Point Process Models for Multivariate High-Frequency Irregularly Spaced Data

March 17, 2014

Stephen Crowley
scrowley@canaccord.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 Weibullfit 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

1. Defintions 2

1.1. Point Processes and Intensities 2
1.1.1. Stochastic Integrals 3
1.2. The Hawkes Process 3
1.2.1. Linear Self-Exciting Processes 3
1.2.2. The Hawkes(1) Model 6
1.2.3. Maximum Likelihood Estimation 6
1.2.4. The Hawkes Process in Quantum Theory 7
1.3. Predicting When The Next Event Of A Process Will Occur 7
1.3.1. Predicting White Noise Processes 8
1.3.2. The case of P=1 and the Lambert W Function 8
1.3.3. The Case of Any Order 9
1.4. Alternative Kernels for the Hawkes Process 12
1.4.1. A Generalized Exponential Kernel 12
1.4.2. The Hawkes Process Having a Weibull Kernel 13
1.4.3. The Hawkes Process Having a Gamma Kernel 15
1.4.4. A Hyperbolic Kernel 18
1.5. Assessing Goodness of Fit with Graphical Methods 19
1.6. Combining the ACD and Hawkes Models 23
1.6.1. Comparison of Hawkes-Exp vs ACD+Hawkes-Exp Model Fits 24
1.7. Multivariate Hawkes Models 27
1.7.1. The Compensator 28
1.7.2. Log-Likelihood 29
1.7.3. Projection of the Next Occurance Time 30

2. Numerical Methods 31

2.1. The Nelder-Mead Algorithm 31
2.1.1. Starting Points for Optimizing the Hawkes Process of Order 31

3. Examples 31

3.1. Millisecond Resolution Trade Sequences 31
3.1.1. Adjusting for the Deterministic Daily Intensity Variation 31
3.1.2. Univarate Hawkes model fit to SPY (SPDR S&P 500 ETF Trust) 32
3.1.3. Multivariate SPY Data for 2012-08-14 37
3.1.4. Multivariate SPY Data for 2012-11-19 38

Bibliography 39

4. Appendix 40

4.1. The Lambert W Function W(k,x) 40

1. DEFINTIONS

1.1. Point Processes and Intensities.

Consider a dimensional multivariate point process. Let denote the *counting process* associated with the -th point process which is simply the number of events which have occured by time . Let denote the filtration of the pooled process of point processes consisting of the set denoting the history of arrival times of each event type associated with the point processes. At time , the most recent arrival time will be denoted . A process is said to be simple if no points occur at the same time, that is, there are no zero-length durations. The right-continuous(cadlag) counting process[8, 4.1.1.2] can be represented as a sum of Heaviside step functions

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

The counting function(process) jumps at the occurrence of each point and its value is the number of points occuring up to the point in time of the jump, inclusively. The left-continuous counting function does not include the time of the most recent jump, it counts the number of events occuring *before* t and is denoted by

$$\breve{N}_t^k = \sum_{t_i^k < t} \theta(t - t_i^k) \tag{2}$$

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

$$\lambda^k(t|F_t) = \lim_{\Delta t \to 0} \frac{\Pr\left(N_{t+\Delta t}^k - N_t^k > 0|F_t\right)}{\Delta t} \tag{3}$$

For small values of we have

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

so that

$$E((N_{t+\Delta t}^k - N_t^k) - \lambda^k(t|F_t)\Delta t) = o(\Delta t) \tag{5}$$

and (5) will be uncorrelated with the past of as . Next consider

$$\begin{aligned}
&\lim_{\Delta t\to 0}\sum_{j=1}^{\frac{(s_1-s_0)}{\Delta t}}\left(N^k_{s_0+j\Delta t}-N^k_{s_0+(j-1)\Delta t}\right)-\lambda^k(s_0+j\Delta t|F_t)\Delta t\\
&=\lim_{\Delta t\to 0}(N^k_{s_0}-N^k_{s_1})-\sum_{j=1}^{\frac{(s_1-s_0)}{\Delta t}}\lambda^k(j\Delta t|F_t)\Delta t\\
&=(N^k_{s_0}-N^k_{s_1})-\int_{s_0}^{s_1}\lambda^k(t|F_t)\mathrm{d}t
\end{aligned}\tag{6}$$

which will be uncorrelated with , that is

$$E\left(\int_{s_0}^{s_1}\lambda^k(t|F_t)\mathrm{d}t\right)=N^k_{s_0}-N^k_{s_1}\tag{7}$$

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

$$\Lambda^k(s_0,s_1)=\int_{s_0}^{s_1}\lambda^k(t|F_t)\mathrm{d}t\tag{8}$$

Let denote the time interval, or duration, between the -th and -th arrival times. The *-conditional survivor function* for the -th process is given by

$$S_k(x^k_i)=P_k(t^k_i>x^k_i|F_{t_{i-1}+\tau})\tag{9}$$

Let

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

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

$$S^k(x^k_i)=e^{-\int_{t_{i-1}}^{t_i}\lambda^k(t|F_t)\mathrm{d}t}=e^{-\tilde{\mathcal{E}}^k_i}\tag{10}$$

and is an i.i.d. exponential random variable with unit mean and variance. Since the random variable

$$\mathcal{E}^k_{N(t)}=1-\tilde{\mathcal{E}}_{N(t)}\tag{11}$$

has zero mean and unit variance. Positive values of indicate that the path of conditional intensity function under-predicted the number of events in the time interval and negative values of indicate that over-predicted the number of events in the interval. In this way, (9) can be interpreated as a generalized residual. The *backwards recurrence time* given by

$$U^{(k)}(t)=t-t_{N^k(t)}\tag{12}$$

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

1.1.1. Stochastic Integrals.

The *stochastic Stieltjes integral*[1, 2.1][9, 2.2] of a measurable process, having either locally bounded or nonnegative sample paths, with respect to exists and for each we have

$$\int_{(0,t]}X(s)\mathrm{d}N^k_s=\sum_{i\geqslant 1}\theta(t-t^k_i)X(t^k_i)\tag{13}$$

1.2. The Hawkes Process.

1.2.1. Linear Self-Exciting Processes.

A (univariate) linear self-exciting (counting) process is one that can be expressed as [19][7][18][3][8, 11.3]

$$\begin{aligned}\lambda(t) &= \lambda_0(t)\kappa + \int_{-\infty}^{t} \nu(t-s)\mathrm{d}N_s \\ &= \lambda_0(t)\kappa + \sum_{t_k<t} \nu(t-t_k)\end{aligned} \tag{14}$$

where is a deterministic base intensity, see (159), expresses the positive influence of past events on the current value of the intensity process, and takes the place of the constant in the referenced papers. For comparison with the multivariate case see Equation (125). The (exponential) Hawkes process of order is a linear self-exciting process defined by the exponential kernel

$$\nu(t) = \sum_{j=1}^{P} \alpha_j e^{-\beta_j t} \tag{15}$$

which has the survivor function

$$\begin{aligned}S(t) &= \int_t^{\infty} \nu(s)\mathrm{d}s \\ &= \int_t^{\infty} \sum_{j=1}^{P} \alpha_j e^{-\beta_j s}\mathrm{d}s \\ &= \frac{\sum_{j=1}^{P} \alpha_j e^{-\beta_j t} \prod_{k=1,k\neq j}^{P} \beta_k}{\prod_{j=1}^{P} \beta_j}\end{aligned} \tag{16}$$

where the product denotes that is excluded so that the hazard function is written

$$\begin{aligned}\bar{\nu}(t) &= \frac{\nu(t)}{S(t)} \\ &= \frac{\sum_{j=1}^{P} \alpha_j e^{-\beta_j t}}{\frac{\sum_{j=1}^{P} \alpha_j e^{-\beta_j t} \prod_{k=1,k\neq j}^{P} \beta_k}{\prod_{j=1}^{P} \beta_j}} \\ &= \frac{\sum_{j=1}^{P} \alpha_j e^{-\beta_j t} \prod_{j=1}^{P} \beta_j}{\sum_{j=1}^{P} \alpha_j e^{-\beta_j t} \prod_{k=1,k\neq j}^{P} \beta_k}\end{aligned} \tag{17}$$

so that the "instantaneous" hazard rate is

$$\lim_{t\to 0} \bar{\nu}(t) = \frac{\sum_{j=1}^{P} \alpha_j \prod_{j=1}^{P} \beta_j}{\sum_{j=1}^{P} \alpha_j \prod_{k=1,k\neq j}^{P} \beta_k} \tag{18}$$

The intensity of the exponential Hawkes process is written as

$$\begin{aligned}\lambda(t) &= \lambda_0(t)\kappa + \int_0^t \sum_{j=1}^{P} \alpha_j e^{-\beta_j(t-s)}\mathrm{d}N_s \\ &= \lambda_0(t)\kappa + \sum_{k=0}^{N_t-1} \sum_{j=1}^{P} \alpha_j e^{-\beta_j(t-t_k)} \\ &= \lambda_0(t)\kappa + \sum_{j=1}^{P} \sum_{k=0}^{N_t-1} \alpha_j e^{-\beta_j(t-t_k)} \\ &= \lambda_0(t)\kappa + \sum_{j=1}^{P} \alpha_j \sum_{k=0}^{N_t-1} e^{-\beta_j(t-t_k)} \\ &= \lambda_0(t)\kappa + \sum_{j=1}^{P} \alpha_j B_j(N_t) \qquad \forall t \in t_i\end{aligned} \tag{19}$$

where is given recursively by

$$\begin{aligned} B_j(i) &= \sum_{k=1}^{i-1} e^{-\beta_j(t_i - t_k)} \\ &= e^{-\beta_j(t_i - t_{i-1})} \sum_{k=1}^{i-1} e^{-\beta_j(t_{i-1} - t_k)} \\ &= e^{-\beta_j(t_i - t_{i-1})} \left(1 + \sum_{k=1}^{i-2} e^{-\beta_j(t_{i-1} - t_k)}\right) \\ &= e^{-\beta_j(t - t_{i-1})}(1 + B_j(i-1)) \end{aligned} \tag{20}$$

since . A univariate Hawkes process is stationary if the branching ratio is less than one.

$$\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} < 1 \tag{21}$$

If a Hawkes process is stationary then the unconditional mean is

$$\begin{aligned} \mu = E[\lambda(t)] &= \frac{\lambda_0}{1 - \int_0^\infty \nu(t)\mathrm{d}t} \\ &= \frac{\lambda_0}{1 - \int_0^\infty \sum_{j=1}^{P} \alpha_j e^{-\beta_j t}\mathrm{d}t} \\ &= \frac{\lambda_0}{1 - \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}} \end{aligned} \tag{22}$$

For consecutive events, we have the compensator (8)

$$\begin{aligned} \Lambda(t_{i-1}, t_i) &= \int_{t_{i-1}}^{t_i} \lambda(t)\mathrm{d}t \\ &= \int_{t_{i-1}}^{t_i} \left(\lambda_0(t) + \sum_{j=1}^{P} \alpha_j B_j(N_t)\right)\mathrm{d}t \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \int_{t_{i-1}}^{t_i} \sum_{j=1}^{P} \alpha_j \sum_{k=0}^{i-1} e^{-\beta_j(t - t_k)}\mathrm{d}t \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \sum_{j=1}^{P} \alpha_j \sum_{k=0}^{i-1} \int_{t_{i-1}}^{t_i} e^{-\beta_j(t - t_k)}\mathrm{d}t \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \sum_{k=0}^{i-1} \int_{t_{i-1}}^{t_i} \nu(t - t_k)\mathrm{d}t \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \sum_{k=0}^{i-1} \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\left(e^{-\beta_j(t_{i-1} - t_k)} - e^{-\beta_j(t_i - t_k)}\right) \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\left(1 - e^{-\beta_j(t_i - t_{i-1})}\right)A_j(i-1) \end{aligned} \tag{23}$$

compared with the multivariate compensator in Equation (133)where there is the recursion

$$\begin{aligned} A_j(i) &= \sum_{t_k \leqslant t_i} e^{-\beta_j(t_i - t_k)} \\ &= \sum_{k=0}^{i-1} e^{-\beta_j(t_i - t_k)} \\ &= 1 + e^{-\beta_j(t_i - t_{i-1})} A_j(i-1) \end{aligned} \tag{24}$$

with since the integral of the exponential kernel (15) is

$$\begin{aligned}\int_{t_{i-1}}^{t_i} \nu(t)\mathrm{d}t \; &= \int_{t_{i-1}}^{t_i} \sum_{j=1}^{P} \alpha_j e^{-\beta_j (t-t_k)}\mathrm{d}t \\ &= \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} (e^{-\beta_j t_i} - e^{-\beta_j t_{i-1}})\end{aligned} \tag{25}$$

If then (23) simplifies to

$$\begin{aligned}\Lambda(t_{i-1}, t_i) \; &= (t_i - t_{i-1})\lambda_0 + \sum_{k=0}^{i-1} \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\left(e^{-\beta_j(t_{i-1}-t_k)} - e^{-\beta_j(t_i-t_k)}\right) \\ &= (t_i - t_{i-1})\lambda_0 + \sum_{k=0}^{i-1} \int_{t_{i-1}-t_k}^{t_i-t_k} \nu(t)\mathrm{d}t \\ &= (t_i - t_{i-1})\lambda_0 + \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\left(1 - e^{-\beta_j(t_i-t_{i-1})}\right)A_j(i-1)\end{aligned} \tag{26}$$

Similiarly, another parametrization is given by

$$\begin{aligned}\Lambda(t_{i-1}, t_i) \; &= \int_{t_{i-1}}^{t_i} \kappa\lambda_0(s)\mathrm{d}s + \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\left(1 - e^{-\beta_j(t_i-t_{i-1})}\right)A_j(i-1) \\ &= \kappa\int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\left(1 - e^{-\beta_j(t_i-t_{i-1})}\right)A_j(i-1) \\ &= \kappa\Lambda_0(t_{i-1}, t_i) + \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\left(1 - e^{-\beta_j(t_i-t_{i-1})}\right)A_j(i-1)\end{aligned} \tag{27}$$

where scales the predetermined baseline intensity . In this parameterization the intensity is also scaled by

$$\lambda(t) \; = \kappa\,\lambda_0(t) + \sum_{j=1}^{P} \alpha_j\,B_j(N_t) \tag{28}$$

this allows to precompute the deterministic part of the compensator .

