Hawkes Processes and Their Applications to High-Frequency Data Modeling

Chapter · April 2016

DOI: 10.1002/9781118593486.ch7

2 authors:

Baron Law
Agam Capital
2 PUBLICATIONS **7** CITATIONS
SEE PROFILE

Frederi Viens
Purdue University
84 PUBLICATIONS **1,243** CITATIONS
SEE PROFILE

Hawkes Processes and their Applications to High Frequency Data Modeling

Baron Law and Frederi Viens

Purdue University

8 Oct, 2015

1 Introduction

This short paper introduces and surveys an emerging class of stochastic point processes used in modeling the evolution of high-frequency data on stock markets at a high level of quantitative detail.

The information contained in a stock market's *Limit Order Books (LOB)* is a multivariate time series which records the order arrival times and volumes at each price level of thousands of stocks trading on the exchange. An LOB exhibits a number of distinctive characteristics [1–3] including

1. irregular time interval between arrivals
2. discrete state space of price ticks and volume lot sizes
3. intraday seasonality (more activities around market open and close)
4. arrival clustering
5. self excitation from its own history
6. cross (mutual) excitation from the history of other assets
7. long memory of excitation effect

Consequently, classical time series models with fixed time intervals such as ARIMA and GARCH are not suitable to model High Frequency (HF) financial data. A standard approach commonly used in practice is to re-sample the data in 5-minutes intervals [4, 5], thereby avoiding the time scale for liquid stocks where many of the characteristics listed above can be observed, but this may amount to discarding more than 99% of the data for such stocks. On the other hand, Poisson processes, which are widely used in the market microstructure literature [6, 7], fail to depict the above prevalent features of HF data.

This survey paper, on the current research in HF financial data modeling, concentrates on the use of so-called Hawkes processes, a family of point processes designed to model self and cross excitation. In section 2, we offer an informal introduction to point processes while a more technical review can be found in Appendix A. All the presented material about point processes is thoroughly covered in major textbooks including our reference [8–12]. Sections 3 and 4 introduce Hawkes processes and their statistical inference; a brief history thereof is provided in Appendix B. Section 5 presents the applications of Hawkes processes to HF data modeling.

2 Point Processes

This section provides an informal introduction to point processes. A more rigorous treatment using random measures and martingale theory can be found in Appendix A.

A point process is a random countable set of points $\{x_1, x_2, ...\}$ on a set X. The $\{x_n\}$'s are X-valued random variables; they can have a highly non-trivial dependence structure. For example, $\{x_n\}$'s can be locations of earthquakes on $X = \mathbb{R}^2$, and if self excitation (property 5 above) is built into the model, which is the case with Hawkes processes as we will discuss, then $\{x_n\}$'s can exhibit random patterns of aftershocks surrounding major earthquakes.

Let $N(A)$ be the number of points inside a region $A \subseteq$ X. If $N(A)$ is known for all subsets A of X, this essentially determines the locations of all the points and this representation is the principle tool to describe a point process. A point process N is called *simple* if each location has at most one point.

When X is the positive half-line $\mathbb{R}_+$, this typically represents the time axis and the points are regarded as the times of event occurrences. In this case, $N(t) = N((0,t])$ denotes the number of occurrences at or before time t. For our purpose of modeling HF data, we will mainly deal with point processes on $\mathbb{R}_+$.

When an event happens at $t_n \in \mathbb{R}_+$, it may carry an additional information y_n (its mark). For instance in finance, each order arrival is associated with an order quantity (volume); in seismology, each earthquake is reported with a magnitude. A point process with marks is called a marked point process (MPP). Let Y be the mark space (i.e. $y_n \in Y$) and $A \subseteq Y$; then $N(t,A) = N((0,t] \times A)$ denotes the number of events that happened at or before time t such that the marks fall within the set A. The ground process $N_g(t) = N((0,t] \times Y)$ is the count of all events in $(0,t]$ regardless of the marks[1]. The marks of a MPP are called *unpredictable* if y_n is independent of $\{(t_i, y_i)\}_{i<n}$ and they are called *independent* if y_n is independent of $\{(t_i, y_i)\}_{i\neq n}$[2]. A MPP N is called simple if N_g is simple.

If a point process has multiple occurrences in the same location, we can treat it as a simple marked point process with the mark being the number of points in each location. Hence, without loss of generality, most of the results to be presented will be based on simple point processes.

N is called a multivariate point process if the mark space $Y = \{1,..,d\}$. In this case, $N_i(\bullet) = N(\bullet \times \{i\})$ is called the marginal process for the points of type i. A simple d-variate point process is different from a d-dimensional ($\mathbb{R}^d$-valued) simple point process as the former cannot have any common jump times. In addition, if the multivariate point process also carries some extra information $w_n \in W$ at each point, the mark space will become $Y = \{1,..,d\} \times W$. In this case, the marginal processes are $N_i(t,A) = N((0,t] \times \{i\} \times A)$ and the marginal ground processes are $N_i(t) = N((0,t] \times \{i\} \times Y)$.

Many point processes can be modeled in term of their stochastic intensities[3] $\lambda(t)$, which can be defined informally as the expected number of arrivals per unit of time, at time t, conditioned on all the information just before time t, i.e.

$$\lambda(t) = \lim_{h \to 0^+} \frac{\mathbb{E}(N((t-h,t])|\mathscr{F}_{t-h})}{h} \tag{1}$$

where $\mathscr{F}_t$, which represents the information contained in all the events happened up to and including time t, denotes the filtration of N. For a multivariate point process, $\lambda_i(t)$ is the intensity of the marginal process $N_i(t)$.

The most well-known point processes are Poisson processes, where the intensities are deterministic functions of time. When one generalizes the intensities to become stochastic processes on their own, they play a role similar to stochastic volatility in diffusions. One important use of stochastic intensity is to allow it to change according to how events unfold over time. The Hawkes processes are directly exploring this feature for the purpose of modeling self excitation, as we are about to see.

3 Hawkes Processes

A Hawkes process [13] is a point process where its stochastic intensity has an autoregressive form . For a nonlinear multivariate marked Hawkes process, the intensity $\lambda(t) = (\lambda_1(t),..,\lambda_d(t))$ is given by [4,5]

$$\lambda_i(t) = \Phi_i\left(\sum_{j=1}^{d} \int_{(-\infty,t)\times Y} \gamma_{ij}(t-s,y) N_j(ds \times dy),\ t\right) = \Phi_i\left(\sum_{t_n<t} \gamma_{i,w_n}(t-t_n, y_n),\ t\right) \tag{2}$$
$$\Phi_i : \mathbb{R} \times \mathbb{R}_+ \longrightarrow \mathbb{R}_+, \quad \gamma_{ij} : \mathbb{R}_+ \times Y \longrightarrow \mathbb{R}, \quad N_j : \mathscr{B}(\mathbb{R}_+ \times Y) \longrightarrow \mathbb{N}$$

where $w_n \in \{1,..,d\}$ denotes the type of t_n and Φ_i is known as rate function. Consider the special case

$$\lambda_i(t) = \mu_i(t) + \sum_{j=1}^{d} \int_{(-\infty,t)\times Y} \gamma_{ij}(t-s,y) N_j(ds \times dy) = \mu_i(t) + \sum_{t_n<t} \gamma_{i,w_n}(t-t_n, y_n) \tag{3}$$
$$\mu_i : \mathbb{R}_+ \longrightarrow \mathbb{R}_+, \quad \gamma_{ij} : \mathbb{R}_+ \times Y \longrightarrow \mathbb{R}_+, \quad N_j : \mathscr{B}(\mathbb{R}_+ \times Y) \longrightarrow \mathbb{N}$$

[1] We may refer to the ground process of a MPP without the subscript "g" if the meaning is clear from the input parameter.
[2] Notice that if marks are independent, future location t_{n+1} cannot depend on previous mark y_n.
[3] See Appendix A for formal definition and existence condition of stochastic intensity.
[4] Notice that some authors use γ_{ji}, so that the first index is the source type and the second index is the destination type.
[5] Hawkes process only specifies the intensity without any restriction on the mark distribution.

i.e. $\Phi_i(x,t) = \mu_i(t) + x$. Such a Hawkes process determined by (3) is called linear, and $\mu_i(t)$ is called the base or background rate. The function γ_{ij} is called (marked) decay/exciting/fertility kernel and often $\gamma_{ij}(t,y)$ takes the separable form $\gamma_{ij}(t)g_{ij}(y)$ where g_{ij} is called mark impact kernel. Popular choices of decay kernel $\gamma_{ij}(t)$ include exponential $\alpha_{ij}e^{-\beta_{ij}t}$ [13], power law $\alpha_{ij}(c_{ij}+t)^{-\beta_{ij}}$ [14] or Laguerre-type polynomial $\sum_{k=0}^{K}\alpha_{ijk}t^k e^{-\beta_{ijk}t}$ [15].

If the decay function is exponential with $\beta_{ij} = \beta_i$, the intensity $\lambda(t)$ and the vector $(N(t),\lambda(t))$ are both Markov processes[6] [16, 17]. Moreover, provided that $\mu_i(t) = \mu_i$, then $(\lambda_1(t),..,\lambda_d(t))$ satisfies the system of stochastic differential equations (SDE)

$$d\lambda_i(t) = \beta_i(\mu_i - \lambda_i(t))dt + \sum_{j=1}^{d}\alpha_{ij}dN_j(t) \tag{4}$$

This specification has the simple interpretation that the events of N_j which happened just before time t increase the intensity $\lambda_i(t)$ by $\alpha_{ij} \geq 0$ and thus trigger further events. Yet if the intensity $\lambda_i(t)$ is higher than μ_i, the first term becomes negative ($\beta_i > 0$) and draws the intensity back to the equilibrium level μ_i. In other words, the intensity $\lambda_i(t)$ is a mean-reverting process driven by its own point process. The Markov property and this intuitive interpretation may explain why the exponential decay kernels are so widely used.

For linear Hawkes processes, $\mu_i(t), \gamma_{ij}(t)$ and $g_{ij}(t)$ must be non-negative for all t, in order to ensure the positivity of $\lambda_i(t)$. As a result, unlike nonlinear Hawkes processes, linear Hawkes processes cannot model inhibitory effect (negative excitation). Nonetheless, the linear Hawkes processes are easier to handle, their properties are better understood and most importantly, they have a branching structure representation, which is extremely useful in simulation, estimation and interpretation of the models.

3.1 Branching Structure Representation

Linear Hawkes processes have a very elegant branching structure representation [18]. We describe here the version for the multivariate Hawkes processes with unpredictable marks [19].

