

# POINT PROCESS MODELS FOR MULTIVARIATE HIGH-FREQUENCY IRREGULARLY SPACED DATA

01-30-2013

Stephen Crowley

stephen.crowley@hushmail.com

**ABSTRACT.** Definitions from the theory of point processes are recalled. Models of intensity function parametrization and maximum likelihood estimation from data are explored. Closed-form log-likelihood expressions are given for the (exponential) Hawkes (univariate and multivariate) process, Autoregressive Conditional Duration(ACD), with both exponential and Weibull distributed errors, and a hybrid model combining the ACD and the exponential Hawkes models. Formulas are also derived, however without the elegant recursions of the exponential kernels, for kernels of the Weibull and Gamma type and comparison of the Weibull fit vs exponential kernel fits via QQ and probability plots are provided. The additional complexity of the Hawkes-Weibull or the ACD-Hawkes appears to not be worth the tradeoff. Diurnal, or daily, adjustment of the deterministic predictable part of the intensity variation via piecewise polynomial splines is discussed. Data from the symbol SPY on three different electronic markets is used to estimate model parameters and generate illustrative plots. The parameters were estimated without diurnal adjustments, a repeat of the analysis with adjustments is due in a future version of this article. The connection of the Hawkes process to quantum theory is briefly mentioned. Prediction of the next point of a Hawkes process is briefly discussed and a closed-form expression in terms of the Lambert W function for the standard exponential kernel with  $P=1$  is calculated.

## TABLE OF CONTENTS

|  |           |
|--|-----------|
| <b>1. Defintions</b>   | <b>2</b>  |
| 1.1. Point Processes and Intensities                                 | 2         |
| 1.1.1. Stochastic Integrals  | 3         |
| 1.2. The Exponential Autoregressive Conditional Duration(EACD) Model | 3         |
| 1.2.1. The Weibull-ACD Model   | 5         |
| 1.3. The Hawkes Process  | 6         |
| 1.3.1. Linear Self-Exciting Processes                                | 6         |
| 1.3.2. The Hawkes(1) Model   | 9         |
| 1.3.3. Maximum Likelihood Estimation                                 | 9         |
| 1.3.4. Prediction  | 10        |
| 1.3.5. The Hawkes Process in Quantum Theory                          | 11        |
| 1.4. Alternative Kernels for the Hawkes Process                      | 11        |
| 1.4.1. A Generalized Exponential Kernel                              | 11        |
| 1.4.2. The Hawkes Process Having a Weibull Kernel                    | 12        |
| 1.4.3. The Hawkes Process Having a Gamma Kernel                      | 14        |
| 1.4.4. A Hyperbolic Kernel   | 17        |
| 1.5. Assessing Goodness of Fit with Graphical Methods                | 18        |
| 1.6. Combining the ACD and Hawkes Models                             | 22        |
| 1.6.1. Comparison of Hawkes-Exp vs ACD+Hawkes-Exp Model Fits         | 23        |
| 1.7. Multivariate Hawkes Models                                      | 26        |
| 1.7.1. The Compensator   | 27        |
| 1.7.2. Log-Likelihood  | 28        |
| <b>2. Numerical Methods</b>  | <b>29</b> |
| 2.1. The Nelder-Mead Algorithm                                       | 29        |

|   |    |
|---|----|
| 2.1.1. Starting Points for Optimizing the Hawkes Process of Order $P$ . . . . . | 29 |
| <b>3. Examples</b> . . . . .  | 29 |
| 3.1. Millisecond Resolution Trade Sequences . . . . .                           | 29 |
| 3.1.1. Adjusting for the Deterministic Daily Intensity Variation . . . . .      | 29 |
| 3.1.2. Univariate Hawkes model fit to SPY (SPDR S&P 500 ETF Trust) . . . . .    | 30 |
| 3.1.3. Multivariate SPY Data for 2012-08-14 . . . . .                           | 36 |
| 3.1.4. Multivariate SPY Data for 2012-11-19 . . . . .                           | 37 |
| <b>Bibliography</b> . . . . .   | 38 |
| <b>4. Appendix</b> . . . . .  | 39 |
| 4.1. The Lambert W Function $W(k, x)$ . . . . .                                 | 39 |

## 1. DEFINITIONS

### 1.1. Point Processes and Intensities.

Consider a  $K$  dimensional multivariate point process. Let  $N_t^k$  denote the *counting process* associated with the  $k$ -th point process which is simply the number of events which have occurred by time  $t$ . Let  $F_t$  denote the filtration of the pooled process  $N_t$  of  $K$  point processes consisting of the set  $t_0^k < t_1^k < t_2^k < \dots < t_i^k < \dots$  denoting the history of arrival times of each event type associated with the  $k=1 \dots K$  point processes. At time  $t$ , the most recent arrival time will be denoted  $t_{N_t^k}^k$ . A process is said to be simple if no points occur at the same time, that is, there are no zero-length durations. The counting process can be represented as a sum of Heaviside step functions  $\theta(t) = \begin{cases} 0 & t < 0 \\ 1 & t \geq 0 \end{cases}$

$$N_t^k = \sum_{t_i^k \leq t} \theta(t - t_i^k) \quad (1)$$

The *conditional intensity function* gives the conditional probability per unit time that an event of type  $k$  occurs in the next instant.

$$\lambda^k(t|F_t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(N_{t+\Delta t}^k - N_t^k > 0 | F_t)}{\Delta t} \quad (2)$$

For small values of  $\Delta t$  we have

$$\lambda^k(t|F_t)\Delta t = E(N_{t+\Delta t}^k - N_t^k | F_t) + o(\Delta t) \quad (3)$$

so that

$$E((N_{t+\Delta t}^k - N_t^k) - \lambda^k(t|F_t)\Delta t) = o(\Delta t) \quad (4)$$

and (4) will be uncorrelated with the past of  $F_t$  as  $\Delta t \rightarrow 0$ . Next consider

$$\begin{aligned} & \lim_{\Delta t \rightarrow 0} \sum_{j=1}^{\frac{(s_1-s_0)}{\Delta t}} (N_{s_0+j\Delta t}^k - N_{s_0+(j-1)\Delta t}^k) - \lambda^k(s_0+j\Delta t|F_t)\Delta t \\ &= \lim_{\Delta t \rightarrow 0} (N_{s_0}^k - N_{s_1}^k) - \sum_{j=1}^{\frac{(s_1-s_0)}{\Delta t}} \lambda^k(j\Delta t|F_t)\Delta t \\ &= (N_{s_0}^k - N_{s_1}^k) - \int_{s_0}^{s_1} \lambda^k(t|F_t)dt \end{aligned} \quad (5)$$

which will be uncorrelated with  $F_{s_0}$ , that is

$$E\left(\int_{s_0}^{s_1} \lambda^k(t|F_t)dt\right) = N_{s_0}^k - N_{s_1}^k \quad (6)$$

The integrated intensity function is known as the *compensator*, or more precisely, the  $F_t$ -*compensator* and will be denoted by

$$\Lambda^k(s_0, s_1) = \int_{s_0}^{s_1} \lambda^k(t|F_t)dt \quad (7)$$

Let  $x_k = t_i^k - t_{i-1}^k$  denote the time interval, or duration, between the  $i$ -th and  $(i-1)$ -th arrival times. The  $F_t$ -*conditional survivor function* for the  $k$ -th process is given by

$$S_k(x_i^k) = P_k(t_i^k > x_i^k | F_{t_{i-1}+\tau}) \quad (8)$$

Let

$$\tilde{\mathcal{E}}_i^k = \int_{t_{i-1}}^{t_i} \lambda^k(t|F_t)dt = \Lambda^k(t_{i-1}, t_i)$$

then provided the survivor function is absolutely continuous with respect to Lebesgue measure (which is an assumption that needs to be verified, usually by graphical tests) we have

$$S^k(x_i^k) = e^{-\int_{t_{i-1}}^{t_i} \lambda^k(t|F_t)dt} = e^{-\tilde{\mathcal{E}}_i^k} \quad (9)$$

and  $\tilde{\mathcal{E}}_{N(t)}$  is an i.i.d. exponential random variable with unit mean and variance. Since  $E(\tilde{\mathcal{E}}_{N(t)}) = 1$  the random variable

$$\mathcal{E}_{N(t)}^k = 1 - \tilde{\mathcal{E}}_{N(t)} \quad (10)$$

has zero mean and unit variance. Positive values of  $\mathcal{E}_{N(t)}$  indicate that the path of conditional intensity function  $\lambda^k(t|F_t)$  under-predicted the number of events in the time interval and negative values of  $\mathcal{E}_{N(t)}$  indicate that  $\lambda^k(t|F_t)$  over-predicted the number of events in the interval. In this way, (8) can be interpreted as a generalized residual. The *backwards recurrence time* given by

$$U^{(k)}(t) = t - t_{N^k(t)} \quad (11)$$

increases linearly with jumps back to 0 at each new point.

### 1.1.1. Stochastic Integrals.

The *stochastic Stieltjes integral* [1, 2.1][10, 2.2] of a measurable process, having either locally bounded or nonnegative sample paths,  $X(t)$  with respect to  $N^k$  exists and for each  $t$  we have

$$\int_{(0,t]} X(s)dN_s^k = \sum_{i \geq 1} \theta(t - t_i^k) X(t_i^k) \quad (12)$$

### 1.2. The Exponential Autoregressive Conditional Duration(EACD) Model.

Letting  $p_i$  be the family of conditional probability density functions for arrival time  $t_i$ , the log likelihood of the (exponential) ACD model can be expressed in terms of the conditional densities or intensities as [13]

$$\begin{aligned} \ln \mathcal{L}(\{t_i\}_{i=0 \dots n}) &= \sum_{i=0}^n \log p_i(t_i | t_0, \dots, t_{i-1}) \\ &= \left( \sum_{i=1}^n \log \lambda(t_i | i-1, t_0, \dots, t_{i-1}) \right) - \int_{t_0}^{t_n} \lambda(u | n, t_0, \dots, t_{N_u}) du \\ &= \left( \sum_{i=1}^n \log \lambda(t_i | i-1, t_0, \dots, t_{i-1}) - \int_{t_{i-1}}^{t_i} \lambda(u | n, t_0, \dots, t_{N_u}) du \right) \\ &= \left( \sum_{i=1}^n \log \lambda(t_i | i-1, t_0, \dots, t_{i-1}) - \tilde{\mathcal{E}}_i \right) \\ &= \int_{t_0}^{t_n} \ln \lambda(t) dN_t - \int_{t_0}^{t_n} \lambda(t) dt \end{aligned} \quad (13)$$

We will see that  $\lambda$  can be parameterized in terms of

$$\lambda(t|N_t, t_1, \dots, t_{N_t}) = \omega + \sum_{i=1}^{N_t} \pi_i(t_{N_t+1-i} - t_{N_t-i}) \quad (14)$$

so that the impact of a duration between successive events depends upon the number of intervening events. Let  $x_i = t_i - t_{i-1}$  be the interval between consecutive arrival times; then  $x_i$  is a sequence of durations or “waiting times”. The conditional density of  $x_i$  given its past is then given directly by

$$E(x_i|x_{i-1}, \dots, x_1) = \psi_i(x_{i-1}, \dots, x_1; \theta) = \psi_i \quad (15)$$

Then the ACD models are those that consist of the assumption

$$x_i = \psi_i \varepsilon_i \quad (16)$$

where  $\varepsilon_i$  is independently and identically distributed with density  $p(\varepsilon; \phi)$  where  $\theta$  and  $\phi$  are variation free. ACD processes are limited to the univariate setting but later we will see that this model can be combined with a Hawkes process in a multivariate framework. [7] The conditional intensity of an ACD model can be expressed in general as

$$\lambda(t|N_t, t_1, \dots, t_{N_t}) = \lambda_0\left(\frac{t - t_{N_t}}{\psi_{N_t+1}}\right) \frac{1}{\psi_{N_t+1}} \quad (17)$$

where  $\lambda_0(t)$  is a deterministic baseline hazard, so that the past history influences the conditional intensity by both a multiplicative effect and a shift in the baseline hazard. This is called an *accelerated failure time* model since past information influences the rate at which time passes. The simplest model is the exponential ACD which assumes that the durations are conditionally exponential so that the baseline hazard  $\lambda_0(t) = 1$  and the conditional intensity is

$$\lambda(t|x_{N_t}, \dots, x_1) = \frac{1}{\psi_{N_t+1}} \quad (18)$$

The compensator for consecutive events of the ACD model in the case of constant baseline intensity  $\lambda_0(t) = 1$  is simply

$$\begin{aligned} \tilde{\mathcal{E}}_i &= \Lambda^k(t_{i-1}, t_i) \\ &= \int_{t_{i-1}}^{t_i} \lambda(t|x_i, \dots, x_1) dt \\ &= \int_{t_{i-1}}^{t_i} \frac{1}{\psi_{N_t+1}} dt \\ &= \int_{t_{i-1}}^{t_i} \frac{1}{\psi_i} dt \\ &= \frac{t_{i-1} - t_i}{\psi_i} \\ &= \frac{x_i}{\psi_i} \end{aligned} \quad (19)$$

where  $x_i = t_i - t_{i-1}$ . A general model without limited memory is referred to as ACD( $m, q$ ) where  $m$  and  $q$  refer to the order of the lags so that there are  $(m + q + 1)$  parameters.

$$\psi_i = \omega + \sum_{j=1}^m \alpha_j x_{i-j} + \sum_{j=1}^q \beta_j \psi_{i-j} \quad (20)$$

where  $\omega \geq 0, \alpha_j \geq 0, \beta_j \geq 0$  and  $\psi_i = \frac{\omega}{1 - \sum_{j=1}^q \beta_j}$  for  $i = 1 \dots \max(m, q)$  so the conditional intensity is then written

$$\lambda(t|x_{N_t}, \dots, x_1) = \frac{1}{\omega + \sum_{j=1}^m \alpha_j x_{N_t+1-j} + \sum_{j=1}^q \beta_j \psi_{N_t+1-j}} \quad (21)$$

The log-likelihood for the ACD(m,q) model is then written in terms of the durations  $x_i = t_i - t_{i-1}$

$$\begin{aligned} \ln \mathcal{L}(\{x_i\}_{i=1, \dots, n}) &= \left( \sum_{i=1}^n \ln \lambda(t_i | i-1, t_0, \dots, t_{i-1}) - \tilde{\mathcal{E}}_i \right) \\ &= \sum_{i=1}^n \ln \left( \frac{S(x_i)}{\psi_i} \right) \\ &= \sum_{i=1}^n \ln \left( \frac{e^{-\tilde{\mathcal{E}}_i}}{\psi_i} \right) \\ &= \sum_{i=1}^n \ln \left( \frac{e^{-\frac{x_i}{\psi_i}}}{\psi_i} \right) \\ &= \sum_{i=1}^n \ln \left( \frac{1}{\psi_i} \right) - \frac{x_i}{\psi_i} \end{aligned} \quad (22)$$

An ACD process is stationary if

$$\sum_{i=1}^m \alpha_j + \sum_{i=1}^q \beta_j < 1 \quad (23)$$

in which case the unconditional mean exists and is given by

$$\mu = E[x_i] = \frac{\omega}{1 - (\sum_{i=1}^m \alpha_j + \sum_{i=1}^q \beta_j)} \quad (24)$$

The goodness of fit can be checked by testing that residuals  $\tilde{\mathcal{E}}_i$  have mean and variance equal to 1 and no autocorrelation.

### 1.2.1. The Weibull-ACD Model.

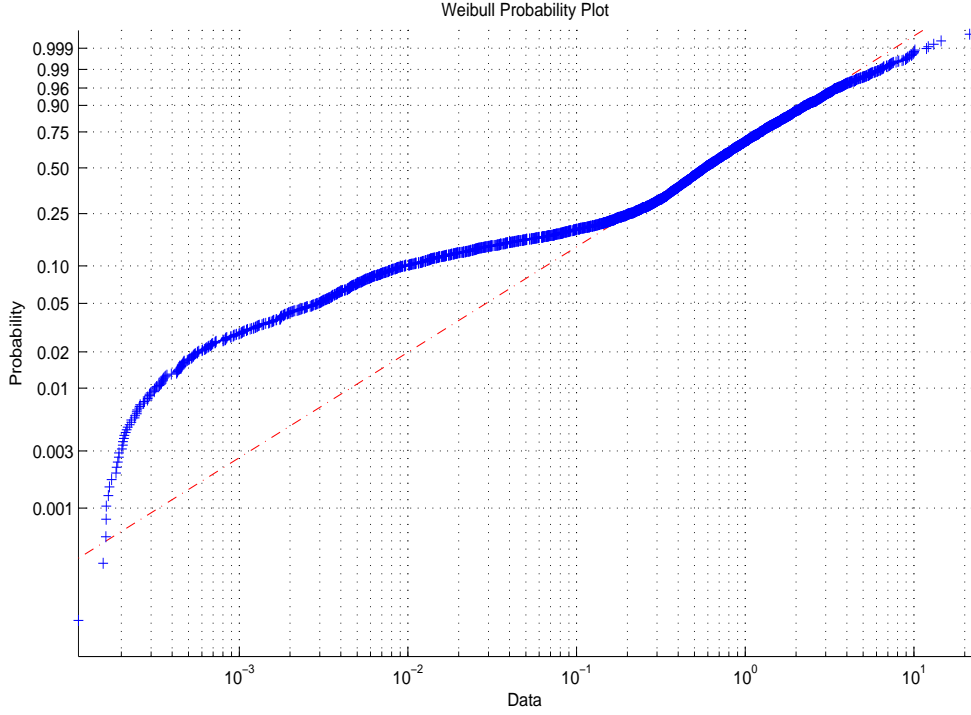
The WACD(Weibull-ACD) model extends the EACD model by assuming a Weibull distribution for the residuals  $\varepsilon_i$  in (16) instead of an exponential. We have the intensity given by

$$\lambda(t|x_{N_t}, \dots, x_1) = \left( \frac{\Gamma\left(1 + \frac{1}{\gamma}\right)}{\psi_{N_t+1}} \right)^\gamma (t - t_{N_t})^{\gamma-1} \gamma \quad (25)$$

and log-likelihood by

$$\ln \mathcal{L}(\{x_i\}_{i=1, \dots, n}) = \sum_{i=1}^n \ln \left( \frac{\gamma}{x_i} \right) + \gamma \ln \left( \frac{\Gamma\left(1 + \frac{1}{\gamma}\right) x_i}{\psi_i} \right) - \left( \frac{\Gamma\left(1 + \frac{1}{\gamma}\right) x_i}{\psi_i} \right)^\gamma \quad (26)$$

The goodness of fit can be checked by testing that the mean of  $\tilde{\mathcal{E}}_i$  is equal to 1 and graphically checking what is known as a weibull plot. If it is a good fit, the empirical curve will be near the straight line. In the example shown below, the weibull does better than the exponential but it is still not a great fit.



