# ON THE HAWKES' PROCESSES AND ITS APPLICATION ON A-H SHARES

**Qian LIU**

MPhil

The Hong Kong Polytechnic University
2019

The Hong Kong Polytechnic University
Faculty of Applied Science and Textiles
Department of Applied Mathematics

# On the Hawkes' Processes and Its Application on A-H Shares

**Qian Liu**

A thesis submitted in partial fulfilment of the requirements for
the degree of Master of Philosophy
April 2018

# Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____(signed)

Qian Liu (name of student)

# Dedication

Dedicate to my husband and my daughter , those beloved, and the very interest I hold on to financial mathematics.

# Abstract

There are many indicators for measuring the financial market liquidity. The thesis explores recent academic literature related to one-dimensional and two-dimensional Hawkes' processes applied in financial market. Hawkes' processes are derived from a model of multivariate point processes and has recently been sought-after for the applications in high frequency financial models in the past decade. In this thesis, the Hawkes process with exponential response function is applied to model the arrival process of the trades, where the expected intensity and flow branching ratio are calculated from the estimated values of the model parameters. The expected intensity represents the transaction intensity, which is an indicator of the adequacy of liquidity; and another liquidity indicator is the branching ratio, which is the measure of the degree of aggregation of transactions.

After giving a brief overview of the main definitions on properties and characteristics of how Hawkes' processes are applied, we describe procedures for simulating and deriving maximum likelihood estimation (MLE) functions for parameter estimation on one dimensional process and extend to two dimensional case. Then we survey various empirical studies using one-dimensional and two-dimensional Hawkes' processes in A-H shares both in Mainland and in the Hong Kong stock markets. Comparable studies of both the Mainland and Hong Kong markets are beneficial because both markets differ, so we would like to focus on different investors investment behavior. We choose two pairs of A-H shares both in Mainland and the Hong Kong stock markets in different industries classification. It seems intuitive to analyze the different regulated markets simultaneously since the process of market evolution can be observed as well as the differences and similarities of the two markets. Our model accounts for the arrival of bid-ask orders intensities on both two markets that influence activities, trigger one-sided or two-sided clustering of trades. We use branching ratio and expected

intensity as measurement of activity on A-H shares that provides a direct way to access their level of endogeneity and also relates to market liquidity. From our numerical results, We can observe that two-dimensional Hawkes' processes have better explanatory power on market liquidity than one-dimensional Hawkes' processes. Also the resulting from confusion matrix, the overall accuracy on trading strategy would be more than 70%. Furthermore the performance on H-shares in Hong Kong stock market did better than A-shares in Mainland China stock market.

There are still many applications of Hawkes' process, as we discuss in relating with two-dimensional Hawkes' processes its bid-ask side, calculate its branching ratio and expected intensity. Another contribution of this thesis surrounds the bid-ask expectation intensity we proposed a trading strategy on confusion matrix. Through confusion matrix comparison, we conclude that the forecast performance of expected intensity of H-shares in Hong Kong market is obviously better than that of A-shares in the domestic market.

# Acknowledgements

When endeavoring to carry out this research much gratitude is towards the following individuals for their various support through my M.Phil program. I hereby acknowledge the support they provided.

To my much appreciated supervisor, Prof. Ka-fai Yiu, Cedric, for his continued patience and guidance and encouragement for without which my work would not have been completed. I am truly grateful for his unwavering support and I have learnt much from his rigorous and positive attitude and passion to my scientific research.

Also at the forefront of my research I am grateful for the kindness of other professors who have been a source of guidance and inspiration and mentorship.

I am thankful for my husband and parents for their constant love and encouragement and support. I appreciate especially the support and understanding of my lovely baby daughter Natalie, "I love you and I wish you grow to be a healthy and happy girl."

I give a special thank you to my friends for their love, support and encouragement as well as my fellow researchers at Hong Kong Polytechnic University.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Transactions in modern-day financial markets have reached a scale of unprecedented volume. Over the last decade we have observed not only more transactions but also increases in the volume of trade exchanges around the world, with an increase occurrences of such transactions. The occurrences of the New York stock market disaster in 1987 led to a deeper thinking on the micro-price formation mechanism of the market. The day it happened, the Dow Jones index dropped by 22.6%. During the period of stable economic growth in the United States, the New York Stock Exchange suddenly suffered such a huge shock. This incidence fully exposed the fragility of the financial market, and also prompted financial workers to focus their attention to this issue. Research has found that the main reason for stock market shock is financial market information asymmetry. How to analyze the investor investment behavior and design a reasonable strategy for market microstructure through historical transaction data, and how to make the market more efficient have become popular research topics in finance.

With the experience and technological advance, trading activities have been taken to a new standard, which is known as High Frequency Trading (HFT). In year 2000, HFT accounted for a relatively small amount of trading activity in equity market. Nowadays it has grown to be the dominant force [40]. With a substantial growth in HFT it has had a massive impact on the microstructural aspects of financial markets. The US Securities and Exchange Commission (SEC) recommends reviewing the behavior of the stock market to check whether market rules are in sync with changing in trading techniques and practices.

SEC also outlines the speed at which the financial markets increased. For example, small execution of orders such as NTSE in 2005 with an average of 10.1 seconds which by 2009 had been reducing further to 0.7 seconds. Over the same period, the combined number of traders in NYSE jumped from 2.9 million in 2005 to 22.1 million in 2009. These impressive figures set a new paradigm and challenges to the finance community. It is well known that trading activity is not evenly distributed through the trading day, and it occurs more frequently at the beginning and end of the day. In addition, trading activity tends to occur in clusters. As one looks closer at trading frequency during the day, it seems that there are short periods intensity activities. This makes the duration between trades irregular, in this contrast to the standard econometric techniques and this notably poses a challenge. However, some attempts have been made to model the irregular gaps between trades know as irregularly spaced and auto-correlated inter-trade durations. One was the Autoregressive Conditional Duration (ACD) Model [23], which was initiated in modeling transaction data. A point process is a technique that uses a set of stochastic processes. The simplest case is the homogenous Poisson process that described the random arrival of a new event at a constant rate of $\lambda$. This process has many other applications which could be used in different situations; for example, it can be used to model the arrival rate of radiation at a Geiger counter, or the arrived of clients in a shop. A more sophisticated point process is a non-homogenous Poisson process where events are related and determined by time, i.e $\mu = \mu(t)$. This process is widely used to model the arrival rate of aircraft, containerships and telephone calls. However, the analysis of more random signals that arise from events such as earthquakes or the stock market requires more sophisticated models than the non-homogenous point process. Earthquakes and stock market data present what is called an endogenous clustering effect which cannot be modelled using the simple point process. The ACD model is one of the first attempt to include all the stylized factors into an auto-correlated model. However, there exists no joint points that can be used to couple the processes making it difficult to estimate contemporaneous correlations between the individual events. It is also difficult to account for time-varying covariates in a discrete-time duration framework. Those are the drawbacks of ACD model. We will give some details about ACD model in next coming section.

There are other stylized factors related model that is able to reproduce facts looked like stock market data for clustering of order arrivals, the Hawkes model [37]. Hawkes' model is based on point processes but is self-exciting. It find many applications in fields

such as seismology, epidemiology, neurophysiology, and also in finance. This model becomes popular because of Daley and Vere-Jones [18], who used model to led the clustering effect of parsimonious data as well as a linear representation of conditional intensity. The model is said to be self-exciting because the Hawkes models events that arrive at a rate, which is possibly time-varying events when they occur those events affect third events passes on and on. In this thesis, we work with Hawkes' model and focus on high frequency stock market data on the duration of trades and quotes. Closely following the estimation method by Filimonov and Sornette [28], who have already estimated the mid-price changes from duration data for the E-mini $S\&P500$ contract with Hawkes' process. Time-stamp data is considered to be very important, they have rounded to the nearest second, it is common to see multiple trades happened in a single second. Filimonov and Sornette [28] use random number intervals (one 10th of a second) to randomly offset the original time-stamp. We assert this is a strong assumption. We are going to assess the impact of these random time-stamp estimators against the Hawkes model. In this chapter, we introduce basic point processes, ACD model and Hawkes' process and some applications in finance.

## 1.1 Point Process

Point processes are composed of a random collection of points that fall in space or time and provide a statistical pattern that describe the timing and properties of events. Occurrences that fitted into point process find several possible applications. It can be applied in geophysical events such as earthquakes, in which one can indicate the likelihood of another earthquake happening in that or a similar location in the future. In an ecological sense the location of a certain species that has been observed can help forecasting a certain set of habitat locations likely in the future. Even social media events could be modeled over time where a set of properties related to user habits, topics of interest, and connections to the surrounding networks. In finance, a buy or a sell transaction could influence prices and volumes of a similar transaction in future.

The point process is known as a stochastic process where random processes composed of a time-series events that occur continuously in time, as proposed by Daley and Vere-Jones [18]. These are used to describe localized data in a set of time points that are finite. In

contrast to continuous-valued processes, which can take any number of values at each point in time, a point process only take one of two possible value to indicate an event has occurred at that point in time. There is a sense in which using probability models make it much easier to express a point process data mathematically. Having said that point process data is often inappropriately analyzed initially because the standard signal-processing techniques are primarily designed for continuous-valued data. A good understanding of probability theory on point processes is vital for appropriate analysis of point process data.

A point process containing additional variables beyond the simple time values makes a marked point process which is a function containing multiple inputs. The marked point processes have application for example in finance where the function would contain not only the times of trades in the market but also the volumes of them, quoted prices and who trade or not trade. Now let turn to focus on what the point process is.

Let $t$ denotes calendar time and let $t_{i\in 1,2,...}$ be a random sequence of increasing events arrival times $0 \leq t_i \leq t_{i+1}$, then the sequence $\{t_i\}$ is called a point process on $[0, \infty)$. In particular, the variable $t_i$ can represent the time of occurrence of transactions, or arrival of orders in an order book. Simultaneously the assumption $W_i$ indicates the value sequence, and $W_i \in 1, 2, ..., K$ represents the $i$th event, $K$ is non-negative integers. If $t_i$ and $W_i$ are combined together, a new sequence $t_i, W_i$ can be obtained. This process is called marked point process. The sequence contains the time of the current event and the specific content of the event (K dimension value information), can also be said to be a K-dimensional point process. The content of the marked value could be the transaction information of the buy and sell, and data information in many transaction aspect, such as the price of buying and selling. The occurrence time of these standard values can be recorded as $t_i^k$.

A common kernel concept in the point process theory is the intensity function. Here, a stochastic process $N(t)$ is assumed to be a counting process, and $F_t$ is a domain flow process. $N(t)$ is adapted to $F_t$, and $N(t)$ is a sub-martingale process. According to Doob-Meyer's decomposition theory, the sub-martingale $N(t)$ overrange partial solution is shown as Eqn. (1.1), that is, a martingale process $M(t)$ and an incremental process $\Lambda(t)$, in which $\Lambda(t)$ is called the compensation process of $N(t)$. Details of the compensation process are discussed

in Chapter 3,

$$\begin{aligned} E[N(t)|F_s] &\geq N(s) \\ N(t) &= M(t) + \Lambda(t) \\ \Lambda(t) &= \int_0^t \lambda(u)du \end{aligned} \tag{1.1}$$

Eqn. (1.1) shows $\lambda(t)$ is a scalar, and a left continuous right limit process situation, then $\lambda(t)$ is called the $N(t)$ process intensity function based on the $F(t)$ domain flow.

### 1.1.1 Basic Poisson Process

The simplest type of point process is Poisson process. Suppose a basically counting process $N(a, b]$ ,where $a < t \leq b$ expresses the events happened in this period of time. From the mathematical aspect describing the Poisson process, it has the following properties:

- the number of occurrences in different time intervals are independent of one another;
- two points almost certainly won't arrive at the same time;
- within a particular time interval, the probability of an occurrence must be proportional to the length of the interval.

Because the $\lambda$ is the number of the number of events per unit time, so it can be defined as the intensity. If $\Delta t \to 0$, then we got

$$\begin{cases} Pr(N[t, t + \Delta t] = 1) = \lambda \Delta t \\ Pr(N[t, t + \Delta t] >= 2) \to 0 \end{cases} \tag{1.2}$$

For any time interval likes $(a, b]$, We first can be divided into multiple small time period, and then by the independence of the different time periods, using binomial distribution to calculate the probability distribution, reoccupy approximate Poisson distribution

$$\begin{cases} Pr(N(a, b] = k) = \frac{n!}{k!(n-k)!} \\ \Downarrow \\ (\lambda \Delta t)^k (1 - \lambda \Delta t)^{((\frac{(b-a)}{\Delta(t)})-k)} \to (\lambda(b-a))^k \frac{e^{(-\lambda(b-a))}}{k!} \end{cases} \tag{1.3}$$

Here we can see $N(a,b]$ approximately obey that parameters for $\lambda(b-a)$ Poisson distribution. So it is easy to know the time interval $t_{interval}$ between two events follows the exponential distribution

$$\begin{cases} Pr(interval > t_{interval}) = Pr(N(t, t+t_{interval}] = 0) = e^{-\lambda t_{interval}} \\ \Downarrow \\ f(interval = t_{interval}) = \lambda e^{-\lambda t_{interval}} \end{cases} \tag{1.4}$$

It is interesting to simulate a Poisson process on the basis of Poisson qualities. Naturally, we follow simulation algorithm which was raised by Lewis and Shedler [42]. This method now and then is called "interarrival scheduling". It assumes a random number generator uniform $(0,1)$ that generates random variables uniformly distributed on $(0,1)$, Table 1.1.1.

| **Algorithm 1.1** Simulation of a Homogenous Poisson process algorithm set with Rate $\lambda$,on $[0, \mathbf{T}]$ | |
|---|---|
| **1Input:**$\lambda, \mathbf{T}$ | |
| **2 While True do** | |
| **3 Generate** $u \sim$ uniform(0,1); | |
| **4 Let** $\omega = -\ln \frac{u}{\lambda}$; | so that $\omega \sim$ exponential($\lambda$) |
| **5 Set** $t_{n+1} = t_n + \omega$ ; | |
| **6 If** $t_{n+1} > \mathbf{T}$ **then** | |
| **7 return** $\{t_k\}_{k=1,2,\ldots,n}$ | |
| **8 else** | |
| **9 Set** $n = n+1$ | |
| **10 end If** | |
| **11 end** | |

Table 1.1: A simple simulation method of a Homogenous Poisson process.

We can simulate events that follow the Poisson distribution, simulation takes advantage of the fact that the time interval obeys an exponential distribution. The $k$th event occurred time is the random variable of parameters $\lambda$ with an exponential distribution of the $(k-1)$th event time:

According to this simple case, we simulate from $t = 0$ to $t = 100$ with $\lambda = 0.5$, there are 50 events occurred between $t = 0$ and $t = 100$, see Fig 1.2.

A point process in its simplest form is homogeneous Poisson process. The process where the probability of one event occurring in the next(very tiny) time interval is proportionally to a constant $\mu(\lambda(t)) = \mu$. The very structure of a simple Poisson process makes it unable to observe the stock market such as clustering of orders on arrival. Another special feature

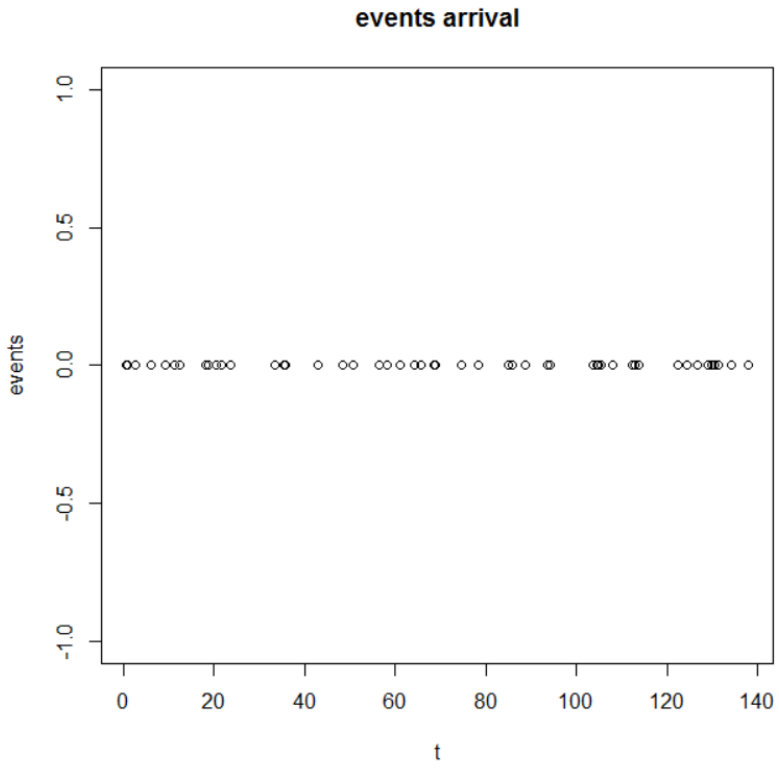



Figure 1.1: A simple random Poisson process with $\lambda = 0.5$ and the total events is 50.

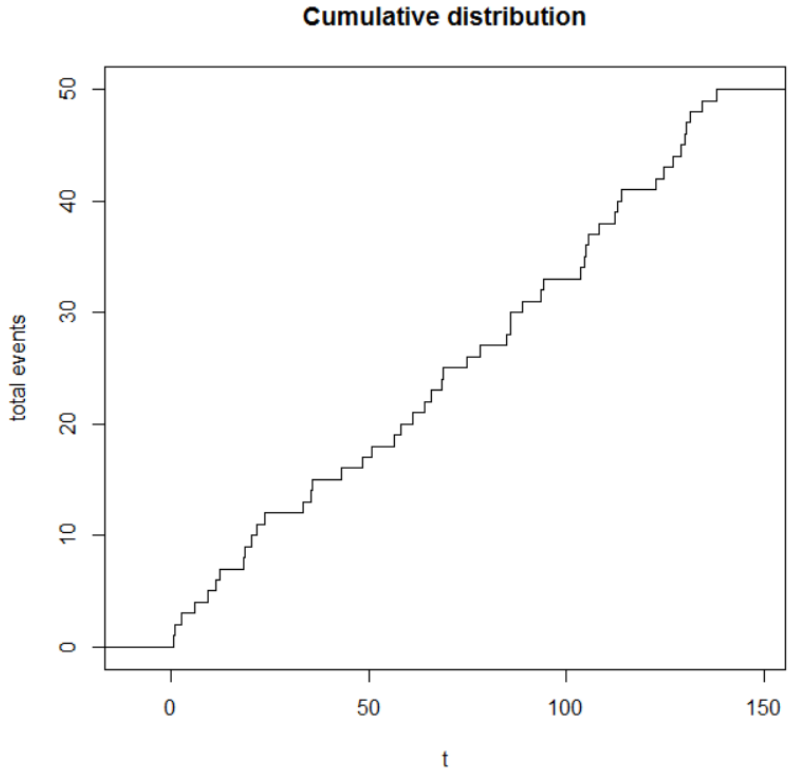



Figure 1.2: Cumulative events relationship with time.

of the homogeneous Poisson process is that it has no memory, this problem is significant as intra-event duration does not depend on the previous events, thus if one event reproduced some specific stylized factors such as clustering of arrivals orders, then some correlation structuring would not be corporated into the process. In the Chapter 3, we present a process where correlation structures are an additive to the Poisson process in a model known as the Hawkes model.

### 1.1.2 Duration ACD Model

In essence, financial market transactions occur at unequal intervals, but traditional econometrics are based on observations at the same time interval. Engle and Russell [22] mentioned equal time interval processing method had a problem, because the transaction frequency changes with time, the higher the transaction frequency, the more information it contains. Equal time interval results in many aggregated with observations do not provide any information. To make up for this deficiency, one effective method that is to record all the data for each transaction. There are two types of transaction data: one is the arrival time interval of the transaction, that is the durations; the other category includes the transaction price, the transaction volume, and the bid-ask spread, the sequence process that combines any two is called marked point process, in addition, all of these are called high frequency data. The application of high-frequency data helps to compensate the lack of information at the same time interval of statistical modeling. It gives insight into the deeper observed in the market trading process, and it also becomes an important tool on analyzing and revealing the impact of asymmetric information on financial market. Engle and Russell [23] proposed the Autoregressive Conditional Duration Model (ACD) model for financial market events clustering modeling for renewal. The clustering of duration means that short durations are often followed by short durations, and long durations are often followed by long durations. The GARCH model is a clustering of volatility, and the ACD model is a clustering of duration. The duration of clustering means that transactions tend to be more frequent over a period of time, while transactions are of relatively low frequency in another period of time, which means that short durations are often followed by short durations, and long durations are often followed by the long durations. Therefore, the ACD model has a very similar form to the GARCH model. Both are multiplication errors based on latent

variables model (Multiplicative Error Model, MEM).

We assume $t$ denotes the time when the $i$th transaction occurred, so $i = 1, 2, ..., n_t$ which is the number of transactions in the transaction day. Let $x_i = t_i - t_{i-1}$, then $x_i$ denote the time interval between the $i$th transaction and the $(i-1)$th transaction, that is, the duration of transaction. The transaction duration $x_i$ is not an independent and irrelevant point process, but a conditional point process related to past transaction information. Therefore , Engle and Russell give the following Autoregressive conditional duration model ACD(p,q)

$$\begin{aligned} x_i &= \psi_i \varepsilon_i \\ \psi_i &= \omega + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} \end{aligned} \tag{1.5}$$

.

where $\psi_i = E(x_i|x_{i-1}, ..., x_1)$ is the expectation of the conditional duration and there is a dynamic linear relationship. $\varepsilon_i$ is a random disturbance of independent identically distributed(i.i.d) and $E(\varepsilon_i) = 1$. Since $x_i$ is defined as a duration, therefore $x_i > 0$, $\psi_i > 0$, $\varepsilon_i > 0$, and the parameters $\alpha_j > 0$, $\beta_j > 0$, $\omega_j > 0$, $\sum(\alpha_j + \beta_j) < 1$. Here are some related theories about point processes. A common nuclear concept in point process theory is the intensity function. This assumes that $N(t)$ is a simple point process, $F_t$ is a domain flow process. $N_t$ is adapted to $F_t$, and $N(t)$ is a under the martingale process,

$$E[N(t)|F_s] \geq N(s). \tag{1.6}$$

According to Doob-Meyer's decomposition theory, the under the martingale $N(t)$ can be decomposed as

$$N(t) = M(t) + \Lambda(t). \tag{1.7}$$

That is, one martingale process $M(t)$ and one incremental process $\Lambda(t)$. Among them, $\Lambda(t)$ is called the compensation process of $N(t)$ calculated as

$$\Lambda(t) = \int_0^t \lambda(u)du. \tag{1.8}$$

It note that $\lambda(t)$ is a scalar, and a left continuous right limit process situation, then

$\lambda(t)$ is called the $N(t)$ process intensity function based on the $F(t)$ domain flow, and can also be expressed equivalently as

$$\begin{aligned} E[N(s) - N(t)|F_t] &= E[\int_t^s \lambda(u)du|F_t] \\ \lambda(t) \approx \lambda(t+) &:= \lim_{\Delta \downarrow 0} \frac{1}{\Delta} E[N(t+\Delta) - N(t)|F_t]. \end{aligned} \tag{1.9}$$

Where $\lambda(t+) := \lim_{\Delta \downarrow 0} \lambda(t+\Delta)$, under the smooth point process, $\lambda := E[dN(t)/dt] = E[\lambda(t)]$ is a constant.

Another important theory in ACD duration modeling is Hazard function. This function measures the event characteristics of a point process from another aspect. The theoretical knowledge in the survival analysis usually used to discuss the hazard function implied in the distribution function of different random disturbances, that is, the conditional intensity function. It reflects the possibility of failure after the duration $x_i$ of the $i$th transaction, and more accurately describes the conditional intensity characteristics of the duration, which is more explanatory than the probability density function under absolute conditions. Assume that $f(t)$ and $F(t)$ are the probability density functions and cumulative distributions of the random disturbance term respectively. The function, whose hazard function $h$ is defined as follows

$$h(t|N(t), t_1, ..., t_n) = \lim_{\Delta t \to 0} \frac{P(N(t+\Delta t) > N(t)|N(t), t_1, ..., t_n)}{\Delta t} = \frac{f(t)}{1 - F(t)}. \tag{1.10}$$

Let $t = \varepsilon_i = \frac{x_i}{\varphi_i}$, then we get the hazard function of duration $x_i$ is

$$h(x_i|F) = h_0(\frac{x_i}{\varphi_i})\frac{1}{\varphi_i}. \tag{1.11}$$

Where $h_0$ is the baseline risk rate function for the random disturbance term $\varepsilon_i$ .

The benchmark hazard function under various distributions can be seen from the following Table 1.2. The exponential distribution hazard function is a constant, while the Weibull distribution contains the Exp distribution ($\gamma = 1$), which is the hazard rate function with monotonicity. That is, when $\gamma > 1$, its hazard function is monotonically increasing, and when $\gamma < 1$, it is monotonically decreasing.

It should be noted that the distribution of random perturbation terms represents the

| | Probability density function(pdf) | hazard function |
|---|---|---|
| Exp Distribution | $f(t) = \lambda exp(-\lambda t)$ | $h(t) = \lambda$ |
| Weibull Distribution | $f(t) = \kappa\gamma(\kappa t)^{\gamma-1}exp(-(\kappa t)^{\gamma})$ | $h(t) = \gamma\kappa^{\gamma}t^{\gamma-1}$ |
| Burr Distribution | $f(t) = \frac{\mu\kappa t^{\kappa-1}}{(1+\sigma^2\mu t^{\kappa})^{\frac{1}{\sigma^2}+1}}$ | $h(t) = \frac{\mu\kappa t^{\kappa-1}}{1+\sigma^2\mu t^{\kappa}}$ |

Table 1.2: Probability density function of different random disturbances and its hazard function

behavior of the trader. When $\gamma = 1$, it degenerates into an Exp distribution, which means that the trader is conducting a random transaction, that is, the transaction of the uninformed trader is skillful. When $\gamma < 1$, it is a function of decreasing slope, it represents the trading behavior of informed traders. When $\gamma > 1$, the hazard rate function is a trend graph with the slope increasing upwards, representing the trading behavior of the technical trader.

Grammig and Maurer [32] found in the empirical data that the distribution histogram of the stock volume duration is not completely monotonous, but there is an inverted "U" type distribution feature. Using the Exponential distribution or the Weibull distribution for ACD modeling existed estimation bias, then the Burr distribution under the non-monotonic hazard function is introduced. As can be seen from its expression, the Burr distribution function contains two shape parameters, so it is not a simple monotonic mapping. We show in Figure 1.3 the inverted "U" type form the H-stock 1398(ICBC) is confirmed and also we performs a diurnal adjustment of the durations,such as removes a daily seasonal component. The Detail about Qusai-Maximum Likelihood Estimation on ACD Model and how is the difference between ACD Model and Hawkes' model are discussed in Chapter 3.

### 1.1.3 Hawkes' Model

Hawkes' process, a self-exciting process, known as epidemic models since a number of occurrence events increases the probability for further events. In natural sciences, the Hawkes model play an essential role in modeling the next coming events.

Hawkes' process describes a series of individual events that happened either exogenously (known as immigrant) events or endogenous (known as descendant) events, but all due to past events. Endogenous events are described as basic rate, those events can be created multiple events that follow on to subsequent generations. We can refer to the events as

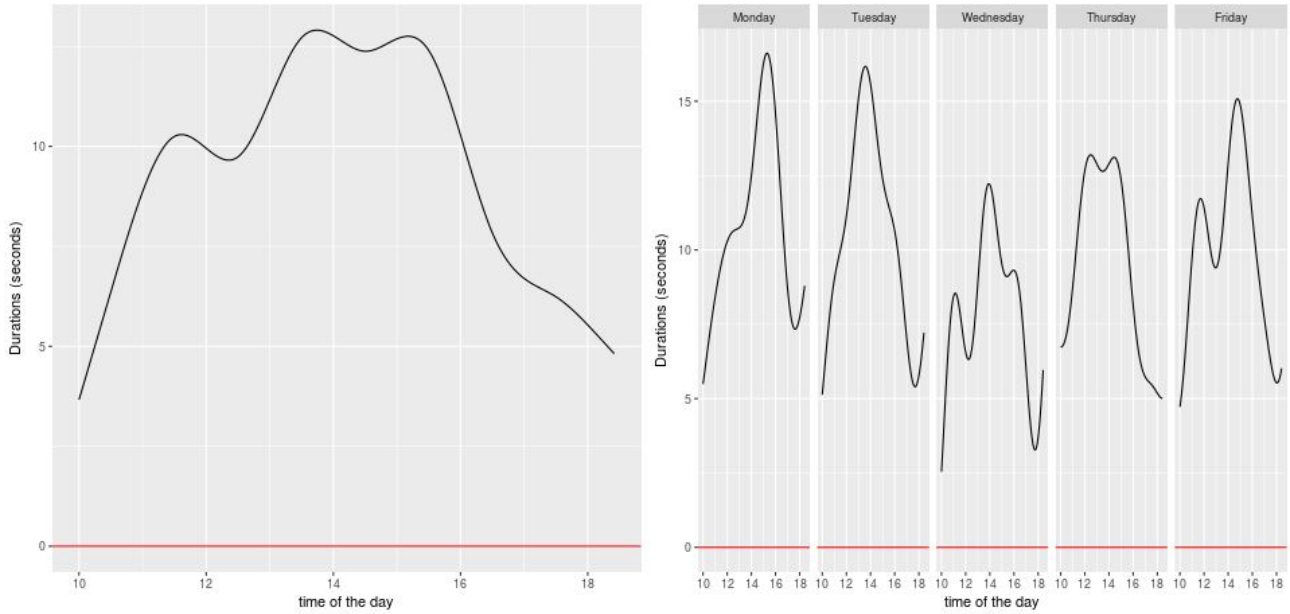



Figure 1.3: Daily and week day duration pattern from H-1398 ICBC Sep 2017

mother and daughter events where the mother events can statistically be expected to create a defined number of daughter events, the daughter events then create subsequent predictable generations of events. Previous to this models, we considered to describe events that arrived independently to another one at either a constant rate(the homogenous Poisson process) or decided by a specific variable rate (for the non-homogenous Poisson). But for certain applications, an event arrived probability observed by another events in the near future such as the expectation of aftershocks when modeling seismology. The model that A.G Hawkes [38] proposed has become an excellent model to describe the combination of arrival times of events. It was initially applied to a study in earthquakes but has gained attention in the financial sector in recent years. The reason for this is mostly down to its self-exciting ability, and it has been the most consistent at dynamic asset pricing. The model can be observed for example in the real stock market when situations of crisis are approaching it is still able to continue to watch the decline in asset prices. Since many types of financial market events extreme return occurrences or order submissions are clustered in time, Hawkes' process can be a good tool in finance too.

In this section, we begin to introduce a specific class of Hawkes' process where the event arrived rate explicitly dependent on the past events, the most widely used and most well-known self-exciting process. Because the Hawkes process is a self-exciting process, the arrival rate of viewing events $\lambda(t)$, a measure of how likely a viewing event will occur in a

infinitesimal interval around time $t$ as

$$\lambda(t) = \mu(t) + \int_0^t g(t-u)dN(u) = \mu(t) + \sum_{t_i<t} g(t-t_i). \tag{1.12}$$

where $\mu(t)$ is a deterministic time function, and in the Hawkes' model, $\mu(t)$ also can be defined as a parameterized covariant function, $g(s)$ is a weight function. The further step of $\lambda(t)$ can be expanded as

$$\lambda(t) = \phi(\mu(t) + \sum_{t_i<t} g(t-t_i)). \tag{1.13}$$

Where $\phi$ is a non-linear function that is used to ensure that the final result is positive, avoiding the density function taking a meaningless negative value. The most widely used weight function $g$ is proposed by Hawkes [39]

$$g(t) = \sum_{j=1}^{K} \alpha_j e^{-\beta_j t}. \tag{1.14}$$

Among them $\alpha_j \geq 0$, $\beta_j \geq 0$,$j = 1, 2, ..., K$, $K$ represents the order value of the whole process. For process with $K \geq 1$, the distance weights in the density function are in the form of exponential failure. In order to better illustrate the rules, this thesis imposes constraints on the coefficient $\beta_1 > \beta_2 > ... > \beta_p$. At the same time, for the sake of the stability requirement, we need $0 < \int_0^\infty g(s)ds < 1$, that is $\sum_{j=1}^{K} \frac{\alpha_j}{\beta_j} < 1$. When the Hawkes' model is used to describe the financial market data, there are a large number of artificial trading time intervals in actual stock trading, such as trading suspension days, weekends, holidays, especially the lunch break of Mainland China stock market and so on. In order to ignore these effects, we remove all non-trading periods, and connect all trading material together.

