.264759348 31.787530464 1.52151744 1.47490657
> log1
```

```
[1] -221.8798


> length(dat$time[10*3600*1E9<dat$time & dat$time<10.5*3600*1E9])
[1] 9006
> par2
            [,1]       [,2]      [,3]        [,4]       [,5]       [,6]
[1,] 0.004039457 0.02517031 0.0010863 0.001726316 0.01605339 0.03311215
[2,] 1.430616084 1.34816272 1.2595922 1.061854922 8.53461690 1.35207982
[3,] 1.570226379 1.49599919 1.5992835 2.025451069 9.17493895 1.50872039
> log2
[1] -1291.815


> length(dat$time[10.5*3600*1E9<dat$time & dat$time<11*3600*1E9])
[1] 10287
> par3
           [,1]       [,2]         [,3]        [,4]       [,5]        [,6]
[1,] 0.04723964 0.01446308  0.005945633 0.001698968 0.04450376 0.007983855
[2,] 1.32909732 1.30255001 18.684952577 1.245060034 1.36152465 1.307397565
[3,] 1.48171004 1.58356131 23.040053503 1.602013627 1.52195196 1.585467795
> log3
[1] -524.9006


> length(dat$time[9.5*3600*1E9<dat$time & dat$time<11*3600*1E9])
[1] 30261
> par4
           [,1]      [,2]       [,3]         [,4]       [,5]     [,6]
[1,] 0.08782117 0.0991005 0.00522651  0.008411268 0.08593423 0.116820
[2,] 1.34270251 1.2942871 5.63502178 19.895732585 1.19173552 1.268226
[3,] 1.43119496 1.4090893 8.09151145 44.105696613 1.63138702 1.411082
> log4
[1] -327.4118
```

for time 09:30-11:00, and results for 12:00-13:30 are

```
> length(dat$time[13*3600*1E9<dat$time & dat$time<13.5*3600*1E9])
[1] 8479
> par5
            [,1]       [,2]         [,3]         [,4]        [,5]       [,6]
[1,]  0.01690415 0.01792175 7.368783e-04  0.001132344 0.002854535 0.01546516
[2,]  6.18585556 1.30750049 8.911088e+00 12.975053466 1.415425270 1.29537972
[3,] 12.31131102 1.58022134 2.478936e+01 38.216586578 1.572589485 1.58441689
> log5
```

```
[1] -538.8574


> length(dat$time[12*3600*1E9<dat$time & dat$time<13.5*3600*1E9])
[1] 25208
> par6
           [,1]       [,2]        [,3]         [,4]       [,5]       [,6]
[1,] 0.01275504 0.05539768  0.003064901   0.00332229 0.02115197 0.05354709
[2,] 1.37934667 1.35897288 19.002654840  46.67868443 1.32293745 1.32840168
[3,] 1.61649743 1.52071825 46.466680025 118.35881264 1.46398508 1.48310556
> log6
[1] -417.3114
```

Ultimately, tests based on 14:00-15:30 result in

```
> length(dat$time[15*3600*1E9<dat$time & dat$time<15.5*3600*1E9])
[1] 15651
> par7
           [,1]       [,2]        [,3]         [,4]       [,5]       [,6]
[1,] 0.07827659 0.03784188  0.001546297  0.002458917 0.08510174 0.03688887
[2,] 1.20845586 1.33261975 14.848304291 29.409582831 1.18588859 1.35682946
[3,] 1.61755595 1.48251157 33.278523558 67.312996167 1.64429042 1.51514049
> log7
[1] -324.3509