**Figure 1.** Weibull plot for WACD(1,1) model fit to SPY INET on 2012-11-30

### 1.3. The Hawkes Process.

#### 1.3.1. Linear Self-Exciting Processes.

A (univariate) linear self-exciting (counting) process  $N_t$  is one that can be expressed as [20][8][19][3][9, 11.3]

$$\begin{aligned}\lambda(t) &= \lambda_0(t)\kappa + \int_{-\infty}^t \nu(t-s)dN_s \\ &= \lambda_0(t)\kappa + \sum_{t_k < t} \nu(t-t_k)\end{aligned}\tag{27}$$

where  $\lambda_0(t)$  is a deterministic base intensity, see (132),  $\nu: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  expresses the positive influence of past events  $t_i$  on the current value of the intensity process, and  $\kappa$  takes the place of the  $\lambda_0$  constant in the referenced papers. The (exponential) Hawkes process of order  $P$  is a linear self-exciting process defined by the exponential kernel

$$\nu(t) = \sum_{j=1}^P \alpha_j e^{-\beta_j t}\tag{28}$$

which has the survivor function

$$\begin{aligned}S(t) &= \int_t^\infty \nu(s)ds \\ &= \int_t^\infty \sum_{j=1}^P \alpha_j e^{-\beta_j s} ds \\ &= \frac{\sum_{j=1}^P \alpha_j e^{-\beta_j t} \prod_{k=1, k \neq j}^P \beta_k}{\prod_{j=1}^P \beta_j}\end{aligned}\tag{29}$$

where the product  $\prod_{k=1, k \neq j}^P$  denotes that  $k = j$  is excluded so that the hazard function is written

$$\begin{aligned}\bar{\nu}(t) &= \frac{\nu(t)}{S(t)} \\ &= \frac{\sum_{j=1}^P \alpha_j e^{-\beta_j t}}{\sum_{j=1}^P \alpha_j e^{-\beta_j t} \prod_{k=1, k \neq j}^P \beta_k} \\ &= \frac{\prod_{j=1}^P \beta_j}{\sum_{j=1}^P \alpha_j e^{-\beta_j t} \prod_{k=1, k \neq j}^P \beta_k}\end{aligned}\quad (30)$$

so that the “instantaneous” hazard rate is

$$\lim_{t \rightarrow 0} \bar{\nu}(t) = \frac{\sum_{j=1}^P \alpha_j \prod_{j=1}^P \beta_j}{\sum_{j=1}^P \alpha_j \prod_{k=1, k \neq j}^P \beta_k} \quad (31)$$

The intensity of the exponential Hawkes process is written as

$$\begin{aligned}\lambda(t) &= \lambda_0(t)\kappa + \int_0^t \sum_{j=1}^P \alpha_j e^{-\beta_j(t-s)} dN_s \\ &= \lambda_0(t)\kappa + \sum_{k=0}^{N_t-1} \sum_{j=1}^P \alpha_j e^{-\beta_j(t-t_k)} \\ &= \lambda_0(t)\kappa + \sum_{j=1}^P \sum_{k=0}^{N_t-1} \alpha_j e^{-\beta_j(t-t_k)} \\ &= \lambda_0(t)\kappa + \sum_{j=1}^P \alpha_j \sum_{k=0}^{N_t-1} e^{-\beta_j(t-t_k)} \\ &= \lambda_0(t)\kappa + \sum_{j=1}^P \alpha_j B_j(N_t) \quad \forall t \in t_i\end{aligned}\quad (32)$$

where  $B_j(i)$  is given recursively by

$$\begin{aligned}B_j(i) &= \sum_{k=1}^{i-1} e^{-\beta_j(t_i-t_k)} \\ &= e^{-\beta_j(t_i-t_{i-1})} \sum_{k=1}^{i-1} e^{-\beta_j(t_{i-1}-t_k)} \\ &= e^{-\beta_j(t_i-t_{i-1})} \left( 1 + \sum_{k=1}^{i-2} e^{-\beta_j(t_{i-1}-t_k)} \right) \\ &= e^{-\beta_j(t_i-t_{i-1})} (1 + B_j(i-1))\end{aligned}\quad (33)$$

since  $e^{-\beta_j(t_{i-1}-t_{i-1})} = e^{-\beta_j 0} = e^0 = 1$ . A univariate Hawkes process is stationary if the branching ratio is less than one.

$$\sum_{j=1}^P \frac{\alpha_j}{\beta_j} < 1 \quad (34)$$

If a Hawkes process is stationary then the unconditional mean is

$$\begin{aligned}\mu = E[\lambda(t)] &= \frac{\lambda_0}{1 - \int_0^\infty \nu(t) dt} \\ &= \frac{\lambda_0}{1 - \int_0^\infty \sum_{j=1}^P \alpha_j e^{-\beta_j t} dt} \\ &= \frac{\lambda_0}{1 - \sum_{j=1}^P \frac{\alpha_j}{\beta_j}}\end{aligned}\quad (35)$$

For consecutive events, we have the compensator (7)

$$\begin{aligned}
\Lambda(t_{i-1}, t_i) &= \int_{t_{i-1}}^{t_i} \lambda(t) dt \\
&= \int_{t_{i-1}}^{t_i} \left( \lambda_0(t) + \sum_{j=1}^P \alpha_j B_j(N_t) \right) dt \\
&= \int_{t_{i-1}}^{t_i} \lambda_0(s) ds + \int_{t_{i-1}}^{t_i} \sum_{j=1}^P \alpha_j \sum_{k=0}^{i-1} e^{-\beta_j(t-t_k)} dt \\
&= \int_{t_{i-1}}^{t_i} \lambda_0(s) ds + \sum_{j=1}^P \alpha_j \sum_{k=0}^{i-1} \int_{t_{i-1}}^{t_i} e^{-\beta_j(t-t_k)} dt \\
&= \int_{t_{i-1}}^{t_i} \lambda_0(s) ds + \sum_{k=0}^{i-1} \int_{t_{i-1}}^{t_i} \nu(t-t_k) dt \\
&= \int_{t_{i-1}}^{t_i} \lambda_0(s) ds + \sum_{k=0}^{i-1} \sum_{j=1}^P \frac{\alpha_j}{\beta_j} (e^{-\beta_j(t_{i-1}-t_k)} - e^{-\beta_j(t_i-t_k)}) \\
&= \int_{t_{i-1}}^{t_i} \lambda_0(s) ds + \sum_{j=1}^P \frac{\alpha_j}{\beta_j} (1 - e^{-\beta_j(t_i-t_{i-1})}) A_j(i-1)
\end{aligned} \tag{36}$$

where there is the recursion

$$\begin{aligned}
A_j(i) &= \sum_{\substack{t_k \leq t_i \\ k=0 \\ i-1}} e^{-\beta_j(t_i-t_k)} \\
&= \sum_{k=0}^{i-1} e^{-\beta_j(t_i-t_k)} \\
&= 1 + e^{-\beta_j(t_i-t_{i-1})} A_j(i-1)
\end{aligned} \tag{37}$$

with  $A_j(0) = 0$  since the integral of the exponential kernel (28) is

$$\begin{aligned}
\int_{t_{i-1}}^{t_i} \nu(t) dt &= \int_{t_{i-1}}^{t_i} \sum_{j=1}^P \alpha_j e^{-\beta_j(t-t_k)} dt \\
&= \sum_{j=1}^P \frac{\alpha_j}{\beta_j} (e^{-\beta_j t_i} - e^{-\beta_j t_{i-1}})
\end{aligned} \tag{38}$$

If  $\lambda_0(t) = \lambda_0$  then (36) simplifies to

$$\begin{aligned}
\Lambda(t_{i-1}, t_i) &= (t_i - t_{i-1}) \lambda_0 + \sum_{k=0}^{i-1} \sum_{j=1}^P \frac{\alpha_j}{\beta_j} (e^{-\beta_j(t_{i-1}-t_k)} - e^{-\beta_j(t_i-t_k)}) \\
&= (t_i - t_{i-1}) \lambda_0 + \sum_{j=1}^P \frac{\alpha_j}{\beta_j} (1 - e^{-\beta_j(t_i-t_{i-1})}) A_j(i-1)
\end{aligned} \tag{39}$$

Similiarly, another parametrization is given by

$$\begin{aligned}
\Lambda(t_{i-1}, t_i) &= \int_{t_{i-1}}^{t_i} \kappa \lambda_0(s) ds + \sum_{j=1}^P \frac{\alpha_j}{\beta_j} (1 - e^{-\beta_j(t_i-t_{i-1})}) A_j(i-1) \\
&= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(s) ds + \sum_{j=1}^P \frac{\alpha_j}{\beta_j} (1 - e^{-\beta_j(t_i-t_{i-1})}) A_j(i-1) \\
&= \kappa \Lambda_0(t_{i-1}, t_i) + \sum_{j=1}^P \frac{\alpha_j}{\beta_j} (1 - e^{-\beta_j(t_i-t_{i-1})}) A_j(i-1)
\end{aligned} \tag{40}$$

where  $\kappa$  scales the predetermined baseline intensity  $\lambda_0(s)$ . In this parameterization the intensity is also scaled by  $\kappa$

$$\lambda(t) = \kappa \lambda_0(t) + \sum_{j=1}^P \alpha_j B_j(N_t) \quad (41)$$

this allows to precompute the deterministic part of the compensator  $\Lambda_0(t_{i-1}, t_i) = \int_{t_{i-1}}^{t_i} \lambda_0(s) ds$ .

### 1.3.2. The Hawkes(1) Model.

The simplest case occurs when the baseline intensity  $\lambda_0(t) = 1$  is constant and  $P = 1$  where we have

$$\lambda(t) = \kappa + \sum_{t_i < t} \alpha e^{-\beta(t-t_i)} \quad (42)$$

which has the unconditional mean

$$E[\lambda(t)] = \frac{\kappa}{1 - \frac{\alpha}{\beta}} \quad (43)$$

### 1.3.3. Maximum Likelihood Estimation.

The log-likelihood of a simple point process is written as

$$\begin{aligned} \ln \mathcal{L}(N(t)_{t \in [0, T]}) &= \int_0^T (1 - \lambda(s)) ds + \int_0^T \ln \lambda(s) dN_s \\ &= T - \int_0^T \lambda(s) ds + \int_0^T \ln \lambda(s) dN_s \end{aligned} \quad (44)$$

which in the case of the Hawkes model of order  $P$  can be explicitly written [15] as

$$\begin{aligned} \ln \mathcal{L}(\{t_i\}_{i=1 \dots n}) &= T - \Lambda(0, T) + \sum_{i=1}^n \ln \lambda(t_i) \\ &= T + \sum_{i=1}^n \ln \lambda(t_i) - \Lambda(t_{i-1}, t_i) \\ &= T - \Lambda(0, T) + \sum_{i=1}^n \ln \lambda(t_i) \\ &= T - \Lambda(0, T) + \sum_{i=1}^n \ln \left( \kappa \lambda_0(t_i) + \sum_{j=1}^P \sum_{k=1}^{i-1} \alpha_j e^{-\beta_j(t_i - t_k)} \right) \\ &= T - \Lambda(0, T) + \sum_{i=1}^n \ln \left( \kappa \lambda_0(t_i) + \sum_{j=1}^P \alpha_j R_j(i) \right) \\ &= T - \int_0^T \kappa \lambda_0(s) ds - \sum_{i=1}^n \sum_{j=1}^P \frac{\alpha_j}{\beta_j} (1 - e^{-\beta_j(t_n - t_i)}) \\ &\quad + \sum_{i=1}^n \ln \left( \kappa \lambda_0(t_i) + \sum_{j=1}^P \alpha_j R_j(i) \right) \end{aligned} \quad (45)$$

where  $T = t_n$  and we have the recursion[14]

$$\begin{aligned} R_j(i) &= \sum_{k=1}^{i-1} e^{-\beta_j(t_i - t_k)} \\ &= e^{-\beta_j(t_i - t_{i-1})} (1 + R_j(i-1)) \end{aligned} \quad (46)$$

If we have constant baseline intensity  $\lambda_0(t) = 1$  then the log-likelihood can be written

$$\begin{aligned} \ln \mathcal{L}(\{t_i\}_{i=1\dots n}) &= T - \kappa T - \sum_{i=1}^n \sum_{j=1}^P \frac{\alpha_j}{\beta_j} (1 - e^{-\beta_j(t_n - t_i)}) \\ &\quad + \sum_{i=1}^n \ln \left( \lambda_0 + \sum_{j=1}^P \alpha_j R_j(i) \right) \end{aligned} \quad (47)$$

Note that it was necessary to shift each  $t_i$  by  $t_1$  so that  $t_1 = 0$  and  $t_n = T$ . Also note that  $T$  is just an additive constant which does not vary with the parameters so for the purposes of estimation can be removed from the equation.

### 1.3.4. Prediction.

The next event arrival time of a point process can be predicted by solving for the unknown  $t_{n+1}$  in the equation

$$\varepsilon_{n+1} = \Lambda(t_n, t_{n+1}) = \int_{t_n}^{t_{n+1}} \lambda(s; \mathfrak{F}_s) ds \quad (48)$$

where  $\mathfrak{F}_s$  is the filtration up to time  $s$  and the parameters of  $\lambda$  are fixed. Now generate a “large” sample of random variables exponentially distributed with mean 1 and for each value of the sample solve the equation

$$t_{n+1,i} : \Lambda(t_n, t_{n+1}) = \varepsilon_{n+1,i} \quad (49)$$

then average the resulting collection of inversions so that

$$t_{n+1} = \frac{1}{h} \sum_{i=1}^h t_{n+1,i} \quad (50)$$

If  $P=1$  the inverse equation actually has a closed form in terms of the Lambert W function (4.1), that is, let

$$a(n) = \sum_{k=0}^n e^{\beta_1(t_k + 2t_n)} \quad (51)$$

$$b(n) = \sum_{k=0}^n e^{\beta_1(t_k + t_n)} \quad (52)$$

$$c(n) = \sum_{k=0}^n e^{\beta_1 t_k} \quad (53)$$

$$d(n) = \sum_{k=0}^n \sum_{l=0}^n e^{\beta_1(t_l + t_k + t_n)} \quad (54)$$

then

$$\begin{aligned} \Lambda^{-1}(x; t_0 \dots t_n) &= \frac{a(n)t_n e^{-\beta_1 t_n}}{b(n)} + \\ &\quad \frac{a(n)e^{-\beta_1 t_n} W \left( \frac{\alpha_1 b(n) e^{-\frac{((2\kappa t_n + x)\beta_1 e^{\beta_1 t_n} - \alpha_1 c(n))e^{-\beta_1 t_n}}{\kappa}}}{\kappa} \right)}{\frac{\beta_1 a(n)}{\kappa}} + \\ &\quad \frac{1}{\kappa} \left( \frac{a(n)e^{-\beta_1 t_n} x}{b(n)} - \frac{e^{-\beta_1 t_n} \alpha_1 d(n)}{\beta_1 b(n)} \right) \end{aligned} \quad (55)$$

so that the expectation of the occurrence of the next point is given by the integral

$$\int_0^\infty \Lambda^{-1}(x; t_0 \dots t_n) e^{-x} dx = \int_0^\infty \left( \frac{a(n)t_n e^{-\beta_1 t_n}}{b(n)} + \frac{a(n)e^{-\beta_1 t_n} W \left( \frac{\frac{1}{\kappa} \alpha_1 b(n) e^{-\frac{((2\kappa t_n + x)\beta_1 e^{\beta_1 t_n} - \alpha_1 c(n))e^{-\beta_1 t_n}}{\kappa}}}{\kappa} \right)}{\beta_1 a(n)} + \frac{1}{\kappa} \left( \frac{a(n)e^{-\beta_1 t_n} x}{b(n)} - \frac{e^{-\beta_1 t_n} \alpha_1 d(n)}{\beta_1 b(n)} \right) \right) e^{-x} dx \quad (56)$$

which, as far as I am aware, must be calculated with numeric methods which are actually far more efficient than Monte Carlo sampling as other studies have suggested. It would be nice to have expressions like this involving the Lambert W function for  $P > 1$  but neither Maple nor Mathematica were able to find any solutions in terms of “known” functions for  $P > 1$ .

### 1.3.5. The Hawkes Process in Quantum Theory.

The Hawkes process arises in quantum theory by considering feedback via continuous measurements where the quantum analog of a self-exciting point process is a source of irreversibility whose strength is controlled by the rate of detections from that source. [21].

## 1.4. Alternative Kernels for the Hawkes Process.

### 1.4.1. A Generalized Exponential Kernel.

In the paper of [14] there is made a reference to [16], which appears to not be available online. Nevertheless, [14] contains the necessary recursive formulas. Here, the exponential kernel changes from (28)

$$\nu(t) = \alpha e^{-\beta t} \quad (57)$$

to

$$\nu(t) = \sum_{j=0}^P \alpha_j t^j e^{-\beta t} \quad (58)$$

with a recursive structure that involves binomial coefficients. The intensity of the generalized exponential Hawkes process is written as

$$\begin{aligned} \lambda(t) &= \lambda_0(t)\kappa + \int_0^t \sum_{j=0}^P \alpha_j t^j e^{-\beta(t-s)} dN_s \\ &= \lambda_0(t)\kappa + \sum_{k=0}^{N_t-1} \sum_{j=0}^P \alpha_j t^j e^{-\beta(t-t_k)} \\ &= \lambda_0(t)\kappa + \sum_{j=0}^P \sum_{k=0}^{N_t-1} \alpha_j t^j e^{-\beta(t-t_k)} \\ &= \lambda_0(t)\kappa + \sum_{j=0}^P \alpha_j t^j \sum_{k=0}^{N_t-1} e^{-\beta(t-t_k)} \\ \lambda(t_i) &= \lambda_0(t)\kappa + \sum_{j=0}^P \alpha_j R_j(i) \end{aligned} \quad (59)$$