A univariate Hawkes' process is a linear self-exciting point process with an intensity given by

$$\begin{aligned} \lambda(t) =& \mu + \int_{0<t} g(t-s)dN_s \\ =& \mu + \sum_{i<t} \alpha_i e^{-\beta(t-i)}. \end{aligned} \tag{1.15}$$

where $\mu$ is a baseline intensity describing the arrival of exogenous events and the second term is a weighted sum over past events. The kernel $g(t-t_i)$ describes the impact on the current intensity of a previous event that took place at time $t_i$ . The most widely used kernel

of Hawkes' model is known as the exponential kernel, this property in a stochastic process is stated to predict the probability that a daughter event is created by a mother event decreased exponentially over time. This kernel is named to form a power-law and commonly applied to work in geophysical simulations. In contrast, empirical kernel described a simulation where occurred events cause new events at a much later time, that is to say the mother event still trigger a daughter event after a long period. This is known as a "long memory kernel." In a financial context we would forecast from live market data the number of events to come, to do this we need to estimate the parameters. This can be done by using the Maximum Likelihood Estimation method. If we knew the clustering, we would easily to calculate the average number of endogenous daughter events triggered based on immigrant mother events. To do this, we assess Hawkes' process by using the Maximum Likelihood (MLE) estimator. The estimation procedure in Hawkes' self-exciting process was described by Ozaki [48]. Ogata [45] researched on the asymptotic property of the MLE estimator for the Hawkes process.

Two-dimensional Hawkes' processes have exhibited exciting features that could contribute to financial modeling. In self and mutually exciting point processes that model the arrival times of an order book such as bid-ask order books are difficult. A fuller econometric framework was derived by Bowsher [9] who simultaneously analyzed trade times and mid-quote changes with a multivariate method. We are more interested in modeling events, and since traders usually minimize their market impact by splitting up their orders according to liquidity. The trades may contain information or time interval in order books dynamically.

Fig 1.4 shows a typical realization of a multivariate Hawkes' processes. In this thesis, we show that a two-dimensional Hawkes' process do well of A-H stock market data. The cross-influences of bid-ask both side trades are weaker according to Bowsher [9], since the maximum likelihood calibrations improve the statistical performance with deterministic time-dependent based intensity.

The thesis might serve as a road-map for the necessary material to learn how Hawkes' process works and its flexibility, and its potential utilization in many different fields. With the ever expanding amount of electronic data available in society today, it will be most likely developed and demanded for rapid and precise high intensity event processing modeling.

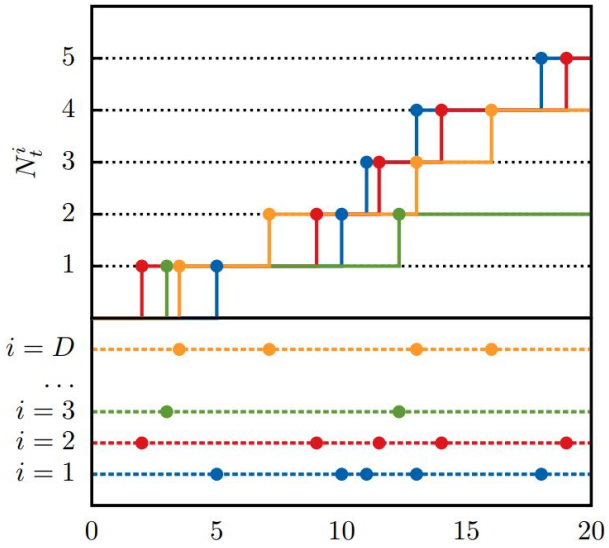



Figure 1.4: A realization of a multivariate Hawkes' process. The dots represent individual events, while different rows refer to different $i$ coordinates, and horizontal axis shows time $t$.

## 1.2 Order Analysis

In this thesis, we choose two set of $A-H$ stocks as Industrial and Commercial Bank of China Limited (ICBC) and Huaneng Power International Inc. four months trading data. We want to understand statistical characteristics of order flow. What is the interval between orders and orders? How long is the time interval What's the distribution of the time interval? To answer these questions, we do some statistical analysis of the interval between orders. The implication of liquidity is when you want to trade, you have the ability to do large transactions quickly. It consists of three elements, speed, depth, and width. The depth is mainly related to the number of orders. The more orders both parties have, the more they can afford the buy or sell pressure.

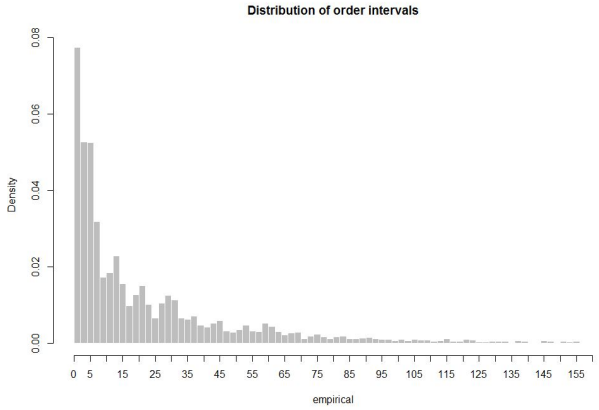
We choose Commercial Bank of China Limited (ICBC) as an example to analyze trading orders interval $\Delta T$. Firstly, we have a look at H-share 1398 orders performance. It can

| *Min.* | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|---|---|---|---|---|---|
| 1.00 | 4.00 | 13.00 | 37.90 | 35.00 | 50.00 |

Table 1.3: H-share 1398 orders performance

be seen from Table 1.3 that the 75 quantiles are around 35 seconds, which means that the trading of H shares 1398 is not very frequent and the liquidity is not very good. Above is the layout of order interval, due to a lot of intervals less than 38 seconds, we select the interval

(a) Distribution of order intervals




(b) Empirical vs Exponential

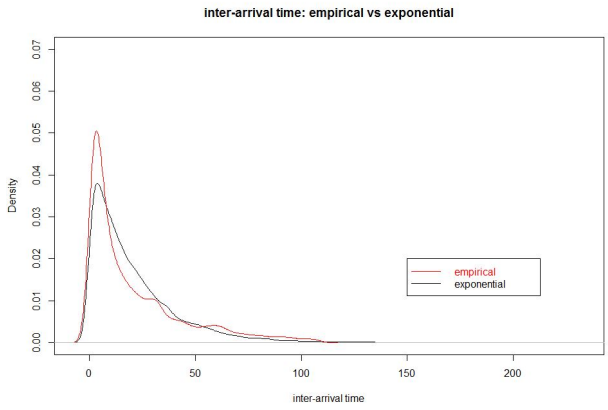



Figure 1.5: Panel (a) shows distribution of order intervals Panel (b) shows empirical vs exponential inter-arrival time. All trading orders are from H-share 1398 Aug to Nov 2017 among four months.

of less than 38s as visualization, we can observe distribution is highly biased, most orders intervals are less than 35 seconds. The exponential distribution Fig 1.5 Panel(b) is used to fit the order interval of less than 135 seconds. It is found that the speed of the actual data attenuation is much higher than that of the corresponding optimal parameter simulation exponential distribution. Since exponential distributions decay at a very fast rate, it is a little bit more telling of how small the spacing is, and that the core of the exponential distribution is fit for the Hawkes process.

Now let focus on another A shares 601398 monthly data on Table 1.4. It can be seen

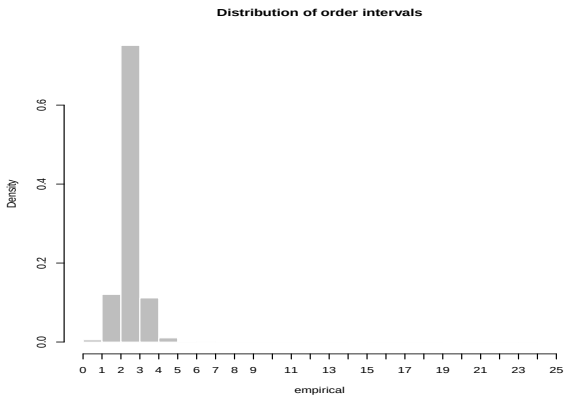
| *Min.* | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|---|---|---|---|---|---|
| 0.00 | 2.00 | 3.00 | 23.90 | 5.00 | 5411.00 |

Table 1.4: H-share 601398 orders performance

from Table 1.4 the 75 quantiles are around 5 seconds, which means that the trading of H shares 1398 is frequent and the liquidity looks good. It is apparent from the figure that the trading of A-share 601398 is more frequently than H shares 1033, and the liquidity also seems a little better. The maximum time interval is around 5411 seconds. We use a two-point search to extract the transaction time for each trading day, but did not propose a lunch break. We all know that A-shares have a one-and-a-half hour lunch break, so this huge interval is the lunch break. Since we choose to analyze the transaction time interval less than the mean as a visual analysis, the maximum impact is not significant. We can observe distribution is highly biased, most orders intervals are around 3 seconds. The fitted order interval distribution of A shares 601398 is in Fig 1.6, most order intervals are around

1.5 seconds. Actually it quite likes Poisson distribution. However, the distributions also decay at a fast rate. There are more evidences about the scale of the small gap, which can be said the arrival of events has a high degree of agglomeration and cannot be characterized merely by the Poisson process.

(a) Distribution of A-601398 order intervals




(b) Empirical vs Exponential

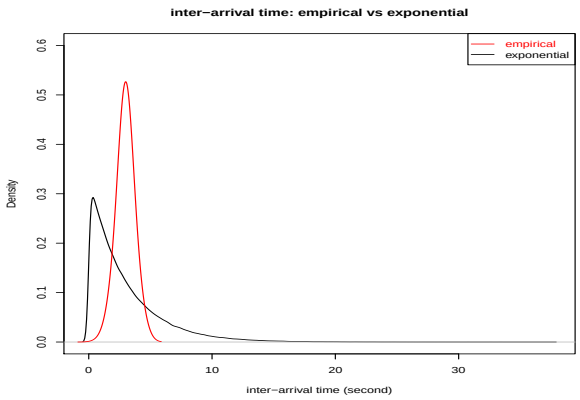



Figure 1.6: Panel (a) shows distribution of order intervals Panel (b) shows empirical vs exponential inter-arrival time. All trading orders are A-share 601398 from Aug to Nov 2017 among four months.

## 1.3 Branching Ratio

The ever-increasing branching ratio has attracted more attention in recent years. Several financial literatures have paid some attention on high frequency trades activities related with branching ratio, and try to find some consistency. However, several measures of financial market quality such as price discovery, volatility and liquidity are influenced by high frequency trades, these theories were discovered and analyzed by Zhang [56], Brogaard [13], Hendershott and Riordan [40] and many offer papers. In general, the conclusions of these studies in the sense that high frequency trades associate with increased liquidity, decreased volatility and prices are becoming more informative.

A model-independent approximation for branching ratio can be introduced with Hawkes' self-exciting process. The estimator knows only the mean and variance of the event counting in advanced over a large time window. Statistics are readily available from empirical data for this. The method we proposed significantly simplified the estimation of the process as a proxy for market endogeneity and can formerly estimate using the numerical likelihood maximization.

If we interpret branching ratio as derived from Hawkes' process, then $n = \int_0^\infty g(\nu)d\nu$ provided a direct quantification of the kernel and defined precisely of diverging trading activity in the absence of any external driving. If we interpret branching ratio as a measure of market quality, it makes sense that a low branching ratio reflects a healthier market where price changes are driven by exogenous events, while a high branching ratio reflects a less healthier market where price changes are driven by "positive feedback mechanisms" and herding behavior, as proposed in Filimonov and Sornette [28].

The branching ratio can be defined as $n$, where $n$ is the average number of daughter events triggered by a mother event. A branching ratio smaller than one is known as sub-critical. Therefore, $n$ quantifies market reflexivity in an elegant way. Three regimes exist depending on the branching ratio value: a system can be critical ($n = 1$), super-critical ($n > 1$) or sub-critical ($n < 1$):

- a sub-critical regime ($n < 1$), where families die out almost surely,
- the critical regime ($n = 1$), where one family lives indefinitely without exploding. In the language of Hawkes process, this requires $\mu = 0$ to be properly defined and it is equivalent to the Hawkes' process without ancestors *Brémaud* and *Massoulié* [10],
- the explosive regime ($n > 1$), where a single event triggers an infinite family with a strictly positive probability.

In the subcritical regime, immigrant events (mother events) can generate a limited number of daughter events and eventually the system won't die out. However, a single immigrant could lead to an explosion (of infinite events) as each event statistically could create multiple daughter events. Low branching ratio indicates that the price changes more influenced by external factors than internal, an asymmetric information leads to different trading strategies performed by diverse types of investors. Stocks with a low branching ratio are more appropriate for speculation than those having a high branching ratio, as they are more likely to be volatile due to external factors than their own fundamentals. We try to explore the connecion between A-H stock market endogeneity and comparing the intra-day dynamics of the branching ratio with a "crude" measure of stock liquidities, which is defined as the ratio of orders per trading day.

## 1.4 Bid-ask spreads ("BAS")

The content of the transaction data in the estimation model is not only about transaction price but also about the duration of the trades, even including the rate of return, the volume of the transaction and the duration size of the transaction. A bid-ask spread closely related to the price and also data transaction content, so studying on high-frequency is actually the study of discrete variables of different intensities over time.

From the order aspect, buyers and sellers negotiate and bargain for a stock relied on computer networks for communications technology as a basis for the realization of the online transactions. All series of traders' transaction activities (buy, sell or cancel) regarding the stock are recorded in order book. Best ask price (or ask quote) is the price at which buyers most willingly to accept, whereas best bid price (or bid quote) reflects the situation on the opposite side. The unconformity of ask and bid quote is known as the bid-ask spread. When best ask equals to the best bid price, it is called that BBO (best bid and offer) is attained. It should note that the best ask has spread with the best bid, if not, everyone would buy orders at the best bid and sell them at the best ask. Therefore, the existed arbitrage opportunities result in BBO almost attained in quite a short time according to the low latency trading frequency.

The BAS affect is closely related to the liquidity of the A-H stock markets, but first it is crucial to understand the mechanism behind the way the spread fluctuates. In addition Roll [50] claimed that in a stock market studying BAS are determined efficiently by the market-maker through their function. When predicting BAS, it thought that a short-term analytical tool is more useful. As experience found in the most recent financial crisis, liquidity would break down all of a sudden and unhedged credit risk in very short term. BAS may provide us a useful tool to predict liquidity levels in the short run.

## 1.5 Outline of the thesis

This work differs from previous studies in that, firstly, they are built directly from the transaction arrival process itself. Secondly, in recent years, Hawkes' process has been applied in modeling of the transaction arrival orders in financial industry. Specifically, we first

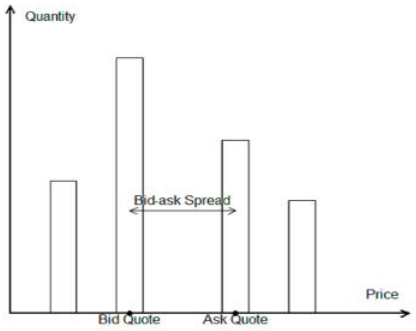



Figure 1.7: A simplified order book.

simulated the arrival of a transaction using a univariate Hawkes' process with an exponential response equation and used this model to establish the expected intensity and branching ratio. The third is to use the two-dimensional Hawks model to analyze the trading capabilities of transactions to verify whether Hawkes' model constructed here can efficiently explain the liquidity of A-H shares. The thesis is organized as follows.

Chapter 2, we present a review on point process first, and then introduce a Hawkes' process with its applications.

In Chapter 3, we build up an comprehensive one-dimensional Hawkes's model, we also describe a simple design of the memory kernel and results on events simulation. We drive the likelihood of thinning algorithms and describes consistent and asymptotic properties of the maximum-likelihood estimator. At last of this chapter, we introduce predicting setup. At last, we compare differences between ACD Model and Hawkes' model, and show Hawkes' advantage.

In Chapter 4, we develop a two-dimensional Hawkes model and it includes both cross-exciting and self-exciting parts as the corresponding influence of bid-ask market orders on the subsequent orders.

In Chapter 5, application of the one and two-dimensional Hawkes' processes on A-H shares in Mainland and Hong Kong stock market including the underlying branching ratio structure are given. Moveover, we propose a confusion matrix related to bid-ask expected intensity trading strategy based on two dimensional Hawkes' model.

Chapter 6 includes summary and introduction of future works based on this thesis.

# Chapter 2

# Literature Review

There are many research on point process. The simplest case of point process is the Poisson process, which is divided into homogeneity and non-homogeneity. Poisson process uses a fixed ratio or a time-dependent deterministic function to describe the arrival rate of events. The Poisson process does not capture the aggregation effect of the transaction, while Hawkes' process can do it. In countries where HFT is well developed, Hawkes' process is applied to trading order modeling. Hawkes' process is a point process with the characteristics of self-exciting. In Hawkes' process, events arrive at a certain rate. This rate can be fixed, or can also be time-varying, and an arrival of an event can trigger the arrival of a second event, and then trigger the arrival of a third event. This is a cycle of self-motivation. This feature of the Hawkes' process makes it possible to capture transactions agglomeration. Hawkes' model was used to model data from the New York stock exchange and the results demonstrate Hawkes' model has the ability to analyze trading and quotation behavior. According to order classification, we could believe that Hawkes' processes have significant advantages in the fitting and prediction of financial data, both in theory and in practice. The development of the point process will be reviewed first.

## 2.1 Review on Point Process

A random point process is defined as a model that can effectively generate random events. There are many aspects of the event which can be described by the time, location,

state etc, or can be expressed as a specific point in space. A process of self-exciing is when the probability of an event occurrence depends or is influenced by the previous event. This type of process can be regarded as a typical Poisson process. The characteristics of this type of self-exciting process follows a propagation trend more popular. The strength of the momentum is not constant but depends on itself and the previous trend as this will dictate subsequent event behavior. In a Poisson process, however the intensity is constant and the events follow one another with 'no memory' affect. In reality, there are many situations that do not conform to this assumption. For example criminal behavior does not follow this as many criminals tend to continue committing crimes subsequent to a successful crime where they were perhaps not apprehended in a neighborhood. Indeed the police anticipates that criminals re-offend and often bring them to justice in order to reduce crime rates. In finance trends are tracked by traders who aspire to make the markets grow in a larger number of orders quickly.

Until recently there are only a few models capable of accepting high frequency variation. They most rely on discrete time models based on either aggregated dynamics of the intervals or take into account of the continuous time events. These models are often referred to as "trade time" or "business time" models. Hasbrouck [35], Engle and Russell [23] through the nineties, suggested that modeling financial data at the transaction level has an advantage known as "point processes." These applications in finance were ongoing and became a very popular research topic. We would refer to a recent review of Bowens and Hautsch [7] on point process models.

In finance, point processes are mainly of in financial markets because of some stylized facts related to the microstructure. Point processes offer a parsimonious way of modeling the duration between events. These have been used extensively to model the arrival rate of quotes and prices in different markets. There is a distinguished difference between those using continuous point processes such as in ACD and ACI models and those using discrete time point processes like the Hawkes model.

ACD model was first proposed in the context of market microstructure by Engle and Russell [23]. Russell [51] proposed a mode for directly intensity calculations as the autoregressive conditional intensity model. The ACD model together with durations still remain today and the most well used model in high frequency econometrics today. The set of models

and the process is sometimes defined by means of its "hazard function", that is to specify certain condition of inter-events (or the intervals). Furthermore, the ACD model can also represent the "intensity function" representing the conditional probability model density of the event occurrence and subsequent events (see Daley [17] for fuller details on mathematical applications). The difference between ACD model and Hawkes' model is that the Hawkes' model provides a simple description of the conditional intensity but it is not the same for ACD model. The Hawkes' model has the advantage that it is able to treat multi-variable point processes easier than ACD model. More details and discussions on the disadvantages can be found in Russell [51] and Bowsher [9].

## 2.2 Review on Hawkes' Process

Ogata [47] had previously used Hawkes' process to analyze earthquake data. In 2004, Sornett applied Hawkes' process to model book sales through Amazon data, and studying the dynamics of YouTube video views.

Even though Hawkes model had been around since the 1970s, but only recently it has applied in finance. Bowsher [9] first applied Hawkes' process to trading events related to financial markets. Bowsher [9] developed a generalized Hawkes' model described in terms of conditional intensity, and applies bivariate versions of them to explain the trades and mid-quote changes in one stock (General Motors Corporation) over 40 trading days in the year 2000. Bowsher shows there is a two-way interaction between trades and mid-price changes which increases trade intensity. Bowsher's work used data from Trade and Quotes (TAQ) database where the timestamps are stored to the precision of one second. As it happened, there are some issues with multiple events within the same time-stamp, and common problem is that high frequency data will have a lot rounded nearest second. We need distinguish them to solve random uniform components of time. In the past, the number of events within one time stamp represented a very low percentage of around 0.26% of all events changes. However if one uses more recent data, when high frequency traders increased the number of events within a given second, problems will arise. For example, using TAQ on Yahoo stock in one trading day in February 2010 around 30% of all prices changes occurred within the same second. Another earlier application of Hawkes' process that related to finance

was by Hewlett [41]. He used bivariate Hawkes' model to the flow of orders in the FX market. He envisioned that the model could predict future trading intensity rate based on historical trades. Sometimes, takers are looking for a way to fill larger orders in transactions by splitting the orders through running of the market-makers. To split orders up, they must identify the pattern of buy and sell to get better deals, otherwise, if they fill the large orders at once, this will influence the market price significantly.

In a notable research, Bowsher [9], identified the flexible advantage of applying a class of multivariate counting processing specified by a conditional intensity vector, transpiring to a bivariate Hawkes' processes to model the joint dynamics of trades on New York stock exchanges including mid-prices changes. He noted that by involving the intensity vector that a linear function can be generated from past events. The popularity of the Hawkes process that the process can explain the demand in high frequency finance by its simplicity and flexibility, this was anticipated by Bowsher [9]. Hawkes' model therefore can easily account for more types of events interactions which are influenced by some intensive factors and show some statistical stationarity events over time. By using formulae on parameters with complex dynamics changes, straightforward interprets clustering of the behavior of modern day markets.

The length of the asset price duration cannot be characterized by traditional jump diffusion according to Ait-Sahalia [1], due to the extent of its spread. He constructed the Hawkes' jump diffusion model in a study of stock index performed in five difference global financial regions. Filimonov and Sornette (2012) used Hawkes' process to get a measure of George Soros's "market reflexivity" and made an empirical study with small $S\&P500$ index future. The reflexivity theory states is a link between the market participants' perceptions and the market itself, there is an interaction between the impact of the performance of the asset itself, it suggested that Hawkes' process can do quantitative analysis. Bormetti [8] combined a model to study the synchronization of 20 stock prices with Hawkes' process on the Italian stock exchange. Ait-Sahalia [3] had used Hawkes' process for credit risk and combined it with impulse response analysis in an attempt to examine how credit events affect the prices of the European credit default swaps (CDS). Grothe [33] combined an extreme value theory with Hawkes' process to characterize the combination of extreme event arrival times in the financial time series. Dramatically, Hawkes' process is applied to other financial

topics such as VaR estimation by Chavez-Demoulin and McGill [15], trade-through modeling by Muni Toke and Pomponio [54], portfolio credit risk by Errais [24] or financial contagion across regions by $\ddot{Ait} - Sahalia$ [2] and across assets by Bormetti [34].

Related to branching ratio, which is drived by Hawkes' model, Filimonov and Sornette [28] proposed to measure market endogeneity which was named as "reflexivity" by George Soro. By using their estimation of maximum likelihood on a data quota on E-mini $S\&P500$ stock they spanned from the year 1998 until 2010 and found that before 2000 the market endogeneity was relatively lower with a branching ratio of 0.3, however after 2004 it increased to 0.9 Filimonov and Sornotte [28] observed that the branching ratio remained stable growth through the periods of market stress. Filimonov and Sornette [28] took notice of that in the general branching ratio average coincides with the rise of activities by high frequency traders. Even in the Flash-Crash itself there was no relation to HFT but to the speed of the automated trading system that exacerbated the conditions in the market on that day. Evidence presented from Filimonov and Sornette [28] said that the branching ratio increased over time and it is approximately from 0.2 at the beginning of the sample data in 1998 and 0.8 by the end of the sample data in 2010. They concluded that there was correlation between the increase in HFT activity and the size of the branching ratio.

## 2.3 Measurement of Market Liquidity

Liquidity is a relatively broad concept and it can be understood from different perspectives. Amihud and Mendelson [4] completed the transaction and found the time or a cost of needed for an ideal price as a measurement of liquidity. Lippman and McCall [43] argue that if an asset can be sold quickly at the expected price, the asset has liquidity. Schwartz [36] had a similar view. He believed that liquidity has ability to quickly implement with a reasonable order price on the market. Massimb and Phelps [44] studied with two characteristics of the liquidity profile descriptions were: firstly, the timeliness, namely the ability to execute order in a timely manner; second is the market depth, that is, the ability to execute orders without causing significant price fluctuation. Currently, most of the liquidity indicators are constructed on these three main aspects, such as the bid-ask difference belongs to the transaction cost aspect, the turnover rate used by Amibud [5] belongs to the transaction

speed aspect.

Amihud and Mendelson [4] investigated liquidity in the asset pricing framework. Amihud and Mendelson used Fama-MacBeth's [27] method to construct the portfolio of article. They discussed the relationship between bid-ask differences and found a positive relationship between portfolio annual rates of return and bid-ask spreads. Eleswarapu and Reinganum [21] studied the effect of seasonal effects on bid-ask spreads and yields. They constructed the combination according to the criteria of Amihud and Mendelson [4], and made certain amendments to reduce the errors, but excluded small company stocks. Eleswarapu and Reinganum [21] found that the relationship between the bid-ask spread and the returns on assets were mainly limited. Eleswarapu [20] selected Nasdaq data from 1973-1990 as a sample to verify the liquidity premium proposed by Amihud and Mendelson [4]. The article gave positive support result. Eleswarapu [20] speculated that Nasdaq's dealers do not face competition from limit orders and floor traders, so as compared to the New York stock exchange. Research showed that the liquidity indicator is statistically significant on the New York Stock Exchange and the American Stock Exchange, but not on the Nasdaq. In spite of this, there are certain problems with the transaction volume as a liquidity indicator. Out of the difference, the Nasdaq dealer's internal spread can better represent the actual transaction costs.

Brennan and Subrahmanyam [12] divided transaction costs into fixed and variable costs ingredient. The portion of fixed costs is only related to the length of time, while the portion of variable costs is related to the size and length of the transaction. Brennan and Subrahmanyam [12] added liquidity variables under the framework of Fama and French [26]. Result showed that the changes in transaction costs are very significant and the fixed parts are not significant. One of the possible reasons they believed that this result showed they did not choose the proxy variable. Brennan's and Subrahmanyam's speculations had guided the later research to explore the agency variables than bid-ask spreads, such as transaction volume and turnover rate. Brennan, Chordia and Subrahmanyam [11] were still under the framework of Fama and French [26], but the liquidity indicator they used was not the bid-ask spread, but the transaction volume.

Datar, Naik and Radcliffe [19] changed the turnover rate (the trading volume divided by the total number of outstanding shares) as liquidity proxy variables. They used the

natural logarithm of the company's stock market value for the previous month as a scale factor. The $\frac{B}{M}$ factor was constructed just like Fama and French [25] and the combined $\beta$ was the same as Amihud and Mendelson [4]. The results of the article showed that liquidity can still partly explain the cross-sectional stock returns even under the control of scale $\frac{B}{M}$, and $\beta$. In addition, unlike Eleswarapu and Reinganum [21], they found that the liquidity effects not only existed in January but were reflected in various months of the whole year. Chui and Wei [53] studied stocks on the New York Stock Exchange, American Stock Exchange, and Nasdaq Market using variables and methods similar to those of Datar, Naik and Radcliffe [19], and demonstrated the turnover rate and $\frac{B}{M}$ variable is significant when interpreting the cross-sectional yield. However, their discovered in the seasonal aspect of liquidity effects is different from previous studies. They found that significant liquidity effects existed seasonally. Amihud [5] divided the absolute value of daily stock returns by amount of the transactions, and treated it as liquidity. This article found that in the time series, the lack of expected liquidity also affects the stock's excess return. So, the expected excess returns is a compensation for the low liquidity of stocks. Amihud's indicator has been widely used in the subsequent research.

In addition to common indicators, such as bid-ask spreads, turnover, and trading volume, scholars have more details of high frequency on orders and transactions were studied. Garvey and Wu [30] conducted a detailed of the execution orders during the trading day. They made orders submitted during the middle period were now executed slower than orders submitted during the opening and closing hours, but orders submitted during the middle period had lower execution costs. There is an offsetting relationship between execution speed and execution cost of orders, and this relationship has different expressions at different time intervals within a trading day.

In the face of a variety of liquidity indicators, Goyenko, Holden, and Trzcinka [31] made a series of comparisons. They used the high-frequency index as a comparison standard, compared and analyzed all widely used liquidity indicators from the transaction cost and price aspects. The results showed that different indicators have different performance under different conditions, and should be selected according to the research questions. In recent years, Mainland China financial market has experienced unprecedented development. Not only has the total volume of transactions reached an unprecedented level, but the transaction

frequency has also multiplied. The transaction itself contains information on liquidity, such as the number of transactions occur within a unit of time, or the probability of the next transaction occurring after one transaction occurs. To better explain maekrt liquidity, the next question in research should be, how to model the arrival of transactions.

# Chapter 3

# One-Dimensional Hawkes' Self-exciting Process

## 3.1 Introduction on Hawkes' Model

Random point process can effectively describe the generation of random events. Many random phenomena occurs at the moment, place, state, etc, can be expressed as a point in a certain space. The self-exciting Hawkes process describes the occurrence probability of the current event depends on the occurrence of previous events. In natural sciences, Hawkes' model play an important role in modeling the emission of particles from a radiating body. Hawkes' process can be regarded as a point process that satisfies the typical stochastic process. Its intensity rate is not a constant, but depends on itself previous events and will affect the follow-up events. In homogeneous Poisson process, the intensity is constant and even non-homogeneous Poisson the intensity not constant but the occurrence of the events follow the principle of "no memory". So Poisson process does not capture the aggregate effect of the transaction but Hawkes' process could. Especially in the stock market, the traders will track the big order trend that make the whole market into a larger number of orders with a short time. In Hawkes' process, events arrive at a certain rate, which can be fixed time-varying, and when one event arrives, it can trigger the arrival of the second event, and then trigger the arrival of the third event. This is called self-exciting, and because of this characteristic, Hawkes' process could captures the clustering effect of transactions.

This chapter provides an overview introduction on Hawkes' self-exciting process, for modeling discrete, inter-dependent events over continuous time. We start by reviewing the definitions and the key concepts on one-dimensional Hawkes' process. We then introduce Hawkes' process its event intensity function, as well as schemes for events simulation, parameters estimation, describes consistent and asymptotic properties of the maximum-likelihood estimator and the goodness of fit on Hawkes' model. We also describe a simple designed of the memory kernel, and some results on estimating parameters and predicting.

