Applied Probability Trust (7 March 2011)

TITLE Multivariate Hawkes Processes: an Application to Financial Data.

PAUL EMBRECHTS,* *ETH Zurich and Swiss Finance Institute*
THOMAS LINIGER,* *ETH Zurich*
LU LIN,* *ETH Zurich*

Abstract

A Hawkes process is also known under the name of a self-exciting point process and has numerous applications throughout science and engineering. We derive the statistical estimation (MLE) and goodness-of-fit (mainly graphical) for multivariate Hawkes processes with possibly dependent marks. As application, we analyze two data sets from finance.

Keywords: multivariate Hawkes process; point process; clustering; self-exciting.

2000 Mathematics Subject Classification: Primary 60G55
Secondary 91B28

1. Introduction

Numerous approaches have been proposed to model clusters of extremes in univariate as well as multivariate time series models, this for a variety of examples throughout the sciences and economics. A particularly useful class of stochastic processes in this context consists of self-exciting or Hawkes processes; see also [6] and the references therein. For early references to Hawkes processes, see [9] and numerous papers by Yoshihiko Ogata on seismology and earthquake modelling ([15], [16]). For applications of Hawkes process models to finance, see for instance [1], [2], [5] and [8] and the numerous references therein. Our paper is mainly based on [11], [10] and [12]. The novelty of our approach is the intrinsic multivariate character coupled with the possibility of the model to allow for dependent marks through the notion of copulas. The difficulties for using multivariate Hawkes processes are their mathematical complexity and the typically large number of parameters to be estimated.

The paper is structured as follows: In Section 2 we introduce the concept of a multivariate Hawkes process. Section 3 contains parameter estimation using maximum likelihood, together with graphical goodness-of-fit tests. Section 4 gives two examples of Hawkes processes fitted to financial data, one example involving a multivariate model and the other one a univariate model with vector-valued marks. The choice of these parameterizations stresses more the numerical fitting feasibility, rather than offering an in depth search for a parsimonious model. Mathematical details in Sections 2 and 3 are mostly omitted; the interested reader is referred to [11].

2. Multivariate Hawkes Process

In the context of marked point processes, the terms multivariate and vector-valued refer to two different mathematical concepts. The points of a *multivariate* point process

* Postal address: RiskLab, Department of Mathematics, ETH Zurich, Rämistrasse 101, 8092 Zurich, Switzerland. Email address: embrechts@math.ethz.ch

(with scalar-valued marks) are triples of the form (t_i, d_i, x_i), with $i \in \Im$, countable. The component $t_i \in \mathbb{R}$ is time, $d_i \in \{1, \ldots, d\}$ is the component index and $x_i \in \mathbb{R}$ is the mark value of the i-th point. In contrast, the points of a (univariate) point process with *vector-valued* marks are tuples of the form $(t_i, x_{i,1}, \ldots, x_{i,d})$, where $x_{i,j} \in \mathbb{R}$, for $i \in \Im$, $j \in \{1, \ldots, d\}$. The component $t_i \in \mathbb{R}$ again denotes time and $x_{i,j}$ are the components of the vector-valued mark $\boldsymbol{x}_i \in \mathbb{R}^d$.

The index d_i, appearing in the multivariate case, is not a mark in addition to x_i; it assigns t_i to the component d_i of the point process. In contrast, a point process with vector-valued marks has d mark values $x_{i,j}$, one for each of the d components.

Hawkes [9] originally introduced self-exciting point processes via intensity functions; see also Oakes [14] for an early discussion.

Definition 1. (*Hawkes Process.*) Consider either a multivariate marked point process N whose j-th component N_j, $j \in \{1, \ldots, d\}$, has intensity

$$\lambda_j(t) := \eta_j + \sum_{k=1}^{d} \vartheta_{jk} \int_{(-\infty,t)\times\mathbb{R}} \omega_j(t-s) g_k(x) N_k(ds \times dx),\ t \in \mathbb{R}, \tag{1}$$

where $\omega_j : \mathbb{R}_+ \to \mathbb{R}_+$, $g_k : \mathbb{R} \to \mathbb{R}_+$ and $\eta_j, \vartheta_{jk} \geq 0$; or consider a point process with vector-valued marks with intensity

$$\lambda(t) := \eta + \vartheta \int_{(-\infty,t)\times\mathbb{R}^d} \omega(t-s) g(\boldsymbol{x}) N(ds \times d\boldsymbol{x}),\ t \in \mathbb{R}, \tag{2}$$

where $\omega : \mathbb{R}_+ \to \mathbb{R}_+$, $g : \mathbb{R}^d \to \mathbb{R}_+$ and $\eta, \vartheta \geq 0$. Then this point process is called a *Hawkes process* and the corresponding intensity is called a *Hawkes intensity*.

For simplicity, we refer to the process with intensity (1) as the *multivariate* case, whereas for (2) as the *vector-valued* case.

