Point Process Models for Multivariate High-Frequency Irregularly Spaced Data

12-31-2012

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 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 Hawkes models. 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. The Hawkess process with a Weibull kernel is also briefly mentioned and will be explored more in the future.

Table of contents

1. Defintions

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 occured by time t. Let F_t denote the filtration of the pooled process N_t of K point processes consisting of the set $t_0^k < t_1^k < t_2^k < \ldots < t_i^k < \ldots$ denoting the history of arrival times of each event type associated with the $k = 1 \ldots K$ 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 $\theta(t) = \left\{ \begin{smallmatrix} 0 & t<0 \\ 1 & t\geqslant 0 \end{smallmatrix} \right.$

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

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

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

For small values of Δt we have

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

so that

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

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

$$\begin{aligned} &\lim_{\Delta t \to 0} \sum_{j=1}^{\frac{(s_1 - s_0)}{\Delta t}} \left(N_{s_0 + j\Delta t}^k - N_{s_0 + (j-1)\Delta t}^k\right) - \lambda^k(s_0 + j\Delta t | F_t)\Delta t \\ &= \lim_{\Delta t \to 0} (N_{s_0}^k - N_{s_1}^k) - \sum_{j=1}^{\frac{(s_1 - s_0)}{\Delta t}} \lambda^k(j\Delta t | F_t)\Delta t \\ &= (N_{s_0}^k - N_{s_1}^k) - \int_{s_0}^{s_1} \lambda^k(t|F_t)\mathrm{d}t \end{aligned} \tag{5}$$

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

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

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

$$\Lambda^k(s_0, s_1) = \int_{s_0}^{s_1} \lambda^k(t|F_t)\mathrm{d}t \tag{7}$$

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

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

Let

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

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

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

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

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

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 interpreated as a generalized residual. The *backwards recurrence time* given by

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

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

1.1.1. Stochastic Integrals.

The *stochastic Stieltjes integral*[1, 2.1][8, 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

$$\int_{(0,t]} X(s)\mathrm{d}N^k_s = \sum_{i \geqslant 1} \theta(t - t^k_i) X(t^k_i) \tag{12}$$

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 [11]

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

We will see that λ can be parameterized in terms of

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

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

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

Then the ACD models are those that consist of the assumption

$$x_i = \psi_i \varepsilon_i \tag{16}$$

where ε_i is independently and identically distributed with density $p(\varepsilon;\ \phi)$ where θ and ϕ 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. [6] The conditional intensity of an ACD model can be expressed in general as

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

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

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

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

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

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.

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

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

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

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

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

An ACD process is stationary if

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

in which case the unconditional mean exists and is given by

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

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 ε_i in (16) instead of an exponential. We have the intensity given by

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

and log-likelihood by

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

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.

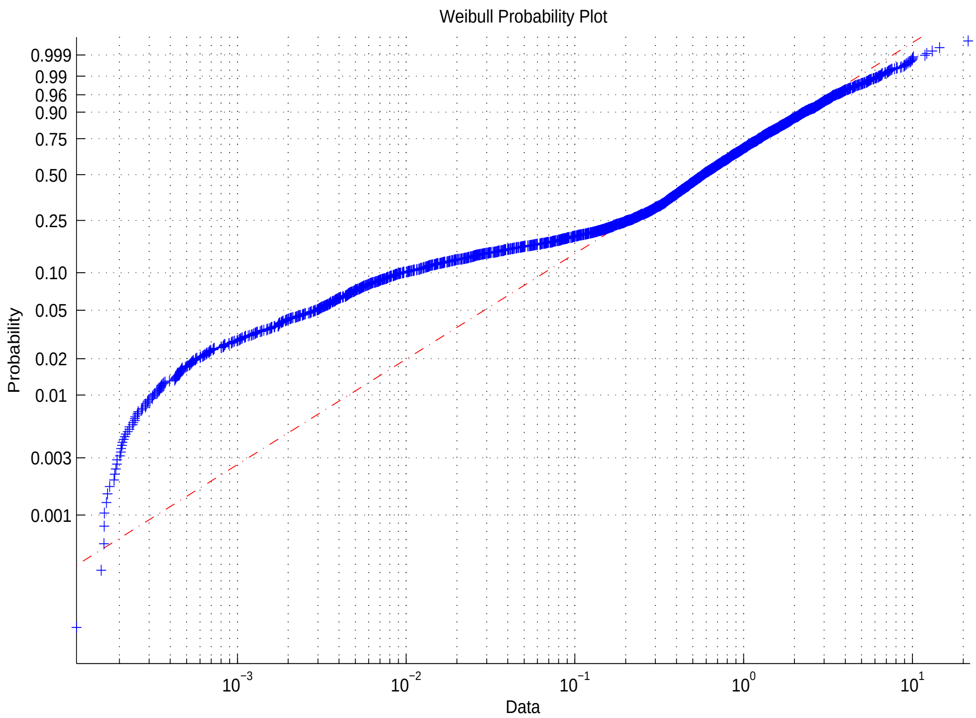

Figure 1. Weibull plot for WACD(1,1) model fit to SPY INET on 2012-11-30

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 [15][7][14][3]

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

where $\lambda_0(t)$ is a deterministic base intensity, see (77), $\nu: \mathbb{R}_+ \to \mathbb{R}_+$ expresses the positive influence of past events t_i on the current value of the intensity process, and κ takes the place of the λ_0 constant in the referenced papers. The (exponential) Hawkes process of order P is a linear self-exciting process defined by the exponential kernel