Note that  $\beta$  does not have a subscript in this parametrization. For consecutive events, we have the compensator (7)

$$\begin{aligned} \Lambda(t_{i-1}, t_i) &= \int_{t_{i-1}}^{t_i} \lambda(t) dt \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(s) ds + \sum_{k=0}^{i-1} \int_{t_{i-1}}^{t_i} \nu(t-t_k) dt \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(t)\kappa + \sum_{j=0}^P \alpha_j t^j \sum_{k=0}^{N_t-1} e^{-\beta(t-t_k)} dt \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(s) ds + \int_{t_{i-1}}^{t_i} \sum_{j=0}^P \alpha_j t^j \sum_{k=0}^{N_t-1} e^{-\beta(t-t_k)} dt \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(s) ds + \sum_{k=0}^{i-1} \hat{\nu}(t_i - t_k) - \hat{\nu}(t_{i-1} - t_k) \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(s) ds + \sum_{k=0}^{i-1} \end{aligned} \quad (60)$$

where  $\hat{\nu}(t)$  is the antiderivative of the kernel

$$\begin{aligned}
\hat{\nu}(t) &= \int \nu(t) dt \\
&= \int \sum_{j=0}^P \alpha_j t^j e^{-\beta t} dt \\
&= \frac{e^{-\beta t}}{\beta^{P+1}} \sum_{k=0}^P \alpha_k \sum_{j=0}^k -\beta^{P-(k-j)} (k-j)! \binom{k}{k-j} t^j \\
&= \frac{e^{-\beta t}}{\beta^{P+1}} \sum_{k=1}^P \frac{\alpha_k \beta^{-k} (\beta t)^k e^{\beta t} (2\beta^{P+1} - \beta^{P+2}) \Gamma(k+1, \beta t)}{(\beta t)^{k+1} (\beta - 2)} \\
&= -\frac{1}{\beta} \sum_{k=0}^P \alpha_k \beta^{-k} \Gamma(k+1, \beta t)
\end{aligned} \tag{61}$$

where  $\Gamma(k, \beta t)$  is the incomplete Gamma function (91). The log-likelihood for the Hawkes process having this generalized exponential kernel is given by

$$\begin{aligned}
\ln \mathcal{L}(\{t_i\}_{i=1 \dots n}) &= \sum_{i=1}^n \ln \left( \lambda_0(t_i) \kappa + \sum_{j=0}^P \alpha_j R_j(i) \right) - \kappa \int_0^T \lambda_0(t) dt - \sum_{i=1}^n \sum_{j=0}^P \alpha_j S_j(T - t_i) \\
&= \sum_{i=1}^n \ln \left( \lambda_0(t_i) \kappa + \sum_{j=0}^P \alpha_j R_j(i) \right) - \sum_{i=1}^n \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \sum_{j=0}^P \alpha_j S_j(T - t_i) \\
&= \sum_{i=1}^n \ln \left( \lambda_0(t_i) \kappa + \sum_{j=0}^P \alpha_j R_j(i) \right) - \left( \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \sum_{j=0}^P \alpha_j S_j(T - t_i) \right)
\end{aligned} \tag{62}$$

where  $R_j(i)$  and  $S_j(t)$  are given recursively and  $T = t_n$  is the time of the last point in the sample. Let

$$t_0 = 0 \tag{63}$$

$$R_0(1) = 0 \tag{64}$$

$$S_0(t) = \frac{1 - e^{-\beta t}}{\beta} \tag{65}$$

then

$$A_k(t) = t^k e^{-\beta t} \tag{66}$$

$$\begin{aligned}
R_j(i) &= A_j(t_i - t_{i-1}) + \sum_{k=0}^j \binom{j}{k} A_{j-k}(t_i - t_{i-1}) R_k(i-1) \\
&= (t_i - t_{i-1})^j e^{-\beta(t_i - t_{i-1})} + \sum_{k=0}^j \binom{j}{k} (t_i - t_{i-1})^{j-k} e^{-\beta(t_i - t_{i-1})} R_k(i-1)
\end{aligned} \tag{67}$$

and

$$\begin{aligned}
S_j(t) &= \frac{j S_{j-1}(t) - A_j(t)}{\beta} \\
&= \frac{j S_{j-1}(t) - t^j e^{-\beta t}}{\beta}
\end{aligned} \tag{68}$$

#### 1.4.2. The Hawkes Process Having a Weibull Kernel.

The exponential kernel of the Hawkes process can be replaced with that of a Weibull kernel. [12, 6.3] Recall that the intensity of a Hawkes process is defined by (27)

$$\begin{aligned}
\lambda(t) &= \lambda_0(t) \kappa + \int_{-\infty}^t \psi(t-s) dN_s \\
&= \lambda_0(t) \kappa + \sum_{t_i < t} \psi(t - t_i)
\end{aligned} \tag{69}$$

where the exponential kernel  $\psi(t) = \sum_{j=1}^P \alpha_j e^{-\beta_j t}$  is replaced by the Weibull kernel

$$\psi(t) = \sum_{j=1}^P \alpha_j \left( \frac{v_j}{\omega_j} \right) \left( \frac{t}{\omega_j} \right)^{v_j-1} e^{-\left( \frac{t}{\omega_j} \right)^{v_j}} \quad (70)$$

with parameter vectors  $\alpha_j, v_j, \omega_j$  so the Hawkes-Weibull intensity is written as

$$\begin{aligned} \lambda(t) &= \lambda_0(t)\kappa + \int_0^t \sum_{j=1}^P \alpha_j \left( \frac{v_j}{\omega_j} \right) \left( \frac{t-s}{\omega_j} \right)^{v_j-1} e^{-\left( \frac{t-s}{\omega_j} \right)^{v_j}} dN_s \\ &= \lambda_0(t)\kappa + \sum_{k=0}^{N_t-1} \sum_{j=1}^P \alpha_j \left( \frac{v_j}{\omega_j} \right) \left( \frac{t-t_k}{\omega_j} \right)^{v_j-1} e^{-\left( \frac{t-t_k}{\omega_j} \right)^{v_j}} \\ &= \lambda_0(t)\kappa + \sum_{j=1}^P \sum_{k=0}^{N_t-1} \alpha_j \left( \frac{v_j}{\omega_j} \right) \left( \frac{t-t_k}{\omega_j} \right)^{v_j-1} e^{-\left( \frac{t-t_k}{\omega_j} \right)^{v_j}} \\ &= \lambda_0(t)\kappa + \sum_{j=1}^P \alpha_j \left( \frac{v_j}{\omega_j} \right) \sum_{k=0}^{N_t-1} \left( \frac{t-t_k}{\omega_j} \right)^{v_j-1} e^{-\left( \frac{t-t_k}{\omega_j} \right)^{v_j}} \\ &= \lambda_0(t)\kappa + \sum_{j=1}^P \alpha_j \left( \frac{v_j}{\omega_j} \right) C_j(N_t) \end{aligned} \quad (71)$$

and  $C_j(n)$  is given by

$$C_j(n) = \sum_{k=0}^{n-1} \left( \frac{t-t_k}{\omega_j} \right)^{v_j-1} e^{-\left( \frac{t-t_k}{\omega_j} \right)^{v_j}} \quad (72)$$

where the branching ratio is

$$\begin{aligned} \int_0^\infty \psi(t) dt &= \int_0^\infty \sum_{j=1}^P \alpha_j \left( \frac{v_j}{\omega_j} \right) \left( \frac{t}{\omega_j} \right)^{v_j-1} e^{-\left( \frac{t}{\omega_j} \right)^{v_j}} dt \\ &= \sum_{j=1}^P \alpha_j \end{aligned} \quad (73)$$

The survivor function of the Weibull kernel is given by

$$\begin{aligned} S(t) &= \int_t^\infty \nu(s) ds \\ &= \int_t^\infty \sum_{j=1}^P \alpha_j \left( \frac{v_j}{\omega_j} \right) \left( \frac{s}{\omega_j} \right)^{v_j-1} e^{-\left( \frac{s}{\omega_j} \right)^{v_j}} ds \\ &= \sum_{j=1}^P \alpha_j e^{-t^{v_j} \omega_j^{-v_j}} \end{aligned} \quad (74)$$

which is surprisingly quite a bit less complicated than survivor function of the exponential kernel (29). The hazard function is the quotient of the kernel over the survivor function

$$\begin{aligned} \bar{\psi}(t) &= \frac{\psi(t)}{S(t)} \\ &= \frac{\sum_{j=1}^P \alpha_j \left( \frac{v_j}{\omega_j} \right) \left( \frac{t}{\omega_j} \right)^{v_j-1} e^{-\left( \frac{t}{\omega_j} \right)^{v_j}}}{\sum_{j=1}^P \alpha_j e^{-t^{v_j} \omega_j^{-v_j}}} \end{aligned} \quad (75)$$

Now, similar to (60), the compensator is calculated

$$\begin{aligned}
\tilde{\mathcal{E}}_i &= \Lambda^k(t_{i-1}, t_i) \\
&= \int_{t_{i-1}}^{t_i} \lambda(t) dt \\
&= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \int_{t_{i-1}}^{t_i} \sum_{k=0}^{i-1} \psi(t - t_k) dt \\
&= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \sum_{k=0}^{i-1} \psi(t_i - t_k) - \psi(t_{i-1} - t_k) \\
&= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \int_{t_{i-1}}^{t_i} \sum_{k=0}^{i-1} \sum_{j=1}^P \alpha_j \left( \frac{v_j}{\omega_j} \right) \left( \frac{t - t_k}{\omega_j} \right)^{v_j-1} e^{-\left( \frac{t - t_k}{\omega_j} \right)^{v_j}} dt \\
&= \int_{t_{i-1}}^{t_i} \lambda_0(t) \kappa + \int_{t_{i-1}}^{t_i} \sum_{j=1}^P \alpha_j \left( \frac{v_j}{\omega_j} \right) \sum_{k=0}^{i-1} \left( \frac{t - t_k}{\omega_j} \right)^{v_j-1} e^{-\left( \frac{t - t_k}{\omega_j} \right)^{v_j}} dt \\
&= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \sum_{k=0}^{i-1} \sum_{j=1}^P -\alpha_j \left( e^{-\left( \frac{t_i - t_k}{\omega_j} \right)^{v_j}} - e^{-\left( \frac{t_{i-1} - t_k}{\omega_j} \right)^{v_j}} \right)
\end{aligned} \tag{76}$$

where the integral of the Weibull kernel over consecutive events is given by an application of the first fundamental theorem of calculus

$$\begin{aligned}
\int_{t_{i-1}}^{t_i} \psi(t) dt &= \int_{t_{i-1}}^{t_i} \sum_{j=1}^P \alpha_j \left( \frac{v_j}{\omega_j} \right) \left( \frac{t}{\omega_j} \right)^{v_j-1} e^{-\left( \frac{t}{\omega_j} \right)^{v_j}} dt \\
&= \hat{\psi}(t_i) - \hat{\psi}(t_{i-1}) \\
&= \sum_{j=1}^P -\alpha_j \left( e^{-\left( \frac{t_i}{\omega_j} \right)^{v_j}} - e^{-\left( \frac{t_{i-1}}{\omega_j} \right)^{v_j}} \right)
\end{aligned} \tag{77}$$

where  $\hat{\psi}(t)$  is the antiderivative of the kernel  $\psi(t)$  given by

$$\begin{aligned}
\hat{\psi}(t) &= \int \psi(t) dt \\
&= \int \sum_{j=1}^P \alpha_j \left( \frac{v_j}{\omega_j} \right) \left( \frac{t}{\omega_j} \right)^{v_j-1} e^{-\left( \frac{t}{\omega_j} \right)^{v_j}} dt \\
&= \sum_{j=1}^P -\alpha_j e^{-\left( \frac{t}{\omega_j} \right)^{v_j}}
\end{aligned} \tag{78}$$

The change-of-variables in (76) can be made, let

$$s = -\left( \frac{t - t_k}{\omega_j} \right)^{v_j} \tag{79}$$

then

$$ds_k = -ds \left( \frac{v_j}{\omega_j} \right) \left( \frac{t - t_k}{\omega_j} \right)^{v_j-1} \tag{80}$$

so (76) can be written as

$$\begin{aligned}
\tilde{\mathcal{E}}_i &= \Lambda^k(t_{i-1}, t_i) \\
&= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \sum_{j=1}^P \sum_{k=0}^{i-1} \alpha_j \int_{\left( \frac{t_{i-1} - t_k}{\omega_j} \right)^{v_j}}^{\left( \frac{t_i - t_k}{\omega_j} \right)^{v_j}} e^{-s} ds
\end{aligned} \tag{81}$$

When  $v_j = 1$ ,  $\omega_j = 1$  the Hawkes-Weibull process reduces to the standard exponential Hawkes process.

### 1.4.3. The Hawkes Process Having a Gamma Kernel.

Another parametrization we can try is having a kernel given by a mixture of gamma distributions

$$\begin{aligned}\lambda(t) &= \lambda_0(t)\kappa + \int_{-\infty}^t \nu(t-s) dN_s \\ &= \lambda_0(t)\kappa + \sum_{t_k < t} \nu(t-t_k)\end{aligned}\tag{82}$$

with

$$\nu(t) = \sum_{j=1}^P \frac{\alpha_j}{b_j \Gamma(c_j)} \left( \frac{t}{b_j} \right)^{c_j-1} e^{-\frac{t}{b_j}}\tag{83}$$

having parameter vectors  $\alpha_j, b_j, c_j$  so that the intensity is written

$$\begin{aligned}\lambda(t) &= \lambda_0(t)\kappa + \int_0^t \sum_{j=1}^P \frac{\alpha_j}{b_j \Gamma(c_j)} \left( \frac{t-s}{b_j} \right)^{c_j-1} e^{-\frac{t-s}{b_j}} dN_s \\ &= \lambda_0(t)\kappa + \sum_{i=0}^{N_t-1} \sum_{j=1}^P \frac{\alpha_j}{b_j \Gamma(c_j)} \left( \frac{t-t_i}{b_j} \right)^{c_j-1} e^{-\frac{t-t_i}{b_j}} \\ &= \lambda_0(t)\kappa + \sum_{j=1}^P \sum_{i=0}^{N_t-1} \frac{\alpha_j}{b_j \Gamma(c_j)} \left( \frac{t-t_i}{b_j} \right)^{c_j-1} e^{-\frac{t-t_i}{b_j}} \\ &= \lambda_0(t)\kappa + \sum_{j=1}^P \frac{\alpha_j}{b_j \Gamma(c_j)} \sum_{i=0}^{N_t-1} \left( \frac{t-t_i}{b_j} \right)^{c_j-1} e^{-\frac{t-t_i}{b_j}}\end{aligned}\tag{84}$$

where the branching ratio is

$$\begin{aligned}\int_0^\infty \nu(t) dt &= \int_0^\infty \sum_{j=1}^P \frac{\alpha_j}{b_j \Gamma(c_j)} \left( \frac{t}{b_j} \right)^{c_j-1} e^{-\frac{t}{b_j}} dt \\ &= \sum_{j=1}^P \alpha_j\end{aligned}\tag{85}$$

The compensator is given by

$$\begin{aligned}\tilde{\mathcal{E}}_i &= \Lambda^k(t_{i-1}, t_i) \\ &= \int_{t_{i-1}}^{t_i} \lambda(t) dt \\ &= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \sum_{k=0}^{i-1} \int_{t_{i-1}}^{t_i} \nu(t-t_k) dt \\ &= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \int_{t_{i-1}}^{t_i} \sum_{k=0}^{i-1} \nu(t-t_k) dt \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(t)\kappa + \sum_{j=1}^P \frac{\alpha_j}{b_j \Gamma(c_j)} \sum_{k=0}^{i-1} \left( \frac{t-t_k}{b_j} \right)^{c_j-1} e^{-\frac{t-t_k}{b_j}} dt \\ &= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \int_{t_{i-1}}^{t_i} \sum_{j=1}^P \frac{\alpha_j}{b_j \Gamma(c_j)} \sum_{k=0}^{i-1} \left( \frac{t-t_k}{b_j} \right)^{c_j-1} e^{-\frac{t-t_k}{b_j}} dt \\ &= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \sum_{k=0}^{i-1} \int_{t_{i-1}}^{t_i} \sum_{j=1}^P \frac{\alpha_j}{b_j \Gamma(c_j)} \left( \frac{t-t_k}{b_j} \right)^{c_j-1} e^{-\frac{t-t_k}{b_j}} dt \\ &= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \sum_{k=0}^{i-1} \hat{\nu}(t_i-t_k) - \hat{\nu}(t_{i-1}-t_k) \\ &= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \sum_{k=0}^{i-1} \sum_{j=1}^P -\alpha_j \frac{(c_j+1)}{\Gamma(c_j+2)} \dots \\ &\dots \left( \Gamma\left(c_j+1, \frac{(t_i-t_k)}{b_j}\right) - \Gamma\left(c_j+1, \frac{(t_{i-1}-t_k)}{b_j}\right) - e^{-\frac{(t_i-t_k)}{b_j}} (t_i-t_k)^{c_j} b_j^{-c_j} + e^{-\frac{(t_{i-1}-t_k)}{b_j}} (t_{i-1}-t_k)^{c_j} b_j^{-c_j} \right)\end{aligned}\tag{86}$$

since the integral of the gamma kernel over consecutive events is given by

$$\begin{aligned}
\int_{t_{i-1}}^{t_i} \nu(t) dt &= \int_{t_{i-1}}^{t_i} \sum_{j=1}^P \frac{\alpha_j}{b_j \Gamma(c_j)} \sum_{i=0}^{N_t-1} \left( \frac{t-t_i}{b_j} \right)^{c_j-1} e^{-\frac{t-t_i}{b_j}} dt \\
&= \hat{\nu}(t_i) - \hat{\nu}(t_{i-1}) \\
&= \sum_{j=1}^P -\frac{\alpha_j(c_j+1)}{\Gamma(c_j+2)} \left( \Gamma\left(c_j+1, \frac{t_i}{b_j}\right) - \Gamma\left(c_j+1, \frac{t_{i-1}}{b_j}\right) - e^{-\frac{t_i}{b_j}} t_i^{c_j} b_j^{-c_j} + e^{-\frac{t_{i-1}}{b_j}} t_{i-1}^{c_j} b_j^{-c_j} \right)
\end{aligned} \tag{87}$$

where  $\hat{\nu}(t)$  is the integral of  $\nu(t)$  given by

$$\begin{aligned}
\hat{\nu}(t) &= \int \nu(t) dt \\
&= \int \sum_{j=1}^P \frac{\alpha_j}{b_j \Gamma(c_j)} \left( \frac{t}{b_j} \right)^{c_j-1} e^{-\frac{t}{b_j}} dt \\
&= \sum_{j=1}^P \frac{\alpha_j b_j^{c_j} e^{-\frac{t}{b_j}} b_j^{-c_j}}{\Gamma(c_j+2)} \left( \left( \frac{t}{b_j} \right)^{\frac{c_j}{2}} M_{\frac{c_j}{2}, \frac{c_j}{2} + \frac{1}{2}} \left( \frac{t}{b_j} \right) + \left( \frac{t}{b_j} \right)^{c_j} e^{-\frac{t}{b_j}} (c_j+1) \right) \\
&= \sum_{j=1}^P \frac{\alpha_j}{\Gamma(c_j+2)} \left( \Gamma(c_j+2) + \left( \left( \frac{t}{b_j} \right)^{c_j} e^{-\frac{t}{b_j}} - \Gamma\left(c_j+1, \frac{t}{b_j}\right) \right) (c_j+1) \right)
\end{aligned} \tag{88}$$

where

$$M_{\mu, \nu}(z) = e^{-\frac{z}{2}} z^{\nu + \frac{1}{2}} {}_1F_1\left(\frac{1}{2} - \mu + \nu; z\right) \tag{89}$$

is the Whittaker M function which solves the equation  $y'' + \left(-\frac{1}{4} + \frac{\mu}{z} + \frac{\frac{1}{4} - \nu^2}{z^2}\right)y = 0$  and

$${}_1F_1\left(\frac{a}{b}; z\right) = \sum_{k=0}^{\infty} \frac{\Gamma(a+k)\Gamma(b)z^k}{\Gamma(a)\Gamma(k+1)\Gamma(b+k)} \tag{90}$$

is the confluent hypergeometric function[17] and

$$\Gamma(a, z) = \Gamma(a) - \frac{z^a {}_1F_1\left(\frac{a}{1+a}; -z\right)}{a} \tag{91}$$

is the incomplete Gamma function. When  $b_j = 1$  and  $c_j = 1$  the Hawkes-Gamma model reduces to

the standard Hawkes model with an exponential kernel.