Remark 1. It is indeed possible to define a marked point process by specifying its intensity process. To this end, the intensity process needs to satisfy some formal requirements; see e.g. Definition 6.10 and Definition 6.13 in [11]. On the other hand, the pure specification of an intensity process leaves open questions of existence and uniqueness. In the case of a Hawkes process, these questions can be answered successfully. Sufficient conditions are given in Condition 1 below.

The form of (1) and (2) is closely related to the underlying branching structure. Every point of a Hawkes process is either an immigrant or a descendant. The *immigration intensities* η govern the frequency at which new immigrants arrive. Whenever a point event occurs, be it an immigrant or a descendant, the intensity is increased temporarily, i.e. points arrive at a higher frequency for some time. This intensity increase causes secondary point events, which in turn can spawn descendants of their own. How fast this effect decays in time is governed by the *decay functions* ω. The amount by which the intensity increases does not only depend on the time lag, but also on the mark value of the triggering point. The impact of a point event is determined by the *impact functions* g. Moreover, the *branching coefficients* ϑ specify the mean expected number of descendants of a given point event.

Remark 2. (*Branching Structure.*) Let $d \geq 1$ and $j, k \in \{1, \ldots, d\}$.

1. *Immigration intensities.* As the name implies, the immigration intensities η_j, η govern the frequency at which new immigrants arrive.

2. *Decay functions.* In the multivariate case, ω_j admits the following interpretation: Assume a point has occurred at time $s \in \mathbb{R}$ in component k and fix some $t > s$, such that $\Delta t := t - s$ is the time lag. The intensities of all other components $j \in \{1, \ldots, d\}$ at time t are increased proportionally to $\omega_j(\Delta t)$. An analogous interpretation applies to the vector-valued case.

3. Interpretation of g_k in the multivariate case: Assume the triggering point event is part of component d_i and has mark value x_i. The intensity of all other components is then increased proportionally to $g_{d_i}(x_i)$.

   Interpretation of g in the vector-valued case: Assume the triggering point event has the vector-valued mark $\boldsymbol{x}_i \in \mathbb{R}^d$. The intensity is then increased proportionally to $g(\boldsymbol{x}_i)$.

4. *Branching coefficients.* Interpretation of ϑ_{jk} in the multivariate case: Given that the point event belongs to component k, the intensity of component j is increased proportionally to ϑ_{jk}. In the vector-valued case, every point event increases the intensity proportionally to ϑ.

Definition 2. (*Branching Matrix.*) In the multivariate case define the $d\times d$-matrix $Q := \big(\vartheta_{jk}; j, k \in \{1, \ldots, d\}\big)$; and in the vector-valued case define the 1×1-matrix $Q := \vartheta$. The matrix Q is called the *branching matrix*.

The intensity process given in Definition 1 is the *time*-intensity process. It describes the dynamics of the ground process only, i.e. the process without the marks. For a full specification of a marked point process, one needs to know the (*time*,*space*)-intensity process. But in our situation, it suffices to specify the mark distribution; see also Definition 6.4.III.(b) in [6].

Definition 3. (*Mark Distribution.*) Let F_j be a family of distribution functions on $\mathbb{R}$ with corresponding densities f_j, for $j \in \{1, \ldots, d\}$.

1. *Multivariate case.* Given that a point event is part of component j, the associated mark X_j is independent of the past of the process and has density f_j.

2. *Vector-valued case.* Any newly generated point has a vector-valued mark $\boldsymbol{X} \in \mathbb{R}^d$ attached to it; this mark $\boldsymbol{X}$ is independent of the past of the process and has joint density f. We assume the joint density is given through a copula C with density c, namely,

$$f(\boldsymbol{x}) = c\big(F_1(x_1), \ldots, F_d(x_d)\big) \prod_{i=1}^{d} f_i(x_i), \quad \boldsymbol{x} \in \mathbb{R}^d.$$

   For an introduction to copulas, see Chapter 5 in [12].

Condition 1. (*Normalizing Conditions.*)

1. *Multivariate case.* For all $j, k \in \{1, \ldots, d\}$ assume that:

$$\int_0^\infty \omega_j(t)dt = 1 \qquad \text{and} \qquad \int_{-\infty}^{\infty} g_k(x) f_k(x) dx = 1.$$

2. *Vector-valued case.* Assume that:

$$\int_0^\infty \omega(t)dt = 1 \qquad \text{and} \qquad \int_{\mathbb{R}^d} g(\boldsymbol{x})f(\boldsymbol{x})d\boldsymbol{x} = 1.$$

The purpose of Condition 1 is two-fold: It allows one to formulate Proposition 1 below in a nice and compact form and improves the stability of numerical calculations. On the negative side, imposing these conditions will in general lead to more complicated impact functions g; see Section 1.3 in [11]. From now on we assume Condition 1 is satisfied.

Proposition 1. (Existence and Uniqueness.) *Suppose that the following conditions hold:*

1. *The spectral radius of the branching matrix satisfies:* $\mathrm{Spr}(Q) < 1$.
2. *The decay functions satisfy:*

$$\int_0^\infty t\omega_j(t)dt < \infty, \text{ for all } j \in \{1,\dots,d\}, \text{ in the multivariate case and}$$
$$\int_0^\infty t\omega(t)dt < \infty, \text{ in the vector-valued case.}$$

Then there exists a unique point process with associated intensity process as in Definition 1. Existence here means that one can find a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ *which is rich enough to support such a process. Uniqueness means that any two processes complying with Definition 1 and the above conditions have the same distribution.*

Proof. See Theorem 6.55 in [11].