> length(dat$time[14*3600*1E9<dat$time & dat$time<15.5*3600*1E9])
[1] 47809
> par8
          [,1]      [,2]         [,3]        [,4]      [,5]      [,6]
[1,] 0.2440416 0.1614121  0.005494301  0.02296209 0.2297313 0.1307383
[2,] 1.2207989 1.2888991 10.188796651 37.10680949 1.2338307 1.3599439
[3,] 1.4084833 1.3716093 21.415938091 46.70364874 1.3734815 1.3600412
> log8
[1] -117.1763
```

Here explains the meanings of detected results above. The number in second line with respect to results of each test, for example, 10968 in test 1, suggests total amount of events that occur during the period of test 1. Following is a 6*3 matrix named after "par1" which represents the parametric estimations: the first row of matrix denotes estimated $\{\mu^S, \mu^B, \mu^{MS}, \mu^{MB}, \mu^{CS}, \mu^{CB}\}$ in order, and rest two lines are results for α's and β's in analogously ordered sequences. The negative number that always comes at last shows the log-likelihood value of each single simulation. Log-likelihood function value is an intuitive way to see whether an empirical model fits well or not. Combining properties of logarithm and likelihood function, the more the log-likelihood value approaches zero, the higher probability of maximized likelihood it will obtain, and furthermore, the better the empirical model will fit as is rudimentarily considered.

4.3 Hawkes Intensities Achieving and Statistical Detection

Empirical Intensity Processes

Being acquired parametric estimations, one is able to draw portraits of intensities by exploiting Equation 3.46 and 3.48.

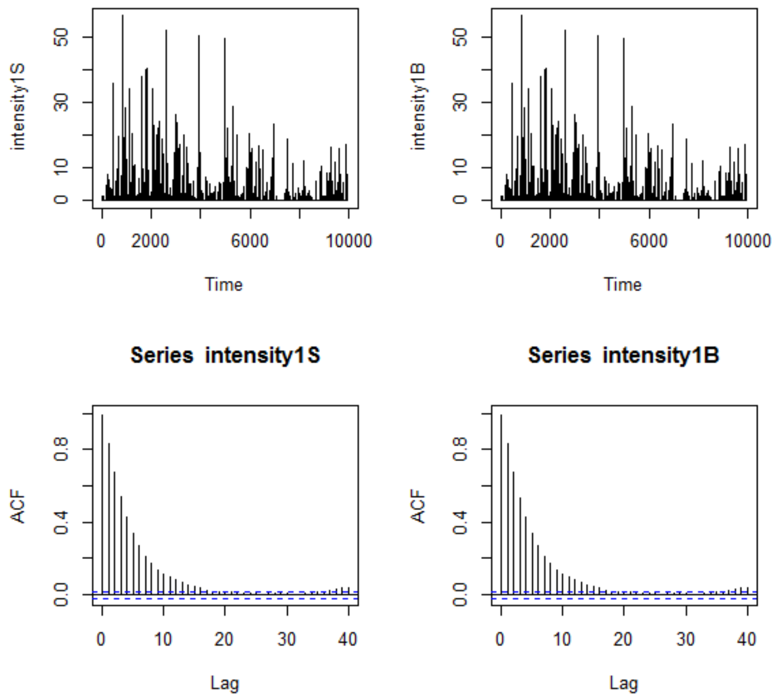

Figure 4.1: Intensities of Orders at Side of Sell and Buy (09:30-10:00)

Figures 4.1, 4.2 and 4.3 amply depict the intensities of orders, market trades and cancellations during the half hour period starting from 09:30 separately. Here the upper rows of figures are time series of intensity functions which are observed by length of 10,000 and the lower ones represent the corresponding discrete autocorrelation functions. Since the amount of market buy and sells are at a lower level, intensities changes are not as frequently as those of orders and cancellations do. Although in each figure, one side intensity for same event type seem to be same as the counterparty, the differences are obvious when enlarging the plots.

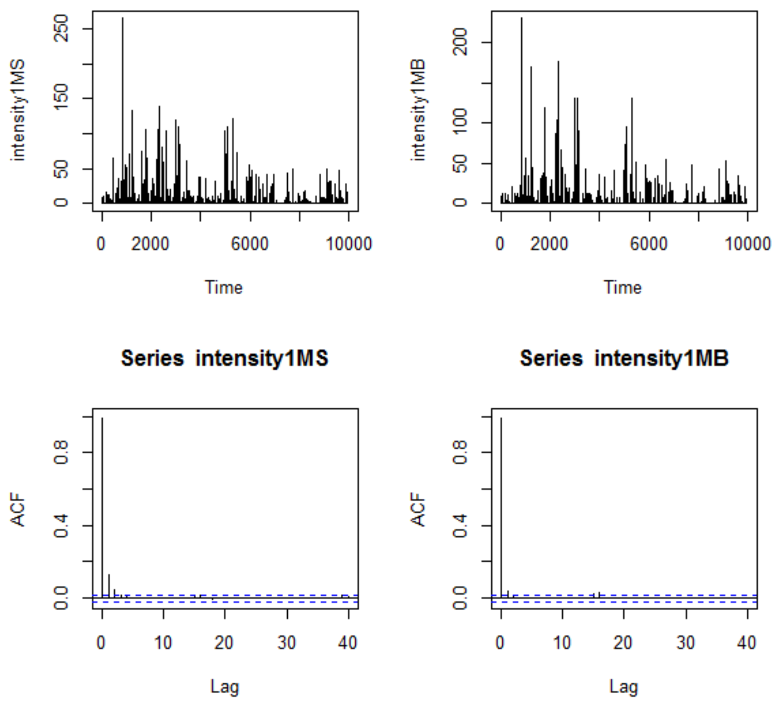

Figure 4.2: Intensities of Market Sell and Buy (09:30-10:00)

Sequently come the discussions of the ACF plots. Most ACFs' decay fiercely after the first lag but jump out of significance interval after every following several lags which indicates an event's successive and decrescendo impact to future recurrences. Exceptionally, ACFs' of market bid and ask seem not jump out of significance bound because of low quantities of market trades or low activeness. Plots for the other seven tests are similar to those of test 1 and are presented in Appendix A.

Spectrum

Although the ACF plots do intuitively reflect the self-excitation and exponentially decaying of Hawkes process, they are discrete distributed. Instead, people could observe other continuous properties of each intensity by straightforwardly looking into the theoretical model itself.

Joseph Fourier was the first one who decomposes a stationary function into sum of sinus and cosine functions and this skill are widely used in spectral analysis which emphasizes on time series' frequency instead of time. Subsequently, spectral density $f(\omega)$ of autocovariance function $\gamma(n)$ with respect to random process X_n is defined as

$$f(\omega) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} cos\,(n\omega)\,\gamma(n)\,.$$

Conversely, the autocovariance function $\gamma(n)$ can be derived as

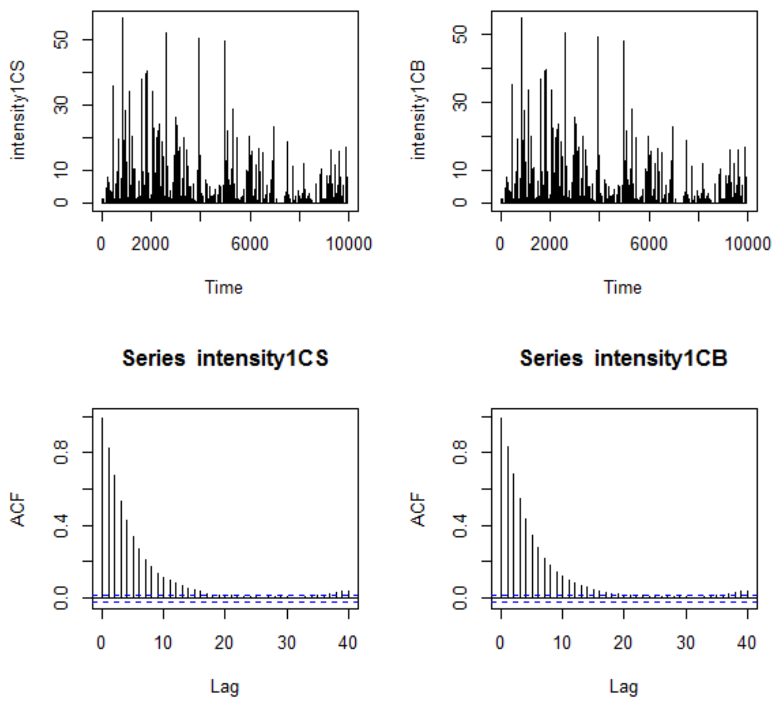

Figure 4.3: Intensities of Cancellations Sell and Buy (09:30-10:00)

$$\begin{aligned}\int_{-\pi}^{\pi} e^{in\omega} f(\omega)\, d\omega &= \int_{-\pi}^{\pi} e^{in\omega} \cdot \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} cos(k\omega)\, \gamma(k)\, d\omega \\ &= \int_{-\pi}^{\pi} e^{in\omega} \cdot \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\omega} \gamma(k)\, d\omega \\ &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) \int_{-\pi}^{\pi} e^{i(n-k)\omega} \gamma(k)\, d\omega \\ &= \frac{1}{2\pi} \sum_{k\neq n} \gamma(k) \cdot 0 + \frac{1}{2\pi} \cdot \gamma(n=k) \cdot 2\pi \\ &= \gamma(n) \quad \forall n.\end{aligned} \tag{4.7}$$

Thus, by relationship between ACF and ACVF, autocorrelation function $\rho(n)$ is

$$\begin{aligned}\rho(n) &= \frac{\gamma(n)}{\gamma(0)} \\ &= \frac{\int_{-\pi}^{\pi} e^{in\omega} f(\omega)\, d\omega}{\int_{-\pi}^{\pi} f(\omega)\, d\omega}.\end{aligned} \tag{4.8}$$

Readers could look through Brockwell PJ and Davis RA. (2002) and find more knowledges about Fourier representation and analysis in the frequency domain. By case of the Hawkes process, Hawkes (1971 and 1974) deduces the formula for spectral intensity which is continuously extended

$$f(\omega) = \frac{\mu \cdot \beta}{2\pi \cdot (\beta - \alpha)} \left\{ 1 + \frac{\alpha (2\beta - \alpha)}{(\beta - \alpha)^2 + \omega^2} \right\}. \tag{4.9}$$

As a consequence, ACF can be presented as

$$\begin{aligned} \rho(n) &= \frac{\int_{-\pi}^{\pi} e^{in\omega} f(\omega)\, d\omega}{\int_{-\pi}^{\pi} f(\omega)\, d\omega} \\ &= \frac{\int_{-\pi}^{\pi} e^{in\omega} f(\omega)\, d\omega}{\int_{-\pi}^{\pi} \frac{\mu\cdot\beta}{2\pi\cdot(\beta-\alpha)} \left\{ 1 + \frac{\alpha(2\beta-\alpha)}{(\beta-\alpha)^2+\omega^2} \right\} d\omega} \\ &= \frac{\int_{-\pi}^{\pi} e^{in\omega} f(\omega)\, d\omega}{\left[\frac{\mu\cdot\beta}{2\pi\cdot(\beta-\alpha)} \cdot 2\pi + \left[\frac{\mu\cdot\beta}{2\pi\cdot(\beta-\alpha)} \cdot \alpha(2\beta-\alpha) \right] \frac{-2\omega}{\left((\beta-\alpha)^2+\omega^2\right)^2} \right]_{-\pi}^{\pi}} \\ &= \frac{\int_{-\pi}^{\pi} e^{in\omega} f(\omega)\, d\omega}{\frac{\beta\mu}{\beta-\alpha} + \frac{\mu\alpha\beta(2\beta-\alpha)}{2\pi(\beta-\alpha)} \cdot \frac{-4\pi}{\left((\beta-\alpha)^2+\omega^2\right)^2}} \\ &= \frac{\int_{-\pi}^{\pi} e^{in\omega} f(\omega)\, d\omega}{constant}. \end{aligned} \tag{4.10}$$

Thus, spectral function theoretically supplies autocorrelation function, and

$$\begin{aligned} f(\omega) &= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\omega} \gamma(n) \\ &= \frac{\gamma(0)}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\omega} \frac{\gamma(n)}{\gamma(0)} \\ &= \frac{\gamma(0)}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\omega} \rho(n) \\ &= constant \cdot \rho(n) \sum_{n=-\infty}^{\infty} e^{-in\omega}, \end{aligned} \tag{4.11}$$

suggests that $f(\omega)$ a one-to-one function of theoretical ACF. As a result, trajectory of spectral function exactly reflects properties of detected model's from the aspect of frequency. For instance, by definition of Hawkes process, conditions of stationarity (see 3.45) requires $\sum_i \frac{\alpha_i}{\beta_i} \leq 1$ which is fulfilled by all results in test1-test8. Hence, spectrum plots of the model are obtainable.

Graphs of theoretical spectral density $0 \leq \omega \leq \pi$ for test 1 are displayed in Figure 4.3. Observe that for each intensity function the spectral density is large for low frequencies

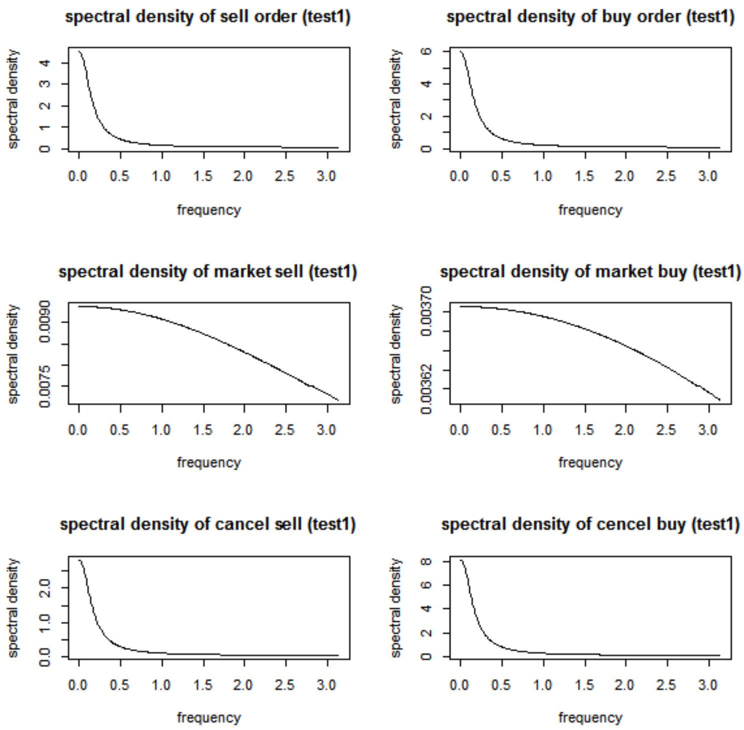

Figure 4.4: Spectral Density of Model (09:30-10:00)

and get smoother with relatively less high-frequency components. By theorem of spectral analysis, this indicates that the process has a positive autocorrelation function initially with a large value at lag one, leading a large contribution from low-frequency components as reflected by the size of the spectral density near null frequency, and then descends with smaller contributions. Situations of test 2-8 are similar to that of test 1 and the spectral densities are attached in Appendix B. Although spectral density does not illustrate corresponding ACFs directly, but it is continuously ranged and since spectrum is an important specialization in depicting a specific time series, one can express the model's ACF quantitatively and judge the trend of ACF through studying homologous spectral density instead.

Accuracy and Efficiency

An intuitive way to observe the accuracy is to detect value of log-likelihood which is mentioned in Section 4.2. Meanwhile, there exist unexpected sources of error, except the modelling error, wrong data, etc.

First problem is that the raw timestamps are recorded as amount of nanoseconds after midnight which are huge numbers, and one obtains MLE by minimizing the log-likelihood function 3.52, 3.53 or 3.56, 3.57. Notice that there are negative scaled exponential functions with respect to event arrival times, misleading computer evaluate it as zero when substituting timestamps with unit of nanosecond. Thus, one tackles this problem by changing the unit, for instance, into second. We throughout exploit a statistical programming software named R which is scientifically precise only six numbers after the decimal point. Thus, the accuracy gets lower when transforming the unit. Additionally, there exist supernumerary errors when doing optimizations which can not be avoided.

