UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI "M.FANNO"

DIPARTIMENTO DI SCIENZE STATISTICHE

CORSO DI LAUREA MAGISTRALE IN

ECONOMICS AND FINANCE – ECONOMIA E FINANZA

TESI DI LAUREA

"INTRODUCTION TO HAWKES PROCESS AND AN APPLICATION WITH FINANCIAL DATA"

RELATORE:

CH.MO PROF. MASSIMILIANO CAPORIN

LAUREANDO/A: YIWEI CAO

MATRICOLA N. 1106069

ANNO ACCADEMICO 2015-2016

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.
Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

Firma dello studente

_____

Abstract

This thesis provides an introduction to self-exciting stochastic process called Hawkes process. The basic analytic framework will be illustrated and simulations and estimations will be carried out by using Matlab. An example of univariate unmarked Hawkes process will be implemented to model the occurrence of price jumps for some stocks.

Acknowledgements

I would like to express my gratitude to my supervisor Prof. M. Caporin for stimulating research interest and for providing continuously guidance and support.

A special thank to my parents who made life changing decisions in order to support unconditionally my education.

Finally, I would like to express appreciation and thank to my girlfriend Yiwei and my friends for the continuous encouragement throughout last months. A particular thank to Francesca, who is at the same time friend and mentor, for providing advice and encouragement during all years of university.

Contents

Chapter 1

Introduction

This thesis concerned with theoretical and empirical application of self-exciting stochastic process called Hawkes process. In particular, our analysis will be focused on the special case of univariate unmarked Hawkes process.

The application of self-exciting stochastic models have been shown to be effective to represent occurrences of earthquakes or dynamics of financial assets. This work aims to contribute to the growing literature of applications of Hawkes process in finance. Our main object is to investigate its application in modelling price jumps in the stock market and to provide the theoretical background for the analysis.

The underlying theory will be illustrated in Chapter 2 starting from an introduction to point process and counting measure. The second part of the chapter will provide an analytical background of Hawkes process and will give a representation based both on the intensity function and on Poisson cluster process. Then, we will illustrate a simulation method for Hawkes process based on Ogata's modified thinning algorithm.

The estimation of parameters will be subject of Chapter 3. We will illustrate three approaches for the estimation: direct numerical maximization (DNM) of log-likelihood function; the exact expectation-maximization (EM) algorithm; and an approximate EM algorithm. It has been showed that the first two methods produce the same estimates, hence for the practical applications of our analysis only DNM method and approximate EM algorithm will be used.

We will use Hawkes process to model the occurrences of price jumps for five stocks in Chapter 4. The first part of the chapter will provide an analytical background of stochastic process with jumps and will illustrate the jump detection method proposed by Corsi, Pirino, and Reno (2010) based on the separation of the continuous component of stochastic process from

the discrete one. The result of our analysis will be presented in the second part of the chapter. Even if for only 3 of 5 stocks considered seems legitimate the use of univariate unmarked Hawkes model, we still conclude that Hawkes processes can be an useful tool in modelling financial assets.

The simulations and the estimations are carried out by using Matlab. The most relevant codes used in this work are collected in the Appendix.

Chapter 2

Hawkes process: Analytical framework

Point process with clustering property were introduced by Hawkes (1971a) and Hawkes (1971b), now they are commonly referred to as the Hawkes processes. In the recent years, Hawkes process has become very popular in many field such as finance (Bacry et al., 2013), seismology (Ogata, 1988) and biology (Reynaud-Bouret, Schbath, et al., 2010). This chapter will illustrate the analytical framework of Hawkes process: the first section will introduce point process in general in order to provide the basis for the later analysis, the second section will specifically deal with Hawkes process and its properties.

2.1 Introduction to point process

Point process in general describes the random scattering of points in a certain space (usually a subset of $\mathbb{R}^d$) or in a certain interval of time, in this thesis the specific points are associated with the time at which certain random events occurs, so only temporal point process will be considered. A simple point process on the non negative time line is a stochastic phenomena whose realisations consist of the times $t_1, t_2, ...$ of events scattered along the time. An example of the realisation of such a stochastic process is illustrated in 2.1 where t_i is the time at which i-th event of interest occurred. A simple point process could be used for instance to record the times of earthquakes or the arrival time of orders for a stock.

A marked point process is obtained if at each occurrence time a mark of the event is also registered, in this case the realisation of the stochastic process is formed by elements $(t_1, m_1), (t_2, m_2), ...$ The marks could be used to

record the magnitude of an event, for instance the magnitude of an earthquakes or the quantity of an order in the stock market, or it could be used to indicate the type of an event, for example to distinguish the sell orders from the buy orders.

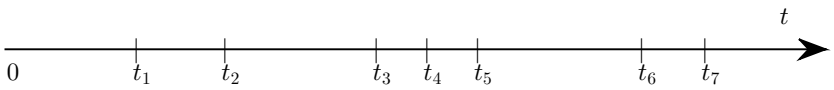

FIGURE 2.1: Realisation of a simple point process, where t_i is the occurrence time of an event of interest.

2.1.1 Counting measure of simple point process

There are many different ways to define mathematically simple point process as a sequence of random variables in a given probability space, for example the Definition 2.1.1 in Jacobsen (2006) or the definition in page 6 of Reiss (2012). Here, we will give a representation of simple point process based mainly on Chapter 3.1 of Daley and Vere-Jones (2002) which focuses on the counting measure of the process.

Let A denote any subset of non negative real line and let $N(A)$ denote the number of occurrences of the process in the set A; More precisely.

$$\begin{aligned} N(A) &= \text{number of indices } i \text{ for which } t_i \text{ lies in } A \\ &= \#\{i : t_i \in A\} \end{aligned} \tag{2.1}$$

t_i represents the occurrence time of an event of interest with i in some suitable index set. In our case we will make the logical assumption that $i \in \mathbb{N}$. Then, if $N(A)$ is finite for bounded sets A, it follows that $N(A)$ is a non-negative integer and this characteristic distinguishes it from other more general non-negative measures as a *counting* measure.

If A satisfies the consistency requirement, i.e. A can be expressed as the union of the disjoint sets $A_1, A_2, ..., A_k$, namely,

$$A = \bigcup_{i=1}^{k} A_i, \quad \text{where } A_i \cap A_j = \emptyset \text{ for } i \neq j,$$

it follows from 2.1 that

$$N(A) = N\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} N(A_i).$$

If A is the half-open interval $(a, b]$ with $0 < a < b$, we will use the abbreviation $N(a, b]$ for the set function $N((a, b])$. In a similar way, we will use the contraction $N(t) = N(0, t] = N((0, t])$ and $N(\mathrm{d}t) = N(t, t + \mathrm{d}t]$ unless otherwise stated.

$N(t)$ is a $\mathbb{R}_+ \to \mathbb{N}$ step function with $N(0) = 0$. Note that if $0 < s < t$ then $0 \leq N(s) \leq N(t)$ is true. So $N(t)$ is a non-decreasing, right-continuous, integer-valued and non-negative function.

2.1.2 Stationarity and orderliness

The idea of stationarity for point process is essentially the same for other stochastic process. Considering the counting measure $N(A)$ and using the Definition 3.2.I of Daley and Vere-Jones (2002), a point process is stationary when for every $r = 1, 2, ...$ and all bounded subsets $A_1, ..., A_r$ of the domain, the joint distribution of

$$\{N(A_1 + t), N(A_2 + t), ..., N(A_r + t)\}$$

does not depend on t. This definition focuses on the structure of process which is invariant with respect to the translation of the time axis.

Weakly stationarity could be defined in a similar way. A point process is weakly stationary if for any set A of the domain, the mean and the variance of the distribution of $N(A)$ are invariant under translations of t, that is

$$\mathbb{E}[N(A)] = \mathbb{E}[N(A + t)] \quad \text{and} \quad Var(N(A)) = Var(N(A + t)).$$

The stationarity of point process could be also defined based on the stationarity of intervals between two occurrence time, for example in Definition 3.2.II. of Daley and Vere-Jones (2002).

Following Cox and Isham (1980), a process is orderly if it satisfies the condition that, for $h \to 0^+$,

$$\Pr\{N(t, t + h] > 1\} = o(h) \quad t \in \mathbb{R}.$$

Whereas a process is said to have no multiple simultaneous occurrence if

$$\Pr\{N([t]) > 1 \quad \text{for some } t \in \mathbb{R}\} = 0,$$

Where the set t consists of the singleton $t \in \mathbb{R}$. [1] The orderliness of a process can be shown to imply no multiple simultaneous occurrence (Daley, 1974).

In the following work, we will always assume property of orderliness to be true. Whereas it is unnecessary to assume the stationarity of processes. The counting measure $N(a, b]$ can be written now as

$$\begin{aligned} N(a, b] &= \int_{(a,b]} N(\mathrm{d}s) \\ &= \sum_{j:t_j \in (a,b]} 1 \end{aligned}$$

where $N(\mathrm{d}s) = 1$ when there is an event of interest in $(s, s+\mathrm{d}s]$ and $N(\mathrm{d}s) = 0$ otherwise. The orderliness property excludes the possibility of multiple occurrence in an infinitesimal interval $\mathrm{d}s$. More generally,

$$\int_{(a,b]} g(s)N(\mathrm{d}s) = \sum_{j:t_j \in (a,b]} g(t_j)$$

where $g(t)$ is a $\mathbb{R}_+ \to \mathbb{R}$ function.

2.1.3 Conditional intensity function

The behaviour of a point process is typically described by its conditional intensity λ. Intuitively we can imagine λ as the probability with which events are expected to occur around a specific point in time, conditioned on the prior history of the point process.

The conditional intensity associated with a point process N could be defined as

$$\lambda(t|\mathcal{F}_t) = \lim_{h \to 0^+} \frac{\Pr\{N(t, t+h] > 0|\mathcal{F}_t\}}{h} \tag{2.2}$$

where the filtration $\mathcal{F}_t$ is the entire history of the point process up to time t.

Another way to define λ is by limiting the conditional expectation

$$\lambda(t|\mathcal{F}_t) = \lim_{h \to 0^+} \frac{\mathbb{E}\{N(t, t+h]|\mathcal{F}_t\}}{h} \tag{2.3}$$

[1] Note in this case the notation $N([t])$ does not mean $N((0, t])$, but $N([t, t])$.

Intuitively, the definition through (2.3) provides the overall intensity of the process including multiplicities, whereas the definition through (2.2) gives the intensity of the process of instants at which points occur, without considering their multiplicities. These two definitions are equal if a process has no multiple simultaneous occurrence (Leadbetter, 1968). The orderliness of point process will be always assumed in this work and the definition through (2.2) will be preferred. A rigorous definition of conditional intensity could be find e.g. in the Chapter 7.2 of (Daley and Vere-Jones, 2002).

2.1.4 An example of point process: Poisson process

The archetypal and the simplest point process is the homogeneous Poisson process. The intensity λ of this process is defined by the requirements that for all t, as $h \to 0^+$,

$$\begin{cases} \Pr\{N(t, t+h] = 1 | \mathcal{F}_t\} = \lambda h + o(h) \\ \Pr\{N(t, t+h] > 1 | \mathcal{F}_t\} = o(h) \\ \Pr\{N(t, t+h] = 0 | \mathcal{F}_t\} = 1 - \lambda h + o(h) \end{cases} \tag{2.4}$$

The essential element in (2.4) is that λ is constant, so it does not depend on t and on the history of the process $\mathcal{F}_t$. In particular, the probability of an occurrence in the interval $(t, t+h]$ is independent on whether there have been relatively few or relatively many points before t, or if there is an occurrence exactly at t. Note that the second equation in (2.4) implies no multiple simultaneous occurrence. An example of realisation of a homogeneous Poisson with $\lambda = 0.2$ is represented in Figure 2.2. The counting measure $N(t)$ is represented on the vertical axis, whereas the horizontal axis gives times of occurrence.

There are situations in which the intensity depends on t, for instance in order to incorporate the time trend or the cyclical fluctuation in the intensity of occurrence. A Poisson process with this characteristic is called non-homogeneous Poisson process. Compared to the homogeneous case, the intensity λ in (2.4) is replace by $\lambda(t)$. Note that in both homogeneous and non-homogeneous Poisson process the intensity does not depend on the history of the process $\mathcal{F}_t$.

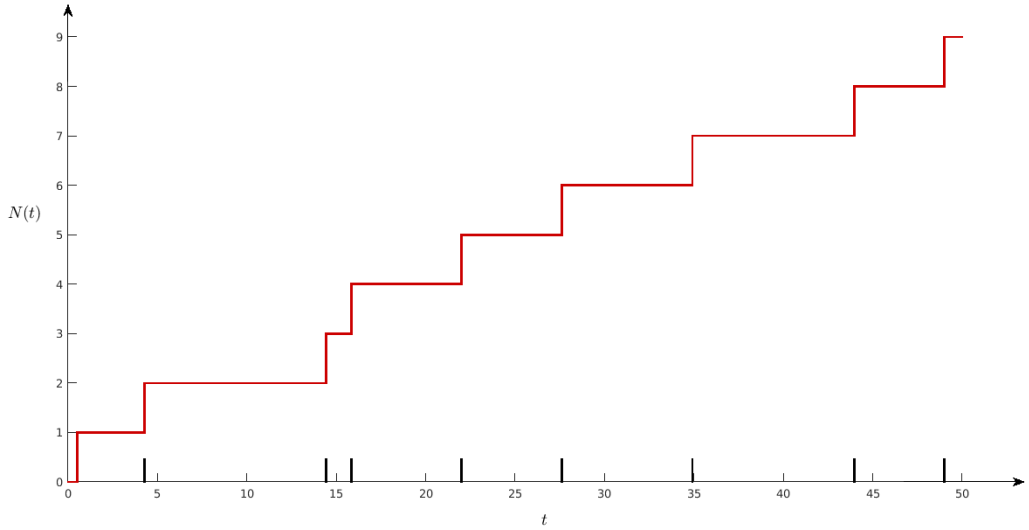

FIGURE 2.2: A simulated realisation of a homogeneous Poisson process in the interval $(0, 50]$ with $\lambda = 0.2$. The vertical lines on the horizontal axis indicate times of occurrence and the vertical axis indicates the counting measure.

2.1.5 Marked point process

In many cases, a point process is not itself the main object of the study, but it contributes as a component of a more complex stochastic model. A temporal marked point process (MPP) is a point process in $\mathbb{R}^+ \times \mathcal{M}$ space, where every realisation t_i is associated with a mark $m_i \in \mathcal{M}$ (For a rigorous definition see e.g. Chapter 2.1 of Jacobsen (2006) or Chapter 1.4 of Karr (1991)). The temporal point process associated with a MPP is referred to as the ground process and it is denoted by notation N_g, whereas the whole MPP is indicated by notation N. A temporal MPP N, with point occurrences in $\mathbb{R}^+$ and marks in $\mathcal{M}$, is a point process $\{(t_i, m_i) : i \in \mathbb{N}\}$ on $\mathbb{R}^+ \times \mathcal{M}$ with the additional property that the ground process N_g associated to it is itself a temporal point process on $\mathbb{R}^+$ (Lapham, 2014). A MPP could be used in many applications due to the great variety of forms that can be taken by the marks and the variety of dependence relations that can exist between the marks themselves and the ground process. For instance, it can be used to model a point process with multiple occurrences in which marks are used to indicate the number of occurrences at each time; In seismology, marks can be used to register the magnitude of earthquakes in addition to the point process which register time of occurrence; A mark could be also a dummy

variable in order to differentiate buy order from sell order in finance. Despite different useful applications of MPP this work will focus on the application of simple point process.

2.2 Hawkes process

Hawkes process is a particular case of point process whose conditional intensity depends on the history of the events. The versatility and the increasing popularity of this model is due to its possibility to combine in one model both a cluster process representation and a simple conditional intensity representation (Daley and Vere-Jones, 2002). In this section we will focus on the conditional intensity representation in order to derive first and second moment measures, whereas a detailed cluster representation could be find e.g. in Chapter 4 of Liniger (2009).

In this section we will make a further contraction in the notation. The realisation of a counting measure will be indicated with $N_t = N(t)$ and the conditional intensity at time t on the filtration $\mathcal{F}_t$ will be indicated with $\lambda_t = \lambda(t|\mathcal{F}_t)$.

2.2.1 Definition and intensity function

Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space. The univariate Hawkes process N with conditional intensity λ_t can be defined for all $t > 0$ and $h \to 0^+$ as following:

$$\begin{cases} \Pr\{N_{t+h} - N_t = 1|\mathcal{F}_t\} = \lambda_t h + o(h) \\ \Pr\{N_{t+h} - N_t > 1|\mathcal{F}_t\} = o(h) \\ \Pr\{N_{t+h} - N_t = 0|\mathcal{F}_t\} = 1 - \lambda_t h + o(h) \end{cases} \tag{2.5}$$

Where $\mathcal{F}_t$ is a filtration on the underlying probability space $(\Omega,\mathcal{F},\mathbb{P})$ and it represents the history of the process. The original definition could be find in (Hawkes, 1971b).

The main characteristic of Hawkes process is that λ_t is dependent on the history of the process $\mathcal{F}_t$, in particular, it is a function of time t and number of jumps N_t. The two most common models of Hawkes processes are those with exponential decay function and those with power law function. In this

work only Hakes processes with exponential decay function will be considered. Models with power law function are more common in seismology, see (Ogata, 1988) for reference.