There are d types of immigrants arriving according to Poisson processes with rates $\mu_1,...,\mu_d$. Each individual (descendant or immigrant) will carry an unpredictable mark when born or arrived. An individual of type j born at time t_n with mark y_n will give birth to an individual of type i according to a non-homogeneous Poisson process with rate $\gamma_{ij}(t-t_n,y_n)$. All the non-homogeneous Poisson processes are independent of one another.

Let $N_i(t)$ be the total number of individuals of type i born/arrived at or before time t under the above scenario, then $N(t) = (N_1(t),..,N_d(t))$ will follow the linear marked Hawkes process (3). This representation forms the basis of the Expectation Maximization (EM) algorithm in Section 4.2.2 and we will also see how it is used to measure the endogeneity of a point process in Section 5.4.

3.2 Stationarity

Considering a Hawkes process N with intensity (2) such that $\Phi_i(x,t) = \Phi_i(x)$, N has an unique stationary version[7] if either of the following conditions are satisified [18, 20]:

1. $\Phi_i(x)$ is k_i-Lipschitz[8] and the spectral radius[9] $\rho(A) < 1$ for the $d \times d$ matrix $A = [k_i \int_0^\infty |\gamma_{ij}(t)|dt]_{i,j}$
2. $\Phi_i(x)$ is Lipschitz, $\Phi_i(x) \leq M$, $\int_0^\infty |\gamma_{ij}(t)|dt < \infty$ and $\int_0^\infty t|\gamma_{ij}(t)|dt < \infty$

Technically speaking, N may have other non-stationary versions together with the stationary one; however, the non-stationary version will converges weakly to the stationary version when $t \to \infty$ (see [21] for exact meaning). Since the Hawkes process starts at $-\infty$, $N((0,t])$ will have the stationary distribution for all $t > 0$.

For the case of an exponential decay kernel $\alpha_{ij}e^{-\beta_{ij}t}$, we have a simpler result. Let $A = [\int_0^\infty \alpha_{ij}e^{-\beta_{ij}t}dt]_{i,j} = [\alpha_{ij}/\beta_{ij}]_{i,j}$, then N has an unique stationary version under either of the following conditions [22]:

1. $\Phi_i(x) = \mu_i + x$, $\alpha_{ij} \geq 0$, $\beta_{ij}, \mu_i > 0$, $\rho(A) < 1$ (linear Hawkes process)
2. $\Phi_i(x) = \max(\mu_i + x, \varepsilon_i)$, $\alpha_{ij} \in \mathbb{R}$, $\beta_{ij}, \mu_i > 0$, $\varepsilon_i > 0$, $\rho(A) < 1$ (T-Hawkes process)
3. $\Phi_i(x) = \min(\mu_i + \exp(x), M_i)$, $\alpha_{ij} \in \mathbb{R}$, $\beta_{ij} > 0$, $M_i > \mu_i > 0$ (E-Hawkes process)

[6] N itself is not a Markov process as its intensity at time t depends on its full history before time t.

[7] See Appendix A for definition of stationarity of point processes.

[8] $f : \mathbb{R} \longrightarrow \mathbb{R}$ is called k-Lipschitz (k>0) if $|f(x) - f(y)| \leq k|x-y|$ $\forall x, y \in \mathbb{R}$.

[9] $\rho(A) = \max_i\{|\pi_i|\}$, $\{\pi_i\}$ are eigenvalues of A.

For the univariate linear case with $\mu = 0$, if there exists $r, R > 0, c \in (0, 1/2)$ such that $\int_0^\infty \gamma(t)dt = 1$, $\sup_{t\geq 0} t^{1+c}\gamma(t) \leq R$, $\lim_{t\to\infty} t^{1+c}\gamma(t) = r$, Brémaud and Massoulié [23] show that there exists a unique stationary non-trivial Hawkes process having such an intensity and he calls it the critical Hawkes process or Hawkes process without ancestors ($\mu = 0$).

3.3 Convergence

In this section, we state the various results about the convergence of Hawkes processes. A properly scaled linear Hawkes process will converge weakly to a Brownian diffusion when the spectral radius of decay functions' L^1-norm is less than one [24]. When the spectral radius is close to one in a certain sense, it converges to the integrated Cox-Ingersoll-Ross (CIR) process [25]. For the non-linear Hawkes processes, we only have the result for the univariate case and the sufficient conditions depends on the Lipschitz constant of Φ [26].

3.3.1 Law of Large Numbers for Multivariate Linear Hawkes processes

Assuming the model (3) without marks, if the spectral radius $\rho(A) < 1$ where $A = [\int_0^\infty \gamma_{ij}(t)dt]_{i,j}$, then [24]

$$\sup_{t\in[0,1]} \left\| \frac{N(nt)}{n} - t(I_d - A)^{-1}\mu \right\| \xrightarrow[n\to\infty]{\text{a.s./}L^2} 0^{10,11} \tag{5}$$

where $\mu = (\mu_1, .., \mu_d)$. When $d = 1$ and we take $t = 1$, it implies

$$\frac{N(T)}{T} \xrightarrow[T\to\infty]{\text{a.s./}L^2} \frac{\mu}{1 - \int_0^\infty \gamma(t)dt} \tag{6}$$

3.3.2 Functional Central Limit Theorem for Multivariate Linear Hawkes processes

Assuming the model (3) without marks, $N = (N_1, .., N_d)$, if the spectral radius $\rho(A) < 1$ where $A = [\int_0^\infty \gamma_{ij}(t)dt]_{i,j}$ and $\int_0^\infty \sqrt{t}\gamma_{ij}(t)dt < \infty \; \forall i, j$, then [24]

$$\sqrt{n}\left(N(\bullet n)/n - \bullet(I_d - A)^{-1}\mu\right) \xrightarrow[n\to\infty]{\text{weak}} (I_d - A)^{-1}\Sigma^{1/2}W(\bullet)^{12} \tag{7}$$

$$\Sigma = \text{diag}((I_d - A)^{-1}\mu)^{13},\ W \text{ is standard } d-\text{dimensional Brownian Motion} \tag{8}$$

3.3.3 Functional Central Limit Theorem for Univariate Non-linear Hawkes processes

Assuming the model (2) without marks and $d = 1$, if $\gamma(t)$ is decreasing, $\int_0^\infty t\gamma(t)dt < \infty$, $\Phi(x,t) = \Phi(x)$ is increasing and k-Lipschitz, $\int_0^\infty k\gamma(t)dt < 1$ then [26]

$$\sqrt{n}\left(N(\bullet n)/n - \bullet\nu\right) \xrightarrow[n\to\infty]{\text{weak}} \sigma W(\bullet) \tag{9}$$

$$\sigma^2 = \mathbb{E}((N([0,1]) - \nu)^2) + 2\sum_{n=1}^{\infty} \mathbb{E}((N([0,1]) - \nu)(N([n,n+1]) - \nu),\ \nu = \mathbb{E}(N([0,1])) \tag{10}$$

3.3.4 Convergence of Nearly Unstable Univariate Linear Hawkes processes

Considering the linear model (3) without marks and $d = 1$, $N(T)/T$ converges to $\mu/(1 - \int_0^\infty \gamma(t)dt)$ when $\int_0^\infty \gamma(t)dt < 1$ by (6), while it explodes when $\int_0^\infty \gamma(t)dt = 1$. However, Jaisson and Rosenbaum [25] find that the properly scaled

[10] A sequence of random variables $X_n \xrightarrow[n\to\infty]{a.s.} X$ if $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$

[11] A sequence of random variables $X_n \xrightarrow[n\to\infty]{L^2} X$ if $\lim_{n\to\infty} E(|X_n - X|^2) = 0$

[12] A sequence of probability measure P_n converges weakly to P if $\int_\Omega f dP_n \to \int_\Omega f dP$ for all bounded continuous function f. A sequence of stochastic process $X_n : \Omega \to D[0,1]$ converges weakly (in distribution) to X if the law of $X_n(P_n \circ X_n^{-1})$ converges weakly to law of $X(P \circ X^{-1})$ in the sense of probability measure, D[0,1] is the Skorokhod space of càdlàg (right continuous with left limits) functions (see [27, 28]).

[13] $v \in \mathbb{R}^d$, $\text{diag}(v) = [a_{ij}]_{d\times d}$, $a_{ii} = v_i$, $a_{ij} = 0 \; \forall i \neq j$.

Hawkes process converges to the integrated Cox-Ingersoll-Ross (CIR) process when one has a sequence of decay kernel $\gamma^{(n)}(t)$ whose integral converges to one at the speed of n^{-1} (see 13). More precisely, let

$$\lambda^{(n)}(t) = \mu + \int_{(0,t)} \gamma^{(n)}(t-s) dN^{(n)}(s), \quad \mu > 0, \quad \gamma^{(n)}(t) = \alpha^{(n)} \gamma(t) \tag{11}$$

$$\gamma : \mathbb{R}_+ \longrightarrow \mathbb{R}_+, \quad \int_0^\infty \gamma(t)dt = 1, \quad \int_0^\infty t\gamma(t)dt = m < \infty, \quad \int_0^\infty |\gamma'(t)|dt < \infty, \sup_{t\in[0,\infty)} |\gamma'(t)| < \infty \tag{12}$$

$$\alpha^{(n)} \in [0,1), \quad \lim_{n\to\infty} \alpha^{(n)} = 1, \quad \lim_{n\to\infty} n(1-\alpha^{(n)}) = c > 0 \tag{13}$$

$$\psi^{(n)}(t) = \sum_{k=1}^{\infty} \gamma^{(n)\otimes k}(t), \quad \rho^{(n)}(t) = \frac{n\psi^{(n)}(nt)}{\int_0^\infty \psi^{(n)}(t)dt}, \quad |\rho^{(n)}(t)| \le M \ \forall n \ \forall t \tag{14}$$

where $f^{\otimes k}$ denotes the k-fold self-convolution of f. If the sequence of Hawkes process $N^{(n)}$ has intensity $\lambda^{(n)}$ satisfying (11-14), then the scaled intensity converges to the CIR process and the scaled Hawkes process converges to the integrated CIR process [25] as follows:

$$(1-\alpha^{(n)})\lambda^{(n)}(n\bullet) \xrightarrow[n\to\infty]{\text{weak}} X(\bullet) \tag{15}$$

$$(1-\alpha^{(n)})\frac{N^{(n)}(n\bullet)}{n} \xrightarrow[n\to\infty]{\text{weak}} \int_0^\bullet X(s)ds \tag{16}$$

$$dX_t = \frac{c}{m}(\mu - X_t)dt + \frac{\sqrt{c}}{m}\sqrt{X_t}dW_t, \quad X_0 = 0 \tag{17}$$

4 Statistical Inference of Hawkes Processes

4.1 Simulation

In this section, we will give an overview of the algorithms which simulate Hawkes processes. Assume we know all the parameters in the functional form of $\mu(t)$ and $\gamma(t,y)$, our goal is to simulate the points $(t_1,y_1),(t_2,y_2),...$ on the interval [0,T].