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

so that the intensity is written as

$$\begin{aligned}\lambda(t) &= \lambda_0(t)\kappa + \int_0^t \sum_{j=1}^{P} \alpha_j e^{-\beta_j(t-s)}\mathrm{d}N_s \\ &= \lambda_0(t)\kappa + \sum_{i=0}^{N_t-1} \sum_{j=1}^{P} \alpha_j e^{-\beta_j(t-t_i)} \\ &= \lambda_0(t)\kappa + \sum_{j=1}^{P} \sum_{i=0}^{N_t-1} \alpha_j e^{-\beta_j(t-t_i)} \\ &= \lambda_0(t)\kappa + \sum_{j=1}^{P} \alpha_j \sum_{i=0}^{N_t-1} e^{-\beta_j(t-t_k)} \\ &= \lambda_0(t)\kappa + \sum_{j=1}^{P} \alpha_j B_j(N_t)\end{aligned} \tag{29}$$

where $B_j(i)$ is given recursively by

$$\begin{aligned}B_j(i) &= \sum_{k=0}^{i-1} e^{-\beta_j(t-t_k)} \\ &= (1+B_j(i-1))e^{-\beta_j(t-t_i)}\end{aligned} \tag{30}$$

A univariate Hawkes process is stationary if

$$\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} < 1 \tag{31}$$

If a Hawkes process is stationary then the unconditional mean is

$$\begin{aligned}\mu = E[\lambda(t)] &= \frac{\lambda_0}{1 - \int_0^\infty \nu(t)\mathrm{d}t} \\ &= \frac{\lambda_0}{1 - \int_0^\infty \sum_{j=1}^{P} \alpha_j e^{-\beta_j t}\mathrm{d}t} \\ &= \frac{\lambda_0}{1 - \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}}\end{aligned} \tag{32}$$

For consecutive events, we have the compensator (7)

$$\begin{aligned}\Lambda(t_{i-1},t_i) &= \int_{t_{i-1}}^{t_i} \lambda(t)\mathrm{d}t \\ &= \int_{t_{i-1}}^{t_i} \left(\lambda_0(t) + \sum_{j=1}^{P} \alpha_j B_j(N_t)\right)\mathrm{d}t \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \sum_{k=0}^{i-1}\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\big(e^{-\beta_j(t_{i-1}-t_k)} - e^{-\beta_j(t_i-t_k)}\big) \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\big(1 - e^{-\beta_j(t_i-t_{i-1})}\big)A_j(i-1)\end{aligned} \tag{33}$$

where there is the recursion

$$\begin{aligned}A_j(i) &= \sum_{t_k \leqslant t_i} e^{-\beta_j(t_i-t_k)} \\ &= \sum_{k=0}^{i-1} e^{-\beta_j(t_i-t_k)} \\ &= 1 + e^{-\beta_j(t_i-t_{i-1})}A_j(i-1)\end{aligned} \tag{34}$$

with $A_j(0)=0$. If $\lambda_0(t)=\lambda_0$ then (33) simplifies to

$$\begin{aligned}\Lambda(t_{i-1},t_i) &= (t_i - t_{i-1})\lambda_0 + \sum_{k=0}^{i-1}\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\big(e^{-\beta_j(t_{i-1}-t_k)} - e^{-\beta_j(t_i-t_k)}\big) \\ &= (t_i - t_{i-1})\lambda_0 + \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\big(1 - e^{-\beta_j(t_i-t_{i-1})}\big)A_j(i-1)\end{aligned} \tag{35}$$

Similiarly, another parameterization is given by

$$\begin{aligned}\Lambda(t_{i-1},t_i) &= \int_{t_{i-1}}^{t_i} \kappa\lambda_0(s)\mathrm{d}s + \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\big(1 - e^{-\beta_j(t_i-t_{i-1})}\big)A_j(i-1) \\ &= \kappa\int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\big(1 - e^{-\beta_j(t_i-t_{i-1})}\big)A_j(i-1) \\ &= \kappa\Lambda_0(t_{i-1},t_i) + \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\big(1 - e^{-\beta_j(t_i-t_{i-1})}\big)A_j(i-1)\end{aligned} \tag{36}$$

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

$$\lambda(t) = \kappa\,\lambda_0(t) + \sum_{j=1}^{P} \alpha_j B_j(N_t) \tag{37}$$

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

1.3.2. The Hawkes(1) Model.

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

$$\lambda(t) = \lambda_0 + \sum_{t_i<t} \alpha e^{-\beta\,(t-t_i)} \tag{38}$$

which has the unconditional mean

$$E[\lambda(t)] = \frac{\lambda_0}{1 - \frac{\alpha}{\beta}} \tag{39}$$

1.3.3. Maximum Likelihood Estimation.

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

$$\begin{aligned} \ln\mathcal{L}(N(t)_{t\in[0,T]}) &= \int_0^T (1-\lambda(s))\mathrm{d}s + \int_0^T \ln\lambda(s)\mathrm{d}N_s \\ &= T - \int_0^T \lambda(s)\mathrm{d}s + \int_0^T \ln\lambda(s)\mathrm{d}N_s \end{aligned} \tag{40}$$

which in the case of the Hawkes model of order P can be explicitly written [13] as