The dynamics of conditional intensity of a Hawkes process with exponential decay function follow the ODE:

$$\mathrm{d}\lambda_t = \beta(\lambda_\infty - \lambda_t)\mathrm{d}t + \alpha \mathrm{d}N_t \tag{2.6}$$

where:
$\beta > 0$ is the constant rate of decay;
$\lambda_\infty > 0$ is the background intensity;
$\alpha > 0$ is the magnitude of self-excited jump.

The time effect on the intensity is represented by $\beta(\lambda_\infty - \lambda_t)$ and it has a negative contribution if $\beta > 0$. On the contrary, a jump at a given time in the counting measure N_t will increase the intensity by α and it increases the probability of another jump in $(t, t+h]$ through the equation (2.5). This causes the clustering effect of the Hawkes process and it explains the dynamics of "self-exciting" effect. However, there is a finite M such that $\lambda_t < M$ if $\alpha < \beta$, so the process does not blow up. The stationarity condition $\frac{\alpha}{\beta} < 1$ will be showed later in eq. (2.10) and in the proof of the Proposition 2.3

The stochastic process λ_t is a Markov process, from ODE (2.6). $\mathrm{d}\lambda_t$ depends only upon λ_t and $\mathrm{d}N_t$, but from (2.5), $\mathrm{d}N_t$ is defined by λ_t.

In order to find the full conditional intensity equation we can apply Ito's lemma to $e^{\beta t}\lambda_t$, and the solution of ODE (2.6) is:

$$\lambda_t = c(t) + \int_0^t \alpha e^{-\beta(t-s)} dN_s$$

where

$$\begin{aligned} c(t) &= \lambda_0 e^{-\beta t} + \beta\lambda_\infty e^{-\beta t}\int_0^t e^{\beta s} ds \\ &= \lambda_0 e^{-\beta t} + \lambda_\infty e^{-\beta t}(e^{\beta t} - 1) \\ &= e^{-\beta t}(\lambda_0 - \lambda_\infty) + \lambda_\infty \end{aligned}$$

Then, the full equation for λ_t is:

$$\lambda_t = e^{-\beta t}(\lambda_0 - \lambda_\infty) + \lambda_\infty + \int_0^t \alpha e^{-\beta(t-s)} dN_s \tag{2.7}$$

Where $\lambda_0 > 0$ is the initial intensity at $t = 0$.

Note that the impact of a given jump on the intensity decay exponentially as time passes. If $\lambda_0 = \lambda_\infty$ or if $t \to +\infty$ the equation (2.7) become:

$$\lambda_t = \lambda_\infty + \int_0^t \alpha e^{-\beta(t-s)} dN_s \tag{2.8}$$

The equation (2.8) could be expressed as:

$$\lambda_t = \lambda_\infty + \sum_{j:t_j<t} \alpha e^{-\beta(t-t_j)} \tag{2.9}$$

In particular, in the Appendix, the above equation will be used to implement a simulation in Matlab of Hawkes process with $\lambda_0 = \lambda_\infty$.

2.2.2 Simulation of a Hawkes process

A simulation algorithm for the non-homogeneous Poisson process based on thinning method was introduced by Lewis and Shedler (1979). The idea is to simulate a homogeneous Poisson process, and then remove excess points stochastically so that the remaining points satisfy the conditional intensity λ_t. This algorithm requires the conditional intensity to be upper bounded, so that there is a finite M for which $\lambda_t \leq M$ for all t. Hawkes process, as the case non-homogeneous Poisson process, is defined by its conditional intensity process, and a generalization of Shedler-Lewis thinning algorithm, called 'Ogata's modified thinning algorithm', was introduced by Ogata (1981) and it requires only the local boundedness of conditional intensity. In case of Hawkes processes with exponential rate of decay $\beta > 0$, λ_t is a non-increasing function in the interval between two adjacent occurrences. This implies $\lambda_t \leq \lambda_{t_i+}$ for $t \in (t_i, t_{i+1})$, where t_i+ indicates a time just after t_i. So a local bound M_t could be set equal to λ_{t_i+} in the interval (t_i, t_{i+1}) and it has to be updated after each occurrence. However Daley and Vere-Jones (2002) stated that this algorithm is inefficient, and they introduced a simple modification by setting $M_t = \lambda_{t+}$ regardless of whether or not t is a point of the process and adding a function of time interval of length $L_t = \kappa\lambda_{t+}$ for an arbitrary κ. M_t is updated if a new point of the process occurs or if the time frame L_t has elapsed. Ogata's modified thinning algorithm based on Algorithm 7.5.IV of Daley and Vere-Jones (2002) is showed in Algorithm 1 and the pseudo-code converted in Matlab is showed in the Appendix.

Set $t = 0$, $i = 0$, $\kappa = 1/2$, $H = \emptyset$;
while $t < T$ **do**
  Compute $M_t = \lambda_{t+}$ and $L_t = \kappa\lambda_{t+}$;
  Generate an exponential r.v. R with mean $1/M_t$;
  if $R > L_t$ **then**
    Set $t = t + L_t$;
  else
    Generate a r.v. U uniformly distributed on $(0, 1)$;
    if $U > \lambda_{t+R}/M_t$ **then**
      Set $t = t + R$;
    else
      Set $i = i + 1$, $t_i = t + R$, $t = t_i$;
      $H = H \cup t_i$
    end
  end
end
return H;

Algorithm 1: Simulation of a Hawkes process on the interval $(0, T)$, where κ has been set equal to 0.5 according to Daley and Vere-Jones (2002). The output is the vector H containing the times of occurrences $(t_1, t_2, ..., t_n)$.

An example of a Hawkes process with $\lambda_\infty = 0.1$, $\alpha = 0.2$ and $\beta = 0.4$ over the time interval $(0, 100]$ is showed in Figure 2.3. The initial intensity λ_0 is assumed equal to the background intensity λ_∞. The conditional intensity shows vertical jumps at each occurrence time giving the self-exciting effect. Compared to homogeneous Poisson process, where the events are uniformly distributed along the time interval, Hawkes process presents a clustering effect of the occurrences. In the example of Figure 2.3 the number of total events is 14 ($N_{100} = 14$) and we can note a clustering effect around $t = 20$ and $t = 90$.

Ogata's thinning algorithm is used here to give a representation of Hawkes process for the simplicity of its implementation. However, other simulation algorithm in the literature could be found. For example, Ozaki (1979) adopted a simulation based on the conditional Hazard function of the process, which can be defined as:

$$\Lambda(t|\mathcal{F}_t, \theta) = \frac{f(t|\mathcal{F}_t, \theta)}{1 - F(t|\mathcal{F}_t, \theta)}$$

Where $f(.)$ is the probability density function and $F(.)$ is the cumulative

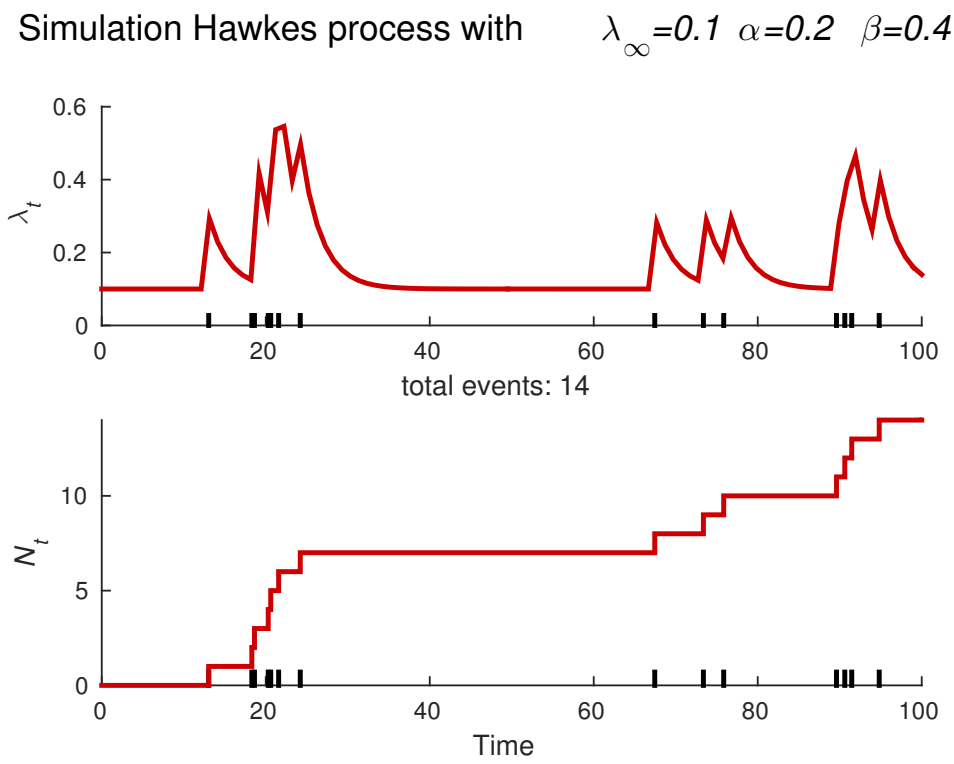

FIGURE 2.3: A simulated realisation of a Hawkes process over the interval $(0, 100]$ using Ogata's thinning algorithm where $\lambda_\infty = 0.1$, $\alpha = 0.2$, $\beta = 0.4$ and $\lambda_0 = \lambda_\infty$. The upper graph shows the conditional intensity over the time, whereas the lower graph shows its counting measure.

distribution function. Møller and Rasmussen (2005) introduced a perfect simulation method based on the cluster representation of the process Same authors formulated also an approximate simulation method which is less computation intensive (Møller and Rasmussen, 2006). Dassios, Zhao, et al. (2013) introduced a numerically efficient simulation algorithm for Hawkes process with exponentially decaying intensity.

2.2.3 Poisson cluster process representation

An alternative interpretation of Hawkes process is provided by Hawkes and Oakes (1974), who showed that all stationary self-exciting point process with finite intensity may be presented as a Poisson cluster process, called otherwise generalised branching Poisson process (Lewis, 1969). Intuitively, a Hawkes process can be seen as a combination of two stochastic mechanisms. First, a background homogeneous Poisson process with intensity λ_∞

which continuously and uniformly generates events. Second, an endogenous feedback mechanism, where any generated event has the potential to lead directly to some future event (Olson and Carley, 2013). The background effect λ_∞ and the feedback effect $\sum_{j:t_j<t} \alpha e^{-\beta(t-t_j)}$ are clearly separable according to eq. (2.9). Similarly the point events of a Hawkes process can be separated into two types: *immigrants* (or in seismology *main events*) without extant parents, which are generated by the background process; *offspring* (or *after shocks*) that are produced by existing point events (Daley and Vere-Jones, 2002). The immigrants points are cluster centres and each point event can generates offspring.

Fig. 2.4 illustrates an example of cluster representation of a Hawkes process. The branching structure is represented in the upper part of the figure. The immigrants points are labelled with zeroes and other numbers $(1, 2, 3, 4)$ indicate the generation that an offspring belong to, for example a point labelled with "1" means that the point is the direct offspring of an immigrant, and a point labelled with "2" means that the point is offspring of an offspring of generation 1. In the example of Fig. 2.4 there are 4 immigrant points and each of them is the centre of a cluster (different colours are used to represents different clusters). The blue cluster has the biggest number of offspring, meanwhile the purple immigrant do not have any offspring. The lower part of Fig. 2.4 indicates the correspondent realisations of point events on the time axis. The branching structure is usually unobservable, we can only observe the times of realisation, but we cannot know whether a point is an immigrant or it is an offspring.

In a Hawkes process with exponential decay and constant background intensity, the immigrant points are generated by the homogeneous Poisson process with intensity λ_∞, whereas each existing point at time t can generate an offspring according to a non-homogeneous Poisson process with intensity $\alpha e^{-\beta(t-t_j)}$, for $t > t_j$, where t_j is the occurrence time of point j. Fig. 2.5 illustrates the intensity path of a Hawkes process with $\lambda_\infty = 0.2$, $\alpha = 0.3$ and $\beta = 0.4$. The grey region represents the background intensity and it is always constant; the coloured areas represents intensity produced by each single jump, for instance the yellow area is the intensity caused by the first jump of the process and its contribution die out exponentially. Fig. 2.5 shows also that the exogenous background mechanism (grey area) is separable from the endogenous feedback mechanism (coloured area). Higher is the intensity at a time t, given the same background intensity λ_∞, higher is

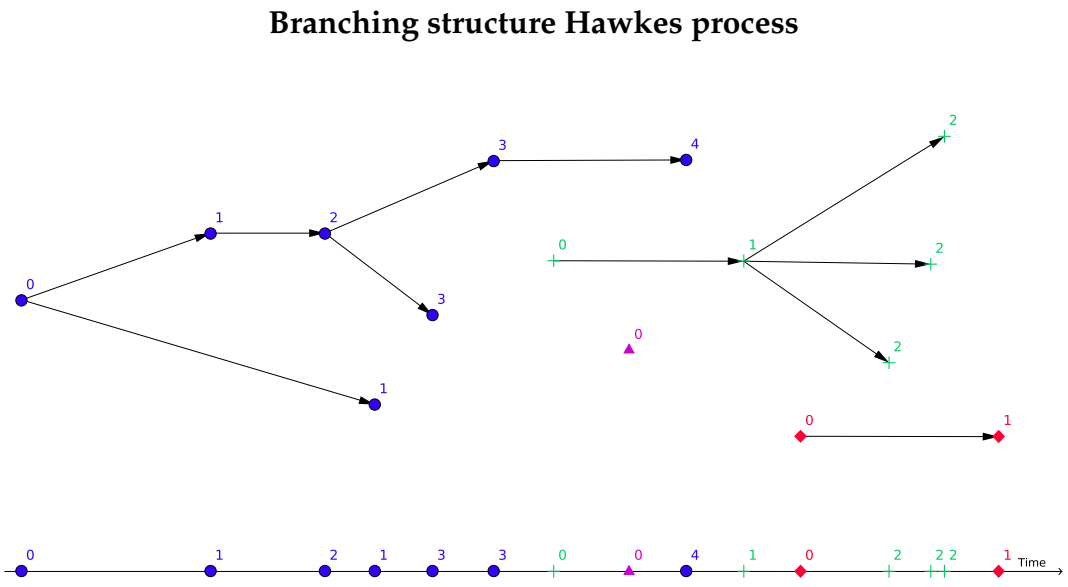

FIGURE 2.4: The branching structure of a realised Hawkes process (upper part) and the realisation on the time axis (lower part). Immigrant points are labelled with zeroes and offspring points are labelled with a number different from zero which indicates the generation that it belongs. Each cluster is represented by a different colour.

the probability that the next point event is an offspring rather than an immigrant. However, the type of a point event and the branching structure of the whole process is unobservable.

An important parameter in a cluster process is the branching ration η, which is defined as the average number of offspring generated by each point event. The branching ratio defines the stationarity of a system and there are three types of cluster process based on η: sub-critical, if $0 < \eta < 1$; critical, if $\eta = 1$; and super-critical, if $\eta > 1$ (Filimonov and Sornette, 2012). In the sub-critical case, there are less than one offspring per event: each immigrant point generates a close cluster with probability 1 and the process is stationary. In the super-critical case, there are more than one offspring per event: this corresponds to an explosive process with the number of events increases exponentially over time. $\eta = 1$ is the critical point separating the two cases and it means on average there is one offspring per each event (Saichev, Helmstetter, and Sornette, 2005).

In a sub-critical cluster process, the branching ratio is the proportion of offspring events respect to the total number of events. A simple way to obtain its theoretical measure given the intensity equation is through the

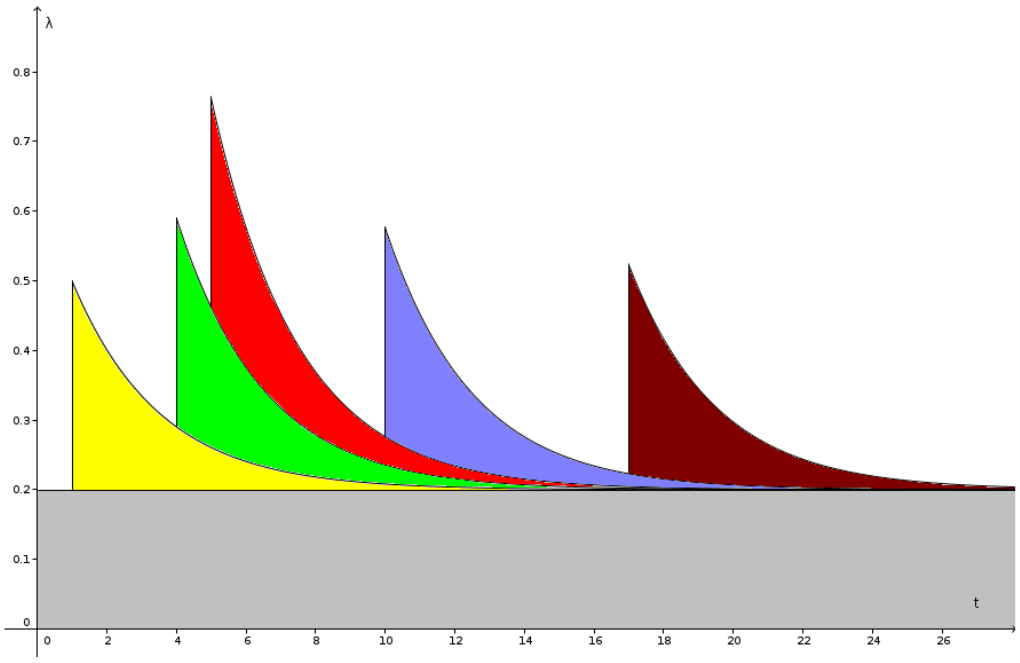

FIGURE 2.5: Intensity contribution of each point event. the background intensity $\lambda_\infty = 0.2$ is represented by the grey area. Each jump increases the intensity by $\alpha = 0.3$ and it dies out with an exponential decay $\alpha e^{-\beta(t-t_j)}$, where $\beta = 0.4$ and t_j is the occurrence time of point j. The contribution of intensity generated by the point events are represented by the coloured area.

definition:

$$\eta = \int_0^\infty \omega(t)dt$$

where, $\omega(t)$ is the endogenous part of the intensity. In a Hawkes process with exponential decay $\omega(t) = \alpha e^{-\beta(t)}$, the theoretical branching ratio is:

$$\eta = \frac{\alpha}{\beta}. \tag{2.10}$$

Note that the condition for a sub-critical cluster process is $\eta < 1$, so the stationarity condition for a Hawkes process is $\frac{\alpha}{\beta} < 1$.