Efficiency, subsequently, is taken into discussion. Both one-variate and multi-variate log-likelihood functions can be employed in the multi-dimensional model in observation, but efficiencies are differentiated largely. It takes in average about 10 minutes' CPU time to estimate MLE based on Equations 3.52, 3.53 and 3.54, but over 72 hours per simulation when 3.56 and 3.57 are in use. It is known that the more the summation functions are, the longer the computation time will spend. Therefore when programming the codes, one has to spare no effort simplifying the log-likelihood formulas and exploit superposition of RPP and programming skills adequately.

4.4 Event Arrival Forecasting

An important application of Hawkes processes is its predictability and reachable forecast. Equations 3.58 and 3.59 support idea of implementing prediction. Next possible event arrival t_{n+1} is proved to satisfy 3.58 or 3.59, where ϵ_{n+1} (or ϵ^k_{n+1}) is an i.i.d. standard exponentially distributed random variable.

To consider the moments probability is the simplest but useful idea to find point estimate. Particularly think over the first sample moment $m_1 = \frac{1}{h}\sum_{i=1}^{h} t_{n+1,i}$ and first population moment $\mu'_1 = \mathbb{E}\left[t_{n+1}|\mathcal{F}_{t_n}\right]$. By central limit theorem (CLT), when the sample size is large enough, i.e., h attains a relatively high level, m_1 can be believed equal to μ'_1. Since task of forecast is to calculate $\mathbb{E}\left[t_{n+1}|\mathcal{F}_{t_n}\right]$, one achieves prediction for next event time by taking average value of h 3.58 or 3.59 solutions.

Being restricted by software, some languages are not compatible with generating i.i.d. standard exponentially distributed random variables directly. As is known to all, the *c.d.f* of exponential distribution is given by

$$F(x) = 1 - e^{-x}, \quad x \geq 0,$$

then solve $u = F(x)$ in term of x and obtain

$$x = F^{-1}(u) = -\ln(1-u)$$

It is clear that the *c.d.f* function is in $[0,1]$. Let $u = F(x)$ be a random variable. By the inverse-transform method, one could generate an i.i.d. random variable which follows exponential distribution via standard uniform random variable U, i.e., U is a random number in $(0,1)$ with also $1 - U \sim Unif(0,1)$:

· Generate a random variable U with $U \sim Unif(0,1)$

· Return $\epsilon_{n+1} = -\ln(U)$.

With all preparations completed, it is sufficient to quest next timestamp for each single type of event. By creating i.i.d. $\exp(1)$ random variable h times, it is able to implement and solve the nonlinear equation 3.58 or 3.59, therefore retrieve result of prediction as the average of all solutions to h trials which are different with each other due to randomness of ϵ_{n+1}'s. In this case, restricted by CPU and time, $h = 10000$ is taken.

09:30-11:00

test	*type*	*last event time*	*next arrival time*	*prediction*	*error*	*variance*
test1						
09:30-10:00	sell order	35993.36	36000.03	36005.52	+5.49	647.94
	buy order	35996.88	36001.76	36001.27	-0.49	167.31
	market sell	35988.76	36020.36	36017.24	-3.12	5144.64
	market buy	35988.12	36067.48	36039.45	-28.03	16947.14
	cancel sell	35993.31	36000.95	36010.9	+9.95	1376.71
	cancel buy	35989.63	36001.95	36003.68	+1.73	110.84
test2						
10:00-10:30	sell order	37765.83	37803.45	37798.4	-5.05	8368.74
	buy order	37786.08	37802.02	37801.96	-0.06	921.57
	market sell	37765.83	37836.07	37921.31	+85.24	31966.04
	market buy	37786.08	37808.4	37944.97	+136.57	23577.37
	cancel sell	37769.49	37803.45	37816.11	+12.66	1771.27
	cancel buy	37770.93	37802.19	37801.13	-1.06	66.35
test3						
10:30-11:00	sell order	39599.91	39601.52	39601.15	-0.37	39.34
	buy order	39599.54	39601.65	39607.17	+5.52	912.38
	market sell	39531.86	39604.9	39532.7	-72.2	309.19
	market buy	39599.54	39602.52	39680.72	+78.2	39940.66
	cancel sell	39599.54	39600.55	39606.96	+6.41	270.32
	cancel buy	39599.65	39602.38	39623.96	+21.58	4405.73
test4						
09:30-11:00	sell order	39599.91	39601.52	39601.57	+0.05	7.93
	buy order	39599.54	39601.65	39600.47	-1.18	10.84
	market sell	39531.86	39604.9	39600.89	-4.01	212.83
	market buy	39599.54	39602.52	39635.19	+32.67	7571.09
	cancel sell	39599.54	39600.55	39604.61	+4.06	84.83
	cancel buy	39599.65	39602.38	39601.47	+0.91	23.07

Table 4.1: Prediction Result (Morning)

12:00-13:30

test	*type*	*last event time*	*next arrival time*	*prediction*	*error*	*variance*
test5						
13:00-13:30	sell order	48599.94	48600.08	48609.55	+9.47	980.94
	buy order	48599.92	48604.08	48600.01	-4.07	1.24
	market sell	48553.05	48695.09	48835.84	+140.75	116278.1
	market buy	48599.86	48695.11	48688.84	-6.27	47703.54
	cancel sell	48599.9	48604.13	48599.97	-4.16	0.006
	cancel buy	48599.95	48606.48	48606.77	+0.29	812.08
test6						
12:00-13:30	sell order	48599.94	48600.08	48600.2	+0.12	31.06
	buy order	48599.92	48604.08	48599.99	-4.09	0.007
	market sell	48553.05	48695.09	48684.15	-10.94	47030.29
	market buy	48599.86	48695.11	48639.94	-55.17	18093.25
	cancel sell	48599.9	48604.13	48599.97	-4.16	0.006
	cancel buy	48599.95	48606.48	48601.9	-4.58	58.63

Table 4.2: Prediction Result (Midday)

14:00-15:30

test	*type*	*last event time*	*next arrival time*	*prediction*	*error*	*variance*
test7						
15:00-15:30	sell order	55794.87	55800.36	55801.14	+0.78	114.93
	buy order	55793.12	55801.67	55804.56	+2.89	487.26
	market sell	55757	55900.62	56039.05	+138.43	93681.21
	market buy	55767.64	55921.48	55802.54	-118.94	16788.81
	cancel sell	55790.58	55801.69	55792.28	-9.41	31.89
	cancel buy	55796.24	55827.04	55807.58	-19.46	475.82
test8						
14:00-15:30	sell order	55794.87	55800.36	55798.83	-1.53	10.57
	buy order	55793.12	55801.67	55795.87	-5.8	23.84
	market sell	55757	55900.62	55869.99	-30.63	26243.99
	market buy	55767.64	55921.48	55800.07	-121.41	2574.44
	cancel sell	55790.58	55801.69	55791.18	-10.51	2.38
	cancel buy	55796.24	55827.04	55800.52	-26.52	33.01

Table 4.3: Prediction Result (Afternoon)

Results of forecast are formulated in Tables 4.4, 4.4 and 4.4, in which the first column of numbers denote the latest event recurrence (seconds after midnight) during observed time period. For example, for 09:30-10:00 in test 1, the first column notes largest numbers before $10 * 3600 * 10^9$ with respect to specific event types. Following column recorded the actual event arrival time t_{n+1} whereas the right hand side of it denotes the estimated next occurrence which is algorithmic average of 10,000 experiments. Since the model is 6-dimensional, each test supplies estimations for six different event types: bid/ask order, market buy/ sell and cancel buy/sell. The column named after "error" means the discrepancy between prediction and true timestamp of recurrence where the positive (negative) symbol indicates an overestimation (underestimation). In the end, "variance" notes the fluctuation of sample, i.e., variance of 10,000 predicted candidates.

When approaching the solution to nonlinear equation 3.58 or 3.59 by exploiting standard Newton-Raphson method with help of functions in R, one has to choose an interval in which the optimizer hunts for the most approacher. We select $[t-360, t+720]$ for all prediction objects, where t is the right border of each test, in test 1 for example, t is $10:00 = 10 \cdot 3600s$. The variances slump if people look into shorter intervals $[t_n, t+720]$ starting from the latest informed event time t_n, thus, variances in Table 4.4, 4.4 and 4.4 could be cut back by curtailing investigated intervals, with forecasts moving upward slightly.

Variance is possibly goes to quite large scale which is always a signal of dangerous. For example, predictions for market sell and buy in test 2 lead relatively great discrepancy with true value with large variances which is 31966.04 and 23577.37 respectively. A main reason why the predictions here are especially inexact is that among quantity of 9006 data (see Section 4.2) only 236 event times are of type market sell and 263 belong to market buy series. Recall that in Section 2.2, Figure 2.4 and 2.3 suggest a much lower proportion executions have than cancellations do. Market sell/ buy events are even a subset of executions, making market trades less than other type of events essentially. Thus, the simulated results for market buy/ sell are expected (but not to a certainty) to be worse than those for other event types generally.

Being irregularly-spaced, the lag between timestamps of the same type floats from 1 second up to over 150 seconds. On the other hand, the intensity series fluctuate severely with time, leading somewhere fit empirical data better whilst somewhere worse. These properties determine that only comparisons between forecasts of the same recurrence (same true next event arrival of same type) meaningful, i.e., comparisons between test 3 and 4, test 5 and 6, test 7 and 8.