If the marks distribution depends only on t_n, we can simply generate y_n conditioned on the generated t_n. Next, t_{n+1} can be generated from the intensity $\lambda(t)$ for $t > t_n$ which depends on $\{(t_1,y_1),...,(t_n,y_n)\}$. If the distribution of y_n also depends on $\{(t_{n-1},y_{n-1}),(t_{n-2},y_{n-2}),...\}$, the algorithms can be modified accordingly.

4.1.1 Inverse CDF Transform

The first simulation algorithm for Hawkes processes appears in Ozaki [29]. Suppose the intensity is governed by the univariate Hawkes model in (3). Let t_n be the arrival time and $\tau_n = t_n - t_{n-1}$ be the inter-arrival time. By (63), $\lambda(t) = h_n(t-t_{n-1})$ for $t \in (t_{n-1},t_n]$ where $h_n(t) = g_n(t)/(1-G_n(t^-))$ and g, G are the conditional pdf, cdf of τ_n given $\mathscr{F}_{t_{n-1}}$. If $G_n(t)$ is continuous, $h_n(t)$ is simply the hazard function, and it can be shown that

$$G_n(\tau_n) = 1 - \exp\left(-\int_0^{\tau_n} h_n(s)ds\right) = 1 - \exp\left(-\int_{t_{n-1}}^{t_{n-1}+\tau_n} \lambda(s)ds\right) \tag{18}$$

Given t_{n-1}, we can generate $t_n = t_{n-1} + \tau_n$ by inverse cdf transform $\tau_n = G_n^{-1}(U)$, $U \sim \text{Unif}(0,1)$. However, the inversion needs to be done numerically, so this method is largely superceded by Ogata's modified thinning which we now discuss.

4.1.2 Ogata's Modified Thinning

Ogata [30] introduces the modified thinning method which does not require numerical inversion. The algorithm is based on the following theorem. Let $N = (N_1,..,N_d)$ be a multivariate point process with intensity $(\lambda_1,..,\lambda_d)$ such that $\sum_{i=1}^d \lambda_i(t) \le \lambda^*(t)$ $\forall t$ *a.s.* ($\lambda^*(t)$ is an exogenously chosen deterministic rate function) and N^* is the univariate

non-homogeneous Poisson process with intensity $\lambda^*(t)$. If each point t_n in N^* is given a mark y_n such that $\mathbb{P}(y_n = i) = \lambda_i(t_n)/\lambda^*(t_n),\ i = 1,..,d$, then $(N_1^*,..,N_d^*)$ has the same distribution as $(N_1,..,N_d)$.

The following algorithm generates a d-dimensional multivariate Hawkes process such that $\lambda_i(t)$ is decreasing between points and $|\lambda_i(t) - \lambda_i(t^-)| \leq \alpha_i\ \forall t$.

Ogata's Modified Thinning [30]

1. $n = 1,\ t_0 = 0$
2. Generate $\tau_n \sim \text{Exp}(\lambda_n^*)$ for some $\lambda_n^* \geq \sum_{i=1}^d (\lambda_i(t_{n-1}) + \alpha_i)$
3. Let $t_n = t_{n-1} + \tau_n$
4. Generate $U_n \sim \text{Unif}(0,1)$
5. if $U_n \in \left(\sum_{i=1}^{k-1} \lambda_i(t_n)/\lambda_n^*, \sum_{i=1}^{k} \lambda_i(t_n)/\lambda_n^*\right]$ for some $k \in \{1,..,d\}$ return t_n and the point is of type k, else discard t_n (but keep the value for use in next generation step 2, 3)
6. $n = n+1$, goto step 2

4.1.3 Simulation by Branching Structure

This method generates points using the branching structure representation of linear marked Hawkes processes. Type-j immigrants arrive according to a non-homogeneous Poisson process with rate $\mu_j(t)$. Next, the type-j parent arriving at t_n with mark y_n produces type-i descendants according to non-homogeneous Poisson process with rate $\gamma_{ij}(t - t_n, y_n)$ and the generation is repeated for each descendant until all of them exceed the pre-defined time T. Since all the non-homogeneous Poisson processes are independent, the generations can be done in parallel.

Simulation by the Branching Structure [31]

1. Generate non-homogeneous Poisson processes with intensities $\mu_i(t),\ i = 1,..,d$ on $[0,T]$
2. For each points t_n, generate $y_n|t_n$
3. Suppose t_n is of type j, generates type-i descendants according to non-homogeneous Poisson process with intensity $\gamma_{ij}(t - t_n, y_n)$ on $[t_n, T],\ i = 1,..,d$
4. repeat step 2, 3 for all descendants

The non-homogeneous Poisson process with intensity $\mu(t)$ on $[0,T]$ can be generated using Lewis' thinning algorithm [32]

1. generate $N \sim \text{Poisson}(\mu^*)$ for some $\mu^* \geq \max_{t\in(0,T]} \mu(t)$
2. generate $U_n \sim \text{Unif}(0,1),\ n = 1,..,N$
3. $T_n = U_{(n)}T,\ n = 1,..,N$ $\left(\{U_{(n)}\}\right.$ is the order statistics of $\left.\{U_n\}\right)$
4. generate $V_n \sim \text{Unif}(0,1),\ n = 1,..,N$
5. return T_n if $V_n \leq (\mu(T_n)/\mu^*),\ n = 1,..N$, otherwise discard T_n

4.2 Estimation

Suppose we observe a point process on $(0,T]$ and collect the event times and marks $\{(t_1,y_1),..,(t_N,y_N)\}$, now we would like to estimate the functions $\mu(t)$ and $\gamma(t,y)$ in the intensity $\lambda(t)$ which drives the process $N(t)$. We will summarize the various methods appearing in the literature, but so far the focus is on unmarked processes. In the special case where the marks are independent and identically distributed (IID), the mark distribution can be estimated separately from the point process.

If we assume $\mu(t)$ and $\gamma(t)$ have some parametric representations, we can use Maximum Likelihood Estimation (MLE), Expectation Maximization (EM), or Generalized Method of Moments (GMM) to estimate the parameters. Otherwise, we need to rely on some advanced non-parametric techniques to estimate the whole function curves.

4.2.1 Maximum Likelihood Estimation (MLE)

The log-likelihood of a Hawkes process is given by [29]

$$\log(L(\theta)) = \sum_{i=1}^{d} \left(-\int_0^T \lambda_i(t;\theta)dt + \int_0^T \log(\lambda_i(t;\theta))dN_i(t) \right) \tag{19}$$

In the case of multivariate linear Hawkes process, it becomes

$$\log(L(\theta)) = -\int_0^T \left(\sum_{i=1}^d \mu_i(t;\theta)\right) dt - \sum_{n=1}^N \int_{t_n}^T \left(\sum_{i=1}^d \gamma_{i,w_n}(t-t_n;\theta)\right) dt + \sum_{n=1}^N \log\left(\mu_{w_n}(t_n;\theta) + \sum_{t_m<t_n} \gamma_{w_n,w_m}(t_n-t_m;\theta)\right) \quad (20)$$

The parameters θ can be estimated by maximizing the log-likelihood. However, the numerical optimization is problematic as the log-likelihood function is usually quite flat (see [33, fig.2,3]) and may have a lot of local maxima (see [33, fig.4]).

4.2.2 Expectation Maximization (EM)

For linear Hawkes process, the estimation can also be done via Expectation Maximization (EM) [34, 35] as in [33, 36–39]. EM is a variant of MLE where part of the data is missing. In the branching structure representation, the missing data is the parents which produce the descendants. Let z_n denotes the index of the parent of t_n and w_{z_n} represents the type of the parent of t_n. If $z_n = m$ and $w_{z_n} = j$, that means t_n is produced by the type j point t_m. When z_n is 0, t_n is an immigrant. Also we define $w_0 = 0$, $\gamma_{i,0}(t) = \mu_i(t)$ and $t_0 = 0$ to simplify the expression. Suppose $\{t_n, w_n, z_n\}$ are known, since each generation is an independent Poisson process, the complete data log-likelihood is

$$\log(L(\theta)) = \sum_{n=0}^N \sum_{i=1}^d \left\{ -\int_{t_n}^T \gamma_{i,w_n}(t-t_n;\theta)dt + \sum_{t_m>t_n} \log(\gamma_{i,w_n}(t_m-t_n;\theta))\mathbb{1}(z_m=n)\mathbb{1}(w_m=i) \right\} \quad (21)$$

$$Q(\theta|\theta^{(k)}) = \mathbb{E}^{\theta^{(k)}}(\log(L(\theta))|\{(t_k,w_k)\})$$
$$= \sum_{n=0}^N \sum_{i=1}^d \left\{ -\int_{t_n}^T \gamma_{i,w_n}(t-t_n;\theta)dt + \sum_{t_m>t_n} \log(\gamma_{i,w_n}(t_m-t_n;\theta))\mathbb{P}^{\theta^{(k)}}(z_m=n|\{(t_k,w_k)\})\mathbb{1}(w_m=i) \right\} \quad (22)$$

$$\mathbb{P}^{\theta^{(k)}}(z_m=n|\{(t_k,w_k)\})\mathbb{1}(w_m=i) = \frac{\gamma_{i,w_n}(t_m-t_n;\theta^{(k)})\mathbb{1}(w_m=i)}{\sum_{l=0}^{m-1}\gamma_{i,w_l}(t_m-t_l;\theta^{(k)})} \quad (23)$$

The EM algorithm can be implemented as follows:

1. $k=0$ and choose an initial guess $\theta^{(0)}$
2. E-step: compute $Q(\theta|\theta^{(k)}) = \mathbb{E}^{\theta^{(k)}}(\log(L(\theta))|\{(t_k,w_k)\})$
3. M-step: compute $\theta^{(k+1)} = \text{argmax}_\theta\ Q(\theta|\theta^{(k)})$
4. $k=k+1$, repeat E-step and M-step until $\theta^{(k)}$ converges (e.g. $||\theta^{(k+1)}-\theta^{(k)}|| < \varepsilon$)

In general, the optimization in M-step need to be solve numerically but when the decay kernel has the exponential form $\alpha_{ij}\beta_{ij}\exp(-\beta_{ij}t)$, Olson and Carley [39] suggest a closed form approximate iteration.