In the context of Hawkes process applied to finance, Filimonov and Sornette (2012) and Lorenzen (2012) claim that the branching ratio is key in predicting movements and important events in the stock markets. Filimonov and Sornette (2012) showed also that during the "Flash-crash" of May 6,

2010, η increased significantly and approached to 1. The authors believe that knowing the level of endogeneity could help to predict when a market is critical.

2.2.4 Analytical framework

The derivation of analytical results follows Errais, Giesecke, and Goldberg (2010) and Chapter 2.1 of Da Fonseca and Zaatour (2014). The main idea is to exploit a two-dimensional stochastic process X_t consisting of Hawkes process and its conditional intensity, such that $X_t = (\lambda_t, N_t)$. We will derive the distributional properties of Hawkes process by taking the advantage of particular properties of X_t.

While the point process N_t itself does not have the Markov property, the process $X_t = (\lambda_t, N_t)$ is a Markov process in the state space $D = \mathbb{R}_+ \times \mathbb{N}$, since the process λ_t has Markov property as we saw in the Section 2.2.1. In the context of piecewise deterministic Markov process theory and using the results in Davis (1984), X_t has an infinitesimal generator $\mathcal{D}$, defined at a function $g : D \to \mathbb{R}$ with continuous partial derivative $\frac{\partial g}{\partial \lambda}(x)$, such that:

$$\mathcal{D}g(x) = \lim_{h \to 0} \frac{\mathbb{E}_t^x[g(X_{t+h})] - g(x)}{h}$$

with $\mathbb{E}_t^x = \mathbb{E}^x[.|\mathcal{F}_t]$ and $X_t = x$.

In the specific case of process $X_t = (\lambda_t, N_t)$, the infinitesimal generator computed using the definition (2.5), the dynamic (2.11) and Taylor series is:

$$\mathcal{D}g(x) = \beta(\lambda_\infty - \lambda_t)\frac{\partial g}{\partial \lambda}(x) + \lambda_t[g(\lambda_t + \alpha, N_t + 1) - g(x)] \tag{2.11}$$

Moreover, the process:

$$M_t = g(X_t) - g(X_0) - \int_0^t \mathcal{D}f(X_u)du$$

is a martingale relative to its natural filtration (A proof is giving in the appendix of Errais, Giesecke, and Goldberg (2010) or proposition 1.6 of chapter VII in Revuz and Yor (1999)). By the martingale property, we have for $s > t$:

$$\mathbb{E}_t\left[g(X_s) - \int_0^s \mathcal{D}f(X_u)du\right] = g(X_t) - \int_0^t \mathcal{D}f(X_u)du$$

which bring to the equation:

$$\mathbb{E}_t[g(X_s)] = g(X_t) + \mathbb{E}_t\left[\int_t^s \mathcal{D}f(X_u)du\right] \tag{2.12}$$

The equation (2.12) is known in the literature also as Dynkin's formula.

The following derivation of moments of the process X_t and its autocovariance will rely on the infinitesimal generator (2.11) and Dynkin's formula (2.12). It will be first computed the expected number of jumps and the expected intensity as they turn to be useful in the later derivation.

Lemma 2.1 *Given a Hawkes process $X_t = (\lambda_t, N_t)$ with dynamic give by (2.6), the expected number o jumps $\mathbb{E}[N_t]$ and the expected intensity $\mathbb{E}[\lambda_t]$ satisfy the set of ordinary differential equation:*

$$d\mathbb{E}[N_t] = \mathbb{E}[N_t]dt \tag{2.13}$$

$$d\mathbb{E}[\lambda_t] = (\beta\lambda_\infty + (\alpha - \beta)\mathbb{E}[\lambda_t])dt \tag{2.14}$$

Proof. Consider $g(X_t) \equiv N_t$, a function from $D \to \mathbb{N}$ with $\dfrac{\partial g}{\partial \lambda}(x) = 0$. The infinitesimal generator from (2.11) is:

$$\mathcal{D}g(X_t) = \lambda_t[(N_t + 1) - N_t] = \lambda_t$$

Applying Dynkin's formula (2.12) we obtain:

$$\mathbb{E}[N_t] = N_0 + \mathbb{E}\left[\int_0^t \lambda_s ds\right]$$

Then using Fubini's theorem we have:

$$\mathbb{E}[N_t] = N_0 + \int_0^t \mathbb{E}\left[\lambda_s\right] ds \tag{2.15}$$

Which differentiating with respect to t gives the ODE (2.13).

To obtain the ODE (2.14) consider now $g(X_t) \equiv \lambda_t$. The infinitesimal generator (2.11) in this case is:

$$\mathcal{D}g(X_t) = \beta(\lambda_\infty - \lambda_t) + \lambda_t[(\lambda_t + \alpha) - \lambda_t] = \beta(\lambda_\infty - \lambda_t) + \alpha\lambda_t$$

and Dynkin's formula (2.12) and Fubini's theorem results in:

$$\begin{aligned}\mathbb{E}[\lambda_t] &= \lambda_0 + \mathbb{E}\left[\int_0^t \beta(\lambda_\infty - \lambda_s) + \alpha\lambda_s ds\right] \\ &= \lambda_0 + \beta\lambda_\infty t + (\alpha - \beta)\int_0^t \mathbb{E}[\lambda_s] ds\end{aligned}$$

Which leads to (2.14) after differentiating with respect to t. ■

Lemma 2.2 *Given a Hawkes process $X_t = (\lambda_t, N_t)$ with dynamic give by (2.6), $\mathbb{E}[N_t^2]$, $\mathbb{E}[\lambda_t N_t]$ and $\mathbb{E}[\lambda_t^2]$ satisfy the set of ordinary differential equation:*

$$d\mathbb{E}[N_t^2] = 2\mathbb{E}[\lambda_t N_t]dt + \mathbb{E}[\lambda_t] \tag{2.16}$$

$$\mathbb{E}[\lambda_t N_t] = \beta\lambda_\infty \mathbb{E}[N_t]dt + (\alpha - \beta)\mathbb{E}[\lambda_t N_t]dt + \mathbb{E}[\lambda_t^2]dt + \alpha\mathbb{E}[\lambda_t]dt \tag{2.17}$$

$$d\mathbb{E}[\lambda_t^2] = (\alpha^2 + 2\beta\lambda_\infty)\mathbb{E}[\lambda_t]dt + 2(\alpha - \beta)\mathbb{E}[\lambda_t^2]dt \tag{2.18}$$

Proof. Consider the function $g(X_t) \equiv N_t^2$. The infinitesimal generator (2.11) is:

$$\mathcal{D}g(X_t) = 0 + \lambda_t[(N_t + 1)^2 - N_t] = 2\lambda_t N_t + \lambda_t$$

Dynkin's formula and Fubuni's theorem leads to:

$$\begin{aligned}\mathbb{E}[N_t^2] &= N_0^2 + \mathbb{E}\left[\int_0^t 2\lambda_u N_u + \lambda_u du\right] \\ &= N_0^2 + 2\int_0^t \mathbb{E}[\lambda_u N_u]du + \int_0^t \mathbb{E}[\lambda_u]du\end{aligned} \tag{2.19}$$

Which leads to (2.16) after differentiating with respect to t.

To obtain ODE (2.17) and (2.18) we can use the same procedure for $g(X_t) \equiv \lambda_t N_t$ and $g(X_t) \equiv \lambda_t^2$. ■

Proposition 2.3 *Given a Hawkes process $X_t = (\lambda_t, N_t)$ with dynamic given by (2.6), then the expected value of the number of jumps in a time interval τ for $t \to \infty$ is:*

$$\lim_{t\to\infty} \mathbb{E}[N_{t+\tau} - N_t] = \frac{\beta\lambda_\infty}{\beta - \alpha}\tau = \Lambda\tau \tag{2.20}$$

Where $\Lambda = \lim_{t\to\infty} \mathbb{E}[\lambda_t] = \frac{\lambda_\infty}{1-\alpha/\beta}$ is the stationary regime expected intensity.

Proof. To compute the expected number of jumps during a time interval τ we first need to find the expression of $\mathbb{E}[\lambda_t]$. To do so, we can find the

solution of ODE (2.14) with initial condition $\mathbb{E}[\lambda_0] = \lambda_0$ and integral factor $e^{-(\alpha-\beta)t}$. It results in:

$$\mathbb{E}[\lambda_t] = \lambda_\infty \beta \frac{e^{(\alpha-\beta)t} - 1}{\alpha - \beta} + e^{(\alpha-\beta)t}\lambda_0 \tag{2.21}$$

Note that if $\beta < \alpha$, $\lim_{t\to\infty} \mathbb{E}[\lambda_t] = +\infty$. In the above equation we can recognize the condition for the stability of Hawkes process, which is $\frac{\alpha}{\beta} < 1$, coherent with what we found in eq. (2.10). And the stationary regime expected intensity is:

$$\lim_{t\to\infty} \mathbb{E}[\lambda_t] = \frac{\lambda_\infty}{1 - \alpha/\beta} = \Lambda \tag{2.22}$$

Then, we can use (2.21) with (2.15) to find the expression for the mean number of jumps:

$$\mathbb{E}[N_t] = N_0 + \frac{e^{(\alpha-\beta)t} - 1 - (\alpha-\beta)t}{(\alpha-\beta)^2}\lambda_\infty\beta + \frac{e^{(\alpha-\beta)t} - 1}{\alpha-\beta}\lambda_0 \tag{2.23}$$

Using the above computation, we can now obtain the the expected value of the number of jumps during a time interval of length τ:

$$\mathbb{E}[N_{t+\tau} - N_t] = \frac{-\lambda_\infty\beta\tau}{\alpha-\beta} + \frac{e^{(\alpha-\beta)t}}{(\alpha-\beta)^2}(e^{(\alpha-\beta)\tau} - 1)[\lambda_\infty\beta + (\alpha-\beta)\lambda_0] \tag{2.24}$$

Under the stability condition $\frac{\alpha}{\beta} < 1$ and for limit of $t \to \infty$, the above expression does not depend on the initial value of the intensity λ_0 and it results in (2.20). ■

Proposition 2.4 *Given a Hawkes process $X_t = (\lambda_t, N_t)$ with dynamic given by (2.6), then the variance of the number of jumps during a time interval of length τ for $t \to \infty$ is:*

$$\begin{aligned} V(\tau) &= \lim_{t\to\infty} \left\{\mathbb{E}[(N_{t+\tau} - N_t)^2] - \mathbb{E}[N_{t+\tau} - N_t]^2\right\} \\ &= \Lambda\left(\tau\kappa^2 + (1-\kappa^2)\frac{1 - e^{-\tau\gamma}}{\gamma}\right) \end{aligned} \tag{2.25}$$

Where: $\Lambda = \frac{\lambda_\infty}{1-\alpha/\beta}$, $\kappa = \frac{1}{1-\alpha/\beta}$ *and* $\gamma = \beta - \alpha$

Proof. We first need to compute the second moment of number of jumps during a time interval τ in order to find the variance:

$$\mu_2 = \mathbb{E}\left[(N_{t+\tau} - N_t)^2\right] = \mathbb{E}\left[\mathbb{E}_t[N_{t+\tau}^2] - 2N_t\mathbb{E}_t[N_{t+\tau}] + N_t^2\right] \tag{2.26}$$

From equation (2.19) and (2.15) we know that:

$$\mathbb{E}_t[N_{t+\tau}^2] = N_t^2 + 2\int_t^{t+\tau} \mathbb{E}_t[\lambda_u N_u]du + \int_t^{t+\tau} \mathbb{E}_t[\lambda_u]du$$
$$\mathbb{E}_t[N_{t+\tau}] = N_t + \int_t^{t+\tau} \mathbb{E}_t[\lambda_u]du$$

Using the expressions above, the second moment of the number of jumps in a given interval (2.26) become:

$$\mu_2 = 2\int_t^{t+\tau} \mathbb{E}[\lambda_u N_u]du + \int_t^{t+\tau} \mathbb{E}[\lambda_u]du - 2\mathbb{E}\left[N_t \int_t^{t+\tau} \mathbb{E}_t[\lambda_u]du\right] \tag{2.27}$$

The first integral of (2.27) can be computed from ODE (2.17) and it leads to:

$$\begin{aligned} &\int_t^{t+\tau} \mathbb{E}[\lambda_u N_u]du = \\ = \quad &\int_t^{t+\tau} e^{(\alpha-\beta)(u-t)}\mathbb{E}[\lambda_t N_t]du \\ &+ \int_t^{t+\tau}\int_t^u e^{(\alpha-\beta)(u-s)}\{\beta\lambda_\infty\mathbb{E}[N_s] + \mathbb{E}[\lambda_s^2] + \alpha\mathbb{E}[\lambda_s]\}dsdu \end{aligned}$$

In the similar way, the third term of (2.27) can be computed from ODE (2.14) and it gives:

$$\begin{aligned} &\mathbb{E}\left[N_t\int_t^{t+\tau}\mathbb{E}_t[\lambda_u]du\right] = \\ = \quad &\mathbb{E}\left[N_t\left(\int_t^{t+\tau} e^{(\alpha-\beta)(u-t)}\lambda_t du + \int_t^{t+\tau}\int_t^u e^{(\alpha-\beta)(u-r)}\beta\lambda_\infty dr du\right)\right] \\ = \quad &\int_t^{t+\tau} e^{(\alpha-\beta)(u-t)}du\mathbb{E}[\lambda_t N_t] + \int_t^u e^{(\alpha-\beta)(u-r)}drdu\beta\lambda_\infty\mathbb{E}[N_t] \end{aligned}$$

Substituting $\mathbb{E}[N_s] = \mathbb{E}[N_t] + \int_t^s \mathbb{E}[\lambda_r]dr$ in the result of first integral of (2.27) and after simplification, the expression for the second moment is:

$$\begin{aligned}\mu_2 =& 2\int_t^{t+\tau}\int_t^u e^{(\alpha-\beta)(u-s)}\left\{\beta\lambda_\infty\int_t^s \mathbb{E}[\lambda_r]dr + \mathbb{E}[\lambda_s^2] + \alpha\mathbb{E}[\lambda_s]\right\}dsdu \\ &+ \int_t^{t+\tau}\mathbb{E}[\lambda_u]du\end{aligned}$$

Note that the above expression depends only on $\mathbb{E}[\lambda_t]$ and $\mathbb{E}[\lambda_t^2]$. These two expectations depend on the initial intensity λ_0, but for $t \to \infty$, the contribution of λ_0 become insignificant. The second moment of the number of jumps over time interval of length τ for $t \to \infty$ is:

$$\begin{aligned}\lim_{t\to\infty}\mathbb{E}[(N_{t+\tau}-N_t)^2] = \lim_{t\to\infty}\tau\Lambda + 2\beta\lambda_\infty\Lambda\int_t^{t+\tau}\int_t^u e^{(\alpha-\beta)(u-s)}\int_t^s drdsdu \\ + 2(\Lambda_2+\alpha\Lambda)\int_t^{t+\tau}\int_t^u e^{(\alpha-\beta)(u-s)}dsdu\end{aligned}$$

Where $\Lambda = \lim_{t\to\infty}\mathbb{E}[\lambda_t] = \frac{\lambda_\infty}{1-\alpha/\beta}$ and $\Lambda_2 = \lim_{t\to\infty}\mathbb{E}[\lambda_t^2] = \Lambda\left(\frac{\alpha^2+2\beta\lambda_\infty}{2(\beta-\alpha)}\right)$.

$$\begin{aligned}\int_t^{t+\tau}\int_t^u e^{(\alpha-\beta)(u-s)}\int_t^s drdsdu &= -\frac{\tau^2}{2(\alpha-\beta)} - \frac{\tau}{(\alpha-\beta)^2} + \frac{e^{(\alpha-\beta)\tau}-1}{(\alpha-\beta)^3} \\ \int_t^{t+\tau}\int_t^u e^{(\alpha-\beta)(u-s)}dsdu &= -\frac{\tau}{\alpha-\beta} + \frac{e^{(\alpha-\beta)\tau}-1}{(\alpha-\beta)^2}\end{aligned}$$