## 3.2 Theoretical One-Dimensional Hawkes Model

### 3.2.1 Hawkes' Process

Hawkes process, a self-exciting process, serve as epidemic models, since a number of events occurrence increased the probability of further events. When using socio-economic systems in modeling complex point process, a Hawkes' process has become the best approach because of its flexible and modest structure. In the modern world, it is applied on earthquake modeling, genomic events along DNA, brain seizures and the spread of crime and extreme events, also widely used in the financial modelling such as in the probabilities of credit defaults occurring. It is also used for financial process especially in high frequency finance, but lower frequency data is also possible. The general method counts arrival of events, such as a Poisson process with one parameter. And even more, a simple Poisson process on the other hand does not take the history of events into account. For our purposes, however, Poisson process is too simple as we need a way to explain the clustering of events occurrence. Based on this, an extension of the basic Poisson process which aim to explain clustering of events coming is called Hawkes' self-exciting process. Self-exciting models like Hawkes' are widely used in miscellaneous sciences, some examples are in seismology (modeling of earthquakes and volcanic eruptions), ecology (wildfire assessment), neuroscience (modelling of brain spike trains which bunch together), even modeling of eruption of violence (on modeling civilian deaths in Iraq, and on crime forecasting), and finance market trading orders. The Hawkes process calculates the time-varying intensity of event occurrence rate, which is determined by the history happened events of the process.

Inorder to describe the aggregation of the occurrence events. We need to extend the traditional Poisson model, by fixing $\lambda$ to a certain value, but let $\lambda$ itself become a function of time, namely the $\lambda(t)$. For any $\Delta t \to 0$ and the rest of the hypothesis is similar to Poisson process

$$Pr(N(t, t + \Delta t] = 1) = \lambda(t)\Delta t. \tag{3.1}$$

a generalized intensity function of point process is

$$\lambda^i(t) = \mu_i + \sum_{j=1}^{K} \int_{u<t} g_{ij}(t-u) dN_u^j, \quad \mu > 0. \tag{3.2}$$

Where $\lambda^i(t)$ represents the instantaneous intensity at time $t$, $\mu_i$ is the base rate of Hawkes' process,$\alpha_j \geq 0$, $\beta_j \geq 0$,$j = 1, 2, ..., K$, K represents the order value of the whole process. For the $K \geq 1$, the distance weights in the density function are in the form of exponential failure

$$\int_0^\infty g(\nu) d\nu < 1. \tag{3.3}$$

such that negative-valued probabilities could be restricted to a relative low level and $g(\nu)$ expresses the positive effects on the current value motivated by past events of Hawkes' self-exciting process.

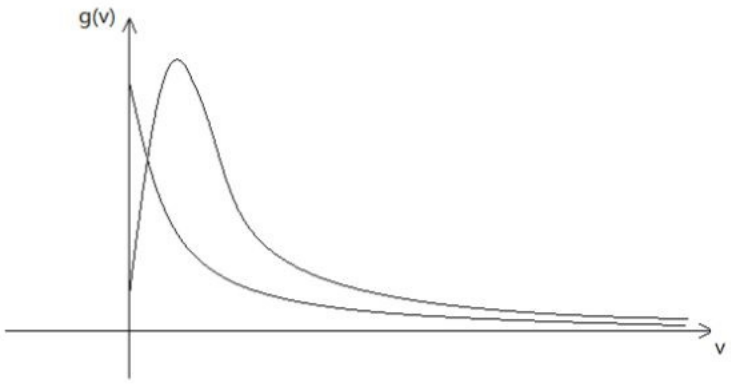



Figure 3.1: Samples of $g(\nu)$

Now we focus on function (3.2) whose intensity function restrictions guarantee a positively respond function $g(\nu)$ beginning at zero value, when $\nu < 0$. By observing on Fig 3.1, as $t$ is moving away from 0, the quicker $g(\nu)$ declines, the more larger the integral term in

function (3.3) with the intensity will be. Exceptionally if there exists a hump in $g(\nu)$, the more height and thickness a hump be, 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 a hump effects the intensities of longer time period hereafter than a $g(\nu)$ without a hump, and the specific properties of the hump determined the extent of "longer" time. When $g(\nu) = 0$, the Hawkes' process degenerates into a Poisson process with constant intensity rate $\mu$. When $g(\nu) > 0$, it is shown that the Hawkes' intensity process has self-excitation, which usually manifests as aggregation of arrival time. If Hawkes' process is extended to high-dimensional, it can also depict the time gathering in the arrival space. In fact, the Hawks process can be viewed as a branch process with immigrants. Among them, the event caused by the reference intensity arrives called "immigrants", and any event that arrives at $u$ time-stamp takes the intensity as $\int_{-\infty}^{t} g(t-u)dN$, resulting in "offspring".

Hawkes' model allows estimating dependencies without imposing parametric time series structures. The dependence is rather specified in terms of the elapsed time since past events. Hence, the marginal contribution of previous events on the current intensity is independent of the number of intervening events. The exponentially decaying response function

$$g_{ij}(t) = \sum_{k=1}^{K} \alpha_{ijk} e^{-\beta_{ijk}(t)}, \quad \mu > 0$$

and

$$\sum_{k=1}^{K} \frac{\alpha_{ijk}}{\beta_{ijk}} < 1.$$

It is because, on one side, 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 $\alpha$ and $\beta$ denote the scale and location parameters respectively which determine the "portrait" of intensity function. $K$ is the order of Hawkes' process which is picked up exogenously and means the number of superposed different response functions with divers parameters. Here we follow Venables and Ripley [55] proposed a good selection on $K = 1$. They compared selection on different $K$'s, and found that who has $K = 1$ is the minimum AIC from a whole selection of Hawkes' model. The

simplest version is when $K = 1$, and Hawkes' model is defined as

$$\begin{aligned} \lambda^i(t) &= \mu_i + \sum_{k=1}^{K} \sum_{j=1}^{I} \int_{u<t} \alpha_{ijk} e^{-\beta_{ijk}(t-u)} dN_u^j \\ &\Downarrow \\ \lambda^i(t) &= \mu_i + \sum_{j=1}^{I} \int_{u<t} \alpha_{ijk} e^{-\beta_{ijk}(t-u)} dN_u^j \end{aligned} \tag{3.4}$$

the one-dimensional Hawkes model is

$$\lambda(t) = \mu + \int_{u<t} \alpha_t e^{-\beta(t-u)} dN_u. \tag{3.5}$$

After time discrimination, the one-dimensional Hawkes intensity function can be rewritten as

$$\lambda(t) = \mu + \sum_{i<t} \alpha_i e^{-\beta(t-i)} \qquad \mu > 0, \quad \alpha > 0, \quad \beta > \alpha. \tag{3.6}$$

- $\mu$ = base rate of occurrence.
- $\alpha$ = the intensity jump right after an event occurrence.
- $\beta$ = the exponential intensity decay.

The first term $\mu$ in the Hawkes model is the base intensity rate that determines the rate of arrival of first events per unit of time. The response function then controls how offsprings are generated by first events and it is also the source of clustering in the model. Actually the response function includes two parameters $\alpha$ and $\beta$ which two define the clustering properties of the process. $\alpha$ describes the intensity jump right after one event occurrence, and $\beta$ describes the exponential decay. The parameter $\mu$ can also be explained like the intensity of exogenous events of the total process, for example bad news or good news. So we usually could see that $\frac{\alpha}{\beta} < 1$ guarantees the intensity decreases more rapid than new events increase, but if not the process could explode.

The unconditional expected intensity assumes as $E[\lambda(t)] = C$, which is a constant.

Thus

$$\begin{aligned} C = E[\lambda(t)] &= E[\mu] + E[\int_{-\infty}^{t} g(t-u)dN_u] \\ &= \mu + E[\int_{-\infty}^{t} g(t-u)\lambda(u)du] \\ &= \mu + C\int_{0}^{\infty} g(\nu)d\nu. \end{aligned}$$

which gives

$$\begin{aligned} C &= \frac{\mu}{1-\int_0^\infty g(\nu)d\nu} = \frac{\mu}{1-\int_0^\infty \alpha e^{-\beta t}dt} \\ &= \frac{\mu}{1-(-\frac{\alpha}{\beta}e^{-\beta t}|_0^\infty)} = \frac{\mu}{1-(0-[-\frac{\alpha}{\beta}e^0])} \\ &= \frac{\mu}{1-\frac{\alpha}{\beta}}. \end{aligned} \tag{3.7}$$

The stationarity condition for one dimensional Hawkes' process we show

$$\sum_{k=1}^{K} \frac{\alpha_{ijk}}{\beta_{ijk}} < 1. \tag{3.8}$$

Equation (3.7) gives the average intensity for a stationary one-dimensional Hawkes' process with $K = 1$, so the unconditional expected value of intensity process

$$E[\lambda(t)] = \frac{\mu}{1-\frac{\alpha}{\beta}}. \tag{3.9}$$

## 3.3 Simulated Univariate Hawkes' Process

### 3.3.1 Simulation Method of a Hawkes Process

After theoretical Hawkes' framework set-up, then we simulate one-dimensional Hawkes' process, by using Ogata [46] method, in which the author suggested a method based on Lewis and Shedler [42] thinning procedure of a non-homogeneous Poisson process to simulate the points sequentially. The purport of thinning method used an upper bound of the $\lambda$ first feed to the exponential distribution as a parameter in a given interval, and then the time of the

new event happened would be calculated. But the actual situation in the time interval of $\lambda(t)$ should be constantly decreased (kernel function distance be stretched), so we can only accept the new point by a certain probability. If it be rejected, then we need a new simulated point. In another words, there are some points need to be deleted that the simulated process should be closer to the truth. The simulation of a Hawkes process is mainly done to verify the result of maximum likelihood estimation. We need artificial self-exciting point process data to check the validity of the maximum likelihood method. Suppose that self-exciting point process data $t_1, t_2..., t_k$ are given, then let $F(t \mid t_1, t_2, ..., t_k, \theta)$ be conditional distribution of random variable of the interval between $t_k$ and the next event $t, (t \geq t_k)$ of the process, and let $f(t \mid t_1, t_2, ..., t_k, \theta)$ be its probability density function. The basic idea behind simulation follows from the conditional Hazard function

$$\frac{f(t|t_1, ..., t_n, \theta)}{1 - F(t|t_1, t_2, ..., t_k, \theta)} = \Lambda(t|t_1, t_2, ..., t_k, \theta).$$

Then we have

$$\log(1 - F(u|t_1, t_2, ..., t_k, \theta)) = -\int_{t_k}^{u} \Lambda(t|t_1, t_2, ..., t_k, \theta)dt = -\int_{t_k}^{u} \mu + \sum_{i=1}^{k} g(t - t_i|\theta)dt$$

since the time of $(k+1)st$ event $u$ satisfied the aforementioned equation, and $1-F(u|t_1, t_2, ..., t_k, \theta)$ is distributed uniformly on [0,1], generation of $t_{k+1}$ is performed by a random uniform number $U$, therefore $u$ is solved in the following equation to generate the time of the next event,

$$\log U + \int_{t_k}^{u} (\mu + \sum_{i=1}^{k} g(t - t_i|\theta))dt = 0.$$

Following the exponentially decaying response function $g(t) = \alpha e^{-\beta t}$, the last equation reduces to

$$\log U + \mu(u - t_k) - \frac{\alpha}{\beta}(\sum_{i=1}^{k} e^{-\beta(u-t_i)} - \sum_{i=1}^{k} e^{-\beta(t_k-t_i)}) = 0.$$

Consider the expression in the above equation, then we get

$$A(k) = \sum_{i=1}^{k} e^{-\beta(t_k-t_i)} - \sum_{i=1}^{k} e^{-\beta(u-t_i)}.$$

This can be written as

$$\begin{aligned}
&\sum_{i=1}^{k} e^{-\beta(t_k-t_i)} - \sum_{i=1}^{k} e^{-\beta(t_k-t_i)} e^{-\beta(u-t_k)} \\
&\Downarrow \\
&(1-e^{-\beta(u-t_k)}) \sum_{i=1}^{k} e^{-\beta(t_k-t_i)} \\
&\Downarrow \\
&(1-e^{-\beta(u-t_k)})(1+e^{-\beta(t_k-t_{k-1})}) \sum_{i=1}^{k-1} e^{-\beta(t_{k-1}-t_i)} \\
&\Downarrow \\
&A(k) = S(k)(1-e^{-\beta(u-t_k)}), \qquad S(k) = e^{-\beta(t_k-t_{k-1})} S(k-1) + 1.
\end{aligned} \tag{3.10}$$

By using above-mentioned, the function

$$\log U + \mu(u-t_k) - \frac{\alpha}{\beta}(\sum_{i=1}^{k} e^{-\beta(u-t_i)} - \sum_{i=1}^{k} e^{-\beta(t_k-t_i)}) = 0.$$

can be written as

$$\log U + \mu(u-t_k) + \frac{\alpha}{\beta} S(k)(1-e^{-\beta(u-t_k)}) = 0.$$

Where $S(1) = 1$ and $S(k) = e^{-\beta(t_k-t_{k-1})} S(k-1) + 1$

In order to solve the equation numerically, it requires the first derivative of the above equation. Here is the concrete algorithm which proposed by Ogata [46] as follows:

- Generate a random uniform distribution on interval [0,1].
- The whole process is simulated on time interval $[0, T]$.

  Secondly we could see the algorithm is given as following:

- Algorithm-Initialization

  1. Initialization: set $\lambda^* \leftarrow \lambda_0(0)$ and $n \leftarrow 1$.
  2. First Event: Generate a random uniform number $U$ on $[0, 1]$,
     (a) Set $s \leftarrow -\frac{1}{\lambda^*} \ln U$,

(b) If $s \leq T$, then $t_1 \leftarrow s$.

(c) Else go to last step.

- Algorithm-General routine

1. Main routine: Set $n \leftarrow n + 1$.

   (a) Update maximum intensity: set $\lambda^* \leftarrow \lambda(t_{n-1}) + \alpha$, $\lambda^(*)$ shows a jump of size $\alpha$ as an event has occurred just now, $\lambda$ is left-continuous, but this jump is not counted in $\lambda(t_{n-1})$.

   (b) New event:

      i. Generate a random uniform number $U$ on $[0, 1]$ and set $s \leftarrow s - \frac{1}{\lambda^*} \ln U$.

      ii. If $s \geq T$, then go to last step.

   (c) Rejection test: Generate $D \rightsquigarrow U_{[0,1]}$.

      i. If $D \leq \frac{\lambda(s)}{\lambda^*}$, then set $t_n \leftarrow s$,and go through the general routine again.

      ii. Else update $\lambda^* \leftarrow \lambda(s)$ and try a new date as *step*$(b)$ of general routine.

2. Output: Retrieve the simulated process $t_n$ on $[0, T]$.

The above described algorithm was implemented using R. After the data was generated it is possible to calculate the intensity of the process using one-dimensional Hawkes equation. Note that the parameter $\mu$ determines the intensity of exogenous events (roughly, how many events occur per unit of time) and that $\alpha$ and $\beta$ determine the clustering of the process and the intra-event dynamics. Supposing $\mu = 1.2, \alpha = 0.6, \beta = 0.8$ and model 100 events, we get the following simulated Hawkes' process in Fig 3.2.

From the simulated process we can observed that the self-excitability is visible and the events occur within short time from each other. Also we can see more from the simulated process that every event occurrence augments the chance of another event occurrence which leads to cluster of another events. The actual average of intensity of process is

$$\frac{N(t)}{t} \approx \frac{100}{150} = \frac{2}{3} \tag{3.11}$$

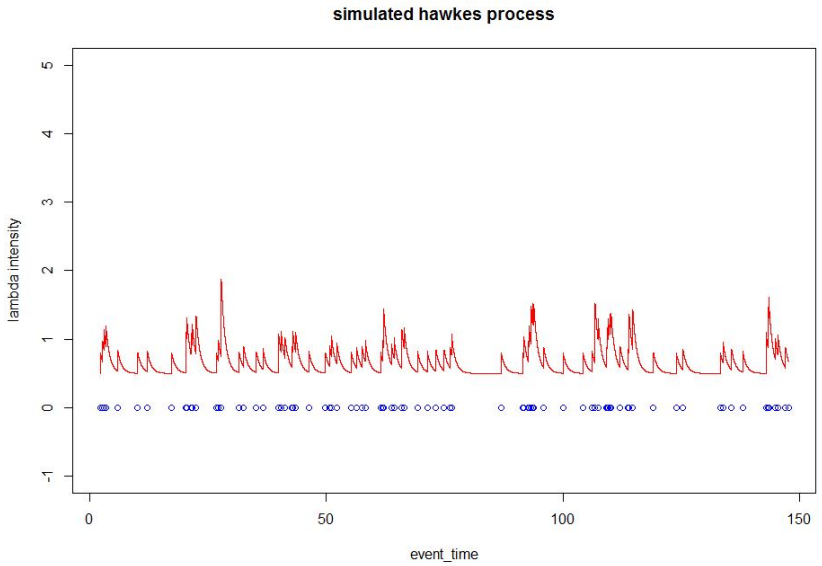



Figure 3.2: Simulated 100 events and its corresponding intensities

and the theoretical expected intensity is derived

$$\begin{aligned} E(\lambda(t)) &= E(\mu) + E(\int_{u<t} \alpha e^{-\beta(t-u)} dN_u) \\ &\Downarrow \\ E(\lambda(t)) &= \frac{\mu}{1-\frac{\alpha}{\beta}} = \frac{2}{3}. \end{aligned} \tag{3.12}$$

The result on simulation processes show it's correct. And also from the result above we can check the stationary condition for the Hawkes process is that $\frac{\alpha}{\beta} < 1$.

### 3.3.2 Compensator of Hawkes Process and Random Time Change Theorem

For a stochastic process the compensator is a deterministic function which is subtracted from the process itself to make it a local martingale. Mathematically, the integral of the intensity over the whole history of the process could be referred to as the compensator

$$\Lambda(t) = \int_0^t \lambda(s) ds. \tag{3.13}$$

Generally speaking a simple Poisson process with intensity $\lambda$, then the compensator can be written as $\Lambda(t) = \lambda t$, since the intensity $\lambda$ of a homogenous Poisson process is a constant.

For a general point process the compensator defined by the Equation (3.13), takes the point process with intensity $\lambda(s)$ to a unit-rate Poisson process. Therefore the durations as defined bellow are exponentially distributed

$$\Lambda(t_i, t_{i+1}) = \int_{t_i}^{t_{i+1}} \lambda(s)ds. \tag{3.14}$$

and can be used to test the simulated process via a quantile-quantile plot. This equation performs a random change in the time-scale of the process. The resulting process from this random time change is called the residual process. Inserting the intensity for a Hawkes process in Equation (3.14)

$$\Lambda(t_i, t_{i+1}) = \int_{t_i}^{t_{i+1}} \mu(s)ds + \int_{t_i}^{t_{i+1}} \sum_{t_k<s} \alpha exp(-\beta(s - t_k))ds. \tag{3.15}$$

and noting that the summation is over the (discrete) event times that are smaller or equal to $t_i$ to get (assuming $\mu(s) = \mu$)

$$\Lambda(t_{i+1}, t_i) = \mu(t_i - t_{i+1}) + \sum_{k=1}^{i} \int_{t_i}^{t_{i+1}} \alpha exp(-\beta(s - t_k))ds \tag{3.16}$$

which leads to

$$\Lambda(t_i, t_{i+1}) = \mu(t_{i+1} - t_i) - \sum_{k=1}^{i} \frac{\alpha}{\beta}[exp(-\beta(t_{i+1} - t_k)) - exp(-\beta(t_i - t_k))]. \tag{3.17}$$

The residual process derived from the model has been commonly used to measure the goodness-of-fit of Hawkes' model. The time change property of point process ensures that Hawkes' process after integrated is a Poisson process with unit rate. Based on this property, the durations of the whole integrated process distributed exponentially with unit rate. We follow Ogata [47], letting $\xi_i = \Lambda(t_{i-1}, t_i)$ and making use of Equation (3.14) are main requirement to assure the one-to-one transformation of the point process possibly described by the events $t_i$ into the random time changed set $\xi_i$. Then we obtain the residual process $\xi_i$ of unit rate Poisson, which is described by using the time change property. Let $U_k = 1 - exp\Lambda(t_{i-1}, t_i)$, then $U_k$ distributes as a uniform random variable in range $[0, 1)$.

As a result of the previously estimated parameters $\hat{\theta} = (\hat{\mu}, \hat{\alpha}, \hat{\beta})$, there is simple way of

assessing the goodness of fit of Hawkes' model by calculated the estimated $\hat{\Lambda}(t_i)$, from which basis we can obtain the estimated residual process $\hat{\xi}_i$. The general view on residual process that is, if the Hawkes model has a good presentation of our data then the residual process will follow a unit rate Poisson process, or equivalently the durations of the residual process will follow a unit rate exponential distribution.

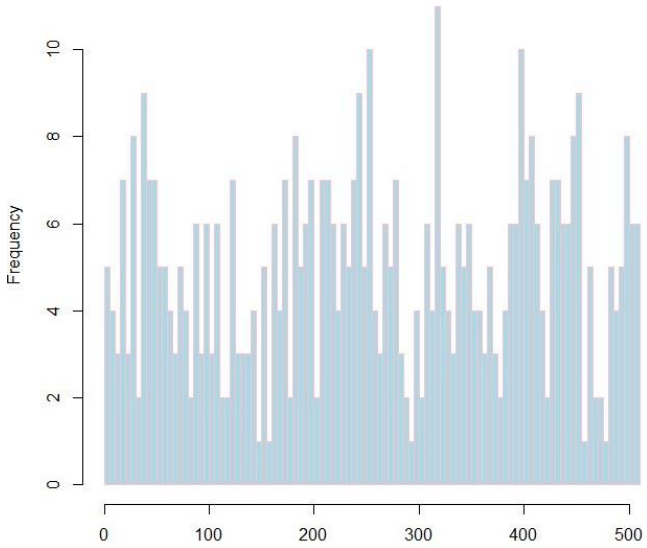



Figure 3.3: Time changed process

Applying time change theorem to the original process and splitting the time interval in to steps of 5 units, Fig 3.3 shows a result in a frequency count of 5. The above histogram confirms that the parameters estimation via $MLE$ is correct.

## 3.4 Estimation of Hawkes's Parameters

### 3.4.1 Maximum-likehood Estimation Method

Previous section describes simulation method of Hawkes' process kernel function, but when we put up the number of events, such as from 100 to 10,000, the program can be found running time is greatly increased, this is because when calculate the strength of the corresponding intensity need calling all the events in front of the moment, so using a recursive structure to optimize the calculation process. The estimation procedure of Hawkes' self-exciting process was built on by Ozaki [48]. Moreover, the asymptotic properties of the Maximum-Likelihood estimators of Hawkes's process were established by Ogata [45].

Given the occurrence observations $t_1, t_2, t_3, ..., t_n$ for an interval $[0, T]$, $T > t_n$, and based on equation (3.6) before. The log-likelihood of a Hawkes process with an intensity function is written as

$$\lambda(t|\theta) = \mu + \int_{-\infty}^{t} g(t-u|\theta)\, dN(u).$$

the log-likelihood function is formed

$$\log L(t_1, t_2, ..., t_n|\theta) = -\int_0^T \lambda(t|\theta)dt + \int_0^T \log \lambda(t|\theta)dt.$$

where $\theta = (\theta_1, \theta_2, ..., \theta_n)$. This is the familiar expression $\log(\lambda t e^{-\lambda t}) = -\lambda t + \log(\lambda t)$. The above log-likelihood is defined under the assumption that the occurrence observations are observed from time $t = 0$ to a given time $T = t_n$. But actually in most identification problems, only $t_1, t_2, t_3, ..., t_n$ are given and $T$ is not specified. so here we suppose $T = t_n$.

The log-likelihood function for Hawkes' self-exciting model is

$$\log L(t_1, t_2, ..., t_n|\theta) = -\int_0^T [\mu + \int_{-\infty}^{t} \alpha e^{-\beta(t-u)} dN(u)]dt + \int_0^T \log(\mu + \int_{-\infty}^{t} \alpha e^{-\beta t} dN(u))dt.$$

Exchanging the variables $u$ and $t$ in the integrals

$$\log L(t_1, t_2, ..., t_n|\theta) = -\mu t_n - \int_0^T [\int_u^{t_n} \alpha e^{-\beta(t-u)} dt]dN(u) + \int_0^T \log(\mu + \int_{-\infty}^{t} \alpha e^{-\beta t} dN(u))dt.$$

$$\Downarrow$$

$$\log L(t_1, t_2, ..., t_n|\theta) = -\mu t_n + \int_0^T [\frac{\alpha}{\beta}(e^{-\beta(t_n-u)} - 1)]dN(u) + \int_0^T \log(\mu + \int_{-\infty}^{t} \alpha e^{-\beta t} dN(u))dt$$

$$\Downarrow$$

$$\log L(t_1, t_2, ..., t_n|\theta) = -\mu t_n + \sum_{i=1}^{n} [\frac{\alpha}{\beta}(e^{-\beta(t_n-t_i)} - 1)] + \sum_{i=1}^{n} \log(\mu + A_i).$$

where $A_i = \alpha \sum_{t_j < t_i} e^{-\beta(t_i - t_j)}$ is a recursive structure and $t_i, (i \geq 2)$ represents the $ith$ event happened time. So we can easily understand $A(1) = 0$ and the follows

$$A_i = (A_{i-1} + \alpha)e^{-\beta(t_i - t_{i-1})}.$$

According to this property, the recursive structure can largely reduce the computational

complexity. By fitting data into the mathematical model via optimization methods we find out the optimal values of model parameters.

Here we conduct two simple estimation exercises by created 30 and 1000 different samples by using simulation method showed in previous section. It begins with a small sample of 100 events and increase the sample size by 100 observations every time a new sample is simulated, the last sample we draw has 30000 and 100000 observations. For this exercise, the parameters are using true values $\theta = (\mu = 1.2, \alpha = 0.6, \beta = 0.8)$.

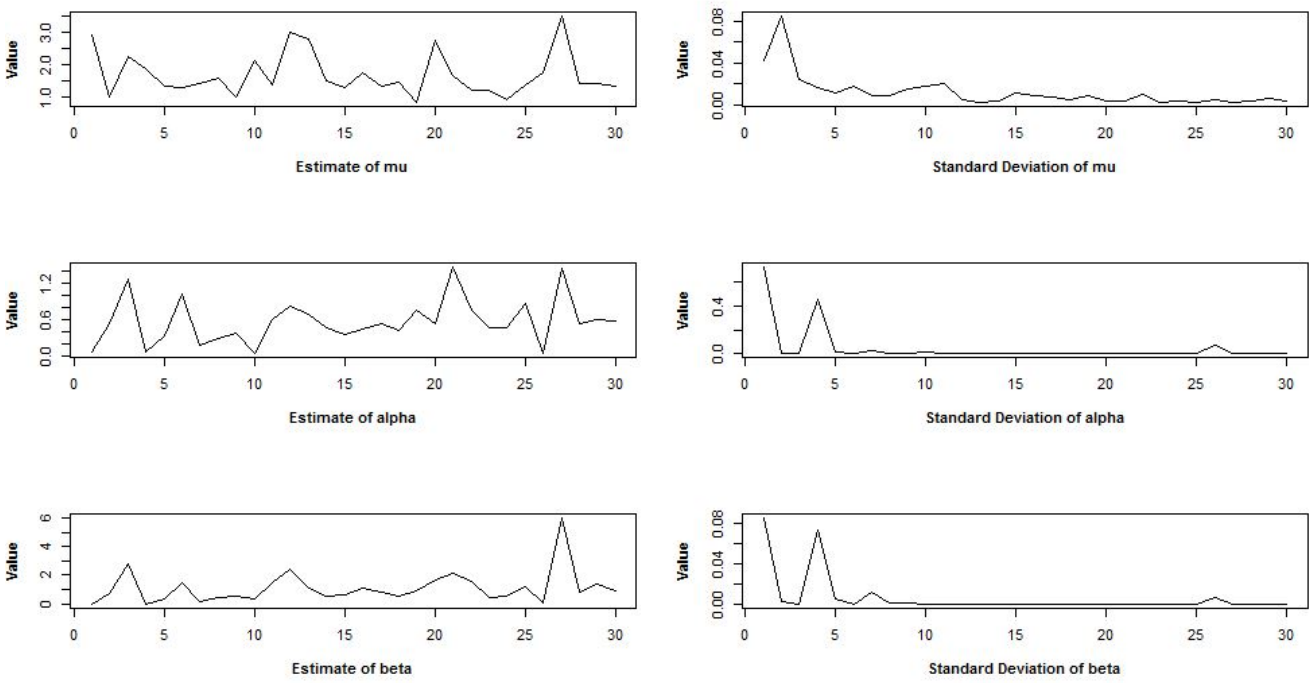



Figure 3.4: Estimation on 30 sample size $\hat{\theta} = (\hat{\mu} = 1.31(0.0110), \hat{\alpha} = 0.56(0.0115), \hat{\beta} = 0.88(0.0466))$. Numbers in parentheses are standard deviations.

We can see the estimated parameters are very close to the real parameters we posted before, which showed good performance that the maximum likelihood estimation is very effective. we use simulation method to verify the possibility of the maximum likelihood estimation of Hawkes' self-exciting point model with its response function. It is obviously to see benefit as the intensity process. Our results give a promising insight in the statistical identification of general Hawkes's process for further development.

Another two significant quantities in the concept of a Hawkes process are also of inter-

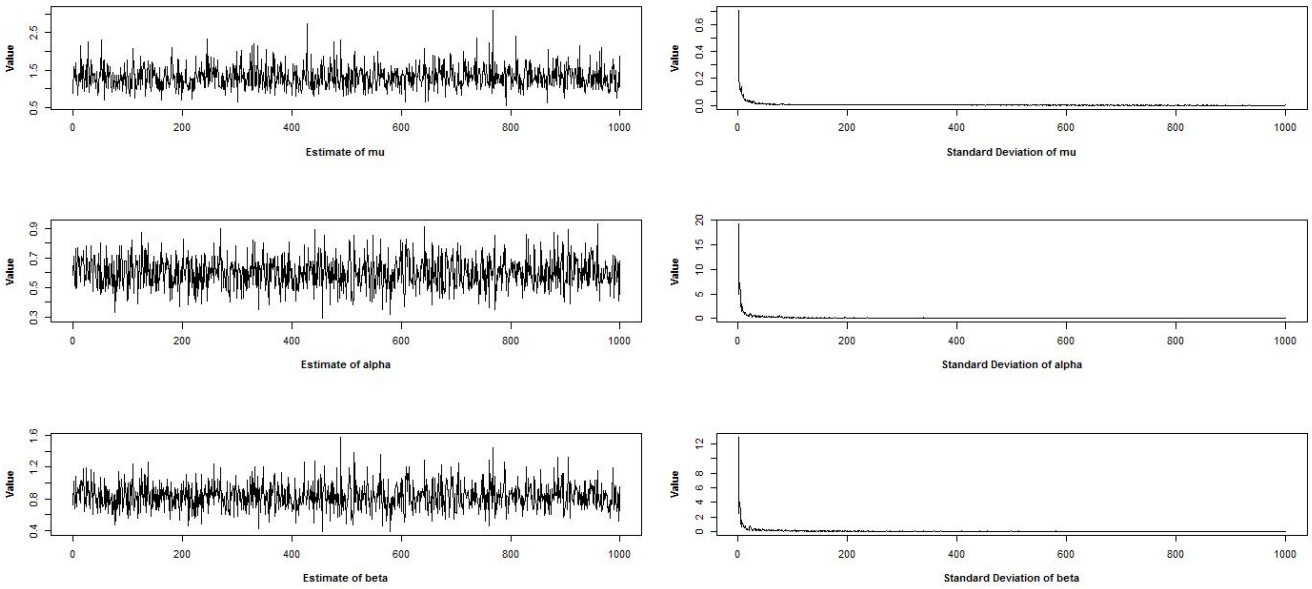



Figure 3.5: Estimation on 1000 sample size $\hat{\theta} = (\hat{\mu} = 1.06(0.9592), \hat{\alpha} = 0.68(0.1140), \hat{\beta} = 0.95(0.5984))$. Numbers in parentheses are standard deviations

est: one is the clustering size $Z$

$$Z = \frac{1}{1 - \frac{\alpha}{\beta}}.$$

and another is Branching Ratio $n$ of the process which is the average number of daughter events per mother event, it represents the average number of offspring produced per event

$$n = \int_0^\infty \alpha e^{-\beta t} dt = \frac{\alpha}{\beta}, \tag{3.18}$$

where both parts of expressions are effective for Hawkes' model exponential kernel. The most interested expression above is the branching ratio. Whenever the base rate of intensity $\mu$ is a constant and the process is in the sub-critical ($n < 1$) or in the critical ($n = 1$) regime the branching ratio can be used as a measure of the proportion of events that are generated inside the model to all events (i.e. events endogenously generated). That is to say, branching ratio can be used to evaluate how much of trading activities are caused by process. When $n > 1$, the whole process $N(t)$ is non-stationary and diverges to infinity for a finite time. When $n = 1$, $N(t)$ is called the critical Hawkes's process. Brémaud and Massoulié [10] discuss its stability , This thesis is to study the events endogenously of the A-H stocks under $n < 1$ situation.