$$\mu_i^{(k+1)} = \frac{\sum_{m=1}^N \mathbb{P}^{\theta^{(k)}}(z_m=0|\{(t_k,w_k)\})\mathbb{1}(w_m=i)}{T} \quad (24)$$

$$\alpha_{ij}^{(k+1)} = \frac{\sum_{n=1}^N\sum_{m=n+1}^N \mathbb{P}^{\theta^{(k)}}(z_m=n|\{(t_k,w_k)\})\mathbb{1}(w_m=i,w_n=j)}{\sum_{n=1}^N \mathbb{1}(w_n=j)} \quad (25)$$

$$\beta_{ij}^{(k+1)} = \frac{\sum_{n=1}^N\sum_{m=n+1}^N \mathbb{P}^{\theta^{(k)}}(z_m=n|\{(t_k,w_k)\})\mathbb{1}(w_m=i,w_n=j)}{\sum_{n=1}^N\sum_{m=n+1}^N (t_m-t_n)\mathbb{P}^{\theta^{(k)}}(z_m=n|\{(t_k,w_k)\})\mathbb{1}(w_m=i,w_n=j)} \quad (26)$$

$$\mathbb{P}^{\theta^{(k)}}(z_m=n|\{(t_k,w_k)\})\mathbb{1}(w_m=i,w_n=j) = \frac{\alpha_{ij}^{(k)}\beta_{ij}^{(k)}\exp(-\beta_{ij}^{(k)}(t_m-t_n))\mathbb{1}(w_m=i,w_n=j)}{\mu_i^{(k)} + \sum_{l=1}^{m-1}\alpha_{i,w_l}^{(k)}\beta_{i,w_l}^{(k)}\exp(-\beta_{i,w_l}^{(k)}(t_m-t_l))} \quad (27)$$

$$\mathbb{P}^{\theta^{(k)}}(z_m=0|\{(t_k,w_k)\})\mathbb{1}(w_m=i) = \frac{\mu_i^{(k)}\mathbb{1}(w_m=i)}{\mu_i^{(k)} + \sum_{l=1}^{m-1}\alpha_{i,w_l}^{(k)}\beta_{i,w_l}^{(k)}\exp(-\beta_{i,w_l}^{(k)}(t_m-t_l))} \quad (28)$$

In addition, the summation $\sum_{l=1}^{m-1} \alpha_{i,w_l}^{(k)} \beta_{i,w_l}^{(k)} \exp(-\beta_{i,w_l}^{(k)}(t_m - t_l))$ can be truncated after $\exp(-\beta_{i,w_l}^{(k)}(t_m - t_l))$ has decayed to a small value. The speed of EM is reported to be 10-100 times faster than MLE and more importantly, MLE does not converges within 500 iterations in practically all test cases while EM does [39].

4.2.3 Generalized Method of Moments (GMM)

Another method for statistical estimation apart from MLE is the *Generalized Method of Moments*[14] [40]. The idea is to find the parameters which minimize the difference between theoretically moments (see Appendix A) in term of the unknown parameters and the empirical moments computed directly from the data. If we have more moments than the number of parameters, the method involves solving a weighted least squares problem.

Da Fonseca and Zaatour [41] obtain the analytic moment expressions by restricting the process to be univariate with exponential kernel and making use of the Markov property in this special case. The authors claim that this method is extremely fast but no speed comparison result is provided.

4.2.4 Nonparametric Estimation

Without assuming any parametric form for $\mu(t)$ nor $\gamma(t)$, some nonparametric methods are developed recently to estimate the whole base rate and decay kernel functions. Similar to parametric estimation, penalized MLE or GMM is used to find the function with desirable characteristics (e.g. smooth functions, sparse coefficients). Nonetheless, the nonparametric method, which involves finding the unknown functions in infinite-dimensional spaces, requires extensive computational effort and the underlying statistical construction is usually much more involved than the parametric counterpart.

To the best of our knowledge, the first attempt in nonparameteric estimation of Hawkes process is by Gusto and Schbath [42] in 2005. The authors express the kernel function of the multivariate Hawkes process using B-splines [43] with equally spaced knots. The log-likelihood function involving the basis coefficients are then maximized numerically and the optimal order for the B-splines basis as well as number of knots are determined using AIC criteria [44].

Instead of B-splines, Reynaud-Bouret and Schbath [45] find the function within the space of piecewise constant functions which minimizes the empirical L^2-norm between the true and estimated kernel functions. The method is later extended to multivariate cases [46] with a Lasso type penalty [47] in the minimization objective.

Instead of EM, Zhou et al. [48] use Minorize-Maximization (MM) algorithm [49], in which EM is a special case. In the E-step of MM algorithm, $Q(\bullet|\theta^{(k)})$ is any lower bound of the objective function $\log(L(\bullet))$ such that $Q(\theta^{(k)}|\theta^{(k)}) = \log(L(\theta^{(k)}))$. It is then iteratively maximized in the M-steps until convergence. In [48], the kernel functions are expressed using a finite number of basis functions which are estimated nonparametrically in M-step by solving the Euler-Lagrange equation.

Another approach is to use moment matching to find the kernel function as in [50–52]. In Bacry and Muzy [52], the authors derive the conditional moment density $\mathbb{E}(dN_i(t)|dN_j(0) = 1, dy)$ of multivariate marked Hawkes process as the solution of Wiener-Hopf equation [53] involving $\mu_i, \gamma_{ij}(t), g_{ij}(y)$ for the case that the mark impact kernel is piecewise constant. The conditional moment density can be estimated by any kernel density estimation technique and the Wiener-Hopf equation can be solved numerically via the Nyström method [54].

4.3 Hypothesis Testing

4.3.1 Random Time Change

The classical method to test the goodness-of-fit of a point process model on $\mathbb{R}_+$ is Ogata's residual analysis [14]. Ogata calls $\{\tilde{t}_n = \int_0^{t_n} \hat{\lambda}(s)ds\}$ the residual process[15] and according to the random time change theorem (see Appendix A), the residual process should be close to a standard Poisson process if the estimated intensity $\hat{\lambda}(t)$ is close to the true intensity $\lambda(t)$. The hypothesis that $\{\tilde{t}_n\}$ is a standard Poisson process can be tested by the following methods:

1. QQ Plot [56] of $\{\tilde{\tau}_n = \tilde{t}_n - \tilde{t}_{n-1}\}$ vs Exp(1).
2. Kolmogorov-Smirnov Test [57–59] to test $\tilde{\tau}_n \sim$ Exp(1)
3. Ljung-Box Test [60] to test the lack of serial correlation of $\{\tilde{\tau}_n\}$

[14] Although GMM is consistent under some mild regularity conditions, unlike MLE, it is not asymptotic efficient among the class of consistent estimators.

[15] The terminology is not standard, Baddeley et al. [55] refer $\{N(t_n) - \int_0^{t_n} \hat{\lambda}(s)ds\}$ as residual in order to extend the concept to higher dimension.

4.3.2 Approximate Thinning

Another method to test goodness-of-fit is by thinning, which doesn't require integration of the intensity function. It is useful if the intensity function is estimated non-parametrically. However, the thinned residual process is only approximately a Poisson process.

By Ogata's modified thinning [30], we know that if there exists $b > 0$ such that $b \leq \lambda(t) \forall t$ and we keep point t_n with probability $b/\lambda(t_n)$, the thinned point process is a homogeneous Poisson process with rate b. However, the infimum b of the intensity function is often close to 0, making the number of points in the thinned process very small and the test to have little power. A remedy is to use approximate thinning [61] as follows: choose an integer $k \ll N$, select one point from $\{t_1, .., t_N\}$ with probability of selecting t_n proportional to $\lambda(t_n)^{-1}$. Repeat the selection (without replacement) until k points are selected. The resulting k points will be approximately a homogeneous Poisson process.

5 Applications of Hawkes processes

After the groundwork of basic theory and statistical inference for Hawkes processes, we now unleash their power to model HF data. First, the readers are reminded how diverse the notion of stock trading frequency can be. According to the Trade And Quote database (TAQ), between 9:30am to 4:00pm on May 2, 2014, there were 11 million quote changes (limit + cancellation + market orders) and 0.3 million trades (market orders) for SPDR S&P 500 ETF (SPY). In other words, on average there are 460 quote changes and 13 trades per second. If we take a snapshot every 5 minutes as in [4, 5], we will only use 0.03% of trade data and 0.0007% of quote data. In comparison, Pathfinder Bancorp (PBHC) only has 306 quote changes and 11 trades on the the same day, which means there is a 35 minutes lag between trades on average and thus the 5 minutes snapshots will just give a series of repeated information. Regardless of the sampling frequency, we are likely to get some misleading result if we analyze the asynchronous data from a portfolio of liquid and illiquid stocks using models with fixed intervals.

The construction of multivariate point processes shows that each variate can have a completely different arrival intensity $\lambda_i(t)$. Nonetheless, the multivariate Hawkes process can still model the dependence structure easily via the $\gamma_{ij}(t)$'s, which are estimated by duly considering all the asynchronous data in the highest frequency without any re-sampling.

Order arrivals and price changes are unarguably two of the most important elements in HF trading. Using Hawkes processes, we can estimate their conditional distributions based on all the historical HF asynchronous data, enabling us to give a more accurate real time prediction of future event occurrences. In the following subsections, we are going to highlight some of the literature which take advantage of Hawkes processes to model HF data.

5.1 Modeling Order Arrivals

Bowsher [22][16] is the first to use Hawkes processes to model order arrivals. He uses nonlinear Hawkes processes to allow for inhibitory effect and he considers two rate functions $\Phi_i(x,t) = \mu_i(t) + \exp(x)$ and $\Phi_i(x,t) = \max(\mu_i(t) + x, \varepsilon_i),\ \varepsilon_i > 0$, where both of them guarantee that the stochastic intensity will be strictly positive at all times. For the deterministic base rate $\mu_i(t)$, he exploits a piecewise linear function with knots at 9:30, 10:00, 11:00,...,16:00 while the decay kernel is the exponential function without marks. In addition, an extra term is included to represent the spillover effects from the previous trading day.

Bowsher uses Maximum Likelihood Estimation (MLE) to estimate the parameters for the bivariate point process of trade and quote of General Motor (GM), trading on NYSE between 5 July 2000 to 29 August 2000. The model is found to be decent according to the goodness-of-fit test using random time change.

Instead of modeling arrivals of all trades and quotes, Large [62] uses Hawkes processes to model only the arrivals of aggressive orders, which are market orders depleting the queue and limit orders falling inside the bid-ask spread, in order to study the *resiliency* of the LOB. A LOB is called resilient if it reverts to its generic shape promptly after large trades. The idea is that when a large trade causes the bid-ask spread to widen, the arrival intensity of aggressive limit orders in a resilient LOB will surge so that the gap will be filled very quickly. In order words, the cross excitation effect $\gamma_{ij}(t)$ from aggressive market orders to aggressive limit orders should be reasonably large in a resilient LOB.

In addition to market orders and limit orders, Large also includes the cancellations of limit orders as well as limit orders falling outside the best quotes. Therefore, he builds a 10-variate linear marked Hawkes process with exponential

[16] Though Bowsher's paper was published in 2007, the first draft appeared in 2002.

decay and mark impact kernel to fit the HF data of Barclays (BARC), trading on LSE between 2 Jan 2002 to 31 Jan 2002. The result shows that the widening of bid-ask spread indeed pumps up the intensities of aggressive limit orders, causing the gap to be filled very quickly and hence making the LOB resilient.