Being compared with forecasts in test 3, the fourth experiment performs better in its entirety. Test 4 substitutes 30261 data into the model, reducing sharply by at least one third of the biases in test 3. Furthermore, variances of almost all types decrease by factor of ten or more. Especially for the cancel buy intensity, when forecasting the next event arrival after 11:00, given one and a half hours' data, the error of prediction reaches +0.91s with variance for 10,000 trials at 23.07, which is much better than indices of the counterparty with +21.58s and 4405.73, making it more precise by 20 and 200 times separately.

Similar conclusions can be mostly summarized when comparing test 5 and 6, test 7 and 8, but some special cases have to be mentioned. In contradistinction to market buy intensity in test 7, 8th experiment's forecast variance subsides whilst the error booms contrarily from +118.94 to -121.41 when observed time interval enlarges. To loose the error is to lose some accuracy, whereas larger oscillation amplitudes are very dangerous when applied in the industry. The mean square error (MSE) is widespread in deciding a better estimation by choosing a smaller $Bias^2 + Variance$. It is simple to work out that $118.94^2 + 16788.81 = 30935.5336 > 17314.8281 = 121.41^2 + 2574.44$, advantages result for market buy in test 8. Analogously one can figure out that all outcomes in test 5 (test 7) are much better than those in test 6 (test 8) based on MSE when above class of special cases engender. In conclusion, when predicting a specific time point, the more backward information is encompassed, the better results will turn out.

Specially when comes to cancel sells in test 5 and 6, the results are exactly the same, intimating that it is sufficient to estimate the exact event arrival time right after 13:30 with only latest half-hour information according to this Hawkes model 4.2. The fairly tiny sample variance (0.006) signals an utmost good approach to specific actual value 48604.13 of this model. Nevertheless, the 4.16s error by this case shows that there remains space retrofitting Model 4.2.

Chapter 5

Trading Strategies

Running away from maddening mathematics for a while, this chapter discusses about general HFT strategies. General trading strategy methods are talked without concrete ones because trading strategies which mostly are combinations or evolutions of the general trading strategies are commercial secrets in each single HFT firm. Since all analyses are based on NASDAQ trading platform, trades follow the trading principles and priority mentioned in Section 1.2.

5.1 High-frequency Trader

As more details can be seen in Durbin (2010), securities traders are generally assorted into four types: the market-makers who earn money from the bid-ask spreads and rebates, meanwhile undertake the market making risks; the predictors who profit from accurately enough capture the future prices change and event arrival time by quantitative means; the investors who possess the securities for relatively long time and the arbitrageurs who seek profit by trading abnormally priced stocks. Liquidity provider (i.e., market-maker) posts new pegged bids and offers to make liquidity, waiting for a liquidity taker to hit the prices and pay fees to the exchange. In this way, the market-maker obtains rebate from the exchange and makes money from spreads. Thus, the relationship between market-makers and investors are quite close and conflicting: market-makers need investors to take their bids such that offer prices would rise so they can earn more from enlarged spreads and higher rebates, whereas investors need market-makers to ensure the existence of liquidity for them to hit and the spread is as narrow as possible in the meantime such that they would not pay too much on the market spread.

Irregularly spaced high-frequency trades occur orderly in pretty short periods of time, leading long-term trading strategies be invalid. Presently the high-frequency trader is a sort hybrid of the former two types, i.e., market-maker and predictor. Particularly the predictor is obviously a short-term predictor owing to the same reason as that for invalidity of long-term strategies. For instance in this thesis, the market-makers are recognized as high-frequency traders and investors are people who want to earn money by participating in trading activities (but not a sort of security traders mentioned above).

5.2 Investor Strategies

Although this thesis is in the trader's angle of view, both tricks that market-maker and investor play at each other are supposed to know.

An investor always trades with small volumes, avoiding pushing the market against him and influencing his forecast, trying to trade whole quantity of securities with most satisfactory prices. He is most willing to trade on BBO such that he would not pay for the market spread but this situation rarely happens. Following are staple investor strategies in use based on the reality.

The simplest strategy is to simply enter the market with a market order, being willing to execute with the instant posted bid/ ask quote. As is specified in Section 1.2 and 5.2, this willingness is not guaranteed to be realized in actual market. Orders stand in the same queue should be executed according to order priority in specific platform, thus they may be matched after the market changing for a while. In supplementation, if the market is expected to move away from him and the order stands still, it would take longer time before execution; conversely if market seems to get approacher, this market order is traded immediately at the willing price. In this strategy, investor should pay the market spread which almost all investors would not like to, hereafter various investor strategies including flexibly exploiting the limit orders are introduced.

Since investor has the right to cancel and change his order fully or partially, he definitely can bargain with the market order. The investor states the bid/ ask order with a price a little bit higher/ lower than the bid/ ask quote and wait for someone else who is satisfied with this price. If this happens, then the investor saves money for spread and trades at a better price than the market order. Or on the contrary if not, the worst case would be the investor has to change his order back to the market order, assuming that the observed side of market price does not change when he moves the order.

NASDAQ permits occurrences of iceberg orders and non-displayed orders. Using iceberg and non-displayed orders are powerful weapons for investor. He partially or fully reserves his order and observe the dynamics of market in the darkness. These orders can not only effectively shun control of market-maker, but also be immediately cancelled, changed and executed. Notice that mostly the iceberg orders have huge sizes but display a rather small part of them. Another equally small slice are displayed once the former slice is fully executed. This step is repeated until the whole order being partially cancelled or digested by market. Since an iceberg order gets easier to be found out after several parts of it being filled, a wise idea for investor is to stop and cancel the order after several executions, then redo the remainder to another market, and so on.

In addition, the investors seek profit by finding opportunities of arbitrage. Based on each stock there exists lots derivatives, e.g., bonds, options and futures. Arbitrage can always be found by deeply look into each two sorts of derivatives. More simply, arbitrage opportunities emerge in crossed markets with non-negligible quantities: firstly, the same security is possible to be posted at difference prices simultaneously in different market; secondly, related portfolios' prices which fluctuate together may cause chance of arbitrage. Investors, or arbitrageurs who take this strategy should have extremely well knowledge in finance and derivative including the relations between each two derivatives, arbitrage and portfolio theory.

5.3 Market-maker Strategies

Strategies towards Earning More

When someone enter the market and takes a trade with a market-maker, the simplest strategy for the market-maker is to wait for another person who trades on the opposite side of his market. Since a market-maker is possibly to possess no share of security, when he waits for another liquidity taker, he would short sell/ buy the security. The prices of securities change unpredictably, making this strategy pretty risky and less profitable for the market-maker. Therefore, most market-makers would not take this strategy solo. If the security price dynamics fluctuates fiercely which happens always, lots of liquidity providers change prices and volumes more frequently whereas the best bid and offer of market-maker who takes passive strategies (for example, just wait for another side being hit) move slower than the others. At that time, this market-maker's offer (bid) price is much smaller (higher) than market price with narrower market spread because active market prices move towards the same orientation. Investors are happy with this passive price and definitely hit it rapidly, leading the passive strategic market-maker earns not much money or even lose money directly out of pocket. Therefore, unless preparing hidden active bid and asks to guard the market which get displayed and take the lead hitting the passive bids/ offers, market-makers would not directly take the passive strategies that is much riskier than active ones.