1.2.2. The Hawkes(1) Model.

The simplest case occurs when the baseline intensity is constant and where we have

$$\lambda(t) = \kappa + \sum_{t_i < t} \alpha e^{-\beta\,(t-t_i)} \tag{29}$$

which has the unconditional mean

$$E[\lambda(t)] = \frac{\kappa}{1 - \frac{\alpha}{\beta}} \tag{30}$$

1.2.3. Maximum Likelihood Estimation.

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

$$\begin{aligned}\ln\mathcal{L}(N(t)_{t\in[0,T]}) \; &= \int_0^T (1 - \lambda(s))\mathrm{d}s + \int_0^T \ln\lambda(s)\mathrm{d}N_s \\ &= T - \int_0^T \lambda(s)\mathrm{d}s + \int_0^T \ln\lambda(s)\mathrm{d}N_s\end{aligned} \tag{31}$$

which in the case of the Hawkes model of order can be explicitly written [14] as

$$\begin{aligned}
\ln \mathcal{L}(\{t_i\}_{i=1\ldots n}) =& T - \Lambda(0,T) + \sum_{i=1}^{n} \ln\lambda(t_i) \\
=& T + \sum_{i=1}^{n} \ln\lambda(t_i) - \Lambda(t_{i-1}, t_i) \\
=& T - \Lambda(0,T) + \sum_{i=1}^{n} \ln\lambda(t_i) \\
=& T - \Lambda(0,T) + \sum_{i=1}^{n} \ln\left(\kappa\lambda_0(t_i) + \sum_{j=1}^{P}\sum_{k=1}^{i-1} \alpha_j e^{-\beta_j(t_i - t_k)} \right) \\
=& T - \Lambda(0,T) + \sum_{i=1}^{n} \ln\left(\kappa\lambda_0(t_i) + \sum_{j=1}^{P} \alpha_j R_j(i) \right) \\
=& T - \int_0^T \kappa\lambda_0(s)\mathrm{d}s - \sum_{i=1}^{n}\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\left(1 - e^{-\beta_j(t_n - t_i)}\right) \\
&+ \sum_{i=1}^{n} \ln\left(\kappa\lambda_0(t_i) + \sum_{j=1}^{P} \alpha_j R_j(i) \right)
\end{aligned} \tag{32}$$

where and we have the recursion[13]

$$\begin{aligned}
R_j(i) =& \sum_{k=1}^{i-1} e^{-\beta_j(t_i - t_k)} \\
=& e^{-\beta_j(t_i - t_{i-1})}(1 + R_j(i-1))
\end{aligned} \tag{33}$$

If we have constant baseline intensity then the log-likelihood can be written

$$\begin{aligned}
\ln \mathcal{L}(\{t_i\}_{i=1\ldots n}) =& T - \kappa T - \sum_{i=1}^{n}\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\left(1 - e^{-\beta_j(t_n - t_i)}\right) \\
&+ \sum_{i=1}^{n} \ln\left(\lambda_0 + \sum_{j=1}^{P} \alpha_j R_j(i) \right)
\end{aligned} \tag{34}$$

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

1.2.4. 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. [20].

1.3. Predicting When The Next Event Of A Process Will Occur.

The next event arrival time(s) of a point process can be predicted by solving for the unknown(s) in the equation

$$\left\{ t_{n+1} : \varepsilon = \Lambda(t_n, t_{n+1}) = \int_{t_n}^{t_{n+1}} \lambda(s; \mathfrak{F}_s)\mathrm{d}s \right\} \tag{35}$$

where is the filtration up to and including time and the parameters of are fixed. The multivariate case is covered in Section (1.7.3). The idea is to integrate over the solution of Equation (35) with all possible values of , distributed according to the unit exponential distribution. The reason for the plural form, time(s), rather than the singular form, time, is that Equation (35) actually has a single real solution and number of complex solutions, where is the number of points that have occured in the process up until the time of prediction. This set of complex expected future event arrival times is the *constellation* of the process, which changes with the arrival of each event(the increasing -algebra filtration), and has somewhat of a multiverse interpretation if thought about from a physical context.

1.3.1. Predicting White Noise Processes.

[10, Preface] was written to be read. :)

1.3.2. The case of P=1 and the Lambert W Function.

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

$$a(n) = \sum_{k=0}^{n} e^{\beta_1(t_k+2t_n)} \tag{36}$$

$$b(n) = \sum_{k=0}^{n} e^{\beta_1(t_k+t_n)} \tag{37}$$

$$c(n) = \sum_{k=0}^{n} e^{\beta_1 t_k} \tag{38}$$

$$d(n) = \sum_{k=0}^{n} \sum_{l=0}^{n} e^{\beta_1(t_l+t_k+t_n)} \tag{39}$$

then

$$\Lambda^{-1}(\varepsilon; t_0 \ldots t_n) = \begin{array}{l} \dfrac{a(n)t_n z}{b(n)} + \\ \dfrac{a(n)zW\left(\dfrac{\alpha_1 b(n) e^{-\frac{\left((2\kappa t_n+\varepsilon)\beta_1 e^{\beta_1 t_n} - \alpha_1 c(n)\right)z}{\kappa}}}{\kappa}\right)}{\beta_1 b(n)} + \\ \dfrac{z}{\kappa b(n)}\left(a(n)\varepsilon - \dfrac{\alpha_1 d(n)}{\beta_1}\right) \end{array} \tag{40}$$

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

$$\int_0^\infty \Lambda^{-1}(\varepsilon; t_0 \ldots t_n) e^{-\varepsilon} \mathrm{d}\varepsilon = \int_0^\infty \left(\frac{a(n)t_n e^{-\beta_1 t_n}}{b(n)} + \frac{a(n)e^{-\beta_1 t_n} W\left(\frac{1}{\kappa}\alpha_1 b(n) e^{\left(-\frac{1}{\kappa}((2\kappa t_n+\varepsilon)\beta_1 e^{\beta_1 t_n} - \alpha_1 c(n))e^{-\beta_1 t_n}\right)}\right)}{\beta_1 b(n)} + \frac{1}{\kappa}\left(\frac{a(n)e^{-\beta_1 t_n}\varepsilon}{b(n)} - \frac{e^{-\beta_1 t_n}\alpha_1 d(n)}{\beta_1 b(n)}\right)\right) e^{-\varepsilon} \mathrm{d}\varepsilon \tag{41}$$

which, as far as I am aware, must be calculated with numeric methods which are actually far more effecient than Monte Carlo sampling as other studies have suggested. Equation (40) suffers from numerical problems due to massive sums of exponentials. We can rescale the equations and drop one of them, let

$$\hat{a}(n) = \frac{a(n)}{e^{\beta_1 2t_n}} = \sum_{k=0}^{n} e^{\beta_1 t_k} \tag{42}$$

$$\hat{b}(n) = \frac{b(n)}{e^{\beta_1 2t_n}} = \sum_{k=0}^{n} e^{\beta_1(t_k-t_n)} \tag{43}$$

$$\hat{d}(n) = \frac{d(n)}{e^{\beta_1 2t_n}} = \sum_{k=0}^{n} \sum_{l=0}^{n} e^{\beta_1(t_l+t_k-t_n)} \tag{44}$$

then Equation (40) can be re-written as

$$\Lambda^{-1}(\varepsilon;t_0\ldots t_n)=\begin{array}{l}\dfrac{\hat{a}(n)t_n z}{\hat{b}(n)}+\\ \dfrac{\hat{a}(n)\,zW\left(\dfrac{\alpha_1\hat{b}(n)e^{\frac{\alpha_1\hat{b}(n)-\beta_1\varepsilon}{\kappa}}}{\kappa}\right)}{\beta_1\hat{b}(n)}+\\ \dfrac{z}{\kappa\hat{b}(n)}\left(\hat{a}(n)\varepsilon-\dfrac{\alpha_1\hat{d}(n)}{\beta_1}\right)\end{array} \tag{45}$$

We have the following recursions with initial conditions and

$$\hat{b}(n)=\hat{b}(n-1)e^{\beta_1(t_{n-1}-t_n)}+1 \tag{46}$$

$$\begin{aligned}\hat{d}(n)\ &=\hat{d}(n-1)e^{\beta_1(t_{n-1}-t_n)}+e^{\beta_1 t_n}+2\sum_{k=0}^{n-1}e^{\beta_1 t_k}\\ &=\hat{d}(n-1)e^{\beta_1(t_{n-1}-t_n)}+e^{\beta_1 t_n}+2\,\hat{a}(n-1)\end{aligned} \tag{47}$$

It would be nice to have expressions like this involving the Lambert W function for but neither Maple nor Mathematica were able to find any solutions in terms of “known” functions for . It is noted that Equation (41) has the form

$$\int_0^\infty (p+qW(re^{-sx+t})+ux)e^{-x}\mathrm{d}x \tag{48}$$

which is a function of 6 variables, , for which it would be a very nice thing to have a closed form expression, in order to avoid a recourse to numerical or Monte Carlo integration. It seems that such an expression is very likely to exist because if we drop the variable from Equation (48) we get a closed-form expression of the form

$$\int_0^\infty (p+qW(re^{-x+t})+ux)e^{-x}\mathrm{d}x=qW(re^t)+\frac{q}{W(re^t)}-q+u+p-\frac{q}{re^t} \tag{49}$$

We can break this problem down into a more manageable one by calculating some more integrals to see if we can find a pattern. Let us begin with the integral

$$\int_0^\infty W(e^{-sx})e^{-x}\mathrm{d}x=W(1)+\left(-\frac{1}{s}\right)^{-\frac{1}{s}}\left(\Gamma\left(\frac{1}{s}\right)-s\Gamma\left(1+\frac{1}{s},-\frac{W(1)}{s}\right)\right) \tag{50}$$

whose closed-form expresssion was found by Vladimir Reshetnikov. [22]