$$\begin{aligned} \ln\mathcal{L}(\{t_i\}_{i=1\ldots n}) &= T - \Lambda(0,T) + \sum_{i=1}^{n} \ln\lambda(t_i) \\ &= T + \sum_{i=1}^{n} \ln\lambda(t_i) - \Lambda(t_{i-1}, t_i) \\ &= T - \Lambda(0,T) + \sum_{i=1}^{n} \ln\lambda(t_i) \\ &= T - \Lambda(0,T) + \sum_{i=1}^{n} \ln\left(\kappa\lambda_0(t_i) + \sum_{j=1}^{P}\sum_{k=1}^{i-1} \alpha_j e^{-\beta_j(t_i - t_k)}\right) \\ &= T - \Lambda(0,T) + \sum_{i=1}^{n} \ln\left(\kappa\lambda_0(t_i) + \sum_{j=1}^{P} \alpha_j R_j(i)\right) \\ &= T - \int_0^T \kappa\lambda_0(s)\mathrm{d}s - \sum_{i=1}^{n}\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\left(1 - e^{-\beta_j(t_n - t_i)}\right) \\ &+ \sum_{i=1}^{n} \ln\left(\kappa\lambda_0(t_i) + \sum_{j=1}^{P} \alpha_j R_j(i)\right) \end{aligned} \tag{41}$$

where $T = t_n$ and we have the recursion[12]

$$\begin{aligned} R_j(i) &= \sum_{k=1}^{i-1} e^{-\beta_j(t_i - t_k)} \\ &= e^{-\beta_j(t_i - t_{i-1})}(1 + R_j(i-1)) \end{aligned} \tag{42}$$

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

$$\begin{aligned} \ln\mathcal{L}(\{t_i\}_{i=1\ldots n}) &= T - \kappa T - \sum_{i=1}^{n}\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\left(1 - e^{-\beta_j(t_n - t_i)}\right) \\ &+ \sum_{i=1}^{n} \ln\left(\lambda_0 + \sum_{j=1}^{P} \alpha_j R_j(i)\right) \end{aligned} \tag{43}$$

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. 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. [16].

1.4. The Hawkes Process Having a Weibull Kernel.

The exponential kernel of the Hawkes process can be replaced with that of a Weibull kernel. [10, 6.3] Recall that the intensity is defined by (27)

$$\begin{aligned}\lambda(t) &= \lambda_0(t)\kappa + \int_{-\infty}^{t} \nu(t-s)\mathrm{d}N_s \\ &= \lambda_0(t)\kappa + \sum_{t_i<t} \nu(t-t_i)\end{aligned} \tag{44}$$

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

$$\nu(t) = \sum_{j=1}^{P} \alpha_j \left(\frac{\kappa_j}{\omega_j}\right)\left(\frac{t}{\omega_j}\right)^{\kappa_j-1} e^{-\beta_j\left(\frac{t}{\omega_j}\right)^{\kappa_j}} \tag{45}$$

so the Weibull-Hawkes intensity is written is

$$\begin{aligned}\lambda(t) &= \lambda_0(t)\kappa + \int_0^t \sum_{j=1}^{P} \alpha_j \left(\frac{\kappa_j}{\omega_j}\right)\left(\frac{t-s}{\omega_j}\right)^{\kappa_j-1} e^{-\beta_j\left(\frac{t-s}{\omega_j}\right)^{\kappa_j}} \mathrm{d}N_s \\ &= \lambda_0(t)\kappa + \sum_{i=0}^{N_t-1}\sum_{j=1}^{P} \alpha_j \left(\frac{\kappa_j}{\omega_j}\right)\left(\frac{t-t_i}{\omega_j}\right)^{\kappa_j-1} e^{-\beta_j\left(\frac{t-t_i}{\omega_j}\right)^{\kappa_j}}\end{aligned} \tag{46}$$

1.5. Combining the ACD and Hawkes Models.

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

$$\lambda(t) = \lambda_0(t) + \frac{1}{\psi_{N_t}} + \int_0^t \nu(t-s)\mathrm{d}N_s \tag{47}$$

where $\lambda_0(t)$ is the determinstic baseline intensity(77) and where the ACD(20) part is

$$\psi_i = \omega + \sum_{j=1}^{m} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} \tag{48}$$

and the Hawkes part has the exponential kernel(28)

$$\nu(t) = \sum_{j=1}^{P} \gamma_j e^{-\varphi_j t} \tag{49}$$

so that

$$\begin{aligned}\int_0^t \nu(t-s)\mathrm{d}N_s &= \int_0^t \sum_{j=1}^{P} \gamma_j e^{-\varphi_j (t-s)} \mathrm{d}N_s \\ &= \sum_{k=0}^{N_t} \nu(t-t_k) \\ &= \sum_{k=0}^{N_t}\sum_{j=1}^{P} \gamma_j e^{-\varphi_j(t-t_k)} \\ &= \sum_{j=1}^{P} \gamma_j \sum_{k=0}^{N_t} e^{-\varphi_j(t-t_k)} \\ &= \sum_{j=1}^{P} \gamma_j B_j(N_t)\end{aligned} \tag{50}$$

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 α and β are also used as parameter names. The Hawkes part of the intensity has a recursive structure similiar to that of the compensator. Let

$$\begin{aligned} B_j(i) &= \sum_{k=0}^{i-1} e^{-\varphi_j(t-t_k)} \\ &= (1+B_j(i-1))e^{-\varphi_j(t-t_i)} \end{aligned} \tag{51}$$

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

$$\lambda(t) = \lambda_0(t) + \frac{1}{\omega + \sum_{j=1}^m \alpha_j x_{N_t-j} + \sum_{j=1}^q \beta_j \psi_{N_t-j}} + \sum_{j=1}^P \gamma_j B_j(N_t) \tag{52}$$

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

$$\begin{aligned} \ln \mathcal{L}(\{t_i\}_{i=1,..,n}) &= \sum_{i=1}^n \left(\ln \lambda(t_i) - \int_{t_{i-1}}^{t_i} \lambda(t) \mathrm{d}t \right) \\ &= \sum_{i=1}^n (\ln \lambda(t_i) - \Lambda(t_{i-1}, t_i)) \\ &= \sum_{i=1}^n \left(\ln \lambda(t_i) - \tilde{\mathcal{E}}_i \right) \end{aligned} \tag{53}$$