#### 1.4.4. A Hyperbolic Kernel.

Another interesting kernel is the hyperbolic kernel having “long memory” features. As usual the intensity is defined by

$$\begin{aligned}\lambda(t) &= \lambda_0(t)\kappa + \int_0^t \nu(t-s) dN_s \\ &= \lambda_0(t)\kappa + \sum_{t_k < t} \nu(t-t_k)\end{aligned}\tag{92}$$

with the hyperbolic kernel

$$\nu(t) = \sum_{j=1}^P \frac{\alpha_j}{(t + \beta_j)^{p_j}}\tag{93}$$

having parameter vectors  $\alpha_j, \beta_j, p_j$  so that the intensity is written

$$\begin{aligned}\lambda(t) &= \lambda_0(t)\kappa + \int_0^t \sum_{j=1}^P \frac{\alpha_j}{(t + \beta_j)^{p_j}} dN_s \\ &= \lambda_0(t)\kappa + \sum_{i=0}^{N_t-1} \sum_{j=1}^P \frac{\alpha_j}{((t-t_i) + \beta_j)^{p_j}} \\ &= \lambda_0(t)\kappa + \sum_{i=0}^{N_t-1} \sum_{j=1}^P \alpha_j ((t-t_i) + \beta_j)^{-p_j} \\ &= \lambda_0(t)\kappa + \sum_{j=1}^P \alpha_j \sum_{i=0}^{N_t-1} ((t-t_i) + \beta_j)^{-p_j}\end{aligned}\tag{94}$$

The compensator is given by

$$\begin{aligned}\tilde{\mathcal{E}}_i &= \Lambda^k(t_{i-1}, t_i) \\ &= \int_{t_{i-1}}^{t_i} \lambda(t) dt \\ &= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \sum_{k=0}^{i-1} \int_{t_{i-1}}^{t_i} \nu(t-t_k) dt \\ &= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \int_{t_{i-1}}^{t_i} \sum_{k=0}^{i-1} \nu(t-t_k) dt \\ &= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \sum_{k=0}^{i-1} \hat{\nu}(t_i-t_k) - \hat{\nu}(t_{i-1}-t_k) \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(t)\kappa + \sum_{j=1}^P h_j \sum_{k=0}^{i-1} ((t-t_k) + \kappa_j)^{-p_j} dt \\ &= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \sum_{k=0}^{i-1} \int_{t_{i-1}}^{t_i} \sum_{j=1}^P \alpha_j ((t-t_k) + \beta_j)^{-p_j} dt \\ &= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \int_{t_{i-1}}^{t_i} \sum_{j=1}^P \alpha_j \sum_{k=0}^{i-1} ((t-t_k) + \beta_j)^{-p_j} dt \\ &= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \sum_{k=0}^{i-1} - \sum_{j=1}^P \frac{\alpha_j ((\beta_j + (t_i-t_k))^{1-p_j} - (\beta_j + (t_{i-1}-t_k))^{1-p_j})}{p_j - 1}\end{aligned}\tag{95}$$

since the integral of the hyperbolic kernel over consecutive events is given by

$$\begin{aligned} \int_{t_{i-1}}^{t_i} \nu(t) dt &= \int_{t_{i-1}}^{t_i} \sum_{j=1}^P \alpha_j (t + \beta_j)^{-p_j} dt \\ &= \hat{\nu}(t_i) - \hat{\nu}(t_{i-1}) \\ &= - \sum_{j=1}^P \frac{\alpha_j ((\beta_j + t_i)^{1-p_j} - (\beta_j + t_{i-1})^{1-p_j})}{p_j - 1} \end{aligned} \quad (96)$$

The antiderivative of the kernel is

$$\hat{\nu}(t) = \int \nu(t) dt = \sum_{j=1}^P \frac{\alpha_j (\beta_j + t)^{1-p_j}}{1-p_j} \quad (97)$$

and definite integral

$$\int_0^s \nu(t) dt = \sum_{j=1}^P - \frac{\alpha_j (\beta_j + s)^{-p_j} \left( \beta_j - \beta_j \left( \frac{\beta_j + s}{\beta_j} \right)^{p_j} + s \right)}{p_j - 1} \quad (98)$$

The branching ratio is given by

$$\int_0^\infty \nu(t) dt = \begin{cases} \frac{\sum_{i=1}^P \beta_i^{1-p_i} \alpha_i \left( 1 - \prod_{\substack{k=1 \\ k \neq i}}^P p_k - \sum_{\substack{k=1 \\ k \neq i}}^P p_k \right)}{\prod_{j=1}^P (p_j - 1)} & \forall p_j > 1 \\ \infty & \exists p_j \leq 1 \end{cases} \quad (99)$$

where  $\exists$  is an existential quantifier which means “at least one of” whatever follows it is true and the product with an excluded index is equivalent to

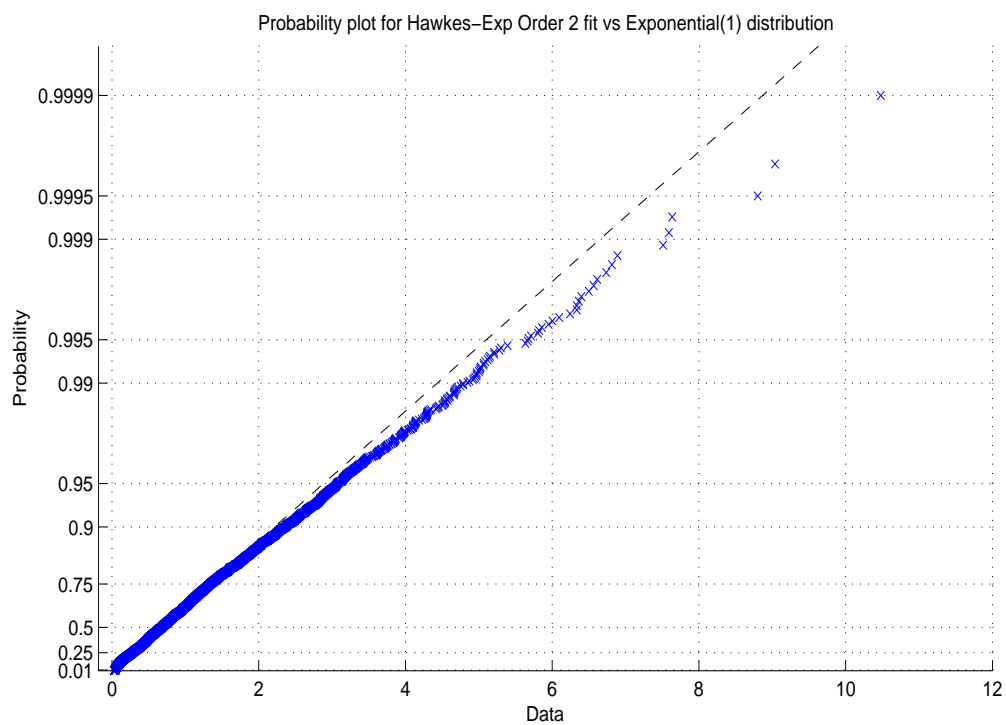
$$\prod_{\substack{k=1 \\ k \neq i}}^P p_k = \prod_{k=1}^{i-1} p_k \prod_{k=i+1}^P p_k \quad (100)$$

and likewise for the sum

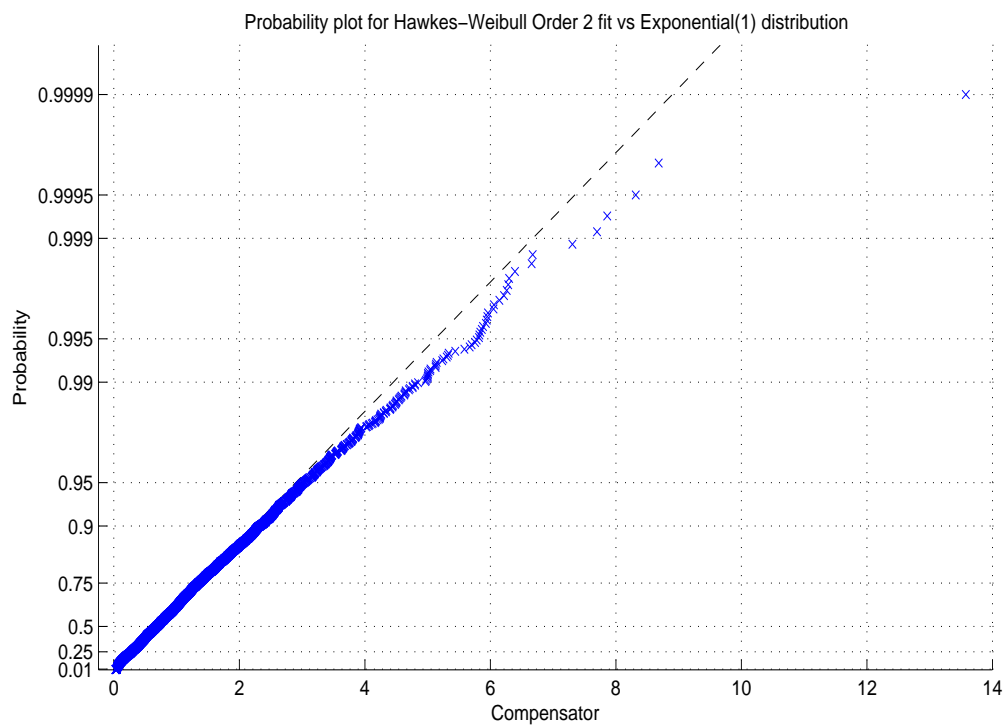
$$\sum_{\substack{k=1 \\ k \neq i}}^P p_k = \sum_{k=1}^{i-1} p_k + \sum_{k=i+1}^P p_k \quad (101)$$

### 1.5. Assessing Goodness of Fit with Graphical Methods.

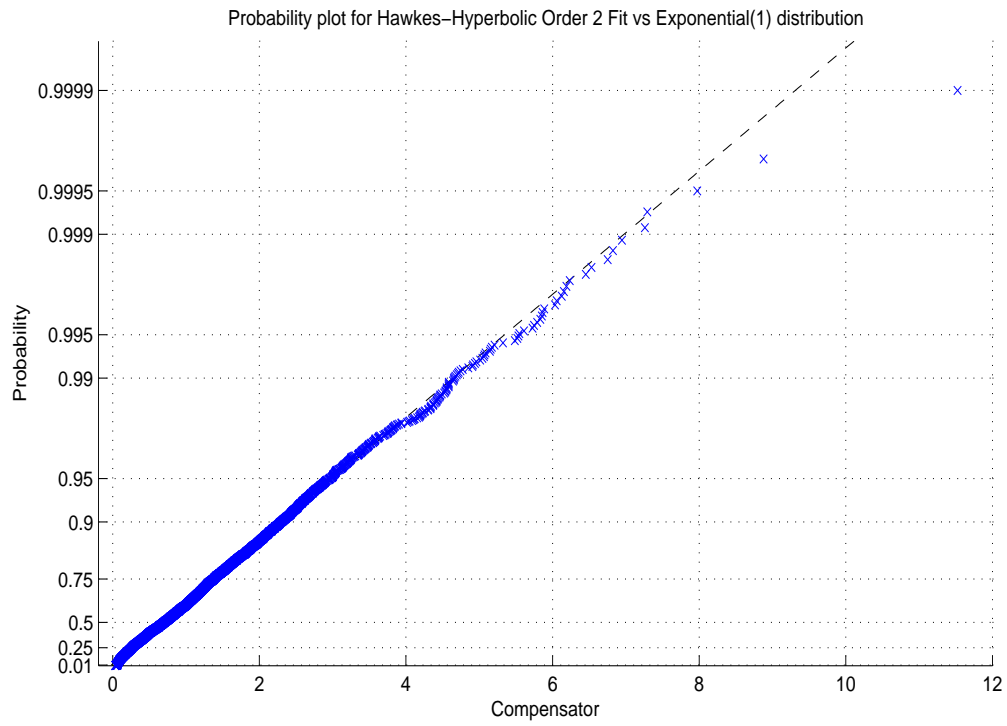
The compensator of a point process, if it is a good fit, will be an i.i.d. exponentially distributed random variable with mean 1 (and thus variance 1 and skewness 2) and no significant autocorrelation at any lag. To demonstrate this, the first 5000 points of SPY on the INET exchange on 2012-11-30 were fit with the standard exponential Hawkes kernel of order  $P = 2$  and the Weibull kernel of order  $P = 2$ . This limited number of 5000 points (accounting for a little over 59 minutes of trading time) was chosen due to the lack of recursion available for the Hawkes kernel and thus increased computational complexity.



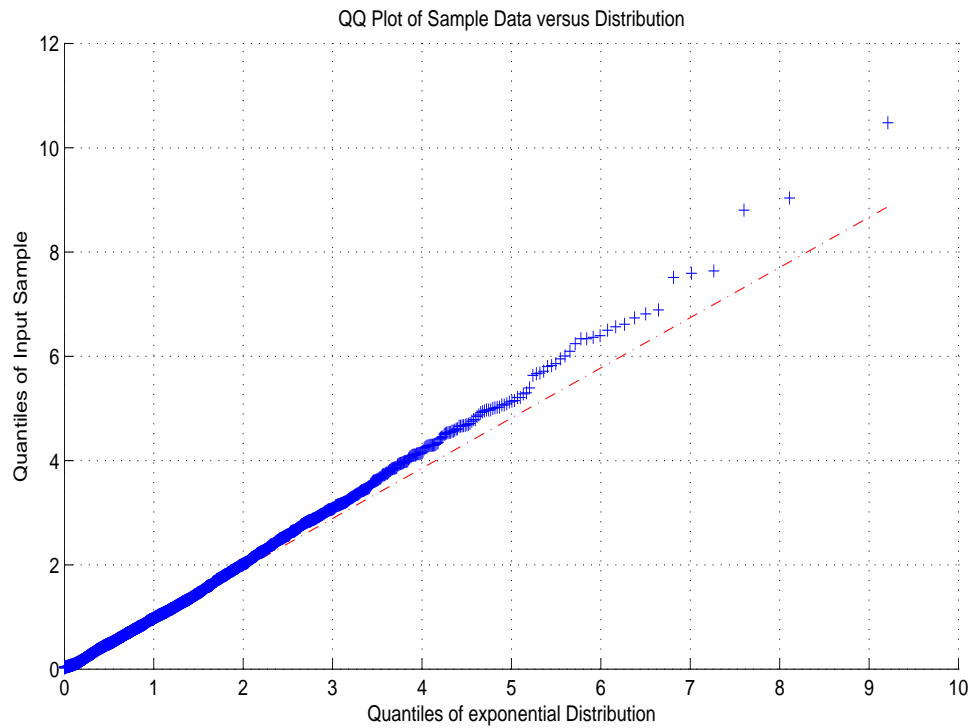
**Figure 2.** Probability plot for Hawkes-Exp Order 2 fit vs Exponential(1) distribution



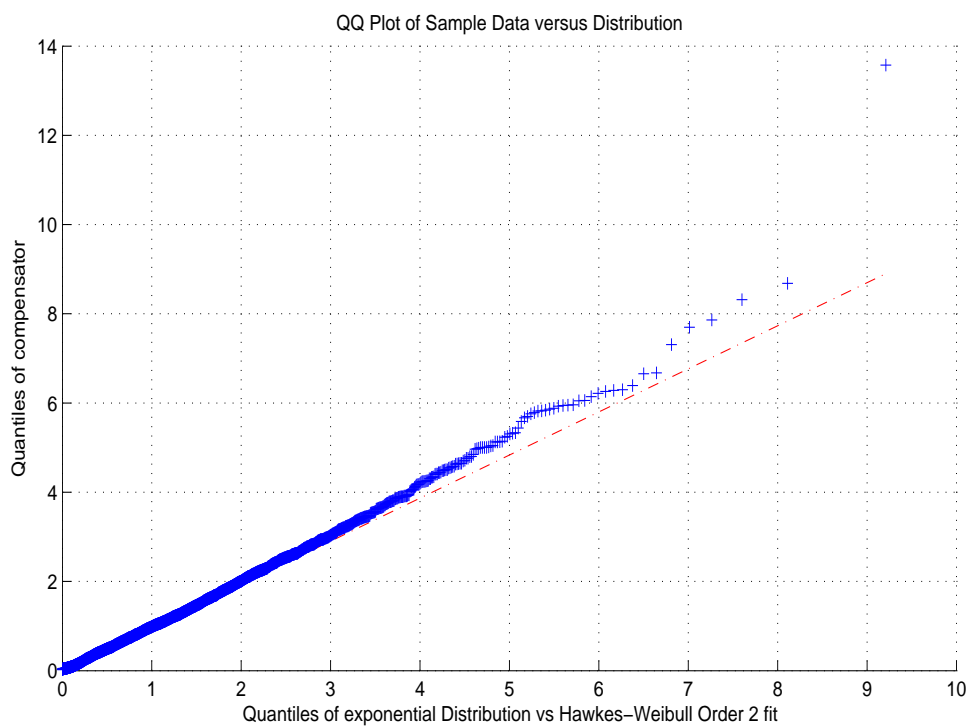
**Figure 3.** Probability plot for Hawkes-Weibull Order 2 fit vs Exponential(1) distribution



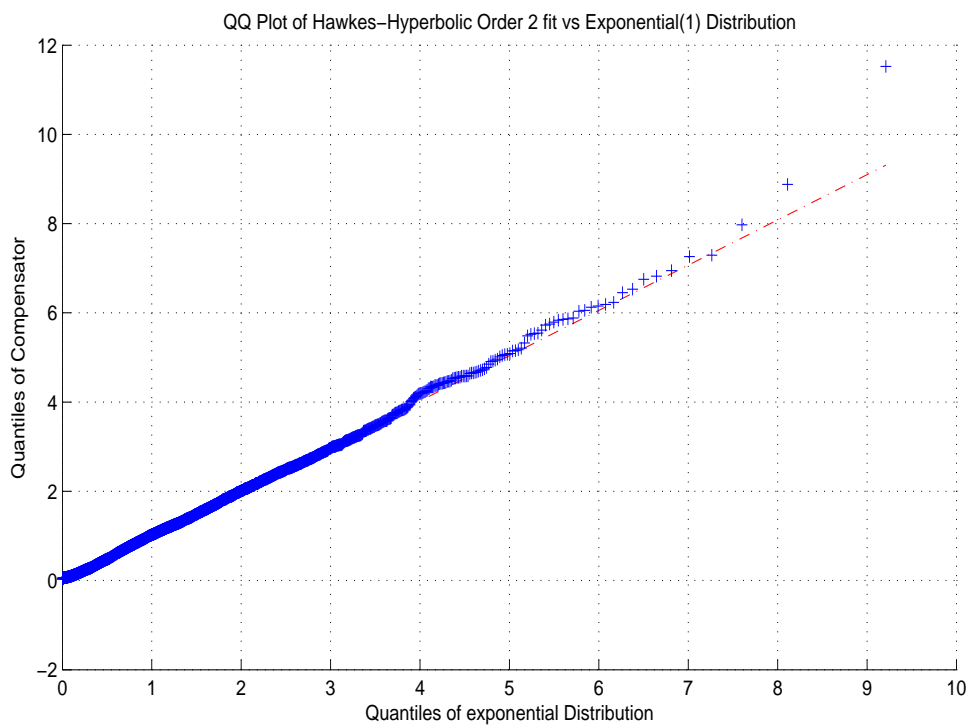
**Figure 4.** Probability plot for Hawkes-Hyperbolic Order 2 fit vs Exponential(1) distribution



**Figure 5.** Quantiles of exponential Distribution vs Hawkes-Exp Order 2 fit



**Figure 6.** Quantiles of exponential Distribution vs Hawkes-Weibull Order 2 fit



**Figure 7.** Quantiles of exponential Distribution vs Hawkes-Hyperbolic Order 2 fit

Table (1) lists the log-likelihood, mean, variance, and skew of the compensator for the exponential and Weibull fits.

|                   | Hawkes – Exp      | Hawkes – Weibull  | Hawkes – Hyperbolic |
|-------------------|-------------------|-------------------|---------------------|
| $\ln \mathcal{L}$ | 935.8106454999223 | 972.8236127532505 | 934.351208722037    |
| Mean              | 1.00000225846881  | 1.00017927572359  | 1.00024283933897    |
| Variance          | 1.08823532290517  | 1.08097250910931  | 1.0294837410725     |
| Skewness          | 2.18458483003574  | 2.33622435438673  | 2.16195765106546    |

**Table 1.**

As can be seen, the rather meager increase of the log-likelihood score gained by switching to the Weibull model and giving up recursion appears to not be worth it, and note that the skew of the Weibull fit is a little higher than the exp fit however the Weibull does fit better and goes further into the tails of the distribution before diverging. The hyperbolic kernel has a slightly lower log-likelihood score than the exponential, however the variance and skew are lower than both the other two kernels and also the QQ and probability plots show a closer fit to the linear slope before the tails diverge.