1.3.3. The Case of Any Order .

For general values of the order , the equation whose root is to be sought is given by the expression

$$\varphi_P(x(\varepsilon))=\begin{array}{l}\left(\prod_{k=1}^{P}\beta_k\right)(\kappa x-(\varepsilon+\kappa T))e^{\sum_{k=1}^{P}\beta_k(x+T)}+\ldots\\ \ldots+\sum_{m=1}^{P}\left(\prod_{k=1}^{P}\left\{\begin{array}{ll}\alpha_k & m=k\\ \beta_k & m\neq k\end{array}\right.\right)\sum_{k=0}^{n}e^{\sum_{j=1}^{P}\beta_j\left(x+\left\{\begin{smallmatrix}T & j\neq m\\ t_k & j=m\end{smallmatrix}\right.\right)}-e^{\sum_{j=1}^{P}\beta_j\left(T+\left\{\begin{smallmatrix}x & j\neq m\\ t_k & j=m\end{smallmatrix}\right.\right)}\end{array} \tag{51}$$

where is arrival time of the most recent point and it is noted that the product of piecewise functions can be written as

$$\begin{aligned}\prod_{k=1}^{P}\left\{\begin{array}{ll}\alpha_k & m=k\\ \beta_k & m\neq k\end{array}\right. &= \alpha_m\left(\prod_{k=1}^{m-1}\beta_k\right)\left(\prod_{k=m+1}^{P}\beta_k\right)\\ &= \alpha_m\prod_{\substack{k=1\\k\neq m}}^{P}\beta_k\end{aligned} \tag{52}$$

and the sums likewise

$$\begin{aligned}\sum_{j=1}^{P}\beta_j\left(x+\left\{\begin{array}{ll}T & j\neq m\\ t_k & j=m\end{array}\right.\right) &= \beta_m(x+t_k)+\sum_{j=1}^{m-1}\beta_j(x+T)+\sum_{j=m+1}^{P}\beta_j(x+T)\\ &= \beta_m(x+t_k)+\sum_{\substack{j=1\\j\neq m}}^{P}\beta_j(x+T)\\ &= \sigma_{m,k}(x,x)\end{aligned} \tag{53}$$

and

$$\begin{aligned}\sum_{j=1}^{P}\beta_j\left(T+\left\{\begin{array}{ll}x & j\neq m\\ t_k & j=m\end{array}\right.\right) &= \beta_m(T+t_k)+\sum_{j=1}^{m-1}\beta_j(x+T)+\sum_{j=m+1}^{P}\beta_j(x+T)\\ &= \beta_m(T+t_k)+\sum_{\substack{j=1\\j\neq m}}^{P}\beta_j(x+T)\\ &= \sigma_{m,k}(x,T)\end{aligned} \tag{54}$$

so that (51) can be rewritten as

$$\varphi_P(x(\varepsilon)) = \tau(x,\varepsilon)+\sum_{j=1}^{P}\phi_j\sum_{k=0}^{N_T}\left(\sigma_{j,k}(x,x)-\sigma_{j,k}(x,T)\right) \tag{55}$$

to be compared with the multivariate case in Equation (148), where

$$\sigma_{m,k}(x,a) = e^{(a+t_k)\beta_m+(x+T)\sum_{\substack{j=1\\j\neq m}}^{P}\beta_j} \tag{56}$$

$$\phi_m = \alpha_m\prod_{\substack{k=1\\k\neq m}}^{P}\beta_k = \prod_{k=1}^{P}\left\{\begin{array}{ll}\alpha_k & k=m\\ \beta_k & k\neq m\end{array}\right. \tag{57}$$

$$\tau(x,\varepsilon) = ((x-T)\kappa-\varepsilon)\upsilon\eta(x) \tag{58}$$

$$\eta(x) = e^{(x+T)\sum_{k=1}^{P}\beta_k} \tag{59}$$

$$\upsilon = \prod_{k=1}^{P}\beta_k \tag{60}$$

The derivative given by

$$\varphi_P'(x(\varepsilon)) = \upsilon(\kappa\eta(x)+\tau(x,\varepsilon))+\sum_{m=1}^{P}\phi_m\sum_{k=0}^{n}\left(\mu\sigma_{m,k}(x)-\mu_m\sigma_{m,k}(T)\right) \tag{61}$$

where

$$\mu = \sum_{k=1}^{P}\beta_k \tag{62}$$

$$\mu_m = \sum_{\substack{j=1\\j\neq m}}^{P}\beta_j \tag{63}$$

is needed so that the Newton iteration can be written

$$
\begin{aligned}
x_{i+1} &= x_i - \frac{\varphi_2(x_i)}{\varphi_2'(x_i)} \\
&= x_i - \frac{\tau(x_i,\varepsilon) + \sum_{m=1}^{P} \phi_m \sum_{k=0}^{n} \left(\sigma_{m,k}(x_i,x_i) - \sigma_{m,k}(x_i,T)\right)}{\upsilon(\kappa\eta(x_i) + \tau(x_i,\varepsilon)) + \sum_{m=1}^{P} \phi_m \sum_{k=0}^{n} \left(\mu\,\sigma_{m,k}(x_i) - \mu_m\,\sigma_{m,k}(T)\right)}
\end{aligned}
\tag{64}
$$

and simplified a bit(at least notationally) if we let

$$
\rho(x,d) = \sum_{m=1}^{P} \phi_m \sum_{k=0}^{n} \left(\sigma_{m,k}(x) \begin{cases} 1 & d=0 \\ \mu & d=1 \end{cases} - \sigma_{m,k}(T) \begin{cases} 1 & d=0 \\ \mu_m & d=1 \end{cases} \right) \tag{65}
$$

then

$$
\begin{aligned}
x_{i+1}(\varepsilon) &= x_i(\varepsilon) - \frac{\varphi_P(x_i(\varepsilon))}{\varphi_P'(x_i(\varepsilon))} \\
&= x_i - \frac{\tau(x_i(\varepsilon),\varepsilon) + \rho(x_i(\varepsilon),0)}{\upsilon(\kappa\eta(x_i(\varepsilon)) + \tau(x_i(\varepsilon),\varepsilon)) + \rho(x_i(\varepsilon),1)}
\end{aligned}
\tag{66}
$$

so that

$$
\Lambda_P^{-1}(\varepsilon;t_0\ldots T) = \lim_{m\to\infty} x_m(\varepsilon) \tag{67}
$$

which leads to the expression for the expected arrival time of the next point

$$
\int_0^\infty \Lambda_P^{-1}(\varepsilon;t_0\ldots T)e^{-\varepsilon}\mathrm{d}\varepsilon = \int_0^\infty \lim_{m\to\infty} x_m(\varepsilon)e^{-\varepsilon}\mathrm{d}\varepsilon \tag{68}
$$

I have a hunch that Fatou's lemma can be invoked so that the order of the limit and the integral in Equation (68) can be exchanged, with perhaps the introduction of another function, which of course would greatly reduce the computational complexity of the equation. The graph of the Newton iteration for a arbitrary chosen set of parameters and times is shown below.

Figure 1. Graph of the Newton iteration with and having parameters is shown below.

We also have the limit

$$\begin{aligned}\lim_{x\to\infty}\frac{\varphi_P(x_i(\varepsilon))}{\varphi_P'(x_i(\varepsilon))} &= \lim_{x\to\infty}\frac{\tau(x_i(\varepsilon),\varepsilon)+\rho(x_i(\varepsilon),0)}{\upsilon(\kappa\eta(x_i(\varepsilon))+\tau(x_i(\varepsilon),\varepsilon))+\rho(x_i(\varepsilon),1)}\\ &=\frac{1}{\mu}\end{aligned}\tag{69}$$

1.4. Alternative Kernels for the Hawkes Process.

1.4.1. A Generalized Exponential Kernel.

In the paper of [13] there is made a reference to [15], which appears to not be available online. Nevertheless, [13] contains the necessary recursive formulas. Here, the expontial kernel changes from (15)

$$\nu(t)=\alpha e^{-\beta t}\tag{70}$$

to

$$\nu(t) = \sum_{j=0}^{P}\alpha_j t^j e^{-\beta t}\tag{71}$$

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

$$\begin{aligned}\lambda(t) &= \lambda_0(t)\kappa+\int_0^t\sum_{j=0}^{P}\alpha_j t^j e^{-\beta(t-s)}\mathrm{d}N_s\\ &= \lambda_0(t)\kappa+\sum_{k=0}^{N_t-1}\sum_{j=0}^{P}\alpha_j t^j e^{-\beta(t-t_k)}\\ &= \lambda_0(t)\kappa+\sum_{j=0}^{P}\sum_{k=0}^{N_t-1}\alpha_j t^j e^{-\beta(t-t_k)}\\ &= \lambda_0(t)\kappa+\sum_{j=0}^{P}\alpha_j t^j\sum_{k=0}^{N_t-1}e^{-\beta(t-t_k)}\\ \lambda(t_i) &= \lambda_0(t)\kappa+\sum_{j=0}^{P}\alpha_j R_j(i)\end{aligned}\tag{72}$$

Note that does not have a subscript in this parametrization. For consecutive events, we have the compensator (8)

$$\begin{aligned}\Lambda(t_{i-1},t_i) &= \int_{t_{i-1}}^{t_i}\lambda(t)\mathrm{d}t\\ &= \int_{t_{i-1}}^{t_i}\lambda_0(s)\mathrm{d}s+\sum_{k=0}^{i-1}\int_{t_{i-1}}^{t_i}\nu(t-t_k)\mathrm{d}t\\ &= \int_{t_{i-1}}^{t_i}\lambda_0(t)\kappa+\sum_{j=0}^{P}\alpha_j t^j\sum_{k=0}^{N_t-1}e^{-\beta(t-t_k)}\mathrm{d}t\\ &= \int_{t_{i-1}}^{t_i}\lambda_0(s)\mathrm{d}s+\int_{t_{i-1}}^{t_i}\sum_{j=0}^{P}\alpha_j t^j\sum_{k=0}^{N_t-1}e^{-\beta(t-t_k)}\mathrm{d}t\\ &= \int_{t_{i-1}}^{t_i}\lambda_0(s)\mathrm{d}s+\sum_{k=0}^{i-1}\hat{\nu}(t_i-t_k)-\hat{\nu}(t_{i-1}-t_k)\\ &= \int_{t_{i-1}}^{t_i}\lambda_0(s)\mathrm{d}s+\sum_{k=0}^{i-1}\end{aligned}\tag{73}$$

where is the antiderivative of the kernel

$$\begin{aligned}\hat{\nu}(t) &= \int \nu(t)\mathrm{d}t\\ &= \int \sum_{j=0}^{P} \alpha_j t^j e^{-\beta t}\mathrm{d}t\\ &= \frac{e^{-\beta t}}{\beta^{P+1}}\sum_{k=0}^{P} \alpha_k \sum_{j=0}^{k} -\beta^{P-(k-j)}(k-j)!\binom{k}{k-j}t^j\\ &= \frac{e^{-\beta t}}{\beta^{P+1}}\sum_{k=1}^{P} \frac{\alpha_k \beta^{-k}(\beta t)^k e^{\beta t}(2\beta^{P+1}-\beta^{P+2})\Gamma(k+1,\beta t)}{(\beta t)^{k+1}(\beta-2)}\\ &= -\frac{1}{\beta}\sum_{k=0}^{P} \alpha_k \beta^{-k}\Gamma(k+1,\beta t)\end{aligned} \tag{74}$$

where is the incomplete Gamma function (104). The log-likelihood for the Hawkes process having this generalized exponential kernel is given by

$$\begin{aligned}\ln\mathcal{L}(\{t_i\}_{i=1\ldots n}) &= \sum_{i=1}^{n} \ln\left(\lambda_0(t_i)\kappa + \sum_{j=0}^{P} \alpha_j R_j(i)\right) - \kappa\int_0^T \lambda_0(t)\mathrm{d}t - \sum_{i=1}^{n}\sum_{j=0}^{P} \alpha_j S_j(T-t_i)\\ &= \sum_{i=1}^{n} \ln\left(\lambda_0(t_i)\kappa + \sum_{j=0}^{P} \alpha_j R_j(i)\right) - \sum_{i=1}^{n} \kappa\int_{t_{i-1}}^{t_i} \lambda_0(t)\mathrm{d}t + \sum_{j=0}^{P} \alpha_j S_j(T-t_i)\\ &= \sum_{i=1}^{n} \ln\left(\lambda_0(t_i)\kappa + \sum_{j=0}^{P} \alpha_j R_j(i)\right) - \left(\kappa\int_{t_{i-1}}^{t_i} \lambda_0(t)\mathrm{d}t + \sum_{j=0}^{P} \alpha_j S_j(T-t_i)\right)\end{aligned} \tag{75}$$

where and are given recursively and is the time of the last point in the sample. Let

$$t_0 = 0 \tag{76}$$

$$R_0(1) = 0 \tag{77}$$

$$S_0(t) = \frac{1-e^{-\beta t}}{\beta} \tag{78}$$

then

$$A_k(t) = t^k e^{-\beta t} \tag{79}$$

$$\begin{aligned}R_j(i) &= A_j(t_i - t_{i-1}) + \sum_{k=0}^{j} \binom{j}{k} A_{j-k}(t_i - t_{i-1}) R_k(i-1)\\ &= (t_i - t_{i-1})^j e^{-\beta(t_i - t_{i-1})} + \sum_{k=0}^{j} \binom{j}{k}(t_i - t_{i-1})^{j-k} e^{-\beta(t_i - t_{i-1})} R_k(i-1)\end{aligned} \tag{80}$$

and

$$\begin{aligned}S_j(t) &= \frac{j S_{j-1}(t) - A_j(t)}{\beta}\\ &= \frac{j S_{j-1}(t) - t^j e^{-\beta t}}{\beta}\end{aligned} \tag{81}$$