### 3.4.2 Consistent and Asymptotic Properties of the Maximum-Likelihood Estimator

We know from Ogata [45] that a point process with its intensity has always asymptotically stationary and he also established the asymptotic properties of the maximum-likelihood estimator of Hawkes' process. Here we first recall the stationarity condition for one dimensional Hawkes' model is showed before Eqn. (3.8). In my thesis the maximum-likelihood estimator of a stationary one dimensional Hawkes' process $\hat{\theta} = (\hat{\mu}, \hat{\alpha}, \hat{\beta})$ properties are

- Asymptotically normal,

$$\sqrt{T}(\hat{\theta}^T - \theta) \to \mathcal{N}(0, (E[\frac{1}{\lambda}\frac{\partial\lambda}{\partial\theta_i}\frac{\partial\lambda}{\partial\theta_j}])_{i,j}). \tag{3.19}$$

- $\hat{\theta}$ asymptotically and efficiently reaches the lower bound of variance.

- Consistency, $\hat{\theta}$ converges in probability to the true values $\hat{\theta} = (\hat{\mu}, \hat{\alpha}, \hat{\beta})$ as $T \to \infty$:

$$\forall\varepsilon > 0, \lim_{T\to\infty} P[|\hat{\theta}^T - \theta| > \varepsilon] = 0. \tag{3.20}$$

Here we conduct two simple exercises by created 30 and 100 different samples to confirm the maximum-likelihood estimator's consistent property. It begins with a small sample of 100 events and increase the sample size by 100 observations every time a new sample is simulated, the last sample we draw has 30000 and 90000 observations. And also I plot the standard errors of each coefficient next to the parameter estimates. Standard errors were constructed by calculating the inverse of the Hessian and multiplying it by minus 1. For this exercise, the parameters use true values $\theta = (\mu = 0.1, \alpha = 1.0, \beta = 2.0)$.

Figure 3.6 and 3.7, show the value of the estimated parameters on the vertical axis and the size of the sample used in the estimation on the horizontal axis, the horizontal lines around the parameters estimates represent the true value of the parameters, but in different simple size. From those two figures result, we can see the more sample big enough the better result is. Moreover, Figure 3.6 and 3.7 obviously realized the asymptotic and consistent properties of the estimator $\hat{\theta} = (\hat{\mu}, \hat{\alpha}, \hat{\beta})$. As can also be seen from the two Table 3.1 and

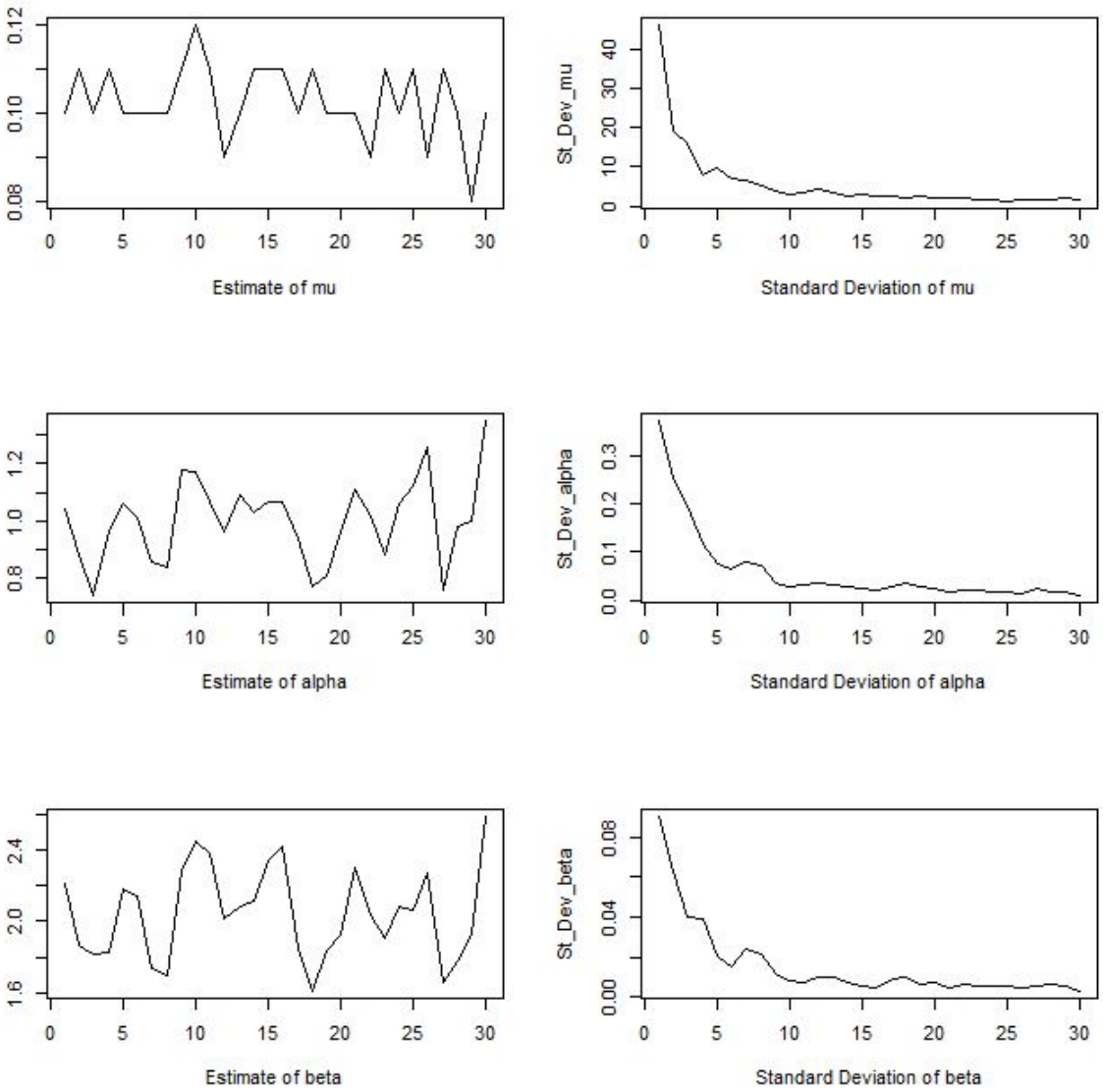



Figure 3.6: Consistency of Estimate with true values $\theta = (\mu = 0.1, \alpha = 1.0, \beta = 2.0)$ based on sample size 30

Table 3.2 shows that maximum likelihood estimator $\hat{\theta} = (\hat{\mu}, \hat{\alpha}, \hat{\beta})$ performed better with a larger sample size.

Descriptive statistics of results obtained are shown in Table 3.3, variances and standard errors are calculated in two different ways: first we calculate the sample variance using the 100 different estimates for each sample we generated (denoted by sample variance in the table); second we calculate the variance for each estimate by calculating the inverse of the Hessian and multiplying it by minus 1 (denoted by estimated variance in the table below). The obtained values, which estimate of the variance-covariance matrix of the parameter estimates, are then averaged over the 100 different variance-covariance matrices. The small

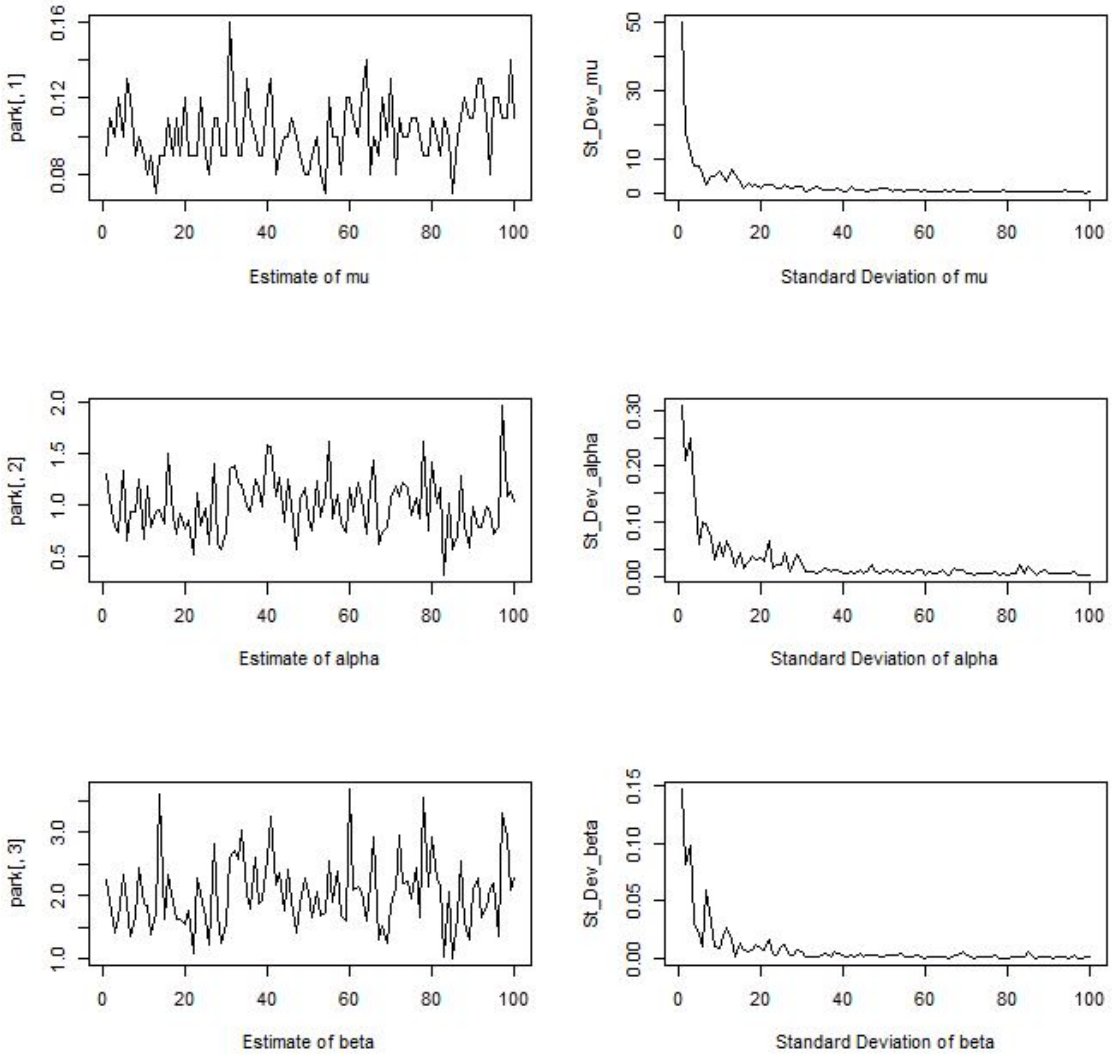



Figure 3.7: Consistency of Estimate with true values $\theta = (\mu = 0.1, \alpha = 1.0, \beta = 2.0)$ based on sample size 100

difference in the values of the variance calculated using these different approaches is due to the small size properties of the sample. The estimations are carried out with true parameters $\theta = (\mu = 0.1, \alpha = 1.0, \beta = 2.0)$.

### 3.4.3 Goodness of Fit

A common way to test how well the model fits the data is by evaluating the residuals, the quality of fit is assessed on the time-deformed defined by Equation (3.14).If the model describes the data correctly, then the residual process should be independent and should have inter-event times (the difference between two residual event time-stamps) expo-

| $T$ | $\mu$ | $\alpha$ | $\beta$ |
|---|---|---|---|
| True value | 0.1 | 1.0 | 2.0 |
| 100 | 0.12(0.7448) | 0.79(0.0233) | 1.43(0.0097) |
| 200 | 0.09(1.3228) | 1.17(0.0113) | 2.16(0.0043) |
| 300 | 0.10(1.6023) | 0.81(0.0180) | 1.89(0.0038) |
| 400 | 0.10(1.6426) | 0.99(0.0128) | 2.20(0.0029) |
| 500 | 0.10(1.4551) | 1.11(0.0113) | 2.27(0.0031) |
| 600 | 0.10(1.2811) | 1.00(0.0142) | 1.97(0.0045) |
| 700 | 0.09(1.4238) | 1.03(0.0156) | 1.81(0.0061) |
| 800 | 0.09(1.8664) | 0.93(0.0145) | 2.06(0.0035) |
| 900 | 0.09(1.4985) | 0.95(0.0165) | 1.81(0.0057) |
| 1000 | 0.10(1.2900) | 1.04(0.0119) | 2.26(0.0028) |

Table 3.1: Maximum likelihood estimation of a one-dimensional Hawkes' process on simulated data. Each estimation is the average result computed on 30 samples of length [0,T]. Standard deviations are given in parentheses.

| $T$ | $\mu$ | $\alpha$ | $\beta$ |
|---|---|---|---|
| True value | 0.1 | 1.0 | 2.0 |
| 100 | 0.12(0.2351) | 0.98(0.0049) | 1.71(0.0020) |
| 200 | 0.09(0.4416) | 1.23(0.0033) | 2.14(0.0013) |
| 300 | 0.10(0.3434) | 1.13(0.0037) | 2.08(0.0014) |
| 400 | 0.09(0.5114) | 1.26(0.0029) | 2.44(0.0010) |
| 500 | 0.09(0.4660) | 0.87(0.0053) | 1.80(0.0015) |
| 600 | 0.10(0.3989) | 0.88(0.0055) | 1.73(0.0016) |
| 700 | 0.09(0.4502) | 0.91(0.0050) | 1.83(0.0015) |
| 800 | 0.09(0.4948) | 0.88(0.0054) | 1.77(0.0016) |
| 900 | 0.10(0.4185) | 1.20(0.0033) | 2.22(0.0012) |

Table 3.2: Maximum likelihood estimation of a one-dimensional Hawkes' process on simulated data. Each estimation is the average result computed on 100 samples of length [0,T]. Standard deviations are given in parentheses.

| Pars | $\mu$ | $\alpha$ | $\beta$ |
|---|---|---|---|
| Average | 0.098 | 0.982 | 1.986 |
| Minimum | 0.090 | 0.79 | 1.43 |
| Maximum | 0.120 | 1.17 | 2.27 |
| Sample Standard Errors | 1.4127 | 0.0151 | 0.0046 |
| Estimated Standard Errors | 1.0302 | 0.0133 | 0.0053 |

Table 3.3: Statistics of Estimated Parameters Result

nentially distributed with unit rate. The maximum-likelihood estimation, by construction, tends to maximize the exponential nature of the durations, but not their independence. This explains why QQ-plots of the resulting durations are visually very satisfying as long as the kernel contains more than one exponential. In order to validate the simulated process we calculate the compensator using simulated data that follows a Hawkes process with expo-

nential response function. The durations given by Equation (3.14) are then exponentially distributed. Figure 3.8 and 3.9 shows the behavior of the integrated simulated process with respect to a unit-rate exponentially distributed process. The fact that both, the simulated process and the theoretical process lies on the same 45° line indicates that both processes have the same distribution. In the plot, the empirical and theoretical axis represent the values for the durations of the simulated and theoretical processes respectively.

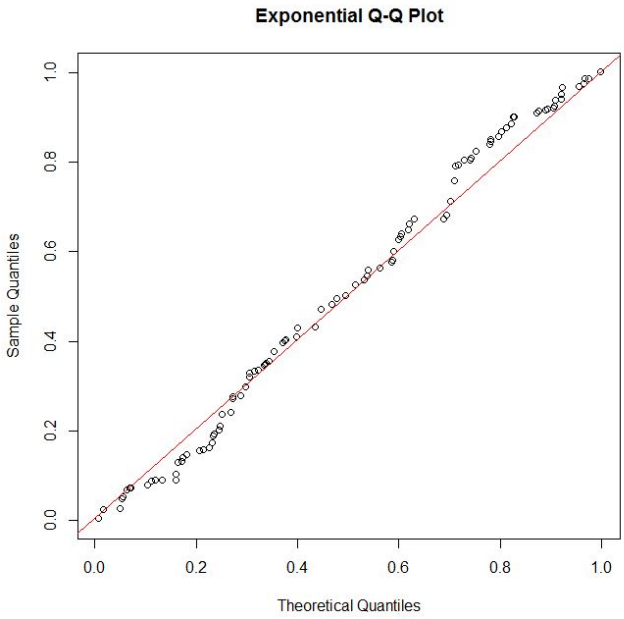



Figure 3.8: QQ-plots of integrated intensities based on 30 sample size of simulated data of Hawkes' process

Now that we know the Hawkes model explains clustering of arrivals well. The next steps would be to apply Hawkes's model to A-H shares trading and then will consider buy and sell arrivals individually and find a way on a given fitted Hawkes' model. These bid-ask intensity predictions can then form a part of a market-making or directional strategy.

## 3.5 Prediction of Univariate Hawkes' Process

It is well known that Hawkes' model can predict changes in a liquid financial market but the quality of the algorithm depends how the parameters have been calibrated. As shown before an the Hawkes process can specify the distinction between the arrival of an event of external origin (exogenous) such as some market news and the internal self-exciting event

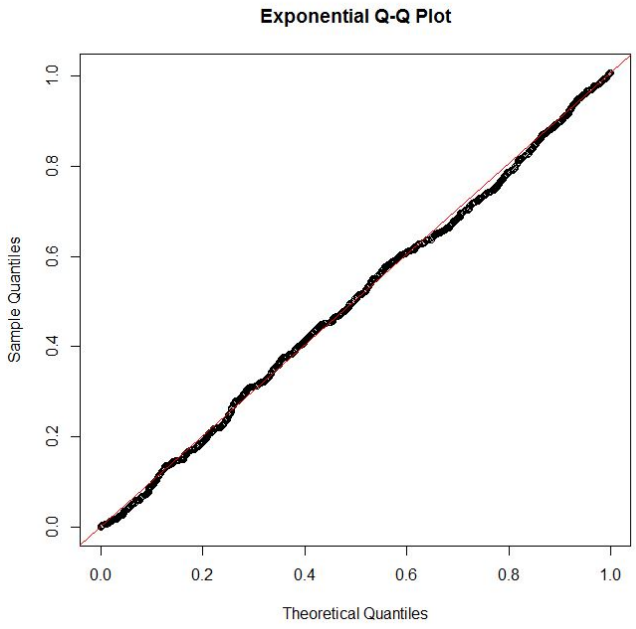



Figure 3.9: QQ-plots of integrated intensities based on on 1000 sample size of simulated data of Hawkes' process

(endogenous) such as the influence on trades by past trades. Being able to do this gives the model its stochastic nature and enables the model simulations to improve the prediction results. The algorithms which forecast the intensity is not dependant on time unlike the real markets however an exogenous intensity may be subject to time so it is essential to de-trend these events in frame-data first before the algorithm is applied.

Understanding the dynamics of the underlying markets is necessary in order to define a forecast model. Traditional economic theories are based on the Efficient Market Hypothesis (EMH), which assumes that markets are fully absorbing all the information instantaneously and that the information must be reflected somehow in the market price. Furthermore large changes in the market occur in response to "big news" that influence the market prices from the outside. However analysis of this has proven the approach wrong. Actually the number of short-term trades are not only based on exogenous events but are influenced from information from the market itself and its participants and past trades. The Hawkes' process is a good model to describe this system in fact, Filimonov and Sornette [28] have demonstrated that the model is a more valid tool to predict parameters than market models based on the EMH. Estimation of the parameters in the Hawkes model is based on a timeframe called The Sample Timeframe. The greater the sample timeframe the better quality algorithm and the better results, however with a higher the complexity of the computations required in the

calculation.

A good sample timeframe must lie in trading hours which includes substantial trading hours as this is when it captures the dynamics of substantial market changes. It must be noted that the sample timeframe cannot be made, for example, at the beginning of the day as it would not capture sufficient samples to reflect the market dynamics. Neither can we use data from the end of the trading day as events could have offspring influences on new events which occur at the next opening and would be affected by events in the period of over 12 hours when the market was closed. However a small number of samples could be taken in the morning however those predictions tend to be less reliable. The sample timeframe is not given any future information in the experimental setup so there is no looking ahead influences.

In the experiment we are finding the difference between the simulation of the average relative deviation between algorithmically predicted events and actual predicted number of events in the timeframe. When the relative deviation has sufficiently reduced and stops fluctuating then the number of simulations needed has been reached. Then a synthetic time series is built which involves setting up a Hawkes process kernel time series with a set of parameters that we choose. The details of such a setup is two-fold.

- The actual parameters of the Hawkes process ($\mu$, $n$ and exponential kernel) are known;
- A prediction algorithm governing the dynamics of the Hawkes process is assumed but other influences can play a role such as real life trading data. In a synthetic time series there is the potential to estimate a simulation of a purely Hawkes' reality.

And the test setup is as follows:

1. A kernel and its corresponding parameters are chosen (exponential kernel);
2. The set length of synthetic time series is defined as $[0, t_{tend}]$;
3. Define a starting time $0 < t_{tend}$ from when wed like to predict and the period length for time frame which ends at $t_{end}$ and will be the input to the algorithm used to predict;
4. A synthetic time series with the set time window $[0, t_{end}]$ its length simulated applying the intensity $\lambda$ of the Hawkes' Poisson model for every event in the sample time window;

5. Simulation the time frame starting at time $t_0$ using the kernel and parameters we used to create the synthetic simulations. We measure the deviation between the number of events our algorithm predicted in the timeframe $[0, t_{end}]$ and the actual number of events in the predicted (synthetic) reality. This step is repeated $N_{sim}$ times and the results are averaged.

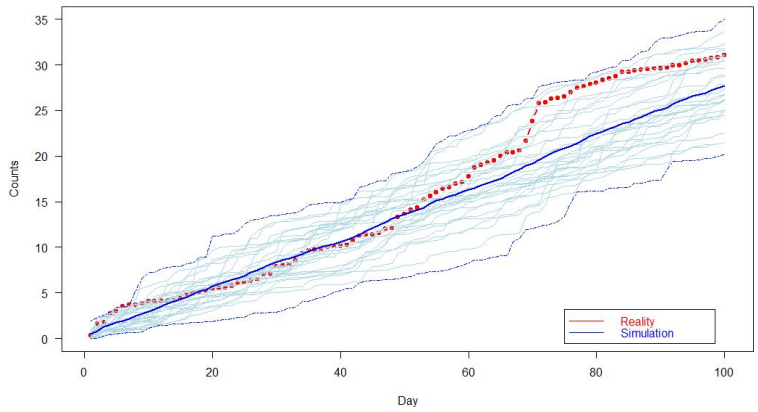



Figure 3.10: A visualisation of the synthetic 100 times simulation.The total averaged number of simulated events (blue) between 0 and $t_{\text{end}}$ are compared with the realization of simulation (red).

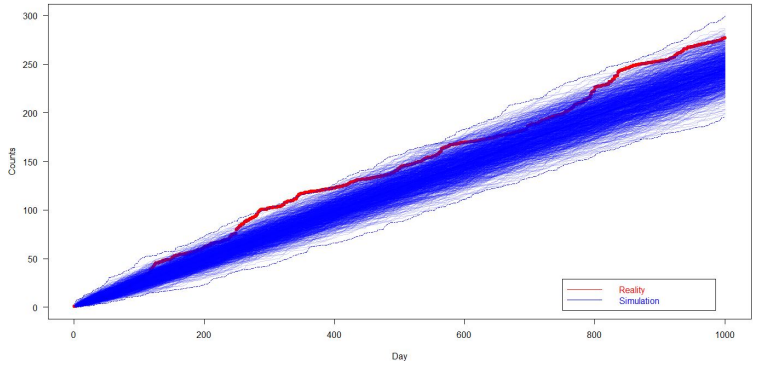



Figure 3.11: A visualisation of the synthetic 1000 times simulation.The total averaged number of simulated events (blue) between 0 and $t_{end}$ are compared with the realization of simulation (red).

The experiment with the synthetic reality is performed with multiple parameter values of the Hawkes intensity function. Even though the deviation across different parameter value combination was variable the dynamics of the fluctuations was similar thought-out all parameter realities. As shown in the figure this qualitative trend is reflected. Over a number of different simulations the expectation value and variance of relative deviations averaged. The figure demonstrates fluctuations when too few simulations were averaged after 100 simulations this fluctuation diminishes and so we chose the average after close to

1000 simulations. Therefore these results is the average over 1000 simulations. This is the point when a compromise is reached between a sufficiently large sample size to reduce the bias of being a stochastic process and a relatively short computational time.

With more simulations using longer samples, the results can improve and get closer to the actual value. Furthermore, the fact that Hawkes' model can explain clustering of the order occurrence, so in Chapter 5 we will use Huaneng Power International Inc. and Industrial and Commercial Bank of China Ltd. two company from Mainland China and Hong Kong stock market to fit both one and two-dimensional Hawkes' processes.

## 3.6 Comparing ACD Model and Hawkes' model

A disadvantage of the ACD model compared to the Hawkes model is that the latter is delineated in wall-clock time, the previous is delineated in event arrived time. The second disadvantage is Hawkes' model provides a straightforward description of the conditional intensity, but this can not be true for the ACD model. This advantage of the Hawkes model makes it easier to treat variable estimation in a whole process. Above all, we now turn to see how does Qusai-Maximum Likelihood Estimation work on ACD model.

### 3.6.1 Qusai-Maximum Likelihood Estimation on ACD Model

The more common estimation method for the ACD model is the quasi-maximum likelihood estimation (Q)MLE whose maximum likelihood function and its first derivative can be expressed as Eqn. (3.21)

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{n} l_i(\theta) = \sum_{i=1}^{n} \log(\frac{1}{\psi_i} f_i(\frac{x_i}{\psi_i})) \\ s(\theta) &= \sum_{i=1}^{n} -(1 + \varepsilon_i \frac{f_i'(\varepsilon_i)}{f_i(\varepsilon_i)}) \frac{d \log(\psi_i)}{d\theta}. \end{aligned} \tag{3.21}$$

The maximum likelihood function of the ACD model under the standard Exp distribution is

$$l(\theta) = -\sum_{i=1}^{n} (\ln \psi_i + \frac{x_i}{\psi_i}), \theta = (\omega, \alpha, \beta). \tag{3.22}$$

On the premise that the conditional mean equation is correctly set, this paper discusses the asymptotic behavior of the quasi-maximum likelihood estimation method (Q)MLE of the ACD(1,1) model in the simple form. Let's take a look at ACD(1,1):

$$\begin{aligned} x_i &= \psi_i \varepsilon_i \\ \psi_i &= \omega + \alpha x_{i-1} + \beta \psi_{i-1}. \end{aligned} \tag{3.23}$$

Assume that the ACD(1,1) model satisfies the following regular condition,

- Hypothesis 1: $E(x_i) \equiv \psi_{0,i} = \omega_0 + \alpha_0 x_{i-1} + \beta_0 \psi_{0,i-1}$.
- Hypothesis 2: The random disturbance term $\varepsilon_i = \frac{x_i}{\psi_i}$, strictly traversed, non-degenerate,and $E(\varepsilon^2 | F_{i-1}) < \infty$, $E(\ln \beta_0 + \alpha_0 \varepsilon_i | F_{i-1}) < 0$.
- Hypothesis 3: $l(\theta) = -\sum_{i=1}^{n} (\ln \psi_i + \frac{x_i}{\psi_i}), \psi_i = \omega + \alpha x_{i-1} + \beta \psi_{i-1}$.
- Hypothesis 4: $\theta_0$ is the only recognizable maximum value of $n^{-1} \sum_{i=1}^{n} E(l_i(\theta) - l_i(\theta_0))$.

Based on these hypotheses, we have $\sqrt{n}(\hat{\theta} - \theta_0) \to N(0, A(\theta_0)^{-1} B(\theta_0) A(\theta_0)^{-1})$

where

$$A(\theta_0) = \lim_{n \to \infty} n^{-1} E(\frac{\partial^2 l}{\partial \theta \partial' \theta}) \qquad B(\theta_0) = \lim_{n \to \infty} n^{-1} E(\frac{\partial l}{\partial \theta} \frac{\partial l}{\partial \theta'}).$$

It is worth noting that even if the random perturbation term of the duration model does not follow an the exponential distribution, the estimation using the (Q)maximum likelihood function can still be obtained. In particular, if the information matrix $A(\theta_0) = B(\theta_0)$ is established, its variance is estimated under (Q)MLE.

Now we estimate the ACD(1,1) model of random disturbance terms under exponential distribution by Monte Carlo simulation. The simulation process as follows:

- First, using the exponential distribution, generate data samples with sample sizes N of 100, 200, 300, 1000, 2000, 3000, 10000, 20000, 30000,respectively.
- Second, the above data samples are estimated using (Q) MLE, where the maximum likelihood objective function of MLE uses a standardized distribution function.

- Third, repeat the above the First and Second step 1000 times to obtain samples of the parameters which need to be estimated.

Simulates a sample from a specified ACD model and error term distribution dist. The error terms can also be sampled from residuals. The possibility of including a diurnal seasonal component in the simulated sample is included.

In the analysis of the results, we present descriptive statistical indicators of the parameters to be estimated, as shown in the table below Table 3.4.

| $T$ | $\omega$ | $\alpha$ | $\beta$ | $AIC$ | $BIC$ |
|---|---|---|---|---|---|
| True value | 0.2 | 0.15 | 0.80 | | |
| N=100 | 0.246(0.2653) | 0.170(0.0923) | 0.755(0.1319) | 427.9894 | 435.8049 |
| N=200 | 0.207(0.1372) | 0.112(0.0594) | 0.764(0.1047) | 610.4480 | 620.3429 |
| N=300 | 0.232(0.1990) | 0.163(0.0640) | 0.740(0.1370) | 1096.0093 | 1107.1206 |
| N=1000 | 0.098(0.0304) | 0.133(0.0203) | 0.822(0.0263) | 2950.8610 | 2965.5840 |
| N=2000 | 0.097(0.0242) | 0.148(0.0182) | 0.803(0.0252) | 6343.6155 | 6360.4182 |
| N=3000 | 0.105(0.0188) | 0.167(0.0156) | 0.779(0.0213) | 9275.2799 | 9293.2991 |
| N=10000 | 0.101(0.0106) | 0.146(0.0079) | 0.804(0.0109) | 31974.7227 | 31996.3537 |
| N=20000 | 0.094(0.0066) | 0.149(0.0053) | 0.803(0.0070) | 63001.4082 | 63025.1187 |
| N=30000 | 0.108(0.0063) | 0.150(0.0046) | 0.795(0.0065) | 95387.0593 | 95411.9861 |

Table 3.4: ACD(1,1) model estimation by (Quasi) Maximum Likelihood

Comparing different sample sizes, the sample mean of the parameter estimate $\alpha$ and $\beta$ of the ACD model gradually approaches the true value of the sample with a large sample size at N=100, 200,..., 30000 sample sizes respectively. And its sample variance is also decreasing, indirectly indicating the consistency of the estimation of various method parameters. It is worth mentioning that b is the model autoregressive coefficient, which means that the model is more accurate in describing the autocorrelation. Since it is difficult to correctly set the true distribution function of the duration model in practical applications, most of them are estimated using the exponential distribution-based (Q)MLE method.

Then we do a QQ-plot of the residuals and the theoretical quantiles implied by the model estimates. Figure 3.12 shows the QQ-plot of the standardized residuals versus a exponential distribution with shape parameter 0.195 and scale parameter 1. The quantiles of the exponential distribution are generated using a random sample of $N = 30000$ observations. A straight line is imposed on the plot to aid interpretation. From the plot Figure 3.13, except for a few large residuals, the assumption of a exponential distribution seems reasonable. In

this particular example, we chose the exponential distribution for its simplicity. Finally, for the ACD(1,1) model, the estimated shape parameter $\alpha$ is less than one, indicating that the hazard function of the adjusted durations is monotonously decreasing. This seems reasonable for the adjusted durations of the heavily traded in stock.