### 1.6. Combining the ACD and Hawkes Models.

The ACD and Hawkes models can be combined to provide a model for intraday volatility. [2] Let

$$\lambda(t) = \lambda_0(t) + \frac{1}{\psi_{N_t}} + \int_0^t \nu(t-s) dN_s \quad (102)$$

where  $\lambda_0(t)$  is the deterministic baseline intensity(132) and where the ACD(20) part is

$$\psi_i = \omega + \sum_{j=1}^m \alpha_j x_{i-j} + \sum_{j=1}^q \beta_j \psi_{i-j} \quad (103)$$

and the Hawkes part has the exponential kernel(28)

$$\nu(t) = \sum_{j=1}^P \gamma_j e^{-\varphi_j t} \quad (104)$$

so that

$$\begin{aligned} \int_0^t \nu(t-s) dN_s &= \int_0^t \sum_{j=1}^P \gamma_j e^{-\varphi_j (t-s)} dN_s \\ &= \sum_{k=0}^{N_t} \nu(t-t_k) \\ &= \sum_{k=0}^{N_t} \sum_{j=1}^P \gamma_j e^{-\varphi_j (t-t_k)} \\ &= \sum_{j=1}^P \gamma_j \sum_{k=0}^{N_t} e^{-\varphi_j (t-t_k)} \\ &= \sum_{j=1}^P \gamma_j B_j(N_t) \end{aligned} \quad (105)$$

where we have replaced  $\alpha = \gamma$  and  $\beta = \varphi$  in the Hawkes part so that the parameter names do not conflict with the ACD part where  $\alpha$  and  $\beta$  are also used as parameter names. The Hawkes part of the intensity has a recursive structure similar to that of the compensator. Let

$$\begin{aligned} B_j(i) &= \sum_{k=0}^{i-1} e^{-\varphi_j (t-t_k)} \\ &= (1 + B_j(i-1)) e^{-\varphi_j (t-t_i)} \end{aligned} \quad (106)$$

where  $B_j(0) = 0$ . Then we have

$$\lambda(t) = \lambda_0(t) + \frac{1}{\omega + \sum_{j=1}^m \alpha_j x_{N_t-j} + \sum_{j=1}^q \beta_j \psi_{N_t-j}} + \sum_{j=1}^P \gamma_j B_j(N_t) \quad (107)$$

The log-likelihood for this hybrid model can be written as

$$\begin{aligned} \ln \mathcal{L}(\{t_i\}_{i=1, \dots, n}) &= \sum_{i=1}^n \left( \ln \lambda(t_i) - \int_{t_{i-1}}^{t_i} \lambda(t) dt \right) \\ &= \sum_{i=1}^n (\ln \lambda(t_i) - \Lambda(t_{i-1}, t_i)) \\ &= \sum_{i=1}^n (\ln \lambda(t_i) - \tilde{\mathcal{E}}_i) \end{aligned} \quad (108)$$

By direct calculation, combining (19) and (60), and letting  $x_i = t_i - t_{i-1}$  we have the compensator

$$\begin{aligned} \tilde{\mathcal{E}}_i &= \Lambda(t_{i-1}, t_i) \\ &= \int_{t_{i-1}}^{t_i} \lambda(t) dt \\ &= \int_{t_{i-1}}^{t_i} \left( \lambda_0(t) + \frac{1}{\psi_{N_t+1}} + \int_0^t \nu(t-s) dN_s \right) dt \\ &= \frac{x_i}{\psi_i} + \int_{t_{i-1}}^{t_i} \left( \lambda_0(t) + \int_0^t \nu(t-s) dN_s \right) dt \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \frac{x_i}{\psi_i} + \sum_{k=0}^{i-1} \sum_{j=1}^P \frac{\gamma_j}{\varphi_j} (e^{-\varphi_j(t_{i-1}-t_k)} - e^{-\varphi_j(t_i-t_k)}) \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \frac{x_i}{\psi_i} + \sum_{j=1}^P \frac{\gamma_j}{\varphi_j} (1 - e^{-\varphi_j x_i}) A_j(i-1) \end{aligned} \quad (109)$$

where  $\psi_i$  is defined by (103) and

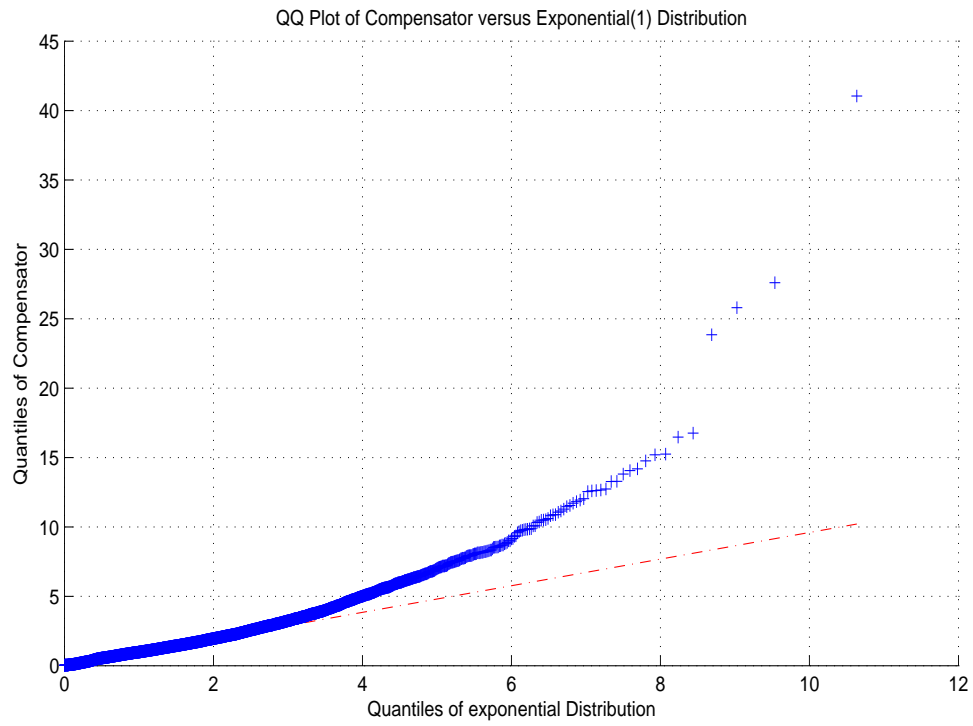
$$A_j(i) = 1 + e^{-\varphi_j x_i} A_j(i-1) \quad (110)$$

is given by (37) so that (108) can be written as

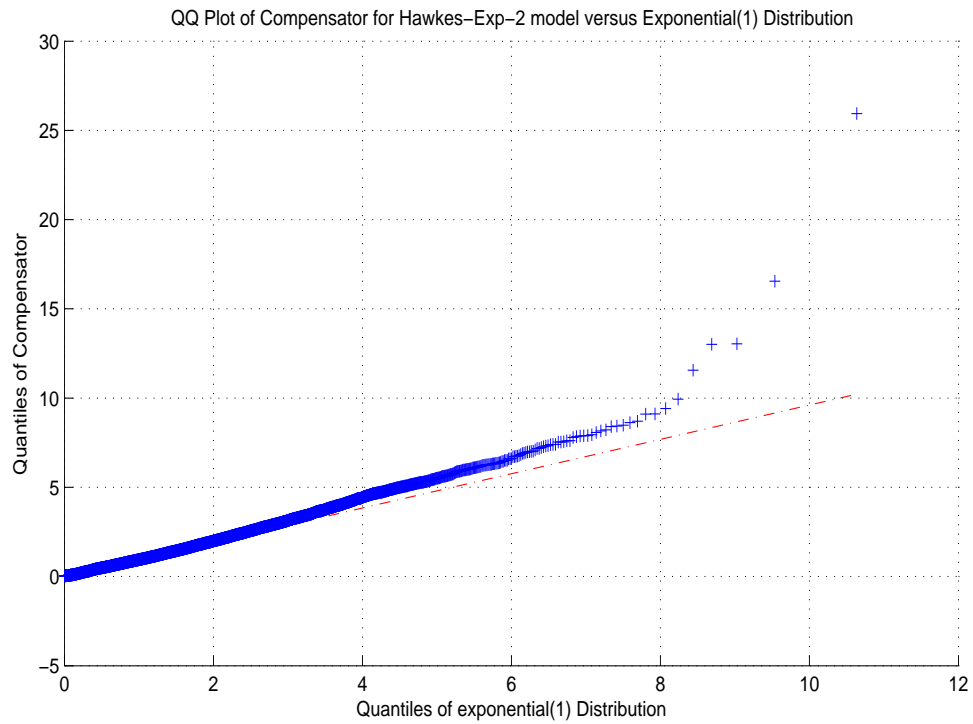
$$\begin{aligned} \ln \mathcal{L}(\{t_i\}_{i=0, \dots, n}) &= \sum_{i=1}^n (\ln \lambda(t_i) - \tilde{\mathcal{E}}_i) \\ &= \sum_{i=1}^n \left( \ln \lambda(t_i) - \left( \frac{x_i}{\psi_i} + \sum_{j=1}^P \frac{\gamma_j}{\varphi_j} (1 - e^{-\varphi_j x_i}) A_j(i-1) \right) \right) \\ &= \sum_{i=1}^n \ln \left( \frac{1}{\psi_i} + \sum_{k=0}^{i-1} \sum_{j=1}^P \gamma_j e^{-\varphi_j(t_i-t_k)} \right) - \left( \frac{x_i}{\psi_i} + \sum_{j=1}^P \frac{\gamma_j}{\varphi_j} (1 - e^{-\varphi_j x_i}) A_j(i-1) \right) \\ &= \sum_{i=1}^n \ln \left( \frac{1}{\psi_i} + \sum_{j=1}^P \gamma_j B_j(i) \right) - \left( \frac{x_i}{\psi_i} + \sum_{j=1}^P \frac{\gamma_j}{\varphi_j} (1 - e^{-\varphi_j x_i}) A_j(i-1) \right) \end{aligned} \quad (111)$$

### 1.6.1. Comparison of Hawkes-Exp vs ACD+Hawkes-Exp Model Fits.

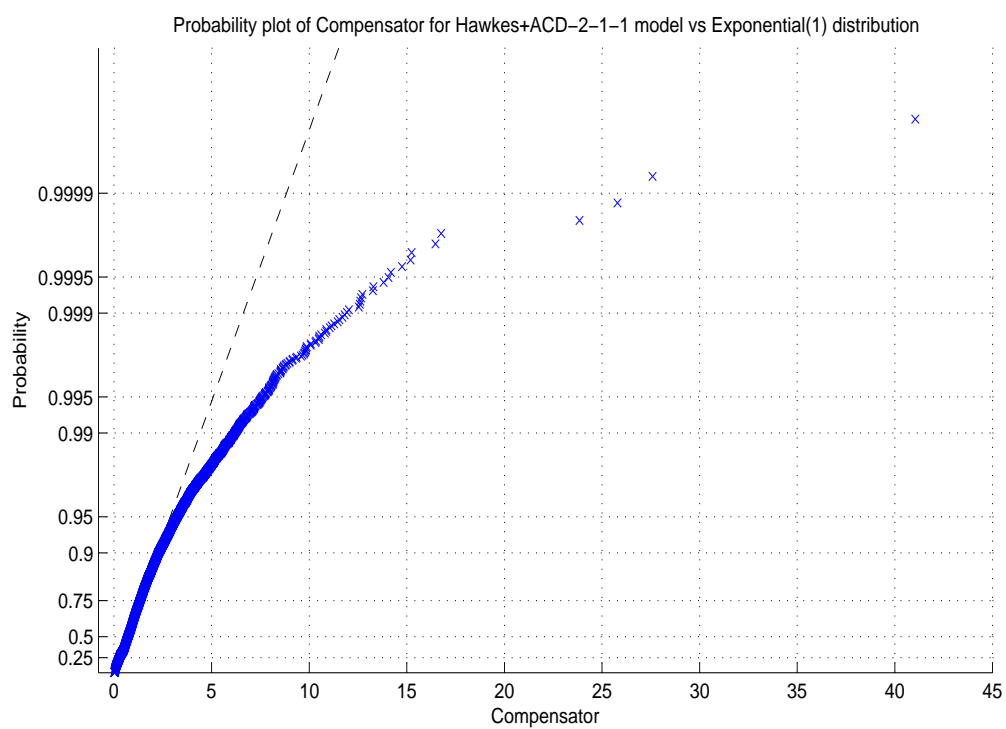
Now the Hawkes-Exp model of order 2 fit to the entire days worth of data on 2012-11-30 for SPY on the INET exchange will be compared against the ACD+Hawkes-Exp hybrid model of Hawkes order 2 and ACD params  $p=1, q=1$ .



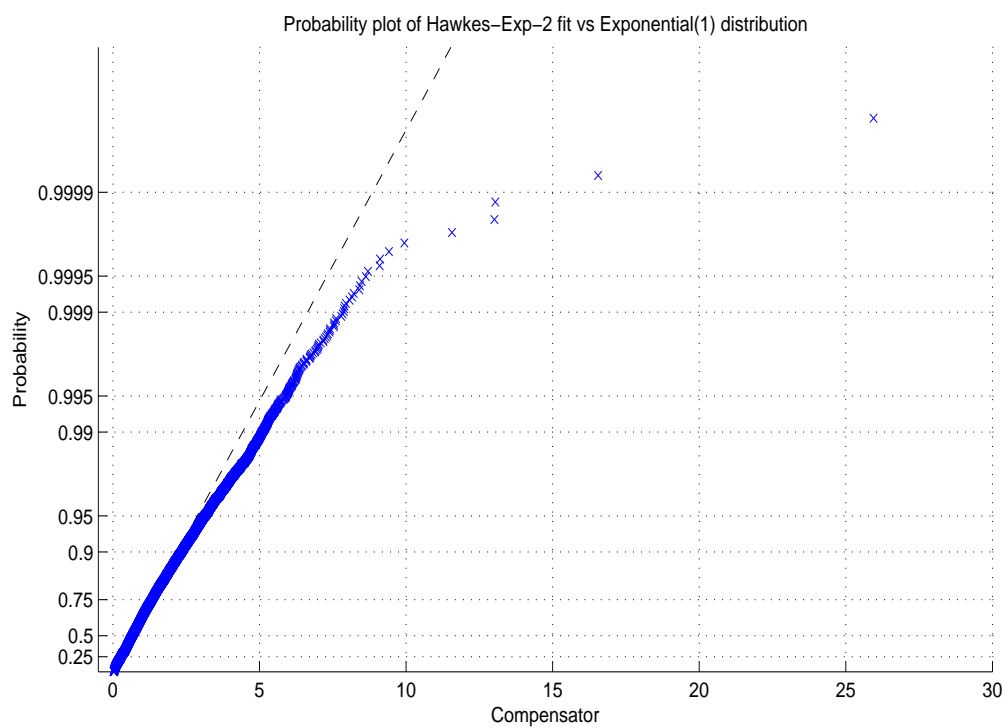
**Figure 8.** QQ Plot of Hawkes-ACD-2-1-1 compensator vs Exponential(1) Distribution



**Figure 9.** QQ Plot of Hawkes-Exp-2 compensator vs Exponential(1) Distribution



**Figure 10.** Probability Plot of Hawkes-ACD-2-1-1 compensator vs Exponential(1) Distribution



**Figure 11.** Probability Plot of Hawkes-Exp-2 compensator vs Exponential(1) Distribution

From the looks of it, the ACD+Hawkes-Exp model is quite a bit worse than plain Hawkes-Exp, either that or I have made a mistake in deriving the hybrid model expressions.

### 1.7. Multivariate Hawkes Models.

Let  $M \in \mathbb{N}^*$  and  $\{(t_i^m)\}_{m=1, \dots, M}$  be an  $M$ -dimensional point process. The associated counting process will be denoted  $N_t = (N_t^1, \dots, N_t^M)$ . A multivariate Hawkes process [8][6][11] is defined with intensities  $\lambda^m(t)$ ,  $m = 1 \dots M$  given by

$$\begin{aligned}
\lambda^m(t) &= \lambda_0^m(t) \kappa^m + \sum_{n=1}^M \int_0^t \sum_{j=1}^P \alpha_j^{m,n} e^{-\beta_j^{m,n}(t-s)} dN_s^n \\
&= \lambda_0^m(t) \kappa^m + \sum_{n=1}^M \sum_{j=1}^P \sum_{t_k^n < t} \alpha_j^{m,n} e^{-\beta_j^{m,n}(t-t_k^n)} \\
&= \lambda_0^m(t) \kappa^m + \sum_{n=1}^M \sum_{j=1}^P \alpha_j^{m,n} \sum_{t_k^n < t} e^{-\beta_j^{m,n}(t-t_k^n)} \\
&= \lambda_0^m(t) \kappa^m + \sum_{n=1}^M \sum_{j=1}^P \alpha_j^{m,n} \sum_{t_k^n < t} e^{-\beta_j^{m,n}(t-t_k^n)} \\
&= \lambda_0^m(t) \kappa^m + \sum_{n=1}^M \sum_{j=1}^P \alpha_j^{m,n} \sum_{k=0}^{N_t^n-1} e^{-\beta_j^{m,n}(t-t_k^n)} \\
&= \lambda_0^m(t) \kappa^m + \sum_{n=1}^M \sum_{j=1}^P \alpha_j^{m,n} B_j^{m,n}(N_t^n)
\end{aligned} \tag{112}$$

where in this parameterization  $\kappa$  is a vector which scales the baseline intensities, in this case, specified by piecewise polynomial splines (132). We can write  $B_j^{m,n}(i)$  recursively

$$\begin{aligned}
B_j^{m,n}(i) &= \sum_{k=0}^{i-1} e^{-\beta_j^{m,n}(t-t_k^n)} \\
&= (1 + B_j^{m,n}(i-1)) e^{-\beta_j^{m,n}(t-t_i^n)}
\end{aligned} \tag{113}$$

In the simplest version with  $P=1$  and  $\lambda_0^m(t)=1$  constant we have

$$\begin{aligned}
\lambda^m(t) &= \kappa^m + \sum_{n=1}^M \int_0^t \alpha^{m,n} e^{-\beta^{m,n}(t-s)} dN_s^n \\
&= \kappa^m + \sum_{n=1}^M \sum_{k=0}^{N_t^n-1} \alpha^{m,n} e^{-\beta^{m,n}(t-t_k^n)} \\
&= \kappa^m + \sum_{n=1}^M \alpha^{m,n} \sum_{k=0}^{N_t^n-1} e^{-\beta^{m,n}(t-t_k^n)} \\
&= \kappa^m + \sum_{n=1}^M \alpha^{m,n} B_1^{m,n}(N_t^n)
\end{aligned} \tag{114}$$

Rewriting (114) in vectorial notion, we have

$$\lambda(t) = \kappa + \int_0^t G(t-s) dN_s \quad (115)$$

where

$$G(t) = (\alpha^{m,n} e^{-\beta^{m,n}(t-s)})_{m,n=1\dots M} \quad (116)$$

Assuming stationarity gives  $E[\lambda(t)] = \mu$  a constant vector and thus

$$\begin{aligned} \mu &= \frac{\kappa}{I - \int_0^\infty G(u) du} \\ &= \frac{\kappa}{I - \left(\frac{\alpha^{m,n}}{\beta^{m,n}}\right)} \\ &= \frac{\kappa}{I - \Gamma} \end{aligned} \quad (117)$$

A sufficient condition for a multivariate Hawkes process to be stationary is that the spectral radius of the branching matrix

$$\Gamma = \int_0^\infty G(s) ds = \frac{\alpha^{m,n}}{\beta^{m,n}} \quad (118)$$

be strictly less than 1. The spectral radius of the matrix  $G$  is defined as

$$\rho(G) = \max_{a \in \mathcal{S}(G)} |a| \quad (119)$$

where  $\mathcal{S}(G)$  denotes the set of eigenvalues of  $G$ .