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 (14)

$$\begin{aligned}\lambda(t) &= \lambda_0(t)\kappa + \int_{-\infty}^{t} \psi(t-s)\mathrm{d}N_s\\ &= \lambda_0(t)\kappa + \sum_{t_i<t} \psi(t-t_i)\end{aligned} \tag{82}$$

where the exponential kernel is replaced by the Weibull kernel

$$\psi(t) = \sum_{j=1}^{P} \alpha_j \left(\frac{\upsilon_j}{\omega_j}\right)\left(\frac{t}{\omega_j}\right)^{\upsilon_j - 1} e^{-\left(\frac{t}{\omega_j}\right)^{\upsilon_j}} \tag{83}$$

with parameter vectors so the Hawkes-Weibull intensity is written as

$$\begin{aligned}
\lambda(t) &= \lambda_0(t)\kappa + \int_0^t \sum_{j=1}^{P} \alpha_j \left(\frac{\upsilon_j}{\omega_j}\right)\left(\frac{t-s}{\omega_j}\right)^{\upsilon_j - 1} e^{-\left(\frac{t-s}{\omega_j}\right)^{\upsilon_j}} \mathrm{d}N_s \\
&= \lambda_0(t)\kappa + \sum_{k=0}^{N_t-1} \sum_{j=1}^{P} \alpha_j \left(\frac{\kappa_j}{\omega_j}\right)\left(\frac{t-t_k}{\omega_j}\right)^{\upsilon_j - 1} e^{-\left(\frac{t-t_k}{\omega_j}\right)^{\upsilon_j}} \\
&= \lambda_0(t)\kappa + \sum_{j=1}^{P} \sum_{k=0}^{N_t-1} \alpha_j \left(\frac{\upsilon_j}{\omega_j}\right)\left(\frac{t-t_k}{\omega_j}\right)^{\upsilon_j - 1} e^{-\left(\frac{t-t_k}{\omega_j}\right)^{\upsilon_j}} \\
&= \lambda_0(t)\kappa + \sum_{j=1}^{P} \alpha_j \left(\frac{\upsilon_j}{\omega_j}\right) \sum_{k=0}^{N_t-1} \left(\frac{t-t_k}{\omega_j}\right)^{\upsilon_j - 1} e^{-\left(\frac{t-t_k}{\omega_j}\right)^{\upsilon_j}} \\
&= \lambda_0(t)\kappa + \sum_{j=1}^{P} \alpha_j \left(\frac{\upsilon_j}{\omega_j}\right) C_j(N_t)
\end{aligned} \tag{84}$$

and is given by

$$C_j(n) = \sum_{k=0}^{n-1} \left(\frac{t-t_k}{\omega_j}\right)^{\upsilon_j - 1} e^{-\left(\frac{t-t_k}{\omega_j}\right)^{\upsilon_j}} \tag{85}$$

where the branching ratio is

$$\begin{aligned}
\int_0^\infty \psi(t)\mathrm{d}t &= \int_0^\infty \sum_{j=1}^{P} \alpha_j \left(\frac{\upsilon_j}{\omega_j}\right)\left(\frac{t}{\omega_j}\right)^{\upsilon_j - 1} e^{-\left(\frac{t}{\omega_j}\right)^{\upsilon_j}} \mathrm{d}t \\
&= \sum_{j=1}^{P} \alpha_j
\end{aligned} \tag{86}$$

The survivor function of the Weibull kernel is given by

$$\begin{aligned}
S(t) &= \int_t^\infty \nu(s)\mathrm{d}s \\
&= \int_t^\infty \sum_{j=1}^{P} \alpha_j \left(\frac{\upsilon_j}{\omega_j}\right)\left(\frac{s}{\omega_j}\right)^{\upsilon_j - 1} e^{-\left(\frac{s}{\omega_j}\right)^{\upsilon_j}} \mathrm{d}s \\
&= \sum_{j=1}^{P} \alpha_j e^{-t^{\upsilon_j}\omega_j^{-\upsilon_j}}
\end{aligned} \tag{87}$$

which is surprisingly quite a bit less complicated than survivor function of the exponential kernel (16). The hazard function is the quotient of the kernel over the survivor function

$$\begin{aligned}
\bar{\psi}(t) &= \frac{\psi(t)}{S(t)} \\
&= \frac{\sum_{j=1}^{P} \alpha_j \left(\frac{\upsilon_j}{\omega_j}\right)\left(\frac{t}{\omega_j}\right)^{\upsilon_j - 1} e^{-\left(\frac{t}{\omega_j}\right)^{\upsilon_j}}}{\sum_{j=1}^{P} \alpha_j e^{-t^{\upsilon_j}\omega_j^{-\upsilon_j}}}
\end{aligned} \tag{88}$$

Now, similiar to (23),the compensator is calculated

$$\begin{aligned}\tilde{\mathcal{E}}_i =&\Lambda^k(t_{i-1},t_i)\\ =&\int_{t_{i-1}}^{t_i}\lambda(t)\mathrm{d}t\\ =&\kappa\int_{t_{i-1}}^{t_i}\lambda_0(t)\mathrm{d}t+\int_{t_{i-1}}^{t_i}\sum_{k=0}^{i-1}\psi(t-t_k)\mathrm{d}t\\ =&\kappa\int_{t_{i-1}}^{t_i}\lambda_0(t)\mathrm{d}t+\sum_{k=0}^{i-1}\psi(t_i-t_k)-\psi(t_{i-1}-t_k)\\ =&\kappa\int_{t_{i-1}}^{t_i}\lambda_0(t)\mathrm{d}t+\int_{t_{i-1}}^{t_i}\sum_{k=0}^{i-1}\sum_{j=1}^{P}\alpha_j\left(\frac{\upsilon_j}{\omega_j}\right)\left(\frac{t-t_k}{\omega_j}\right)^{\upsilon_j-1}e^{-\left(\frac{t-t_k}{\omega_j}\right)^{\upsilon_j}}\mathrm{d}t\\ =&\int_{t_{i-1}}^{t_i}\lambda_0(t)\kappa+\int_{t_{i-1}}^{t_i}\sum_{j=1}^{P}\alpha_j\left(\frac{\upsilon_j}{\omega_j}\right)\sum_{k=0}^{i-1}\left(\frac{t-t_k}{\omega_j}\right)^{\upsilon_j-1}e^{-\left(\frac{t-t_k}{\omega_j}\right)^{\upsilon_j}}\mathrm{d}t\\ =&\kappa\int_{t_{i-1}}^{t_i}\lambda_0(t)\mathrm{d}t+\sum_{k=0}^{i-1}\sum_{j=1}^{P}-\alpha_j\left(e^{-\left(\frac{t_i-t_k}{\omega_j}\right)^{\nu_j}}-e^{-\left(\frac{t_{i-1}-t_k}{\omega_j}\right)^{\nu_j}}\right)\end{aligned} \tag{89}$$

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

$$\begin{aligned}\int_{t_{i-1}}^{t_i}\psi(t)\mathrm{d}t =&\int_{t_{i-1}}^{t_i}\sum_{j=1}^{P}\alpha_j\left(\frac{\upsilon_j}{\omega_j}\right)\left(\frac{t}{\omega_j}\right)^{\upsilon_j-1}e^{-\left(\frac{t}{\omega_j}\right)^{\upsilon_j}}\mathrm{d}t\\ =&\hat{\psi}(t_i)-\hat{\psi}(t_{i-1})\\ =&\sum_{j=1}^{P}-\alpha_j\left(e^{-\left(\frac{t_i}{\omega_j}\right)^{\nu_j}}-e^{-\left(\frac{t_{i-1}}{\omega_j}\right)^{\nu_j}}\right)\end{aligned} \tag{90}$$

where is the antiderivative of the kernel given by

$$\begin{aligned}\hat{\psi}(t) =&\int\psi(t)\mathrm{d}t\\ =&\int\sum_{j=1}^{P}\alpha_j\left(\frac{\upsilon_j}{\omega_j}\right)\left(\frac{t}{\omega_j}\right)^{\upsilon_j-1}e^{-\left(\frac{t}{\omega_j}\right)^{\upsilon_j}}\mathrm{d}t\\ =&\sum_{j=1}^{P}-\alpha_j e^{-\left(\frac{t}{\omega_j}\right)^{\nu_j}}\end{aligned} \tag{91}$$

The change-of-variables in (89) can be made, let

$$s=-\left(\frac{t-t_k}{\omega_j}\right)^{\upsilon_j} \tag{92}$$

then

$$\mathrm{d}s_k=-\mathrm{d}s\left(\frac{\upsilon_j}{\omega_j}\right)\left(\frac{t-t_k}{\omega_j}\right)^{\upsilon_j-1} \tag{93}$$

so (89) can be written as

$$\begin{aligned}\tilde{\mathcal{E}}_i =&\Lambda^k(t_{i-1},t_i)\\ =&\kappa\int_{t_{i-1}}^{t_i}\lambda_0(t)\mathrm{d}t+\sum_{j=1}^{P}\sum_{k=0}^{i-1}\alpha_j\int_{\left(\frac{t_{i-1}-t_k}{\omega_j}\right)^{\upsilon_j}}^{\left(\frac{t_i-t_k}{\omega_j}\right)^{\upsilon_j}}e^{-s}\mathrm{d}s\end{aligned} \tag{94}$$

When 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

$$\begin{aligned}\lambda(t) &= \lambda_0(t)\kappa + \int_{-\infty}^{t} \nu(t-s)\mathrm{d}N_s \\ &= \lambda_0(t)\kappa + \sum_{t_k<t} \nu(t-t_k)\end{aligned} \tag{95}$$

with

$$\nu(t) = \sum_{j=1}^{P} \frac{\alpha_j}{b_j\,\Gamma(c_j)} \left(\frac{t}{b_j}\right)^{c_j-1} e^{-\frac{t}{b_j}} \tag{96}$$

having parameter vectors so that the intensity is written

$$\begin{aligned}\lambda(t) &= \lambda_0(t)\kappa + \int_0^t \sum_{j=1}^{P} \frac{\alpha_j}{b_j\,\Gamma(c_j)} \left(\frac{t-s}{b_j}\right)^{c_j-1} e^{-\frac{t-s}{b_j}} \mathrm{d}N_s \\ &= \lambda_0(t)\kappa + \sum_{i=0}^{N_t-1} \sum_{j=1}^{P} \frac{\alpha_j}{b_j\,\Gamma(c_j)} \left(\frac{t-t_i}{b_j}\right)^{c_j-1} e^{-\frac{t-t_i}{b_j}} \\ &= \lambda_0(t)\kappa + \sum_{j=1}^{P} \sum_{i=0}^{N_t-1} \frac{\alpha_j}{b_j\,\Gamma(c_j)} \left(\frac{t-t_i}{b_j}\right)^{c_j-1} e^{-\frac{t-t_i}{b_j}} \\ &= \lambda_0(t)\kappa + \sum_{j=1}^{P} \frac{\alpha_j}{b_j\Gamma(c_j)} \sum_{i=0}^{N_t-1} \left(\frac{t-t_i}{b_j}\right)^{c_j-1} e^{-\frac{t-t_i}{b_j}}\end{aligned} \tag{97}$$

where the branching ratio is

$$\begin{aligned}\int_0^{\infty} \nu(t)\mathrm{d}t &= \int_0^{\infty} \sum_{j=1}^{P} \frac{\alpha_j}{b_j\,\Gamma(c_j)} \left(\frac{t}{b_j}\right)^{c_j-1} e^{-\frac{t}{b_j}} \mathrm{d}t \\ &= \sum_{j=1}^{P} \alpha_j\end{aligned} \tag{98}$$

The compensator is given by

$$\begin{aligned}\tilde{\mathcal{E}}_i &= \Lambda^k(t_{i-1}, t_i) \\ &= \int_{t_{i-1}}^{t_i} \lambda(t)\mathrm{d}t \\ &= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t)\mathrm{d}t + \sum_{k=0}^{i-1} \int_{t_{i-1}}^{t_i} \nu(t-t_k)\mathrm{d}t \\ &= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t)\mathrm{d}t + \int_{t_{i-1}}^{t_i} \sum_{k=0}^{i-1} \nu(t-t_k)\mathrm{d}t \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(t)\kappa + \sum_{j=1}^{P} \frac{\alpha_j}{b_j\Gamma(c_j)} \sum_{k=0}^{i-1} \left(\frac{t-t_k}{b_j}\right)^{c_j-1} e^{-\frac{t-t_k}{b_j}} \mathrm{d}t \\ &= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t)\mathrm{d}t + \int_{t_{i-1}}^{t_i} \sum_{j=1}^{P} \frac{\alpha_j}{b_j\Gamma(c_j)} \sum_{k=0}^{i-1} \left(\frac{t-t_k}{b_j}\right)^{c_j-1} e^{-\frac{t-t_k}{b_j}} \mathrm{d}t \\ &= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t)\mathrm{d}t + \sum_{k=0}^{i-1} \int_{t_{i-1}}^{t_i} \sum_{j=1}^{P} \frac{\alpha_j}{b_j\Gamma(c_j)} \left(\frac{t-t_k}{b_j}\right)^{c_j-1} e^{-\frac{t-t_k}{b_j}} \mathrm{d}t \\ &= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t)\mathrm{d}t + \sum_{k=0}^{i-1} \hat{\nu}(t_i-t_k) - \hat{\nu}(t_{i-1}-t_k) \\ &= \kappa \int_{t_{i-1}}^{t_i} \lambda_0(t)\mathrm{d}t + \sum_{k=0}^{i-1} \sum_{j=1}^{P} -\alpha_j \frac{(c_j+1)}{\Gamma(c_j+2)} \cdots \\ &\cdots \left(\Gamma\left(c_j+1, \frac{(t_i-t_k)}{b_j}\right) - \Gamma\left(c_j+1, \frac{(t_{i-1}-t_k)}{b_j}\right) - e^{-\frac{(t_i-t_k)}{b_j}} (t_i-t_k)^{c_j} b_j^{-c_j} + e^{-\frac{(t_{i-1}-t_k)}{b_j}} \left(t_{i-1}-t_k\right)^{c_j} b_j^{-c_j}\right)\end{aligned} \tag{99}$$