The long run variance of the number of jumps during a time interval of length τ is deduced from the above expressions and (2.20):

$$\begin{aligned}&\lim_{t\to\infty}\left\{\mathbb{E}[(N_{t+\tau}-N_t)^2] - \mathbb{E}[N_{t+\tau}-N_t]^2\right\} = \\ =&\tau\Lambda + 2\beta\lambda_\infty\Lambda\left[\frac{\tau^2}{2(\beta-\alpha)} - \frac{\tau}{(\beta-\alpha)^2} - \frac{e^{-\tau(\beta-\alpha)}-1}{(\beta-\alpha)^3}\right] \\ &+ 2\Lambda\left[\frac{\alpha^2+2\beta\lambda_\infty}{2(\beta-\alpha)} + \alpha\right]\left[\frac{\tau}{\beta-\alpha} + \frac{e^{-\tau(\beta-\alpha)}-1}{(\beta-\alpha)^2}\right] - \tau^2\Lambda^2 \\ =&\tau\Lambda\left[1 - \frac{2\beta\lambda_\infty}{(\beta-\alpha)^2} + \frac{\alpha^2+2\beta\lambda_\infty}{(\beta-\alpha)^2} + \frac{2\alpha}{\beta-\alpha}\right] \\ &+ \Lambda\left[\frac{1-e^{-\tau(\beta-\alpha)}}{(\beta-\alpha)^2}\right]\left(\frac{2\beta\lambda_\infty}{\beta-\alpha} - \frac{\alpha^2+2\beta\lambda_\infty}{\beta-\alpha} - 2\alpha\right) \\ =&\Lambda\left[\frac{\beta^2}{(\beta-\alpha)^2}\tau + \left(\frac{\alpha^2-2\alpha\beta}{(\beta-\alpha)^2}\right)\left(\frac{1-e^{-\tau(\beta-\alpha)}}{\beta-\alpha}\right)\right]\end{aligned}$$

After rearrangement and substituting $\kappa = \dfrac{1}{1-\alpha/\beta}$ and $\gamma = \beta - \alpha$, we obtain the equation (2.25). ■

2.2.5 Marked Hawkes process

A specification of marked point process of Section 2.1.5 and a generalization of simple Hawkes process is the Marked version of Hawkes process introduced in Hawkes (1972) with a first application in seismology in Hawkes and Adamopoulos (1973). In this case, the conditional intensity of the ground process still depends on the history of the events, but in addition, it can be also influenced by the observed mark values. Assuming $\lambda_\infty = \lambda_0$, a general way to express the conditional intensity of a marked Hawkes process is:

$$\lambda_t = \lambda_\infty + \sum_{j:t_j<t} \omega(t - t_j, m_j)$$

where $\omega(t, m)$ is a generic non negative function and m_j are marks. The contribution of marks on the λ_t can occur in different way. A common model used literature is:

$$\omega(t, m) = g(m)\omega^*(t)$$

where $g(m)$ is a $\mathcal{M} \to \mathbb{R}^+$ function which incorporates the effects of marks on the intensity. Note that if $g(m) = 1$, the conditional intensity is independent from the marks and the process degenerates in an unmarked Hawkes process. The impact of the observed marks usually assumes an exponential form in the literature, such that $g(m) = e^{\gamma m}$. A basic example of conditional intensity of a marked Hawkes process could be:

$$\lambda_t = \lambda_\infty + \alpha \sum_{j:t_j<t} e^{\gamma m_j - \beta(t-t_j)}$$

Initially, marked hawkes process are used to model the occurrences of earthquakes, for example (Hawkes and Adamopoulos, 1973), (Ogata, 1988) and (Vere-Jones and Ozaki, 1982). Recently many applications in finance could be found, for example Chavez-Demoulin, Davison, and McNeil (2005) used a marked hawkes process to model extreme returns in financial time series.

Chapter 3

Estimation of parameters

This Chapter will illustrate two approaches for the estimation of parameters of a Hawkes process. The classic approach, which uses direct numerical maximization of log-likelihood; and the Expectation-Maximization (EM) algorithm.

Early estimations of Hawkes processes are performed by using direct numerical maximization of log-likelihood function. This method was introduced in the works of Vere-Jones (1978) and Ozaki (1979). Examples of applications could be found in Ogata (1988) and more recently, in finance, in the works of Bowsher (2007) and Embrechts, Liniger, Lin, et al. (2011). This inference strategy will be analysed in the first section of this chapter.

The computation of maximum likelihood estimation (MLE) through EM algorithm will be subject of the second section. This algorithm is based on the cluster representation of Hawkes process and the unobservable branching structure is treated as the latent variables. The estimation of Hawkes process using parametric EM algorithm is relatively recent and its description could be found in Veen and Schoenberg (2008) and Marsan and Lengliné (2008) and Olson and Carley (2013).

More recently, non parametric estimations have been proposed by Lewis and Mohler (2011) and Bacry, Dayri, and Muzy (2012). However, this chapter will focus on the classic numerical optimization approach and on the parametric EM algorithm.

The presentation and the setting of this chapter follow closely Chapter 3 of Lapham (2014).

3.1 Direct numerical maximization of likelihood

The likelihood of a point process X_t on $(0, T]$, with realizations $t_1, t_2, ..., t_{N_T}$, could be defined following the proposition 7.2.III of Daley and Vere-Jones (2002) as:

$$L(\theta) = \prod_{i=1}^{N(T)} \lambda_{t_i} \exp\left(-\int_0^T \lambda_s ds\right)$$

where $\theta \in \Theta$ is the vector of parameters. The log-likelihood ratio relative to a Poisson process on $(0, T]$ with constant intensity 1 is:

$$l(\theta) = \log \frac{L(\theta)}{L_0(\theta)} = \sum_{i=1}^{N_T} \log(\lambda_t) + \int_0^T (1 - \lambda_s) ds$$

In the specific case of Hawkes process with exponential kernel and $\lambda_0 = \lambda_\infty$, the log-likelihood is expressible as:

$$l(\theta) = \int_0^T \left(1 - \lambda_\infty - \sum_{i:t_i<s} \alpha e^{-\beta(s-t_i)}\right) ds + \sum_{i=1}^{N_T} \log\left(\lambda_\infty + \sum_{j:t_j<t_i} \alpha e^{-\beta(t_i-t_j)}\right)$$

after solving the following integral:

$$\int_0^T \sum_{i:t_i<s} \alpha e^{-\beta(s-t_i)} ds = \sum_{i=1}^{N_T} \left(\frac{\alpha}{\beta} - \frac{\alpha}{\beta} e^{-\beta(T-t_i)}\right)$$

we can write the log-likelihood as:

$$l(\theta) = T - T\lambda_\infty - \frac{\alpha}{\beta} \sum_{i=1}^{N_T} \left(1 - e^{-\beta(T-t_i)}\right) + \sum_{i=1}^{N_T} \log\left(\lambda_\infty + \alpha \sum_{j:t_j<t_i} e^{-\beta(t_i-t_j)}\right) \tag{3.1}$$

The second summation in the equation (3.1) presents a nested sum, it requires $N_T(N_T - 1)$ operations in order to be solved. This nested sum is the most computational intensive part in the log-likelihood evaluation and it makes the order of the total operations of $(N_T)^2$. In the early applications of Hawkes process model, the computational power was not sufficient for real-world data sets and the numerical estimation of parameters was difficult (Liniger, 2009). The increase of computational speed may be one of the reasons for the increased number of applications (Lapham, 2014). However there are different methods to reduce the computational burden in the

estimation of parameters.

In the specific case of Hawkes process with an exponential decay, the number of operations could be reduced to the order of N_T. As noted by Ogata (1981), the computation of equation (3.1) could be simplified using a recursive formula. We can see that $\forall i > 1$

$$\begin{aligned} A(i) &= \sum_{j=1}^{i-1} e^{-\beta(t_i - t_j)} \\ &= e^{-\beta(t_i - t_{i-1})} \left(1 + \sum_{j=1}^{i-2} e^{-\beta(t_{i-1} - t_j)} \right) \\ &= e^{-\beta(t_i - t_{i-1})} \left(1 + A(i-1) \right) \end{aligned}$$

then, by setting $A(1) = 0$, we can rewrite the log-likelihood function (3.1) as:

$$l(\theta) = T - T\lambda_\infty - \frac{\alpha}{\beta} \sum_{i=1}^{N_T} \left(1 - e^{-\beta(T - t_i)} \right) + \sum_{i=1}^{N_T} \log \left(\lambda_\infty + \alpha A(i) \right). \tag{3.2}$$

Note that the conditional intensity could be expressed as:

$$\lambda_t = \lambda_\infty + \alpha A(i),$$

the simplification through the recursive formula lies on the fact that λ_t is a Markov process.

In the context of univariate Hawkes process with constant background intensity λ_∞, Ogata (1978) proved the log-likelihood estimator $\hat{\theta} = (\widehat{\lambda}_\infty, \hat{\alpha}, \hat{\beta})$, defined as $\hat{\theta} = \arg\max_{\theta \in \Theta} l(\theta)$, to be consistent, i.e. it converges in probability to the true value

$$\forall \epsilon > 0, \lim_{T \to \infty} \Pr[|\hat{\theta} - \theta| > \epsilon] = 0;$$

and asymptotically normal, so that for $T \to \infty$

$$\sqrt{T}(\hat{\theta} - \theta) \to \mathcal{N}(0, I^{-1}(\theta)),$$

where $I^{-1}(\theta) = T\mathbb{E}\left[\frac{1}{\lambda} \frac{\partial \lambda}{\theta_i} \frac{\partial \lambda}{\theta_j} \right]$ $i, j = 1, 2, 3$ is a 3×3 matrix.

However there is no closed-form expression for the computation of $\hat{\theta}$, therefore we have to perform numerical maximization, for example using `fminunc` in Matlab, in order to find the solution.

3.2 Expectation Maximization algorithm

The EM algorithm is an iterative algorithm for computing the maximum likelihood estimates when the observations can be viewed as incomplete data (Dempster, Laird, and Rubin, 1977). Each iteration of the EM algorithm consists of two steps: expectation, or the E-step, and maximization, or the M-step. In the E-step the latent variables are estimated given the observed data and current estimates for the parameters. In the M-step the likelihood function is maximized using the latent variables estimated in the E-step, this step computes new estimates for the next iteration. The likelihood is increasing at each iteration, so the convergence is assured (Borman, 2004). In the context of Hawkes processes, Veen and Schoenberg (2008) proposed an EM algorithm for the parameters estimation in which the unobservable branching structure under the Poisson cluster representation is treated as the latent variables. The computational burden of EM algorithm could be significant and several approximations could be made in order to make the estimation less computationally intensive. We will illustrate both the exact inference and an approximate inference following the work of Veen and Schoenberg (2008), Olson and Carley (2013) and Lapham (2014).

3.2.1 Exact EM algorithm estimation

Consider a general Hawkes process with conditional intensity:

$$\lambda_t = \lambda_\infty + \sum_{j:t_j<t} \omega(t - t_j)$$

where $\omega(.)$ is the endogenous mechanism of the process. Following the work of Lapham (2014), we introduce a variable u_i in order to distinguish immigrant point events from offspring ones as following:

$u_i = i$ if point event i is an immigrant;

$u_i = j$ if point event i is an offspring triggered by the point event j.

The complete data of the process consists of occurrences time and u_i associated with each occurrence $(t_1, u_1), (t_2, u_2), ..., (t_{N_T}, u_{N_T})$. The variable u_i fully describe the branching structure of a Hawkes process.

If the branching structure is observable, we can write the complete data log-likelihood as:

$$l_{CD}(\theta) = \sum_{i:u_i=i} \log(\lambda_\infty) - \lambda_\infty T - \sum_{i=1}^{N_T} \int_{t_i}^{T} \omega(s-t_i)ds + \sum_{i:u_i \neq i} \log \omega(t_i - t_{u_i}) \quad (3.3)$$

Intuitively, the first sum in the eq. (3.3) is related to the contribution of immigrant events which arrive according to the background intensity λ_∞. The last sum in the eq. (3.3) is related to the contribution of offspring events which arrive according to the feedback mechanism $\omega(.)$. The two addends in the middle with negative sign describes how unlikely it was to have not seen additional events (Olson and Carley, 2013).

Expectation step

The E-step consists in the computation of the conditional expectation of log-likelihood with respect to the history of the process $\mathcal{F}_T$ under the current estimate of parameters $\theta^{(k)}$, where k is the number of iterations already performed in the algorithm. The conditional expected value of log-likelihood based on the eq. (3.3) could be rewrite as:

$$\begin{aligned} Q\left(\theta|\theta^{(k)}\right) &= \mathbb{E}\left[l_{CD}(\theta)|\mathcal{F}_T, \theta^{(k)}\right] \\ &= \mathbb{E}\Bigg[\log(\lambda_\infty) \sum_{i=1}^{N_T} I_{\{u_i=i\}} - \lambda_\infty T - \sum_{i=1}^{N_T} \int_{t_i}^{T} \omega(s-t_i)ds \\ &\quad + \sum_{i=1}^{N_T} \sum_{j \neq i} I_{\{u_i=j\}} \log \omega(t_i - t_j)|\mathcal{F}_T, \theta^{(k)}\Bigg] \end{aligned} \quad (3.4)$$

where I is a dummy variable which assume unit value when the condition in the subscript is true.

Given the history of the process $\mathcal{F}_T$ and the current estimate $\theta^{(k)}$ we can define the probability for which whether a point event is an immigrant ($u_i = i$) or it is triggered by the event j ($u_i = j$, where $j < i$):

$$\begin{aligned} \Pr\left\{u_i = i|\mathcal{F}_{t_i}, \theta^{(k)}, t_i\right\} &= \frac{\lambda_\infty^{(k)}}{\lambda_\infty^{(k)} + \sum_{n:t_n<t_i} \omega(t_i - t_n|\theta^{(k)})} \\ \Pr\left\{u_i = j|\mathcal{F}_{t_i}, \theta^{(k)}, t_i\right\} &= \frac{\omega(t_i - t_j|\theta^{(k)})}{\lambda_\infty^{(k)} + \sum_{n:t_n<t_i} \omega(t_i - t_n|\theta^{(k)})} \end{aligned} \quad (3.5)$$

In the complete data log-likelihood, we have the following identities:

$$\begin{aligned}\mathbb{E}\left[I_{\{u_i=i\}}|\mathcal{F}_T,\theta^{(k)}\right] &= \Pr\left\{u_i=i|\mathcal{F}_{t_i},\theta^{(k)},t_i\right\}\\ \mathbb{E}\left[I_{\{u_i=j\}}|\mathcal{F}_T,\theta^{(k)}\right] &= \Pr\left\{u_i=j|\mathcal{F}_{t_i},\theta^{(k)},t_i\right\}.\end{aligned} \tag{3.6}$$

Substituting eq. (3.6) into eq. (3.4) we can rewrite the conditional expected value of log-likelihood as:

$$\begin{aligned}Q(\theta|\theta^{(k)}) = \log(\lambda_\infty)\sum_{i=1}^{N_T}\Pr\left\{u_i=i|\mathcal{F}_{t_i},\theta^{(k)},t_i\right\} - \lambda_\infty T - \sum_{i=1}^{N_T}\int_{t_i}^{T}\omega(s-t_i)ds\\ + \sum_{i=2}^{N_T}\sum_{j=1}^{i-1}\log(\omega(t_i-t_j))\Pr\left\{u_i=j|\mathcal{F}_{t_i},\theta^{(k)},t_i\right\}\end{aligned} \tag{3.7}$$

Maximization step

The objective of M-step is to find the parameters that maximize $Q(\theta|\theta^{(k)})$ in the eq. (3.7). The parameters obtained in this step will be denoted as $\theta^{(k+1)}$ and they will be used in the next iteration of the algorithm.

The maximization is performed by setting the partial derivative of $Q(\theta|\theta^{(k)})$ with respect to each parameter equal to 0. In the exact EM algorithm the closed-form expression could be found only for λ_∞, whereas there is no analytical solution for other parameters of $\theta^{(k+1)}$.

Iteration of EM algorithm

Each iteration of EM algorithm increases the log-likelihood of the eq. (3.3), the details for the monotonicity and the convergence to a stationary value could be found in the Chapter 3 of McLachlan and Krishnan (2007).

Starting from arbitrary set of parameters $\theta^{(0)}$, the EM algorithm is performed through the E and M steps in order to maximize the log-likelihood. The algorithm stops once the convergence criterion is reached, so that $Q(\theta|\theta^{(k+1)}) - Q(\theta|\theta^{(k)}) < \epsilon$.

The EM algorithm could summarized as following:

Step 0 Set $k=0$ and set the starting value of parameters in $\theta^{(0)}$;

Step 1 (E-Step) Estimate the triggering probabilities of eq. (3.5) by using $\theta^{(k)}$ and $\mathcal{F}_T$;

Step 2 (M-Step) Find $\theta^{(k+1)}$ such that eq. (3.7) is maximized;

Step 3 If $Q(\theta|\theta^{(k+1)}) - Q(\theta|\theta^{(k)}) \geq \epsilon$, return to Step 1, otherwise stop the algorithm.