Assumption 2 in Proposition 1 is actually a stronger requirement than what is needed from a minimalistic point of view. It does guarantee a strong coupling property which is desirable for numerical calculations. More about the different types of convergence and stability is explained in Definition 1 in [4] and subsequent remarks therein; see also Proposition 6.43 and the proof of Theorem 6.55 in [11].

3. Parameter Estimation, Goodness-of-Fit Tests and Simulation

The standard way of estimating the parameters of a Hawkes process is the maximum likelihood method. In order to define the likelihood function, let us fix an *observation period* $D := [T_*, T^*]$, i.e. the time interval during which empirical data have been collected.

Definition 4. (*Compensator.*) For all $t \in D$ define the *compensator* in the multivariate and vector-valued case, respectively, by:

$$\Lambda_j(t) := \int_{T_*}^t \lambda_j(s)ds, \text{ for } j \in \{1,\dots,d\} \qquad \text{and} \qquad \Lambda(t) := \int_{T_*}^t \lambda(s)ds.$$

By substituting the definition of the intensity processes, expanded expressions for the compensators are as follows:

1. *Multivariate case.* For $j \in \{1, \dots, d\}$ and $t \in D$:

$$\Lambda_j(t) = \eta_j(t-T_*) + \sum_{k=1}^{d} \vartheta_{jk} \int_{(-\infty,t)\times\mathbb{R}} \Big[\bar{\omega}_j(t-u) - \bar{\omega}_j(T_*-u)\Big] g_k(x) N_k(du \times dx). \tag{3}$$

2. *Vector-valued case.* For $t \in D$:

$$\Lambda(t) = \eta(t-T_*) + \vartheta \int_{(-\infty,t)\times\mathbb{R}^d} \Big[\bar{\omega}(t-u) - \bar{\omega}(T_*-u)\Big] g(\boldsymbol{x}) N(du \times d\boldsymbol{x}). \tag{4}$$

The functions $\bar{\omega}_j$ and $\bar{\omega}$ are defined for $t \geq 0$ by

$$\bar{\omega}_j(t) := \int_0^t \omega_j(s)ds, \text{ for } j \in \{1, \dots, d\} \qquad \text{and} \qquad \bar{\omega}(t) := \int_0^t \omega(s)ds,$$

with the convention $\bar{\omega}_j(t) = \bar{\omega}(t) = 0$, if $t < 0$.

Proposition 2. (Hawkes Likelihood Function.) *Let N be a Hawkes process which has been observed in the time interval $D = [T_*, T^*]$.*

1. Multivariate case. *The log-likelihood is given by*

$$\begin{aligned} \log L = &\sum_{j=1}^{d} \int_{D\times\mathbb{R}} \log \lambda_j(t) N_j(dt \times dx) \\ &+ \sum_{j=1}^{d} \int_{D\times\mathbb{R}} \log f_j(x) N_j(dt \times dx) - \sum_{j=1}^{d} \Lambda_j(T^*). \end{aligned} \tag{5}$$

2. Vector-valued case. *The log-likelihood is given by*

$$\begin{aligned} \log L = &\int_{D\times\mathbb{R}^d} \log \lambda(t) N(dt \times d\boldsymbol{x}) \\ &+ \sum_{i=1}^{d} \int_{D\times\mathbb{R}} \log f_i(x_i) N(dt \times d\boldsymbol{x}) - \Lambda(T^*) \\ &+ \int_{D\times\mathbb{R}} \log c\big(F_1(x_1), \dots, F_d(x_d)\big) N(dt \times d\boldsymbol{x}). \end{aligned} \tag{6}$$

Proof. The general case is proved in [11], Proposition 6.2 and the special case for copula structured marks is derived in [10], Proposition 3.2.

Standard numerical maximization algorithms can now be used to estimate the parameters of a corresponding Hawkes model. It then remains to assess the goodness-of-fit of an estimated Hawkes model, which is what we will discuss next. Bear in mind that the methods mentioned below are not specific to Hawkes processes alone, but can be used for other point processes as well. The basic idea is to construct the so-called *residual process* and compare the observed residual process with its theoretically expected counterpart.

Definition 5. (*Residual Process.*) Assume that one has observed n points of a Hawkes process in the time interval D.

1. *Multivariate case.* Recall that the points are of the form (t_i, d_i, x_i), where $1 \leq i \leq n$. Define the sequence $(\tau_1, \ldots, \tau_n)$ of transformed times by:

$$\tau_i := \Lambda_{d_i}(t_i).$$

   Now for $j \in \{1, \ldots, d\}$, define the j-th *residual process* R_j as the point process consisting of all τ_i which lie in component j, i.e. where $d_i = j$.