More examples of applications of Hawkes processes to order arrivals include the following papers: Muni Toke and Pomponio [63] use similar approach as Large [62] to model trades-through, namely market orders which deplete the best queues and consume at least one share in the second best. Muni Toke [64] designs a more realistic market simulator using Hawkes processes with exponential kernel for order arrivals. Hewlett [65] models the arrival of market orders with Hawkes processes for single period market making. Finally, Alfonsi and Blanc [66], Jaisson [67] tackle the problem of optimal execution with market orders following multivariate Hawkes processes.

5.2 Modeling Price Jumps

5.2.1 Single Asset

Traditionally the events of price jumps are modeled by Poisson processes, which suffer from the drawbacks mentioned in the introduction section. Again, Hawkes processes can be applied to model price jumps, which often delineate clustering, self and cross excitation behaviors.

Bacry et al. [68] use Hawkes processes to model the price jumps, resulting in a model which can reproduce the microstructure noise [69], Epps effect [70] and jump clustering, while maintaining the coarse scale limit of Brownian diffusion. In their model, the trade price $X(t)$ has the dynamics

$$X(t) = N_1(t) - N_2(t) \tag{29}$$

where $N(t) = (N_1(t), N_2(t))$ is a bivariate linear Hawkes process with exponential decay kernel. $N_1(t), N_2(t)$ represents the total number of upward and downward jumps respectively. The authors make additional assumptions that the Hawkes process N has only cross excitation and coefficients are symmetric in order to simplify computation.

$$\lambda_1(t) = \mu + \int_{(0,t)} \gamma(t-s) dN_2(t), \quad \lambda_2(t) = \mu + \int_{(0,t)} \gamma(t-s) dN_1(t), \quad \gamma(t) = \alpha \mathrm{e}^{-\beta t} \tag{30}$$

According to the model, when X jumps up (down), λ_2 (resp. λ_1) increases, causing the probability of jumping down (resp. up) to increase. Such a cross linkage generates the effect of microstructure noise where the trade price is bouncing between best bid and best ask.

Due to the bid-ask bounce, it is well-known that the realized variance (annualized) increases when the sampling frequency increases [71].

$$V(\tau) = \mathbb{E}\left(\frac{1}{T} \sum_{n=0}^{T/\tau} (X((n+1)\tau) - X(n\tau))^2 \right) \tag{31}$$

$$= \frac{2\mu}{1-\alpha/\beta} \left(\frac{1}{(1+\alpha/\beta)^2} + \left(1 - \frac{1}{(1+\alpha/\beta)^2}\right) \frac{1-\mathrm{e}^{-(\alpha+\beta)\tau}}{(\alpha+\beta)\tau} \right) \tag{32}$$

Such an effect can be easily demonstrated by computing the expected realized variance (32) of the jump model (29) and the result with $\mu = 0.16, \alpha = 0.024, \beta = 0.11$ is shown in Fig.1. The authors apply the model to Euro-Bund futures and find a very good fit between the observed and theoretical realized variance under this highly simplified model.

Let $Y(t) = X(nt)$, then $Y(t)$ is a coarse scale version of $X(t)$. For example, if t in X is in micro second and $n = 60000$, then t in Y will be in minute. When we look at the trade price in a low frequency setting, Bacry et al. [24] show that the macroscopic Hawkes jump model goes back to the classical model of Brownian motion due to the functional central limit theorem for linear Hawkes process (7). Assuming that $\int_0^\infty \gamma(t)dt < 1$, then

$$\frac{X(n\bullet)}{\sqrt{n}} \xrightarrow[n\to\infty]{\text{weak}} \sigma W(\bullet), \ \sigma^2 = \frac{2\mu}{(1 - \int_0^\infty \gamma(t)dt)(1 + \int_0^\infty \gamma(t)dt)^2} \tag{33}$$

It is interesting to see how the macroscopic variance σ^2 is related to the microscopic base rate μ and cross excitation $\gamma(t)$. As $\int_0^\infty \gamma(t)dt$ approaches 1, the variance goes to ∞.

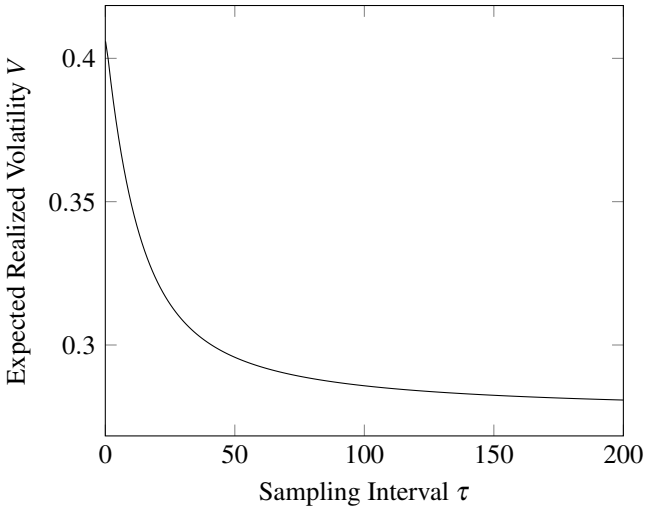

Figure 1: Volatility Signature Plot of Hawkes Jump Model

Jaisson and Rosenbaum [25] extend the model of Bacry et al. [68] to the case of nearly unstable Hawkes process, where $\int_0^\infty \gamma(t)dt \simeq 1$, by constructing a sequence of kernel functions whose integrals converge to one at the speed of n^{-1}. They show that the properly scaled price process converges to Brownian diffusion with Heston stochastic volatility [72]. The full result is stated below.

$$X^{(n)}(t) = N_1^{(n)}(t) - N_2^{(n)}(t) \tag{34}$$

$$\lambda_1^{(n)}(t) = \mu + \int_0^t \gamma_1^{(n)}(t-s)dN_1^{(n)}(s) + \int_0^t \gamma_2^{(n)}(t-s)dN_2^{(n)}(s) \tag{35}$$

$$\lambda_2^{(n)}(t) = \mu + \int_0^t \gamma_2^{(n)}(t-s)dN_1^{(n)}(s) + \int_0^t \gamma_1^{(n)}(t-s)dN_2^{(n)}(s) \tag{36}$$

$$\left(\int_0^\infty \gamma_1^{(n)}(t)dt + \int_0^\infty \gamma_2^{(n)}(t)dt\right) < 1, \quad \gamma_i^{(n)}(t) = \alpha^{(n)}\gamma_i(t) \tag{37}$$

$$\gamma_i : \mathbb{R}_+ \longrightarrow \mathbb{R}_+, \quad \int_0^\infty (\gamma_1(t) + \gamma_2(t))dt = 1, \quad \int_0^\infty t(\gamma_1(t) + \gamma_2(t))dt = m < \infty \tag{38}$$

$$\int_0^\infty |\gamma_i\prime(t)|dt < \infty, \quad \sup_{t\in[0,\infty)} |\gamma_i\prime(t)| < \infty \tag{39}$$

$$\alpha^{(n)} \in [0,1), \quad \lim_{n\to\infty} \alpha^{(n)} = 1, \quad \lim_{n\to\infty} n(1-\alpha^{(n)}) = c > 0 \tag{40}$$

$$\psi^{(n)}(t) = \sum_{k=1}^{\infty} \left(\gamma_1^{(n)} + \gamma_2^{(n)}\right)^{\otimes k}(t), \quad \rho^{(n)}(t) = \frac{n\psi^{(n)}(nt)}{\int_0^\infty \psi^{(n)}(t)dt}, \quad |\rho^{(n)}(t)| \le M \ \forall n \ \forall t \tag{41}$$

Under the conditions of (34 - 41),

$$\frac{X^{(n)}(n\bullet)}{n} \xrightarrow[n\to\infty]{\text{weak}} Y(\bullet) \tag{42}$$

$$dY_t = \frac{\sqrt{V_t}}{1 - \int_0^\infty |\gamma_1(t) - \gamma_2(t)|dt} dW_t^1, \quad Y_0 = 0 \tag{43}$$

$$dV_t = \frac{c}{m}\left(\frac{2\mu}{c} - V_t\right)dt + \frac{\sqrt{V_t}}{m}dW_t^2, \quad V_0 = 0 \tag{44}$$

Conditions (34 - 36) is just a bivariate Hawkes model (with both self and cross excitation) but now we have a different $\gamma_i^{(n)}(t)$ for each n that use to scale the time. The rest are the regularity conditions similar to the univariate

nearly unstable Hawkes process (17) and the most important one is (40) which states that $\alpha^{(n)}$ converges to one at the speed of n^{-1}. However, the interesting result is that instead of converging to an integrated CIR, the price dynamics formed by the difference between two Hawkes processes converges to a stochastic volatility model.

5.2.2 Two Assets

To model the Epps effect, Bacry et al. [68] consider the two-asset case with prices $(X_1(t), X_2(t))$ given by

$$X_1(t) = N_1(t) - N_2(t), \quad X_2(t) = N_3(t) - N_4(t) \tag{45}$$

$$\lambda_i(t) = \mu_i + \sum_{j=1}^{4} \int_{(0,t)} \alpha_{ij} \exp(-\beta(t-s)) dN_j(s), \quad i = 1,..,4 \tag{46}$$

$(N_1(t),..,N_4(t))$ is a 4-variate Hawkes process with exponential kernel where $\beta_{ij} = \beta$. The coupling of excitation effects is constrained to have the form

$$\alpha = \begin{pmatrix} 0 & \alpha_{12} & \alpha_{13} & 0 \\ \alpha_{12} & 0 & 0 & \alpha_{13} \\ \alpha_{31} & 0 & 0 & \alpha_{34} \\ 0 & \alpha_{31} & \alpha_{34} & 0 \end{pmatrix} \tag{47}$$

In this case, there is a closed form representation for the realized correlation, which vanishes when the sampling interval goes to zero (Epps effect).