By direct calculation, combining (19) and (33), and letting $x_i = t_i - t_{i-1}$ we have the compensator

$$\begin{aligned} \tilde{\mathcal{E}}_i &= \Lambda(t_{i-1}, t_i) \\ &= \int_{t_{i-1}}^{t_i} \lambda(t) \mathrm{d}t \\ &= \int_{t_{i-1}}^{t_i} \left(\lambda_0(t) + \frac{1}{\psi_{N_t+1}} + \int_0^t \nu(t-s) \mathrm{d}N_s \right) \mathrm{d}t \\ &= \frac{x_i}{\psi_i} + \int_{t_{i-1}}^{t_i} \left(\lambda_0(t) + \int_0^t \nu(t-s) \mathrm{d}N_s \right) \mathrm{d}t \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(t) \mathrm{d}t + \frac{x_i}{\psi_i} + \sum_{k=0}^{i-1} \sum_{j=1}^P \frac{\gamma_j}{\varphi_j} (e^{-\varphi_j(t_{i-1}-t_k)} - e^{-\varphi_j(t_i-t_k)}) \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(t) \mathrm{d}t + \frac{x_i}{\psi_i} + \sum_{j=1}^P \frac{\gamma_j}{\varphi_j} (1 - e^{-\varphi_j x_i}) A_j(i-1) \end{aligned} \tag{54}$$

where ψ_i is defined by (48) and

$$A_j(i) = 1 + e^{-\varphi_j x_i} A_j(i-1) \tag{55}$$

is given by (34) so that (53) can be wriitten as

$$\begin{aligned} \ln \mathcal{L}(\{t_i\}_{i=0,..,n}) &= \sum_{i=1}^n (\ln \lambda(t_i) - \tilde{\mathcal{E}}_i) \\ &= \sum_{i=1}^n \left(\ln \lambda(t_i) - \left(\frac{x_i}{\psi_i} + \sum_{j=1}^P \frac{\gamma_j}{\varphi_j} (1 - e^{-\varphi_j x_i}) A_j(i-1) \right) \right) \\ &= \sum_{i=1}^n \ln \left(\frac{1}{\psi_i} + \sum_{k=0}^{i-1} \sum_{j=1}^P \gamma_j e^{-\varphi_j(t_i-t_k)} \right) - \left(\frac{x_i}{\psi_i} + \sum_{j=1}^P \frac{\gamma_j}{\varphi_j} (1 - e^{-\varphi_j x_i}) A_j(i-1) \right) \\ &= \sum_{i=1}^n \ln \left(\frac{1}{\psi_i} + \sum_{j=1}^P \gamma_j B_j(i) \right) - \left(\frac{x_i}{\psi_i} + \sum_{j=1}^P \frac{\gamma_j}{\varphi_j} (1 - e^{-\varphi_j x_i}) A_j(i-1) \right) \end{aligned} \tag{56}$$

1.6. 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[7][5][9] is defined with intensities $\lambda^m(t), m=1...M$ given by

$$\begin{aligned}\lambda^m(t) &= \lambda_0^m(t)\kappa^m + \sum_{n=1}^{M}\int_0^t \sum_{j=1}^{P} \alpha_j^{m,n} e^{-\beta_j^{m,n}(t-s)} \mathrm{d}N_s^n \\ &= \lambda_0^m(t)\kappa^m + \sum_{n=1}^{M}\sum_{j=1}^{P}\sum_{t_k^n<t} \alpha_j^{m,n} e^{-\beta_j^{m,n}(t-t_k^n)} \\ &= \lambda_0^m(t)\kappa^m + \sum_{n=1}^{M}\sum_{j=1}^{P} \alpha_j^{m,n} \sum_{t_k^n<t} e^{-\beta_j^{m,n}(t-t_k^n)} \\ &= \lambda_0^m(t)\kappa^m + \sum_{n=1}^{M}\sum_{j=1}^{P} \alpha_j^{m,n} \sum_{t_k^n<t} e^{-\beta_j^{m,n}(t-t_k^n)} \\ &= \lambda_0^m(t)\kappa^m + \sum_{n=1}^{M}\sum_{j=1}^{P} \alpha_j^{m,n} \sum_{k=0}^{N_t^n-1} e^{-\beta_j^{m,n}(t-t_k^n)} \\ &= \lambda_0^m(t)\kappa^m + \sum_{n=1}^{M}\sum_{j=1}^{P} \alpha_j^{m,n}\, B_j^{m,n}(N_t^n)\end{aligned} \tag{57}$$

where in this parameterization κ is a vector which scales the baseline intensities, in this case, specified by piecewise polynomial splines (77). We can write $B_j^{m,n}(i)$ recursively

$$\begin{aligned}B_j^{m,n}(i) &= \sum_{k=0}^{i-1} e^{-\beta_j^{m,n}(t-t_k^n)} \\ &= (1+B_j^{m,n}(i-1))e^{-\beta_j^{m,n}(t-t_i^n)}\end{aligned} \tag{58}$$

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

$$\begin{aligned}\lambda^m(t) &= \kappa^m + \sum_{n=1}^{M}\int_0^t \alpha^{m,n} e^{-\beta^{m,n}(t-s)} \mathrm{d}N_s^n \\ &= \kappa^m + \sum_{n=1}^{M}\sum_{k=0}^{N_t^n-1} \alpha^{m,n} e^{-\beta^{m,n}(t-t_k^n)} \\ &= \kappa^m + \sum_{n=1}^{M} \alpha^{m,n} \sum_{k=0}^{N_t^n-1} e^{-\beta^{m,n}(t-t_k^n)} \\ &= \kappa^m + \sum_{n=1}^{M} \alpha^{m,n}\, B_1^{m,n}(N_t^n)\end{aligned} \tag{59}$$