In the Hawkes process with exponential decay the log-likelihood is:

$$\begin{aligned} l(\theta) = \sum_{i:u_i=i} \log(\lambda_\infty) - \lambda_\infty T - \sum_{i=1}^{N_T} \int_{t_i}^{T} \alpha e^{-\beta(s-t_i)} ds \\ + \sum_{i:u_i \neq i} (\log \alpha - \beta(t_i - t_{u_i})). \end{aligned} \tag{3.8}$$

The probability that an event i is an immigrant event or it is triggered by the event j is:

$$\begin{aligned} \Pr\left\{u_i = i | \mathcal{F}_{t_i}, \theta^{(k)}, t_i\right\} &= \frac{\lambda_\infty^{(k)}}{\lambda_\infty^{(k)} + \sum_{n:t_n<t_i} \alpha^{(k)} e^{-\beta^{(k)}(t_i - t_n)}} \\ \Pr\left\{u_i = j | \mathcal{F}_{t_i}, \theta^{(k)}, t_i\right\} &= \frac{\alpha^{(k)} e^{-\beta^{(k)}(t_i - t_j)}}{\lambda_\infty^{(k)} + \sum_{n:t_n<t_i} \alpha^{(k)} e^{-\beta^{(k)}(t_i - t_n)}} \end{aligned} \tag{3.9}$$

The conditional expect log-likelihood in terms of triggering probabilities is:

$$\begin{aligned} Q(\theta|\theta^{(k)}) = \log(\lambda_\infty) \sum_{i=1}^{N_T} \Pr\left\{u_i = i | \mathcal{F}_{t_i}, \theta^{(k)}, t_i\right\} - \lambda_\infty T + \frac{\alpha}{\beta} \sum_{i=1}^{N_T} \left(e^{-\beta(T-t_i)} - 1\right) \\ + \sum_{i=2}^{N_T} \sum_{j=1}^{i-1} (\log \alpha - \beta(t_i - t_j)) \Pr\left\{u_i = j | \mathcal{F}_{t_i}, \theta^{(k)}, t_i\right\} \end{aligned} \tag{3.10}$$

In order to find the parameters that maximize eq. (3.10), we set the partial derivatives of $Q(\theta|\theta^{(k)})$ with respect to each parameter equal to 0:

$$\frac{\partial Q}{\partial \lambda_\infty} = \frac{1}{\lambda_\infty} \sum_{i=1}^{N_T} \Pr\left\{u_i = i | \mathcal{F}_{t_i}, \theta^{(k)}, t_i\right\} - T = 0$$

$$\frac{\partial Q}{\partial \alpha} = \frac{1}{\beta} \sum_{i=1}^{N_T} \left(e^{-\beta(T-t_i)} - 1\right) + \frac{1}{\alpha} \sum_{i=2}^{N_T} \sum_{i=1}^{i-1} \Pr\left\{u_i = j | \mathcal{F}_{t_i}, \theta^{(k)}, t_i\right\} = 0$$

$$\frac{\partial Q}{\partial \beta} = \frac{\alpha}{\beta}\left[\sum_{i=1}^{N_T}\left(1-e^{-\beta(T-t_i)}\right)/\beta - \sum_{i=1}^{N_T}(T-t_i)e^{-\beta(T-t_i)}\right] - \sum_{i=2}^{N_T}\sum_{j=1}^{i-1}(t_i-t_j)\Pr\left\{u_i=j|\mathcal{F}_{t_i},\theta^{(k)},t_i\right\}=0$$

From the equations above we obtain a system of equations for the computation of $\theta^{(k+1)}$:

$$\lambda_\infty^{(k+1)} = \frac{\sum_{i=1}^{N_T}\Pr\left\{u_i=i|\mathcal{F}_{t_i},\theta^{(k)},t_i\right\}}{T} \tag{3.11}$$

$$\alpha^{(k+1)} = \frac{\beta^{(k+1)}\sum_{i=2}^{N_T}\sum_{i=1}^{i-1}\Pr\left\{u_i=j|\mathcal{F}_{t_i},\theta^{(k)},t_i\right\}}{\sum_{i=1}^{N_T}\left(1-e^{-\beta^{(k+1)}(T-t_i)}\right)} \tag{3.12}$$

$$\beta^{(k+1)} = \frac{\alpha^{(k+1)}\left[\sum_{i=1}^{N_T}\left(1-e^{-\beta^{(k+1)}(T-t_i)}\right)/\beta^{(k+1)} - \sum_{i=1}^{N_T}(T-t_i)e^{-\beta^{(k+1)}(T-t_i)}\right]}{\sum_{i=2}^{N_T}\sum_{j=1}^{i-1}(t_i-t_j)\Pr\left\{u_i=j|\mathcal{F}_{t_i},\theta^{(k)},t_i\right\}} \tag{3.13}$$

Only the eq. (3.11) could be solved analytically, whereas the solutions for $\alpha^{(k+1)}$ and $\beta^{(k+1)}$ could be found by using a computer software. For example by substituting the expression of $\alpha^{(k+1)}$ in (3.13), we obtain an equation only in $\beta^{(k+1)}$ which could be solved with a root-finding algorithm.

The computation burden of the exact EM algorithm is significant, because there is no analytical solution for all the parameters in the M-Step. Many approximation are introduced in the literature in order to reduce the computational intensity of the estimation.

3.2.2 An approximate EM algorithm estimation

The objective of an approximate EM algorithm is to reduce the computation burden during the M-Step. Lewis and Mohler (2011) and Olson and Carley (2013) illustrated an approximation for exponential decay Hawkes processes such that a closed-form equation could be found for all parameters in the M-Step.

The idea underlying the approximation is that the intensity contribution generated by each event dies out exponentially. Consider the following expression in the conditioned expected log-likelihood in the eq. (3.10):

$$\frac{\alpha}{\beta}\sum_{i=1}^{N_T}\left(e^{-\beta(T-t_i)}-1\right)$$

We know that if $t_1 \ll T$, $e^{-\beta(T-t_1)} \to 0$ due to the exponential decay rate β. The only events that could have a significant impact on the above expression are those observed close to T. However for large data and for big enough decay ratio β, we can approximate

$$\frac{\alpha}{\beta}\sum_{i=1}^{N_T}\left(e^{-\beta(T-t_i)}-1\right) \approx -\frac{\alpha}{\beta}N_T. \tag{3.14}$$

The approximation is more accurate when there are few observations near the end of interval $(0, T]$ and if $\beta^{-1} \ll T$ (Lewis and Mohler, 2011).

Using the approximation (3.14), we can rewrite the log-likelihood (3.8) and conditioned expected log-likelihood in terms of triggering probabilities (3.10) as:

$$\widetilde{l}(\theta) = \sum_{i:u_i=i}\log(\lambda_\infty) - \lambda_\infty T - \frac{\alpha}{\beta}N_T + \sum_{i:u_i\neq i}(\log\alpha - \beta(t_i - t_{u_i})) \tag{3.15}$$

and

$$\begin{aligned}\widetilde{Q}(\theta|\theta^{(k)}) = \log(\lambda_\infty)\sum_{i=1}^{N_T}\Pr\left\{u_i = i|\mathcal{F}_{t_i},\theta^{(k)},t_i\right\} - \lambda_\infty T - \frac{\alpha}{\beta}N_T \\ + \sum_{i=2}^{N_T}\sum_{j=1}^{i-1}(\log\alpha - \beta(t_i - t_j))\Pr\left\{u_i = j|\mathcal{F}_{t_i},\theta^{(k)},t_i\right\}\end{aligned}. \tag{3.16}$$

Setting the partial derivatives of the above equation with respect to each parameter equal to 0, we obtain the following estimates of $\theta^{(k+1)}$:

$$\lambda_\infty^{(k+1)} = \frac{\sum_{i=1}^{N_T}\Pr\left\{u_i = i|\mathcal{F}_{t_i},\theta^{(k)},t_i\right\}}{T} \tag{3.17}$$

$$\alpha^{(k+1)} = \frac{\beta^{(k+1)}\sum_{i=2}^{N_T}\sum_{i=1}^{i-1}\Pr\left\{u_i = j|\mathcal{F}_{t_i},\theta^{(k)},t_i\right\}}{N_T} \tag{3.18}$$

$$\beta^{(k+1)} = \frac{\sum_{i=2}^{N_T}\sum_{i=1}^{i-1}\Pr\left\{u_i = j|\mathcal{F}_{t_i},\theta^{(k)},t_i\right\}}{\sum_{i=2}^{N_T}\sum_{j=1}^{i-1}(t_i - t_j)\Pr\left\{u_i = j|\mathcal{F}_{t_i},\theta^{(k)},t_i\right\}} \tag{3.19}$$

The analytical solutions can be found for all the three parameters of a Hawkes process. The stop criterion is now $\widetilde{Q}(\theta|\theta^{(k+1)}) - \widetilde{Q}(\theta|\theta^{(k)}) < \epsilon$.

3.3 Simulation and comparison

In this section direct numerical maximization (DNM) method will be compared to EM algorithm using a Monte Carlo simulation. The exact EM algorithm estimation gives the same result as DNM method according to Lapham (2014), so the comparison will be carried out between the approximate EM algorithm and DNM method. Matlab is used to perform the estimation and the code could be found in the Appendix.

Two Monte Carlo simulations with 100 sampling each are performed over the interval $(0, 1000]$. The data are generated by using Hawkes process with given parameters and Ogata's thinning algorithm described in 2.2.2 . Two sets of given parameters used in the simulations are:

Set 1	$\lambda_\infty = 0.15$	$\alpha = 0.25$	$\beta = 0.5$;
Set 2	$\lambda_\infty = 0.05$	$\alpha = 0.04$	$\beta = 0.06$.

The reason of choosing these two sets of parameters is to analyse the scale effect, especially for the approximate EM estimation case.

The results are reported in table 3.1 and table 3.2. In the Monte Carlo simulation with parameters from Set 1, both methods give similar results. The average of estimates for λ_∞ is very close to the given value 0.15 and two approaches give almost the same result both in terms of average and standard deviation. The estimates for α are close to the given value 0.25 for both methods, however estimates from DNM methods give closer result to the true value on average. The estimates for β are slightly upper biased in both approaches. The estimates of DNM are proved by Ogata (1978) to be consistent, so the bias could be caused by the simulation method for which data are generated. However the true value of all 3 parameters are within a standard deviation from the average of estimates for both methods.

In the Monte Carlo simulation with parameters from Set 2, the DNM method gives estimates slightly upper biased, but still close to the true value of parameters. The results of approximate EM algorithm are more upper biased both respect to the DNM method and respect to the same algorithm applied to parameters of Set 1. This positive bias is predictable, because the approximation (3.14) works better for big value of β, but in the Set 2 $\beta = 0.06$, so the feedback effects of final points are not negligible. However the true value of parameters are within a standard deviation from the average of estimates for both methods.

	λ_∞	α	β	Execution
Given value	0.1500	0.2500	0.5000	time
DNM method bias [1]	0.0016	0.0066	0.0254	0.2424 s
DNM std.	0.0195	0.0513	0.1205	
Approx EM algorithm bias	0.0018	0.0092	0.0375	1.7428 s
Approx EM std.	0.0196	0.0509	0.1202	

TABLE 3.1: Estimates of 100 simulations over the interval $(0, 1000]$ with $\lambda_\infty = 0.15$, $\alpha = 0.25$ and $\beta = 0.5$ (Set 1). The execution times is the average time in seconds required to complete a estimation.

	λ_∞	α	β	Execution
Given value	0.0500	0.0400	0.0600	time
DNM method bias	0.0063	0.0002	0.0080	0.1327 s
DNM std.	0.0208	0.0152	0.0285	
Approx EM algorithm bias	0.0127	0.0058	0.0252	0.6981 s
Approx EM std.	0.0215	0.0157	0.0276	

TABLE 3.2: Estimates of 100 simulations over the interval $(0, 1000]$ with $\lambda_\infty = 0.05$, $\alpha = 0.04$ and $\beta = 0.06$ (Set 2). The execution times is the average time required to complete a estimation

Comparing two simulations, we found that smaller are the value of true parameters, relatively bigger are the standard deviation of estimates. This result is coherent with the results of Lewis and Mohler (2011).

From the last column of the table 3.1 and table 3.2 we can notice that the DNM method is less computationally intensive compared to the approximate EM algorithm. The first set of parameters $(0.15, 0.25, 0.50)$ requires longer execution time due to the greater number of occurrences. According to eq. (2.23), the expect number of events for the set 1 is 299, whereas the set 2 is expected to produce 145 occurrences in the same interval $(0, 1000]$.

[1] Bias of an estimator = $\mathbb{E}_\theta[\tilde{\theta}] - \theta$

Chapter 4

An application of Hawkes process

This chapter will illustrate an application of Hawkes process with financial data. The object of our analysis is the occurrences of price jumps in the stock market.

The dynamics of stock return can be decomposed into continuous Brownian motion and discrete jump component. Many authors showed the presence of jumps in finance for stock market or for short term interest rate, see for instance Aıt-Sahalia (2004) and Johannes (2004). Being able to distinguish between continuous diffusion and jumps mechanism is important as it has implications for risk management and asset allocation. (Barndorff-Nielsen and Shephard, 2006). The early implementation of jumps in finance focused on Poisson jumps, the main work in this direction has been provided by Merton (1976). Other works using stochastic differential equations with jumps could be found for example in Duffie, Pan, and Singleton (2000), Kou (2002), Schönbucher (2003) or Andersen, Bollerslev, and Diebold (2007).

The first section of this chapter will illustrate the analytical background of stochastic process with jumps and the statistical test proposed by Corsi, Pirino, and Reno (2010) for the jumps detection. The second section will illustrate the results that we obtained by using Hawkes process to model the occurrence of jumps.

4.1 Analytical background of process with jumps

Our analysis will follow Barndorff-Nielsen and Shephard (2006) and Corsi, Pirino, and Reno (2010) and it will be based on the special case where X_t, the logarithmic price of a stock, is a member of the Brownian semimartingale

with jumps ($\mathcal{BSMJ}$) class:

$$X_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s + \sum_{j=1}^{N_T} c_j \tag{4.1}$$

Where μ is a predictable locally bounded drift, the volatility σ is a càdlàg, W is a standard Brownian motion, N_t is a counting process whose intensity is an adapted stochastic process λ_t with times of jumps at $(\tau_j)_{j=1,2,...,N_T}$ and c_js are i.i.d. random variables measuring the magnitude of the jumps. The first two terms of (4.1) represents the continuous part of the process and the last term represents the discrete part.

The dynamics of the process X_t is:

$$dX_t = \mu_t dt + \sigma_t dW_t + c_j dN_t \tag{4.2}$$

The object of interest in our study is the quadratic variation $[X]$ of the process which can be defined as:

$$[X]_t = X_t^2 - X_0 - 2\int_0^t X_{s-} dX_s \tag{4.3}$$

and the quadratic variation over a time interval of length T is:

$$[X]_t^{t+T} = X_{t+T}^2 - X_t^2 - 2\int_t^{t+T} X_{s-} dX_s \tag{4.4}$$

where t indexes the day, as we follow the model proposed by Corsi, Pirino, and Reno (2010). For example, if $T = 1$, then $[X]_t^{t+1}$ measures the quadratic variation of the logarithmic price between the day t and the day $t + 1$.

In order to estimate the quadratic variation of $[X]_t^{t+T}$, we divide the interval $[t, t + T]$ in n equal sized subintervals of length δ, such that $\delta = T/n$. In this interval we have n log-return of prices defined as:

$$\Delta_{t,j} X = X_{t+j\delta} - X_{t+(j-1)\delta}, \qquad j = 1, ..., n \tag{4.5}$$

In the following work the subscript t will be omitted for simplicity and we will use the contraction $\Delta_j X = \Delta_{j,t} X$.

The quadratic variation $[X]_t^{t+T}$ could be estimated by using the realized volatility of the process:

$$\mathrm{RV}_\delta(X)_t = \sum_{j=1}^{T/\delta} (\Delta_j X)^2 \tag{4.6}$$

The realized volatility converge in probability to quadratic variation for $\delta \to 0$:

$$p - \lim_{\delta \to 0} \mathrm{RV}_\delta(X)_t \to [X]_t^{t+T} \tag{4.7}$$

See for example Barndorff-Nielsen and Shephard (2002) for more details concerning the estimation of quadratic variation.

The detection of the presence of jumps could be carry out by decomposing the quadratic variation $[X]_t^{t+T}$ into its continuous and discrete component:

$$[X]_t^{t+T} = [X^c]_t^{t+T} + [X^d]_t^{t+T} \tag{4.8}$$

where $[X^c]_t^{t+T} = \int_t^{t+T} \sigma_s^2 ds$ is the integrated variance and $[X^d]_t^{t+T} = \sum_{j=N_t}^{N_{t+T}} c_j^2$, where c_j measures the size of the jump at time τ_j. If there is no jump in the interval $[t, t+T]$, the discrete component of quadratic variation is 0, so $[X]_t^{t+T} = [X^c]_t^{t+T}$. The basic idea of jumps detection is to test whether $[X]_t^{t+T} - [X^c]_t^{t+T}$ is equal to 0.

There are different models in the literature to estimate the continuous component of the quadratic variation. We will follow threshold multipower variation (TMPV) proposed by Corsi, Pirino, and Reno (2010), which could be viewed as combination of multipower variation (MPV) introduced by Barndorff-Nielsen and Shephard (2004) and threshold model of Mancini (2009).

4.1.1 Estimation by using multipower variation

In order to decompose the continuous quadratic variation from the discrete one, Barndorff-Nielsen and Shephard (2004) introduced MPV, which is defined as:

$$\mathrm{MPV}_\delta(X)_t^{[r_1,\ldots,r_M]} = \delta^{1-\frac{1}{2}(r_1,\ldots,r_M)} \sum_{j=M}^{T/\delta} \prod_{k=1}^{M} |\Delta_{j-k+1} X|^{r_k} \tag{4.9}$$

where $r_k > 0$ are parameters indicate the power variation of order r_k-th. For the practical applications in general $M \leq 4$. In the special case of MPV where $M = 2$, it assumes the specific name of bipower variation, if $M = 3$ it is called tripower variation, and so on. MPV is used to estimate the continuous component of quadratic variation and Barndorff-Nielsen, Shephard, and Winkel (2006) show that:

$$p - \lim_{\delta \to 0} \text{MPV}_\delta(X)_t^{[r_1,\ldots,r_M]} = \left(\prod_{k=1}^{M} \mu_{r_k} \right) \int_t^{t+T} \sigma_s^{r_1+\ldots+r_M} ds \tag{4.10}$$

where $\mu_r = \mathbb{E}(|u|^r) = 2^{r/2} \dfrac{\Gamma(\frac{r+1}{2})}{\Gamma(1/2)}$, and $u \sim \mathcal{N}(0,1)$.