If we assume $\mu_1 = \mu_2,\ \mu_3 = \mu_4,\ \alpha_{12} = \alpha_{34} = 0,\ (\int_0^\infty \gamma_{13}(t)dt)(\int_0^\infty \gamma_{31}(t)dt) < 1$, then the macroscopic bivariate asset prices converges to correlated Brownian diffusion [24]

$$\frac{1}{\sqrt{n}} \begin{pmatrix} X_1(n\bullet) \\ X_2(n\bullet) \end{pmatrix} \xrightarrow[n\to\infty]{\text{weak}} \frac{\sqrt{2} \begin{pmatrix} \sqrt{\nu_1} W_1(\bullet) + \sqrt{\nu_2} \int_0^\infty \alpha_{13}(t) dt W_2(\bullet) \\ \sqrt{\nu_1} \int_0^\infty \alpha_{31}(t) dt W_1(\bullet) + \sqrt{\nu_2} W_2(\bullet) \end{pmatrix}}{(1 - (\int_0^\infty \gamma_{13}(t)dt)(\int_0^\infty \gamma_{31}(t)dt))^{3/2}} \tag{48}$$

$$\nu_1 = \mu_1 + \left(\int_0^\infty \gamma_{13}(t)dt\right)\mu_3,\ \nu_2 = \mu_3 + \left(\int_0^\infty \gamma_{31}(t)dt\right)\mu_1 \tag{49}$$

(W_1, W_2) is standard 2-dimensional Brownian motion

This convergence result gives us an explicit formula to estimate the macroscopic correlation from the asynchronous HF data.

As a final remark, under this jump representation, the observed trade price is not some hidden continuous fair value process plus some microstructure noise as in [73]. It is the result of the trading interactions between buyers (N_1, N_3) and sellers (N_2, N_4) on a fixed price grid. There is no such thing as HF volatility or correlation since prices are not diffusions but pure jump processes in the HF scale. Volatility and correlation are only meaningful when we look at the coarse scale diffusion approximation, but those low frequency representation parameters can be computed directly from the HF jump model characteristics.

5.3 Modeling Jump-Diffusion

Duffie et al. [74, 75] propose the affine jump-diffusion $X(t)$, which has the following structure.[17]

$$dX(t) = (k_0(t) + k_1(t)X(t))dt + (h_0(t) + h_1(t)X(t))dW(t) + \zeta dN(t) \tag{50}$$

$$\lambda(t) = a_0(t) + a_1(t)X(t) \tag{51}$$

The jump intensity $\lambda(t)$ of $N(t)$ is an affine function of $X(t)$, which depends on the Brownian motion $W(t)$ and the jump process $N(t)$, with jump size ζ drawn from a fixed distribution. When $k_0 = \beta\theta, k_1 = -\beta, h_0 = h_1 = 0, a_0 = 0, a_1 = 1, \zeta = \alpha$, we can see that $\lambda(t) = X(t)$ and $d\lambda(t) = \beta(\theta - \lambda(t))dt + \alpha dN(t)$. Hence in this case, $N(t)$ is the Hawkes process with exponential kernel.

[17] We only show the one dimensional case for simplicity.

Zhu [76] derives some convergence results when ζ is a constant and the diffusion part is a CIR process.

$$dX(t) = \beta(\mu - X(t))dt + \sigma\sqrt{X(t)}dW(t) + \alpha dN(t) \tag{52}$$

$$\lambda(t) = a_0 + a_1 X(t) \tag{53}$$

Aït-Sahalia et al. [77] model the contagion of financial crisis with the Hawkes jump-diffusion where the price dynamic $X_i(t)$ is given by

$$dX_i(t) = \mu_i dt + \sqrt{V_i(t)}dW_i^X(t) + Z_i(t)dN_i(t) \tag{54}$$

$$dV_i(t) = \kappa_i(\theta_i - V_i(t))dt + \eta_i\sqrt{V_i(t)}dW_i^V(t) \tag{55}$$

The diffusion part is the Heston stochastic volatility model and the jump part is a multivariate Hawkes process modeling the clustering and propagation of jumps among multiple assets. $Z_i(t)$ corresponds to the jump size and direction.

5.4 Measuring Endogeneity (Reflexivity)

In term of the Hawkes branching structure representation of events arrivals, Filimonov et al. [78, 79] portray immigrants as exogenous news whereas the descendants are endogenous incidents. In the context of price movements in the stock market, immigrants are the price discovery due to orders from informed traders, who react to external information, whereas the descendants are the destabilizing ripples created by noise traders, who engage in herding [80], momentum trading [81] and parasite trading [82] etc.

Under the univariate linear Hawkes model with exponential decay kernel and constant base rate, the expected number of direct descendants per individual (branching coefficient) is given by

$$n = \int_0^\infty \gamma(s)ds = \int_0^\infty \alpha e^{-\beta s}ds = \alpha/\beta \tag{56}$$

For a given immigrant, the expected number of descendants in all generations is $n + n^2 + n^3 + ... = n/(1-n)$ if $n < 1$, so the ratio of descendants (non-immigrants) vs total population is

$$\frac{\text{descendants}}{\text{descendants + immigrant}} = \frac{n/(1-n)}{n/(1-n)+1} = n \tag{57}$$

Therefore, the branching coefficient n characterizes the amount of endogenous feedback activities while the base rate μ measures the arrival rate of exogenous information.

Using E-mini S&P futures as proxy, Filimonov and Sornette [78] find that the level of endogeneity (reflexivity[18]) n in the US market has gone from 0.3 in 1998 to 0.7 in 2007. Moreover, in the flash crash of May 6, 2010, n reached a peak of 0.95.

Nonetheless, using the power law decay kernel, Hardiman et al. [84] challenge the result of Filimonov and Sornette [78] by reporting that the branching ratio n has always been close to one since 1998 and that the market could be a critical Hawkes process [23], but Filimonov and Sornette [85] refute that the power law kernel is sensitive to outliers in addition to other counter arguments. Later Hardiman and Bouchaud [86] devised a nonparametric estimation of the branching ratio in term of moments, but the result depends heavily on the window size used in the empirical moment computation.

Appendix A: Point Processes

Definition

Let X (state space) be a locally compact Hausdorff second countable topological space[19], $\mathscr{B}_X$ be the Borel sets on X and $\mathscr{B}$ be the collection of bounded (relatively compact) borel sets on X. A Borel measure μ on $(X, \mathscr{B}_X)$ is

[18] Filimonov and Sornette [78] borrow this term from Soros [83].

[19] Some textbooks use complete separable metric space, but locally compact Hausdorff second countable space has a complete separable metrization and all the results here do not depend on any particular choice of metric [8, p.11]. In most cases, $X = \mathbb{R}^m$.

called locally finite if $\mu(B) < \infty \ \forall B \in \mathscr{B}$. Let $\mathfrak{N}(X)$[20] be the set of (positive) locally finite Borel counting (integer-valued) measure on $(X, \mathscr{B}_X)$ and $\mathscr{N}(X)$[21] be the σ-algebra of $\mathfrak{N}(X)$ generated by the set of evaluation functionals $\{\Phi_B : \mathfrak{N}(X) \longrightarrow \mathbb{N} |\ B \in \mathscr{B}\}$ where $\Phi_B(\mu) = \mu(B)$ and $\mathbb{N} = \{0, 1, 2, ...\}$.

A point process N on X is defined as a measurable mapping from a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ to $(\mathfrak{N}(X), \mathscr{N}(X))$; thus a point process is formally a measure-valued random element. However, for any point process N, there exists random variables $b_i \in \mathbb{Z}_+ = \{1, 2, ..\},\ x_i \in X,\ n \in \overline{\mathbb{Z}}_+ = \mathbb{Z}_+ \cup \{\infty\}$ such that $N(\bullet) = \sum_{i=1}^{n} b_i \delta_{x_i}(\bullet)$ where δ_x is the Dirac measure $(\delta_x(A) = \mathbb{1}(x \in A))$ [8, p.20]. If we think of b_i as the number of points at x_i, we can see that the point process N is indeed the random counting measure showing the total number of points in any given region and this matches our intuition that a point process is a random set of points $\{x_i\}$ on X.

The point process N is called simple if $\mathbb{P}(N(\{x\}) > 1) = 0\ \forall x \in X$[22]; that is, each location has at most one point. In this case, $N(\bullet) = \sum_{i=1}^{n} \delta_{x_i}(\bullet)$.

Suppose X is also a topological vector space (e.g. $\mathbb{R}^m$), the shift operator $S_t : \mathscr{B}_X \longrightarrow \mathscr{B}_X$ is defined as $S_t(A) = A + t = \{(s + t) \in X | s \in A\}$. A point process N is called stationary if the shifted process $N \circ S_t$[23] has the same distribution as $N\ \forall t \in X$.

Moments

Let $k \in \mathbb{Z}_+$, the k^{th} moment measure[24] $M^k : \mathscr{B}_X^{\otimes k} \longrightarrow [0, \infty]$ of a point process N is defined as

$$M^k(A_1, .., A_k) = \mathbb{E}(N(A_1)...N(A_k)) = \mathbb{E}\left(\sum_{x_1} .. \sum_{x_k} \delta_{(x_1,..,x_k)}(A_1 \times ... \times A_k)\right) \tag{58}$$

The first moment measure is also called mean (intensity) measure and denoted as $M(\bullet)$. The covariance measure is defined as

$$C^2(A_1, A_2) = \mathrm{Cov}(N(A_1), N(A_2)) = M^2(A_1, A_2) - M(A_1)M(A_2) \tag{59}$$

The second and higher moment measures have concentration along diagonals, so we also have the k^{th} *factorial* moment measure.

$$M^{(k)}(A_1, .., A_k) = \mathbb{E}\left(\sum_{x_1 \neq .. \neq x_k} .. \sum \delta_{(x_1,..,x_k)}(A_1 \times ... \times A_k)\right) \tag{60}$$

The name factorial comes from the fact that $M^{(k)}(A, .., A) = \mathbb{E}(N(A)(N(A) - 1)...(N(A) - k + 1))$. Obviously, $M(A) = M^{(1)}(A)$ and for $k = 2$, we have $M^2(A_1, A_2) = M^{(2)}(A_1, A_2) + M(A_1 \cap A_2)$.

If $X = \mathbb{R}^m$ and N is stationary, it can be shown that $M(A) = \lambda |A|$ where $\lambda = M((0, 1]^m)$ and $|\bullet|$ is Lebesgue measure. That implies the mean measure M of a stationary point process is absolutely continuous with respect to Lebesgue measure with constant density $M((0, 1]^m)$. If the covariance factorial moment measure $C^{(2)}$ is also absolutely continuous, we denote its density function as $c^{(2)}(x, y)$. Since N is stationary, $c^{(2)}(x, y) = c^{(2)}(y - x)$ and $c^{(2)}(\bullet)$ is called reduced covariance density. The covariance measure C^2 is usually not absolutely continuous but for simple point process N on $\mathbb{R}_+$, the quantity below is still called (reduced) covariance density, and is useful in estimation:

$$c^2(dx) = \mathbb{E}(N(x + dx)N(x))/dx^2 - \lambda^2 = \lambda \delta(dx) + c^{(2)}(dx) \quad \left(\int_{-\infty}^{\infty} \delta(x)dx = 1\right) \tag{61}$$

Marked Point Processes

When an event happens, it may carry an additional information (mark). For instance, each order arrival is associated with an order quantity (volume) and each earthquake is reported with a magnitude. A point process with marks is called marked point process.