2. *Vector-valued case.* Recall that the points are of the form $(t_i, x_{i,1}, \ldots, x_{i,d})$. Define the sequence of transformed times $(\tau_1, \ldots, \tau_n)$ by:

$$\tau_i := \Lambda(t_i).$$

   The *residual process* R is then defined to be the point process consisting of the n transformed times τ_i, $i = 1, \ldots, n$.

Proposition 3. (Random Time Change.) *Take the observation period $D := [T_*, \infty)$ and consider a Hawkes process with strictly positive immigration intensities $\eta_j, \eta > 0$.*

1. Multivariate case. *The residual processes R_j, $j = 1, \ldots, d$, are independent Poisson processes with unit intensity.*

2. Vector-valued case. *The residual process R is a Poisson process with unit intensity.*

Proof. This statement goes back to [18], [13] and [17]. Modern formulations are given in [3], Theorem T16 of Section II.6 and in [6], Theorem 7.4.I.

The concept of random time change is not only valuable in goodness-of-fit analysis, but it also forms the basis for a simulation algorithm:

Remark 3. (*Simulation of a Hawkes Process.*) Consider first the multivariate case: One way to simulate a multivariate Hawkes process is by using a multivariate extension of Algorithm 7.5.IV in [6]. It is commonly called *Ogata's modified thinning algorithm.* The adaption of this algorithm to Hawkes processes is given in [11], Algorithm 1.21. A corresponding algorithm for the vector-valued case is more straightforward, since one basically has to simulate a one-dimensional Hawkes process plus a vector-valued mark from a d-dimensional copula with given marginal distributions; see Algorithm 3.1-3.3 in [10].

4. Two Illustrative Examples

The examples below highlight the statistical fitting procedures and goodness-of-fit tests. The second example mainly stresses the differences between the multivariate and vector-valued cases rather than giving a detailed statistical analysis. The software package underlying the applications in this section is written in R and C++ and is available from the second author upon request.

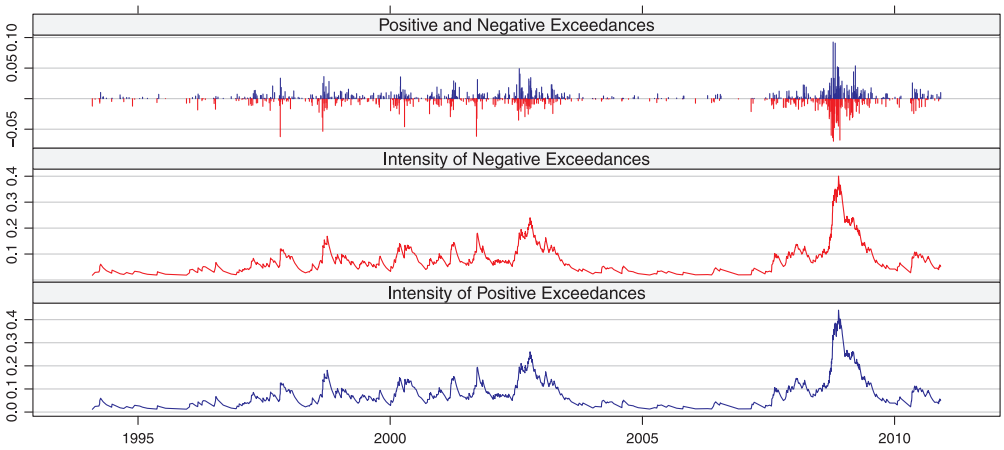

FIGURE 1: Estimated Hawkes process, multivariate case (Example 4.1).

4.1. Multivariate Hawkes Process

As an example we fit a multivariate Hawkes process to daily stock market index data. More precisely, we consider daily closing values from the Dow Jones Industrial Average from 1994-01-01 to 2010-12-31. Our aim is to capture extremal clustering in the context of multivariate Hawkes processes. The Hawkes process model is very versatile and offers many possibilities to calibrate the model to the characteristics of the data. To keep the exposition manageable, we focus only on one specific model. We did however look at other specifications which turned out to be suitable as well. We make the following assumptions on decay and impact functions and mark distributions:

- *Point process structure.* Starting with the daily log-return data, we fix two thresholds, given by the 10% and 90% quantiles. We retain only those days for which the return is either below the 10% threshold or above the 90% threshold. All other days are removed. This procedure leads to two point process components ($d = 2$), corresponding to the negative and positive excesses. The mark values attached to the times are the absolute values of the excesses.
- *Decay function.* We use exponential decay functions $\omega_j(t) = \delta_j \exp\{-\delta_j t\}$, but make the additional assumption that the decay speeds of the two components coincide, i.e. that $\delta = \delta_1 = \delta_2 > 0$. As a side remark: Choosing this decay function has the nice property that the Hawkes process becomes a Markov process; see [11], Remark 1.22. This is an advantage for numerical computations.
- *Mark distributions.* We assume the marks, i.e. the threshold excesses, have an exponential distribution with parameter $\lambda_j > 0$, for $j \in \{1, 2\}$. This choice is partly based on extreme value theory as discussed in Section 3.4 in [7] and Section 7.3.2 in [12], where the exponential distribution appears as the excess limit in the Gumbel case. Furthermore, the exponential distribution facilitates the numerics in the construction of parametric bootstrap confidence intervals below. Another relevant example (with slower convergence for the bootstrap) would have been the Pareto distribution.
- *Impact functions.* In order to cover constant and linear impact effects, we take