### 1.7.1. The Compensator.

The compensator of the  $m$ -th coordinate of a multivariate Hawkes process between two consecutive events  $t_{i-1}^m$  and  $t_i^m$  of type  $m$  is given by

$$\begin{aligned} \Lambda^m(t_{i-1}^m, t_i^m) &= \int_{t_{i-1}^m}^{t_i^m} \lambda^m(s) ds \\ &+ \int_{t_{i-1}^m}^{t_i^m} \sum_{n=1}^M \sum_{t_k^n < t_i^m} \sum_{j=1}^P \alpha_j^{m,n} e^{-\beta_j^{m,n}(s-t_k^n)} ds \\ &= \int_{t_{i-1}^m}^{t_i^m} \lambda^m(s) ds \\ &+ \sum_{n=1}^M \sum_{t_k^n < t_{i-1}^m} \sum_{j=1}^P \frac{\alpha_j^{m,n}}{\beta_j^{m,n}} [e^{-\beta_j^{m,n}(t_{i-1}^m - t_k^n)} - e^{-\beta_j^{m,n}(t_i^m - t_k^n)}] \\ &+ \sum_{n=1}^M \sum_{t_{i-1}^m \leq t_k^n < t_i^m} \sum_{j=1}^P \frac{\alpha_j^{m,n}}{\beta_j^{m,n}} [1 - e^{-\beta_j^{m,n}(t_i^m - t_k^n)}] \end{aligned} \quad (120)$$

To save a considerable amount of computational complexity, note that we have the recursion

$$\begin{aligned} A_j^{m,n}(i) &= \sum_{t_k^n < t_i^m} e^{-\beta_j^{m,n}(t_i^m - t_k^n)} \\ &= e^{-\beta_j^{m,n}(t_i^m - t_{i-1}^m)} A_j^{m,n}(i-1) + \sum_{t_{i-1}^m \leq t_k^n < t_i^m} e^{-\beta_j^{m,n}(t_i^m - t_k^n)} \end{aligned} \quad (121)$$

and rewrite (120) as

$$\begin{aligned} \Lambda^m(t_{i-1}^m, t_i^m) &= \kappa^m \int_{t_{i-1}^m}^{t_i^m} \lambda_0^m(s) ds + \int_{t_{i-1}^m}^{t_i^m} \sum_{n=1}^M \sum_{j=1}^P \sum_{t_k^n < s} \alpha_j^{m,n} e^{-\beta_j^{m,n}(s - t_k^n)} ds \\ &= \kappa^m \int_{t_{i-1}^m}^{t_i^m} \lambda_0^m(s) ds \\ &\quad + \sum_{n=1}^M \sum_{j=1}^P \frac{\alpha_j^{m,n}}{\beta_j^{m,n}} \left[ (1 - e^{-\beta_j^{m,n}(t_i^m - t_{i-1}^m)}) \times A_j^{m,n}(i-1) + \sum_{t_{i-1}^m \leq t_k^n < t_i^m} (1 - e^{-\beta_j^{m,n}(t_i^m - t_k^n)}) \right] \\ &= \kappa^m \int_{t_{i-1}^m}^{t_i^m} \lambda_0^m(s) ds \\ &\quad + \sum_{n=1}^M \sum_{j=1}^P \frac{\alpha_j^{m,n}}{\beta_j^{m,n}} \left[ (1 - e^{-\beta_j^{m,n}(t_i^m - t_{i-1}^m)}) \times \left( \sum_{t_k^n < t_{i-1}^m} e^{-\beta_j^{m,n}(t_{i-1}^m - t_k^n)} \right) + \sum_{t_{i-1}^m \leq t_k^n < t_i^m} (1 - e^{-\beta_j^{m,n}(t_i^m - t_k^n)}) \right] \end{aligned} \quad (122)$$

where we have the initial conditions  $A_j^{m,n}(0) = 0$ .

### 1.7.2. Log-Likelihood.

The log-likelihood of the multivariate Hawkes process can be computed as the sum of the log-likelihoods for each coordinate. Let

$$\ln \mathcal{L}(\{t_i\}_{i=1, \dots, N_T}) = \sum_{m=1}^M \ln \mathcal{L}^m(\{t_i\}) \quad (123)$$

where each term is defined by

$$\ln \mathcal{L}^m(\{t_i\}) = \int_0^T (1 - \lambda^m(s)) ds + \int_0^T \ln \lambda^m(s) dN_s^m \quad (124)$$

which in this case can be written as

$$\begin{aligned} \ln \mathcal{L}^m(\{t_i\}) &= T - \Lambda^m(0, T) + \sum_{i=1}^{N_T} z_i^m \ln \left( \lambda_0^m(t_i) \kappa^m + \sum_{n=1}^M \sum_{j=1}^P \sum_{t_k^n < t_i} \alpha_j^{m,n} e^{-\beta_j^{m,n}(t_i - t_k^n)} \right) \\ &= T - \Lambda^m(0, T) + \sum_{i=1}^{N_T^m} \ln \left( \lambda_0^m(t_i^m) \kappa^m + \sum_{n=1}^M \sum_{j=1}^P \sum_{t_k^n < t_i^m} \alpha_j^{m,n} e^{-\beta_j^{m,n}(t_i^m - t_k^n)} \right) \end{aligned} \quad (125)$$

where again  $t_{N_T} = T$  and

$$z_i^m = \begin{cases} 1 & \text{event } t_i \text{ of type } m \\ 0 & \text{otherwise} \end{cases} \quad (126)$$

and

$$\Lambda^m(0, T) = \int_0^T \lambda^m(t) dt = \sum_{i=1}^{N_T^m} \Lambda^m(t_{i-1}^m, t_i^m) \quad (127)$$

where  $\Lambda^m(t_{i-1}^m, t_i^m)$  is given by (122). Similiar to to the one-dimensional case, we have the recursion

$$R_j^{m,n}(i) = \sum_{t_k^n < t_j^m} e^{-\beta_j^{m,n}(t_i^m - t_k^n)} = \begin{cases} e^{-\beta_j^{m,n}(t_i^m - t_{i-1}^m)} R_j^{m,n}(i-1) + \sum_{t_{i-1}^m \leq t_k^n < t_i^m} e^{-\beta_j^{m,n}(t_i^m - t_k^n)} & \text{if } m \neq n \\ e^{-\beta_j^{m,n}(t_i^m - t_{i-1}^m)} (1 + R_j^{m,n}(i-1)) & \text{if } m = n \end{cases} \quad (128)$$

so that (125) can be rewritten as

$$\begin{aligned} \ln \mathcal{L}^m(\{t_i\}) &= T - \kappa^m \int_0^T \lambda_0^m(t) dt - \dots \\ &\dots - \sum_{i=1}^{N_T^m} \sum_{n=1}^M \sum_{j=1}^P \frac{\alpha_j^{m,n}}{\beta_j^{m,n}} \left[ (1 - e^{-\beta_j^{m,n}(t_i^m - t_{i-1}^m)}) \times A_j^{m,n}(i-1) + \sum_{t_{i-1}^m \leq t_k^n < t_i^m} (1 - e^{-\beta_j^{m,n}(t_i^m - t_k^n)}) \right] + \dots \\ &\dots + \sum_{i=1}^{N_T^m} \ln \left( \lambda_0^m(t_i^m) \kappa^m + \sum_{n=1}^M \sum_{j=1}^P \alpha_j^{m,n} R_j^{m,n}(i) \right) \end{aligned} \quad (129)$$

with initial conditions  $R_j^{m,n}(0) = 0$  and  $A_j^{m,n}(0) = 0$  where  $T = t_N$  where  $N$  is the number of observations,  $M$  is the number of dimensions, and  $P$  is the order of the model. Again,  $T$  can be dropped from the equation for the purposes of optimization.

## 2. NUMERICAL METHODS

### 2.1. The Nelder-Mead Algorithm.

The Nelder-Mead simplex algorithm[5] was used to optimize the likelihood expressions given above.

#### 2.1.1. Starting Points for Optimizing the Hawkes Process of Order $P$ .

A starting point for the optimization of a Hawkes process of order  $P$  with an “exact” unconditional intensity was chosen as the most reasonable starting point, but it is by no means claimed to be the best. Let  $x_i = t_i - t_{i-1}$  be the interval between consecutive arrival times as in the ACD model (16). Then set the initial value of  $\lambda_0$  to  $\frac{0.5}{E[x_i]}$ ,  $\alpha_{1...P} = \frac{1}{P}$  and  $\beta_{1...P} = 2$ . This gives an unconditional mean of  $E[x_i]$  for these parameters used as a starting point for the Nelder-Mead algorithm.

## 3. EXAMPLES

### 3.1. Millisecond Resolution Trade Sequences.

The source data has resolution of milliseconds but the data is transformed prior to estimation by dividing each time by 1000 so that the unit of time is seconds. Also, trades occurring at the same price within 2ms of each other are dropped from the analysis. Further work will be done to find the optimal level of time aggregation, ideally the data would be timestamped with nanosecond resolution and this will be done in the future.

#### 3.1.1. Adjusting for the Deterministic Daily Intensity Variation.

It is a well known fact that arrival rates (and the closely related volatility) have daily “seasonal” or “diurnal” patterns where trading activity peaks after open and before close and has a low around the middle of the day known as the “lunchtime effect”. In order to account for this we will fit a cubic spline with 14 knot points spaced every 30 minutes, including the opening and closing times of  $t = 0$  and  $t = 6.5 \times 60 \times 60 = 23400$  respectively since  $t$  has units of seconds. Let the adjusted durations be defined

$$\tilde{x}_i = \phi(t_i) x_i \quad (130)$$

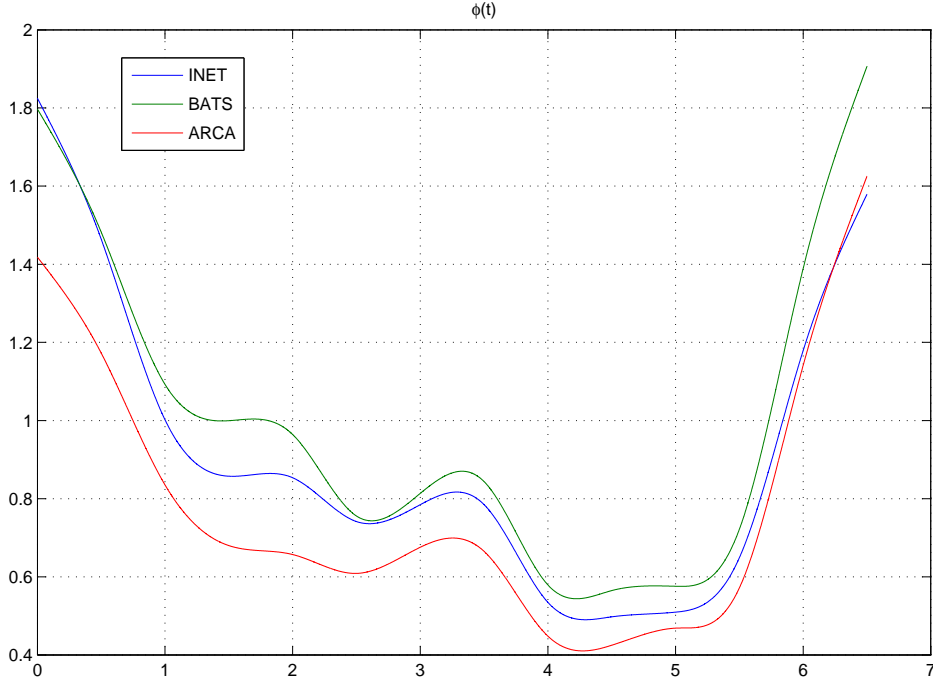
where  $x_i = t_i - t_{i-1}$  is the unadjusted duration and  $\phi(t_i)$  is a (piecewise polynomial) cubic spline with knot points at  $t(zj)$  with values given by  $P_j$

$$P_j = \frac{1}{(N_{t(zj)+w} - N_{t(zj)-w})} \sum_{i=N_{t(zj)-w}}^{N_{t(zj)+w}} \frac{1}{x_i} \text{ for } j = 0 \dots 13 \quad (131)$$

where  $z = 60 \times 30 = 1800$  is the number of seconds in a half-hour and  $j = 0 \dots (6.5 \times 2)$ . The first and last knots have a “window” of 30-minutes whereas the interior knot points have a window of 1 hour looking forward and backward in time 30-minutes, the first knot point only looks forward and the last knot point only looks backward. This gives us the “deterministic baseline intensity” which is a piecewise polynomial cubic spline function whose exact form is not mentioned here since it is not the focus of the paper.

$$\lambda_0(t) = f(t, P_0, \dots, P_j) \quad (132)$$

The following figure shows the “deterministic part” of the intensity estimated for SPY on 2012-11-30 for INET, BATS, and ARCA.



**Figure 12.** Interpolating spline  $\phi(t)$  for SPY on 2012-11-30

### 3.1.2. Univariate Hawkes model fit to SPY (SPDR S&P 500 ETF Trust).

Consider these parameter estimates for the (univariate) Hawkes model of various orders fitted to data generated by trades of the symbol SPY traded on the NASDAQ on Nov 30th, 2012. The unconditional sample mean intensity for this symbol on this day on this exchange was 0.8882491159065832 trades per second where the number of samples is  $n = 20787$ . The data presented here has not been “deseasonalized”, the analysis with deterministic diurnal variation accounted for will be presented in the next section. As can be seen,  $P = 6$  provides the best likelihood but a more rigorous method to choose  $P$  would be to use some information criterion like Bayes or Akaike to decide the order  $P$ . Error bars are not provided, but presumably they could be estimated with derivative information. Note that the closer  $E[\lambda(t)]$  to 0.8882491159065832 and  $E[\Lambda]$  and  $\text{Var}[\Lambda]$  to 1.0 the better, since  $\Lambda$  should be exponentially distributed with mean 1 by design and for a Poisson process the mean and variance are equal. The next thing to check

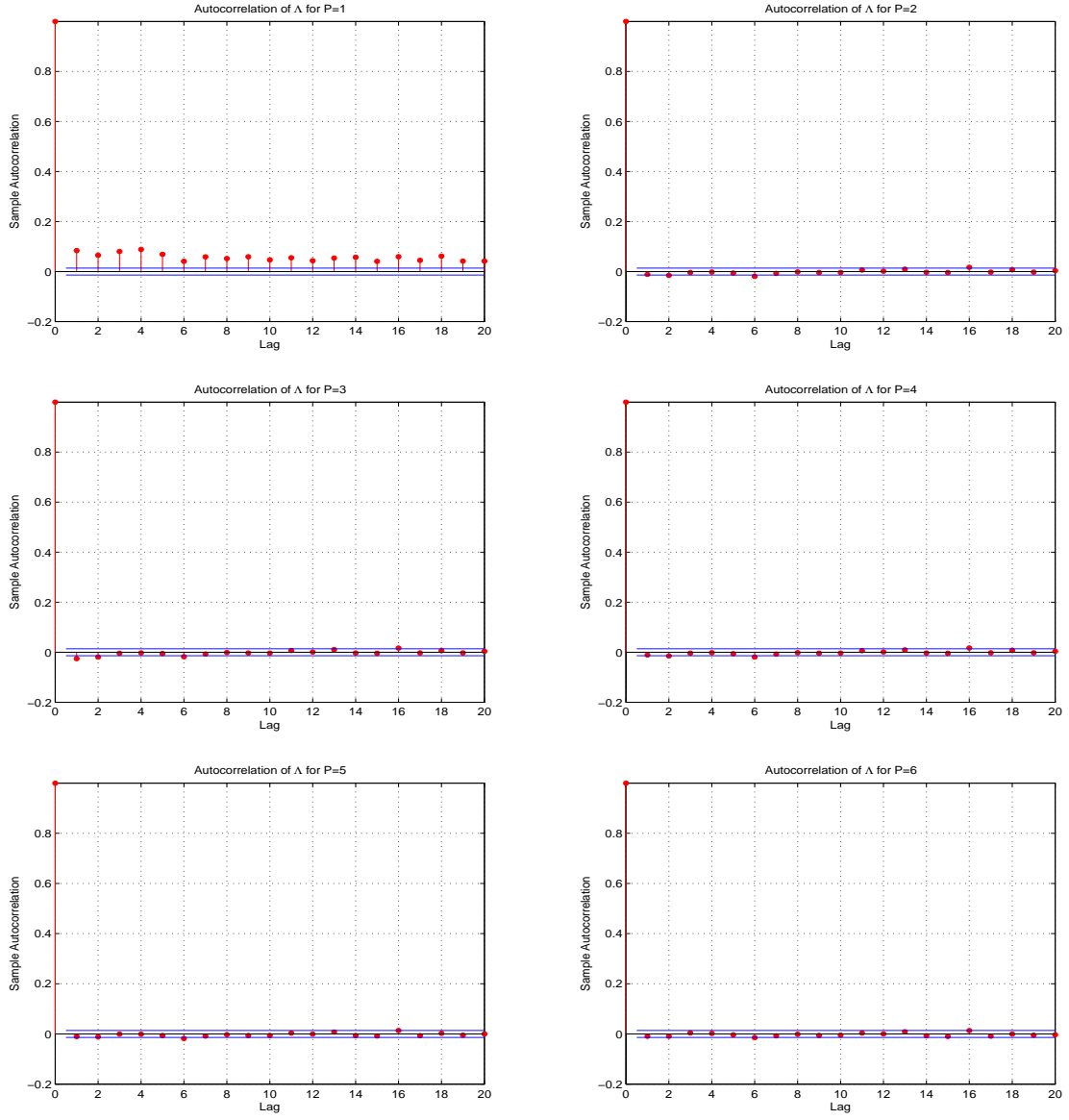
is that the  $\Lambda$  series is not autocorrelated.