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

$$\begin{aligned}
\int_{t_{i-1}}^{t_i} \nu(t)\mathrm{d}t &= \int_{t_{i-1}}^{t_i} \sum_{j=1}^{P} \frac{\alpha_j}{b_j \Gamma(c_j)} \sum_{i=0}^{N_t-1} \left(\frac{t-t_i}{b_j}\right)^{c_j-1} e^{-\frac{t-t_i}{b_j}} \mathrm{d}t \\
&= \hat{\nu}(t_i) - \hat{\nu}(t_{i-1}) \\
&= \sum_{j=1}^{P} -\frac{\alpha_j (c_j+1)}{\Gamma(c_j+2)} \left(\Gamma\left(c_j+1, \frac{t_i}{b_j}\right) - \Gamma\left(c_j+1, \frac{t_{i-1}}{b_j}\right) - e^{-\frac{t_i}{b_j}} t_i^{c_j} b_j^{-c_j} + e^{-\frac{t_{i-1}}{b_j}} t_{i-1}^{c_j} b_j^{-c_j}\right)
\end{aligned} \tag{100}$$

where is the integral of given by

$$\begin{aligned}
\hat{\nu}(t) &= \int \nu(t)\mathrm{d}t \\
&= \int \sum_{j=1}^{P} \frac{\alpha_j}{b_j \, \Gamma(c_j)} \left(\frac{t}{b_j}\right)^{c_j-1} e^{-\frac{t}{b_j}} \mathrm{d}t \\
&= \sum_{j=1}^{P} \frac{\alpha_j b_j^{c_j} e^{-\frac{t}{2b_j}} b_j^{-c_j}}{\Gamma(c_j+2)} \left(\left(\frac{t}{b_j}\right)^{\frac{c_j}{2}} M_{\frac{c_j}{2}, \frac{c_j}{2}+\frac{1}{2}}\left(\frac{t}{b_j}\right) + \left(\frac{t}{b_j}\right)^{c_j} e^{-\frac{t}{2b_j}} (c_j+1)\right) \\
&= \sum_{j=1}^{P} \frac{\alpha_j}{\Gamma(c_j+2)} \left(\Gamma(c_j+2) + \left(\left(\frac{t}{b_j}\right)^{c_j} e^{-\frac{t}{b_j}} - \Gamma\left(c_j+1, \frac{t}{b_j}\right)\right)(c_j+1)\right)
\end{aligned} \tag{101}$$

where

$$M_{\mu,\nu}(z) = e^{-\frac{z}{2}} z^{\nu+\frac{1}{2}} {}_1F_1\left(\begin{matrix} \frac{1}{2}-\mu+\nu \\ 1+2\nu \end{matrix}; z\right) \tag{102}$$

is the Whittaker M function which solves the equation and

$${}_1F_1\left(\begin{matrix} a \\ b \end{matrix}; z\right) = \sum_{k=0}^{\infty} \frac{\Gamma(a+k)\Gamma(b) z^k}{\Gamma(a)\Gamma(k+1)\Gamma(b+k)} \tag{103}$$

is the confluent hypergeometric function[16] and

$$\Gamma(a,z) = \Gamma(a) - \frac{z^a \, {}_1F_1\left(\begin{matrix} a \\ 1+a \end{matrix}; -z\right)}{a} \tag{104}$$

is the incomplete Gamma function. When and 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

$$\begin{aligned}\lambda(t) &= \lambda_0(t)\kappa + \int_{\infty}^{t} \nu(t-s)\mathrm{d}N_s \\ &= \lambda_0(t)\kappa + \sum_{t_k<t} \nu(t-t_k)\end{aligned} \tag{105}$$

with the hyperbolic kernel

$$\nu(t) = \sum_{j=1}^{P} \frac{\alpha_j}{(t+\beta_j)^{p_j}} \tag{106}$$

having parameter vectors so that the intensity is written

$$\begin{aligned}\lambda(t) &= \lambda_0(t)\kappa + \int_0^t \sum_{j=1}^{P} \frac{\alpha_j}{(t+\beta_j)^{p_j}}\mathrm{d}N_s \\ &= \lambda_0(t)\kappa + \sum_{i=0}^{N_t-1}\sum_{j=1}^{P} \frac{\alpha_j}{((t-t_i)+\beta_j)^{p_j}} \\ &= \lambda_0(t)\kappa + \sum_{i=0}^{N_t-1}\sum_{j=1}^{P} \alpha_j((t-t_i)+\beta_j)^{-p_j} \\ &= \lambda_0(t)\kappa + \sum_{j=1}^{P}\alpha_j\sum_{i=0}^{N_t-1} ((t-t_i)+\beta_j)^{-p_j}\end{aligned} \tag{107}$$

The compensator is given by

$$\begin{aligned}\tilde{\mathcal{E}}_i &= \Lambda^k(t_{i-1},t_i) \\ &= \int_{t_{i-1}}^{t_i} \lambda(t)\mathrm{d}t \\ &= \kappa\int_{t_{i-1}}^{t_i} \lambda_0(t)\mathrm{d}t + \sum_{k=0}^{i-1}\int_{t_{i-1}}^{t_i} \nu(t-t_k)\mathrm{d}t \\ &= \kappa\int_{t_{i-1}}^{t_i} \lambda_0(t)\mathrm{d}t + \int_{t_{i-1}}^{t_i}\sum_{k=0}^{i-1} \nu(t-t_k)\mathrm{d}t \\ &= \kappa\int_{t_{i-1}}^{t_i} \lambda_0(t)\mathrm{d}t + \sum_{k=0}^{i-1} \hat{\nu}(t_i-t_k) - \hat{\nu}(t_{i-1}-t_k) \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(t)\kappa + \sum_{j=1}^{P} h_j\sum_{k=0}^{i-1} ((t-t_k)+\kappa_j)^{-p_j}\mathrm{d}t \\ &= \kappa\int_{t_{i-1}}^{t_i} \lambda_0(t)\mathrm{d}t + \sum_{k=0}^{i-1}\int_{t_{i-1}}^{t_i}\sum_{j=1}^{P} \alpha_j((t-t_k)+\beta_j)^{-p_j}\mathrm{d}t \\ &= \kappa\int_{t_{i-1}}^{t_i} \lambda_0(t)\mathrm{d}t + \int_{t_{i-1}}^{t_i}\sum_{j=1}^{P}\alpha_j\sum_{k=0}^{i-1} ((t-t_k)+\beta_j)^{-p_j}\mathrm{d}t \\ &= \kappa\int_{t_{i-1}}^{t_i} \lambda_0(t)\mathrm{d}t + \sum_{k=0}^{i-1} -\sum_{j=1}^{P} \frac{\alpha_j((\beta_j+(t_i-t_k))^{1-p_j} - (\beta_j+(t_{i-1}-t_k))^{1-p_j})}{p_j-1}\end{aligned} \tag{108}$$

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

$$\begin{aligned}\int_{t_{i-1}}^{t_i} \nu(t)\mathrm{d}t &= \int_{t_{i-1}}^{t_i} \sum_{j=1}^{P} \alpha_j (t+\beta_j)^{-p_j}\mathrm{d}t \\ &= \hat{\nu}(t_i) - \hat{\nu}(t_{i-1}) \\ &= -\sum_{j=1}^{P} \frac{\alpha_j((\beta_j+t_i)^{1-p_j} - (\beta_j+t_{i-1})^{1-p_j})}{p_j-1}\end{aligned} \tag{109}$$

The antiderivative of the kernel is

$$\hat{\nu}(t) = \int \nu(t)\mathrm{d}t = \sum_{j=1}^{P} \frac{\alpha_j(\beta_j+t)^{1-p_j}}{1-p_j} \tag{110}$$

and definite integral

$$\int_0^s \nu(t)\mathrm{d}t = \sum_{j=1}^{P} -\frac{\alpha_j(\beta_j+s)^{-p_j}\left(\beta_j - \beta_j\left(\frac{\beta_j+s}{\beta_j}\right)^{p_j} + s\right)}{p_j-1} \tag{111}$$

The branching ratio is given by

$$\int_0^\infty \nu(t)\mathrm{d}t = \begin{cases} \dfrac{\sum_{i=1}^{P} \beta_i^{1-p_i}\alpha_i\left(1-\prod_{\substack{k=1\\k\neq i}}^{P} p_k - \sum_{\substack{k=1\\k\neq i}}^{P} p_k\right)}{\prod_{j=1}^{P}(p_j-1)} & \forall p_j > 1 \\ \infty & \exists p_j \leqslant 1 \end{cases} \tag{112}$$

where 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

$$\prod_{\substack{k=1\\k\neq i}}^{P} p_k = \prod_{k=1}^{i-1} p_k \prod_{k=i+1}^{P} p_k \tag{113}$$

and likewise for the sum

$$\sum_{\substack{k=1\\k\neq i}}^{P} p_k = \sum_{k=1}^{i-1} p_k + \sum_{k=i+1}^{P} p_k \tag{114}$$

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 (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 and the Weibull kernel of order . 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.

Table 1.

As can be seen, the rather meager increase of the log-likehood 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

$$\lambda(t) = \lambda_0(t) + \frac{1}{\psi_{N_t}} + \int_0^t \nu(t-s)\mathrm{d}N_s \tag{115}$$

where is the determinstic baseline intensity(159) and where the ACD(21) part is

$$\psi_i = \omega + \sum_{j=1}^{m} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} \tag{116}$$

and the Hawkes part has the exponential kernel(15)

$$\nu(t) = \sum_{j=1}^{P} \gamma_j e^{-\varphi_j t} \tag{117}$$

so that

$$\begin{aligned}\int_0^t \nu(t-s)\mathrm{d}N_s &= \int_0^t \sum_{j=1}^{P} \gamma_j e^{-\varphi_j (t-s)}\mathrm{d}N_s \\ &= \sum_{k=0}^{N_t} \nu(t-t_k) \\ &= \sum_{k=0}^{N_t} \sum_{j=1}^{P} \gamma_j e^{-\varphi_j(t-t_k)} \\ &= \sum_{j=1}^{P} \gamma_j \sum_{k=0}^{N_t} e^{-\varphi_j(t-t_k)} \\ &= \sum_{j=1}^{P} \gamma_j B_j(N_t)\end{aligned} \tag{118}$$

where we have replaced and in the Hawkes part so that the parameter names do not conflict with the ACD part where and are also used as parameter names. The Hawkes part of the intensity has a recursive structure similiar to that of the compensator. Let

$$\begin{aligned}B_j(i) &= \sum_{k=0}^{i-1} e^{-\varphi_j(t-t_k)} \\ &= (1 + B_j(i-1))e^{-\varphi_j(t-t_i)}\end{aligned} \tag{119}$$

where . Then we have

$$\lambda(t) = \lambda_0(t) + \frac{1}{\omega + \sum_{j=1}^{m} \alpha_j x_{N_t - j} + \sum_{j=1}^{q} \beta_j \psi_{N_t - j}} + \sum_{j=1}^{P} \gamma_j B_j(N_t) \tag{120}$$

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

$$\begin{aligned}
\ln \mathcal{L}(\{t_i\}_{i=1,..,n}) &= \sum_{i=1}^{n} \left(\ln \lambda(t_i) - \int_{t_{i-1}}^{t_i} \lambda(t) \mathrm{d}t \right) \\
&= \sum_{i=1}^{n} \left(\ln \lambda(t_i) - \Lambda(t_{i-1}, t_i) \right) \\
&= \sum_{i=1}^{n} \left(\ln \lambda(t_i) - \tilde{\mathcal{E}}_i \right)
\end{aligned} \tag{121}$$

By direct calculation, combining (20) and (23), and letting we have the compensator

$$\begin{aligned}
\tilde{\mathcal{E}}_i &= \Lambda(t_{i-1}, t_i) \\
&= \int_{t_{i-1}}^{t_i} \lambda(t) \mathrm{d}t \\
&= \int_{t_{i-1}}^{t_i} \left(\lambda_0(t) + \frac{1}{\psi_{N_t+1}} + \int_0^t \nu(t-s) \mathrm{d}N_s \right) \mathrm{d}t \\
&= \frac{x_i}{\psi_i} + \int_{t_{i-1}}^{t_i} \left(\lambda_0(t) + \int_0^t \nu(t-s) \mathrm{d}N_s \right) \mathrm{d}t \\
&= \int_{t_{i-1}}^{t_i} \lambda_0(t) \mathrm{d}t + \frac{x_i}{\psi_i} + \sum_{k=0}^{i-1} \sum_{j=1}^{P} \frac{\gamma_j}{\varphi_j} \left(e^{-\varphi_j (t_{i-1} - t_k)} - e^{-\varphi_j (t_i - t_k)} \right) \\
&= \int_{t_{i-1}}^{t_i} \lambda_0(t) \mathrm{d}t + \frac{x_i}{\psi_i} + \sum_{j=1}^{P} \frac{\gamma_j}{\varphi_j} (1 - e^{-\varphi_j x_i}) A_j(i-1)
\end{aligned} \tag{122}$$

where is defined by (116) and

$$A_j(i) = 1 + e^{-\varphi_j x_i} A_j(i-1) \tag{123}$$

is given by (24) so that (121) can be wriitten as

$$\begin{aligned}
\ln \mathcal{L}(\{t_i\}_{i=0,..,n}) &= \sum_{i=1}^{n} (\ln \lambda(t_i) - \tilde{\mathcal{E}}_i) \\
&= \sum_{i=1}^{n} \left(\ln \lambda(t_i) - \left(\frac{x_i}{\psi_i} + \sum_{j=1}^{P} \frac{\gamma_j}{\varphi_j} (1 - e^{-\varphi_j x_i}) A_j(i-1) \right) \right) \\
&= \sum_{i=1}^{n} \ln \left(\frac{1}{\psi_i} + \sum_{k=0}^{i-1} \sum_{j=1}^{P} \gamma_j e^{-\varphi_j (t_i - t_k)} \right) - \left(\frac{x_i}{\psi_i} + \sum_{j=1}^{P} \frac{\gamma_j}{\varphi_j} (1 - e^{-\varphi_j x_i}) A_j(i-1) \right) \\
&= \sum_{i=1}^{n} \ln \left(\frac{1}{\psi_i} + \sum_{j=1}^{P} \gamma_j B_j(i) \right) - \left(\frac{x_i}{\psi_i} + \sum_{j=1}^{P} \frac{\gamma_j}{\varphi_j} (1 - e^{-\varphi_j x_i}) A_j(i-1) \right)
\end{aligned} \tag{124}$$

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 .