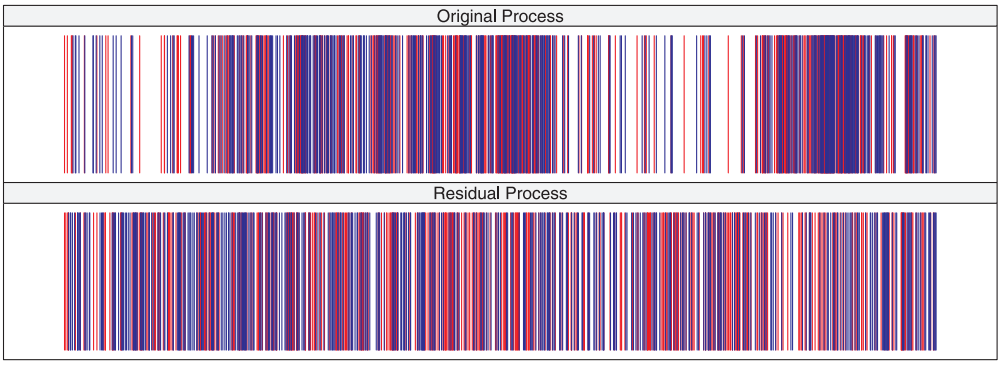

FIGURE 2: Barcode plot (Example 4.1).

the impact functions $\tilde{g}_k(x) = \alpha_k + \beta_k x$, where $\alpha_k, \beta_k \geq 0$, for $k \in \{1, 2\}$. Since these functions do not satisfy Condition 1, we actually need to take the following *normalized impact functions* instead:

$$g_k(x) = \frac{\tilde{g}_k(x)}{\mathbb{E}[g_k(X)]} = \frac{\alpha_k + \beta_k x}{\alpha_k + \beta_k \mathbb{E}[X]} = \frac{\lambda_k^2}{\alpha_k \lambda_k^2 + \beta_k \lambda_k}(\alpha_k + \beta_k x).$$

A closer examination shows that g_k has only one degree of freedom in the parameters (α_k, β_k). Hence we may set $\alpha_k = 1$, wlog.

The full Hawkes model has 6 more parameters, namely two immigration intensities η_1, η_2 and four branching coefficients ϑ_{jk}, for $j, k = 1, 2$. This leads in total to the following 11 parameters:

$$\eta_1 \quad \eta_2 \quad \vartheta_{11} \quad \vartheta_{12} \quad \vartheta_{21} \quad \vartheta_{22} \quad \beta_1 \quad \beta_2 \quad \delta \quad \lambda_1 \quad \lambda_2.$$

By using maximum likelihood parameter estimation, as explained in Section 3, the following parameter estimates are obtained:

$$\begin{array}{llllll} \eta_1 = 0.018 & \eta_2 = 0.012 & \vartheta_{11} = 0.74 & \vartheta_{12} = 0 & \vartheta_{21} = 0.83 & \vartheta_{22} = 0 \\ \delta = 0.021 & \beta_1 = 47 & \beta_2 = 74 & \lambda_1 = 109 & \lambda_2 = 122. & \end{array}$$

In order to obtain approximate confidence intervals, we applied a parametric bootstrapping method for the more important parameters. Based on the estimated parameters, 1000 sample paths of the same length as the original data set are generated, and then the parameters re-estimated. This leads to the following 95%-confidence intervals:

$$\begin{array}{llll} \vartheta_{11} \in [0.42, 0.82] & \vartheta_{12} \in [0, 0.21] & \vartheta_{21} \in [0.62, 0.94] & \vartheta_{22} \in [0, 0.19] \\ \eta_1 \in [0.012, 0.032] & \eta_2 \in [0.005, 0.022] & \delta \in [0.017, 0.030]. & \end{array}$$

These show a fair amount of uncertainty in the parameter estimates, indicating that the data set is still relatively small for an accurate fitting of a Hawkes model. The above parameter estimates yield a point estimate of $\mathrm{Spr}(Q) = 0.74$, with a confidence interval of $I := [0.52, 0.83]$. Since $I \subseteq [0, 1)$, the Hawkes process is well-defined. The