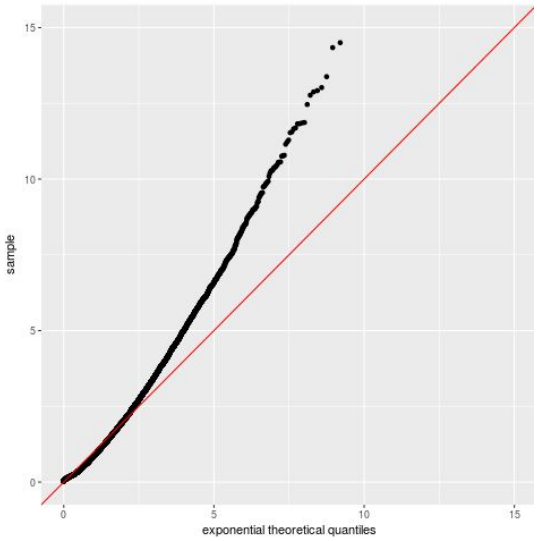



Figure 3.12: Quantile-Quantile plot for the ACD(1,1) simulation process.

Moreover, QML robust correlations shows in Table 3.5:

| | $\omega$ | $\alpha$ | $\beta$ |
|---|---|---|---|
| $\omega$ | 1.000 | 0.442 | -0.762 |
| $\alpha$ | 0.442 | 1.000 | -0.902 |
| $\beta$ | -0.762 | -0.902 | 1.000 |

Table 3.5: QML robust correlations

### 3.6.2 Differences between ACD Model and Hawkes' model

The ACD model and the Hawkes model both define the stochastic arrival time of events, however, these models have some variations in the function of their numerical analysis and diagnostic on how endogenous and exogenous events are analyzed. The benchmark, which called the "Efficient Market Hypothesis"(EMH), argues that in an ideal securities market, asset prices reflect all available information completely, quickly, and accurately. At this point, the dynamic process of price will be fully expressed as for the reflection of external information, there is no possibility of an internal self-feedback mechanism. However, in

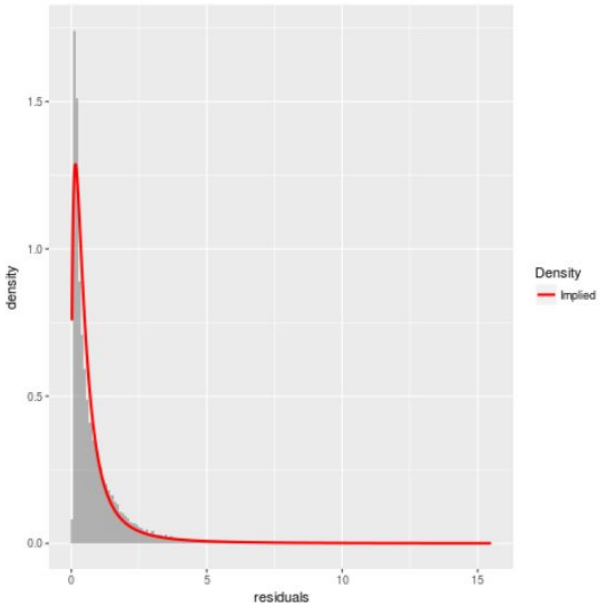



Figure 3.13: Plots a density histogram of the residuals and superimposes the density implied by ACD model estimates.

practice, the efficient market hypothesis is often difficult to establish. By studying financial high-frequency data, it found that changes in asset prices are not always linked to certain information. This shows that the price dynamic process has its own inherent self-feedback mechanism, namely endogeneity. Filimonov and Sornette [28] believe that the endogeneity of the price dynamic process is the embodiment of the "reflexivity" theory. The theory holds that human cognitive activities are incomplete, and investors' expectations and decisions exist deviations. The accumulation of these deviations caused the market to deviate from equilibrium, resulting in the market is invalid. From this point of view, comparing the exogenous and endogenous properties of securities have practical significance, and it can be used as a reference for judging whether the market is effective.

We thus wish to demonstrate that when comparing a Hawkes model to the ACD model, the Hawkes model performs better. ACD is closely mapped with the numerical side and the Hawkes model depends on the "branching ratio" to illuminate the characteristic. Because the Hawkes model has the branching ratio which can better model the behavior, and it considers how fragile and susceptible a market is to shocks. The reason highlighted is due to the so-called "branching ration" which directly measures the endogeneity, even could interpret these events and their affect as a whole.

In the Hawkes model, events $t_i$ are either an immigrant or descendant event. The rate

at which immigration occurs is determined by the background intensity $\mu(t)$ an exogenous random process. A zeroth-order event, which can be called the mother event, triggers one or more first-order events, which are called daughter events. Each of these trigger second-order events called grand-daughter of the initial mother who can generate descendants in clusters, which we called descendent or aftershock events. These are known as first-second or higher-order event clusters and occurs due to the feedback effect of the self-exciting point process model, in the class of mutually exciting nonlinear point process. These immigrant and aftershock events are a sub-critical regime in the case of a constant background intensity and its branching ration is exactly equal to that fraction of the average number of events generated as an affect.

In the following example if we count separately both the exogenous and endogenous events, then the rate of exogenous immigrants (zeroth-order events) is equal to the background activity rate: $C_{exo} = \mu$ then we find that each immigrant each independently triggers on average to $n$ daughter and first-order events as $c_1 = \mu n$ and in turn second-order events as $c_2 = nc_1 = \mu n^2$ and continuing to an infinitum. By summing over all generation we will obtain the rate of all endogenous descendants

$$C_{endo} = \sum_{i=1}^{\infty} r_i = \mu \sum_{i=1}^{\infty} n^i = \frac{\mu n}{1-n}. \tag{3.24}$$

Where $n < 1$, the global rate is the sum of the rates of descendants and immigrants which is equal to

$$C = C_{exo} + C_{endo} = \mu + \frac{\mu n}{1-n} = \frac{\mu}{1-n}. \tag{3.25}$$

The fraction of descendant (endogenously driven events) in the whole system is equal to the branching ratio

$$\frac{C_{endo}}{C} = n. \tag{3.26}$$

So $n$ on the data is calibrated to provide a direct quantities estimated degree of the endogeneity of the system.

Because the ACD model shares lots of similarities with the Hawkes model and their point processes exhibit similar degrees of clustering. We know the duration of ACD model defined before as $x_i = t_i - t_{i-1}$ and the expectation of the conditional duration as $\psi_i =$

$E(x_i|x_{i-1}, ..., x_1)$. Recalled ACD(p,q) following:

$$\begin{aligned} x_i &= \psi_i \varepsilon_i \\ \psi_i &= \omega + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j}. \end{aligned} \tag{3.27}$$

so the conditional intensity of model is given by

$$\lambda(t|x_{N(t)}, ..., x_1) = \lambda_0(\frac{t - t_{N(t)}}{\psi_{N(t)+1}})\frac{1}{\psi_{N(t)+1}}. \tag{3.28}$$

This equation can be continuously simplified in the case that under the durations, which are conditional exponentially distributed,

$$\lambda(t|x_{N(t)}, ..., x_1) = \frac{1}{\psi_{N(t)+1}}. \tag{3.29}$$

Based on this, the expected conditional duration of ACD(p,q) can be rewritten as,

$$E(x_i) = \frac{\omega}{1 - \sum(\alpha_j + \beta_j)}. \tag{3.30}$$

As we compared, the expected conditional duration of ACD model is quite similar to the expected intensity for Hawkes' process in Eqn. (3.31),

$$E(\lambda(t)) = \frac{\mu}{1 - \frac{\alpha}{\beta}}. \tag{3.31}$$

In particular, the branching ratio for ACD(p,q) defined by combining parameters,

$$\eta = \sum_{j=1}^{p} \alpha_j + \sum_{k=1}^{q} \beta_k \tag{3.32}$$

for ACD(1,1) reduces to:

$$\eta = \alpha + \beta \tag{3.33}$$

where $\eta$ plays a similar role to the branching ratio $n$ of Hawkes' process with an exponential kernel. Some important differences however between the combined ACD model and the Hawkes model are that the background rate $\mu$ is estimated by the Hawkes model, whereas

the ACD model is controlled by a combined parameter $\eta$ with indirect consequence also relating to the duration. In contrast to the background rate $\mu$ completely describes the exogenous event's impact in the Hawkes model. Added to that, the parameter $\omega$ of the ACD model is not the single factor affected by exogenous events and there is no tight decoupling between the driver of the event $\omega$ endogenous level $\eta$ occurrence as in the parameters $\alpha$ and $\beta$ of the Hawkes model, which happens to include the self-exciting factor thus giving it an important difference.

So by focusing on the element of endogeneity of a ACD process, we consider the introduction of the composite parameter $\eta$ and its association with parameters $\alpha$ and $\beta$ to control the expected arrival of an event. Given this consequence, $\eta$ has an indirect effect of the complete description of the exogenous impact to the system, where as in the Hawkes model the rate at which the background rate $\mu(t)$ does completely describes the exogenous impact directly. The parameter $\omega$ of the ACD model does not the only factor embodying the exogenous activity ,and there is no strict decoupling between the exogenous driver $\omega$ and endogenous level $\eta$, just as occurs for the parameters $\mu$ and $n$ of the Hawkes model. Filimonov and Wheatley [29] summarized that, in contrast to the Hawkes model, the ACD in its classical form does not provide a clean distinction between exogenous and endogenous activities.

# Chapter 4

# Two-Dimensional Hawkes' Self-exciting Processes

As the bid-ask spread widely increased following the financial crisis (2008-2009) during the time period of illiquidity in the financial market, it is important to question how bid-ask spread could affect liquidity and price fluctuations. It can be observed out that whenever the liquidity evaporates, all market participants then find it more difficult to complete transactions. Therefore, the bid-ask influence is closely related to the liquidity of stock market, and this helps to understand how transactional liquidity behaved. We have already modelled one dimensional self-exciting Hawkes' process in last chapter and it performs well. Based on this, we develop a two-dimensional Hawkes model which includes both cross-exciting and self-exciting parts as the influence of bid-ask market orders on the subsequent orders. We hope we can discover some principles, and it will help traders to better understand the stock market liquidity by separated bid side and ask side expected in the model.

## 4.1 Model Framework

We define a basic Hawkes model as $(t_i^A)_{i\geq 1}$ for the point process of the trades to occur at the ask side and $(t_i^B)_{i\geq 1}$ the point process which might occur at the bid side. Let $N_A$ and $N_B$ signify the relevant counting process. It is assumed that both of these two processes form a two-dimensional Hawkes process where the underlying intensity process is denoted

by $\lambda_A(t)$ and $\lambda_B(t)$ for defined parameters $(\alpha_{ij}, \beta_i)_{(i,j)\in\{A,B\}}$ as shown below

$$\begin{cases} \lambda_A(t) = \mu_A + \int_0^t \upsilon_{AA}(t-s)dN_A(s) + \int_0^t \upsilon_{AB}(t-s)dN_B(s) \\ \lambda_B(t) = \mu_B + \int_0^t \upsilon_{BA}(t-s)dN_A(s) + \int_0^t \upsilon_{BB}(t-s)dN_B(s). \end{cases} \tag{4.1}$$

It usually needs a further parameterization, so we consider

$$\upsilon_{ij}(s) = \alpha_{ij}e^{-\beta_i s}.$$

Based on this, the standard bivariate Hawkes model with $K = 1$ is

$$\begin{cases} \lambda_A(t) = \mu_A + \sum_{t_i<t} \alpha_{AA}e^{-\beta_A(t-t_i)} + \sum_{t_j<t} \alpha_{AB}e^{-\beta_A(t-t_j)} \\ \lambda_B(t) = \mu_B + \sum_{t_i<t} \alpha_{BA}e^{-\beta_B(t-t_i)} + \sum_{t_j<t} \alpha_{BB}e^{-\beta_B(t-t_j)}. \end{cases} \tag{4.2}$$

We shall model a bivariate Hawkes' process as we look at the occurrences of order book. There is significant evidence to show empirical bid-ask side orders do appear in a clustered way, so it makes sense to model them bothside within two-dimensional Hawkes' processes. We define a basic function as follows

$$\begin{cases} \mu_A = \mu_B = \mu \\ \alpha_{AA} = \alpha_{BB} = \alpha_{self} \\ \alpha_{AB} = \alpha_{BA} = \alpha_{cross} \\ \beta_A = \beta_B = \beta_{self}. \end{cases} \tag{4.3}$$

## 4.2 Simulation of Two-Dimensional Hawkes Processes

### 4.2.1 Simulation Method

We have simulated one-dimensional Hawkes' process before, here we generalize one-dimensional algorithm in a two-dimensional setting. Firstly we recall the following procedure:

- Generate a random uniform distribution on interval [0,1]

- The whole process is simulated on time interval $[0, T]$,and we defined

$$I^K(t) = \sum_{n=1}^{K} \lambda^n(t),$$

the sum of intensities of the $K$ components of the multivariate process. So $I^2(t) = \sum_{n=1}^{2} \lambda^2(t)$ is the total intensities of bivariate Hawkes' process and we set initial value $I^0 = 0$.

- Algorithm-Initialization

1. Set $i \leftarrow 1$, $i^1 \leftarrow 1$,...,$i^M \leftarrow 1$ and

$$I^* \leftarrow I^M(0) = \sum_{n=1}^{M} \lambda_o^i(0).$$

2. First Event: Generate a random uniform number $U$ on $[0, 1]$
   (a) Set $s \leftarrow -\frac{1}{\lambda^*} \ln U$,
   (b) If $s > T$, then go to last step.
   (c) Attribution test:
      i. Generate $D \rightsquigarrow U_{[0,1]}$ and set $t_1^{n_0} \leftarrow s$.
      ii. Where $n_0$ is such that $\frac{I^{n_0-1}(0)}{I^*} < D \leq \frac{I^{n_0}(0)}{I^*}$.
   (d) Set $t_1 \leftarrow t_1^{n_0}$.

- Algorithm-Main routine

1. Main routine: Set $i^{n_0} \leftarrow i^{n_0} + 1$ and $i \leftarrow i + 1$.
   (a) Update maximum intensity: set $I^* \leftarrow I^M(t_{i-1}) + \sum_{n=1}^{M} \sum_{j=1}^{K} \alpha_j^{nn_0}$.
   (b) New event:
      i. Generate a random uniform number $U$ on $[0, 1]$ and set $s \leftarrow s - \frac{1}{I^*} \ln U$.
      ii. If $s > T$, then go to last step.
   (c) Attribution Rejection test: Generate $D \rightsquigarrow U_{[0,1]}$.
      i. If $D \leq \frac{I^M(s)}{I^*}$, then set $t_{i^{n_0}}^{n_0} \leftarrow s$,
      ii. Where $n_0$ is such that $\frac{I^{n_0-1}(s)}{I^*} < D \leq \frac{I^{n_0}(s)}{I^*}$.
      iii. And $t_i \leftarrow t_{i^{n_0}}^{n_0}$ go through the general routine again.

iv. Else update $I^* \leftarrow I^M(s)$ and try a new date as $step(b)$ of general routine.

2. Output: Retrieve the simulated process $(\{t_i^n\}_i)_{n=1,2,...,M}$ on $[0,T]$.

Actually we simulate cross-correlated intensity processes by using the thinning method again. We thus simulated a bivariate Hawkes' processes with $K = 1$ and the following parameters:

$$\begin{cases} \mu_A = 1.2, \alpha_{AA} = 0.6, \alpha_{AB} = 0, \quad \beta_A = 0.8 \\ \mu_B = 1.2, \alpha_{BA} = 0, \quad \alpha_{BB} = 0.6, \beta_B = 0.8 \end{cases} \tag{4.4}$$

Sample path simulated two side market order activity rates. In Fig 4.2.1, when influential trades arrive, the activity of both buy and sell orders increase but by differing amounts.

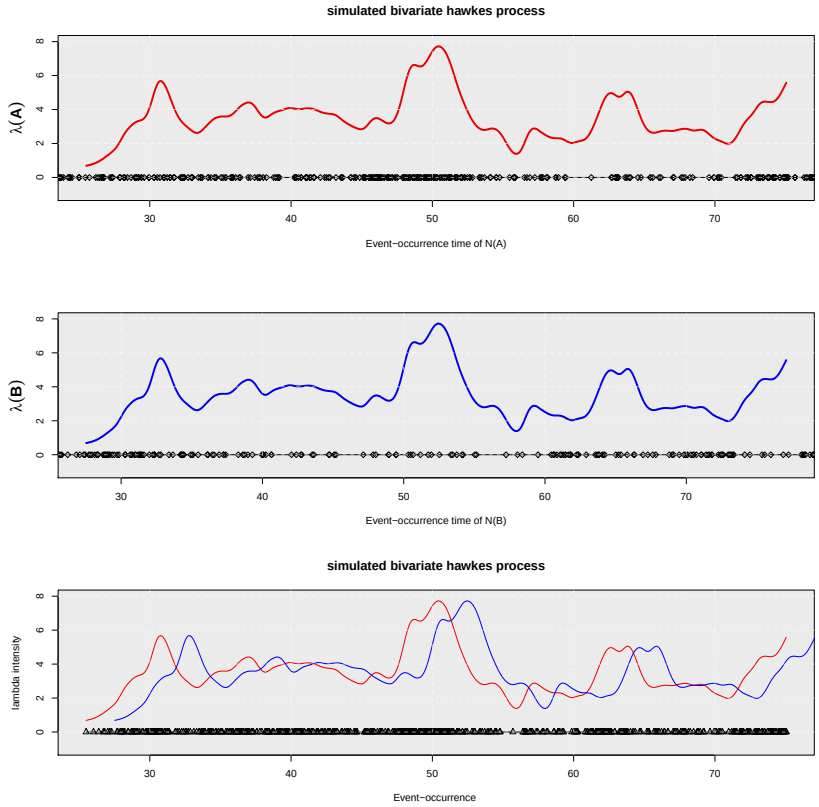



Figure 4.1: Simulated intensities of two-dimensional Hawkes' processes. The black points represent events occurence.

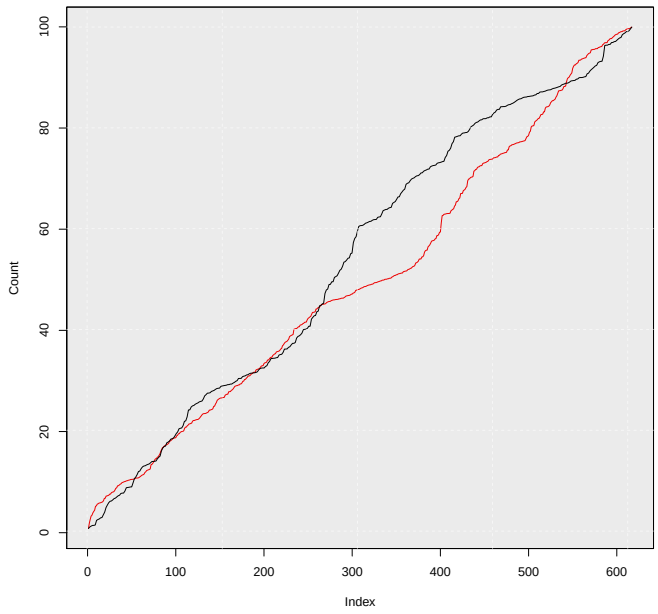



Figure 4.2: Counting of two-dimensional Hawkes, process

## 4.3 Maximum Likelihood Estimation of Two-Dimensional Hawkes Processes

### 4.3.1 Maximum likelihood estimation method

Based on data set $(t_i^A)_{i\geq 1}$ and $(t_i^B)_{i\geq 1}$, the log-likelihood function for two-dimensional Hawkes can be written as

$$L_T(\mu_A, \mu_B, \alpha_{AA}, \alpha_{AB}, \alpha_{BA}, \alpha_{BB}, \beta_A, \beta_B) = L_T^A(\mu_A, \alpha_{AA}, \alpha_{AB}, \beta_A) + L_T^B(\mu_B, \alpha_{BA}, \alpha_{BB}, \beta_B).$$

The first part of MLE function can be expressed as

$$\begin{aligned} L_T^A(\mu_A, \alpha_{AA}, \alpha_{AB}, \beta_A) = &\sum_{i=2}^{n} log\{\mu_A + \alpha_{AA} R_{AA}(i) + \alpha_{AB} R_{AB}(i)\} \\ &- \mu_A T - \frac{\alpha_{AA}}{\beta_A} \sum_{i=1}^{n} \{1 - e^{-\beta_A (T - t_i^{(A)})}\} \\ &- \frac{\alpha_{AB}}{\beta_A} \sum_{j=1}^{m} \{1 - e^{-\beta_A (T - t_j^{(B)})}\}. \end{aligned} \tag{4.5}$$

Similarly the second part of MLE function can be rewritten

$$\begin{aligned} L_T^B(\mu_B, \alpha_{BA}, \alpha_{BB}, \beta_B) = &\sum_{j=2}^{m} log\{\mu_B + \alpha_{BA} R_{BA}(i) + \alpha_{BB} R_{BB}(i)\} \\ &- \mu_B T - \frac{\alpha_{BA}}{\beta_B} \sum_{i=1}^{n} \{1 - e^{-\beta_B (T - t_i^A)}\} \\ &- \frac{\alpha_{BB}}{\beta_B} \sum_{j=1}^{m} \{1 - e^{-\beta_B (T - t_j^B)}\}, \end{aligned} \tag{4.6}$$

where $R(i, j)$ are given recursively,

$$R_{AA}(1) = R_{AB}(1) = R_{BA}(1) = R_{BB}(1) = 0.$$

So we can obtain

$$R_{AA}(i) = e^{-\beta_A (t_i^{(A)} - t_{i-1}^{(A)})} \{1 + R_{AA}(i-1)\},$$

where $R_{AA}(i)$ can recursively express as

$$\begin{aligned} R_{AA}(i) &= \sum_{i'=1}^{i} e^{-\beta_A (t_i^{(A)} - t_{i'}^{(A)})} \\ &= e^{-\beta_A (t_i^{(A)} - t_0^{(A)})} + e^{-\beta_A (t_i^{(A)} - t_1^{(A)})} + e^{-\beta_A (t_i^{(A)} - t_2^{(A)})} + ... + e^{-\beta_A (t_i^{(A)} - t_{i-1}^{(A)})} \\ &= e^{-\beta_A (t_i^{(A)} - t_{i-1}^{(A)})} e^{-\beta_A (t_{i-1}^{(A)} - t_{i-2}^{(A)})} ... e^{-\beta_A (t_1^{(A)} - t_0^{(A)})} + e^{-\beta_A (t_i^{(A)} - t_{i-1}^{(A)})} e^{-\beta_A (t_{i-1}^{(A)} - t_{i-2}^{(A)})} ... e^{-\beta_A (t_2^{(A)} - t_1^{(A)})} \\ &+ ... + e^{-\beta_A (t_i^{(A)} - t_{i-1}^{(A)})} \\ &= e^{-\beta_A (t_i^{(A)} - t_{i-1}^{(A)})} (e^{-\beta_A (t_{i-1}^{(A)} - t_{i-2}^{(A)})} ... e^{-\beta_A (t_1^{(A)} - t_0^{(A)})} + e^{-\beta_A (t_{i-1}^{(A)} - t_{i-2}^{(A)})} ... e^{-\beta_A (t_2^{(A)} - t_1^{(A)})} + ... + 1) \\ &= e^{-\beta_A (t_i^{(A)} - t_{i-1}^{(A)})} (1 + \sum_{i'=1}^{i-1} e^{-\beta_A (t_i^{(A)} - t_{i'}^{(A)})}) \\ &= e^{-\beta_A (t_i^{(A)} - t_{i-1}^{(A)})} \{(1 + R_{AA}(i-1))\}. \end{aligned} \tag{4.7}$$

Similarly we can recursively get expression of $R_{AB}(i)$ as

$$R_{AB}(i) = e^{-\beta_A (t_i^{(A)} - t_{i-1}^{(A)})} (R_{AB}(i-1)) + \sum_{\{j: t_{i-1}^{(A)} \le t_j^{(B)} < t_i^{(A)}\}} e^{-\beta_A (t_i^{(A)} - t_j^{(B)})}. \tag{4.8}$$

$$
\begin{aligned}
R_{AB}(i) &= \sum_{j'=1}^{i} e^{-\beta_A(t_i^{(A)}-t_{j'}^{(B)})} \\
&= e^{-\beta_A(t_i^{(A)}-t_0^{(A)})} + e^{-\beta_A(t_i^{(A)}-t_1^{(A)})} + ... + e^{-\beta_A(t_i^{(A)}-t_{j^*-1}^{(B)})} + e^{-\beta_A(t_i^{(A)}-t_{j^*}^{(B)})} \\
&= e^{-\beta_A(t_i^{(A)}-t_0^{(A)})} + e^{-\beta_A(t_i^{(A)}-t_1^{(A)})} + ... + e^{-\beta_A(t_i^{(A)}-t_{j^*-1}^{(B)})} + \sum_{\{j':t_{i-1}^{(A)}\le t_{j'}^{(B)}<t_i^{(A)}\}} e^{-\beta_A(t_i^{(A)}-t_{j'}^{(B)})} \\
&= e^{-\beta_A(t_i^{(A)}-t_{i-1}^{(A)})}e^{-\beta_A(t_{i-1}^{(A)}-t_{i-2}^{(A)})}...e^{-\beta_A(t_1^{(A)}-t_0^{(A)})} + e^{-\beta_A(t_i^{(A)}-t_{i-1}^{(A)})}e^{-\beta_A(t_{i-1}^{(A)}-t_{i-2}^{(A)})}...e^{-\beta_A(t_2^{(A)}-t_1^{(A)})} \\
&+ e^{-\beta_A(t_i^{(A)}-t_{i-1}^{(A)})}e^{-\beta_A(t_{i-1}^{(A)}-t_{j^*-1}^{(B)})} + \sum_{\{j':t_{i-1}^{(A)}\le t_{j'}^{(B)}<t_i^{(A)}\}} e^{-\beta_A(t_i^{(A)}-t_{j'}^{(B)})} \\
&= e^{-\beta_A(t_i^{(A)}-t_{i-1}^{(A)})}(e^{-\beta_A(t_{i-1}^{(A)}-t_{i-2}^{(A)})}...e^{-\beta_A(t_1^{(A)}-t_0^{(A)})} + e^{-\beta_A(t_{i-1}^{(A)}-t_{i-2}^{(A)})}... \\
&e^{-\beta_A(t_2^{(A)}-t_1^{(A)})} + ... + e^{-\beta_A(t_{i-1}^{(A)}-t_{j^*-1}^{(B)})}) + \sum_{\{j':t_{i-1}^{(A)}\le t_{j'}^{(B)}<t_i^{(A)}\}} e^{-\beta_A(t_i^{(A)}-t_{j'}^{(B)})} \\
&= e^{-\beta_A(t_i^{(A)}-t_{i-1}^{(A)})}(R_{AB}(i-1)) + \sum_{\{j':t_{i-1}^{(A)}\le t_{j'}^{(B)}<t_i^{(A)}\}} e^{-\beta_A(t_i^{(A)}-t_{j'}^{(B)})}.
\end{aligned}
\tag{4.9}
$$

$$
R_{BA}(j) = e^{-\beta_B(t_j^{(B)}-t_{j-1}^{(B)})}(R_{BA}(j-1)) + \sum_{\{i:t_{j-1}^{(B)}\le t_i^{(A)}<t_j^{(B)}\}} e^{-\beta_B(t_j^{(B)}-t_i^{(A)})}. \tag{4.10}
$$

and

$$
R_{BB}(j) = e^{-\beta_B(t_j^{(B)}-t_{j-1}^{(B)})}\{1 + R_{BB}(j-1)\}. \tag{4.11}
$$

Estimation of the resulting likelihood can be computationally intensive. Hawkes' process provides a suitable estimation of high-quantile based measures for financial time series. we conduct this two-dimensional estimation method for empirical in the following Chapter 5.

### 4.3.2 Result of Estimation

Table 4.1 and Table 4.2 summarized the maximum likelihood estimation of a two-dimensional Hawkes processes for simulated data from both the bid side and the ask side. When we run the MLE program each estimation results are a statistic summary with 100 computed samples of length $[0, 100]$ by 1000 times simulation of the bivariate process. Therefore we assigned the original value for each parameter at the beginning of simulated process as Equation 4.4.

When we compute the maximum-likelihood estimates of parameters for our model we

| Pars | $\mu_A$ | $\alpha_{AA}$ | $\alpha_{AB}$ | $\beta_A$ |
|---|---|---|---|---|
| Average | 1.554 | 0.749 | 0.027 | 1.100 |
| Median | 1.570 | 0.725 | 0.030 | 1.120 |
| Minimum | 1.260 | 0.660 | 0.010 | 0.920 |
| Maximum | 1.810 | 0.990 | 0.050 | 1.420 |
| Standard Deviation | 0.1869 | 0.1042 | 0.0141 | 0.1694 |

Table 4.1: Statistics summary for the maximum-likelihood estimation of the ask side of Hawkes' model

| Pars | $\mu_B$ | $\alpha_{BA}$ | $\alpha_{BB}$ | $\beta_B$ |
|---|---|---|---|---|
| Average | 1.462 | 0.023 | 0.619 | 0.847 |
| Median | 1.410 | 0.020 | 0.630 | 0.865 |
| Minimum | 1.290 | 0.000 | 0.500 | 0.660 |
| Maximum | 1.870 | 0.060 | 0.690 | 0.960 |
| Standard Deviation | 0.1680 | 0.0188 | 0.0662 | 0.0943 |

Table 4.2: Statistics summary for the maximum-likelihood estimation of the bid side of Hawkes' model

use algorithm resulted in section 4.3. We take into account that there are huge variations of trade frequency through the day so we restrict our observations to $9:30AM$ to $4:00PM$. At noon the A-shares were divided. Transaction data cleaning makes the stationary assumptions we will make in section 4.3 more realistic.

The Table 4.1 and Table 4.2 show the estimation values found at the ask and bid sides. They shown the maximum likelihood estimation where the true parameters converge. The results observed are a very large variation in the numerical values of the maximum likelihood. However, what size are the parameters, so that it is clear to observe the cross-exciting effect any given side of the trades through on the opposite side and is much weaker than self-exciting effect, which deals with the clustering of trades through. The average value of $\alpha_{AB}$ is nearly 28 times smaller than the average value of $\alpha_{AA}$, while in the meantime the associated exponential decay $\alpha_{BA}$ is also 27 times smaller than the average $\alpha_{BB}$.

Our estimation results show that a simple bivariate Hawkes process fits nicely on observations of bid-ask both side trades. We note that the cross-influence of bid and ask both side are weak, such as the weak cross-exciting effect. We will thus use our two-dimensional model to A-H stock market and even discover trading strategy in a future work.

# Chapter 5

# Application

Research in finance field is moving toward modeling high-frequency data recently. The reason is that high-frequency data contains a lot of information that have significant effect on the liquidity of financial markets. That is, the best available data from A-H shares in terms of frequency (almost tick by tick from Bloomberg), has the invaluable advantage of containing trading information. However, there is a lack of information on high-frequency data model and analysis of liquidity of China financial market especially in A-H stock in Mainland China and Hong Kong stock market.

In this chapter, we employ the Hawkes model to compare one-dimensional and two-dimensional empirical performances. The proposed modelling framework provides important theoretical support on testing A-H shares. The dynamics transaction data display strong seasonality. Our results suggested that the two-dimensional Hawkes model would be obviously superior to one-dimensional model work, especially on explaining A-H shares trading data. This chapter is structured as follows: First, we describe the data source. Next, we describe the one-dimensional and two-dimensional model parameters and then focus on model fitting. Lastly, we discuss the implications of our findings.