Let Y (mark space) be a locally compact Hausdorff second countable space, $(Y, \mathscr{B}_Y)$ be a measurable space and ν (mark distribution) be a probability measure on $(Y, \mathscr{B}_Y)$. A marked point process (MPP) N is a measurable mapping

[20] On locally compact Hausdorff second countable space, all locally finite Borel measures are Radon measures.

[21] $\mathscr{N}(X)$ is the same as the Borel σ-algebra generated by the vague topology of $\mathfrak{N}(X)$ [8, p.32].

[22] X is Hausdorff, so all singletons are closed and thus measurable.

[23] $N \circ S_t : \Omega \longrightarrow (\mathfrak{N}(X), \mathscr{N}(X)),\ ((N \circ S_t)(\omega))(A) = (N(\omega))(S_t(A))$, that is N is shifted t unit to the left when $X = \mathbb{R}$.

[24] The notations of moment, covariance, factorial moment, reduced moment vary between authors.

$N:\Omega \longrightarrow (\mathfrak{N}(X\times Y),\mathscr{N}(X\times Y))$ such that the ground measure $N_g(\bullet)=N(\bullet\times Y)$ is a point process (i.e. locally finite)[25]. Hence a marked point process is nothing but a point process on a product space, but usually we treat the location x and mark y differently and we have a few more definitions.

N is called a multivariate point process if $Y=\{1,..,d\}$. In this case, $N_i(\bullet)=N(\bullet\times\{i\})$ is called the marginal process of type i points. A MPP N is called simple if N_g is simple[26]. The marks of a MPP are called *unpredictable* if y_n is independent of $\{(x_i,y_i)\}_{i<n}$ and they are called *independent* if y_n is independent of $\{(x_i,y_i)\}_{i\neq n}$[27].

Stochastic Intensity

In this section, $X=\mathbb{R}_+$[28] and $N_t=N((0,t])$. Let $(\Omega,\mathscr{F},\mathscr{F}_t,\mathbb{P})$ be a filtered complete probability space. A stochastic process $Z:\mathbb{R}_+\times\Omega\longrightarrow\mathbb{R}$ is called $\mathscr{F}$-predictable if it is measurable with respect to the predictable σ-algebra $\mathscr{P}=\sigma(\{(s,t]\times A|0\leq s<t,A\in\mathscr{F}_s\})$. If Z_t is adapted and left-continuous, then Z_t is predictable [10, p.9]. In practice, all the predictable processes we use are in this category. Also if Z_t is predictable, then $Z_t\in\mathscr{F}_{t^-}$; in other words, the value of the predictable process Z_t at time t is "known" just before time t.

We assume the filtration $\{\mathscr{F}_t\}$ satisfies the usual condition (complete and right-continuous) and $\{N_t\}$ is adapted and simple. A stochastic process $A:\mathbb{R}_+\times\Omega\longrightarrow\mathbb{R}_+$ is called a $\mathscr{F}$-compensator of a point process N if A_t is increasing, right-continuous, $\mathscr{F}$-predictable, $A_0=0$ a.s. and (N_t-A_t) is a $\mathscr{F}$-local martingale. If $A_t=\int_0^t\lambda_s ds$ a.s., λ_t is non-negative and $\mathscr{F}$-predictable, then λ_t is called the stochastic or conditional $\mathscr{F}$-intensity of N[29,30,31]. A defining properties of λ_t is that

$$\mathbb{E}\left(\int_s^t\lambda_u du\middle|\mathscr{F}_s\right)=\mathbb{E}(N_t-N_s|\mathscr{F}_s)\quad\text{a.s.}\quad\forall s<t \tag{62}$$

When $s\to t$, this becomes $\lambda_t dt=\mathbb{E}(N(dt)|\mathscr{F}_{t^-})$. We can see that the stochastic intensity λ_t is the instantaneous rate of arrival conditioned on all information just before time t. For a multivariate point process, $\lambda_i(t)$ is the intensity of the marginal process $N_i(t)$.

From the definition, we notice that intensity exists if and only if the compensator is absolutely continuous. In fact, the compensator of a point process can be expressed in term of the conditional inter-arrival time $(t_n-t_{n-1})|\mathscr{F}_{t_{n-1}}$ if the conditional distribution has support over $\mathbb{R}_+$. Under this condition, the intensity exists if and only if the conditional inter-arrival time is absolutely continuous. In this case, the intensity is given by [9, p.70]

$$\lambda_t=h_n(t-t_{n-1})\text{ if }t\in(t_{n-1},t_n] \tag{63}$$

$$h_n(t)=\frac{g_n(t)}{1-G_n(t^-)},\ (t_n-t_{n-1})|\mathscr{F}_{t_{n-1}}\sim G_n \tag{64}$$

Once we know the intensity, we know the conditional distributions of all inter-arrival times and hence the complete distribution of the point process [11, p.233].

$A:\mathbb{R}_+\times\mathscr{B}_Y\times\Omega\longrightarrow\mathbb{R}_+$ is called a compensator of the MPP N if $A(\bullet,B)$ is a compensator of $N(\bullet\times B)\ \forall B\in\mathscr{B}_Y$ and $A(t,\bullet)$ is a measure on $(Y,\mathscr{B}_Y)\ \forall t\in\mathbb{R}_+$. If $A(t,B)=\int_0^t\int_B\lambda(s)\nu(s,dy)ds$ a.s. where $\lambda(t)$ is non-negative and predictable, then $\lambda(t)$ is called the stochastic intensity of the MPP N and $\nu(t_n,dy)=\mathbb{P}(y_n\in dy|\mathscr{F}_{t_n^-})$ is called the conditional mark distribution.

Random Time Change

If the filtration is usual, a point process N on $\mathbb{R}_+$ is simple and adapted, its intensity $\lambda(t)$ exists and $\int_0^\infty\lambda(s)ds=\infty$ a.s., then $\{\tilde{t}_n=\int_0^{t_n}\lambda(s)ds\}$ is a standard Poisson process (rate=1). The above theorem is called random time change theorem [88, 89] and is extremely useful in testing the goodness-of-fit of a stochastic intensity model.

[25] From this definition, Poisson random measure N on $\mathbb{R}^2$ is not MPP on $\mathbb{R}\times\mathbb{R}$ as $N(A\times\mathbb{R})=\infty$.

[26] Any point process can be treated as simple MPP with the mark being the number of points at x_i.

[27] Notice that if marks are independent, future location x_{n+1} cannot depends on previous mark y_n.

[28] The stochastic intensity of point process on $\mathbb{R}_+$ is extended to higher dimension in [87].

[29] Stochastic intensity is unique up to modification [10, p.31].

[30] Notice that stochastic intensity depends on the underlying filtration, so some text use the notation $\lambda(t|\mathscr{F}_t)$ but we will simply use $\lambda(t)$ and call it stochastic intensity or intensity when there is no confusion about the filtration.

[31] if A_t is absolutely continuous with respect to Lebesgue measure and λ_t is the Radon-Nikodym derivative (may not be predictable) then $\mathbb{E}(\lambda_t|F_{t^-})$ is a version of the stochastic intensity. Some authors only require the intensity to be adapted, but using the conditional expectation, one can always find a predictable version of intensity provided that the intensity has finite first moment.

Appendix B: A Brief History of Hawkes processes

Hawkes processes are proposed by Hawkes [13] in 1971 in order to model contagious processes like epidemics, neuron firing and particle emission, where the occurrences of events trigger further events. Although the intensity of Cox processes [90], introduced in 1955, are stochastic, they are determined before the events are unfolded[32]. In order to portray the excitation behavior in contagious processes, Hawkes extends the model in such a way that the intensity is a predictable stochastic process with an intuitive autoregressive form, which allows it to adapt to events that happen over time.

After Hawkes' seminal paper, there are a number of theoretical developments include the branching structure representation by Hawkes and Oakes [18] in 1974, Markov property for intensity with exponential decay kernel by Oakes [16] in 1975, MLE for Hawkes processes by Ozaki [29] in 1979, Ogata's modified thinining simulation algorithm [30] in 1981, nonlinear Hawkes processes by Brémaud and Massoulié [20] in 1996, nonparametric estimation by Gusto and Schbath [42] in 2005, EM for Hawkes processes by Veen and Schoenberg [33] in 2008 and functional central limit theorem for Hawkes processes by Bacry et al. [24], Jaisson and Rosenbaum [25], Zhu [26] in 2013-2015.

Although the first application of Hawkes processes to earthquake occurrences appeared in 1982 [15], it was not until 1988 [14] that Hawkes processes received much attention. Since then, the versatility of Hawkes model is leveraged in seismology [14, 91], finance (risk and credit default modeling) [92, 93], social networks [94, 95], neuroscience [96, 97] etc. (see [98, 99] for more applications).

The use of Hawkes processes in HF financial data modeling starts with Bowsher [22] in 2007[33] and then Large [62] in the same year. Both authors exploit Hawkes processes so as to describe the interactions among different types of order arrivals. Later, Bacry et al. [68] and Aït-Sahalia et al. [77] employ Hawkes processes to reproduce jump clustering in pure jump and jump diffusion representation of stock prices in 2013 and 2015 respectively. On the other hand, an interesting idea from Filimonov and Sornette [78] appears in 2012 which utilizes the branching coefficient of linear Hawkes model to measure the level of endogenous activities in the US stock market, though the debate about the validity of the result is still on going [84–86].

References

[1] J.-P. Bouchaud, J. D. Farmer, and F. Lillo. How markets slowly digest changes in supply and demand. *Handbook of financial markets: dynamics and evolution*, 1:57, 2009.

[2] R. Cont. Statistical modeling of high-frequency financial data. *Signal Processing Magazine, IEEE*, 28(5):16–25, 2011.

[3] A. Chakraborti, I. M. Toke, M. Patriarca, and F. Abergel. Econophysics review: I. empirical facts. *Quantitative Finance*, 11(7):991–1012, 2011.

[4] T. Andersen, T. Bollerslev, F. Diebold, and P. Labys. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453):42–55, 2001.

[5] O. E. Barndorff-Nielsen. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):253–280, 2002.

[6] M. Garman. Market microstructure. *Journal of Fiancial Economics*, 3(3):257–275, 1976.

[7] T. Ho and H. Stoll. Optimal dealer pricing under transactions and return uncertainty. *Journal of Fiancial Economics*, 9(1):47–73, 1981.

[8] O. Kallenberg. *Random measures*. Akademie-Verlag, 1983.

[9] A. Karr. *Point processes and their statistical inference*. CRC press, 1991.

[10] P. Brémaud. *Point processes and queues*. Springer, 1981.

[32] For processes on $\mathbb{R}_+$, it means intensity of Cox process is $\mathscr{F}_0$ measurable.