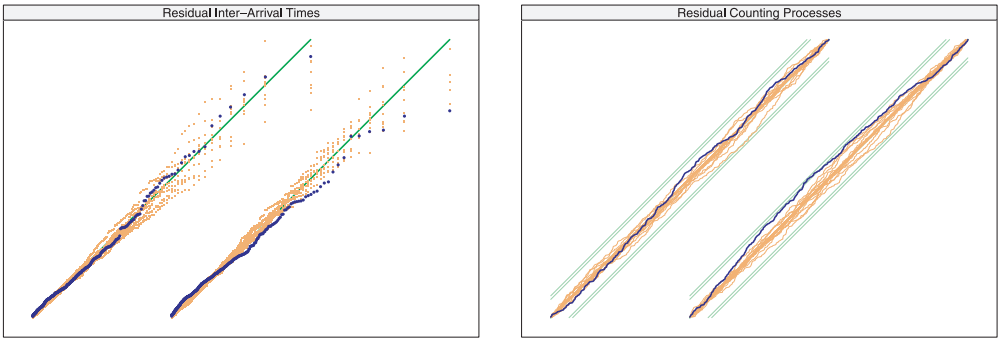

FIGURE 3: Goodness-of-fit plots for residual inter-arrival times and residual counting process (Example 4.1).

construction of asymptotic confidence intervals based on the Fisher information matrix is a topic for further research.

We now look at the estimated Hawkes process graphically: Depicted in the first panel of Figure 1 is the point process consisting of negative (red) and positive (blue) threshold excesses, as explained above. The second panel shows the estimated intensity process of the first component of the Hawkes process, i.e. the component which corresponds to the negative excesses, and the third panel shows the intensity process of the positive excesses. The two intensity processes are clearly quite similar. Indeed, this is expected as the estimated spectral radius is not too far from 1, and hence a strong coupling between the two components exists.

Next we look at some graphical goodness-of-fit tests which are motivated by the random time change as explained in Proposition 3. We start with what we call a *barcode plot*: In the top panel of Figure 2, the original point process is shown. A line has been drawn for each event, in red and blue for the two components N_1 and N_2, respectively. The vertical axis in this plot is for visualization only (all lines have the same height) and has no further meaning. In contrast, the bottom panel shows the two components of the residual process. The theory suggests that it should consist of two independent Poisson processes. Whether this is so may be difficult to assess from this plot. What is clearly visible is that the clusters in the original point process have disappeared and the residual point process exhibits no obvious clusters any more.

Next we look at the inter-arrival times of the residual process. The left hand panel of Figure 3 shows a Q-Q-plot of these inter-arrival times against the standard exponential distribution, separately for the two components of the residual process. In brighter color some genuine exponential random values have been added, in order to give a visual clue of what deviations one can reasonably expect. The right hand panel of Figure 3 shows the associated counting functions of the two components, i.e. the number of events versus time. Some true Poisson processes have been added in a lighter color, again as a visual help. Additionally, there are two confidence bands for the Poisson null hypothesis, i.e. bands which a genuine Poisson process does not cross on a 95% and 99% level. We omit here further tests for the quality of the estimated mark distributions, as such tests are not specific to Hawkes processes.

One final visual test is a comparison between the original point process and a random sample path of the estimated Hawkes process. This has the following reasoning: At

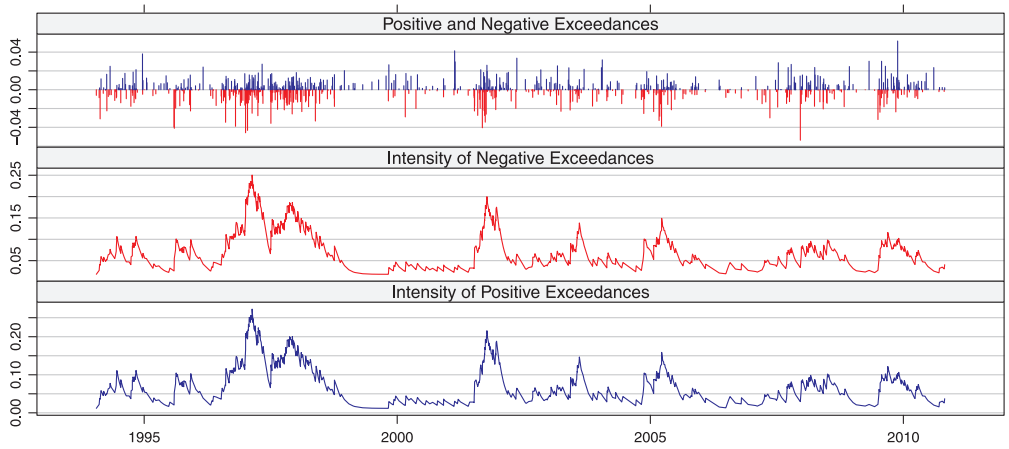

FIGURE 4: Simulated Hawkes process, multivariate case (Example 4.1).

times even an inappropriate choice of a Hawkes process is able to detect the apparent clusters in a data set. At a first glance, there would be no reason to reject such a Hawkes model. But the characteristics of such a Hawkes process may be quite different from what one would expect at first. This is why we take a time interval of the same length as the original data set and simulate a random path on this interval, using the estimated Hawkes process. Note that this has nothing to do with a *prediction*. The only purpose of this is to check whether the clustering behavior of the estimated process is indeed what the estimated intensity process suggests.

A simulated path, see Remark 3, of the estimated Hawkes process is shown in Figure 4. The negative and positive marks are again combined in the first panel and the two intensity processes are shown in the two bottom panels. Note that the simulated version looks comparable to the estimated one in Figure 1, or at least, no serious discrepancy stands out.