Instead, a popular method in use is to take an active strategy: when somebody enters his market, for example on the bid side, the market-maker lifts the ask quote and simultaneously decrease the bid, i.e., to the position of previous second best bid. Analogous measures can be taken when investor enters the other side of market. This strategy would cause the market move with broader bid-ask spread and the liquidity provider profits directly more from next trade on.

Similar to investors, the market-makers also have expectancy about bid and ask prices, but they generally do not set them as best buy and sell in the market at first. Instead, market-makers tend to hide their best bid and ask, putting them to the position of second best bid/ ask and investigate the reactions of market. Being attracted by the competitive bid/ offer, new trades are upcoming to take this liquidity. Meanwhile, the market-maker actively cancel best bid/ ask, making the expectancies best and furthermore, profits starting from prices that are better than market.

When there a huge volume of order enter the market with the bid (ask) quote, market-makers' eyes twinkle. Compare the market to a bone and the new pegged joiner to a doggy. The market-maker tantalizes the doggy's appetite by towing the bone towards the higher (lower) price and investigate the reaction of doggy. If the doggy is attracted, it would follow the bone and jump to the same price level. This step can be replicated for several times if doggy really has desire for the bone. The ideal situation is that the doggy eventually jumps no lower to the initial ask (bid) level and eat up the accumulated shares all this way up. This strategy is really pretty risky, because the doggy may change its mind at any time and walk away, making market-maker sell off (or pay high price to buy) his shares at a loss fully or partially. This method that is called "push the elephant" with the huge volume of order known as an elephant is usually applied in once the market-maker captures an iceberg order.

For trading strategy regarding the rebate, one should familiar with rebate rules in each platform first. The larger (total) quantities of trades there are, the more rebates the market-maker will receive. Theoretically, the exchange gives rebate to market-maker for every fixed amount of shares. If the bid-ask spread is rarely small, the market-maker can possess the whole best ask himself such that other people have to raise the offer when they enter the ask side. When this is done, market-maker then release the previous best ask he owns by inputting it to the bid side. Being no smaller than best bid, previous best ask relocate into market as the new best bid. In this way, even if the market nearly reaches BBO at that time, the market-maker earns rebates on both sale and purchase trades.

Strategies towards Less Risk

Essence of most strategies introduced in last section is to play tricks of compensation such that market keeps active and the market-maker is able to conduct round-trips of bid/ ask interchanges successfully. Unfortunately when the market moves rapidly, completed round-trips are not guaranteed. Things get worse if at the same time the market tends to against market-maker's willingness too quickly so that he cannot close his bids/ asks in time, then his loss is not bounded.

Strategy of taking less risk regarding a passive one is instructed in the same paragraph which effectively protects market-maker's benefits from losing too much directly out of his pocket. Meanwhile, appropriate mutations of other strategies mentioned above can restrict the possible loss of market-maker to a bounded or smaller scope.

A widely used strategy called penny jump can to some extent protect market-maker from loss too much. Begin with a relatively large volume of bid (offer) enters the market at the bid (ask) quote, market-maker pulls a small quantity of bid (ask) 1 cent up (down) which makes the moved bid (ask) is better than the others, being executed in a short time. Then the market-maker stares at trend of market movement, if the market changes towards his willingness, obviously a round-trip is attainable; otherwise, he aborts the round-trip by setting the remaining ask (bid) back to the initial position where initial large quantity of bid (ask) stands where the exchange matches immediately and the market-maker lose $1 \cdot$*moved quantity* cents. For market-makers, it is better to loss less than to wait, therefore, market-makers' loss is limited by applying this strategy.

Also the traders can apply algorithmic means to do short-term prediction, trend following and lag estimation. Empirical significance of algorithmic trading analysis to high-frequency traders are clarified in Section 5.4.

In conclusion, there are many strategies for high-frequency traders to exploit, partially or in combination. Concrete reactions and measures a market-maker take are determined by platform trading rules and personal habit thus is difficult to describe in thesis.

5.4 Significance of Algorithmic Simulation Results

Financial high-frequency market extends at a pretty large magnitude all the time, thus, numerical simulations at least until now cannot accurately predict what the price would be in the following tick. If people get benefit from a model mostly, the model is considered as well fitted. Analogous conclusions for order book modelling regarding timestamps can be summarized with the same criterion.

Event arrival time simulation and forecast plays an important role in high-frequency trading. When being informed of enough data, high-frequency traders can predict next order, execution or cancellation arrival of each security. Being mostly a little bit underestimate next event arrival, these forecast results in Section 4.4 can lead a high-frequency trader earns money at most time. With these results, the market-maker's corresponding strategies can be planned and arrayed before next event's recurrence. For example, starting with better than market prices where market-maker exploits the strategy of hide his best prices, and then move out the bid/ ask quote exactly before the moment of upcoming order occurrence which is predicted by multi-dimensional Hawkes Process. The closer forecast to actual next recurrence is, the more successful the trading strategy will be. Furthermore, almost all strategies are determined when associating with the price upward probability estimations.

Similarly for investors, the market execution times are particularly valued because when grasping the moment of market trades, investors would when to peg his new order at what price level (equal to or better than the market order). Although the forecasts for market trades are relatively unstable and inexactly, when obtaining enough data, this situation is expected to be ameliorated.

Chapter 6

Summary and Further Studies

6.1 Main Achievements and Conclusions

After introducing fundamental concepts about trading, including order, order book, market and limit order, one begin studies with depicting and handling empirical financial high-frequency data. We filter an ultimate dataset in observation associated with major trading time and priority principles in NASDAQ OMX Stockholm platform. The entire thesis only investigates displayed orders and corresponding operations over the time duration 09:30-17:00 from an trader's angle of view.

Currently one means to simulate the series of timestamps via intensities modelling. A detailed system of theoretical instructions and proofs are necessities before empirically applied. Firstly concepts about (simple/ regular) point process and counting process are clarified, forwardly elicit definition of intensity function. With results that are easily economically interpreted, martingale method simplifies conditional expectations' calculations, leading this technique be much prior to other means of analysis. In addition, when exploiting martingale method, all likelihood-based statistical inference can be applied only if Equation 3.18 is satisfied regarding a unique probability space, which is more rigorously deduced by Karr (1986) and lays the foundation of intensity relational inference. Following by all assumptions being satisfied comes an exhaustive derivation on likelihood function (joint *pdf* function) of regular point process. This formula earliest summarized by Rubin (1972) describes the general form of log-likelihood function with respect to a specific RPP whatever intensity process model is under consideration. In accompany with the important conclusion, K-dimensional ($K \geq 2$) superposed intensities are particularly concerned and can be represented by several mono-variate intensities as in Equation 3.17 which simplifies superposed-intensity-based inference and saves numerical computation time/ space to a great extent. An important task for modelling is to forecast the next event arrival time. Martingale calculation rules and principles combined with specific assumptions guarantee the predictability of RPP. Therefore, the next event prediction is root of a nonlinear equation 3.44 obtained by employing Doob-Meyer decomposition theorem (relationship between supermartingale and martingale) and practical significance of birth process.