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 and be an -dimensional point process. The associated counting process will be denoted . A multivariate Hawkes process[7][6][11], compared with the univariate case in Equation (14), is defined with intensities given by

$$\begin{aligned}
\lambda^m(t) &= \lambda_0^m(t)\kappa^m + \sum_{n=1}^{M} \int_0^t \sum_{j=1}^{P} \alpha_j^{m,n} e^{-\beta_j^{m,n}(t-s)} \mathrm{d}N_s^n \\
&= \lambda_0^m(t)\kappa^m + \sum_{n=1}^{M} \sum_{j=1}^{P} \sum_{t_k^n<t} \alpha_j^{m,n} e^{-\beta_j^{m,n}(t-t_k^n)} \\
&= \lambda_0^m(t)\kappa^m + \sum_{n=1}^{M} \sum_{j=1}^{P} \alpha_j^{m,n} \sum_{t_k^n<t} e^{-\beta_j^{m,n}(t-t_k^n)} \\
&= \lambda_0^m(t)\kappa^m + \sum_{n=1}^{M} \sum_{j=1}^{P} \alpha_j^{m,n} \sum_{t_k^n<t} e^{-\beta_j^{m,n}(t-t_k^n)} \\
&= \lambda_0^m(t)\kappa^m + \sum_{n=1}^{M} \sum_{j=1}^{P} \alpha_j^{m,n} \sum_{k=0}^{N_t^n-1} e^{-\beta_j^{m,n}(t-t_k^n)} \\
&= \lambda_0^m(t)\kappa^m + \sum_{n=1}^{M} \sum_{j=1}^{P} \alpha_j^{m,n} B_j^{m,n}(N_t^n)
\end{aligned} \tag{125}$$

where in this parameterization is a vector which scales the baseline intensities, in this case, specified by piecewise polynomial splines (159). We can write recursively

$$\begin{aligned}
B_j^{m,n}(i) &= \sum_{k=0}^{i-1} e^{-\beta_j^{m,n}(t-t_k^n)} \\
&= (1+B_j^{m,n}(i-1))e^{-\beta_j^{m,n}(t-t_i^n)}
\end{aligned} \tag{126}$$

In the simplest version with and constant we have

$$\begin{aligned}
\lambda^m(t) &= \kappa^m + \sum_{n=1}^{M} \int_0^t \alpha^{m,n} e^{-\beta^{m,n}(t-s)} \mathrm{d}N_s^n \\
&= \kappa^m + \sum_{n=1}^{M} \sum_{k=0}^{N_t^n-1} \alpha^{m,n} e^{-\beta^{m,n}(t-t_k^n)} \\
&= \kappa^m + \sum_{n=1}^{M} \alpha^{m,n} \sum_{k=0}^{N_t^n-1} e^{-\beta^{m,n}(t-t_k^n)} \\
&= \kappa^m + \sum_{n=1}^{M} \alpha^{m,n} B_1^{m,n}(N_t^n)
\end{aligned} \tag{127}$$

Rewriting (127) in vectorial notion, we have

$$\lambda(t) = \kappa + \int_0^t G(t-s)\mathrm{d}N_s \tag{128}$$

where

$$G(t) = (\alpha^{m,n}\, e^{-\beta^{m,n}(t-s)})_{m,n=1\ldots M} \tag{129}$$

Assuming stationarity gives a constant vector and thus

$$\begin{aligned}\mu &= \frac{\kappa}{I - \int_0^\infty G(u)\mathrm{d}u}\\ &= \frac{\kappa}{I - (\frac{\alpha^{m,n}}{\beta^{m,n}})}\\ &= \frac{\kappa}{I - \Gamma}\end{aligned} \tag{130}$$

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

$$\Gamma = \int_0^\infty G(s)\mathrm{d}s = \frac{\alpha^{m,n}}{\beta^{m,n}} \tag{131}$$

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

$$\rho(G) = \max_{a \in \mathcal{S}(G)} |a| \tag{132}$$

where denotes the set of eigenvalues of .

1.7.1. The Compensator.

The compensator of the -th coordinate of a multivariate Hawkes process between two consecutive events and of type , compared with the univariate case in Equation (23), is given by

$$\begin{aligned}\Lambda^m(t_{i-1}^m, t_i^m) &= \int_{t_{i-1}^m}^{t_i^m} \lambda^m(s)\mathrm{d}s + \int_{t_{i-1}^m}^{t_i^m} \sum_{n=1}^{M} \sum_{t_k^n < t_i^m} \sum_{j=1}^{P} \alpha_j^{m,n} e^{-\beta_j^{m,n}(s - t_k^n)}\mathrm{d}s\\ &= \int_{t_{i-1}^m}^{t_i^m} \lambda^m(s)\mathrm{d}s + \int_{t_{i-1}^m}^{t_i^m} \sum_{n=1}^{M} \sum_{k=0}^{\check{N}_{t_i^m}^n} \sum_{j=1}^{P} \alpha_j^{m,n} e^{-\beta_j^{m,n}(s - t_k^n)}\mathrm{d}s\\ &= \int_{t_{i-1}^m}^{t_i^m} \lambda^m(s)\mathrm{d}s\\ &+ \sum_{n=1}^{M} \sum_{t_k^n < t_{i-1}^m} \sum_{j=1}^{P} \frac{\alpha_j^{m,n}}{\beta_j^{m,n}} [e^{-\beta_j^{m,n}(t_{i-1}^m - t_k^n)} - e^{-\beta_j^{m,n}(t_i^m - t_k^n)}]\\ &+ \sum_{n=1}^{M} \sum_{t_{i-1}^m \leqslant t_k^n < t_i^m} \sum_{j=1}^{P} \frac{\alpha_j^{m,n}}{\beta_j^{m,n}} [1 - e^{-\beta_j^{m,n}(t_i^m - t_k^n)}]\\ &= \int_{t_{i-1}^m}^{t_i^m} \lambda^m(s)\mathrm{d}s\\ &+ \sum_{n=1}^{M} \sum_{k=0}^{\check{N}_{t_{i-1}^m}^n - 1} \sum_{j=1}^{P} \frac{\alpha_j^{m,n}}{\beta_j^{m,n}} [e^{-\beta_j^{m,n}(t_{i-1}^m - t_k^n)} - e^{-\beta_j^{m,n}(t_i^m - t_k^n)}]\\ &+ \sum_{n=1}^{M} \sum_{k=\check{N}_{t_{i-1}^m}^n}^{\check{N}_{t_i^m}^n - 1} \sum_{j=1}^{P} \frac{\alpha_j^{m,n}}{\beta_j^{m,n}} [1 - e^{-\beta_j^{m,n}(t_i^m - t_k^n)}]\end{aligned} \tag{133}$$

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

$$\begin{aligned}A_j^{m,n}(i) &= \sum_{t_k^n < t_i^m} e^{-\beta_j^{m,n}(t_i^m - t_k^n)}\\ &= e^{-\beta_j^{m,n}(t_i^m - t_{i-1}^m)} A_j^{m,n}(i-1) + \sum_{t_{i-1}^m \leqslant t_k^n < t_i^m} e^{-\beta_j^{m,n}(t_i^m - t_k^n)}\end{aligned} \tag{134}$$

and rewrite (133) as

$$\begin{aligned}\Lambda^m(t_{i-1}^m, t_i^m) =& \kappa^m \int_{t_{i-1}^m}^{t_i^m} \lambda_0^m(s)\mathrm{d}s + \int_{t_{i-1}^m}^{t_i^m} \sum_{n=1}^{M}\sum_{j=1}^{P}\sum_{t_k^n<s} \alpha_j^{m,n} e^{-\beta_j^{m,n}(s-t_k^n)}\mathrm{d}s \\ =& \kappa^m \int_{t_{i-1}^m}^{t_i^m} \lambda_0^m(s)\mathrm{d}s \\ &+ \sum_{n=1}^{M}\sum_{j=1}^{P} \frac{\alpha_j^{m,n}}{\beta_j^{m,n}}\left[(1-e^{-\beta_j^{m,n}(t_i^m-t_{i-1}^m)})\times A_j^{m,n}(i-1) + \sum_{t_{i-1}^m\leqslant t_k^n<t_i^m}(1-e^{-\beta_j^{m,n}(t_i^m-t_k^n)})\right] \\ =& \kappa^m \int_{t_{i-1}^m}^{t_i^m} \lambda_0^m(s)\mathrm{d}s \\ &+ \sum_{n=1}^{M}\sum_{j=1}^{P} \frac{\alpha_j^{m,n}}{\beta_j^{m,n}}\left[(1-e^{-\beta_j^{m,n}(t_i^m-t_{i-1}^m)})\times\left(\sum_{t_k^n<t_{i-1}^m} e^{-\beta_j^{m,n}(t_{i-1}^m-t_k^n)}\right) + \sum_{t_{i-1}^m\leqslant t_k^n<t_i^m}(1-e^{-\beta_j^{m,n}(t_i^m-t_k^n)})\right]\end{aligned} \tag{135}$$

where we have the initial conditions .

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

$$\ln \mathcal{L}(\{t_i\}_{i=1,\ldots,N_T}) = \sum_{m=1}^{M} \ln \mathcal{L}^m(\{t_i\}) \tag{136}$$

where each term is defined by

$$\ln\mathcal{L}^m(\{t_i\}) = \int_0^T (1-\lambda^m(s))\mathrm{d}s + \int_0^T \ln\lambda^m(s)\mathrm{d}N_s^m \tag{137}$$

which in this case can be written as

$$\begin{aligned}\ln \mathcal{L}^m(\{t_i\}) =& T - \Lambda^m(0,T) + \sum_{i=1}^{N_T} z_i^m \ln\left(\lambda_0^m(t_i)\kappa^m + \sum_{n=1}^{M}\sum_{j=1}^{P}\sum_{t_k^n<t_i} \alpha_j^{m,n} e^{-\beta_j^{m,n}(t_i-t_k^n)}\right) \\ =& T - \Lambda^m(0,T) + \sum_{i=1}^{N_T^m} \ln\left(\lambda_0^m(t_i^m)\kappa^m + \sum_{n=1}^{M}\sum_{j=1}^{P}\sum_{t_k^n<t_i^m} \alpha_j^{m,n} e^{-\beta_j^{m,n}(t_i^m-t_k^n)}\right)\end{aligned} \tag{138}$$

where again and

$$z_i^m = \begin{cases} 1 & \text{event } t_i \text{ of type } m \\ 0 & \text{otherwise}\end{cases} \tag{139}$$

and

$$\Lambda^m(0,T) = \int_0^T \lambda^m(t)\mathrm{d}t = \sum_{i=1}^{N_T^m} \Lambda^m(t_{i-1}^m, t_i^m) \tag{140}$$

where is given by (135). Similiar to to the one-dimensional case, we have the recursion

$$\begin{aligned}R_j^{m,n}(i) =& \sum_{t_k^n<t_j^m} e^{-\beta_j^{m,n}(t_i^m-t_k^n)} \\ =& \begin{cases} e^{-\beta_j^{m,n}(t_i^m-t_{i-1}^m)} R_j^{m,n}(i-1) + \sum_{t_{i-1}^m\leqslant t_k^n<t_i^m} e^{-\beta_j^{m,n}(t_i^m-t_k^n)} & \text{if } m\neq n \\ e^{-\beta_j^{m,n}(t_i^m-t_{i-1}^m)}(1+R_j^{m,n}(i-1)) & \text{if } m=n\end{cases}\end{aligned} \tag{141}$$

so that (138) can be rewritten as

$$\begin{aligned}\ln \mathcal{L}^m(\{t_i\}) \; &= T - \kappa^m \int_0^T \lambda_0^m(t) \mathrm{d}t - \ldots \\ &\ldots - \sum_{i=1}^{N_T^m} \sum_{n=1}^{M} \sum_{j=1}^{P} \frac{\alpha_j^{m,n}}{\beta_j^{m,n}} \left[(1 - e^{-\beta_j^{m,n}(t_i^m - t_{i-1}^m)}) \times A_j^{m,n}(i-1) + \sum_{t_{i-1}^m \leqslant t_k^n < t_i^m} (1 - e^{-\beta_j^{m,n}(t_i^m - t_k^n)}) \right] + \ldots \\ &\ldots + \sum_{i=1}^{N_T^m} \ln \left(\lambda_0^m(t_i^m)\kappa^m + \sum_{n=1}^{M} \sum_{j=1}^{P} \alpha_j^{m,n} R_j^{m,n}(i) \right)\end{aligned} \tag{142}$$

with initial conditions and where where is the number of observations, is the number of dimensions, and is the order of the model. Again, can be dropped from the equation for the purposes of optimization. According to