Rewriting (59) in vectorial notion, we have

$$\lambda(t) = \kappa + \int_0^t G(t-s)\mathrm{d}N_s \tag{60}$$

where

$$G(t) = (\alpha^{m,n}\, e^{-\beta^{m,n}(t-s)})_{m,n=1...M} \tag{61}$$

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

$$\begin{aligned}\mu &= \frac{\kappa}{I - \int_0^\infty G(u)\mathrm{d}u} \\ &= \frac{\kappa}{I - (\frac{\alpha^{m,n}}{\beta^{m,n}})} \\ &= \frac{\kappa}{I-\Gamma}\end{aligned} \tag{62}$$

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

$$\Gamma = \int_0^\infty G(s)\mathrm{d}s = \frac{\alpha^{m,n}}{\beta^{m,n}} \tag{63}$$

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

$$\rho(G) = \max_{a \in \mathcal{S}(G)} |a| \tag{64}$$

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

1.6.1. 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

$$\begin{aligned}\Lambda^m(t_{i-1}^m, t_i^m) &= \int_{t_{i-1}^m}^{t_i^m} \lambda^m(s)\mathrm{d}s \\ &+ \sum_{n=1}^{M} \sum_{j=1}^{P} \sum_{t_k^n < t_{i-1}^m} \frac{\alpha_j^{m,n}}{\beta_j^{m,n}} [e^{-\beta_j^{m,n}(t_{i-1}^m - t_k^n)} - e^{-\beta_j^{m,n}(t_i^m - t_k^n)}] \\ &+ \sum_{n=1}^{M} \sum_{j=1}^{P} \sum_{t_{i-1}^m \leqslant t_k^n < t_i^m} \frac{\alpha_j^{m,n}}{\beta_j^{m,n}} [1 - e^{-\beta_j^{m,n}(t_i^m - t_k^n)}]\end{aligned} \tag{65}$$

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

$$\begin{aligned}A_j^{m,n}(i) &= \sum_{t_k^n < t_i^m} e^{-\beta_j^{m,n}(t_i^m - t_k^n)} \\ &= e^{-\beta_j^{m,n}(t_i^m - t_{i-1}^m)} A_j^{m,n}(i-1) + \sum_{t_{i-1}^m \leqslant t_k^n < t_i^m} e^{-\beta_j^{m,n}(t_i^m - t_k^n)}\end{aligned} \tag{66}$$

and rewrite (65) as

$$\begin{aligned}\Lambda^m(t_{i-1}^m, t_i^m) &= \kappa^m \int_{t_{i-1}^m}^{t_i^m} \lambda_0^m(s)\mathrm{d}s + \int_{t_{i-1}^m}^{t_i^m} \sum_{n=1}^{M} \sum_{j=1}^{P} \sum_{t_k^n < s} \alpha_j^{m,n} e^{-\beta_j^{m,n}(s - t_k^n)} \mathrm{d}s \\ &= \kappa^m \int_{t_{i-1}^m}^{t_i^m} \lambda_0^m(s)\mathrm{d}s \\ &+ \sum_{n=1}^{M} \sum_{j=1}^{P} \frac{\alpha_j^{m,n}}{\beta_j^{m,n}} \left[(1 - e^{-\beta_j^{m,n}(t_i^m - t_{i-1}^m)}) \times A_j^{m,n}(i-1) + \sum_{t_{i-1}^m \leqslant t_k^n < t_i^m} (1 - e^{-\beta_j^{m,n}(t_i^m - t_k^n)}) \right] \\ &= \kappa^m \int_{t_{i-1}^m}^{t_i^m} \lambda_0^m(s)\mathrm{d}s \\ &+ \sum_{n=1}^{M} \sum_{j=1}^{P} \frac{\alpha_j^{m,n}}{\beta_j^{m,n}} \left[(1 - e^{-\beta_j^{m,n}(t_i^m - t_{i-1}^m)}) \times \left(\sum_{t_k^n < t_{i-1}^m} e^{-\beta_j^{m,n}(t_{i-1}^m - t_k^n)} \right) + \sum_{t_{i-1}^m \leqslant t_k^n < t_i^m} (1 - e^{-\beta_j^{m,n}(t_i^m - t_k^n)}) \right]\end{aligned} \tag{67}$$

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

1.6.2. Log-Likelihood.

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

$$\ln \mathcal{L}(\{t_i\}_{i=1,\ldots,N_T}) = \sum_{m=1}^{M} \ln \mathcal{L}^m(\{t_i\}) \tag{68}$$

where each term is defined by

$$\ln \mathcal{L}^m(\{t_i\}) = \int_0^T (1 - \lambda^m(s))\mathrm{d}s + \int_0^T \ln \lambda^m(s) \mathrm{d}N_s^m \tag{69}$$

which in this case can be written as

$$\begin{aligned}\ln \mathcal{L}^m(\{t_i\}) &= T - \Lambda^m(0, T) + \sum_{i=1}^{N_T} z_i^m \ln \left(\lambda_0^m(t_i)\kappa^m + \sum_{n=1}^{M} \sum_{j=1}^{P} \sum_{t_k^n < t_i} \alpha_j^{m,n} e^{-\beta_j^{m,n}(t_i - t_k^n)} \right) \\ &= T - \Lambda^m(0, T) + \sum_{i=1}^{N_T^m} \ln \left(\lambda_0^m(t_i^m)\kappa^m + \sum_{n=1}^{M} \sum_{j=1}^{P} \sum_{t_k^n < t_i^m} \alpha_j^{m,n} e^{-\beta_j^{m,n}(t_i^m - t_k^n)} \right)\end{aligned} \tag{70}$$