The Hawkes process which is the focal point of this thesis states a masterly idea of RPP modelling, based on intensity processes. For instance a six-dimensional Hawkes(1) model runs through the whole paper, composed of sell/ buy order, market sell/ buy and cancel sell/ buy intensities. Here each intensity is mono-variate because of restricted by laptop's computational capability. Reasoning from the general to the particular, since Hawkes processes are generated based on RPP, all statistical inference for RPP can be applied into the empirical model in use. One substitutes Hawkes intensities into general log-likelihood formula for RPP to obtain log-likelihood function of Model 4.2, moreover analogously, the forecasts about next events of respective types can also be approached by implementing Equation 3.58 or 3.59 which is retrieved by substituting Hawkes intensities of Model 4.2 into prediction formula of RPP.

Parametric calibrations are realized subsequently by means of MLE. Hawkes process is known as self-exciting, then as a consequence, the parameters μ's, α's and β's update when informed data extend. Also for convenience of comparisons, in this thesis, Model 4.2 yields 8 sets of parameters numerically over three different sessions (morning, midday and afternoon) including 8 time intervals. Corresponding images of intensities, ACF plots, spectral densities and next event arrival prediction thereafter are worked out and listed on Chapter 4, Appendix A and B. The acquired parameters satisfy $\alpha < \beta$ without any exceptions which guarantees the stationarity of model. Furthermore, it is specified in forecast results that when trying to prognosis a specific upcoming timestamp, the more data are informed, the better the forecast will be. Particularly by this case notice that the predictions of next market buy and sell are most unstable because of much less data under observation.

Generally speaking, the Hawkes model 4.2 captures features of irregularly-spaced financial intensities well and provides estimations of next event arrival times. Methodologically an investor is able to earn money by taking appropriate trading strategies which associates predictions of event arrival and price dynamics.

6.2 Future Research

Although Model 4.2 performs well, a lot of improvements are in need. Notice that in Section 4.4 the special case of next cancel sell recurrence time for test 5 and 6 is mentioned: the variance is quite tiny, but there still exists an error over 4 seconds. It claims that the intensities are impacted by some other variates, i.e., mono-variate intensities are not enough. One shall implement multi-variate Hawkes processes in the future and compare the results with those of mono-variate ones.

Also it is difficult to state the goodness of simulation if one merely stare at parametric estimations. The Cramér-von Mises criterion is well known as a criterion judging the goodness-of-fit, i.e., it tells to what extent the data follow the simulated model. This criterion focuses on a statistic regarding the cumulative distribution function given by the model and the empirical distribution function determined by data. It is also useful in adjusting the form of multi-variate model based on Cramér-von Mises criterion which helps omitting meaningless variates and would raise the efficiency of calculation.

In addition, since this thesis observes order book for only one security in a unit trading day, in further research several securities' order books in successive trading days are considered. Hawkes processes

$$\lambda(t) = s(t)\left[\mu + \sum_{j=1}^{P}\sum_{m=1}^{\check{N}(t)} \alpha_j \cdot \exp\left(-\beta_j \cdot (t - t_m)\right)\right]$$

or

$$\lambda^k(t) = s(t)\left[\mu^k + \sum_{l=1}^{K}\sum_{j=1}^{P}\sum_{m=1}^{\check{N}^l(t)} \alpha_j^{k,l} \cdot \exp\left(-\beta_j^{k,l} \cdot \left(t - t_m^l\right)\right)\right]$$

with seasonality factor $s(t)$ is investigated by this case. Also one shall take correlations between each single security into account. Although it takes more time and effort in numerical calculations, this advanced model theoretically would mimic the market state more accurately.

Ultimately, corresponding trading strategies based on new achieved predictions are discussed. Besides, as trading strategies are taken in association with price dynamics prediction, to choose a proper model to simulate the prices and test the entire algorithmic trading strategies are listed on an conspicuous position in further studies.

Appendix A

Intensities and Corresponding ACF Plots

Here offers intensities of test 2-8, each of whose time is divided into 5000 parts to observe for convenience, hence the exponential decrease of ACF is not obvious anymore. Restricted by layout of sheet, enlarged intensities of the same event type for both sides would not shown.

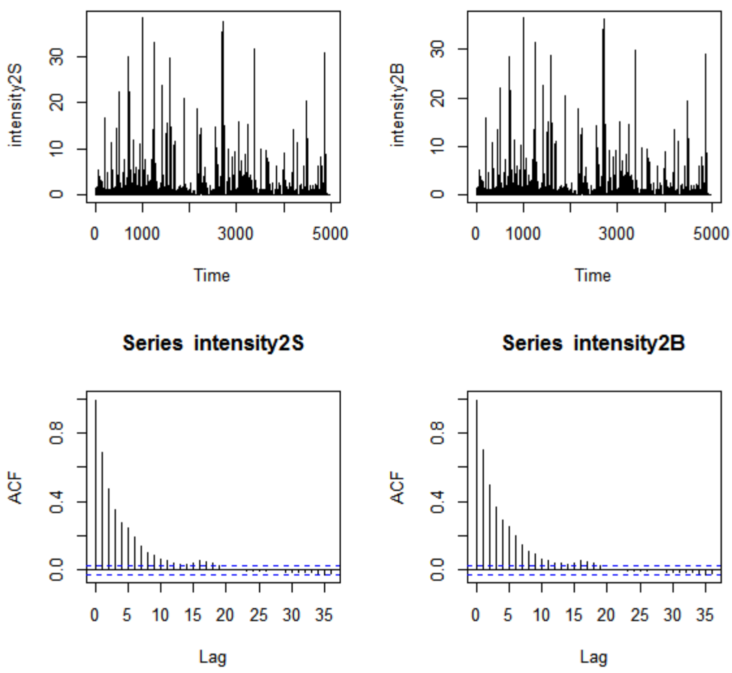

Figure A.1: Intensities of Sell and Buy Orders (10:00-10:30)

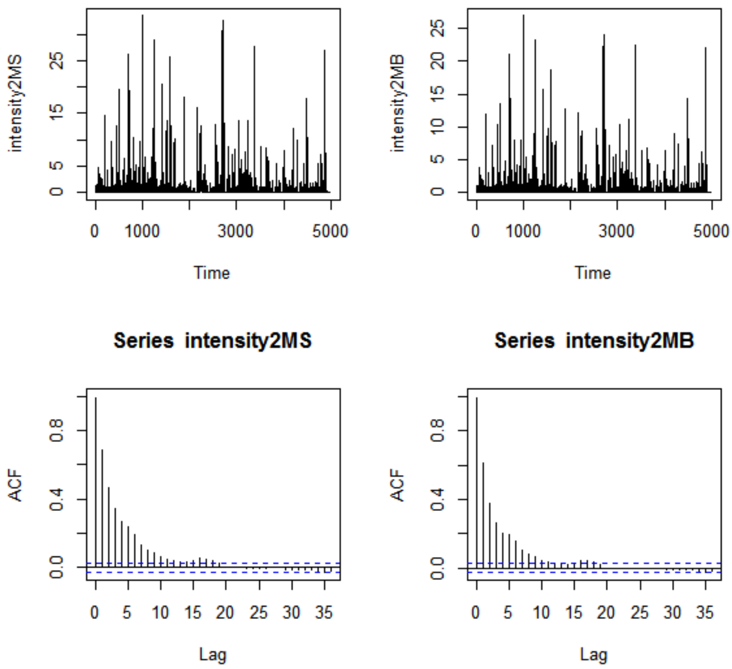

Figure A.2: Intensities of Market Sell and Buy (10:00-10:30)

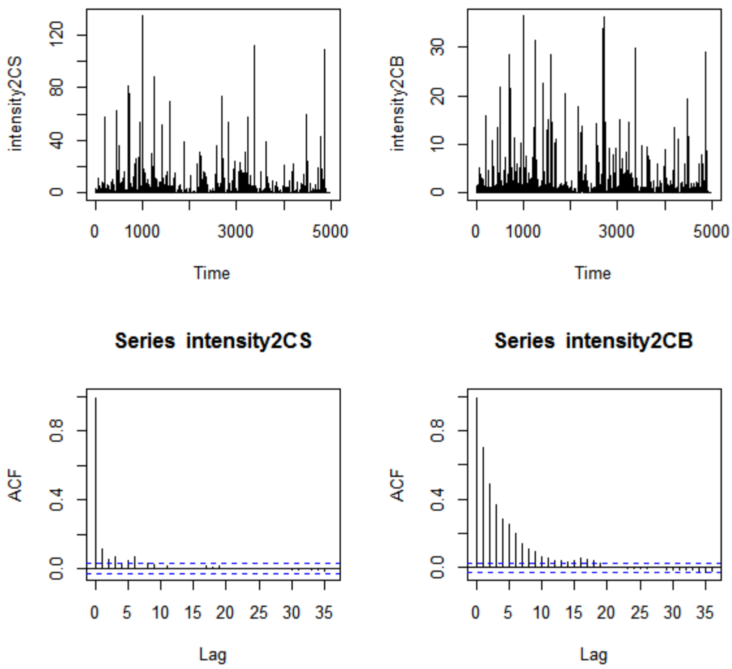

Figure A.3: Intensities of Sell and Buy Cancellations (10:00-10:30)

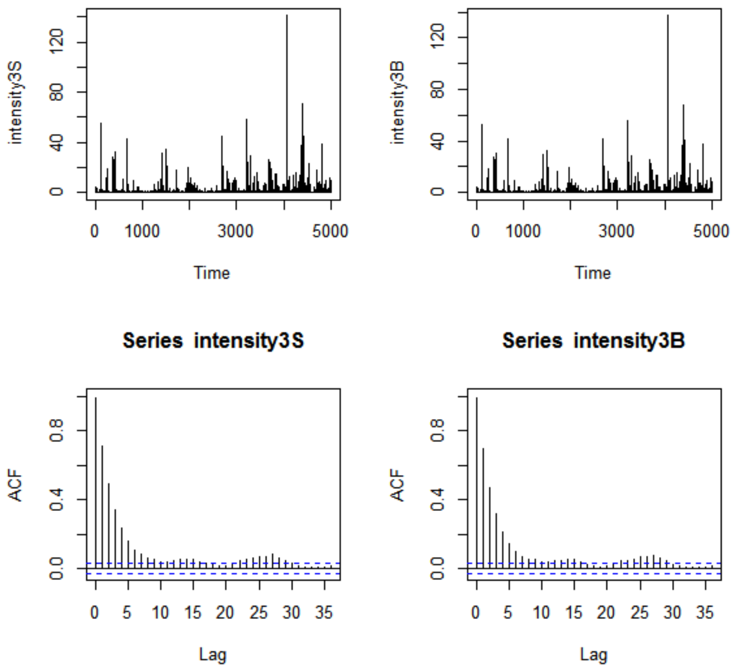

Figure A.4: Intensities of Sell and Buy Orders (10:30-11:00)

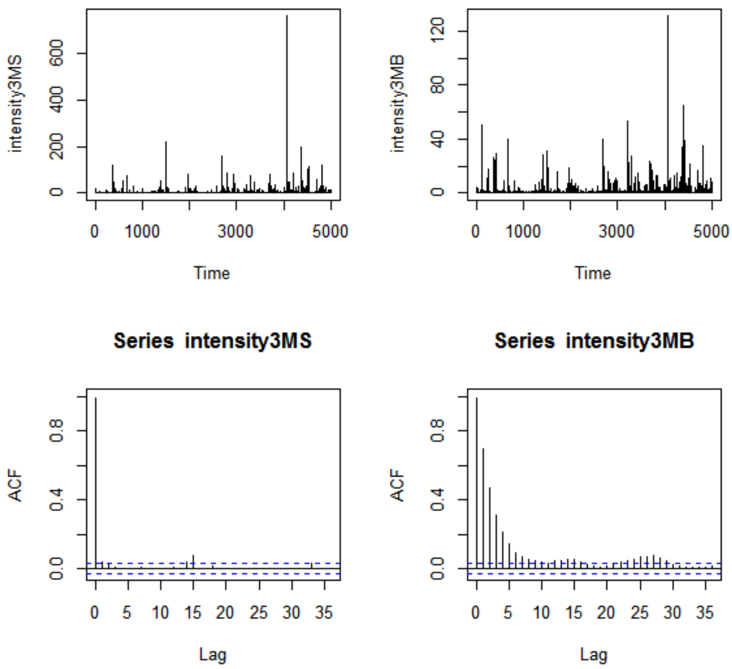

Figure A.5: Intensities of Market Sell and Buy (10:30-11:00)

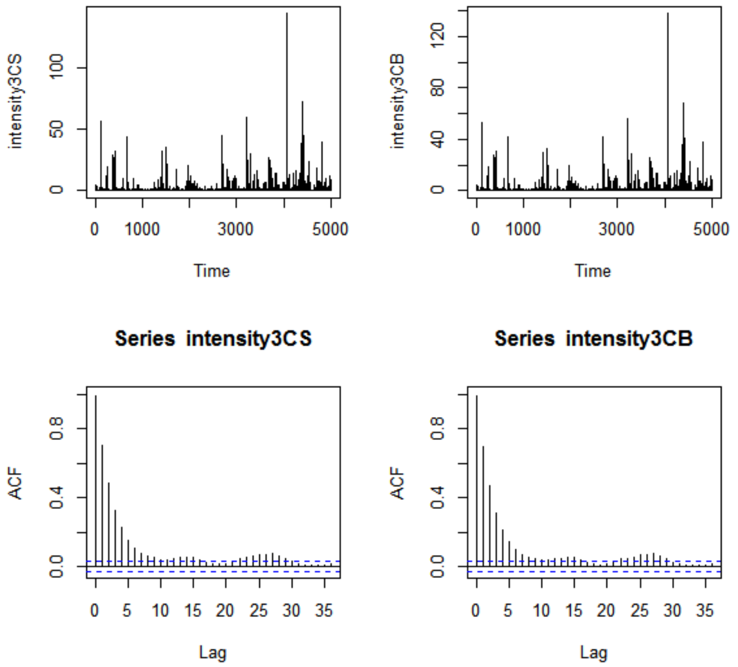

Figure A.6: Intensities of Sell and Buy Cancellations (10:30-11:00)

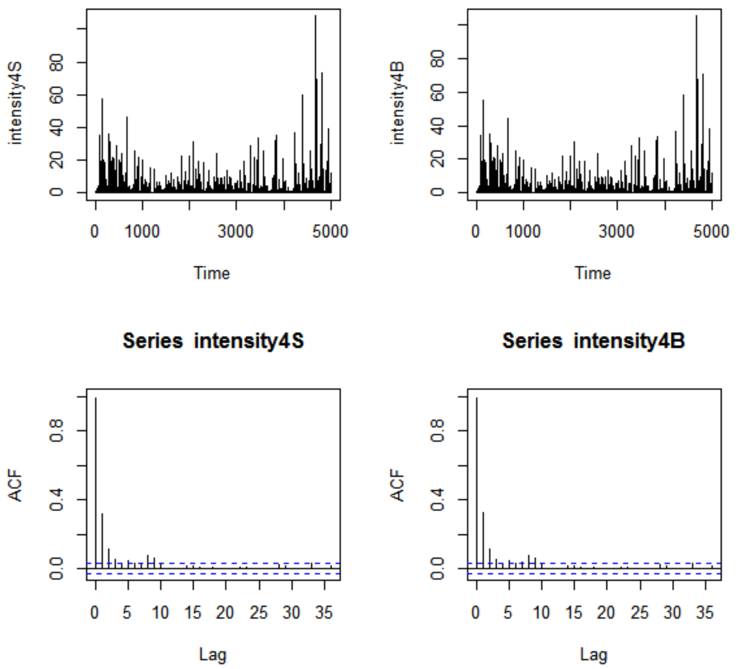

Figure A.7: Intensities of Sell and Buy Orders (09:30-11:00)

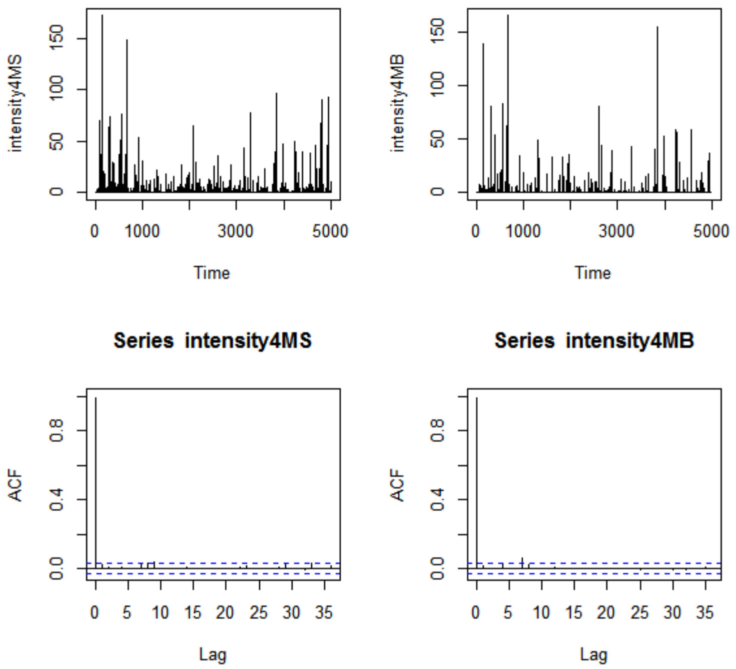

Figure A.8: Intensities of Market Sell and Buy (09:30-11:00)

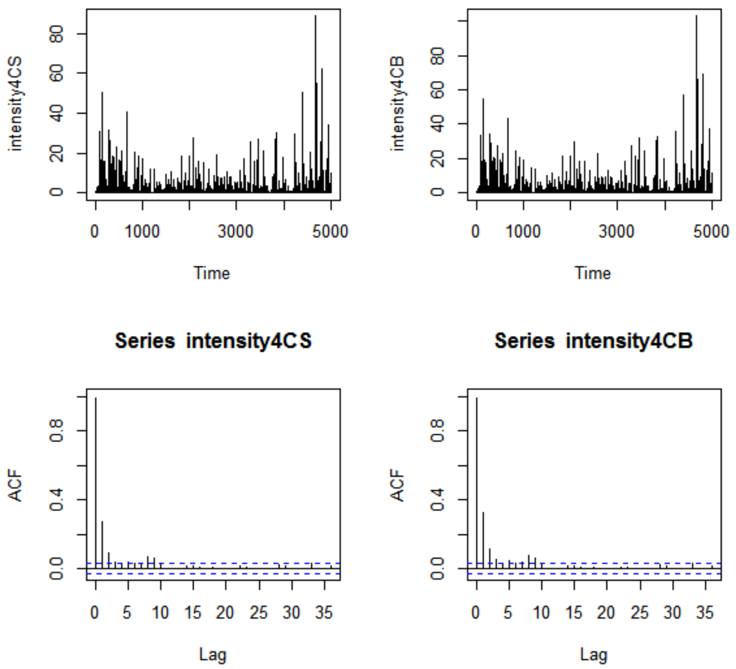

Figure A.9: Intensities of Sell and Buy Cancellations (09:30-11:00)

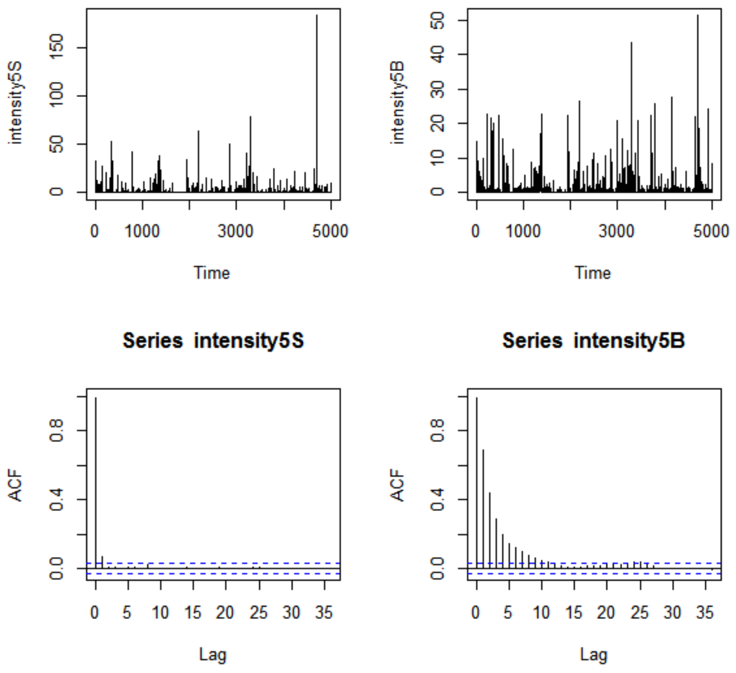

Figure A.10: Intensities of Sell and Buy Orders (13:00-13:30)

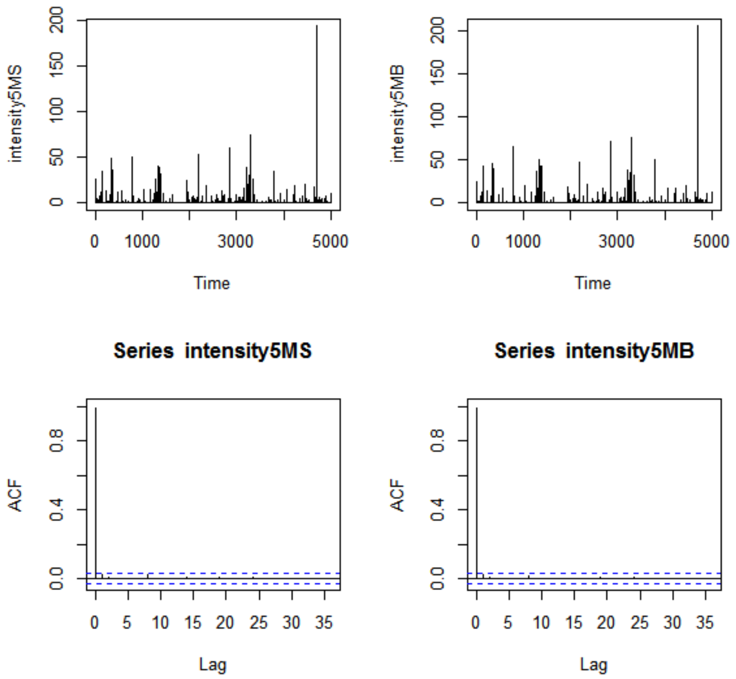

Figure A.11: Intensities of Market Sell and Buy (13:00-13:30)

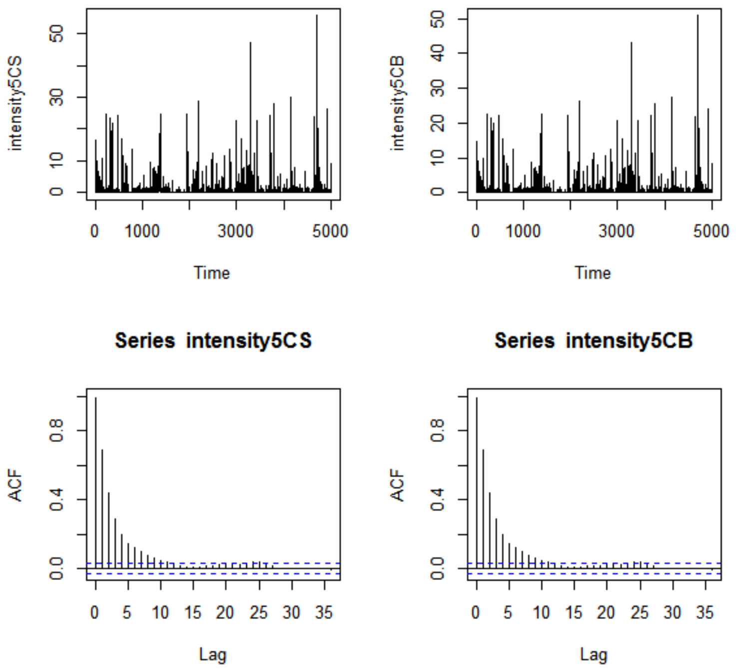

Figure A.12: Intensities of Sell and Buy Cancellations (13:00-13:30)

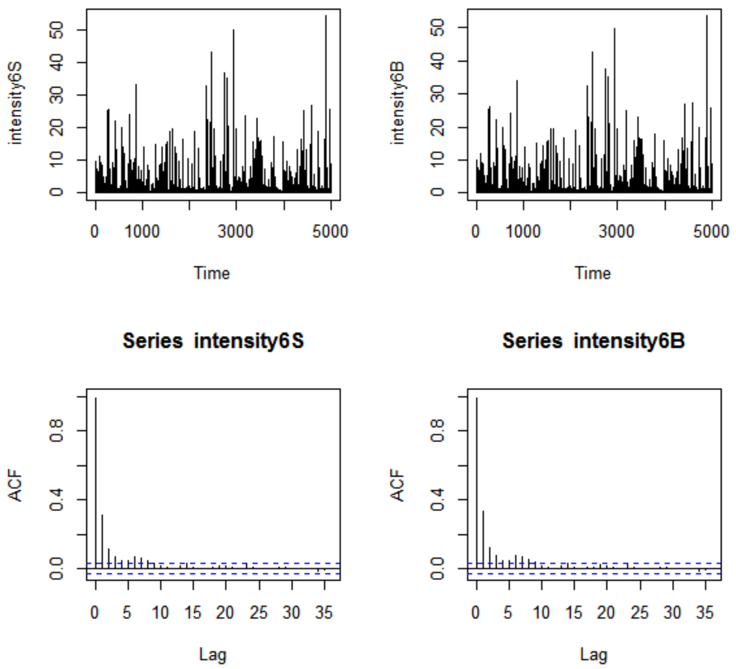

Figure A.13: Intensities of Sell and Buy Orders (12:00-13:30)

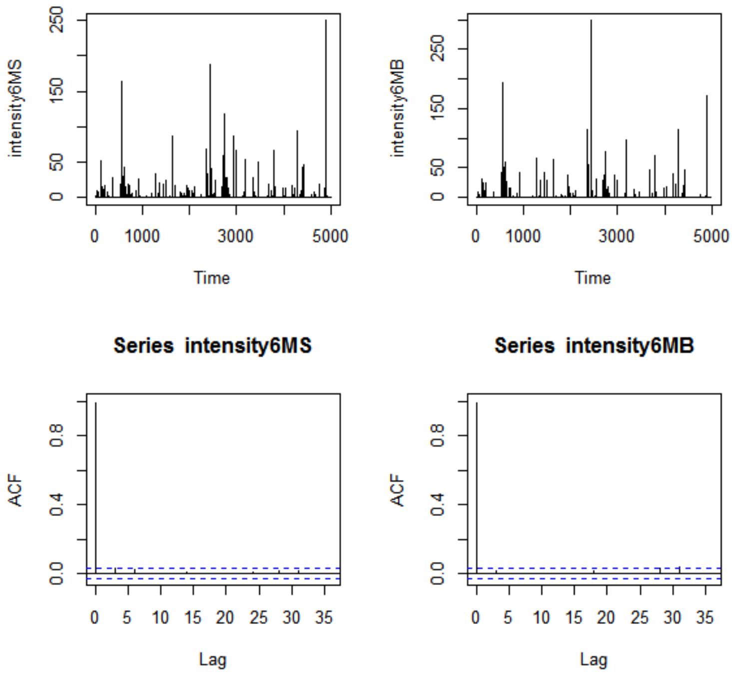

Figure A.14: Intensities of Market Sell and Buy (12:00-13:30)

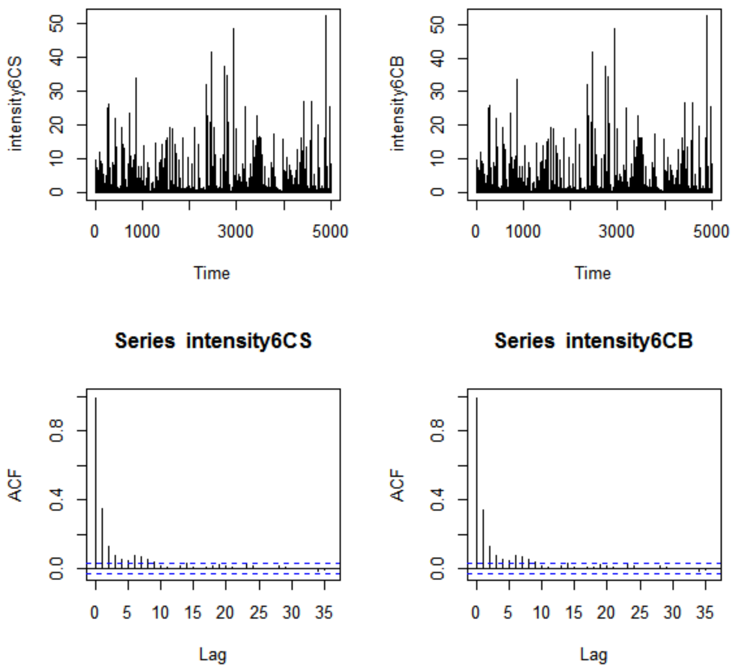

Figure A.15: Intensities of Sell and Buy Cancellations (12:00-13:30)

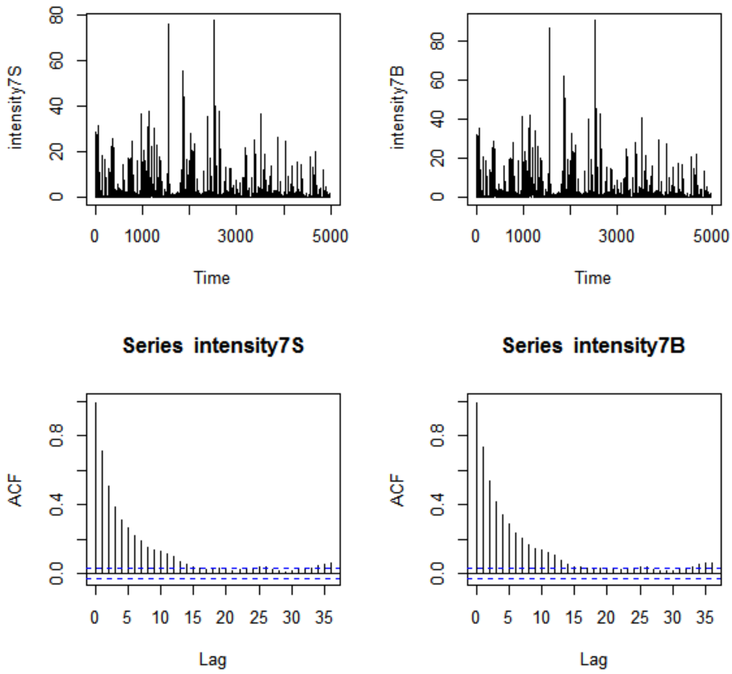

Figure A.16: Intensities of Sell and Buy Orders (15:00-15:30)

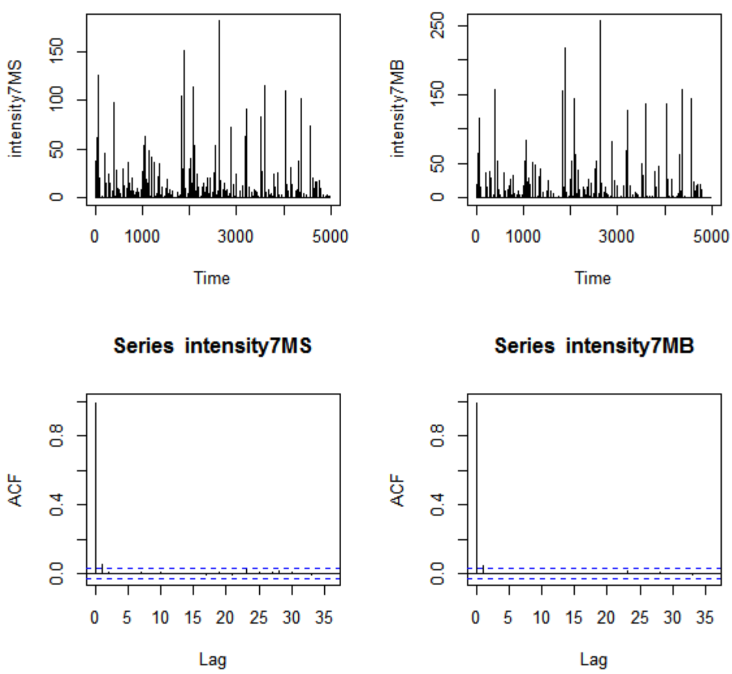

Figure A.17: Intensities of Market Sell and Buy (15:00-15:30)

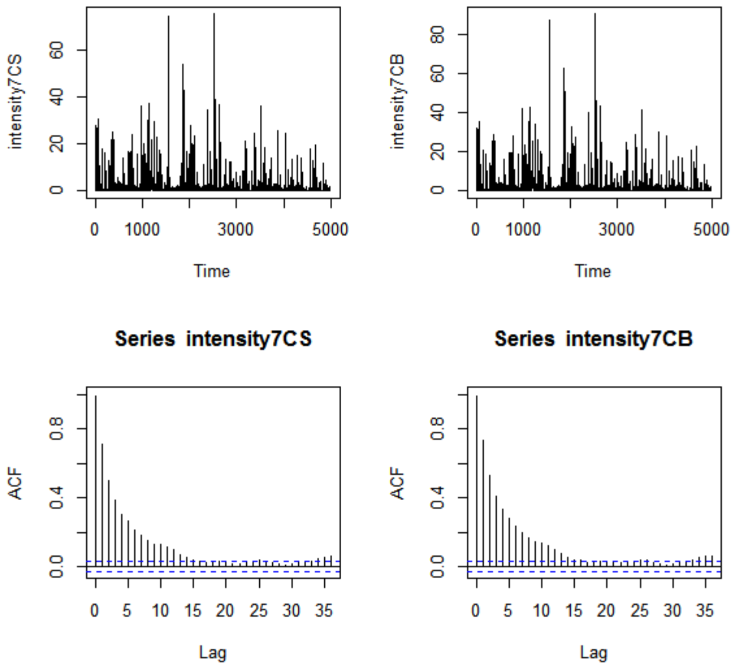

Figure A.18: Intensities of Sell and Buy Cancellations (15:00-15:30)

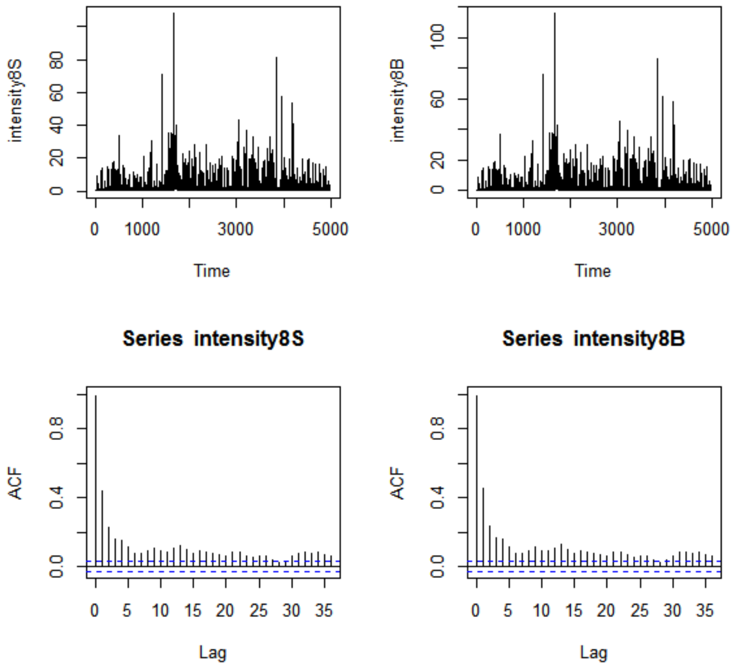

Figure A.19: Intensities of Sell and Buy Orders (14:00-15:30)

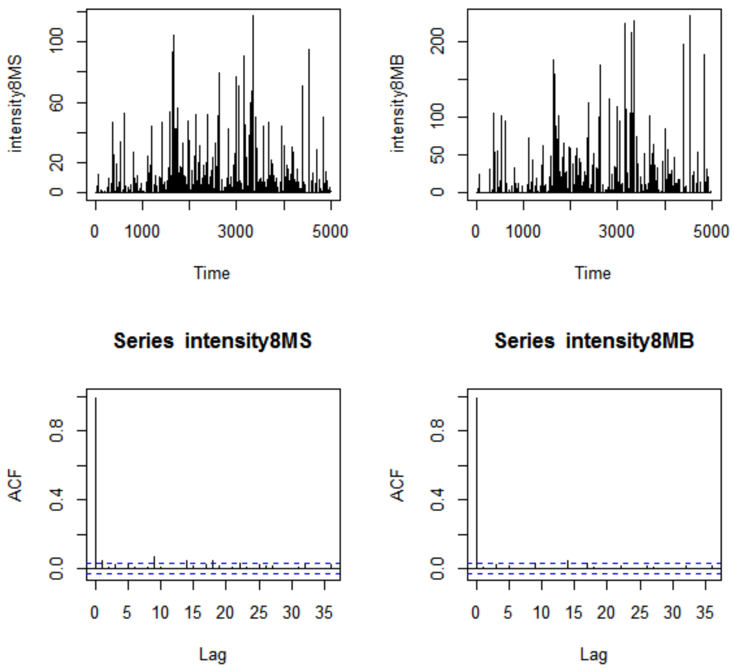

Figure A.20: Intensities of Market Sell and Buy (14:00-15:30)

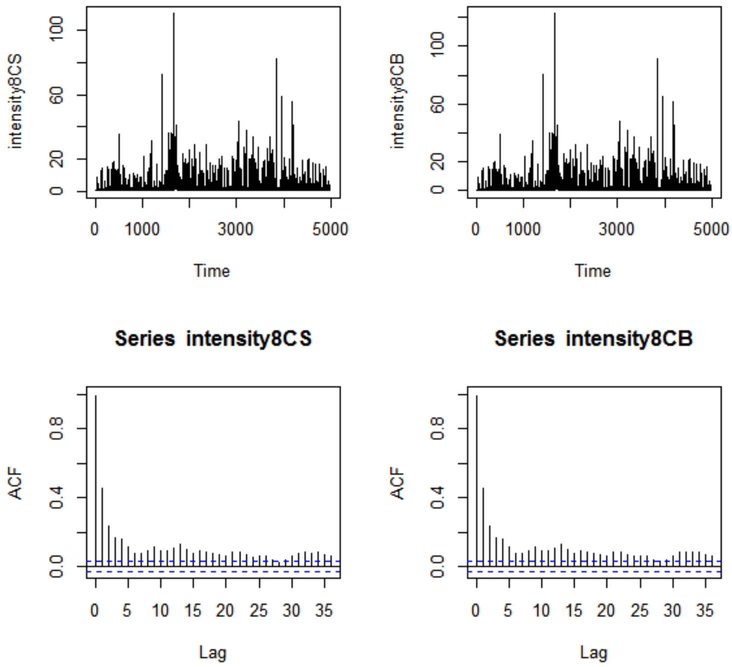

Figure A.21: Intensities of Sell and Buy Cancellations (14:00-15:30)

Appendix B

Spectral Densities

Spectral densities of test 2-8 are displayed, where ω ranges between 0 and π. Analogous explanations for the plots can be summarized as those for test 1 in Section 4.3.

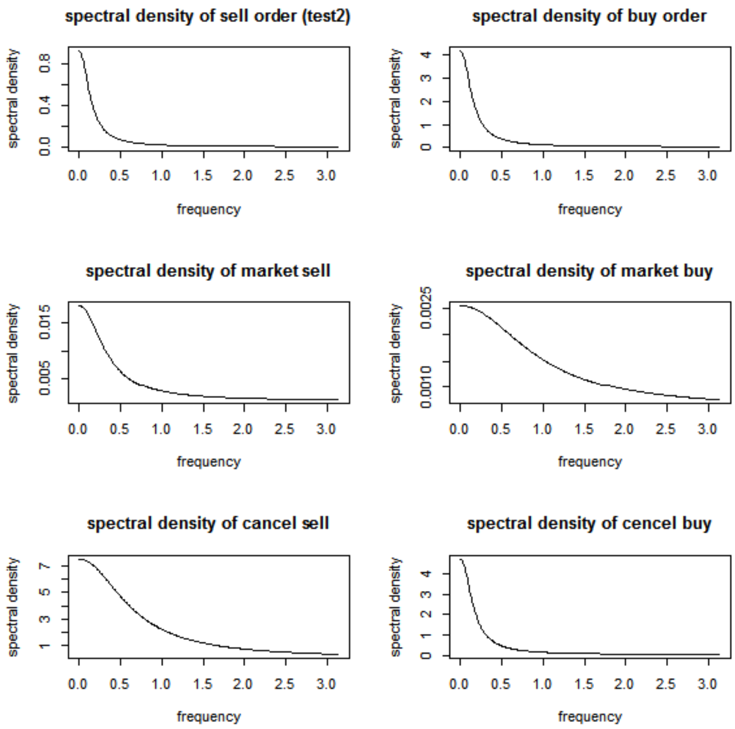

Figure B.1: Spectral Density of Model (10:00-10:30)

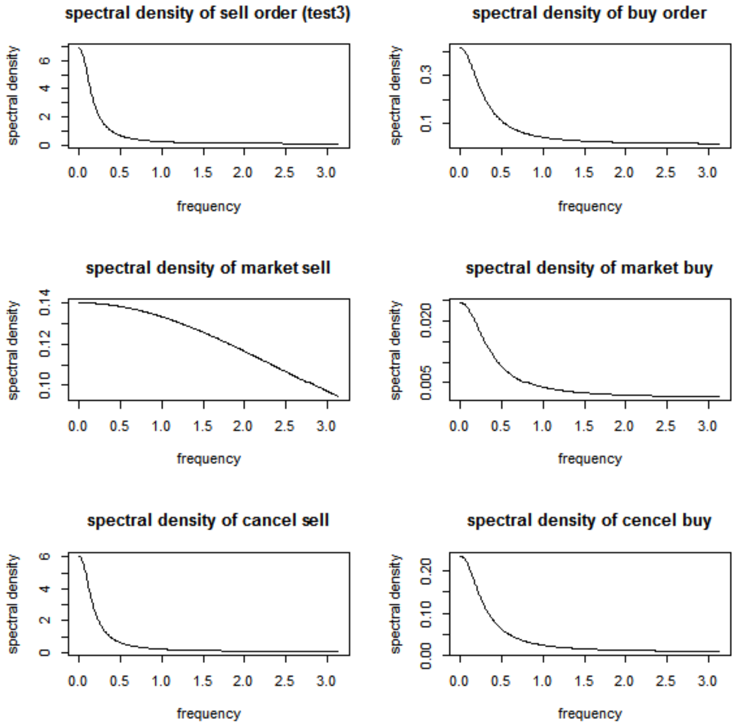

Figure B.2: Spectral Density of Model (10:30-11:00)

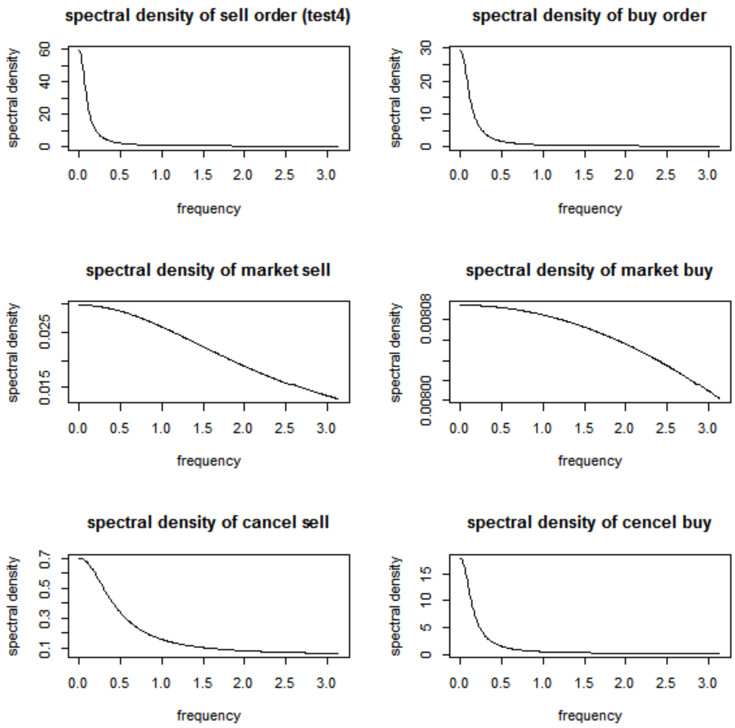

Figure B.3: Spectral Density of Model (09:30-11:00)

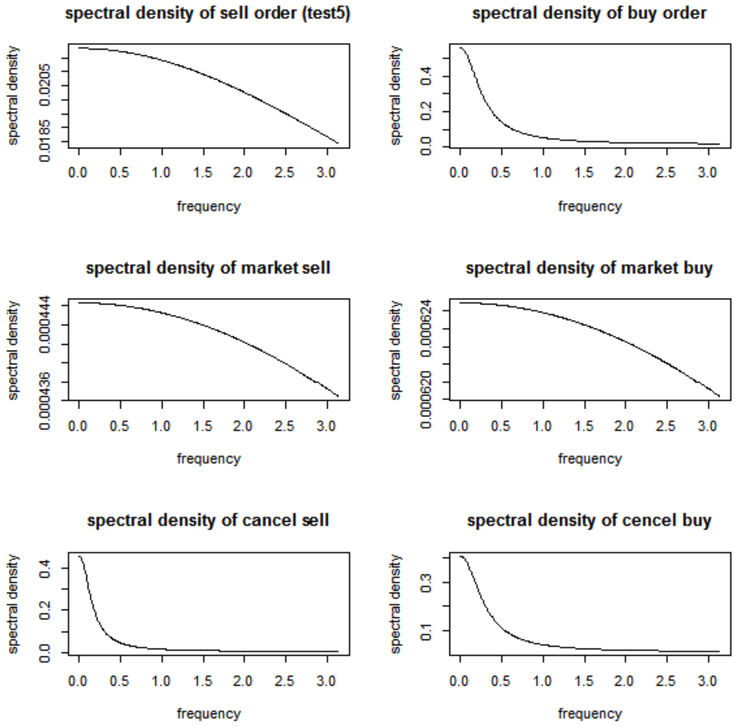

Figure B.4: Spectral Density of Model (13:00-13:30)

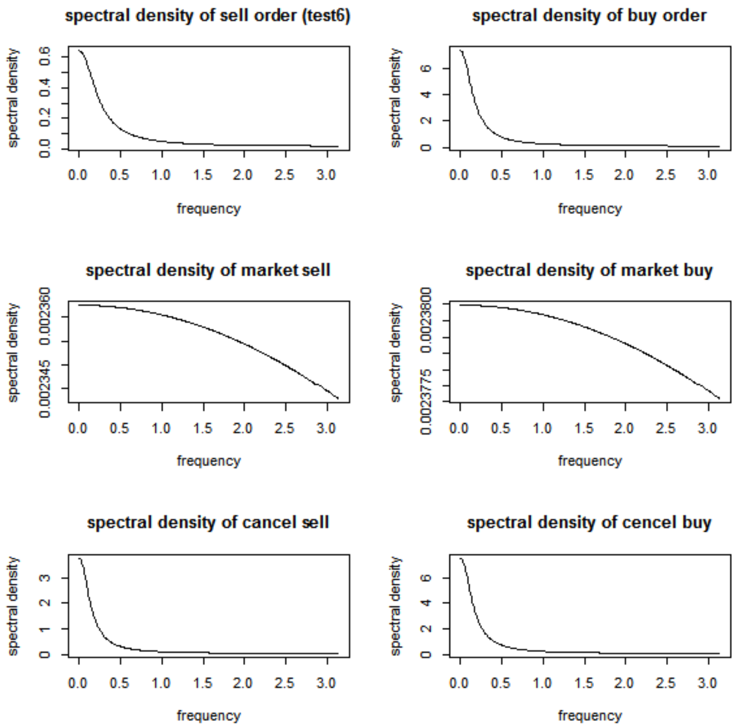

Figure B.5: Spectral Density of Model (12:00-13:30)

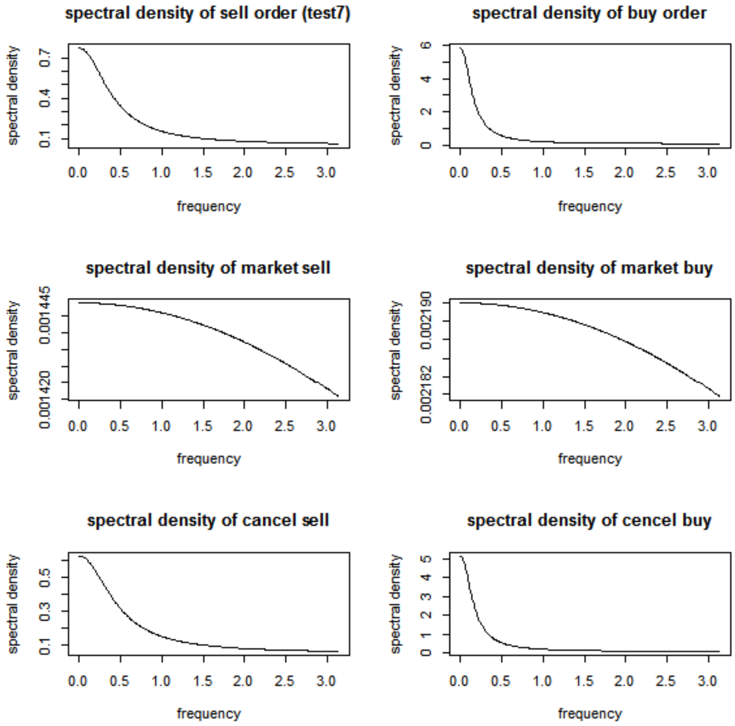

Figure B.6: Spectral Density of Model (15:00-15:30)

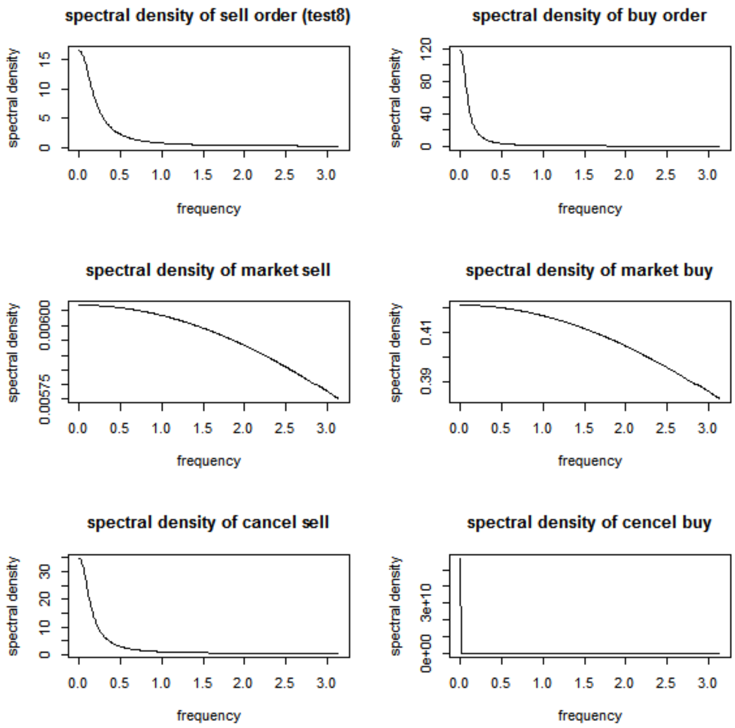

Figure B.7: Spectral Density of Model (14:00-15:30)

Appendix C

R Codes for Plots

Attached are R codes for all plots including empirical data analyses, intensities, ACF and sprectral densities in this essay.