[33] Though Bowsher's paper was published in 2007, the first draft appeared in 2002.

[11] D. Daley and D. Vere-Jones. *An introduction to the theory of point processes: Volume I: Elementary theory and methods*. Springer, 2003.

[12] D. Daley and D. Vere-Jones. *An introduction to the theory of point processes: Volume II: General theory and structure*. Springer, 2007.

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

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

[15] Y. Ogata and H. Akaike. On linear intensity models for mixed doubly stochastic poisson and self-exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 102–107, 1982.

[16] D. Oakes. The markovian self-exciting process. *Journal of Applied Probability*, pages 69–77, 1975.

[17] J. Da Fonseca and R. Zaatour. Clustering and mean reversion in a Hawkes microstructure model. *Journal of Futures Markets*, 2014.

[18] A. G. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, pages 493–503, 1974.

[19] J. Møller and J. G. Rasmussen. Perfect simulation of Hawkes processes. *Advances in applied probability*, pages 629–646, 2005.

[20] P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *The Annals of Probability*, pages 1563–1588, 1996.

[21] P. Brémaud, G. Nappo, and G. Torrisi. Rate of convergence to equilibrium of marked Hawkes processes. *Journal of Applied Probability*, 39(1):123–136, 2002.

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

[23] P. Brémaud and L. Massoulié. Hawkes branching point processes without ancestors. *Journal of Applied Probability*, 38(1):122–135, 2001.

[24] E. Bacry, S. Delattre, M. Hoffmann, and J. Muzy. Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, 2013.

[25] T. Jaisson and M. Rosenbaum. Limit theorems for nearly unstable Hawkes processes. *The Annals of Applied Probability*, 25(2):600–631, 2015.

[26] L. Zhu. Central limit theorem for nonlinear Hawkes processes. *Journal of Applied Probability*, 50(3):760–771, 2013.

[27] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2009.

[28] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*. Springer-Verlag Berlin, 1987.

[29] T. Ozaki. Maximum likelihood estimation of Hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, 31(1):145–155, 1979.

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

[31] J. Zhuang, Y. Ogata, and D. Vere-Jones. Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research*, 109(B5), 2004.

[32] P. A. Lewis and G. S. Shedler. Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413, 1979.

[33] A. Veen and F. P. Schoenberg. Estimation of space–time branching process models in seismology using an EM–type algorithm. *Journal of the American Statistical Association*, 103(482):614–624, 2008.

[34] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977.

[35] G. Mclachlan and T. Krishnan. *The EM algorithm and extensions*. John Wiley & Sons, 2007.

[36] D. Marsan and O. Lengline. Extending earthquakes' reach through cascading. *Science*, 319(5866):1076–1079, 2008.

[37] P. Halpin and P. Boeck. Modelling dyadic interaction with Hawkes processes. *Psychometrika*, 78(4):793–814, 2013.

[38] P. F. Halpin. *A scalable EM algorithm for Hawkes processes*. New developments in quantitative psychology. Springer, 2013.

[39] J. F. Olson and K. M. Carley. Exact and approximate EM estimation of mutually exciting Hawkes processes. *Statistical Inference for Stochastic Processes*, 16(1):63–80, 2013.

[40] L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, pages 1029–1054, 1982.

[41] J. Da Fonseca and R. Zaatour. Hawkes process: Fast calibration, application to trade clustering, and diffusive limit. *Journal of Futures Markets*, 2013.

[42] G. Gusto and S. Schbath. Fado: A statistical method to detect favored or avoided distances between occurrences of motifs using the Hawkes' model. *Statistical Applications in Genetics and Molecular Biology*, 4(1):1119, 2005.

[43] C. De Boor. *A practical guide to splines*. Springer-Verlag New York, 1978.

[44] H. Akaike. Information theory and an extension of the maximum likelihood principle. In *Second international symposium on information theory*, pages 267–281. Akademinai Kiado, 1973.

[45] P. Reynaud-Bouret and S. Schbath. Adaptive estimation for Hawkes processes; application to genome analysis. *The Annals of Statistics*, 38(5):2781–2822, 2010.

[46] N. R. Hansen, P. Reynaud-Bouret, and V. Rivoirard. Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli*, 21(1):83–143, 2015. ISSN 1350-7265.

[47] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[48] K. Zhou, H. Zha, and L. Song. Learning triggering kernels for multi-dimensional Hawkes processes. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 1301–1309, 2013.

[49] D. R. Hunter and K. Lange. A tutorial on MM algorithms. *The American Statistician*, 58(1):30–37, 2004.

[50] E. Bacry, K. Dayri, and J. F. Muzy. Non-parametric kernel estimation for symmetric Hawkes processes. application to high frequency financial data. *The European Physical Journal B*, 85(5), 2012.

[51] E. Bacry and J.-F. Muzy. Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 14(7):1147–1166, 2014.

[52] E. Bacry and J.-F. Muzy. Second order statistics characterization of Hawkes processes and non-parametric estimation. *arXiv preprint arXiv:1401.0903*, 2014.

[53] B. Noble. *Methods based on the Wiener-Hopf technique*. Pergamon Press New York, 1958.

[54] G. Chandlen and I. G. Graham. The convergence of Nyström methods for Wiener-Hopf equations. *Numerische Mathematik*, 52(3):345–364, 1987.

[55] A. Baddeley, R. Turner, J. Møller, and M. Hazelton. Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(5):617–666, 2005.

[56] M. Wilk and R. Gnanadesikan. Probability plotting methods for the analysis of data. *Biometrika*, 55(1):1–17, 1968.

[57] A. N. Kolmogorov. Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, 4(1):83–91, 1933.

[58] N. V. Smirnov. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bull. Math. Univ. Moscou*, 2(2), 1939.

[59] N. Smirnov. Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics*, pages 279–281, 1948.

[60] G. M. Ljung and G. E. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

[61] F. P. Schoenberg. Multidimensional residual analysis of point process models for earthquake occurrences. *Journal of the American Statistical Association*, 98(464), 2003.

[62] J. Large. Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10(1):1–25, 2007.

[63] I. Muni Toke and F. Pomponio. Modelling trades-through in a limited order book using Hawkes processes. *Economics: The Open-Access, Open-Assessment E-Journal*, 6(2012-22), 2012.

[64] I. Muni Toke. "Market Making" in an order book model and its impact on the spread. In *Econophysics of order-driven markets*, pages 49–64. Springer, 2011.

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

[66] A. Alfonsi and P. Blanc. Dynamic optimal execution in a mixed-market-impact Hawkes price model. *arXiv preprint arXiv:1404.0648*, 2014.

[67] T. Jaisson. Market impact as anticipation of the order flow imbalance. *Quantitative Finance*, 15(7):1123–1135, 2015.

[68] E. Bacry, S. Delattre, M. Hoffmann, and J.-F. Muzy. Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13(1):65–77, 2013.

[69] Y. Aït-Sahalia, P. A. Mykland, and L. Zhang. How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies*, 18(2):351–416, 2005.

[70] T. W. Epps. Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74(366a):291–298, 1979.

[71] T. Andersen, T. Bollerslev, F. Diebold, and P. Labys. Great realizations. *Risk Magazine*, 2000.

[72] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6(2):327–343, 1993.

[73] F. M. Bandi and J. R. Russell. Separating microstructure noise from volatility. *Journal of Financial Economics*, 79(3):655–692, 2006.

[74] D. Duffie and R. Kan. A yield-factor model of interest rates. *Mathematical finance*, 6(4):379–406, 1996.

[75] D. Duffie, J. Pan, and K. Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376, 2000.

[76] L. Zhu. Limit theorems for a cox-ingersoll-ross process with Hawkes jumps. *Journal of Applied Probability*, 51 (3):699–712, 2014.

[77] Y. Aït-Sahalia, J. Cacho-Diaz, and R. J. A. Laeven. Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, 117(3):585–606, 2015.

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

[79] V. Filimonov, D. Bicchetti, N. Maystre, and D. Sornette. Quantification of the high level of endogeneity and of structural regime shifts in commodity markets. *Journal of International Money and Finance*, 2013.

[80] R. J. Shiller. *Irrational exuberance*. Princeton University Press, 2005.

[81] J. B. De Long, A. Shleifer, L. H. Summers, and R. J. Waldmann. Positive feedback investment strategies and destabilizing rational speculation. *Journal of Finance*, pages 379–395, 1990.

[82] L. Harris. Order exposure and parasitic traders. *Unpublished manuscript*, 1997.

[83] G. Soros. *The alchemy of finance*. John Wiley Sons, 2003.

[84] S. J. Hardiman, N. Bercot, and J.-P. Bouchaud. Critical reflexivity in financial markets: A Hawkes process analysis. *The European Physical Journal B*, 86(10):1–9, 2013.

[85] V. Filimonov and D. Sornette. Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data. *Quantitative Finance*, 15(8):1293–1314, 2015.

[86] S. J. Hardiman and J.-P. Bouchaud. Branching-ratio approximation for the self-exciting Hawkes process. *Physical Review E*, 90(6):062807, 2014.

[87] F. Papangelou. The conditional intensity of general point processes and an application to line processes. *Probability Theory and Related Fields*, 28(3):207–226, 1974.

[88] P. A. Meyer. Démonstration simplifiée d'un théorème de knight. In *Séminaire de probabilités v université de strasbourg*, pages 191–195. Springer, 1971.

[89] F. Papangelou. Integrability of expected increments of point processes and a related random change of scale. *Transactions of the American Mathematical Society*, 165:483–506, 1972.

[90] D. R. Cox. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B (Methodological)*, 17(2):129–164, 1955.

[91] Y. Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402, 1998.

[92] V. Chavez-Demoulin, A. Davison, and A. Mcneil. Estimating value-at-risk: A point process approach. *Quantitative Finance*, 5(2):227–234, 2005.

[93] E. Errais, K. Giesecke, and L. R. Goldberg. Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, 1(1):642–665, 2010.

[94] R. Crane and D. Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences*, 105(41):15649–15653, 2008.

[95] C. Blundell, J. Beck, and K. A. Heller. Modelling reciprocating relationships with Hawkes processes. In *Advances in Neural Information Processing Systems*, pages 2609–2617, 2012.

[96] M. Krumin, I. Reutsky, and S. Shoham. Correlation-based analysis and generation of multiple spike trains using Hawkes models with an exogenous input. *Frontiers in computational neuroscience*, 4, 2010.

[97] V. Pernice, B. Staude, S. Cardanobile, and S. Rotter. How structure determines correlations in neuronal networks. *PLoS computational biology*, 7(5):e1002059, 2011.

[98] T. Liniger. *Multivariate Hawkes processes*. PhD thesis, Swiss Federal Institute of Technology, 2009.

[99] L. Zhu. *Nonlinear Hawkes processes*. PhD thesis, New York University, 2013.