In general, MPV is used for the estimation of integrated variance, $\int_t^{t+T} \sigma_s^2 ds$, and integrated quarticity, $\int_t^{t+T} \sigma_s^4 ds$. We will illustrate as examples the two most relevant cases of MPV, Bipower variation (BPV) and Tripower variation (TriPV), which are used by Barndorff-Nielsen and Shephard (2006) to construct statistical test for the jump detection.

Bipower variation for the estimation of integrated variance is defined as:

$$\text{BPV}_\delta(X)_t = \mu_1^{-2} \text{MPV}_\delta(X)_t^{[1,1]} = \mu_1^{-2} \sum_{j=2}^{T/\delta} |\Delta_j X| \cdot |\Delta_{j-1} X| \underset{\delta \to 0}{\to} \int_t^{t+T} \sigma_s^2 ds \tag{4.11}$$

where $\mu_1 \simeq 0.7979$. And Tripower variation for the estimation of integrated quarticity is defined as:

$$\begin{aligned} \text{TriPV}_\delta(X)_t &= \mu_{\frac{4}{3}}^{-3} \cdot \text{MPV}_\delta(X)_t^{[\frac{4}{3},\frac{4}{3},\frac{4}{3}]} \\ &= \mu_{\frac{4}{3}}^{-3} \frac{1}{\delta} \sum_{j=3}^{T/\delta} |\Delta_j X|^{\frac{4}{3}} \cdot |\Delta_{j-1} X|^{\frac{4}{3}} \cdot |\Delta_{j-2} X|^{\frac{4}{3}} \underset{\delta \to 0}{\to} \int_t^{t+T} \sigma_s^4 ds \end{aligned} \tag{4.12}$$

where $\mu_{4/3} \simeq 0.8309$. Both BPV and TriPV converge in probability as $\delta \to 0$. Barndorff-Nielsen and Shephard (2006) argue that under sufficient regularity, frictionless market conditions and in the absence of jumps in the price path,

$$z = \delta^{-\frac{1}{2}} \frac{(\text{RV}_\delta(X)_T - \text{BPV}_\delta(X)_T) \times \text{RV}_\delta(X)_T^{-1}}{\sqrt{\left(\frac{\pi^2}{4} + \pi - 5\right) \times \max\left\{1, \frac{\text{TriPV}_\delta(X)_T}{(\text{BPV}_\delta(X)_T)^2}\right\}}} \tag{4.13}$$

where $z \to \mathcal{N}(0,1)$ in distribution as $\delta \to 0$. Hence, an abnormally large

value of standardized difference between realized volatility and bipower variation is an evidence of a jump over the interval considered.

4.1.2 Estimation by using threshold function

An alternative estimator of integrated variance and integrated quatircity based on a threshold function $\Theta(\delta)$ is provided by Mancini (2009). The estimator are defined as follows:

$$\text{TIV}_\delta(X)_t = \sum_{j=1}^{T/\delta} |\Delta_j X|^2 I_{\{|\Delta_j X|^2 \leq \Theta(\delta)\}}, \tag{4.14}$$

and

$$\text{TIQ}_\delta(X)_t = \frac{1}{3\delta} \sum_{j=1}^{T/\delta} |\Delta_j X|^4 I_{\{|\Delta_j X|^2 \leq \Theta(\delta)\}}, \tag{4.15}$$

where $\Theta(\delta)$ is a deterministic threshold function of the lag δ between the observations, such that:

$$\lim_{\delta \to 0} \Theta(\delta) = 0, \qquad \text{and } \lim_{\delta \to 0} \frac{\delta \log \frac{1}{\delta}}{\Theta(\delta)} = 0. \tag{4.16}$$

Mancini (2009) showed that for $\delta \to 0$ the estimators converge in probability to the true value:

$$p - \lim_{\delta \to 0} \text{TIV}_\delta(X)_t = \int_t^{t+T} \sigma_s^2 ds, \qquad p - \lim_{\delta \to 0} \text{TIQ}_\delta(X)_t = \int_t^{t+T} \sigma_s^4 ds.$$

4.1.3 Estimation by using threshold multipower variation

Although BPV (4.11) and TriPV (4.12) are consistent estimators of integrated variance and integrated quarticity for $\delta \to 0$, Corsi, Pirino, and Reno (2010) showed that in finite sample they are significantly upper biased in presence of jumps and this causes an underestimation of the jump component. This problem cannot be solved by simply reducing δ, because for small δ the market microstructure noise would make unreliable the estimation of realized volatility (4.6), which assume the absence of noise (Hansen and Lunde, 2006). The estimators TIV (4.14) and TIQ (4.15) are potentially less biased, but they have the problem of being sensitive to the specification of the threshold. (Corsi, Pirino, and Reno, 2010)

In order to overcome the problems described above, Corsi, Pirino, and Reno (2010) introduced estimators of integrated variance and integrated quarticity by using threshold multipower variation (TMPV), which could be seen as a combination of mulpipower variation and threshold function. The thresold multipower variation is defined as:

$$\text{TMPV}_{\delta}(X)_t^{[r_1,\ldots,r_M]} = \delta^{1-\frac{1}{2}(r_1,\ldots,r_M)} \sum_{j=M}^{T/\delta} \prod_{k=1}^{M} |\Delta_{j-k+1}X|^{r_k} I_{\{|\Delta_{j-k+1}X|^2 \leq \vartheta_{j-k+1}\}} \tag{4.17}$$

where ϑ is the threshold function, but in contrast to $\Theta(\delta)$ of Mancini (2009), ϑ is not a function of δ. The exact expression of ϑ will be provided below.

As in the case of MPV (4.10), Corsi, Pirino, and Reno (2010) showed the following convergence in probability as $\delta \to 0$:

$$p - \lim_{\delta \to 0} \text{TMPV}_{\delta}(X)_t^{[r_1,\ldots,r_M]} \to \left(\prod_{k=1}^{M} \mu_{r_k} \int_t^{t+T} \sigma_s^{r_1,\ldots,r_M} \right) \tag{4.18}$$

Also, they showed that TMPV has the same distribution of MPV if $\delta \to 0$. However, when δ is large enough to avoid the microstructure noise, authors argue that TMPV model has best performance.

The threshold ϑ in eq. (4.17) is defined as a multiple of the local variance. The estimation of local variance is carry out by using a local linear filter of length $2L+1$ adjusted for the presence of jumps and by iterating in R:

$$\hat{V}_t^R = \frac{\sum_{i=-L,\, i\neq -1,0,1}^{L} K\left(\frac{i}{L}\right) (\Delta_{t+i}X)^2 I_{\left\{(\Delta_{t+i}X)^2 \leq c_V^2 \cdot \hat{V}_{t+i}^{R-1}\right\}}}{\sum_{i=-L,\, i\neq -1,0,1}^{L} K\left(\frac{i}{L}\right) I_{\left\{(\Delta_{t+i}X)^2 \leq c_V^2 \cdot \hat{V}_{t+i}^{R-1}\right\}}} \qquad R = 1, 2, \ldots \tag{4.19}$$

where $\hat{V}_t^0 = +\infty$, so that all the observations can be used in the first step; c_V is set by authors equal to 3; L is the number of adjacent observations considered for the estimation of the local variance, with the exclusion of the immediate adjacent returns where $i = -1, 0, 1$; $K(.)$ is the a Gaussian kernel function where:

$$K(y) = -\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right),$$

the choice of the kernel $K(.)$ is not relevant for the application of our model

(Wand and Jones, 1994). Jumps are detected at each iteration by the condition $(\Delta_{t+i}X)^2 > c_V^2 \cdot \hat{V}_{t+i}^{R-1}$ and removed by means of the indicator function. The iterations stop when $\hat{V}^R = \hat{V}^{R-1}$, which always happens after 2 or 3 iterations at most in the high frequency data.

The threshold ϑ of TMPV (4.17) is defined as:

$$\vartheta_t = c_\vartheta^2 \times \hat{V}_t^R \tag{4.20}$$

where c_ϑ is a dimensionless parameter. Corsi, Pirino, and Reno (2010) showed that the estimators are robust to the change of c_ϑ and they set $c_\vartheta = 3$ for the practical applications.

Corsi, Pirino, and Reno (2010) proposed a jump detection test which is a improvement of z statistic test (4.13) proposed by Barndorff-Nielsen and Shephard (2006). here, TMPV estimators are used instead of MPV estimators. However, in the application of TMPV when $|\Delta_j X|^2 > \vartheta_j$, the corresponding return is removed by the means of the indicator function. This leads to a negative bias of TMPV under the null hypothesis of no jumps, because variation larger than the threshold exist also in the absence of jumps. This issue can be effectively overcome by setting $|\Delta_j X|^r$ equal to its expected value when $|\Delta_j X|^2 > \vartheta_j$. Assuming that $\Delta_j X \sim \mathcal{N}(0, \sigma^2)$ we have:

$$\mathbb{E}[|\Delta_j X|^r \,|(\Delta_j X)^2 > \vartheta] = \frac{1}{2N\left(-\frac{\sqrt{\vartheta}}{\sigma}\right)\sqrt{\pi}} \left(2\sigma^2\right)^{\frac{r}{2}} \Gamma\left(\frac{r+1}{2}, \frac{\vartheta}{2\sigma^2}\right) \tag{4.21}$$

where $N(x)$ is the standard normal cumulative function and $\Gamma(\alpha, x)$ is the upper incomplete gamma function [1]. We can use the realized local variance (4.19) for the estimation of σ^2, so that $\hat{\sigma}^2 = \vartheta / c_\vartheta^2$ from eq. (4.20). Then, the correct TMPV (C-TMPV) estimator is defined as:

$$\text{C-TMPV}_\delta(X)_t^{[r_1,\ldots,r_M]} = \delta^{1-\frac{1}{2}(r_1+\ldots+r_M)} \sum_{j=M}^{T/\delta} \prod_{k=1}^{M} Z_{r_k}(\Delta_{j-k+1}X, \vartheta_{j-k+1}) \tag{4.22}$$

[1] Precisely, $N(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$, and $\Gamma(\alpha, x) = \int_x^{+\infty} s^{\alpha-1} e^{-s} ds$

where $Z_r(\Delta X, \vartheta)$ is defined according to (4.21):

$$Z_r = \begin{cases} \frac{1}{2N(-c_\vartheta)\sqrt{\pi}} \left(2\frac{\vartheta}{c_\vartheta^2}\right)^{\frac{r}{2}} \Gamma\left(\frac{r+1}{2}, \frac{c_\vartheta^2}{2}\right) & \text{if } (\Delta X)^2 > \vartheta \\ |\Delta X|^r & \text{if } (\Delta X)^2 \leq \vartheta \end{cases} \tag{4.23}$$

As in the z statistic test proposed by Barndorff-Nielsen and Shephard (2006), the relevant estimators for a corrected jump detection test are corrected threshold bipower variation (C-TBPV) with $[r_1, r_2] = [1, 1]$ and corrected threshold tripower variation (C-TTriPV) with $[r_1, r_2, r_3] = [\frac{4}{3}, \frac{4}{3}, \frac{4}{3}]$. These estimators are defined as following:

$$\text{C-TBPV}_\delta(X)_t = \mu_1^{-2}\text{C-TMPV}_\delta(X)_t^{[1,1]} = \mu_1^{-2} \sum_{j=2}^{T/\delta} Z_1(\Delta_j X, \vartheta_j) Z_1(\Delta_{j-1} X, \vartheta_{j-1}), \tag{4.24}$$

and

$$\begin{aligned} \text{C-TTriPV}_\delta(X)_t &= \mu_{\frac{4}{3}}^{-3}\text{C-TMPV}_\delta(X)_t^{[\frac{4}{3},\frac{4}{3},\frac{4}{3}]} \\ &= \mu_{\frac{4}{3}}^{-3} \frac{1}{\delta} \sum_{j=3}^{T/\delta} Z_{\frac{4}{3}}(\Delta_j X, \vartheta_j) Z_{\frac{4}{3}}(\Delta_{j-1} X, \vartheta_{j-1}) Z_{\frac{4}{3}}(\Delta_{j-2} X, \vartheta_{j-2}). \end{aligned} \tag{4.25}$$

Corsi, Pirino, and Reno (2010) proposed the corrected threshold statistical test (C-Tz) based on z test of Barndorff-Nielsen and Shephard (2006). The C-Tz test statistic is defined as:

$$\text{C-Tz} = \delta^{-\frac{1}{2}} \frac{(\text{RV}_\delta(X)_T - \text{C-TBPV}_\delta(X)_T) \times \text{RV}_\delta(X)_T^{-1}}{\sqrt{\left(\frac{\pi^2}{4} + \pi - 5\right) \times \max\left\{1, \frac{\text{C-TTriPV}_\delta(X)_T}{(\text{C-TBPV}_\delta(X)_T)^2}\right\}}} \tag{4.26}$$

the authors showed that C-Tz has more power and it is less biased compared to z test statistic of (4.13). Also they proved that C-Tz $\to \mathcal{N}(0, 1)$ in law as $\delta \to 0$.

In the following applications we will use C-Tz test statistic for the jumps detection.

4.2 An example of Hawkes model with financial data

As an example of application, we used Hawkes process to model the intensity of price jumps for some stocks. In the literature there is evidence that price jumps are followed by increased volatility, see for example Joulin et al. (2008) or Corsi, Pirino, and Reno (2010). Here, we want to verify if price jumps could be modelled as a self-exciting stochastic process, i.e. if a price jump has positive impact on the probability of future price jumps.

The stocks being considered in our example are: Citigroup (C), General Electric (GE), Pfizer (PFE), Cisco (CSCO) and Microsoft (MSFT). We analysed the period from January 2$^{\text{nd}}$ 1998 to June 5$^{\text{th}}$ 2015, for the total of 4384 trading days. For every trading day, C-Tz statistics (4.26) is used to detect jumps with a confidence level of 99% (C-Tz $> \Phi_{99\%} = 2.58$). Within a single trading day, $\delta = 5$ minutes has been set in order to avoid microstructure noise. Trades start at 9:30 and end at 16:30, but in order to avoid noise in the opening, we consider the first price at 9:35 rather than 9:30. We recorded 77 prices per each day for a total of 76 log-returns.

We used both the approximated EM algorithm and DNM method to estimate the parameters of Hawkes model according to eq. (2.9):

$$\lambda_t = \lambda_\infty + \sum_{j:t_j<t} \alpha e^{-\beta(t-t_j)}$$

the results are reported in the Table 4.1. Whereas EM algorithm gives meaningful estimates for all stocks, DNM method has some issue in the estimation of CSCO and MSFT, so we will not use these results for the comparison. For C, GE and PFE the estimated $\hat{\alpha}$ show similar values for both methods. EM algorithm in general estimates a higher decay rate $\hat{\beta}$ compared to DNM method, but it is compensated by a lower background intensity $\hat{\lambda}_\infty$. This means EM algorithm attribute a relatively bigger account to the background process, which can be also noticed by the lower estimated branching ratio $\hat{\eta}$ and lower expected percentage of endogenous jumps compared to DNM method. However, for C, GE and PFE both methods claim that there is a significant percentage of jumps caused endogenously , around 30% according to EM algorithm and more than 40% according to DNM method.

The Figures 4.1 and 4.2 show the dynamics of intensity λ_t. The red line is

	$\hat{\lambda}_\infty$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\eta}$	N_T	$\mathbb{E}[N_T]$	Exp endogenous jumps %
C							
EM	0.0840	0.0057	0.0185	0.3101	534	531.04	30.62%
DNM	0.0614	0.0053	0.0104	0.5088	534	535.82	49.74%
GE							
EM	0.0908	0.0085	0.0288	0.2944	564	562.13	29.20%
DNM	0.0744	0.0081	0.0188	0.4292	564	566.23	42.39%
PFE							
EM	0.0987	0.0068	0.0245	0.2753	597	594.89	27.27%
DNM	0.0762	0.0058	0.0130	0.4494	597	597.65	44.14%
CSCO							
EM	0.1067	0.0044	0.0393	0.1108	526	525.62	11.01%
~~DNM~~	~~0.1200~~	~~-22227~~	~~332613~~	~~-0.0668~~	~~526~~	~~493.05~~	~~-6.68%~~
MSFT							
EM	0.1186	0.0056	0.0327	0.1704	627	626.10	16.92%
~~DNM~~	~~0.1430~~	~~-31271~~	~~246503~~	~~-0.1269~~	~~627~~	~~556.42~~	~~-12.69%~~

TABLE 4.1: Results of EM and DNM estimation. $\hat{\lambda}_\infty$, $\hat{\alpha}$ and $\hat{\beta}$ are estimated parameters of Hawkes process as in eq. (2.9). η is the branching ratio as in eq. (2.10). N_T is the total number of jumps observed. $\mathbb{E}[N_T]$ is the expected total number of jumps calculated using eq. (2.23). The percentage of expected endogenous jumps is calculated as $(1 - \text{Exp. background jumps}/\mathbb{E}[N_T])$, where Exp. background jumps is $\lambda_\infty \times T$ where $T = 4384$.

the fit of DNM method and the blue line is the one from EM algorithm. Two methods provide similar results graphically for C, GE and PFE. Whereas the best estimation of DNM method for CSCO and MSFT is a homogeneous Poisson process with intensity λ_∞.