```
###example of limit order book###
plot(0:29,0:29,type="n",axes=F,xlab="",ylab="",main="Example of a Limit Order")
abline(h=0,col="black")
abline(v=0,col="black")
ax=c(3,3,5,5,8,8,10,10,16,16,18,18,21,21,23,23)
ay=c(0,13,13,0,0,25,25,0,0,17,17,0,0,11,11,0)
axx=c(9,17)
ayy=c(10,10)
lines(ax,ay)
lines(axx,ayy)

points(9,0,pch=19)
points(17,0,pch=19)
text(9,-0.5,"Bid Quote",)
text(17,-0.5,"Ask Quote")
text(28,2,"Price")
text(3,29,"Quantity")
text(13,11,"Bid-ask Spread")
arrows(16,10,17,10,length=0.1)
arrows(10,10,9,10,length=0.1)
arrows(0,0,30,0,length=0.1,lwd=2)
arrows(0,0,0,30,length=0.1,lwd=2)

###empirical analysis###
cancel<-canceldata[order(canceldata$ordersequence),]
execution<-executiondata[order(executiondata$ordersequence),]
ord<-orderdata[order(orderdata$ordersequence),]

length(orderdata$mstime)
length(executiondata$mstime)
```

```
length(canceldata$mstime)

length(orderdata$mstime[orderdata$display=="Y"])
length(dat$time[dat$type=="3"])+length(dat$time[dat$type=="4"])
length(dat$time[dat$type=="5"])+length(dat$time[dat$type=="6"])

##########################time series#######################
###event
par(mfrow=c(3,1))
tic_order<-orderdata$mstime
toc_order<-tic_order/1E9/3600
hist(toc_order,main="Histogram of order events")
#or
hist(toc_order,probability=TRUE,main="Histogram of order events",xlim=c(8,18))
lines(density(toc_order),col="red",lwd=2)

tic_trade<-executiondata$mstime
toc_trade<-tic_trade/1E9/3600
hist(toc_trade,main="Histogram of trade times")
#or
hist(toc_trade,probability=TRUE,main="Histogram of execution times",xlim=c(8,18))
lines(density(toc_trade),col="red",lwd=2)

tic_cancel<-canceldata$mstime
toc_cancel<-tic_cancel/1E9/3600
hist(toc_cancel,main="Histogram of cancellation times")
#or
hist(toc_cancel,probability=TRUE,main="Histogram of cancellation times",xlim=c(8,18))
lines(density(toc_cancel),col="red",lwd=2)

###volumes
OrderQuantity<-ts(orderdata$quantity,frequency=1)
plot(OrderQuantity)

CancelQuantity<-ts(canceldata$quantity)
plot(CancelQuantity)

ExecuteQuantity<-ts(executiondata$quantity)
plot(ExecuteQuantity)

###bid-ask spreads
spread<-rep(NA,5000)
for(i in 1:5000){
 spread[i]=orderbookdata$AskPrice_1[i]-as.numeric(levels(orderbookdata$BidPrice_1[i])
}
```

```
tsSpread<-ts(spread,start=1,end=5000)
plot(tsSpread)