4.2. Hawkes Process with Vector-Valued Marks

For this example, we consider the following three indices: the Dow Jones Industrial Average, the Nasdaq-100 and the SP500 Composite. The data consists of hourly observations from 1997-10-01 to 2010-03-01. Denote the log-returns of the three indices as $r_{i,1}, r_{i,2}, r_{i,3}$, for $i \in \{1, \dots, n\}$.

- *Time events.* Consider an *equally weighted* portfolio consisting of the three above indices, and define
$$p_i := r_{i,1} + r_{i,2} + r_{i,3}.$$
Similar to the first example, we take the 1% and 99% empirical quantiles for the p_i-series. But this time we are not interested in the excesses, instead we simply record the exceedance times t_i.
- *Mark values.* At each exceedance time t_i, we define the following three-dimensional mark $\boldsymbol{x}_i \in \mathbb{R}^3$:
$$x_{i,1} = |r_{i,1}|, \qquad x_{i,2} = |r_{i,2}|, \qquad x_{i,3} = |r_{i,3}|.$$

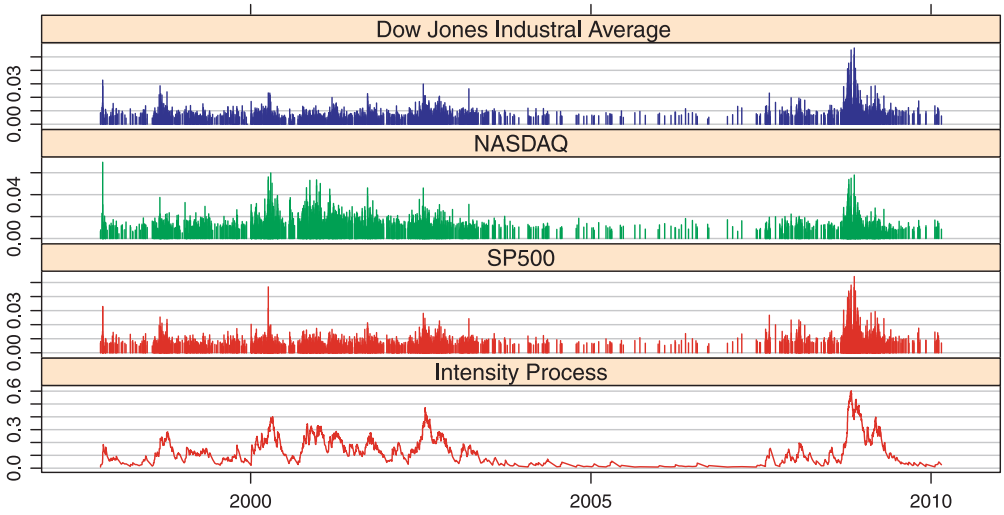

FIGURE 5: Estimated Hawkes process, vector-valued case (Example 4.2).

Hence we obtain the following sequence of observations:

$$(t_1, x_{1,1}, x_{1,2}, x_{1,3}),\ (t_2, x_{2,1}, x_{2,2}, x_{2,3}),\ \dots,\ (t_n, x_{n,1}, x_{n,2}, x_{n,3}).$$

- *Point process structure.* Clearly this sequence has a structure different from that in the first example. Indeed there are now three (simultaneous) marks attached to each point event. This is caused by the fact that the event is not defined separately for each component, but rather through a linear combination. Consequently, we are now dealing with a point process with vector-valued marks.
- *Decay function.* Again we take $\omega(t) = \delta \exp\{-\delta t\}, \delta > 0$.
- *Mark distribution.* We assume the marks follow a *gamma distribution* with parameters $\zeta_j, \sigma_j > 0$:

$$f_j(x) = x^{\zeta_j - 1} \frac{\exp\{-x/\sigma_j\}}{\Gamma(\zeta_j)\sigma_j^{\zeta_j}}, \quad x > 0.$$

  There are in total 6 parameters $\zeta_j, \sigma_j, j = 1, 2, 3$.
- *Impact function.* We choose an additive form for the impact function g, which simplifies the normalizing Condition 1:

$$g(\boldsymbol{x}_i) := \frac{1}{d} \sum_{k=1}^{d} g_k(x_{i,k}),$$

  where $g_k : \mathbb{R} \to \mathbb{R}_+$. More specifically, we choose polynomial functions $\tilde{g}_k(x) = \alpha_k + \beta_k x + \gamma_k x^2$, and associated normalized impact functions

$$g_k(x) = \frac{\tilde{g}_k(x)}{\mathbb{E}[g_k(X)]} = \frac{\alpha_k + \beta_k x + \gamma_k x^2}{\alpha_k + \beta_k \zeta_k \sigma_k + \gamma_k \zeta_k (\zeta_k + 1)\sigma_k^2}.$$

  Hence there are 9 parameters $\alpha_k, \beta_k, \gamma_k \geq 0,\ k = 1, 2, 3$.