## 5.1 Data Description

We now turn to the description of the procedure estimating the Hawkes process by using real transaction data. Our results are obtained making use of The Hang Seng China

AH Premium Index ("HSAHP"). On November 17, 2014, Shanghai stock exchange and Hong Kong stock exchange allowed investors of both areas to buy and sell stocks listed on the other exchange within the specified range through local securities companies. This is the Shanghai-Hong Kong stock connect.This new trading mechanism is conducive to strengthening the connection between the two capital markets and improving the openness of China's capital markets. In the context of the opening of Shanghai-Hong Kong Stock Connect, A-shares are listed in the Shanghai stock exchange and would be largely available to domestic investors. However, H-shares are listed in the Hong Kong stock exchange and would be mostly available to non-Chinese investment. In both cases the original business relates to Mainland China. Prices at which Mainland companies trade in both exchanges are hugely different and interestingly there is no channel to arbitrage. A-shares trade at a premium compared with H-share counterparts (even more so that quoted on the London exchanges). Due to this large imbalance between supply and demand of high quality stock in China they are highly restricted and regulated combined with high demand newfound investment of wealth is increasing the premium prices up. To measure this the Hang Seng China AH Premium Index list the spread between A-shares and H-shares and dual list them in Mainland China. Shares of companies domiciled to Mainland China are tracked on both the average price and the difference between A- and H- Companies with both shares are listed as A-H Companies. The Hang Seng AH Premium Index calculates the weighted average premium (or discount) of A shares against H shares based on the market capitalization of A shares and H shares of component stocks included in the index. The higher the index, the more expensive the A-shares are relative to the H-shares (the higher the premium). Conversely, the lower the index, the cheaper the A-shares are relative to the H-shares.

Before 2000, the price of A shares of many companies was 5 times or even 10 times higher than that of H shares. Moreover, the volatility of the stock market between the Mainland and Hong Kong before 2005 was basically irrelevant. The linkage between the Hong Kong market and the US stock market was obvious. At that time, the A-shares were the most independent stock market in the world, and they completely had their own logic of volatility, and with any country in the world. There is no obvious link between the regional or regional stock markets. However, after 2000, the correlation between the two markets is becoming stronger and stronger. The trend of A-shares and H-shares is a major trend, but the factors that affect the two markets are not the same at different points in time. It is

bound to make A-shares and H-shares. The premium between them is in volatility. Under normal circumstances, the same stock, the issue price of its A shares tend to be higher than the issue price of the H-share market. Given the highly speculative nature of the Mainland market, the turbulent influence of capital on the market, and the relatively vague market pricing model, it may be difficult for the A and H stock market to achieve full interconnection and inter-operability. However, the gradual convergence of the markets in the two places is still a general direction.

For data on A and H shares, the Hawkes model is estimated by using data on two different kinds of industries classification, there are two pairs of financial and utilities industry respectively, A shares and H shares companies four months data on August and November 2017. For both databases we use tick-by-tick data on all trades each trading day. We begin the estimation on one intra-day at the beginning of the regular trading day (09:30 AM for the A shares and 09:00 AM for the H shares - both times are local times) and ending it at the closing of the trading day (03:00 PM for the A shares and 04:00 PM for the H shares).As the data downloaded from Bloomberg, it included lots of different types of data, but in data cleansing procedure, we deleted useless data such as occurred in pre-opening period. In the data pre-processing (for A-shocks), the multiple transactions in the second are merged first, avoiding the impact of zero duration, and eliminating transaction data outside the opening and closing time. Since there is a midday rest system, it is also necessary to subtract the afternoon trading interval for each trading time in the day, so as to ensure that the data time with the noon transaction is continuous and uninterrupted.Information about Hang Seng Stock Connect China AH Premium Index were downloaded from the website of Hang Seng Indexes.

From industry weighting in HSAHP Fig 5.1, we have chosen the largest weight financial industrial classification with a proportion of up to 66.03%. We then chose the smallest weighted Utilities category, only 1.06%. Although the weight of utilities industry is lower than the information technology industry, there are 4 pair stocks in Utilities classification and only 1 pair stock in the information technology classification. Therefore, we chose the utilities category with the lowest proportion as the target for Financial industrial to compare. What is interesting is that Huaneng Power, which has a relatively small utilities industry, has a higher proportion of A/H Price Ratio (157.51%), while the A/H ratio of large-scale

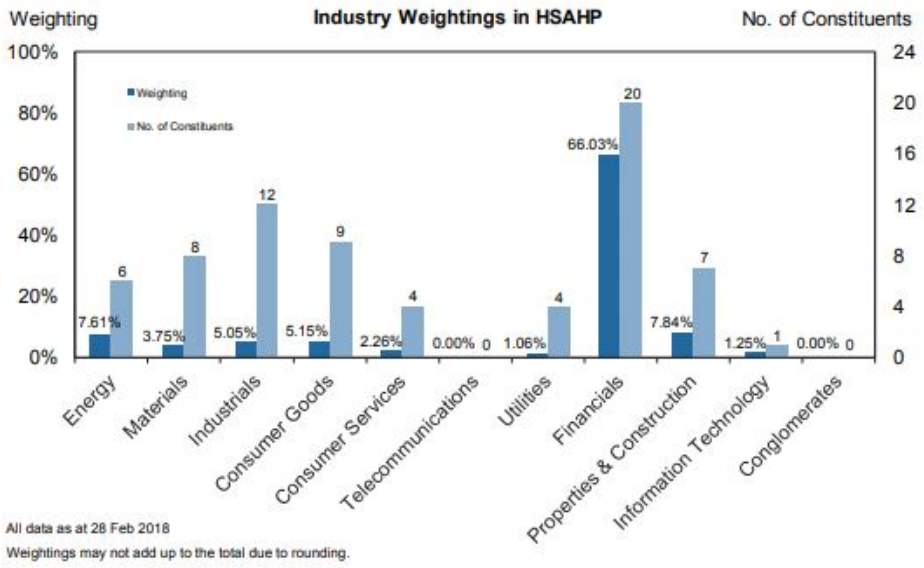



Figure 5.1: Industry Weightings in HSAHP (issued at 28 Feb 2018).

financial industtry ICBC is 122.02%. We firstly present the results when the estimation is

Table 5.1: Summary table of tested A-H shares companies.

| NO. | Company Name | A- | H- | Industrial Classification |
|---|---|---|---|---|
| 1 | Huaneng Power International Inc. | 600011 | 0902 | Utilities Industry |
| 2 | Industrial and Commercial Bank of China Limited | 601398 | 1398 | Financial Industrial |

conducted using Huaneng Power International Inc. A and H shares data. Next, we give a more formal assessment of the performance of the model and calculate the fit of the Hawkes process in QQ-plot. Both estimation procedures and the fitted procedures all have better results.

## 5.2 One-dimensional Hawkes' Process Results on A-H Shares Database

Now we estimate Hawkes' process by using Industrial and Commercial Bank of China Limited (ICBC) and Huaneng Power International Incorporation A-H shares monthly trades with one second time-stamp. In this work, the estimation is carried out by using all observations for a whole trading day at once (from August, 1st, 2017 to November, 30th, 2017).

Daily intervals are long enough to obtain reliable calibrations. In such time intervals, the exogenous activity $\mu$ in Equation (3.6) can be approximated by a constant. As encouraged by Bacry and Muzy [6], fitting longer time windows dependent background intensity is a good way to account for the various activities. The model is also relatively simple to interpret and estimate. The ability to generate endogenous clustering, controlled by the parameters $\alpha$ and $\beta$ when the model has an exponential kernel, makes the Hawkes process be a good choice for financial markets.

### 5.2.1 MLE Estimation Results

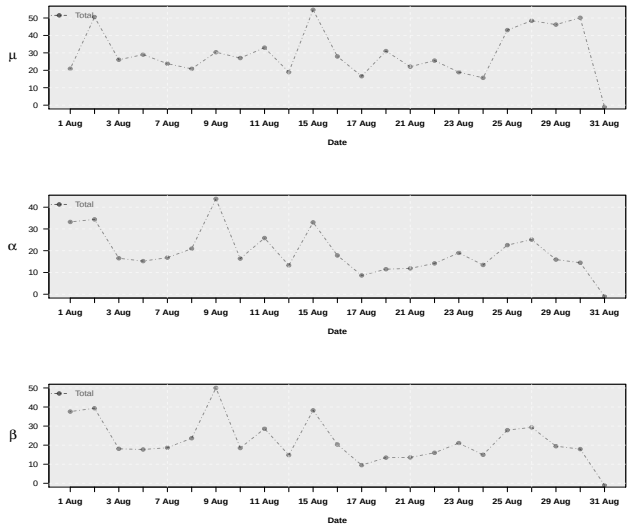
Firstly, we focus on Huaneng Power International Inc.(A-600011 and H-0902), which is a small company by the market capitalization trading in Mainland China shock market, the estimation performance shows in Fig 5.2 and the statistic summary of parameters arisen in Table 5.2.

| $T$ | $\hat{\mu}$ | $\hat{\alpha}$ | $\hat{\beta}$ |
|---|---|---|---|
| August | 29.55(0.0972) | 19.26(17.0638) | 22.07(11.3051) |
| September | 26.53(0.4323) | 14.72(8.9128) | 17.09(4.3431) |
| October | 64.34(0.0460) | 56.68(5.2835) | 61.90(4.4438) |
| November | 33.05(0.2510) | 21.41(4.6401) | 24.16(9.4031) |

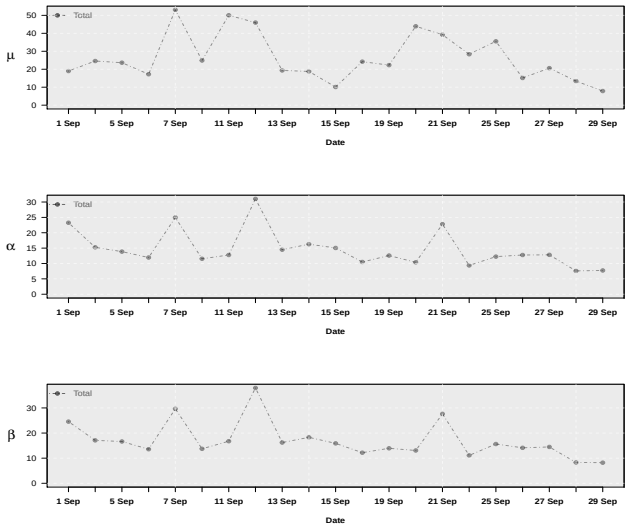
Table 5.2: Monthly parameters estimated summary by Hawkes' model from Huaneng Power International Inc. H-Shares 0902 on Aug-Nov 2017. Standard deviations are put in brackets.

Then focus on the same company which issued in Hong Kong stock market, and its performance shows in Fig 5.3. Also, Table 5.3 summarizes the list of model parameters.

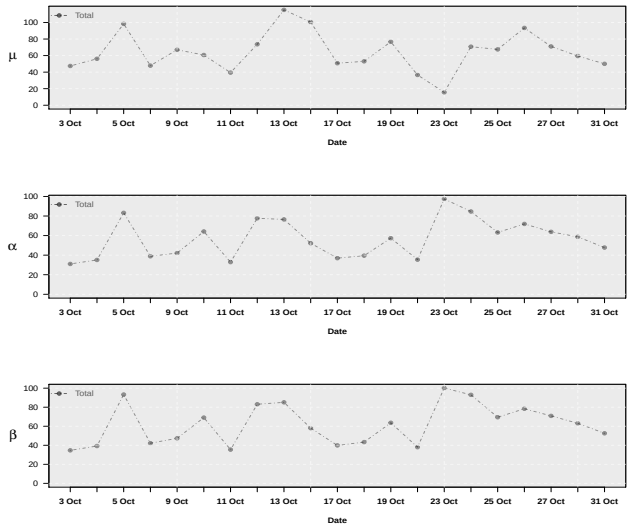
(a) H-0902(Aug 2017)




(b) H-0902(Sep 2017)




(c) H-0902(Oct 2017)




(d) H-0902(Nov 2017)

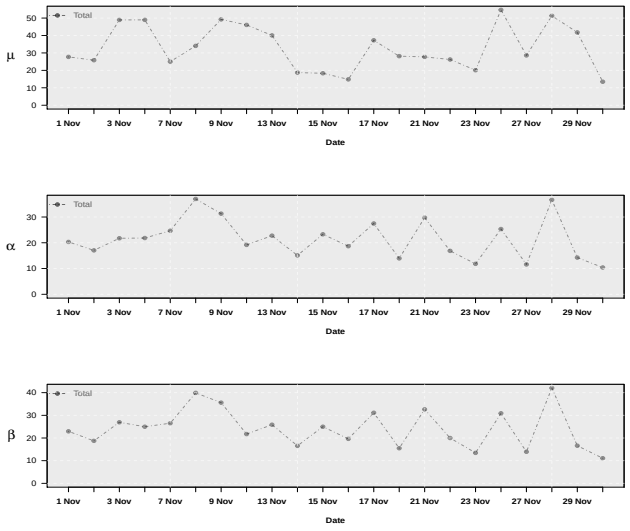
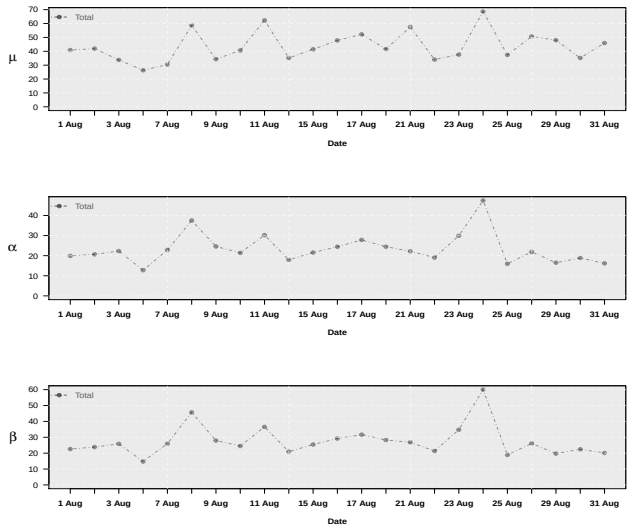
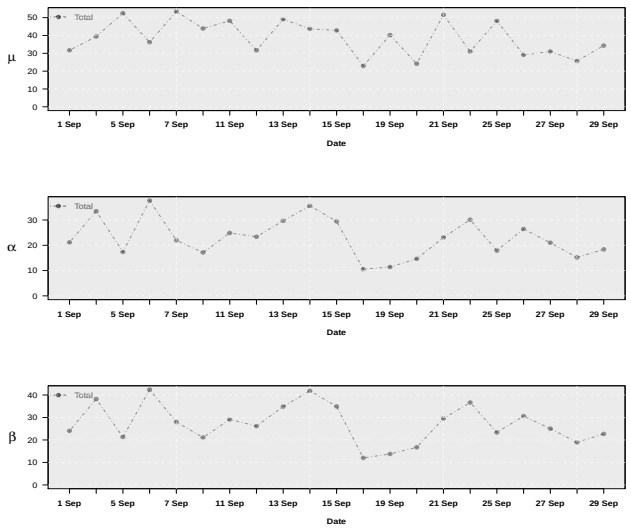



Figure 5.2: Panel (a), (b), (c), (d) are estimation results for H-Shares 0902 from Aug to Nov 2017.

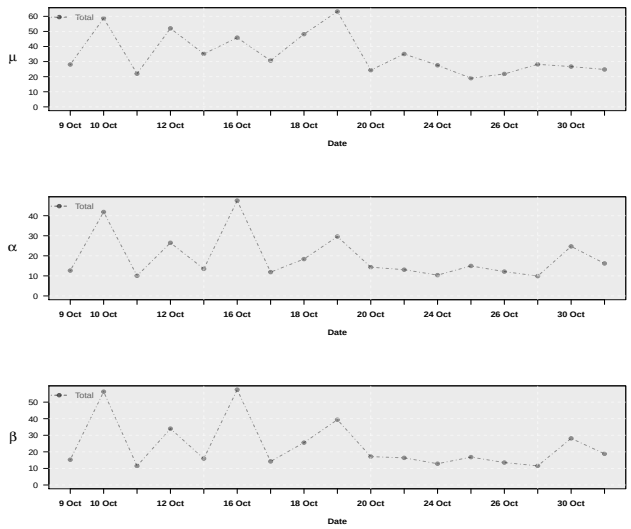
(a) A-600011(Aug 2017)




(b) A-600011(Sep 2017)




(c) A-600011(Oct 2017)




(d) A-600011(Nov 2017)

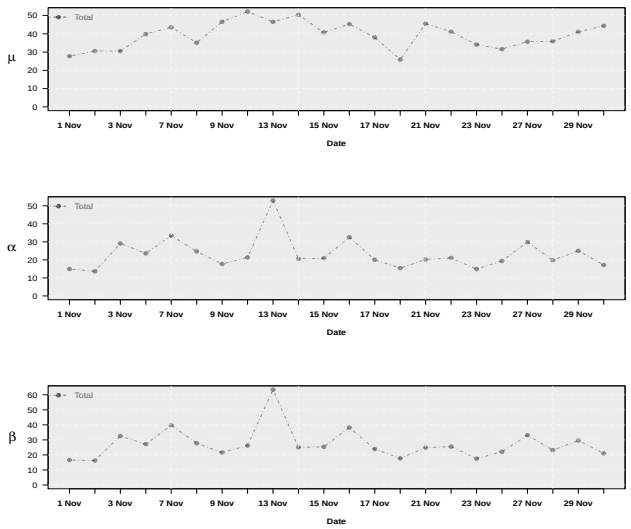



Figure 5.3: Panel (a), (b), (c), (d) are estimation results for A-Shares 600011 from Aug to Nov 2017.

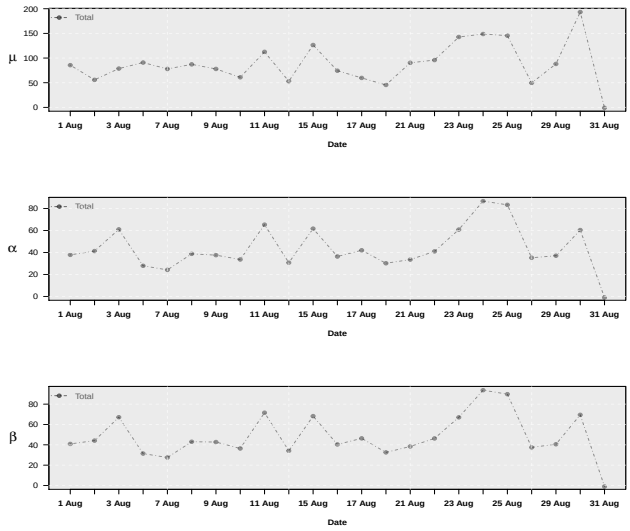
| $T$ | $\hat{\mu}$ | $\hat{\alpha}$ | $\hat{\beta}$ |
|---|---|---|---|
| August | 43.60(0.1144) | 23.32(12.2924) | 27.55(8.1746) |
| September | 38.63(0.1578) | 22.88(7.6576) | 27.21(5.2542) |
| October | 34.78(0.2167) | 19.27(13.1497) | 23.81(10.1391) |
| November | 39.22(0.1194) | 23.11(13.0621) | 27.22(8.9988) |

Table 5.3: Monthly parameters estimated summary by Hawkes' model from Huaneng Power International Inc. A-Shares 600011 on Aug-Nov 2017. Standard deviations are put in brackets.

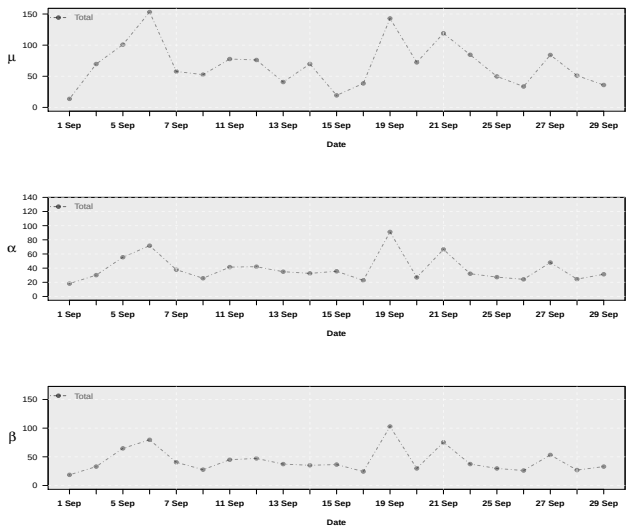
Secondly, we focus on the Industrial and Commercial Bank of China Limited(H-1398 and A-601398), which belongs to financial industry classification and is one of the largest bank (by market capitalization) issued in China stock market. The estimation performance

present in Fig 5.4 and the statistic summary arisen in Table 5.4.

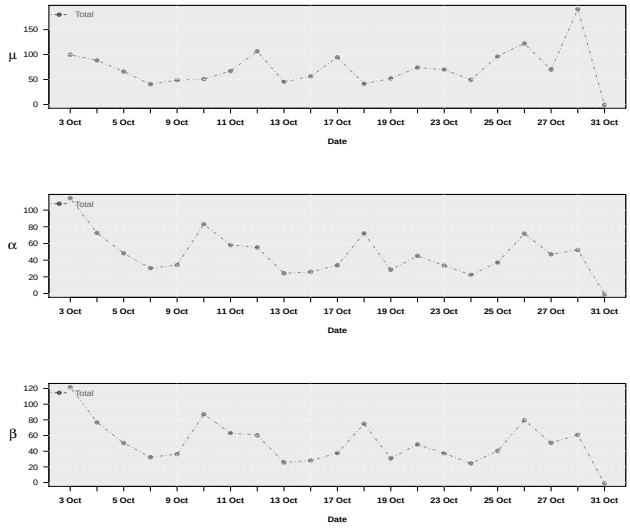
(a) H-1398(Aug 2017)




(b) H-1398(Sep 2017)




(c) H-1398(Oct 2017)




(d) H-1398(Nov 2017)

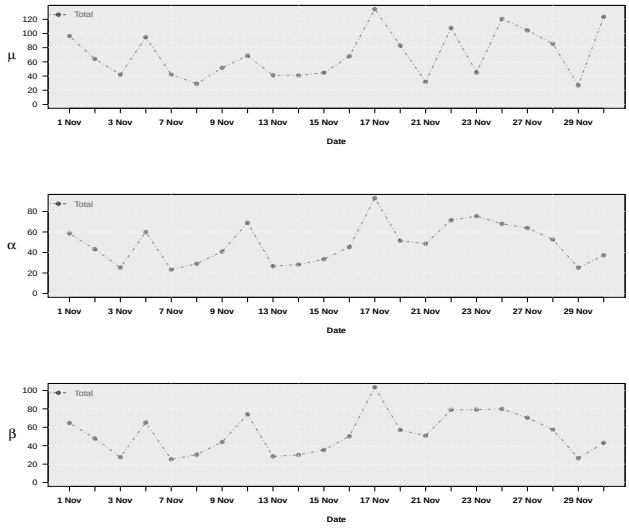


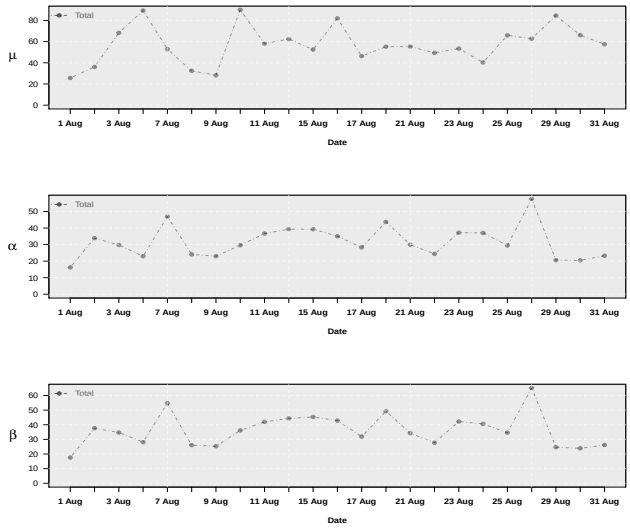

Figure 5.4: Panel (a), (b), (c), (d) are estimation results for H-Shares 1398 from Aug to Nov 2017.

| $T$ | $\hat{\mu}$ | $\hat{\alpha}$ | $\hat{\beta}$ |
|---|---|---|---|
| August | 88.68(0.0183) | 43.71(6.9529) | 48.18(5.3681) |
| September | 68.71(0.0564) | 39.04(15.7209) | 43.03(14.5349) |
| October | 72.83(0.0191) | 47.08(7.4776) | 50.80(5.5905) |
| November | 70.26(0.0305) | 48.62(12.1375) | 53.16(9.2531) |

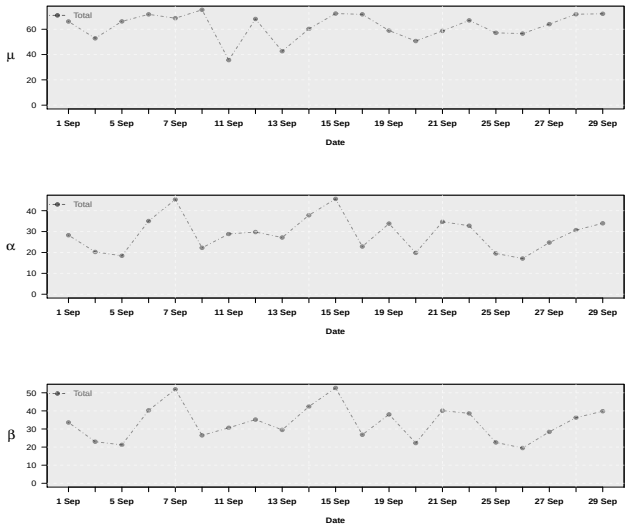
Table 5.4: Parameters estimated by Hawkes's model from Industrial and Commercial Bank of China Limited H-Shares 1398 for the period Aug-Nov 2017. Standard deviations are put in brackets.

Then we look at the same company traded in Hong Kong stock market. The estimated parameters are ploted in Fig 5.5 and the statistic summary are in Table 5.5.

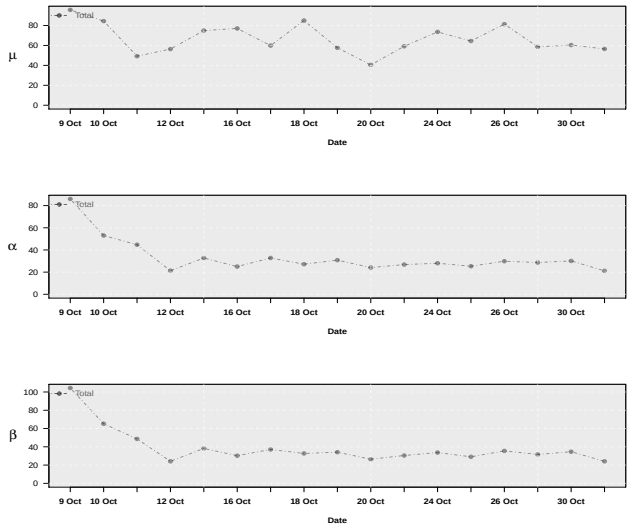
(a) A-601398(Aug 2017)




(b) A-601398(Aug 2017)




(c) A-601398(Aug 2017)




(d) A-601398(Aug 2017)

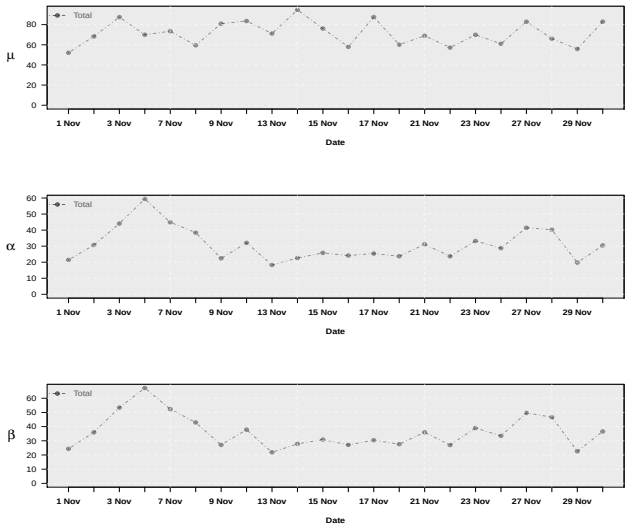



Figure 5.5: Panel (a), (b), (c), (d) are estimation results for A-Shares 601398 from Aug to Nov 2017.

| $T$ | $\hat{\mu}$ | $\hat{\alpha}$ | $\hat{\beta}$ |
|---|---|---|---|
| August | 57.15(0.0479) | 31.69(14.1942) | 36.26(11.4677) |
| September | 62.34(0.0534) | 28.99(5.8753) | 33.30(4.4381) |
| October | 66.73(0.0580) | 33.39(16.7818) | 38.83(13.2960) |
| November | 71.26(0.0485) | 31.00(7.3734) | 36.26(5.2908) |

Table 5.5: Estimated parameters by Hawkes' model from Industrial and Commercial Bank of China Limited A-Shares 601398 for the period 2017. Standard deviations are put in brackets.

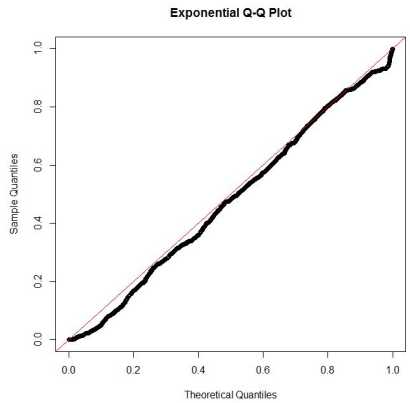
It is not very hard to see, the monthly average parameters Table 5.2 and Table 5.3 are basically performance. However, we have noticed that H shares jumps a lot in October, especially the base rate $\alpha$ as high as 64.34, so we guess this stock in October, there may be some strange movements. We all know that Hong Kong's stock market does not have a golden week in October, but Mainland China does. Could that be the reason? Let's look at another set of comparison data of A and H shares. Table 5.4 and Table 5.5 describe another set of financial stocks. But this group data were relatively stable, especially in October, the average estimation result showed no abnormalities. So it looks like the one-dimensional model works on estimating parameters, but it's not very good at expressing the financial sense. So next, we need to classify these large orders, separate the buy and sell, and then see if we can make some better explanations on financial implication.

### 5.2.2 Goodness of Fit

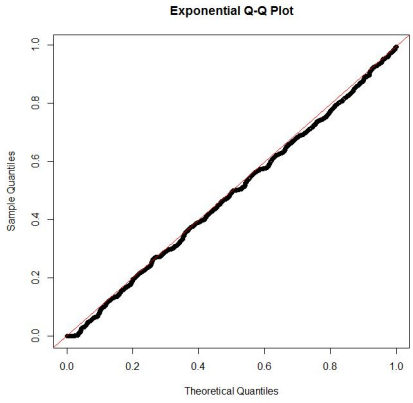
Extending the fits to five days a week requires the handling of weekly seasonality. First and most importantly, A-H shares order book does not operate at weekends, which implies that Mondays and Fridays most likely have a dynamics distinctly different from the other days. Thus, we fit two pairs A-H share orders, which amounts to A-shares 83 fits and H-shares 87 fits it is because 4 more trading days on October(five points per week,four monthly data). We therefore merge all the residuals from the fitting of all four months and construct the QQ-plot against the exponential distribution. The QQ-plot shown on Fig 5.6 visually confirms the accuracy of fit. This also shows that the one-dimensional Hawkes process description of the A-H shares' entire order arrival level is still relatively accurate. Evidence related to empirical transactions indicates that orders appear in clusters, so the stock market transaction orders as we would expect, has been already modelled by one

dimensional self-exciting Hawkes' process and it performs well.

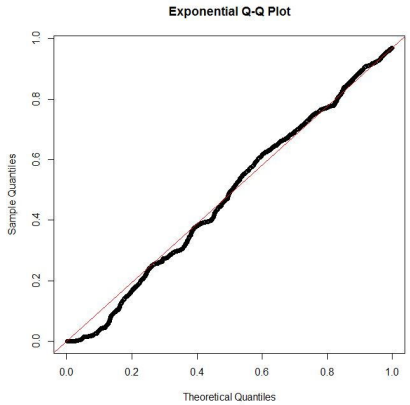
(a) QQ-plot of H-0902




(b) QQ-plot of A-600011




(c) QQ-plot of H-1398




(d) QQ-plot of A-601398

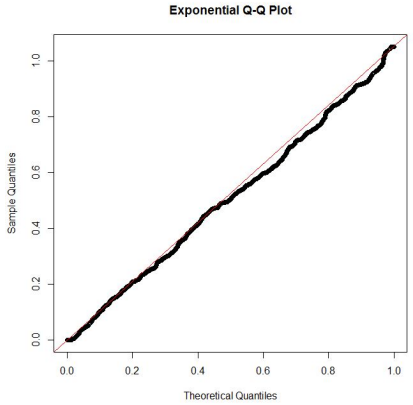



Figure 5.6: Goodness of fit tests of the residuals merged from all intervals under the null hypothesis of exponentially distributed (4 months fits).