###proportions
#all-day
total<-rep(NA,5000)
can<-rep(NA,5000)
exe<-rep(NA,5000)
od<-rep(NA,5000)
proportion_of_cancellation<-rep(NA,5000)
proportion_of_execution<-rep(NA,5000)
proportion_of_order<-rep(NA,5000)
t1<-seq(9*3600*1E9,17*3600*1E9,length=5000)
for (i in 1:5000){
 od[i]= length(orderdata$mstime[orderdata$mstime < t1[i]])
 exe[i]= length(executiondata$mstime[executiondata$mstime < t1[i]])
 can[i]= length(canceldata$mstime[canceldata$mstime < t1[i]])
}
 proportion_of_execution= exe/od
 proportion_of_cancellation=can/od

par(mfrow=c(2,1))
plot.ts(proportion_of_execution)
plot.ts(proportion_of_cancellation)

#normal trading time
t2<-seq(9.5*3600*1E9,16*3600*1E9,length=5000)
for (i in 1:5000){
 od[i]= length(orderdata$mstime[orderdata$mstime < t2[i]])
 exe[i]= length(executiondata$mstime[executiondata$mstime < t2[i]])
 can[i]= length(canceldata$mstime[canceldata$mstime < t2[i]])
}
 proportion_of_execution= exe/od
 proportion_of_cancellation=can/od

par(mfrow=c(2,1))
plot.ts(proportion_of_execution)
plot.ts(proportion_of_cancellation)
###acf
acf(toc_trade)
acf(ExecuteQuantity)

askorder_time1<-orderdata$mstime[orderdata$side=="S"][orderdata$display=="Y"]
ask_order_time1<-askorder_time1[!is.na(askorder_time1)]
bidorder_time1<-orderdata$mstime[orderdata$side=="B"][orderdata$display=="Y"]
```

```
bid_order_time1<-bidorder_time1[!is.na(bidorder_time1)]
ask_order_time<-ask_order_time1[order(ask_order_time1)]
bid_order_time<-bid_order_time1[order(bid_order_time1)]
ask_order_duration<-ask_order_time[-1]-ask_order_time[-length(ask_order_time)]
bid_order_duration<-bid_order_time[-1]-bid_order_time[-length(bid_order_time)]

###A B CA CB MA MB
Ams1<-orderdata$mstime[orderdata$side=="S"][orderdata$display=="Y"]
Ams<-Ams1[!is.na(Ams1)]
Bms1<-orderdata$mstime[orderdata$side=="B"][orderdata$display=="Y"]
Bms<-Bms1[!is.na(Bms1)]

CA_time1<-rep(NA,length(canceldata$mstime))
CB_time1<-rep(NA,length(canceldata$mstime))
MA_time1<-rep(NA,length(executiondata$mstime))
MB_time1<-rep(NA,length(executiondata$mstime))


write(c("CA(time)"), "ask cancel.txt", ncolumns =1, sep = "\t", append = TRUE)
write(c("CB(time)"), "bid cancel.txt", ncolumns =1, sep = "\t", append = TRUE)
for(j in 40001:50000){
 for(i in 1:length(ask_seq)){
   if(canceldata$ordersequence[j]==ask_seq[i]){
    CA_time1[j]<-canceldata$mstime[j]
    write(CA_time1[j], "ask cancel.txt", ncolumns =1, sep = "\t", append = TRUE)
    break}}
 for(l in 1:length(bid_seq)){
   if(canceldata$ordersequence[j]==bid_seq[l]){
    CB_time1[j]<-canceldata$mstime[j]
    write(CB_time1[j], "bid cancel.txt", ncolumns =1, sep = "\t", append = TRUE)
    break}}
}


write(c("MA(time)"), "ask execution.txt", ncolumns =1, sep = "\t", append = TRUE)
write(c("MB(time)"), "bid execution.txt", ncolumns =1, sep = "\t", append = TRUE)

exec<-executiondata[executiondata$liquidity=="R",] #liquidity=R denotes a market buy/
for(k in 10001:length(exec$mstime)){
 for(i in 1:length(ask_seq)){
   if(exec$ordersequence[k]==ask_seq[i]){
    MA_time1[k]<-exec$mstime[k]
    write(MA_time1[k], "ask execution.txt", ncolumns =1, sep = "\t", append = TRUE)
    break}}
 for(l in 1:length(bid_seq)){
```

```
   if(exec$ordersequence[k]==bid_seq[l]){
    MB_time1[k]<-exec$mstime[k]
    write(MB_time1[k], "bid execution.txt", ncolumns =1, sep = "\t", append = TRUE)
    break}}
}

#check
anyDuplicated(executiondata$ordersequence)
anyDuplicated(canceldata$ordersequence)
anyDuplicated(orderdata$ordersequence)

typeA<-rep(1,length(Ams))
typeB<-rep(2,length(Bms))
typeMA<-rep(3,length(MAms))
typeMB<-rep(4,length(MBms))
typeCA<-rep(5,length(CAms))
typeCB<-rep(6,length(CAms))

AO<-cbind(Ams,typeA)
BO<-cbind(Bms,typeB)
MA<-cbind(MAms,typeMA)
MB<-cbind(MBms,typeMB)
CA<-cbind(CAms,typeCA)
CB<-cbind(CBms,typeCB)

initialdat<-data.frame(time=c(AO[,1],BO[,1],MA[,1],MB[,1],CA[,1],CB[,1]),
 type=c(AO[,2],BO[,2],MA[,2],MB[,2],CA[,2],CB[,2]))
check<-edit(initialdat) #visible window
dat<-initialdat[order(initialdat$time),]
check<-edit(dat)

###example of intensity process###
plot(-1:30,-1:30,type="n",axes=F,xlab="",ylab="",main="Example of Intensity Process")
abline(h=0,col="black")
abline(v=0,col="black")
ax=c(0,3)
ay=c(1,4)
lines(ax,ay)
ax=c(3,7)
ay=c(10,3)
lines(ax,ay)
ax=c(7,9)
ay=c(2,1)
lines(ax,ay)
ax=c(15,17)
```

```
ay=c(5,2)
lines(ax,ay)
ax=c(17,19)
ay=c(4,9)
lines(ax,ay)
ax=c(19,21)
ay=c(10,6)
lines(ax,ay)
ax=c(21,22)
ay=c(3,5)
lines(ax,ay)
ax=c(22,24)
ay=c(6,7)
lines(ax,ay)
ax=c(24,27)
ay=c(5,2)
lines(ax,ay)

points(3,4,pch=19)
points(3,10,pch=1)
points(7,3,pch=19)
points(7,2,pch=1)
points(17,2,pch=19)
points(17,4,pch=1)
points(19,9,pch=19)
points(19,10,pch=1)
points(21,6,pch=19)
points(21,3,pch=1)
points(22,5,pch=19)
points(22,6,pch=1)
points(24,7,pch=19)
points(24,5,pch=1)
points(27,2,pch=19)

points(3,0,pch=4)
points(7,0,pch=4)
points(17,0,pch=4)
points(19,0,pch=4)
points(27,0,pch=4)

text(3,-1,"t1")
text(7,-1,"t2")
text(11,-1,"...")
text(11,1,"...")
text(17,-1,"tk")
```

```
text(19,-1,"tk+1")
text(23,-1,"...")
text(27,-1,"tn=t")

arrows(0,0,30,0,length=0.1,lwd=2)
arrows(0,0,0,20,length=0.1,lwd=2)

text(30,2,"time")
text(2.5,21,"intensity")

########################intensity#######################
int=function(param,xx,k,t){
#xx is data,t is time to be substituted, k is nr. of process
  mu=param[1]
  alpha=param[2]
  beta=param[3]
 z=xx$time[xx$type==k]
 zz=z[z<t]
 lambda=mu+sum(alpha*exp((-1)*beta*(t-zz)))
 return(lambda)
}

#draw
par(mfrow=c(2,2))

t=seq(14*3600,15.5*3600,length=5000)
intensity8S=rep(0,length(x))
pp=subset(dat,14*3600*1E9<dat$time & dat$time<15.5*3600*1E9)
pp$time=pp$time/1E9
 for (i in 1: length(t)){
  intensity8S[i]=int(param=par8[,1],
    xx=pp,k=1,t=t[i])
 }
plot.ts(intensity8S)

intensity8B=rep(0,length(x))
 for (i in 1: length(t)){
  intensity8B[i]=int(param=par8[,2],
    xx=pp,k=1,t=t[i])
 }
plot.ts(intensity8B)

acf(intensity8S)
acf(intensity8B)
```

```
intensity8MS=rep(0,length(x))
 for (i in 1: length(t)){
  intensity8MS[i]=int(param=par8[,3],
    xx=pp,k=1,t=t[i])
 }
plot.ts(intensity8MS)


intensity8MB=rep(0,length(x))
 for (i in 1: length(t)){
  intensity8MB[i]=int(param=par8[,4],
    xx=pp,k=1,t=t[i])
 }
plot.ts(intensity8MB)

acf(intensity8MS)
acf(intensity8MB)


intensity8CS=rep(0,length(x))
 for (i in 1: length(t)){
  intensity8CS[i]=int(param=par8[,5],
    xx=pp,k=1,t=t[i])
 }
plot.ts(intensity8CS)


intensity8CB=rep(0,length(x))
 for (i in 1: length(t)){
  intensity8CB[i]=int(param=par8[,6],
    xx=pp,k=1,t=t[i])
 }
plot.ts(intensity8CB)

acf(intensity8CS)
acf(intensity8CB)

###spectrum of Hawkes model###
spec=function(param,w){
  mu=param[1]
  alpha=param[2]
  beta=param[3]
  f=mu/2/pi*beta/(beta-alpha)*(1+alpha*(2*beta-alpha)/((beta-alpha)^2+w^2))
  return(f)
```

```
}

#draw
par(mfrow=c(3,2))

frequency<-seq(0,pi,0.01)

plot(frequency,spec(par8[,1],frequency),"l",ylab="spectral density",
  main="spectral density of sell order (test8)")
plot(frequency,spec(par8[,2],frequency),"l",ylab="spectral density",
  main="spectral density of buy order")
plot(frequency,spec(par8[,3],frequency),"l",ylab="spectral density",
  main="spectral density of market sell")
plot(frequency,spec(par8[,4],frequency),"l",ylab="spectral density",
  main="spectral density of market buy")
plot(frequency,spec(par8[,5],frequency),"l",ylab="spectral density",
  main="spectral density of cancel sell")
plot(frequency,spec(par8[,6],frequency),"l",ylab="spectral density",
  main="spectral density of cencel buy")
```

Bibliography

[1] Adams M and Guillemin V. (1995). *Measure Theory and Integration*, Birkhäuser Boston.

[2] Bacry E, Delattre S, Hoffmann M and Muzy JF. (2012). Scaling Limits for Hawkes Processes and Application to Financial Statistics. *arXiv*, Cite 1202.0842v1.

[3] Bowsher CG. (2007). Modelling Security Market Events in Continuous Time: Intensity Based, Multivariate Point Process Models. *Journal of Econometrics*, Vol. 141, Issue 2, pp. 876-912.

[4] Brockwell PJ and Davis RA. (2002). *Introduction to Time Series and Forecasting*, 2^{nd} ed., Springer-Verlag New York.

[5] Braun WJ and Murdoch DJ. (2007). *A First Course in Statistical Programming with R*, Cambridge University Press.

[6] Brzeźniak Z and Zastawniak T. (1998). *Basic Stochastic Processes: A Course through Exercises*, Springer-Verlag London.

[7] Cai ZW. (2006). *Analysis of Time Series Data Using R*

[8] Casella G and Berger RL. (2002). *Statistical Inference*, 2^{nd} ed., Thomson Learning.

[9] Dacorogna MM, Gençay R, Müller U, Olsen RB and Pictet OV. (2011). *An Introduction to High-frequency Finance*, Academic Press.

[10] Durbin M. (2010). *All about High-frequency Trading*. The McGraw-Hill Companies, Inc.

[11] Hautsch N. (2004). *Modelling Irregularly Spaced Financial Data*. Springer-Verlag Berlin Heidelberg.

[12] Hautsch N. (2011). *Econometrics of Financial High-frequency Data*. Springer-Verlag Berlin Heidelberg.

[13] Hawkes AG. (1971). Spectra of Some Self-Exciting and Mutually Exciting Point Processes. *Biometrika*, Vol. 58, No. 1, pp. 83-90.

[14] Hawkes AG and Oakes D. (1974). A Cluster Process Representation of a Self-Exciting Process. *Journal of Applied Probability*, Vol. 11, No. 3, pp. 493-503.

[15] Karr AF. (1986). *Point Processes and Their Statistical Inference.* Marcel Dekker, cop.

[16] Küchler U and Sørensen. (1997). *Exponential Families of Stochastic Processes*, Springer-Verlag New York.

[17] Lancaster T. (1990) *the Econometric Analysis of Transition Data.* Cambridge University Press.

[18] Møller J and Rasmussen JG. (2005). Perfect Simulation of Hawkes Processes. *Advances in Applied Probability*, Vol. 37, No. 3, pp. 629-646.

[19] NASDAQ OMX Nordic. (2011). *NASDAQ OMX Nordic Market Model 2.1.*

[20] Ogata Y and Akaike H. (1982). On Linear Intensity Models for Mixed Doubly Stochastic Poisson and Self- Exciting Point Processes, *Journal of the Royal Statistical Society*, Series B, Vol. 44, No. 1, pp. 102-107.

[21] Ozaki T. (1979). Maximum Likelihood Estimation of Hawkes' Self-exciting Point Processes. *Annals of the Institute of Statistical Mathematics*, Vol. 31, No. 1, pp. 145-155.

[22] Ross SM. (1996). *Stochastic Processes*, 2^{nd} ed., John Wiley and Sons, Inc.

[23] Rubin I. (1972). Regular Point Processes and Their Detection. *IEEE Transactions on Information Theory*, Vol. IT-18, No. 5, pp. 547-557.

[24] Stirzaker D. (2005). *Stochastic Processes and Models*, Oxford University Press.

[25] Toke IM. (2010). "Market making" in an order book model and its impact on the spread. *Econophysics of Order-driven Markets.* Springer-Verlag Italia. pp. 49-64.

[26] Toke IM and Pomponio F. (2011). Modelling Trades-Through in a Limited Order Book Using Hawkes Processes. *Economics*, No. 2011-32.