where again $t_{N_T} = T$ and

$$z_i^m = \begin{cases} 1 & \text{event } t_i \text{ of type } m \\ 0 & \text{otherwise} \end{cases} \tag{71}$$

and

$$\Lambda^m(0,T) = \int_0^T \lambda^m(t)\mathrm{d}t = \sum_{i=1}^{N_T^m} \Lambda^m(t_{i-1}^m, t_i^m) \tag{72}$$

where $\Lambda^m(t_{i-1}^m, t_i^m)$ is given by (67). Similiar to to the one-dimensional case, we have the recursion

$$\begin{aligned} R_j^{m,n}(i) &= \sum_{t_k^n < t_j^m} e^{-\beta_j^{m,n}(t_i^m - t_k^n)} \\ &= \begin{cases} e^{-\beta_j^{m,n}(t_i^m - t_{i-1}^m)} R_j^{m,n}(i-1) + \sum_{t_{i-1}^m \leqslant t_k^n < t_i^m} e^{-\beta_j^{m,n}(t_i^m - t_k^n)} & \text{if } m \neq n \\ e^{-\beta_j^{m,n}(t_i^m - t_{i-1}^m)} \left(1 + R_j^{m,n}(i-1)\right) & \text{if } m = n \end{cases} \end{aligned} \tag{73}$$

so that (70) can be rewritten as

$$\begin{aligned} \ln \mathcal{L}^m(\{t_i\}) &= T - \kappa^m \int_0^T \lambda_0^m(t)\mathrm{d}t - \ldots \\ &\ldots - \sum_{i=1}^{N_T^m} \sum_{n=1}^{M} \sum_{j=1}^{P} \frac{\alpha_j^{m,n}}{\beta_j^{m,n}} \left[(1 - e^{-\beta_j^{m,n}(t_i^m - t_{i-1}^m)}) \times A_j^{m,n}(i-1) + \sum_{t_{i-1}^m \leqslant t_k^n < t_i^m} (1 - e^{-\beta_j^{m,n}(t_i^m - t_k^n)}) \right] + \ldots \\ &\ldots + \sum_{i=1}^{N_T^m} \ln \left(\lambda_0^m(t_i^m)\kappa^m + \sum_{n=1}^{M} \sum_{j=1}^{P} \alpha_j^{m,n} R_j^{m,n}(i) \right) \end{aligned} \tag{74}$$

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[4] 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 λ_0 to $\frac{0.5}{E[x_i]}$, $\alpha_{1\ldots P} = \frac{1}{P}$ and $\beta_{1\ldots P} = 2$. 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 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 $t = 0$ and $t = 6.5 \times 60 \times 60 = 23400$ respectively since t has units of seconds. Let the adjusted durations be defined

$$\tilde{x}_i = \phi(t_i) x_i \tag{75}$$

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

$$P_j = \frac{1}{(N_{t(zj)+w} - N_{t(zj)-w})} \sum_{i=N_{t(zj)-w}}^{N_{t(zj)+w}} \frac{1}{x_i} \text{ for } j = 0 \ldots 13 \tag{76}$$

where $z = 60 \times 30 = 1800$ is the number of seconds in a half-hour and $j = 0\ldots(6.5 \times 2)$. 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.

$$\lambda_0(t) = f(t, P_0, \ldots, P_j) \tag{77}$$

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

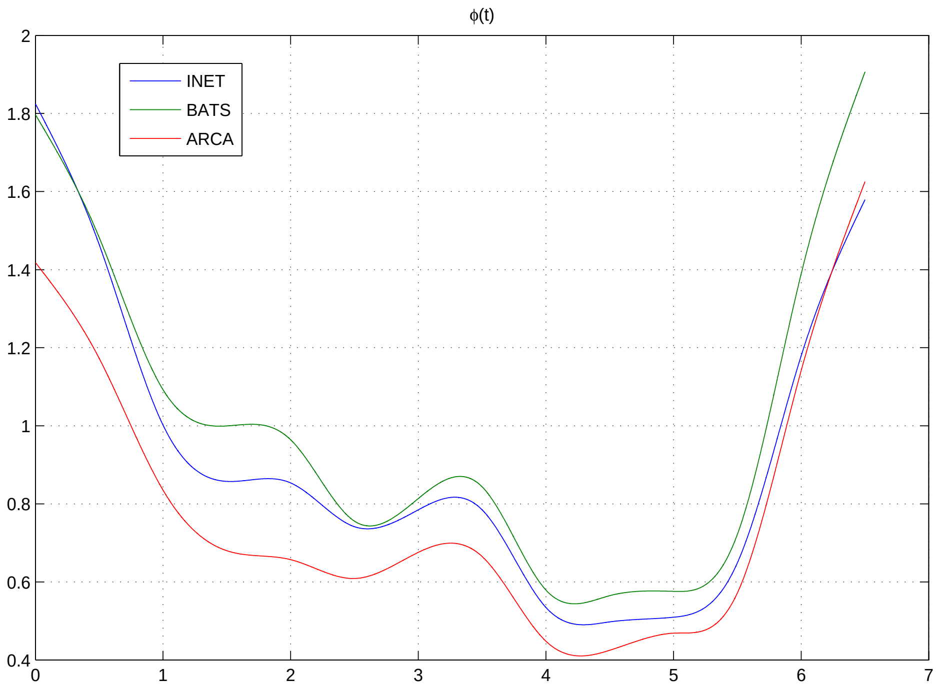

Figure 2. Interpolating spline $\phi(t)$ for SPY on 2012-11-30

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 $n = 20787$. 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 $\mathrm{Var}[\Lambda]$ to 1.0 the better, since Λ 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 Λ series is not autocorrelated.