In order to validate the self-exciting nature of jumps process, we performed Ljung-Box test (Ljung and Box, 1978) and Christoffersen independent test (Christoffersen, 1998) to find the evidence of autocorrelation in the jumps process. We applied both statistical tests to the vector h, which is a binary vector of 4384 elements, where every element is associated to a trading day and $h_t = 1$ if there is a jump in the day t and $h_t = 0$ otherwise. The detailed description of statistical tests could be found in the Appendix and the test results are reported in Table 4.2 and Table 4.3. Ljung-Box test shows evidence of serial correlation for C, GE, and PFE with 95% confidence and supports the presence of self-exciting nature for these stocks, but the null hypothesis of serial independence cannot be rejected for CSCO and MSFT. The latter result is coherent with the estimation of DNM method, according to which jumps of CSCO and MSFT are distributed following a homogeneous Poisson process of intensity λ_∞. The result of Christoffersen independent test reported in Table 4.3 shows that the jumps of all stocks are distributed independently, so this statistical test does not support the existence of self-exciting nature in the jumps process.

Ljung-Box test	p-value	serial independence
C	0.0415	no
GE	0.0021	no
PFE	0.0005	no
CSCO	0.4154	yes
MSFT	0.1649	yes

TABLE 4.2: Ljung-Box test with confidence of 95%. The null of serial independence is rejected at 95% for C, GE and PFE. Whereas the null cannot be rejected for CSCO and MSFT

However, both Christoffersen and Ljung-Box test have some issues in our application. The independent test of Christoffersen is designed for binary data, but it tests the independence of output of day t from day $t-1$. Each realisation of Hawkes process change all the future intensity with an exponential decay rate, but Christoffersen's test cannot capture the effects

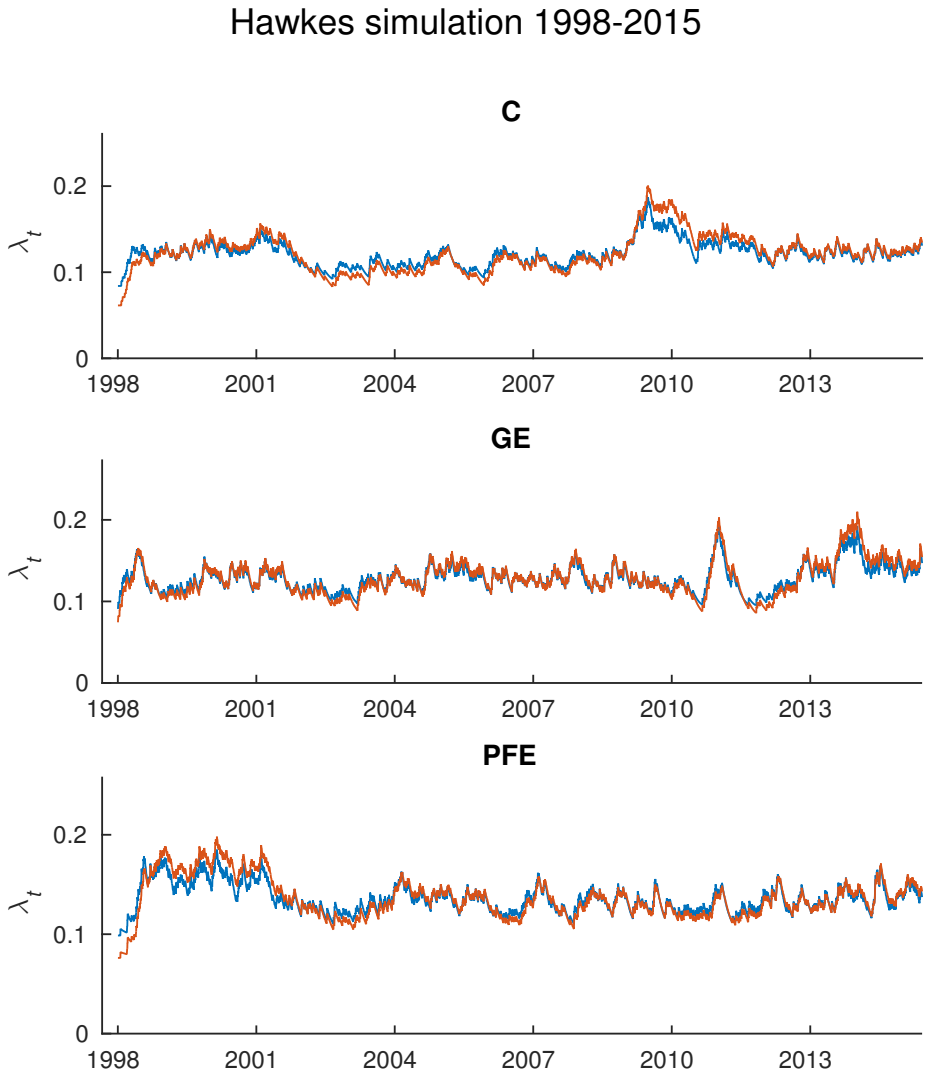

FIGURE 4.1: Intensity dynamics of C, GE and PFE using parameters estimated by EM (blue line) and by DNM (red line).

Christoffersen test	p-value	serial independence
C	0.5915	yes
GE	0.4967	yes
PFE	0.612	yes
CSCO	0.3358	yes
MSFT	0.2687	yes

TABLE 4.3: Christoffersen independent test with confidence of 95%. p-values suggest serial independence for all the stocks.

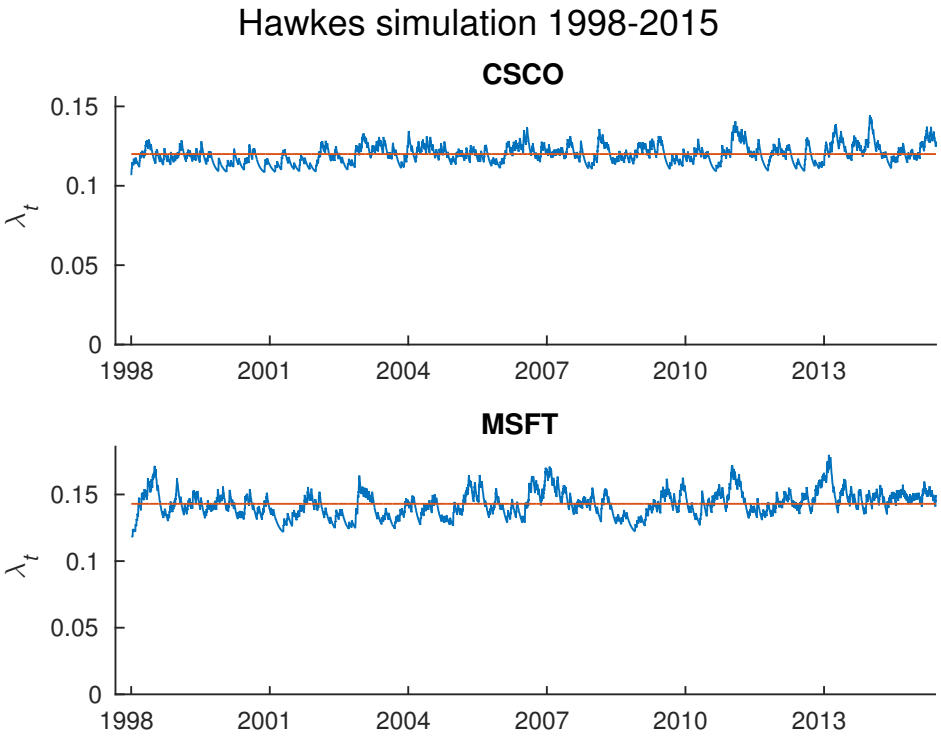

FIGURE 4.2: Intensity dynamics of CSCO and MSFT using parameters estimated by EM (blue line) and by DNM (red line). Note that the intensity of DNM model is constant.

of lags bigger than 1 and this could bring to an underestimation of serial correlation. On the other hand, Ljung-Box tests the overall randomness till 20 lags (Matlab default setting), but it performs poorly with binary data when the probability of output are unequal (Burns, 2002). In our example the probability of a jump is around 12-14%, whereas the probability of no jump is around 86-88%.

Considering the results and the practical issues of independent tests, we conclude that Hawkes processes could be a possible model at least for the stocks C, GE and PFE, whereas there is no strong evidence of self-exciting jumps process in the technology stocks CSCO and MSFT.

Chapter 5

Conclusion

The application of Hawkes process in finance is a very recent subject in the literature. This thesis provides an analytical background of univariate unmarked Hawkes process and it shows an example of its application. Despite the fitness of self-exciting model for the occurrences of price jumps is not robust in our analysis, Hawkes process can still provide valuable informations for the investors. A more accurate model could be accomplished by using multivariate marked hawkes process, for example it can take into account the sign and the size of jumps.

The potential application of Hawkes process in other area of finance seems promising. There is a growing literature of self-exciting model in high frequency trading and financial contagion. Future progress could be made in these directions by reducing the computational burden of estimation and by implementing forecast model of financial market.

Appendix A

All the simulations, estimations and data analysis in this work have been made by using Matlab.

A.1 Simulation of Hawkes process

We assumed that $\lambda_0 = \lambda_\infty$ and the function for the conditional intensity λ_t uses equation 2.9:

$$\lambda_t = \lambda_\infty + \sum_{j:t_j<t} \alpha e^{-\beta(t-t_j)}$$

The code for the computation of λ_t in Matlab is the following:

```
function[lambda] = cond_int(t,H,mu,alpha,Beta)
x = length(t);
lambda = mu*ones(x,1);
    for i = 1:x
        h = H;
        h = h(h < t(i));
        if ~isempty(h)
            lambda(i) = lambda(i) + alpha*sum(exp(-Beta
                *(t(i)-h)));
    end
end
```

Where H represents the history of the process. *mu*, *alpha* and *Beta* are parameters λ_∞, α and β. *Length(t)* is always equal to 1 for the computation of the conditional intensity, but in order to draw the figure 2.3, x evenly spaced points are generated in the interval $(0, T]$ using the function *linspace(0,T,x)*, in this case t and *lambda* are vectors of length x containing the conditional intensity for each of x points in the time axis.

The matlab code for the simulation of Hawkes process using Ogata's thinning algorithm as described in Daley and Vere-Jones (2002) is the following:

```
function[h] = Ogata(T,mu,alpha,Beta)
kappa = 0.5;
t = 0;
h = [];
while t < T
    Mt = cond_int(t + 0.00001,h,mu,alpha,Beta);
    Lt = kappa*Mt;
    R = exprnd (1/Mt);
    if R > Lt
        t = t + Lt;
    else
        U = rand;
        if U > (cond_int(t+R,h,mu,alpha,Beta) / Mt)
            t = t + R;
        else
            t = t + R;
            h = [h t];
        end
    end
end
end
```

Where *Mt* = λ_{t+} and *kappa* has been set equal to 0.5 according to Daley and Vere-Jones (2002) in order to give a reasonable compromise between setting the bound too high, so generating excessive trial points, and setting it too low, thus requiring too many iterations. The output of this function is a vector *h* containing the times of the simulated point events, where *length(h)* = N_T. This code has been used to simulate the Hawkes process showed in Figure 2.3.

A.2 Parameters estimation

A.2.1 DNM of log-likelihood

The log-likelihood is computed according to eq. (3.2):

$$l(\theta) = T - T\lambda_\infty - \frac{\alpha}{\beta}\sum_{i=1}^{N_T}\left(1 - e^{-\beta(T-t_i)}\right) + \sum_{i=1}^{N_T}\log\left(\lambda_\infty + \alpha A(i)\right).$$

and parameters are estimated by minimizing the negative of log-likelihood using `fminunc`.

The computation of $-l(\theta)$ is the following:

```
function [neg_log_like] = log_likelihood(par,H,T)
mu = par(1); alpha = par(2); Beta = par(3);
NT = length(H);
sum1 = sum(1-exp(-Beta*(T-H)));
A = 0;
for i = 2:NT
    A(i) = exp(-Beta*(H(i)-H(i-1)))*(1+A(i-1));
end
sum2 = sum(log(mu+alpha*A));
neg_log_like = - (T - mu*T - alpha/Beta*sum1 + sum2);
```

where *NT* is the counting measure of interval $(0, T]$, *A(i)* is the recursive component, and

$$\text{sum1} = \sum_{i=1}^{N_T}\left(1 - e^{-\beta(T-t_i)}\right)$$

$$\text{sum2} = \sum_{i=1}^{N_T}\log\left(\lambda_\infty + \alpha A(i)\right)$$

The output of the above function is the negative of log-likelihood, which is minimized by the following function to estimate the parameters:

```
function[theta] = DNM(H,T)
x0=[0.1  0.1  0.1];
f = @(theta)log_likelihood(theta,H,T);
[theta]=fminunc(f,x0);
end
```

where *x0* is the initial guess and *theta* is a vector of 3 elements containing estimated parameters $[\hat{\lambda}_\infty\ \hat{\alpha}\ \hat{\beta}]$.

A.2.2 Approximate EM algorithm

The code for approximate EM algorithm is the following:

```
function [theta] = EM (H,T)
mu=0.1; alpha=0.1; Beta=0.1;
NT=length(H);
epsilon=1;
k=2; Q(1)=0;
p(1,1)=1;
while abs(epsilon)>=0.01 | k<100
    for i=2:NT;
        timediff(i,1:i-1)=H(i)-H(1:i-1);
        sumd=sum(alpha*exp(-Beta*timediff(i,1:i-1)));
        p(i,1:i-1)= alpha*exp(-Beta*timediff(i,1:i-1))
            / (mu + sumd);
        p(i,i)= mu / (mu + sumd);
    end
    timediff(:,NT)=0;
    q=tril(p,-1);
    sum1=sum(diag(p));
    sum2=sum(sum(q));
    sum3=sum(sum(q.*timediff));

    mu=sum1/T;
    alpha= sum2^2 / (NT * sum3);
    Beta= sum2 / sum3;

    sum4= sum(sum(q.*(log(alpha)-Beta*timediff)));
    Q(k) = log(mu)*sum1 - mu*T - alpha*NT/Beta + sum4;
    epsilon=Q(k)-Q(k-1);
    k=k+1;
end
theta = [mu alpha Beta];
```

The underlying theory and the notations could be found in Section 3.2.2. In particular, in our code the initial value of iteration index *k* has been set equal to 2 and *Q(1)=0* in order to allow the expression *Q(k)-Q(k-1)*. *p* is the matrix of triggering probability, and the sums in the code equal to the following expressions:

$$\text{sum1} = \sum_{i=1}^{N_T} \Pr\left\{u_i = i|\mathcal{F}_{t_i}, \theta^{(k)}, t_i\right\}$$
$$\text{sum2} = \sum_{i=2}^{N_T}\sum_{i=1}^{i-1} \Pr\left\{u_i = j|\mathcal{F}_{t_i}, \theta^{(k)}, t_i\right\}$$
$$\text{sum3} = \sum_{i=2}^{N_T}\sum_{j=1}^{i-1} (t_i - t_j)\Pr\left\{u_i = j|\mathcal{F}_{t_i}, \theta^{(k)}, t_i\right\}$$
$$\text{sum4} = \sum_{i=2}^{N_T}\sum_{j=1}^{i-1} (\log\alpha - \beta(t_i - t_j))\Pr\left\{u_i = j|\mathcal{F}_{t_i}, \theta^{(k)}, t_i\right\}$$

In the exact EM algorithm and in the most cases of approximate EM algorithm the conditioned expected log-likelihood is an increasing function of number of iterations, so that the condition $\widetilde{Q}(\theta|\theta^{(k+1)}) > \widetilde{Q}(\theta|\theta^{(k)})$ is always true, see for example Figure A.1. But in the approximate EM algorithm, especially when the value of parameters are close to 0, the conditioned expected log-likelihood could be non monotonic as showed in Figure A.2. The maximum reached after around 20 iterations gives apparently the best estimates, but indeed the estimates are closer to the true values when the function approach to the horizontal asymptote as number of iterations increases. The same observation has been made also by Lapham (2014).

In order to make sure that the algorithm does not converge to the apparent best estimates, the stop condition has been set $\widetilde{Q}(\theta|\theta^{(k+1)}) - \widetilde{Q}(\theta|\theta^{(k)}) < 0.01$ and $k \geq 100$, allowing at least 98 iterations (k = number of iteerations -1).