## 5.3 Two-dimensional Hawkes' Process Results on A-H Shares Database

We perform the fit of a bivariate Hawkes model on bid-ask both side market orders on the previously described data: Industrial and Commercial Bank of China Limited (ICBC) and Huaneng Power International Incorporation.

We use two pairs stock monthly tick-by-tick trades data of A-H shares in Mainland China and Hong Kong stock market from August 1st 2017 to November 30 2017. This types of data gives trades (timestamp, price, volume, side of the order book) from opening to the close of the market. For each trading day, we extract the series of timestamps $(t_i^A)_{i\geq 1}$ and

$(t_i^B)_{i\geq 1}$ of the both side trades. Following this, we restrict the research to a trading day interval (for A-shares we need to delete trading orders in lunch time).

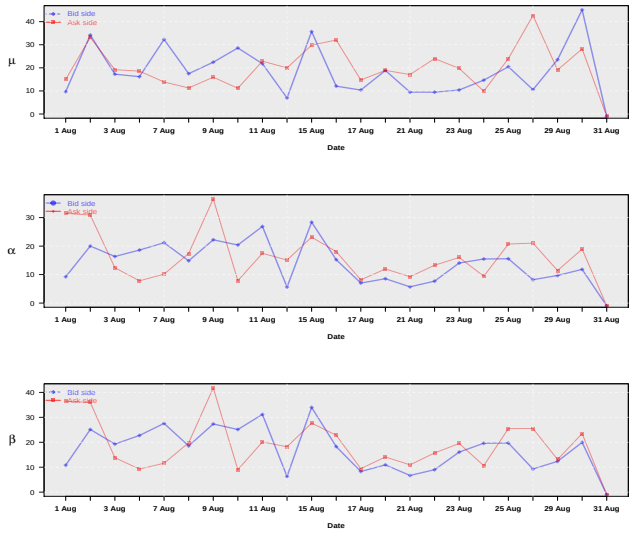
### 5.3.1 MLE Estimation Results

Now we begin estimating two-dimensional Hawkes' process by using Industrial and Commercial Bank of China Limited (ICBC) and Huaneng Power International Inc from A-H shares monthly trades with tome-stamps of one second's precision. In this work, the estimation was carried out by using all observations for a whole trading day. We compute the maximum likelihood estimation for the parameters presented in Chapter 4, Equation (4.2).

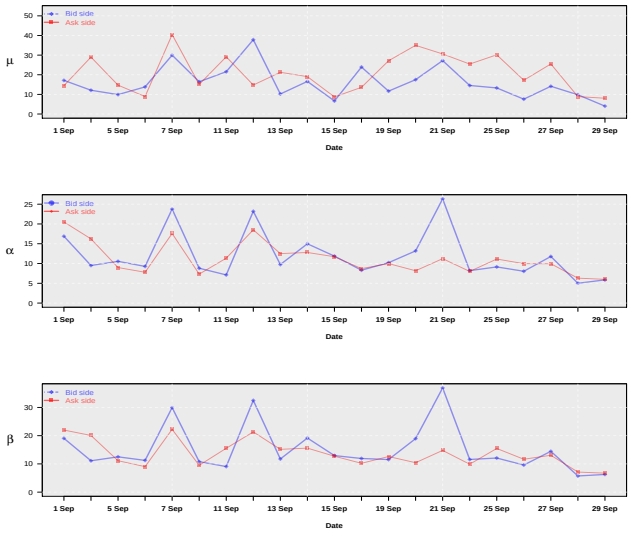
Table 5.6 summarizes the statistics on the H-share 0902 estimated values on the ask and bid sides. They show the MLE where the true parameters converge. The results we observations are a very large variation in the numerical values of the maximum likelihood. However, whatever size the parameters are, it is clear that the cross-exciting effect can be observed, but any given side of the cross-exciting is much weaker than self-exciting effect which on deals with the clustering of tradesthrough. For H-share 0902 the average value of $\alpha_{AB}$ is nearly 14 times smaller that the average value of $\alpha_{AA}$, while at the same time the associated exponential decay $\alpha_{BA}$ is 48 times smaller than the average $\alpha_{BB}$. From the parameters statistics summary, for H-share 0902 the branching ratio $n = \frac{\alpha}{\beta}$, which is equal to the total integrated intensity of an exponential kernel $\int_0^\infty g(\nu)d\nu$, is much weaker in the cross-exciting cases at ask side (taking the average values, $\frac{\alpha_{AB}}{\beta_A} = 0.059$) than in the self-exciting cases (still using the average values $\frac{\alpha_{AA}}{\beta_A} = 0.842$). The same situation with ask side, the cross-excitation cases at ask side (taking the average values, $\frac{\alpha_{BA}}{\beta_B} = 0.016$) than in the self-exciting cases (still using the average values $\frac{\alpha_{BB}}{\beta_B} = 0.817$).

Table 5.7 summarizes the statistics on the A-share 600011 estimated values on the ask and bid sides. For A-share the average value of $\alpha_{AB}$ is nearly 20 times smaller that the average value of $\alpha_{AA}$, while at the same time the associated exponential decay $\alpha_{BA}$ is nearly 25 times smaller than the average $\alpha_{BB}$. From the parameters statistics summary, for A-share 600011 the branching ratio $n = \frac{\alpha}{\beta}$, which is equal to the total integrated intensity of an exponential kernel $\int_0^\infty g(\nu)d\nu$, is much weaker in the cross-excitation cases at ask side

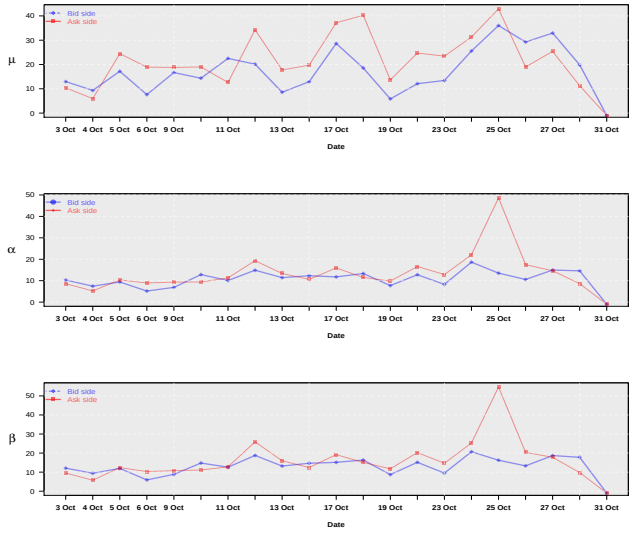
(a)H-0902(Aug 2017)




(b)H-0902(Sep 2017)




(c)H-0902(Oct 2017)




(d)H-0902(Nov 2017)

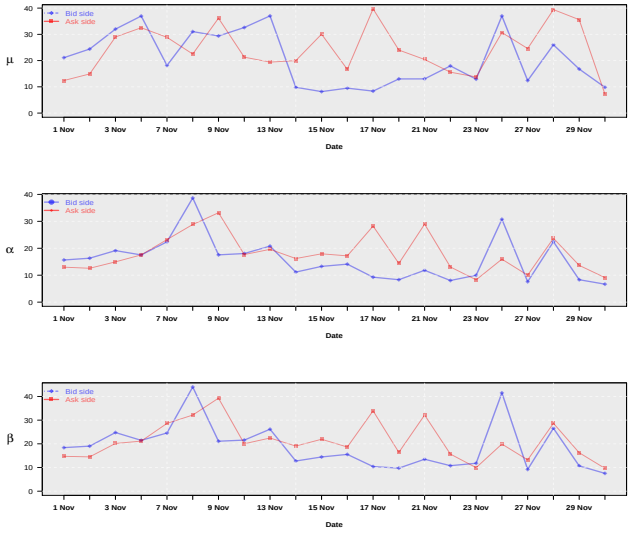
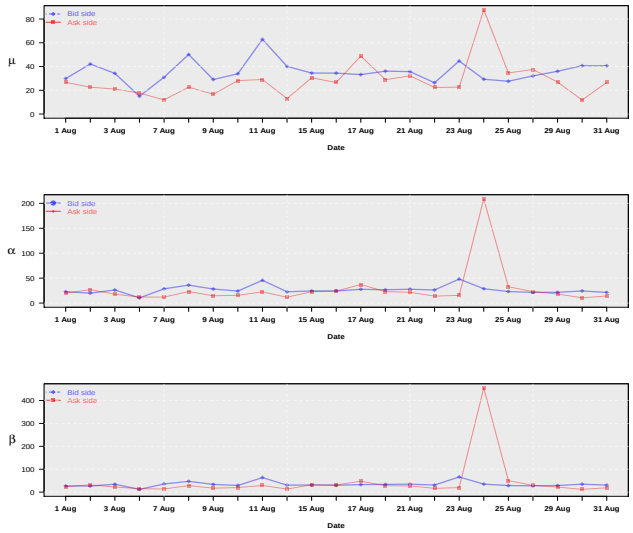
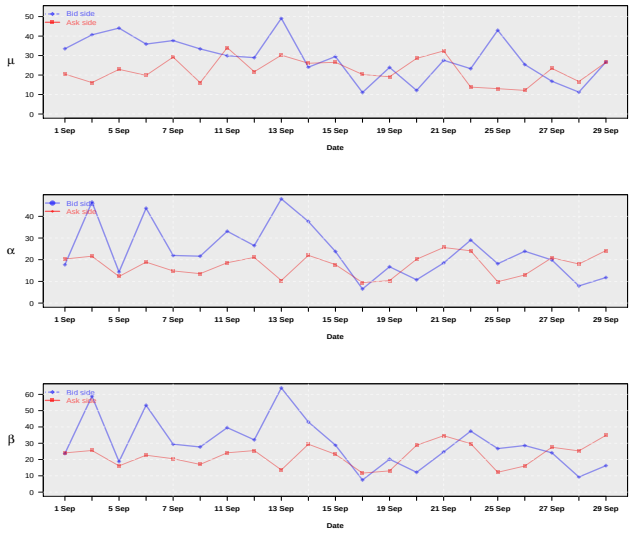



Figure 5.7: Panel (a), (b), (c), (d) are bid-ask side estimation results for H-Shares 0902 from Aug to Nov 2017.

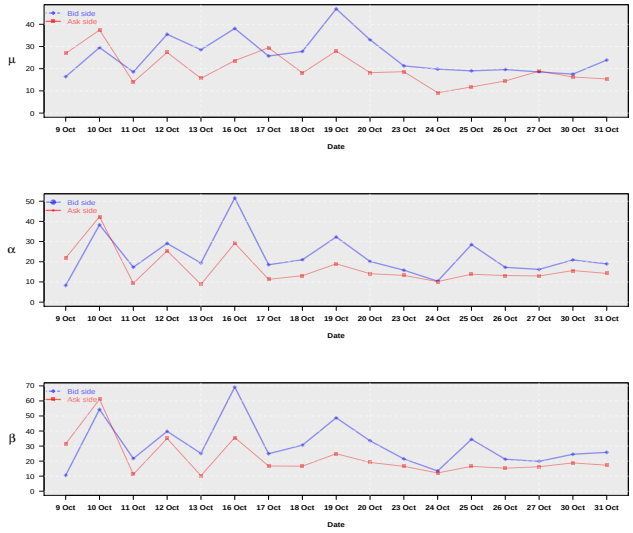
(a)A-600011(Aug 2017)




(b)A-600011(Sep 2017)




(c)A-600011(Oct 2017)




(d)A-600011(Nov 2017)

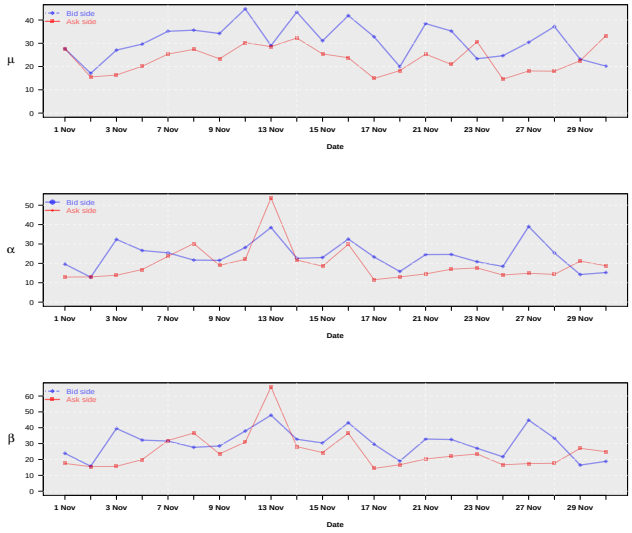



Figure 5.8: Panel (a), (b), (c), (d) are bid-ask side estimation results for H-Shares 600011 from Aug to Nov 2017.

| Pars | $\hat{\mu}_A$ | $\hat{\alpha}_{AA}$ | $\hat{\alpha}_{AB}$ | $\hat{\beta}_A$ | $\hat{\mu}_B$ | $\hat{\alpha}_{BA}$ | $\hat{\alpha}_{BB}$ | $\hat{\beta}_B$ |
|---|---|---|---|---|---|---|---|---|
| August | 20.01 | 15.98 | 1.02 | 18.84 | 18.55 | 0.35 | 13.99 | 17.26 |
| | (0.2202) | (6.0015) | (0.1020) | (4.1153) | (0.2043) | (0.2161) | (5.8104) | (2.1005) |
| September | 20.81 | 11.18 | 1.03 | 13.63 | 15.99 | 0.22 | 11.99 | 15.19 |
| | (0.5885) | (6.5331) | (0.2241) | (3.0566) | (1.3781) | (0.0662) | (4.8800) | (3.6222) |
| October | 21.42 | 13.51 | 1.13 | 15.94 | 17.30 | 0.24 | 10.76 | 12.99 |
| | (0.5375) | (3.3959) | (0.1674) | (2.5649) | (0.3453) | (0.0342) | (9.1313) | (6.5251) |
| November | 24.31 | 18.05 | 0.98 | 21.30 | 20.79 | 0.25 | 15.81 | 18.89 |
| | (0.6059) | (4.9822) | (0.5790) | (4.3827) | (0.5750) | (0.0521) | (4.5473) | (3.4092) |

Table 5.6: Statistics summary of bivariate Hawkes' model on H-0902 from Aug to Nov on 2017. Each estimation is the average result computed of a whole month every trading days. Standard deviations are given in parentheses.

(taking the average values, $\frac{\alpha_{AB}}{\beta_A} = 0.017$) than in the self-excitation cases (still using the average values $\frac{\alpha_{AA}}{\beta_A} = 0.724$ ). The same situation with ask side, the cross-excitation cases at ask side (taking the average values, $\frac{\alpha_{BA}}{\beta_B} = 0.041$) than in the self-excitation cases (still using the average values $\frac{\alpha_{BB}}{\beta_B} = 0.775$).

| Pars | $\hat{\mu}_A$ | $\hat{\alpha}_{AA}$ | $\hat{\alpha}_{AB}$ | $\hat{\beta}_A$ | $\hat{\mu}_B$ | $\hat{\alpha}_{BA}$ | $\hat{\alpha}_{BB}$ | $\hat{\beta}_B$ |
|---|---|---|---|---|---|---|---|---|
| August | 28.11 | 27.90 | 0.61 | 43.74 | 35.63 | 1.35 | 26.53 | 34.13 |
| | (0.2358) | (5.8786) | (0.3621) | (3.4802) | (0.2136) | (1.2060) | (3.1952) | (1.6164) |
| September | 22.34 | 17.48 | 0.50 | 22.65 | 28.95 | 1.61 | 23.72 | 29.18 |
| | (0.3801) | (1.6767) | (0.2241) | (0.8314) | (0.2717) | (2.151) | (7.8984) | (4.3395) |
| October | 20.17 | 16.93 | 0.48 | 22.05 | 29.49 | 1.21 | 27.50 | 37.50 |
| | (0.5550) | (8.5251) | (0.0540) | (6.3272) | (0.3999) | (0.3710) | (3.3978) | (1.9102) |
| November | 23.30 | 19.64 | 0.37 | 24.85 | 31.00 | 1.19 | 23.92 | 30.34 |
| | (0.1946) | (4.3695) | (0.2691) | (2.6208) | (0.4046) | (0.1344) | (8.8939) | (6.1148) |

Table 5.7: Statistics summary of bivariate Hawkes' model on A-600011 from Aug to Nov on 2017. Each estimation is the average result computed of a whole month every trading days. Standard deviations are given in parentheses.

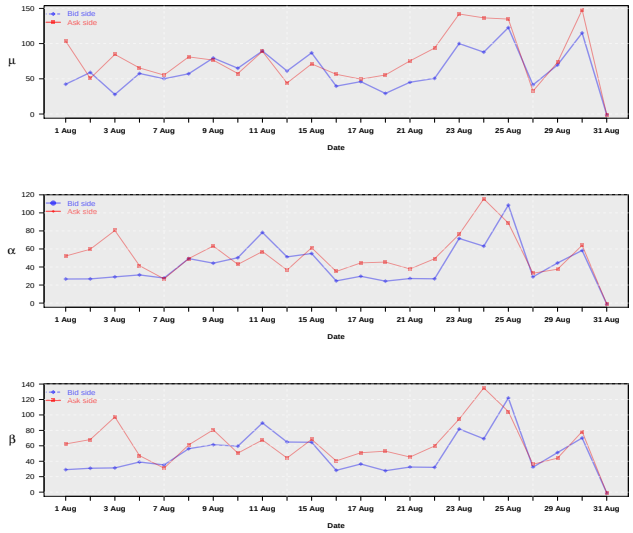
Secondly, we turn to focus on Industrial and Commercial Bank of China Limited(H-1398 and A-601398), which belongs to financial industry classification (actually, it is a large company by the market capitalization). The estimation performance present in Fig 5.9 and Fig 5.10,as related to Equation (4.2). For H-shares 1398 the average value of $\alpha_{AB}$ is nearly 25 times smaller that the average value of $\alpha_{AA}$, while at the same time the associated exponential decay $\alpha_{BA}$ is only 17 times smaller than the average $\alpha_{BB}$. The instantaneous effect is thus much smaller while its half-life is not significantly longer. For A-shares 600011 the average value of $\alpha_{AB}$ is nearly 60 times smaller that the average value of $\alpha_{AA}$, while at the same time the associated exponential decay $\alpha_{BA}$ is only 20 times smaller than the average $\alpha_{BB}$. Table 5.8 summarizes the statistics on the H-share 1398 estimated values on the ask and bid sides. From the parameters statistics summary, for H-share 1398 the branching ratio $n = \frac{\alpha}{\beta}$, which is equal to the total integrated intensity of an exponential kernel $\int_0^\infty g(\nu)d\nu$, is much weaker in the cross-excitation cases at ask side(taking the average values, $\frac{\alpha_{AB}}{\beta_A} = 0.036$ ) than in the self-excitation cases (still using the average values $\frac{\alpha_{AA}}{\beta_A} = 0.855$ ). The same situation with ask side, the cross-excitation cases at ask side(taking the average values, $\frac{\alpha_{BA}}{\beta_B} = 0.038$ ) than in the self-excitation cases (still using the average values $\frac{\alpha_{BB}}{\beta_B} = 0.658$ ).

| Pars | $\hat{\mu}_A$ | $\hat{\alpha}_{AA}$ | $\hat{\alpha}_{AB}$ | $\hat{\beta}_A$ | $\hat{\mu}_B$ | $\hat{\alpha}_{BA}$ | $\hat{\alpha}_{BB}$ | $\hat{\beta}_B$ |
|---|---|---|---|---|---|---|---|---|
| August | 77.21 | 52.06 | 2.59 | 61.74 | 61.85 | 1.35 | 42.47 | 49.76 |
| | (0.0368) | (3.1163) | (6.5821) | (2.1562) | (0.0408) | (0.1161) | (2.9649) | (2.1368) |
| September | 57.08 | 44.63 | 2.35 | 52.12 | 48.70 | 2.22 | 36.02 | 42.56 |
| | (0.1167) | (5.3563) | (4.1358) | (3.9182) | (0.1022) | (1.0662) | (5.9839) | (4.8986) |
| October | 61.32 | 49.52 | 1.89 | 56.34 | 56.61 | 2.24 | 46.67 | 53.55 |
| | (0.0398) | (1.8733) | (0.1884) | (1.2671) | (0.0592) | (3.0342) | (1.3961) | (1.0497) |
| November | 70.63 | 53.53 | 1.69 | 63.30 | 42.68 | 1.85 | 43.66 | 49.76 |
| | (0.0919) | (7.8583) | (0.7790) | (5.8662) | (0.549) | (0.5521) | (2.7598) | (1.9111) |

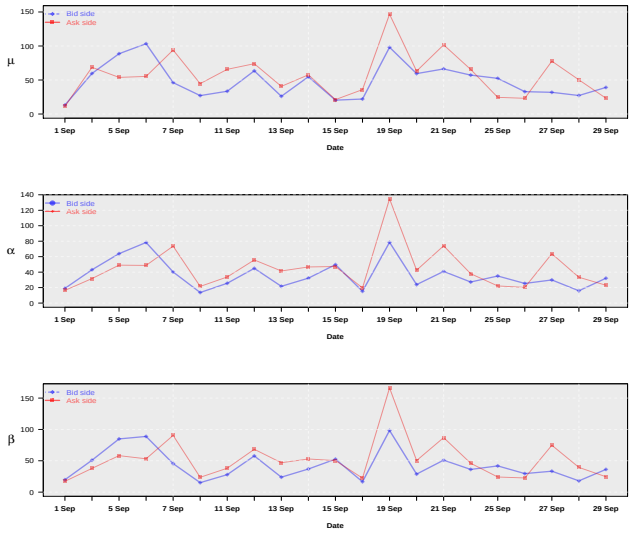
Table 5.8: Statistics summary of bivariate Hawkes' model on H-1398 from Aug to Nov on 2017. Each estimation is the average result computed of a whole month every trading days. Standard deviations are given in parentheses.

Table 5.9 summarizes the statistics on the A-share 601398 estimated values on the ask and bid sides.From the parameters statistics summary, for A-share 601398 the branching ratio $n = \frac{\alpha}{\beta}$, which is equal to the total integrated intensity of an exponential kernel $\int_0^\infty g(\nu)d\nu$, is much weaker in the cross-excitation cases at ask side(taking the average values, $\frac{\alpha_{AB}}{\beta_A} = 0.013$) than in the self-excitation cases (still using the average values $\frac{\alpha_{AA}}{\beta_A} = 0.798$). The same situation with ask side, the cross-excitation cases at ask side (taking the average values,

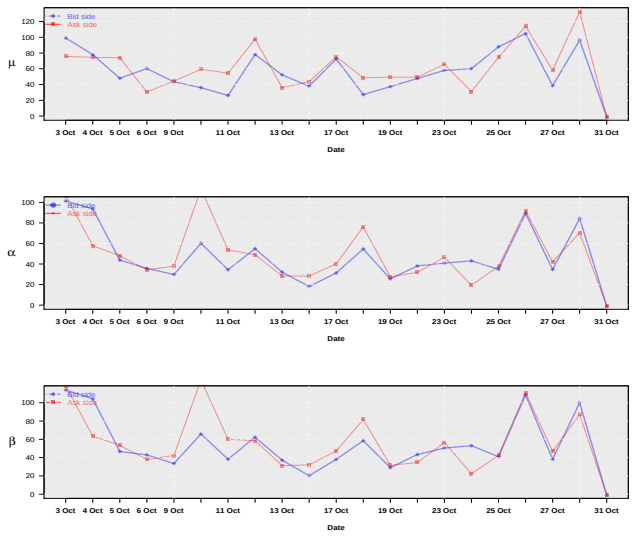
(a)H-1398(Aug 2017)




(b)H-1398(Sep 2017)




(c)H-1398(Oct 2017)




(d)H-1398(Nov 2017)

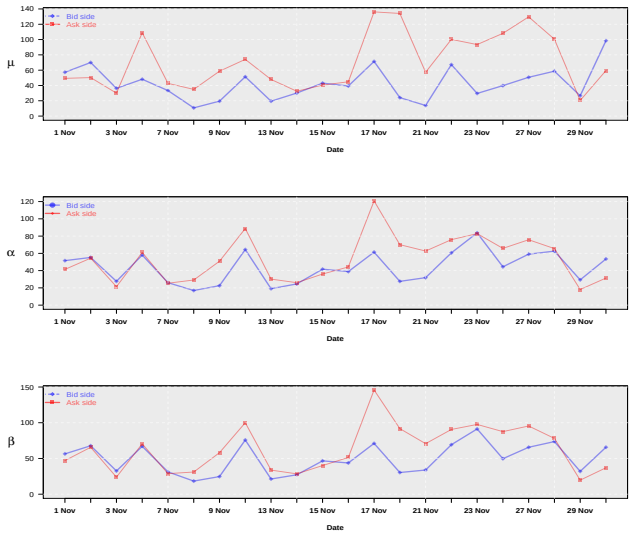
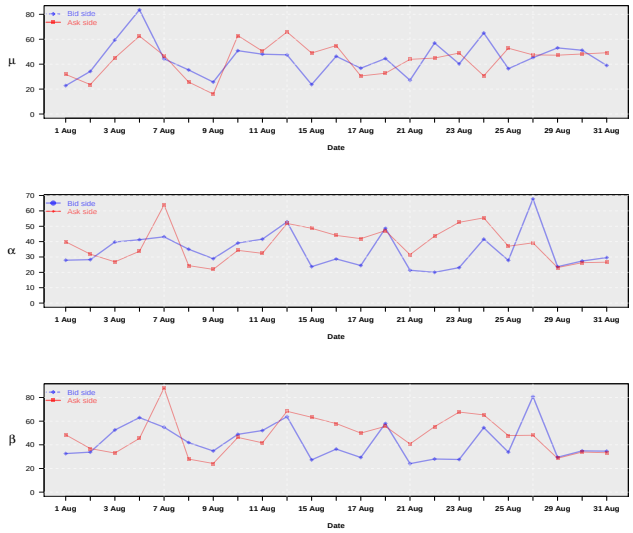
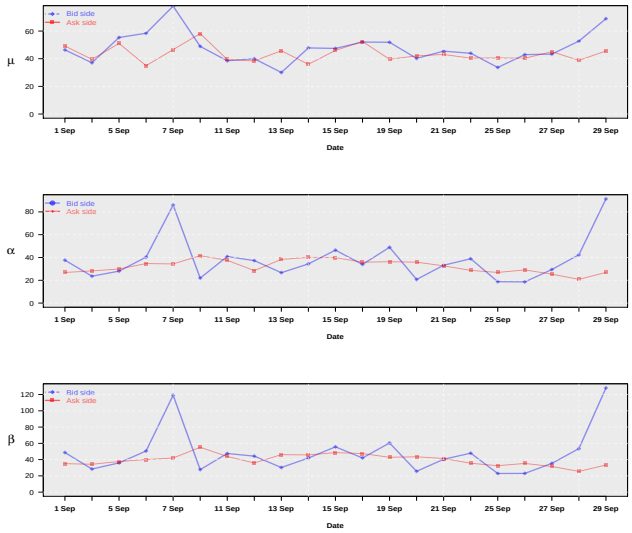



Figure 5.9: Panel (a), (b), (c), (d) are bid-ask side estimation results for H-Shares 1398 from Aug to Nov 2017.

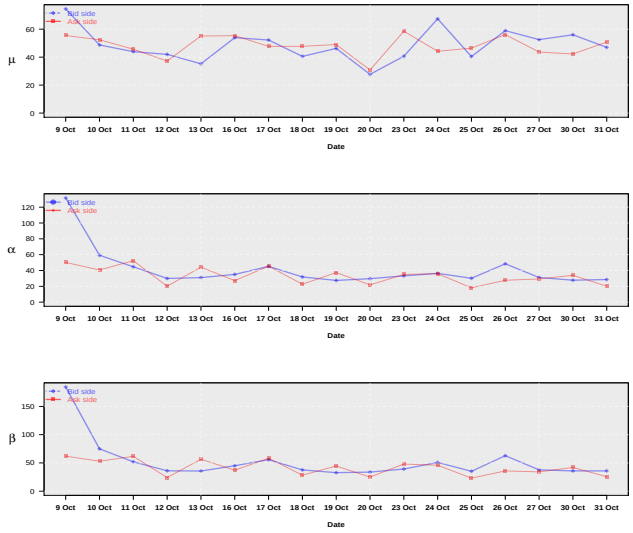
(a)A-601398(Aug 2017)




(b)A-601398(Sep 2017)




(c)A-601398(Oct 2017)




(d)A-601398(Nov 2017)

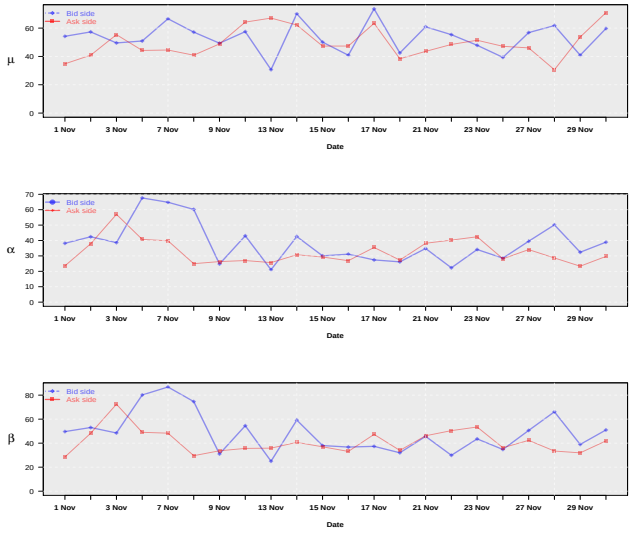



Figure 5.10: Panel (a), (b), (c), (d) are bid-ask side estimation results for H-Shares 601398 from Aug to Nov 2017.