P	κ	$\alpha_{1...P}$	$\beta_{1...P}$	$\ln \mathcal{L}(\{t_i\})$	$E[\lambda(t)]$	$E[\Lambda]$	Var[Λ]
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 1. Parameters and statistics for model fitted to data without diurnal adjustments

P	κ	$\alpha_{1...P}$	$\beta_{1...P}$	$\ln \mathcal{L}(\{t_i\})$	$E[\lambda(t)]$	$E[\Lambda]$	Var[Λ]
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 2. Parameters and statistics for model fitted to data with diurnal adjustments

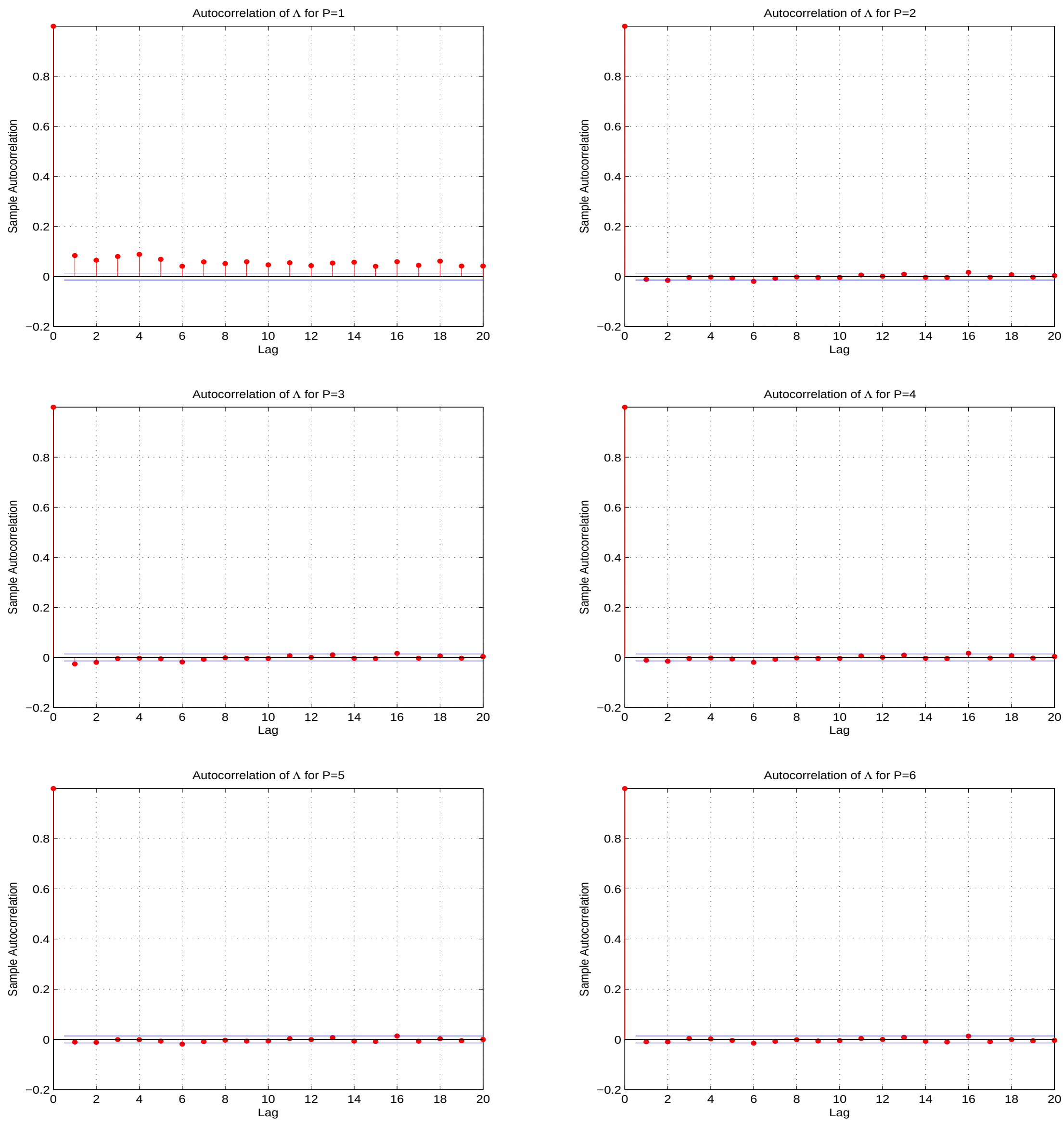

Figure 3. Autocorrelations of $\Lambda(t_{i-1}, t_i)$ for $P = 1...6$ without diurnal adjustments