1.7.3. Projection of the Next Occurance Time .

The next occurance time of the n-th dimension of a multivariate Hawkes process having the usual exponential kernel can be predicted in the same way as the univariate process in Section (1.3), by solving for the unknown in the equation

$$\left\{ t_{n+1}^m \colon \varepsilon = \Lambda^m(t_n^m, t_{n+1}^m) = \int_{t_n^m}^{t_{n+1}^m} \lambda^m(s; \mathfrak{F}_s) \mathrm{d}s \right\} \tag{143}$$

where is the compensator from Equation (133) and is the filtration up to time and the parameters of are fixed. As is the case for the univariate Hawkes process, the idea is to average over all possible realizations of (of which there are an uncountable infinity) weighted according to an exponential unit distribution. Another idea for more accurate prediction is to model the deviation of the generalized residuals from a true exponential distribution and then include the predicted error when calculating this expectation.

Let the most recent arrival time of the pooled and -th processes respectively be given by

$$T = \max\left(T_m \colon m = 1 \ldots M\right) \tag{144}$$

$$T_m = \max\left(t_n^m \colon n = 0 \ldots N^m - 1\right) = t_{N^m - 1}^m \tag{145}$$

and

$$\breve{N}_{T_m}^n = \sum_{k=0}^{\breve{N}^n} \left\{ \begin{array}{ll} 1 & t_k^n < T_m \\ 0 & \end{array} \right. \tag{146}$$

count the number of points occuring in the -th dimension before the most recent point of the -th dimension and

$$\breve{N}(t_j^m < t_k^n) \tag{147}$$

then the next arrival time for a given value of the exponential random variable of the -th dimension of a multivariate Hawkes process having the standard exponential kernel is found by solving for the real root of

$$\varphi_m(x(\varepsilon); \mathcal{F}_T) = \tau_m(x, \varepsilon) + \sum_{l=1}^{P} \sum_{i=1}^{M} \phi_{m,i,l} \sum_{k=0}^{\breve{N}_{T_m}^i} (\sigma_{m,i,l,k}(x, x) - \sigma_{m,i,l,k}(x, T_m)) \tag{148}$$

which is similiar to the univariate case

$$\varphi_P(x(\varepsilon)) = \; \tau(x, \varepsilon) + \sum_{j=1}^{P} \phi_j \sum_{k=0}^{\breve{N}_T} (\sigma_{j,k}(x, x) - \sigma_{j,k}(x, T)) \tag{149}$$

where

$$\mathcal{F}_T = \{\kappa_{\ldots}, \alpha_{\ldots}, \beta_{\ldots}, t_0^1 \ldots t_{N^1}^1 \leqslant T, \ldots, t_0^m \ldots t_{N^m}^m \leqslant T, \ldots, t_0^M \ldots t_{N^M}^M \leqslant T\} \tag{150}$$

is the filtration up to time , to be interpreted as the set of available information, here denoting fitted parameters and observed arrival times of all dimensions, and where

$$\tau_m(x,\varepsilon) = ((x - T_m)\kappa_m - \varepsilon)\upsilon_m \eta_m(x) \tag{151}$$

$$\eta_m(x) = e^{(x+T_m)\sum_{j=1}^{P}\sum_{n=1}^{M}\beta_{m,n,j}} \tag{152}$$

can be seen to be similiar to the univariate equations and and

$$\upsilon_m = \prod_{j=1}^{P}\prod_{n=1}^{M}\beta_{m,n,j} \tag{153}$$

$$\phi_{m,p,k} = \prod_{j=1}^{P}\prod_{n=1}^{M}\begin{cases}\alpha_{m,n,j} & n = p \text{ and } j = k\\ \beta_{m,n,j} & n \neq p \text{ or } j \neq k\end{cases} \tag{154}$$

$$\sigma_{m,i,l,k}(x,a) = e^{\sum_{j=1}^{P}\sum_{n=1}^{M}\beta_{m,n,j}\begin{cases} a+t_k^n & n=i \text{ and } j=l\\ x+T_n & n\neq i \text{ or} j\neq l\end{cases}} \tag{155}$$

For comparison, the univariate case is Equation (55) where

$$\sigma_{m,k}(x,a) = e^{(a+t_k)\beta_m + (x+T)\sum_{\substack{j=1\\ j\neq m}}^{P}\beta_j} = e^{\sum_{j=1}^{P}\beta_j\begin{cases} a+t_k & j=m\\ x+T & j\neq m\end{cases}} \tag{156}$$

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 .

A starting point for the optimization of a Hawkes process of order 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 be the interval between consecutive arrival times as in the ACD model (17). Then set the initial value of to , and . This gives an unconditional mean of 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 so that the unit of time is seconds. Also, trades occuring 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 and respectively since has units of seconds. Let the adjusted durations be defined

$$\tilde{x}_i = \phi(t_i)x_i \tag{157}$$

where is the unadjusted duration and is a (piecewise polynomial) cubic spline with knot points at with values given by

$$P_j = \frac{1}{(N_{t(zj)+w} - N_{t(zj)-w})} \sum_{i=N_{t(zj)-w}}^{N_{t(zj)+w}} \frac{1}{x_i} \text{ for } j = 0 \ldots 13 \tag{158}$$

where is the number of seconds in a half-hour and . 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.

$$\lambda_0(t) = f(t, P_0, \ldots, P_j) \tag{159}$$

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

Figure 12. for SPY on 2012-11-30

3.1.2. Univarate 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 . 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, 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 . Error bars are not provided, but presumably they could be estimated with derivative information. Note that the closer to 0.8882491159065832 and and to the better, since should be exponentially distributed with mean by design and for a Poisson process the mean and variance are equal. The next thing to check is that the series is not autocorrelated.

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

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

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

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

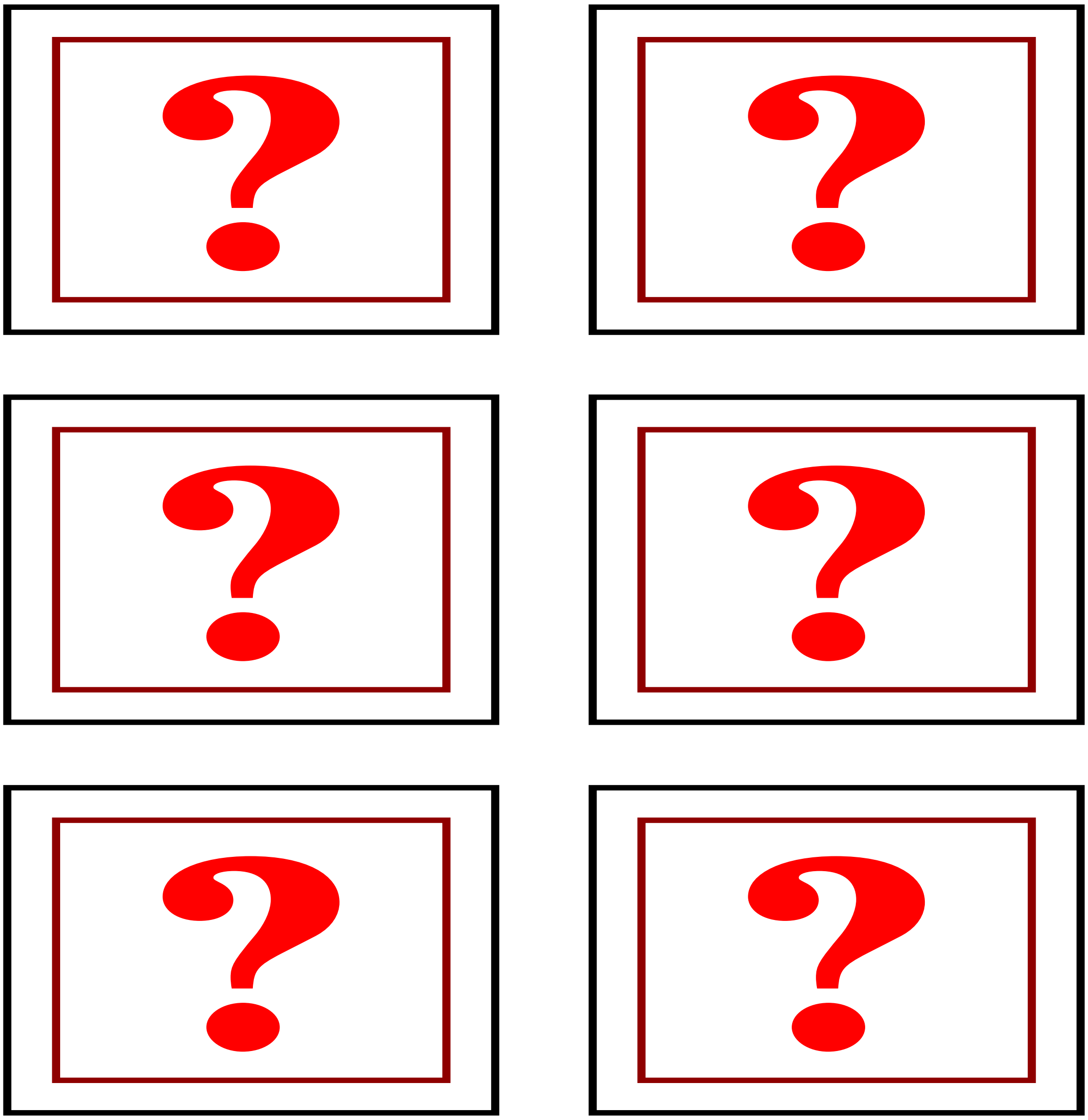

Figure 13. Autocorrelations of for without diurnal adjustments

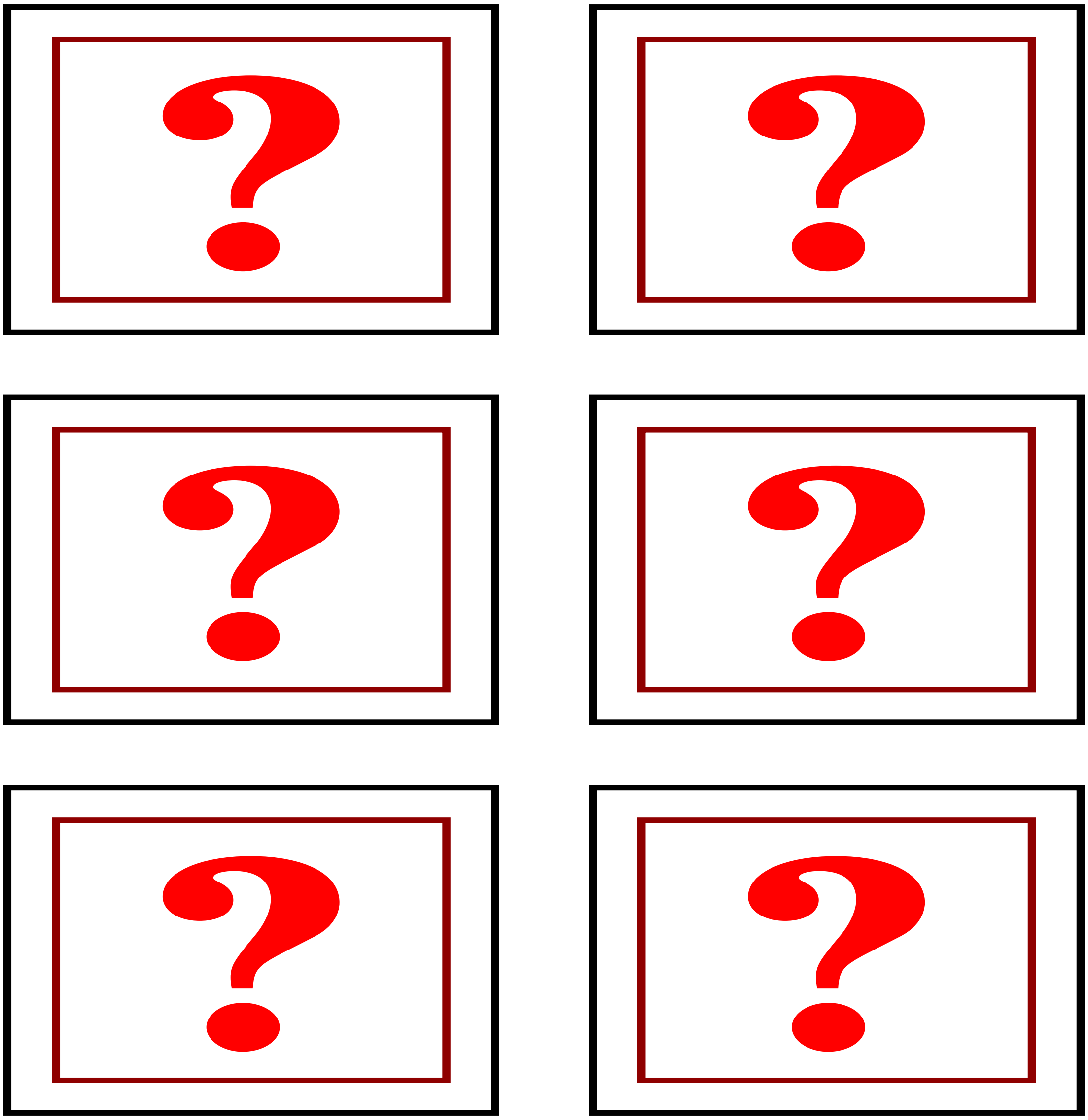

Figure 14. Autocorrelations of for with diurnal adjustments

As can be seen by visually inspecting the autocorrelations, all of the residual series are pretty-much acceptable *without* diurnal adjustments except for 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. in blue and in green