- *Copula.* We assume that our marks admit a 3-dimensional *Gauss-copula* with an equicorrelation structure, i.e. the off-diagonal entries of the correlation matrix are all equal to $\rho \in [0, 1)$.

Finally, there are two more parameters: The branching coefficient ϑ and the immigration intensity η. In total we have 19 parameters to estimate:

$$\vartheta \quad \eta \quad \delta \quad \alpha_1 \quad \beta_1 \quad \gamma_1 \quad \alpha_2 \quad \beta_2 \quad \gamma_2 \quad \alpha_3 \quad \beta_3 \quad \gamma_3 \quad \zeta_1 \quad \sigma_1 \quad \zeta_2 \quad \sigma_2 \quad \zeta_3 \quad \sigma_3 \quad \rho.$$

Using again maximum likelihood, the following parameter estimates are obtained:

$$\begin{array}{lllllll} \vartheta = 0.914 & \eta = 0.008 & \delta = 0.017 & \alpha_1 = 1.821 & \beta_1 = 13.85 & \gamma_1 = 2.163 & \alpha_2 = 5.928 \\ \beta_2 = 13.87 & \gamma_2 = 0.700 & \alpha_3 = 12.83 & \beta_3 = 25.61 & \gamma_3 = 2.601 & \zeta_1 = 2.671 & \sigma_1 = 0.006 \\ \zeta_2 = 5.185 & \sigma_2 = 0.002 & \zeta_3 = 4.13 & \sigma_3 = 0.002 & \rho = 0.475. & & \end{array}$$

The branching coefficient is estimated to be $\vartheta = 0.914$, which again suggests a well-defined Hawkes process. Note that this model allows for two different dependence structures: The frequency dependence caused by the shared intensity process, and the dependence between the different components, modelled through the notion of copula. For the latter dependence, note the estimated value of ρ. Figure 5 shows the 3-dimensional mark values and the estimated intensity process of the fitted Hawkes model.

Similar graphical goodness-of-fit tests, as well as a simulation comparison and the construction of parametric bootstrap confidence intervals, can now be made.

5. Conclusion

In this paper we have shown that multivariate Hawkes processes offer a versatile class of point processes capable of modelling extremal behavior of financial time series. From a mathematical point of view these processes are often perceived as rather involved. We have shown that numerical procedures for estimation and simulation can be successfully implemented and applied. The applicability of these procedures extends well beyond financial applications.

6. Acknowledgment

The authors would like to thank a referee for careful reading of an earlier version of the paper. Paul Embrechts, as Senior SFI Professor, acknowledges financial support of the Swiss Finance Institute.

References

[1] Y. Aït-Sahalia, J. Cacho-Diaz and R. J. A. Laeven, (2010). Modeling financial contagion using mutually exciting jump processes. *The Review of Financial Studies*, to appear.

[2] S. Azizpour, K. Giesecke and G. Schwenkler, (2010). Exploring the sources of default clustering. *Preprint,* Stanford University.

[3] P. Brémaud, (1981). *Point Processes and Queues: Martingale Dynamics*. Springer.

[4] P. Brémaud and L. Massoulié, (1996). Stability of nonlinear Hawkes processes. *The Annals of Probability,* 24(3):1563-1588.

[5] V. Chavez-Demoulin, A. C. Davison and A. J. McNeil, (2005). Estimating Value-At-Risk: A point process approach. *Quantitative Finance,* 5(2):227-234.

[6] D. J. Daley and D. Vere-Jones, 2nd ed., (2003). *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*. Springer.
[7] P. Embrechts, C. Klueppelberg and T. Mikosch, (1997). *Modelling Extremal Events for Insurance and Finance*. Springer.
[8] E. Errais, K. Giesecke and L. R. Goldberg, (2010). Affine point processes and portfolio credit risk. *SIAM J. Financial Math,* 1:642-665.
[9] A. G. Hawkes, (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika,* 58(1):83–90.
[10] L. Lin, (2010). *An Application of Multivariate Hawkes Process to Finance*. Master Thesis, Department of Mathematics, ETH Zurich.
[11] T. Liniger, (2009). *Multivariate Hawkes Processes*. Doctoral Thesis, Department of Mathematics, ETH Zurich.
[12] A. J. McNeil, R. Frey and P. Embrechts, (2005). *Quantitative Risk Management: Concepts, Techniques, Tools*. Princeton University Press.
[13] P.-A. Meyer, (1971). Démonstration simplifiée d'un théorème de Knight. *Séminaire de probabilités de Strasbourg,* 5:191-195.
[14] D. Oakes, (1975). The Markovian self-exciting process. *Journal of Applied Probability,* 12:69-77.
[15] Y. Ogata, (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association,* 83(401):9-27.
[16] Y. Ogata, (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics,* 50(2):379-402.
[17] F. Papangelou, (1972). Integrability of expected increments of point processes and a related change of scale. *Transactions of the American Mathematical Society,* 165:483-506.
[18] S. Watanabe, (1964). On discontinuous additive functionals and Lévy measures of a Markov process. Japanese Journal of Mathematics, 34:53-70.