| $P$ | $\kappa$    | $\alpha_{1...P}$   | $\beta_{1...P}$  | $\ln \mathcal{L}(\{t_i\})$ | $E[\lambda(t)]$ | $E[\Lambda]$ | $\text{Var}[\Lambda]$ |
|-----|-------------|--|--|----------------------------|-----------------|--------------|-----------------------|
| 1   | 0.502711246 | 19.66948678  | 45.315830024   | -3504.24543                | 0.88826610      | 0.9999990    | 1.8638729             |
| 2   | 0.179395347 | 23.8186109<br>0.09959041   | 61.07892017<br>0.243158578   | -1288.3557                 | 0.89489310      | 0.9999972    | 1.1880598             |
| 3   | 0.179558266 | 0.08621919<br>0.22766134<br>28.5616786   | 0.219020402<br>45.23233626<br>55.87754150  | -1586.7082                 | 1.99153298      | 1.11040384   | 1.24678422            |
| 4   | 0.178874698 | 0.09893214<br>0.18481509<br>11.0305006<br>12.5980362                                   | 0.241418546<br>50.59817301<br>66.99771955<br>57.05863369                               | -1283.76240                | 0.88938728      | 0.99874524   | 1.1871400             |
| 5   | 0.153072454 | 8.017991269<br>0.000000005<br>18.28544127<br>1.615965008<br>0.060456987                | 68.68917670<br>79.55782766<br>83.46583667<br>14.45235850<br>0.151551338                | -1051.97938                | 0.99747221      | 1.01670503   | 1.16016527            |
| 6   | 0.132054503 | 0.532479235<br>0.034373403<br>13.04953708<br>4.208599107<br>7.090279453<br>2.291178834 | 4.108969054<br>0.092093459<br>84.86207394<br>81.71142685<br>67.23003519<br>56.20297618 | -991.14436                 | 0.90660986      | 1.00006670   | 1.12981528            |

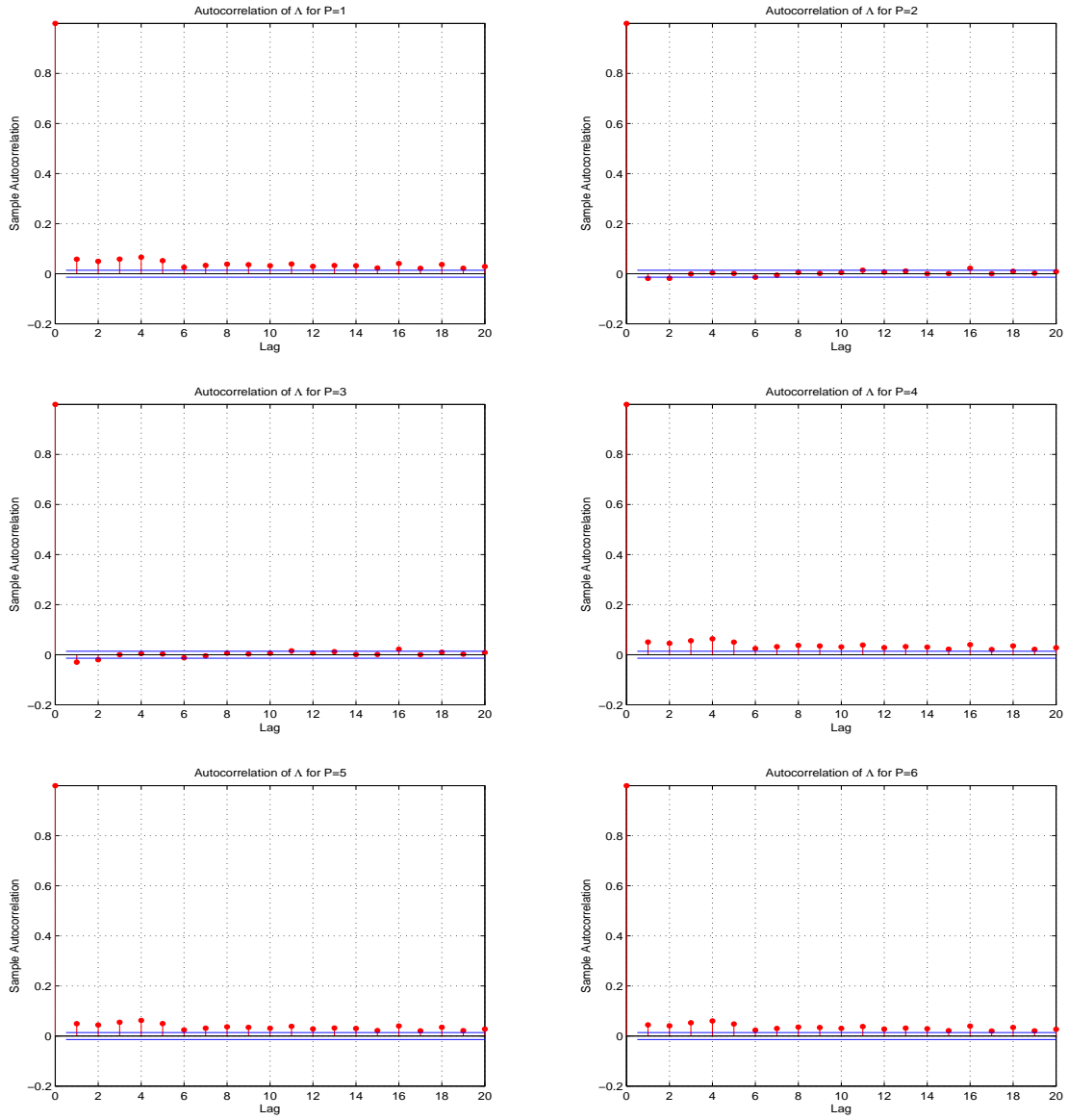
**Table 2.** Parameters and statistics for model fitted to data without diurnal adjustments

| $P$ | $\kappa$     | $\alpha_{1...P}$   | $\beta_{1...P}$  | $\ln \mathcal{L}(\{t_i\})$ | $E[\lambda(t)]$ | $E[\Lambda]$ | $\text{Var}[\Lambda]$ |
|-----|--------------|--|--|----------------------------|-----------------|--------------|-----------------------|
| 1   | 0.5796428053 | 20.7816860009  | 49.181292797   | -2565.16186                |                 | 1.000005090  | 1.64713115            |
| 2   | 0.2972951255 | 24.336309087<br>0.1366737439   | 63.30799040<br>0.426958321   | -1147.38872                |                 | 1.000002078  | 1.15682329            |
| 3   | 0.3105850108 | 29.625207375<br>0.0000000101<br>0.1200815585   | 58.78427931<br>32.16156796<br>0.405484625  | -1422.551267               |                 | 1.108843464  | 1.23286963            |
| 4   | 0.5627834858 | 0.0000000264<br>6.4766935751<br>14.656872968<br>1.8317154168                                 | 40.62190533<br>49.10661802<br>60.00475526<br>21.39853548                               | -2364.699597               |                 | 1.022407180  | 1.59177967            |
| 5   | 0.5506638255 | 0.0725319843<br>0.0507855259<br>6.8528913938<br>15.032951777<br>2.0993068921                 | 26.86479506<br>81.58572968<br>81.58572968<br>60.25583954<br>17.30297034                | -2152.462512               |                 | 1.011487836  | 1.53515842            |
| 6   | 0.5362685399 | 12.459351335<br>8.2747228669<br>0.0000000201<br>2.7582137937<br>0.0041661767<br>1.9821090294 | 77.72815398<br>69.01934786<br>53.74869710<br>47.94942161<br>42.42839207<br>13.72571940 | -1997.336098               |                 | 1.016450670  | 1.48640060            |

**Table 3.** Parameters and statistics for model fitted to data with diurnal adjustments

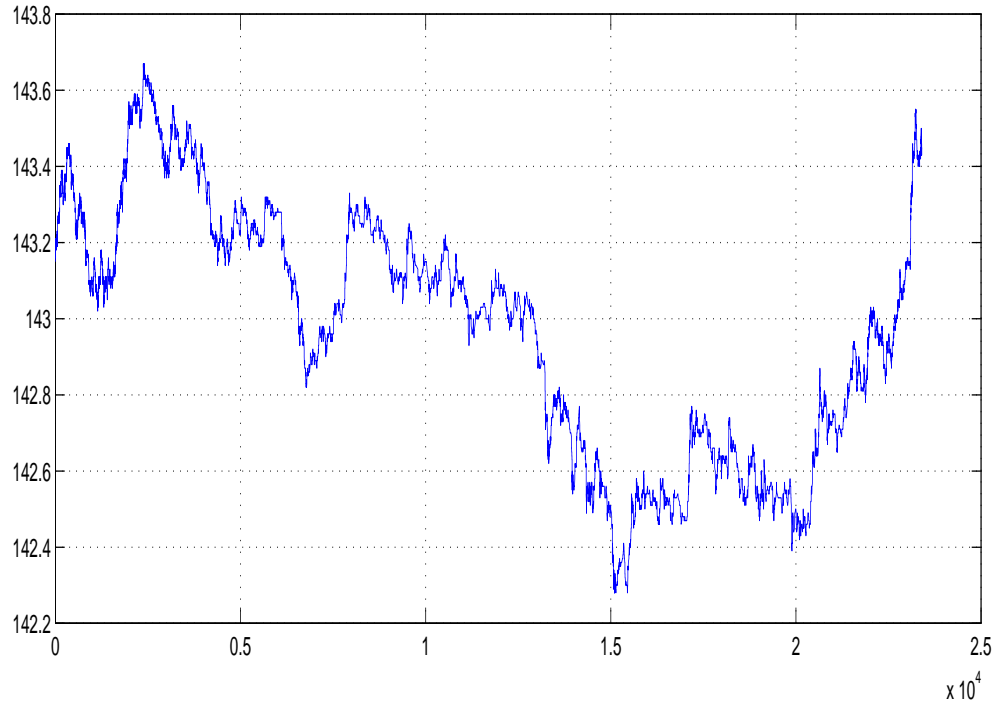


**Figure 13.** Autocorrelations of  $\Lambda(t_{i-1}, t_i)$  for  $P = 1 \dots 6$  without diurnal adjustments

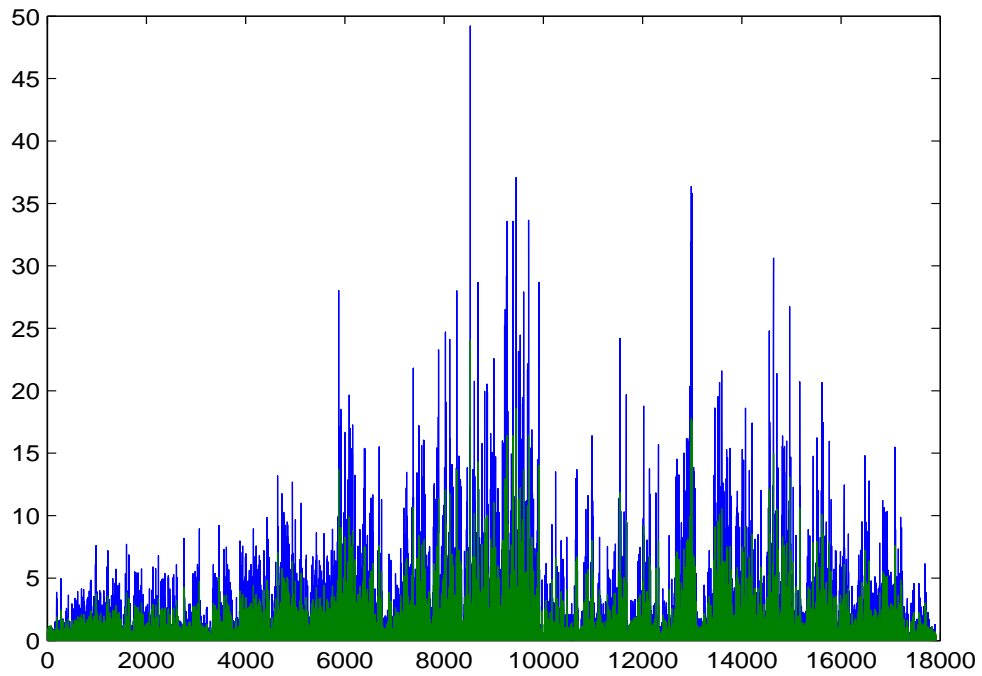


**Figure 14.** Autocorrelations of  $\Lambda(t_{i-1}, t_i)$  for  $P = 1 \dots 6$  with diurnal adjustments

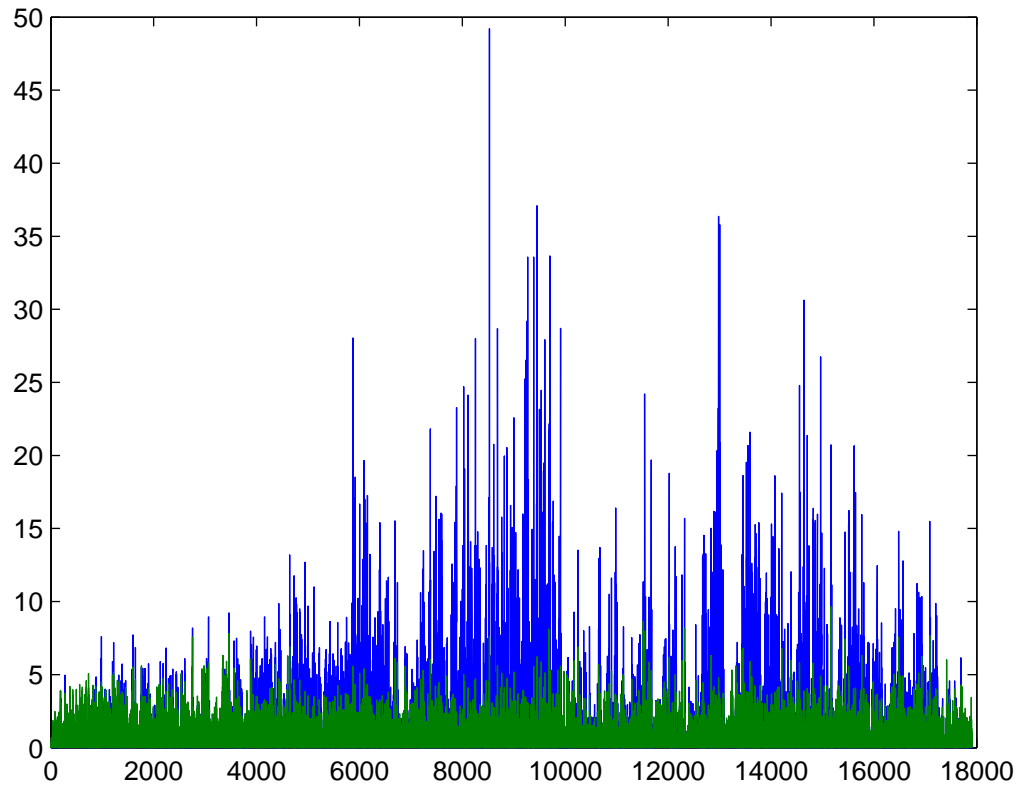
As can be seen by visually inspecting the autocorrelations, all of the residual series are pretty-much acceptable \*without\* diurnal adjustments except for  $P=1$  with still had significant leftover autocorrelation. Strangely, it seems that inclusion of the diurnal adjustment significantly worsens the model fit in nearly all cases. I am tempted to suspect something wrong with the code.



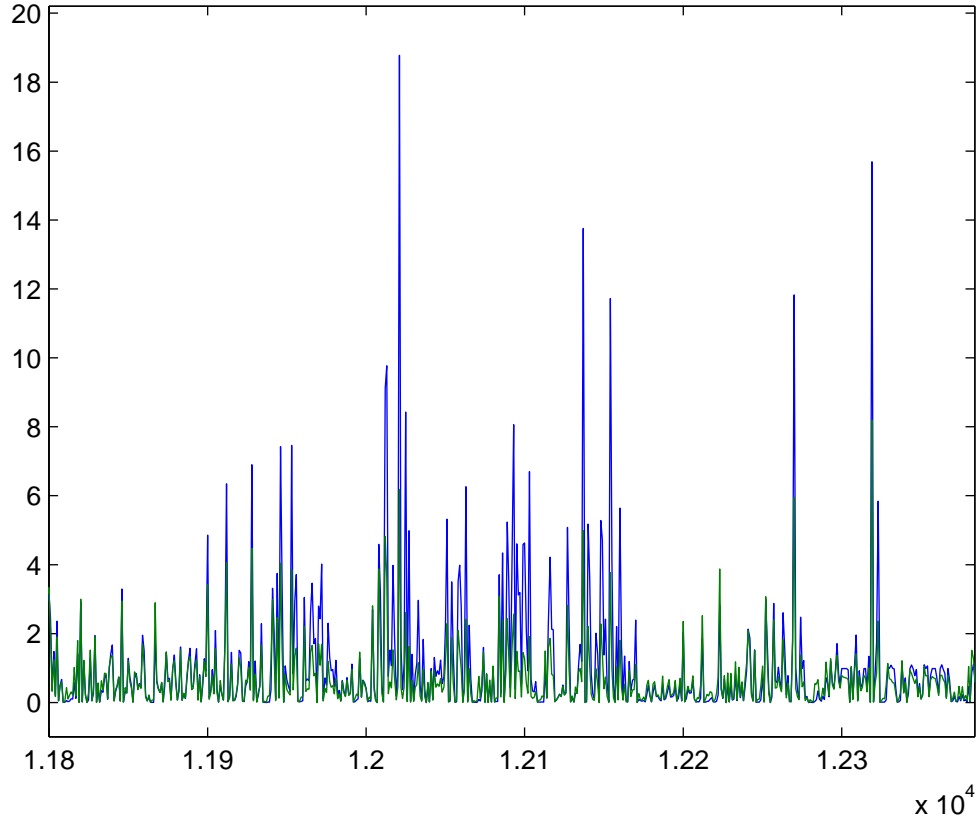
**Figure 15.** Price history for SPY traded on INET on Oct 22nd, 2012



**Figure 16.**  $x_i = t_i - t_{i-1}$  in blue and  $\{\Lambda(t_{i-1}, t_i): P=1\}$  in green



**Figure 17.**  $x_i = t_i - t_{i-1}$  in blue and  $\{\Lambda(t_{i-1}, t_i): P=6\}$  in green



**Figure 18.** Zoomed in view of  $x_i = t_i - t_{i-1}$  in blue and  $\{\Lambda(t_{i-1}, t_i); P=6\}$  in green

### 3.1.3. Multivariate SPY Data for 2012-08-14.

Consider a 5-dimensional multivariate Hawkes model of order  $P=1$  fit to data for SPY from 3 exchanges, INET, BATS, and ARCA on 2012-08-14. Both INET and BATS distinguish buys from sells whereas ARCA does not, hence 5 dimensional, 2 dimensions each for INET and BATS and 1 dimension for ARCA which will naturally have twice as high a rate as that for buys and sells considered separately. The 5 dimensions are organized as follows:

|           |            |           |            |             |
|-----------|------------|-----------|------------|-------------|
| BATS Buys | BATS Sells | INET Buys | INET Sells | ARCA Trades |
|-----------|------------|-----------|------------|-------------|

(133)

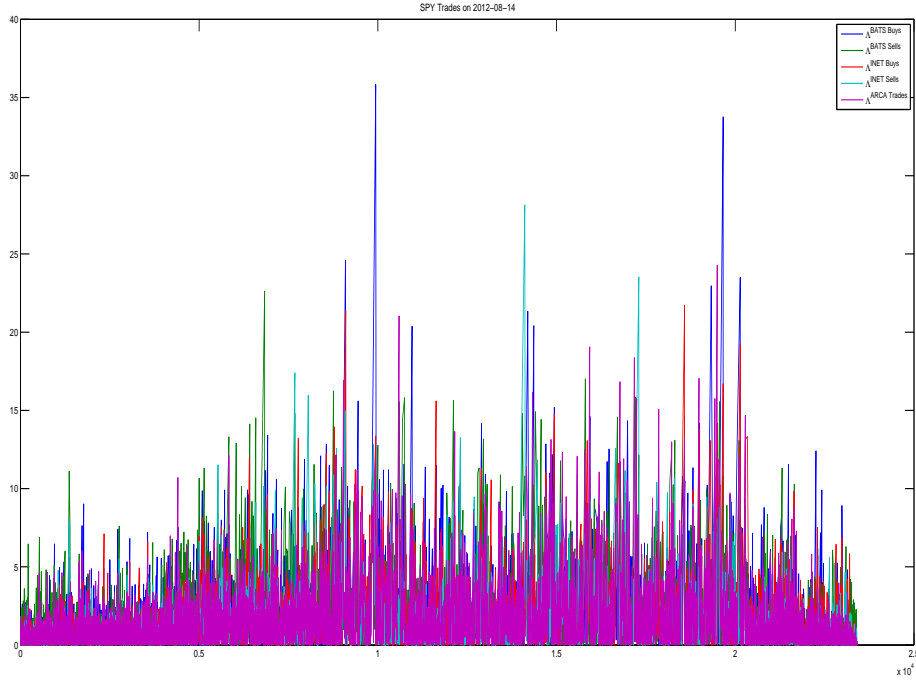


Figure 19.