A.3 Independent tests

A.3.1 Ljung-Box test

Ljung-Box test (Ljung and Box, 1978) is defined as following:

H_0: The data are independently distributed;

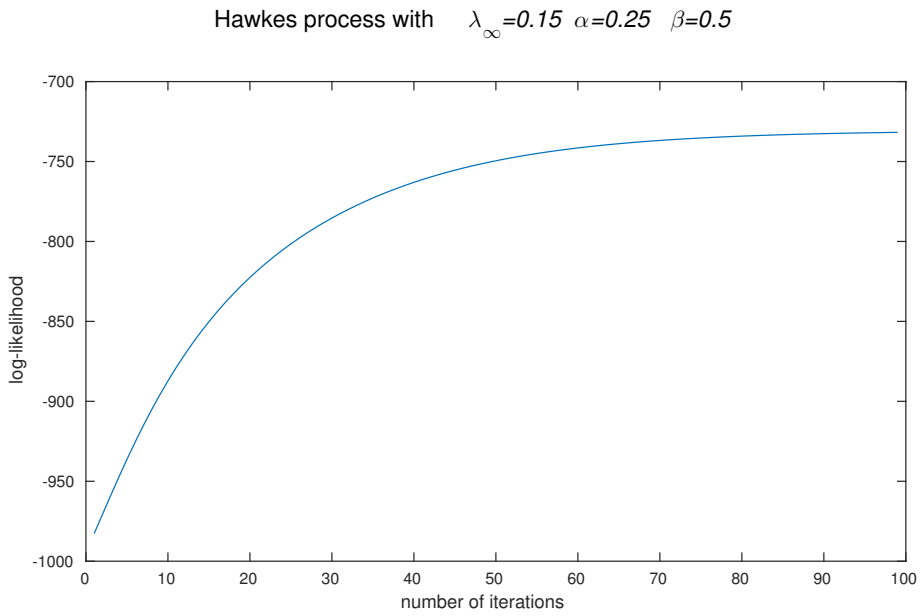

FIGURE A.1: Conditioned expected log-likelihood is strictly increasing in number of iterations.

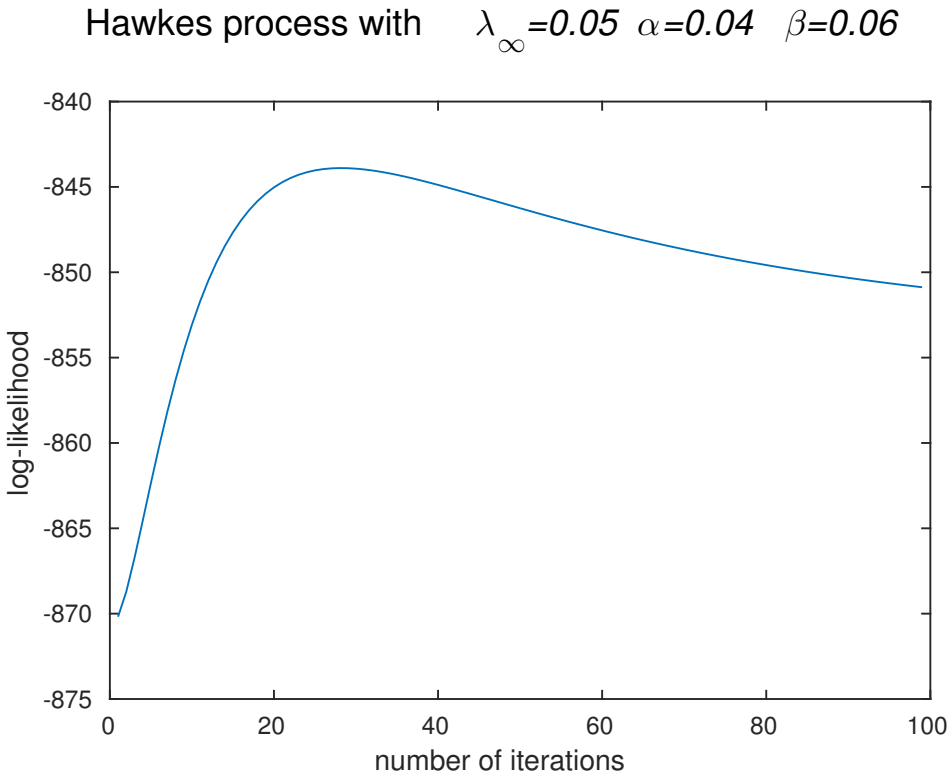

FIGURE A.2: Conditioned expected log-likelihood could be non monotonic.

H_1: The data are not independently distributed. There is serial correlation in the data.

The test statistic is defined as:

$$Q = T(T+2)\sum_{k=1}^{h}\frac{\hat{\rho_k^2}}{T-k} \tag{A.1}$$

where T is the sample size, $\hat{\rho_k}$ is the estimated autocorrelation at lag k, and h is the number of lags being tested. In the default setting of Matlab $h = \min[20, T-1]$, in our analysis T is the total number trading days considered, which equal to 4384, so $h = \min[20, 4383] = 20$.

The estimated autocorrelation is computed as:

$$\rho_k = \frac{\sum_{t=k+1}^{n} r_t r_{t-k}}{\sum_{t=1}^{n} r_t^2}$$

where r_t is the residual of $t-$th element of the sample.

Under the null hypothesis the test statistic is distributed as $\chi^2_{(h)}$. For significant level a, we reject H_0 if:

$$Q > \chi^2_{1-a,h}.$$

A.3.2 Christoffersen independent test

Christoffersen independent test (Christoffersen, 1998) is defined as following:

H_0: There is independence from one period to the next;

H_1: There is no independence from one period to the next. Serial correlation with one lag.

Let ${}^t i = 0$ if there is no jump in the day t and ${}^t i = 1$ if there is a jump in the day t. We define the conditional probability of a jump as:

$$q_0 = \Pr({}^t i = 1 \big|^{t-1} i = 0)$$

$$q_1 = \Pr({}^t i = 1 \big|^{t-1} i = 1)$$

In our analysis there are $T = 4384$ trading days, so there are 4383 consecutive observations $({}^{t-1}i, {}^{t}i)$ such that:

$$a_{00} + a_{01} + a_{10} + a_{11} = 4383$$

where a_{00} is the total number of consecutive observations of type $(0, 0)$, a_{10} is the total number of consecutive observations of type $(1, 0)$ and so on (See Table A.1 for the contingency table). The test want to verify if $\hat{q}_0 = \hat{q}_1 = \hat{q}$, where $\hat{q}$ is the estimated probability of a jump under the null:

$$\hat{q} = \frac{a_{01} + a_{11}}{a_{00} + a_{01} + a_{10} + a_{11}}.$$

If the null does not hold, we estimate the conditional probability of a jump at day t as:

$$\hat{q_0} = \frac{a_{01}}{a_{00} + a_{01}}, \qquad \text{and} \qquad \hat{q_1} = \frac{a_{11}}{a_{10} + a_{11}}.$$

The test statistic is a likelihood ratio test, where the likelihood is defined as:

$$\Lambda = \frac{(1 - \hat{q})^{a_{00}+a_{10}} (\hat{q})^{a_{01}+a_{11}}}{(\hat{q}_0)^{a_{01}} (1 - \hat{q}_0)^{a_{00}} (\hat{q}_1)^{a_{11}} (1 - \hat{q}_1)^{a_{10}}} \tag{A.2}$$

$-2 \log(\Lambda)$ is distributed as $\chi^2_{(1)}$. For significant level a, we reject H_0 if:

$$-2 \log(\Lambda) > \chi^2_{1-a,1}.$$

	${}^{t}i = 0$	${}^{t}i = 1$	
${}^{t-1}i = 0$	a_{00}	a_{01}	$a_{00} + a_{01}$
${}^{t-1}i = 1$	a_{10}	a_{11}	$a_{10} + a_{11}$
	$a_{00} + a_{10}$	$a_{01} + a_{11}$	$a_{00} + a_{01} + a_{10} + a_{11}$

TABLE A.1: Contingency table of Christoffersen independence test

Bibliography

[1] Yacine Aıt-Sahalia. "Disentangling diffusion from jumps". In: *Journal of Financial Economics* 74.3 (2004), pp. 487–528.

[2] Torben G Andersen, Tim Bollerslev, and Francis X Diebold. "Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility". In: *The review of economics and statistics* 89.4 (2007), pp. 701–720.

[3] Emmanuel Bacry, Khalil Dayri, and Jean-François Muzy. "Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data". In: *The European Physical Journal B* 85.5 (2012), pp. 1–12.

[4] Emmanuel Bacry et al. "Modelling microstructure noise with mutually exciting point processes". In: *Quantitative Finance* 13.1 (2013), pp. 65–77.

[5] Ole E Barndorff-Nielsen and Neil Shephard. "Econometrics of testing for jumps in financial economics using bipower variation". In: *Journal of financial Econometrics* 4.1 (2006), pp. 1–30.

[6] Ole E Barndorff-Nielsen and Neil Shephard. "Estimating quadratic variation using realized variance". In: *Journal of Applied econometrics* 17.5 (2002), pp. 457–477.

[7] Ole E Barndorff-Nielsen and Neil Shephard. "Power and bipower variation with stochastic volatility and jumps". In: *Journal of financial econometrics* 2.1 (2004), pp. 1–37.

[8] Ole E Barndorff-Nielsen, Neil Shephard, and Matthias Winkel. "Limit theorems for multipower variation in the presence of jumps". In: *Stochastic processes and their applications* 116.5 (2006), pp. 796–806.

[9] Sean Borman. "The expectation maximization algorithm-a short tutorial". In: *Submitted for publication* (2004), pp. 1–9.

[10] Clive G Bowsher. "Modelling security market events in continuous time: Intensity based, multivariate point process models". In: *Journal of Econometrics* 141.2 (2007), pp. 876–912.

[11] Patrick Burns. "Robustness of the Ljung-Box test and its rank equivalent". In: *Available at SSRN 443560* (2002).

[12] Valérie Chavez-Demoulin, Anthony C Davison, and Alexander J McNeil. "Estimating value-at-risk: a point process approach". In: *Quantitative Finance* 5.2 (2005), pp. 227–234.

[13] Peter F Christoffersen. "Evaluating interval forecasts". In: *International economic review* (1998), pp. 841–862.

[14] Fulvio Corsi, Davide Pirino, and Roberto Reno. "Threshold bipower variation and the impact of jumps on volatility forecasting". In: *Journal of Econometrics* 159.2 (2010), pp. 276–288.

[15] David Roxbee Cox and Valerie Isham. *Point processes*. Vol. 12. CRC Press, 1980.

[16] José Da Fonseca and Riadh Zaatour. "Hawkes process: Fast calibration, application to trade clustering, and diffusive limit". In: *Journal of Futures Markets* 34.6 (2014), pp. 548–579.

[17] Daryl J Daley. "Various concepts of orderliness for point processes". In: *Stochastic Geometry* (1974), pp. 148–161.

[18] DJ Daley and D Vere-Jones. *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. Springer Science & Business Media, 2002.

[19] Angelos Dassios, Hongbiao Zhao, et al. "Exact simulation of Hawkes process with exponentially decaying intensity". In: *Electronic Communications in Probability* 18.62 (2013), pp. 1–13.

[20] Mark HA Davis. "Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models". In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1984), pp. 353–388.

[21] Arthur P Dempster, Nan M Laird, and Donald B Rubin. "Maximum likelihood from incomplete data via the EM algorithm". In: *Journal of the royal statistical society. Series B (methodological)* (1977), pp. 1–38.

[22] Darrell Duffie, Jun Pan, and Kenneth Singleton. "Transform analysis and asset pricing for affine jump-diffusions". In: *Econometrica* 68.6 (2000), pp. 1343–1376.

[23] Paul Embrechts, Thomas Liniger, Lu Lin, et al. "Multivariate Hawkes processes: an application to financial data". In: *Journal of Applied Probability* 48 (2011), pp. 367–378.

[24] Eymen Errais, Kay Giesecke, and Lisa R Goldberg. "Affine point processes and portfolio credit risk". In: *SIAM Journal on Financial Mathematics* 1.1 (2010), pp. 642–665.

[25] Vladimir Filimonov and Didier Sornette. "Quantifying reflexivity in financial markets: Toward a prediction of flash crashes". In: *Physical Review E* 85.5 (2012), p. 056108.

[26] Peter R Hansen and Asger Lunde. "Realized variance and market microstructure noise". In: *Journal of Business & Economic Statistics* 24.2 (2006), pp. 127–161.

[27] AG Hawkes. "Spectra of some mutually exciting point processes with associated variables". In: *Stochastic point processes* (1972), pp. 261–271.

[28] AG Hawkes and L Adamopoulos. "Cluster models for earthquakes-regional comparisons". In: *Bull. Int. Statist. Inst* 45.3 (1973), pp. 454–461.

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

[30] Alan G Hawkes. "Spectra of some self-exciting and mutually exciting point processes". In: *Biometrika* 58.1 (1971), pp. 83–90.

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

[32] Martin Jacobsen. *Point process theory and applications, Probability and its Applications*. 2006.

[33] Michael Johannes. "The Statistical and Economic Role of Jumps in Continuous-Time Interest Rate Models". In: *The Journal of Finance* 59.1 (2004), pp. 227–260.

[34] Armand Joulin et al. "Stock price jumps: news and volume play a minor role". In: *arXiv preprint arXiv:0803.1769* (2008).

[35] Alan Karr. *Point processes and their statistical inference*. Vol. 7. CRC press, 1991.

[36] Steven G Kou. "A jump-diffusion model for option pricing". In: *Management science* 48.8 (2002), pp. 1086–1101.

[37] Brendon M Lapham. "Hawkes processes and some financial applications". PhD thesis. University of Cape Town, 2014.

[38] MR Leadbetter. "On three basic results in the theory of stationary point processes". In: *Proceedings of the American Mathematical Society* 19.1 (1968), pp. 115–117.

[39] Erik Lewis and George Mohler. "A nonparametric EM algorithm for multiscale Hawkes processes". In: *preprint* (2011), pp. 1–16.

[40] PAW Lewis. "Asymptotic properties and equilibrium conditions for branching Poisson processes". In: *Journal of Applied Probability* (1969), pp. 355–371.

[41] Peter A Lewis and Gerald S Shedler. "Simulation of nonhomogeneous Poisson processes by thinning". In: *Naval Research Logistics Quarterly* 26.3 (1979), pp. 403–413.

[42] Thomas Josef Liniger. "Multivariate hawkes processes". PhD thesis. Diss., Eidgenössische Technische Hochschule ETH Zürich, 2009.

[43] Greta M Ljung and George EP Box. "On a measure of lack of fit in time series models". In: *Biometrika* 65.2 (1978), pp. 297–303.

[44] F Lorenzen. "Analysis of order clustering using high frequency data: A point process approach". PhD thesis. Thesis. Tilburg School of Economics and Management Finance Department, 2012.

[45] Cecilia Mancini. "Non-parametric Threshold Estimation for Models with Stochastic Diffusion Coefficient and Jumps". In: *Scandinavian Journal of Statistics* 36.2 (2009), pp. 270–296.

[46] David Marsan and Olivier Lengliné. "Extending earthquakes' reach through cascading". In: *Science* 319.5866 (2008), pp. 1076–1079.

[47] Geoffrey McLachlan and Thriyambakam Krishnan. *The EM algorithm and extensions*. Vol. 382. John Wiley & Sons, 2007.

[48] Robert C Merton. "Option pricing when underlying stock returns are discontinuous". In: *Journal of financial economics* 3.1-2 (1976), pp. 125–144.

[49] Jesper Møller and Jakob G Rasmussen. "Approximate simulation of Hawkes processes". In: *Methodology and Computing in Applied Probability* 8.1 (2006), pp. 53–64.

[50] Jesper Møller and Jakob G Rasmussen. "Perfect simulation of Hawkes processes". In: *Advances in applied probability* (2005), pp. 629–646.

[51] Yoshiko Ogata. "The asymptotic behaviour of maximum likelihood estimators for stationary point processes". In: *Annals of the Institute of Statistical Mathematics* 30.1 (1978), pp. 243–261.

[52] Yosihiko Ogata. "On Lewis' simulation method for point processes". In: *IEEE Transactions on Information Theory* 27.1 (1981), pp. 23–31.

[53] Yosihiko Ogata. "Statistical models for earthquake occurrences and residual analysis for point processes". In: *Journal of the American Statistical association* 83.401 (1988), pp. 9–27.

[54] Jamie F Olson and Kathleen M Carley. "Exact and approximate EM estimation of mutually exciting hawkes processes". In: *Statistical Inference for Stochastic Processes* 16.1 (2013), pp. 63–80.

[55] Tohru Ozaki. "Maximum likelihood estimation of Hawkes' self-exciting point processes". In: *Annals of the Institute of Statistical Mathematics* 31.1 (1979), pp. 145–155.

[56] R-D Reiss. *A course on point processes*. Springer Science & Business Media, 2012.

[57] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion, vol. 293 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. 1999.

[58] Patricia Reynaud-Bouret, Sophie Schbath, et al. "Adaptive estimation for Hawkes processes; application to genome analysis". In: *The Annals of Statistics* 38.5 (2010), pp. 2781–2822.

[59] A Saichev, A Helmstetter, and D Sornette. "Power-law distributions of offspring and generation numbers in branching models of earthquake triggering". In: *Pure and applied geophysics* 162.6-7 (2005), pp. 1113–1134.

[60] Philipp J Schönbucher. *Credit derivatives pricing models: models, pricing and implementation*. John Wiley & Sons, 2003.

[61] Alejandro Veen and Frederic P Schoenberg. "Estimation of space–time branching process models in seismology using an em–type algorithm". In: *Journal of the American Statistical Association* 103.482 (2008), pp. 614–624.

[62] D Vere-Jones and T Ozaki. "Some examples of statistical estimation applied to earthquake data". In: *Annals of the Institute of Statistical Mathematics* 34.1 (1982), pp. 189–207.

[63] David Vere-Jones. "Earthquake prediction-a statistician's view." In: *Journal of Physics of the Earth* 26.2 (1978), pp. 129–146.

[64] Matt P Wand and M Chris Jones. *Kernel smoothing*. Crc Press, 1994.