$\frac{\alpha_{BA}}{\beta_B} = 0.039$) than in the self-excitation cases (still using the average values $\frac{\alpha_{BB}}{\beta_B} = 0.793$). From the statistical results of the four groups of H-shares and A-shares, the cross-excitation

| Pars | $\hat{\mu}_A$ | $\hat{\alpha}_{AA}$ | $\hat{\alpha}_{AB}$ | $\hat{\beta}_A$ | $\hat{\mu}_B$ | $\hat{\alpha}_{BA}$ | $\hat{\alpha}_{BB}$ | $\hat{\beta}_B$ |
|---|---|---|---|---|---|---|---|---|
| August | 43.95 | 38.19 | 0.60 | 48.17 | 44.20 | 2.45 | 34.16 | 42.48 |
| | (0.1068) | (4.4017) | (0.3655) | (2.9532) | (0.1077) | (5.5564) | (4.1374) | (3.1328) |
| September | 43.48 | 32.30 | 0.52 | 39.67 | 47.77 | 1.91 | 38.06 | 48.06 |
| | (0.1151) | (4.2154) | (0.0841) | (2.8707) | (0.1376) | (1.1112) | (0.3187) | (4.8986) |
| October | 48.15 | 33.05 | 0.48 | 41.49 | 48.70 | 1.69 | 41.24 | 52.10 |
| | (0.1059) | (7.4632) | (0.0061) | (4.7429) | (0.1300) | (1.3165) | (3.8948) | (2.5979) |
| November | 49.60 | 32.61 | 0.69 | 41.32 | 53.31 | 1.35 | 38.13 | 48.49 |
| | (0.0978) | (2.8367) | (0.4723) | (1.4723) | (0.1087) | (0.5721) | (1.9581) | (1.2131) |

Table 5.9: Statistics summary of bivariate Hawkes' model on A-601398 from Aug to Nov on 2017. Each estimation is the average result computed of a whole month every trading days. Standard deviations are given in parentheses.

terms are obviously weaker than the self-initiated process. The result also confirmed the very limited cross-excitation effect. The values of cross-excitation term $\frac{\alpha_{AB}}{\beta_A}$ and $\frac{\alpha_{BA}}{\beta_B}$ are so tiny that they are negligible, and of course there are no financial implications. Therefore, we can focus on the calibration and use of a simpler model, where bid and ask both side trades are modelled by two one-dimensional Hawkes' processes, with no cross-excitation

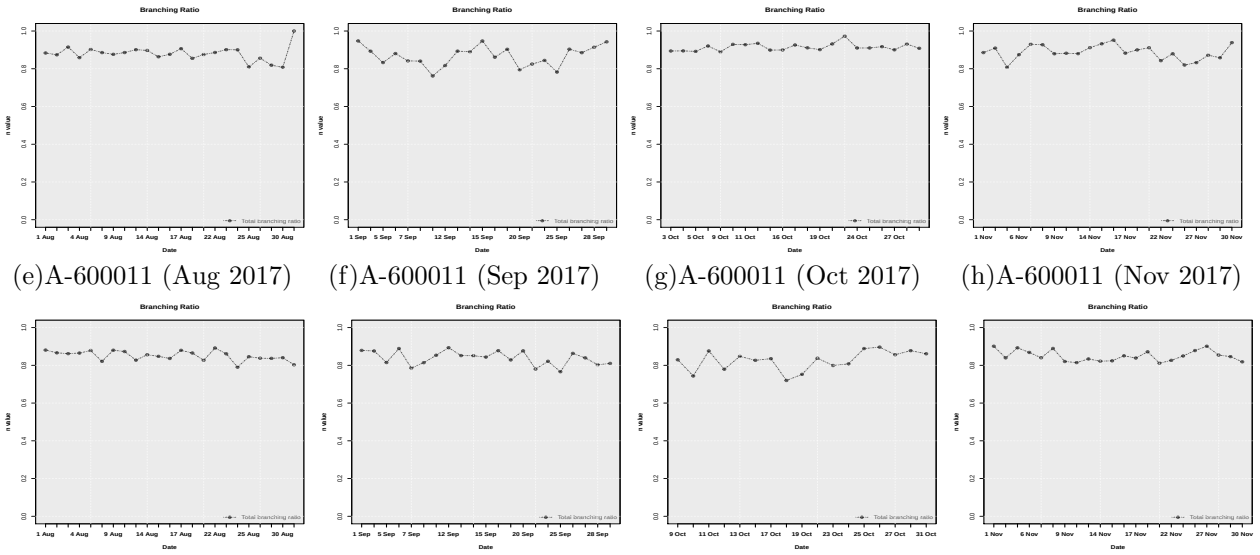
$$\begin{cases} \lambda_A(t) = \mu_A + \sum_{t_i<t} \alpha_{AA} e^{-\beta_A(t-t_i)} \\ \lambda_B(t) = \mu_B + \sum_{t_j<t} \alpha_{BB} e^{-\beta_B(t-t_j)}. \end{cases} \tag{5.1}$$

## 5.4 Branching Ratio

In the framework of Hawkes' process, endogeneity comes from self-excitation, while the baseline activity rate is deemed exogenous. In other words, hawkes' processes provide a straightforward way to measure the importance of endogeneity as inspired by Filimonov and Sornette [28], who considered the level of endogeneity has increased steadily in the last decade. Due to the advent of high-frequency and algorithmic trading developed rapidly, the Hawkes process branching ratio is thought as a tool which could predict flash-crashes on financial markets. In early 2000's, the branching ratio was just around 0.3, and then the branching ratio value was nearly to 0.5 in 2002 and the value was above 0.6 after 2006. It should be noted that the increased in the branching ratio value coincides with the increased

in high frequency trades activities we observed in financial markets. In this way, we would like to detect a closer connection between high frequency traders and the value of branching ratio $n$. Filimonov and Sornette in their research found that the branching ratio has risen quite a bit in recent years. We now turn to explore how branching ratio work with A-H shares related to high frequency trades.

(a)H-0902(Aug 2017) (b)H-0902 (Sep 2017) (c)H-0902 (Oct 2017) (d)H-0902 (Nov 2017)



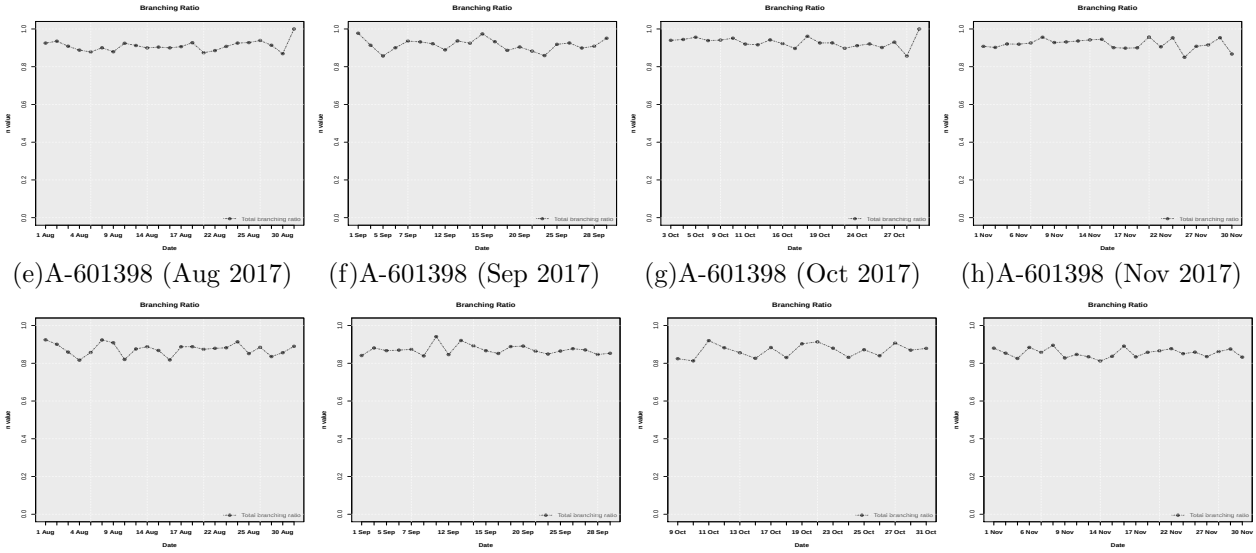

(e)A-600011 (Aug 2017) (f)A-600011 (Sep 2017) (g)A-600011 (Oct 2017) (h)A-600011 (Nov 2017)




Figure 5.11: Panel **(a-d)** show 4 months branching ratio $n$ (Eqn. (3.18)) related to H-0902, and the $n$ value was very high close to 1; and the Panel **(e-h)** shows 4 months branching ratio $n$ (Eqn. (3.18)) related to A-600011. Each Panel shows high n value triggered by exponential kernel.

Evaluating branching ratio n gives a simple measure of the market "distance" criticality. For $n \leq 1$, the process is stationary if $\mu$ is constant. In this case, the branching ratio is also equal to the average proportion of the endogenously generated events among all events. While the total branching ratio oscillates around 0.9 as reported by Fig 5.11 and Fig 5.12. This shows that the endogenous part, which accounts for about 90% of the events, is limited to short time self-reactions in Mainland China and Hong Kong stock markets. This also means that at these time horizons, the reason why it is very hard to fit longer and longer time periods (more than one day), since nothing guarantees the composition of the trader population will be the same for several days in a row. If we interpret branching ratio as a measure of market quality, it makes sense that a low branching ratio reflects a healthier market where price changes are driven by exogenous events and a high branching ratio reflects a less healthier market where price changes are driven by "positive feedback mechanisms" and herding behavior. As derived from one-dimensional Hawkes' process, that $n$ provides a

(a)H-1398 (Aug 2017) (b)H-1398 (Sep 2017) (c)H-1398 (Oct 2017) (d)H-1398 (Nov 2017)




(e)A-601398 (Aug 2017) (f)A-601398 (Sep 2017) (g)A-601398 (Oct 2017) (h)A-601398 (Nov 2017)



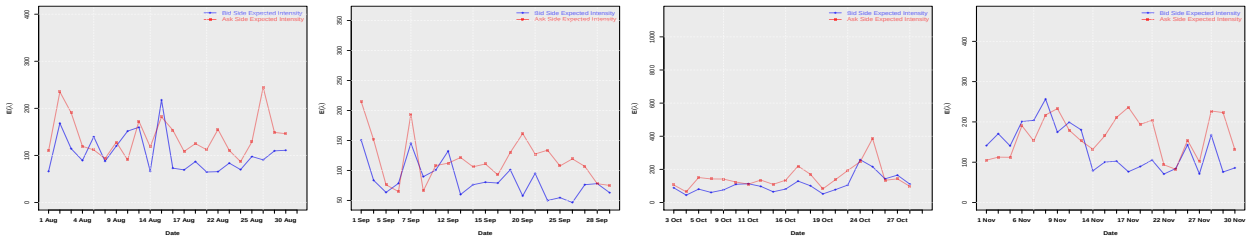

Figure 5.12: Panel **(a-d)** show 4 months branching ratio $n$ (Eqn. (3.18)) related to H-1398, and the $n$ value was very high close to 1; and the Panel **(e-h)** shows 4 months branching ratio $n$ (Eqn. (3.18)) related to A-601398. Each Panel shows high n value triggered by exponential kernel.

direct high quantification of both China and Hong Kong stock market nearly a critical state defined precisely as the limit of diverging trading activity in the absence of any external driving. So the high A-H stock market branching ratio confirms that the number of short-term trades are not only based on exogenous events but are influenced from information from the market itself and its participants and past trades.

### 5.4.1 Expected Intensity

As shown in section 5.3 Equation (5.1)the occurrence of a bid (or respect ask) side has an exciting effect on the stream of buy (or respect sell) orders, conversely the near zero value of $\alpha_{cross}$ tends to indicate that for bivariate Hawkes' processes there is no influence of buy orders or sell orders. Based on these situations, we calculated each pairs of A-H shares expected intensity by both bid and ask side. Result are presented in Fig 5.13 and Fig 5.14.

(a)H-0902(Aug 2017) (b)H-0902 (Sep 2017) (c)H-0902 (Oct 2017) (d)H-0902 (Nov 2017)



(e)A-600011 (Aug 2017) (f)A-600011 (Sep 2017) (g)A-600011 (Oct 2017) (h)A-600011 (Nov 2017)

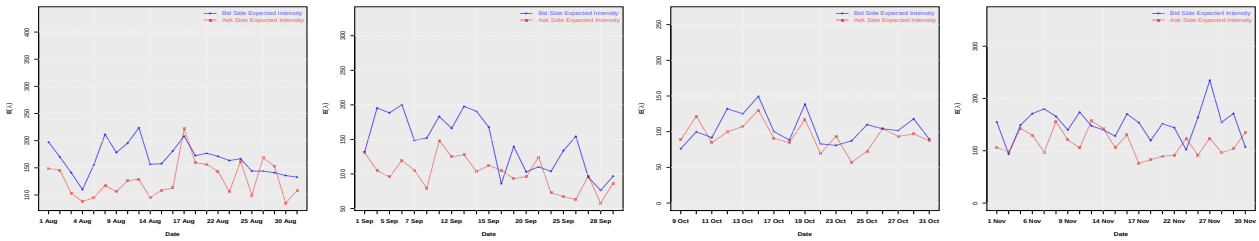
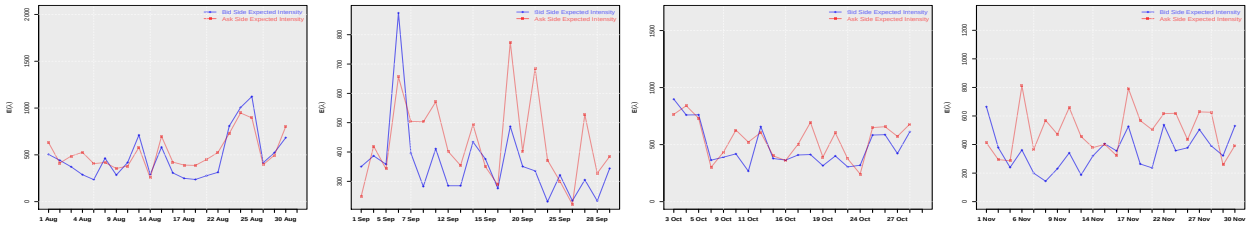


Figure 5.13: Panel **(a-d)** show 4 months bid-ask side expected intensity $E(\lambda_A(t))andE(\lambda_B(t))$ related to H-0902; and the Panel **(e-h)** shows 4 months bid-ask side expected intensity $E(\lambda_A(t))andE(\lambda_B(t))$ related to A-600011.

We note that the ask intensity has spread with the bid intensity, otherwise everybody would buy orders at the best ask and sell them at the best bid, therefore, either H-shares or A-shares there exist arbitrage opportunities, then bid-ask spread is almost attained in quite a short time according to the low latency trading frequency. The bivariate Hawkes' processes based model for ask-bid spread fits the A-H shares data well. In these sense, we believe the Hawkes model itself can be useful tool for stock market liquidity management as well. In the order-driven market, the buying and selling activities excited form prices. Specifically, the bid power is greater than the ask power, and the price shows a rising trend, whereas it shows a downward trend. Simply put, the contrast of the buyer's power can be expressed as the intensity of our description. That is, the buyer's bid and ask intensity trades within the unit time. Based on the observed historical information, when the bid side intensity is

(a)H-1398 (Aug 2017) (b)H-1398 (Sep 2017) (c)H-1398 (Oct 2017) (d)H-1398 (Nov 2017)



(e)A-601398 (Aug 2017) (f)A-601398 (Sep 2017) (g)A-601398 (Oct 2017) (h)A-601398 (Nov 2017)

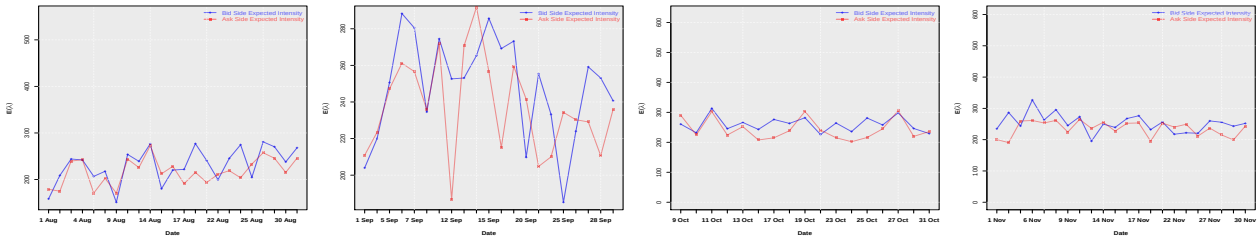


Figure 5.14: Panel **(a-d)** show 4 months bid-ask side expected intensity $E(\lambda_A(t))andE(\lambda_B(t))$ related to H-1398; and the Panel **(e-h)** shows 4 months bid-ask side expected intensity $E(\lambda_A(t))andE(\lambda_B(t))$ related to A-601398.

greater than the ask side intensity, it is bought, and when the ask side intensity is greater than the bid side intensity, it is sold. This is the basic idea based on the buying and selling intensity trading strategies.

## 5.5 Trading Strategy

A famous investing strategy of how to make money in the stock market is:"Buy low and sell high." A principle often sounded for good strategy relying on one's success in the ability to spot the right opportune moment to buy and sell well. What is more surprising than ever is the increasing growth in computing capability and a great deal of activity in the A-H stocks market at trying to profit at buying low and selling high. The best buy (sell) order with the lowest (highest) price is better known as the best bid (best ask). Changes in the price distributions today (see Cvitanic and Kirilenko [16]) and the observed activity in the order book, is coupled, as it can be expected with the advent of the bid and ask intensity of trading conducted in the modern financial market. In this section we develop the goal that the two-dimensional Hawkes model bid and ask intensity is a developing trade strategy where we expect to profit as a result of noticing a model in the bid and ask roundtrip trades in an order book.

There are many studied in why trades are initiated, see for example Sarkar and Schwartz [52]. Sometime because of asymmetric information, differing information or a difference of opinion or even an increased ratio of impatient traders. In the same way, trade clustering results from a number of market events as seen by Cartea and Jaimungal [14]. Some examples which increase market activity are due to the release of public or private data generating an awareness of a fundamental shock to the value of an asset or value of one market that generates responses. There could be a view that as information is released it impounds the stock price. A recent increase in the volume of orders indicates fast traders dominate and it is very difficult to link the activity in the electronic markets with the arrival of news or any other classic ways of explaining motives for trade. There are ultra fast algorithms that make decisions about trades in less than milliseconds, at this speed it is difficult to link with the arrival of private or public information, the types of trade, the liquidity shock or any such market event. So at this point we develop a model where we intent to model in a reduced-form the arrival of market buy or sell order. Here the types chosen are influential orders that excite the market and induce other market traders to submit their orders. For example, the arrival of one influential market sell order often increases the likelihood of observing another market sell, then over the next time steps also (but to a slightly lesser extent) increases likelihood of a market buy order occurring. Thus we model the arrival of market orders and capture trade clustering events, it makes traders feeling they have got positive feedback from market.

### 5.5.1 Confusion Matrix

We can use the confusion matrix to examine the accuracy between real result and its expected intensity. In a confusion matrix information is contained about the actual and predicted result as predicted by the classification system. The performance of a system is often evaluated by data in the matrix. The following table shows the confusion matrix for a two class classifier (see Kohavi and Provost [49]).

With a given classifier instance there are four possible outcomes to the matrix. When a positive instance is correctly classified it is counted as a true positive, if an instance is negative and it counted as false it is a false negative, if it is falsely classified as positive, it is a false positive, and if it is correctly classified as negative then it is a true negative.

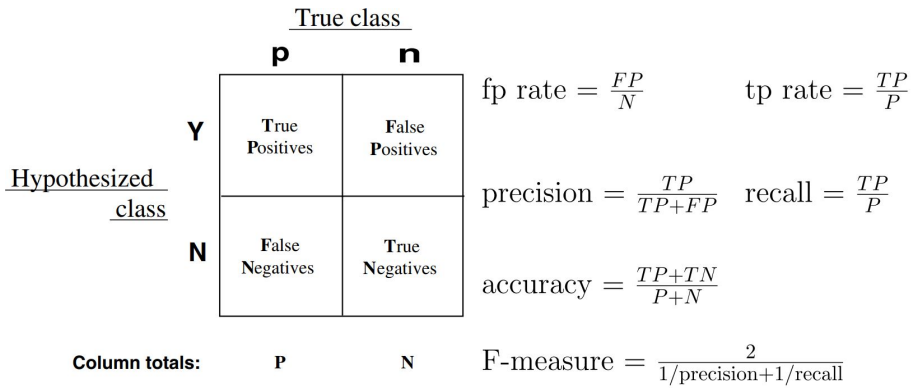



Figure 5.15: Confusion matrix and common performance metrics.

Given an instance set and a classifier the test set, (a two-by-two matrix) is known as a contingency table and can be constructed by four dispositions of instances. The Fig 5.15, shows a confusion matrix with equations of several common metrics that can be calculated by it. Terminology and derivations from a confusion matrix are shown

- P(condition positive) is the number of real positive cases in the data.
- N(condition negative) is the number of real negative cases in the data.
- TP(true positive) is equivalent with hit.
- TN(true negative) is equivalent with correct rejection.
- FP(false positive) is equivalent with false alarm.
- FN (false negative) is equivalent with with miss.
- TPR(true positive rate or hit rate), $TPR = \frac{\sum Truepositive}{\sum Conditionpositive} = \frac{TP}{P}$.
- TNR(true negative rate or specificity), $TNR = \frac{\sum Truenegative}{\sum Conditionnegative} = \frac{TN}{N}$.
- PPV (positive predictive value or precision), $PPV = \frac{TP}{TP+FP}$.
- NPV (negative predictive value), $NPV = \frac{TN}{TN+FN}$.
- ACC(accuracy),$ACC = \frac{\sum Truepositive+\sum Truenegative}{\sum Totalpopulation} = \frac{TP+TN}{P+N}$.

There are a number diagonals which represent the correct decision made and errors or confusions between the various classes. We make specific assumptions that when predicting a

single condition it is either positive or negative (dichotomous). For simplicity we assume that contingency is non-trivial in that both negative and positive states predict a real condition and no marginal sums or probabilities are ever zero.

Based to our model, we assume that whenever an expected bid intensity is greater than an expected ask intensity, then a price will increase, otherwise whenever an expected bid intensity is lower than an expected ask intensity, then it will result the price fall in a stock. Based on this assumption, if we test two pairs of A-H shares bid-ask expected intensitywhich calculated by using the Hawkes model, then we compare with the real stock trading trend. We should highlight that the ACC indicator is the accuracy of our trading strategy, because the investors willing to see profit from the transaction, whenever they execute an optimal trading strategy. We can see the result in Table 5.10, the ACC value on Huaneng Power H-0902 and A–600011 are 0.7925 and 0.5301 respectively. For another stock pair ICBC, the ACC value on H-1398 and A-601398 are 0.7411 and 0.5180 respectively. We notice that the accuracy of prediction on H-shares performs a little bit better than A-shares. Through confusion matrix comparison, we conclude that the forecast performance of expected intensity of H-shares in Hong Kong market is obviously better than that of A-shares in the domestic market, which also indicates that the information in the domestic market is not symmetrical and belongs to the market control rights, while the retail investors trading in the market with little information occupies the dominant position. What's more interesting is that, the small-scale company but high proportion price ratio Huaneng Power's expected intensity prediction accuracy for both H-shares and A-shares is significantly better than the large-scale company ICBC with low proportion price ratio.

| Measure | H-0902 | A-600011 | H-1398 | A-601398 |
|---|---|---|---|---|
| TPR | 0.6389 | 0.4412 | 0.6500 | 0.6750 |
| TNR | 0.8367 | 0.5918 | 0.8222 | 0.3721 |
| PPV | 0.7419 | 0.4286 | 0.7647 | 0.5000 |
| NPV | 0.7593 | 0.6042 | 0.7255 | 0.5517 |
| ACC | 0.7529 | 0.5301 | 0.7411 | 0.5180 |

Table 5.10: Table of the confusion matrix, comparing two companies of A-H share performance.

# Chapter 6

# Summary and Future Works

This chapter summarizes the major findings of this thesis and suggests some directions for further research.

## 6.1 Summary

In this thesis, an overview has been carried out on application of Hawkes' processes in finance. As Hawkes' process builds a specific set of bivariate point process, it has become a popular topic in empirical high frequency finance. We have discussed applications in many different problems and highlighted its great flexibility and versatility, so that it could be successfully involved in a variety of environment with volatilities. In financial markets, this diversity makes the Hawkes process be a good tool accounting for capturing the dynamics of the order book.

Our study shows that the Hawkes processes can be applied to model transaction data and be used to measure the liquidities of stock. We describe the procedure for simulation and derive maximum likelihood function for parameters estimation both on one dimensional and two dimensional Hawkes' processes. We introduce the expected intensity, branching ratio, and trading strategy to assess the performance of Hawkes model. One contribution in the thesis is to show that the branching ratio (see Chapter 5), can account for more than 80% of the endogenous self-exciting events in the A-H stock markets. Especially by comparing with

ACD model, a Hawkes model could be considered to be more practical and likely perform better at determining the expected intensity number of offspring events in financial markets.

Another contribution of this thesis is to extend the Hawkes process on modelling bid-ask expected intensity separately, where we propose a new trading strategy based on bid-ask spread and use the confusion matrix to study the performance. We test two pairs of A-H shares bid-ask intensity, and we find the trading strategy would be more than 70% accuracy rate on measure A-H stock performance, so we result that H-shares performs better than A-shares for both two case.

## 6.2 Future works

We know the kernel of Hawkes model used in this thesis is a fixed exponential form. It assumes that the impact of historical events on future events will decrease exponentially. This assumption simplifies the parameter estimation process but also add more restrictions on the model. Hence, a future work could be to investigate other kernel shapes and compare performance. In other words, changing the magnitude of the parameters $\beta$, could lead to modelling the difference in the frequency of data.

Finally, we can further investigate the time interval based on two-dimensional Hawkes' model to improve the performance of the trading strategy. In the order modelling, this thesis simply divides the order into two categories: bid order and ask order. However, there are many types of orders in practice, such as market orders, limit orders, stop loss orders, limit stop orders, and so on. Orders for different types will also perform differently for executions. Based on this, orders can be further categorized in future studies to allow the study of multi-dimensional Hawkes' processes.

# Bibliography


[1] Yacine Aït-Sahalia, Julio Cacho-Diaz, and Roger JA Laeven. Modeling financial contagion using mutually exciting jump processes. Technical report, National Bureau of Economic Research, 2010.

[2] Yacine Aït-Sahalia, Jianqing Fan, and Dacheng Xiu. High-frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association*, 105(492):1504–1517, 2010.

[3] Yacine Aït-Sahalia and Jean Jacod. *High-frequency financial econometrics*. Princeton University Press, 2014.

[4] Yakov Amihud and Haim Mendelson. Asset pricing and the bid-ask spread. *Journal of financial Economics*, 17(2):223–249, 1986.

[5] Yakov Amihud, Haim Mendelson, and Lasse Heje Pedersen. *Market liquidity: asset pricing, risk, and crises*. Cambridge University Press, 2012.

[6] Emmanuel Bacry and Jean-François Muzy. Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 14(7):1147–1166, 2014.

[7] Luc Bauwens and Nikolaus Hautsch. Modelling financial high frequency data using point processes. In *Handbook of financial time series*, pages 953–979. Springer, 2009.

[8] Giacomo Bormetti, Lucio Maria Calcagnile, Michele Treccani, Fulvio Corsi, Stefano Marmi, and Fabrizio Lillo. Modelling systemic cojumps with hawkes factor models. *arXiv preprint arXiv:1301.6141*, 2013.

[9] Clive G Bowsher. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141(2):876–912, 2007.

[10] Pierre Brémaud and Laurent Massoulié. Hawkes branching point processes without ancestors. *Journal of applied probability*, 38(1):122–135, 2001.

[11] Michael J Brennan, Tarun Chordia, and Avanidhar Subrahmanyam. Alternative factor specifications, security characteristics, and the cross-section of expected stock returns1. *Journal of Financial Economics*, 49(3):345–373, 1998.

[12] Michael J Brennan and Avanidhar Subrahmanyam. Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *Journal of financial economics*, 41(3):441–464, 1996.

[13] Jonathan Brogaard et al. High frequency trading and its impact on market quality. *Northwestern University Kellogg School of Management Working Paper*, 66, 2010.

[14] Alvaro Cartea, Sebastian Jaimungal, and Jason Ricci. Buy low, sell high: A high frequency trading perspective. *SIAM Journal on Financial Mathematics*, 5(1):415–444, 2014.

[15] Valérie Chavez-Demoulin and JA McGill. High-frequency financial data modeling using hawkes processes. *Journal of Banking & Finance*, 36(12):3415–3426, 2012.

[16] Jaksa Cvitanic and Andrei Kirilenko. High frequency traders and asset prices. 2010.

[17] Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media, 2007.

[18] DJ Daley and D Vere-Jones. Basic properties of the poisson process. *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*, pages 19–40, 2003.

[19] Vinay T Datar, Narayan Y Naik, and Robert Radcliffe. Liquidity and stock returns: An alternative test. *Journal of Financial Markets*, 1(2):203–219, 1998.

[20] Venkat R Eleswarapu. Cost of transacting and expected returns in the nasdaq market. *The Journal of Finance*, 52(5):2113–2127, 1997.

[21] Venkat R Eleswarapu and Marc R Reinganum. The seasonal behavior of the liquidity premium in asset pricing. *Journal of Financial Economics*, 34(3):373–386, 1993.

[22] Robert F Engle and Jeffrey R Russell. Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model. *Journal of empirical finance*, 4(2-3):187–212, 1997.

[23] Robert F Engle and Jeffrey R Russell. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, pages 1127–1162, 1998.

[24] Eymen Errais, Kay Giesecke, and Lisa R Goldberg. Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, 1(1):642–665, 2010.

[25] Eugene F Fama and Kenneth R French. The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465, 1992.

[26] Eugene F Fama and Kenneth R French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.

[27] Eugene F Fama and James D MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636, 1973.

[28] Vladimir Filimonov and Didier Sornette. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E*, 85(5):056108, 2012.

[29] Vladimir Filimonov, Spencer Wheatley, and Didier Sornette. Effective measure of endogeneity for the autoregressive conditional duration point processes via mapping to the self-excited hawkes process. *Communications in Nonlinear Science and Numerical Simulation*, 22(1-3):23–37, 2015.

[30] Ryan Garvey and Fei Wu. Intraday time and order execution quality dimensions. *Journal of Financial Markets*, 12(2):203–228, 2009.

[31] Ruslan Y Goyenko, Craig W Holden, and Charles A Trzcinka. Do liquidity measures measure liquidity? *Journal of financial Economics*, 92(2):153–181, 2009.

[32] Joachim Grammig and Kai-Oliver Maurer. Non-monotonic hazard functions and the autoregressive conditional duration model. *The Econometrics Journal*, 3(1):16–38, 2000.

[33] Oliver Grothe, Volodymyr Korniichuk, and Hans Manner. Modeling multivariate extreme events using self-exciting point processes. *Journal of Econometrics*, 182(2):269–289, 2014.

[34] Valérie Gyselinck, Chiara Meneghetti, Monica Bormetti, Eric Orriols, Pascale Piolino, and Rossana De Beni. Considering spatial ability in virtual route learning in early aging. *Cognitive processing*, 14(3):309–316, 2013.

[35] Joel Hasbrouck. The summary informativeness of stock trades: An econometric analysis. *The Review of Financial Studies*, 4(3):571–595, 1991.

[36] Joel Hasbrouck and Robert A Schwartz. Liquidity and execution costs in equity markets. *The journal of portfolio management*, 14(3):10–16, 1988.

[37] Alan G Hawkes. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 438–443, 1971.

[38] Alan G Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[39] Alan G Hawkes and David Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503, 1974.

[40] Terrence Hendershott, Charles M Jones, and Albert J Menkveld. Does algorithmic trading improve liquidity? *The Journal of Finance*, 66(1):1–33, 2011.

[41] Patrick Hewlett. Clustering of order arrivals, price impact and trade path optimisation. In *Workshop on Financial Modeling with Jump processes, Ecole Polytechnique*, pages 6–8, 2006.

[42] Peter A Lewis and Gerald S Shedler. Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics (NRL)*, 26(3):403–413, 1979.

[43] Steven A Lippman and John J McCall. An operational measure of liquidity. *The American Economic Review*, 76(1):43–55, 1986.

[44] Marcel N Massimb and Bruce D Phelps. Electronic trading, market structure and liquidity. *Financial Analysts Journal*, 50(1):39–50, 1994.

[45] Yoshiko Ogata. The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 30(1):243–261, 1978.

[46] Yosihiko Ogata. On lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1):23–31, 1981.

[47] Yosihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association*, 83(401):9–27, 1988.

[48] Tohru Ozaki. Maximum likelihood estimation of hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, 31(1):145–155, 1979.

[49] Foster J Provost, Tom Fawcett, Ron Kohavi, et al. The case against accuracy estimation for comparing induction algorithms. In *ICML*, volume 98, pages 445–453, 1998.

[50] Richard Roll. A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of finance*, 39(4):1127–1139, 1984.

[51] Jeffrey R Russell. Econometric modeling of multivariate irregularly-spaced high-frequency data. *Manuscript, GSB, University of Chicago*, 1999.

[52] Asani Sarkar and Robert A Schwartz. Market sidedness: Insights into motives for trade initiation. *The Journal of Finance*, 64(1):375–423, 2009.

[53] James P Tam, Yi-An Lu, Jin-Long Yang, and Kou-Wei Chiu. An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides. *Proceedings of the National Academy of Sciences*, 96(16):8913–8918, 1999.

[54] Ioane Muni Toke and Fabrizio Pomponio. Modelling trades-through in a limit order book using hawkes processes. *Economics: The Open-Access, Open-Assessment E-Journal*, pages vol–6, 2012.

[55] William N Venables and Brian D Ripley. Tree-based methods. In *Modern Applied Statistics with S*, pages 251–269. Springer, 2002.

[56] Frank Zhang. The effect of high-frequency trading on stock volatility and price discovery. *SSRN eLibrary*, 2010.