We say trades for ARCA because the type sent from the data broker is Unknown, indicating that it is unknown whether it is a buyer or seller initiated trade. We have the following parameter estimates where “large” values of  $\alpha$  ( $>0.1$ ) are highlighted in bold.

$$\lambda = \begin{pmatrix} 0.25380789517348 \\ 0.269289236349466 \\ 0.221292886522613 \\ 0.158954542395839 \\ 0.371572853723448 \end{pmatrix} \quad (134)$$

$$\alpha = \begin{pmatrix} 4.3514 \times 10^{-9} & 0.011879 & \mathbf{0.2648} & 1.917 \times 10^{-8} & \mathbf{0.10771} \\ 0.021881 & 2.6164 \times 10^{-8} & 2.5725 \times 10^{-8} & 0.024946 & \mathbf{0.25138} \\ \mathbf{0.29092} & \mathbf{0.51715} & 1.1254 \times 10^{-8} & 0.0029919 & 0.004607 \\ 0.0041449 & \mathbf{0.52852} & 0.018077 & 3.2535 \times 10^{-9} & 0.0237 \\ 0.021501 & \mathbf{0.71358} & \mathbf{1.0954} & \mathbf{0.15264} & 4.1222 \times 10^{-9} \end{pmatrix} \quad (135)$$

$$\beta = \begin{pmatrix} 1.0954 & 10.803 & 16.665 & 20.188 & 9.6059 \\ 5.6238 & 11.558 & 16.721 & 18.304 & 7.9016 \\ 7.8125 & 15.299 & 16.431 & 14.702 & 6.6458 \\ 8.3083 & 15.758 & 17.749 & 12.953 & 3.1621 \\ 9.4264 & 16.369 & 19.303 & 11.071 & 2.8302 \end{pmatrix} \quad (136)$$

with a log-likelihood score of 39714.1497.

### 3.1.4. Multivariate SPY Data for 2012-11-19.

Consider the same symbol, SPY, as a 5-dimensional Hawkes process as in 3.1.3, for a different day, on 2012-11-19, estimated with order  $P = 2$  for a total of 105 parameters.  $\alpha_j$  coefficients that are  $>0.1$  are highlighted in bold. The parameters listed below resulted in a log-likelihood value of 36543.8529. An interesting pattern emerges in the  $\beta$  coefficients where it takes on some approximate stair-step pattern ranging from 2 to 22. This might be indicative of some fixed-frequency algorithms operating across the different exchanges at approximate 1-second intervals.

$$\lambda = \begin{pmatrix} 0.113371928486215301 \\ 0.116069526955243113 \\ 0.120010488406567112 \\ 0.140864383337674315 \\ 0.236370243964866722 \end{pmatrix} \quad (137)$$

$$\alpha_1 = \begin{pmatrix} 0.000000400520039 & 0.000743243048280 & 0.0730760324025721 & 0.0235425002925593 & \mathbf{0.14994903109} \\ 0.000836306407254 & 0.000048087752871 & 0.0009983197029208 & \mathbf{0.36091325418001} & 0.0303494022034 \\ 0.000007657273830 & 0.008293393618634 & 0.0000346485386433 & \mathbf{0.55279157046563} & 0.0303324666473 \\ 0.000000051209296 & 0.044218944305554 & 0.0165858723488658 & 0.0002898699267899 & \mathbf{0.12041188377} \\ 0.000343063367497 & 0.019728025120072 & \mathbf{0.22664219457110} & \mathbf{0.20883023885464} & 0.0002187148763 \end{pmatrix} \quad (138)$$

$$\alpha_2 = \begin{pmatrix} 0.0247169438667 & 0.045938324942878493 & \mathbf{0.52035195378729} & 0.0015976654768 & 0.0219865625857849 \\ \mathbf{0.10369500283} & 0.000000961851428240 & 0.0058603752158104 & \mathbf{0.17159388407} & 0.0001956826269151 \\ 0.0619247685514 & 0.005680420895898976 & 0.0000041940337011 & 0.0009132788022 & 0.0161550464515489 \\ 0.0073308612563 & \mathbf{0.3760898786954499} & 0.0078995090167169 & 0.0000971358022 & 0.0022020712790430 \\ \mathbf{0.37860663035} & \mathbf{0.8648532461379836} & 0.0096939577784123 & \mathbf{0.23909856627} & 0.0000001318796171 \end{pmatrix} \quad (139)$$

$$\beta_1 = \begin{pmatrix} 2.02691486662775 & 4.58853278669795 & 9.21516653991608 & 14.2039223554899 & 17.7230908440328108 \\ 2.30228990848878 & 5.70815142794409 & 9.75920981324501 & 15.0047495693597 & 17.1640776964259771 \\ 2.71360844613891 & 6.97390906252072 & 10.9112224210093 & 16.3935104902520 & 17.3801721025480269 \\ 3.18861359927744 & 6.93702281997507 & 12.0261860231254 & 17.5228876305459 & 17.8876296984556440 \\ 3.95262799649030 & 7.76155541730819 & 13.5039942724633 & 17.3549525971848 & 18.0730780733303966 \end{pmatrix} \quad (140)$$

$$\beta_2 = \begin{pmatrix} 19.6811983441165 & 20.56326127197891 & 18.53440853276660 & 11.10183435325997 & 5.955287687038747 \\ 20.2253306600591 & 21.39051471260508 & 16.97184115533537 & 9.548598696946248 & 5.459761230875715 \\ 20.2208259457254 & 22.20704300748698 & 17.88989095276187 & 8.724870367131993 & 4.215302773261564 \\ 19.7356631996375 & 21.67330389603866 & 15.76838788843381 & 7.534795006501931 & 3.517163899772246 \\ 20.2972304557004 & 19.06667927692781 & 13.19618799557176 & 6.812943703872132 & 2.825437512911523 \end{pmatrix} \quad (141)$$

## BIBLIOGRAPHY

- [1] C.G. Bowsher. Modelling security market events in continuous time: intensity based, multivariate point process models. *Journal of Econometrics*, 141(2):876–912, 2007.
- [2] Y. Cai, B. Kim, M. Leduc, K. Szczegot, Y. Yixiao and M. Zamfir. A model for intraday volatility. , 2007.
- [3] V. Chavez-Demoulin and JA McGill. High-frequency financial data modeling using hawkes processes. *Journal of Banking & Finance*, , 2012.
- [4] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey and D. E. Knuth. On the lambert w function. *Advances in Computational Mathematics*, 5(1):329–359, Dec 1996.
- [5] JE Dennis and D.J. Woods. Optimization on microcomputers: the nelder-mead simplex algorithm. *New Computing Environments: Microcomputers in Large-Scale Computing*, :116–122, 1987.
- [6] P. Embrechts, T. Liniger and L. Lin. Multivariate hawkes processes: an application to financial data. *Journal of Applied Probability*, 48:367–378, 2011.
- [7] R.F. Engle and J.R. Russell. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, :1127–1162, 1998.
- [8] A.G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.
- [9] N. Hautsch. *Econometrics of financial high-frequency data*. Springer, 2011.
- [10] A. Karr. *Point processes and their statistical inference*, volume 7. CRC, 1991.
- [11] T.J. Liniger. *Multivariate Hawkes Processes*. PhD thesis, Swiss Federal Institute Of Technology Zurich, 2009.
- [12] F. Lorenzen. *Analysis of Order Clustering Using High Frequency Data: A Point Process Approach*. PhD thesis, Tilburg School of Economics and Management, Finance Department, August 2012.
- [13] Y. Ogata. The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 30(1):243–261, 1978.
- [14] Y. Ogata. On lewis’ simulation method for point processes. *Information Theory, IEEE Transactions on*, 27(1):23–31, 1981.

- [15] T. Ozaki. Maximum likelihood estimation of hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, 31(1):145–155, 1979.
- [16] T. Ozaki and H. Akaike. On a generalized parametric model for hawkes' self-exciting point processes. *Institute of Statistical Mathematics Res. Memo*, 149, 1979.
- [17] Kelly Roach. Hypergeometric function representations. In *International Symposium on Symbolic and Algebraic Computation*, pages 301–308. 1996.
- [18] Donald E. Knuth Robert M. Corless, David J. Jeffrey. A sequence of series for the lambert w function. In *International Symposium on Symbolic and Algebraic Computation*, pages 197–204. 1997.
- [19] H. Shek. Modeling high frequency market order dynamics using self-excited point process. *Available at SSRN 1668160*, , 2010.
- [20] Ioane Muni Toke. An introduction to hawkes processes with applications to finance. , [http://fiquant.mas.ecp.fr/ioane\\_files/HawkesCourseSlides.pdf](http://fiquant.mas.ecp.fr/ioane_files/HawkesCourseSlides.pdf).
- [21] HM Wiseman. Quantum theory of continuous feedback. *Physical Review A*, 49(3):2133, 1994.

#### 4. APPENDIX

##### 4.1. The Lambert W Function $W(k, x)$ .

The Lambert W function [4][18] is the inverse of  $xe^x$  given by

$$\begin{aligned} W(z) &= \{x: xe^x = z\} \\ &= W(0, z) \\ &= 1 + (\ln(z) - 1) \exp\left(\frac{i}{2\pi} \int_0^\infty \frac{1}{x+1} \ln\left(\frac{x - i\pi - \ln(x) + \ln(z)}{x + i\pi - \ln(x) + \ln(z)}\right) dx\right) \\ &= \sum_{k=1}^{\infty} \frac{(-k)^{k-1} z^k}{k!} \end{aligned} \quad (142)$$

where  $W(a, z) \forall a \in \mathbb{Z}, z \notin \{0, -e^{-1}\}$  is

$$W(a, z) = 1 + (2i\pi a + \ln(z) - 1) \exp\left(\frac{i}{2\pi} \int_0^\infty \frac{1}{x+1} \ln\left(\frac{x + (2a-1)i\pi - \ln(x) + \ln(z)}{x + (2a+1)i\pi - \ln(x) + \ln(z)}\right) dx\right) \quad (143)$$

A generalization of (142) is solved by

$$\{x: xb^x = z\} = \frac{W(\ln(b)z)}{\ln(b)} \quad (144)$$

The W function satisfies several identities

$$\begin{aligned} W(z) e^{W(z)} &= z \\ W(z \ln(z)) &= \ln(z) & \forall z < 1 \\ |W(z)| &= W(|z|) \\ e^{nW(z)} &= z^n W(z)^{-n} \\ \ln(W(n, z)) &= \ln(z) - W(n, z) + 2i\pi n \\ W\left(-\frac{\ln(z)}{z}\right) &= -\ln(z) & \forall z \in [0, e] \\ \frac{W(-\ln(z))}{-\ln(z)} &= z^{z^{z^{\dots}}} \end{aligned} \quad (145)$$

where  $n \in \mathbb{Z}$ . Some special values are

$$\begin{aligned} W(-1, -e^{-1}) &= -1 \\ W(-e^{-1}) &= -1 \\ W(e) &= 1 \\ W(0) &= 0 \\ W(\infty) &= \infty \\ W(-\infty) &= \infty + i\pi \\ W\left(-\frac{\pi}{2}\right) &= \frac{i\pi}{2} \\ W(-\ln(\sqrt{2})) &= -\ln(2) \\ W(-1, -\ln(\sqrt{2})) &= -2\ln(2) \end{aligned} \quad (146)$$

We also have the limit

$$\lim_{a \rightarrow \pm\infty} \frac{W(a, x)}{a} = 2\pi i \quad (147)$$

and differential

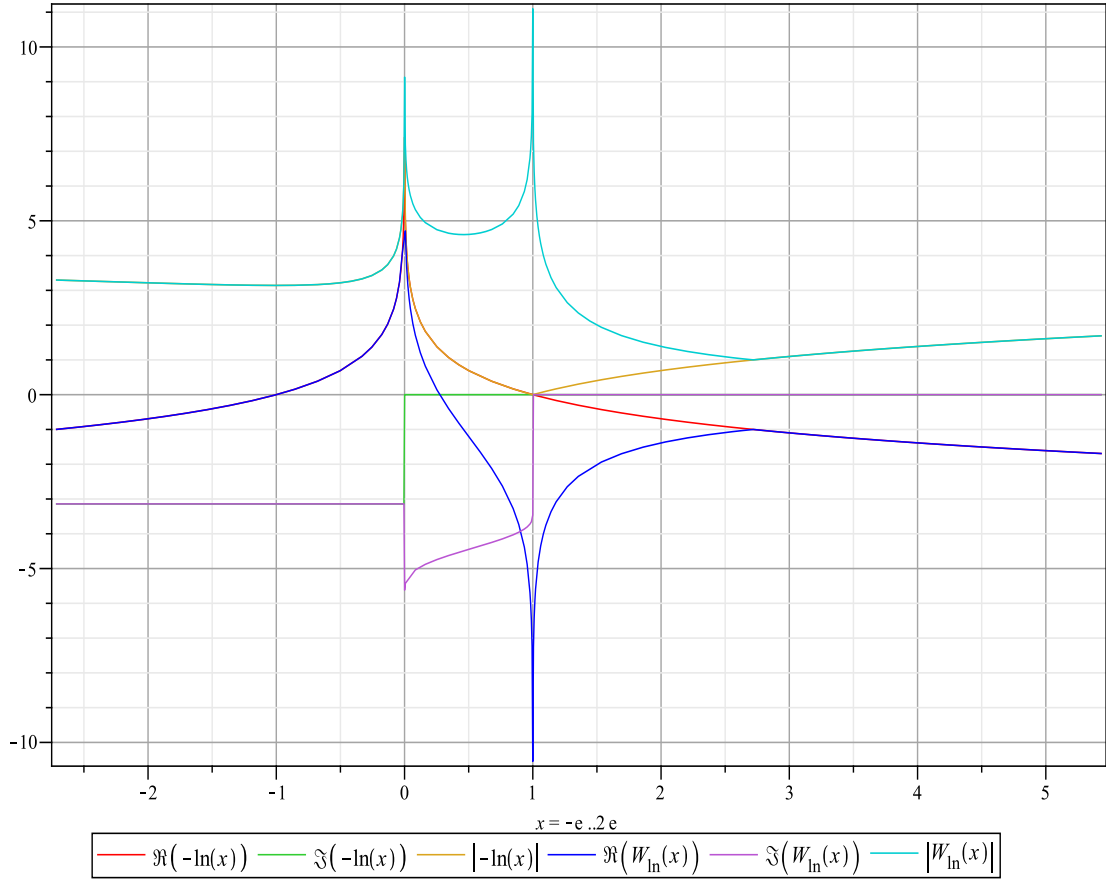
$$\frac{d}{dz} W(a, f(z)) = \frac{W(a, f(z)) \frac{d}{dz} f(z)}{f(z)(1 + W(a, f(z)))} \quad (148)$$

as well as the obvious integral

$$\int_0^1 W\left(-\frac{\ln(x)}{x}\right) dx = \int_0^1 -\ln(x) dx = 1 \quad (149)$$

Let us define, for the sake of brevity, the function

$$\begin{aligned} W_{\ln}(z) &= W\left(-1, -\frac{\ln(z)}{z}\right) \\ &= 1 + \left(\ln\left(-\frac{\ln(z)}{z}\right) - 1 - 2\pi i\right) \exp\left(\frac{i}{2\pi} \int_0^\infty \frac{1}{x+1} \ln\left(\frac{x - 3i\pi - \ln(x) + \ln\left(-\frac{\ln(z)}{z}\right)}{x - i\pi - \ln(x) + \ln\left(-\frac{\ln(z)}{z}\right)}\right) dx\right) \end{aligned} \quad (150)$$



**Figure 20.**  $W\left(-\frac{\ln(x)}{x}\right) = -\ln(x)$  and  $W_{\ln}(x) = W\left(-1, -\frac{\ln(x)}{x}\right)$

Then we have the limits

$$\begin{aligned} \lim_{x \rightarrow -\infty} W_{\ln}(x) &= 0 \\ \lim_{x \rightarrow +\infty} W_{\ln}(x) &= -\infty \end{aligned} \quad (151)$$

and

$$\operatorname{Im}(W_{\ln}(x)) = \begin{cases} -\pi & -\infty < x < 0 \\ \dots & 0 \leq x \leq 1 \\ 0 & 1 < x < \infty \end{cases} \quad (152)$$

$$W_{\ln}(x) = -\ln(x) \quad \forall x \notin [0, e] \quad (153)$$

The root of  $\operatorname{Re}(W_{\ln}(x))$  is given by

$$\begin{aligned} \{x: \operatorname{Re}(W_{\ln}(x)) = 0\} &= \{-x: (x^2)^{\frac{1}{x}} = e^{3\pi}\} \\ &= \frac{2}{3}\pi W\left(\frac{3}{2}\pi\right) \\ &\cong 0.27441063190284810044\dots \end{aligned} \quad (154)$$

where the imaginary part of the value at the root of the real part of  $W_{\ln}(z)$  is

$$\begin{aligned} W_{\ln}\left(\frac{2}{3}\pi W\left(\frac{3}{2}\pi\right)\right) &= W\left(-1, -\frac{\ln\left(\frac{2}{3}\pi W\left(\frac{3}{2}\pi\right)\right)}{\frac{2}{3}\pi W\left(\frac{3}{2}\pi\right)}\right) \\ &= W\left(-1, \frac{3\pi}{2}\right) \\ &= \frac{3\pi i}{2} \\ &\cong i4.712388980384689857\dots \end{aligned} \quad (155)$$