Figure 17. in blue and in green

Figure 18. Zoomed in view of in blue and in green

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

Consider a 5-dimensional multivariate Hawkes model of order 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 seperately. The 5 dimensions are organized as follows:

BATS Buys	BATS Sells	INET Buys	INET Sells	ARCA Trades

(160)

Figure 19.

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

$$\lambda = \begin{pmatrix} 0.25380789517348 \\ 0.269289236349466 \\ 0.221292886522613 \\ 0.158954542395839 \\ 0.371572853723448 \end{pmatrix} \tag{161}$$

$$\alpha = \begin{pmatrix} 4.3514\times10^{-9} & 0.011879 & 0.2648 & 1.917\times10^{-8} & 0.10771 \\ 0.021881 & 2.6164\times10^{-8} & 2.5725\times10^{-8} & 0.024946 & 0.25138 \\ 0.29092 & 0.51715 & 1.1254\times10^{-8} & 0.0029919 & 0.004607 \\ 0.0041449 & 0.52852 & 0.018077 & 3.2535\times10^{-9} & 0.0237 \\ 0.021501 & 0.71358 & 1.0954 & 0.15264 & 4.1222\times10^{-9} \end{pmatrix} \tag{162}$$

$$\beta = \begin{pmatrix} 1.0954 & 10.803 & 16.665 & 20.188 & 9.6059 \\ 5.6238 & 11.558 & 16.721 & 18.304 & 7.9016 \\ 7.8125 & 15.299 & 16.431 & 14.702 & 6.6458 \\ 8.3083 & 15.758 & 17.749 & 12.953 & 3.1621 \\ 9.4264 & 16.369 & 19.303 & 11.071 & 2.8302 \end{pmatrix} \tag{163}$$

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 for a total of 105 parameters. coefficients that are are highlighted in bold. The parameters listed below resulted in a log-likelihood value of 36543.8529. An interesting pattern emerges in the coefficients where it takes on some approximate stair-step pattern ranging from to 22. This might be indicitative of some fixed-frequency algorithms operating across the different exchanges at approximate 1-second intervals.

$$\lambda = \begin{pmatrix} 0.113371928486215301 \\ 0.116069526955243113 \\ 0.120010488406567112 \\ 0.140864383337674315 \\ 0.236370243964866722 \end{pmatrix} \tag{164}$$

$$\alpha_1 = \begin{pmatrix} 0.000000400520039 & 0.000743243048280 & 0.0730760324025721 & 0.0235425002925593 & 0.14994903109 \\ 0.000836306407254 & 0.000048087752871 & 0.0009983197029208 & 0.36091325418001 & 0.0303494022034 \\ 0.000007657273830 & 0.008293393618634 & 0.0000346485386433 & 0.55279157046563 & 0.0303324666473 \\ 0.000000051209296 & 0.044218944305554 & 0.0165858723488658 & 0.0002898699267899 & 0.12041188377 \\ 0.000343063367497 & 0.019728025120072 & 0.22664219457110 & 0.20883023885464 & 0.0002187148763 \end{pmatrix} \tag{165}$$

$$\alpha_2 =$$

$$\begin{pmatrix} 0.0247169438667 & 0.045938324942878493 & 0.52035195378729 & 0.0015976654768 & 0.0219865625857849 \\ 0.10369500283 & 0.000000961851428240 & 0.0058603752158104 & 0.17159388407 & 0.0001956826269151 \\ 0.0619247685514 & 0.005680420895898976 & 0.0000041940337011 & 0.0009132788022 & 0.0161550464515489 \\ 0.0073308612563 & 0.3760898786954499 & 0.0078995090167169 & 0.0000971358022 & 0.0022020712790430 \\ 0.37860663035 & 0.8648532461379836 & 0.0096939577784123 & 0.23909856627 & 0.0000001318796171 \end{pmatrix} \tag{166}$$

$$\beta_1 = \begin{pmatrix} 2.02691486662775 & 4.58853278669795 & 9.21516653991608 & 14.2039223554899 & 17.7230908440328108 \\ 2.30228990848878 & 5.70815142794409 & 9.75920981324501 & 15.0047495693597 & 17.1640776964259771 \\ 2.71360844613891 & 6.97390906252072 & 10.9112224210093 & 16.3935104902520 & 17.3801721025480269 \\ 3.18861359927744 & 6.93702281997507 & 12.0261860231254 & 17.5228876305459 & 17.8876296984556440 \\ 3.95262799649030 & 7.76155541730819 & 13.5039942724633 & 17.3549525971848 & 18.0730780733303966 \end{pmatrix} \tag{167}$$

$$\beta_2 = \begin{pmatrix} 19.6811983441165 & 20.56326127197891 & 18.53440853276660 & 11.10183435325997 & 5.955287687038747 \\ 20.2253306600591 & 21.39051471260508 & 16.97184115533537 & 9.548598696946248 & 5.459761230875715 \\ 20.2208259457254 & 22.20704300748698 & 17.88989095276187 & 8.724870367131993 & 4.215302773261564 \\ 19.7356631996375 & 21.67330389603866 & 15.76838788843381 & 7.534795006501931 & 3.517163899772246 \\ 20.2972304557004 & 19.06667927692781 & 13.19618799557176 & 6.812943703872132 & 2.825437512911523 \end{pmatrix} \tag{168}$$

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

[8] N. Hautsch. *Econometrics of financial high-frequency data*. Springer, 2011.

[9] A. Karr. *Point processes and their statistical inference*, volume 7. CRC, 1991.

[10] Paul Koosis. *The logarithmic integral I*. Cambridge University Press, 1988.

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

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

[15] T. Ozaki and H. Akaike. On a generalized parameteric model for hawkes' self-exciting point processes. *Institute of Statistical Mathematics Res. Memo*, 149, 1979.

[16] Kelly Roach. Hypergeometric function representations. In *International Symposium on Symbolic and Algebraic Computation*, pages 301–308. 1996.

[17] 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.

[18] H. Shek. Modeling high frequency market order dynamics using self-excited point process. *Available at SSRN 1668160*, , 2010.

[19] Ioane Muni Toke. An introduction to hawkes processes with applications to finance. , http://fiquant.mas.ecp.fr/ioane_files/HawkesCourseSlides.pdf.

[20] 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][17] is the inverse of given by

$$\begin{aligned} W(z) &= \{x : x e^x = z\} \\ &= W(0,z) \\ &= 1 + (\ln(z) - 1)\exp\left(\frac{i}{2\pi}\int_0^\infty \frac{1}{x+1}\ln\left(\frac{x - i\pi - \ln(x) + \ln(z)}{x + i\pi - \ln(x) + \ln(z)}\right)\mathrm{d}x\right) \\ &= \sum_{k=1}^{\infty} \frac{(-k)^{k-1} z^k}{k!} \end{aligned} \tag{169}$$

where is

$$W(a,z) = 1 + (2i\pi a + \ln(z) - 1)\exp\left(\frac{i}{2\pi}\int_0^\infty \frac{1}{x+1}\ln\left(\frac{x + (2a-1)i\pi - \ln(x) + \ln(z)}{x + (2a+1)i\pi - \ln(x) + \ln(z)}\right)\mathrm{d}x\right) \tag{170}$$

A generaliztion of (169) is solved by

$$\{x : x b^x = z\} = \frac{W(\ln(b) z)}{\ln(b)} \tag{171}$$

The W function satisifes several identities

$$\begin{aligned} W(z) e^{W(z)} &= z \\ W(z\ln(z)) &= \ln(z) && \forall z < 1 \\ |W(z)| &= W(|z|) \\ e^{nW(z)} &= z^n W(z)^{-n} \\ \ln(W(n,z)) &= \ln(z) - W(n,z) + 2i\pi n \\ W\left(-\frac{\ln(z)}{z}\right) &= -\ln(z) && \forall z \in [0,e] \\ \frac{W(-\ln(z))}{-\ln(z)} &= z^{z^{z^{\cdot^{\cdot^{\cdot}}}}} \end{aligned} \tag{172}$$

where . Some special values are

$$\begin{aligned} W(-1,-e^{-1}) &= -1 \\ W(-e^{-1}) &= -1 \\ W(e) &= 1 \\ W(0) &= 0 \\ W(\infty) &= \infty \\ W(-\infty) &= \infty + i\pi \\ W\left(-\frac{\pi}{2}\right) &= \frac{i\pi}{2} \\ W\left(-\ln(\sqrt{2})\right) &= -\ln(2) \\ W\left(-1, -\ln(\sqrt{2})\right) &= -2\ln(2) \end{aligned} \tag{173}$$

We also have the limit

$$\lim_{a\to\pm\infty}\frac{W(a,x)}{a}=2\pi i \tag{174}$$

and differential

$$\frac{\mathrm{d}}{\mathrm{d}z}W(a,f(z))=\frac{W(a,f(z))\frac{\mathrm{d}}{\mathrm{d}z}f(z)}{f(z)(1+W(a,f(z)))} \tag{175}$$

as well as the obvious integral

$$\int_0^1 W\left(-\frac{\ln(x)}{x}\right)\mathrm{d}x=\int_0^1 -\ln(x)\mathrm{d}x=1 \tag{176}$$

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

$$\begin{aligned}W_{\ln}(z)&=W\left(-1,-\frac{\ln(z)}{z}\right)\\&=1+\left(\ln\left(-\frac{\ln(z)}{z}\right)-1-2\pi i\right)\exp\left(\frac{i}{2\pi}\int_0^\infty\frac{1}{x+1}\ln\left(\frac{x-3i\pi-\ln(x)+\ln\left(-\frac{\ln(z)}{z}\right)}{x-i\pi-\ln(x)+\ln\left(-\frac{\ln(z)}{z}\right)}\right)\mathrm{d}x\right)\end{aligned} \tag{177}$$

Figure 20. and

Then we have the limits

$$\begin{aligned}\lim_{x\to-\infty}W_{\ln}(x)&=0\\\lim_{x\to+\infty}W_{\ln}(x)&=-\infty\end{aligned} \tag{178}$$

and

$$\operatorname{Im}(W_{\ln}(x)) = \begin{cases} -\pi & -\infty < x < 0 \\ \ldots & 0 \leqslant x \leqslant 1 \\ 0 & 1 < x < \infty \end{cases} \tag{179}$$

$$W_{\ln}(x) \;\; = -\ln(x) \quad \forall x \notin [0, e] \tag{180}$$

The root of Re(W$_{\ln}$(x)) is given by

$$\begin{aligned} \{x\colon \operatorname{Re}(W_{\ln}(x)) = 0\} \;\; &= \{-x\colon (x^2)^{\frac{1}{x}} = e^{3\pi}\} \\ &= \frac{2}{3}\pi W\left(\frac{3}{2}\pi\right) \\ &\cong 0.27441063190284810044\ldots \end{aligned} \tag{181}$$

where the imaginary part of the value at the root of the real part of is

$$\begin{aligned} W_{\ln}\left(\frac{2}{3}\pi W\left(\frac{3}{2}\pi\right)\right) \;\; &= W\left(-1, -\frac{\ln\left(\frac{2}{3}\pi W\left(\frac{3}{2}\pi\right)\right)}{\frac{2}{3}\pi W\left(\frac{3}{2}\pi\right)}\right) \\ &= W\left(-1, \frac{3\pi}{2}\right) \\ &= \frac{3\pi i}{2} \\ &\cong i\,4.712388980384689857\ldots \end{aligned} \tag{182}$$