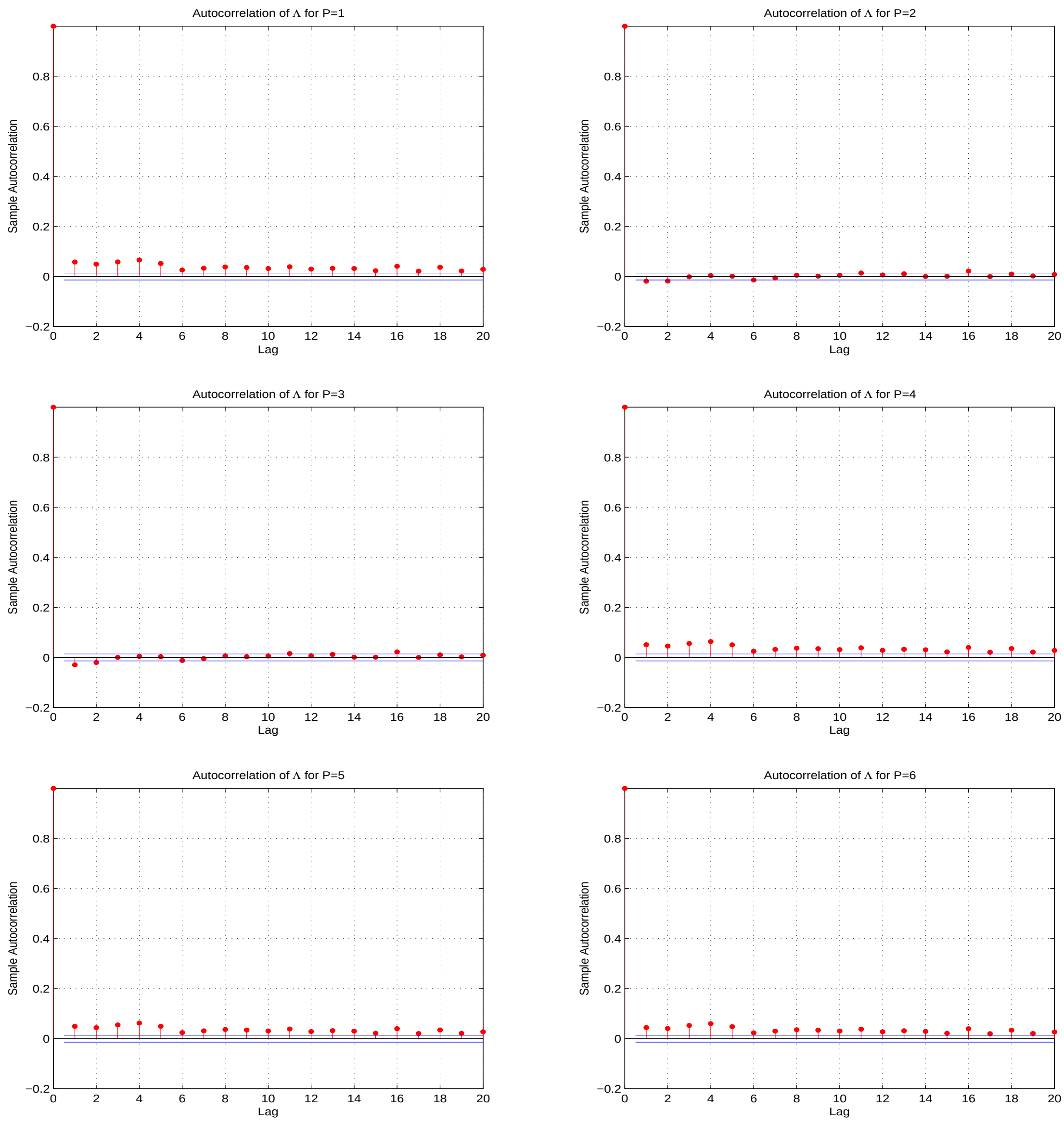

Figure 4. Autocorrelations of $\Lambda(t_{i-1}, t_i)$ for $P = 1 \ldots 6$ with diurnal adjustments

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.

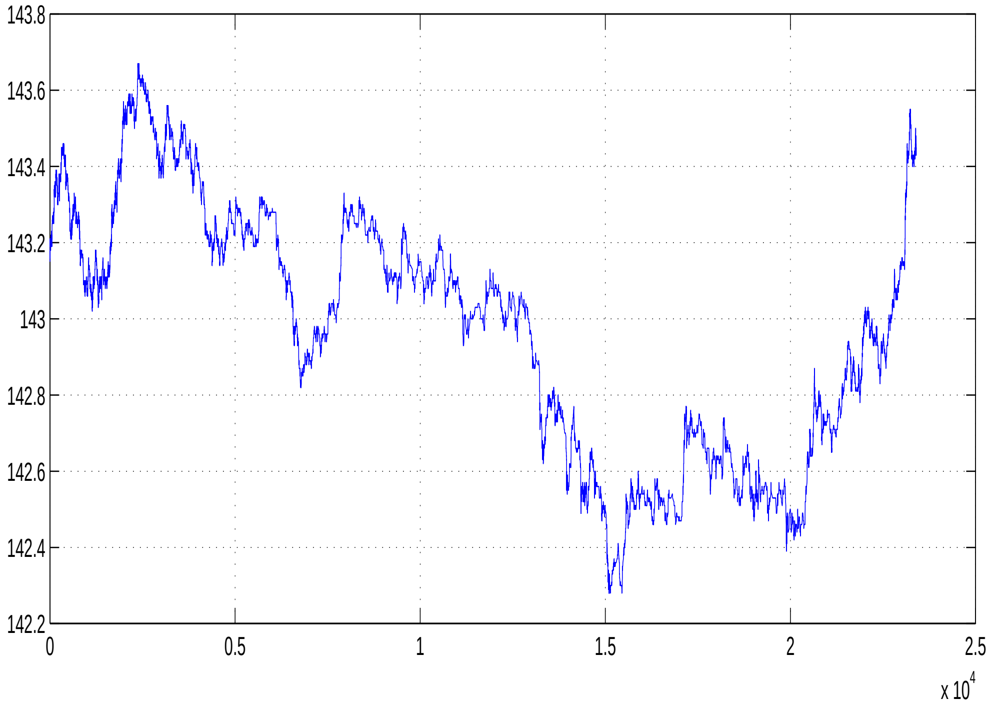

Figure 5. Price history for SPY traded on INET on Oct 22nd, 2012

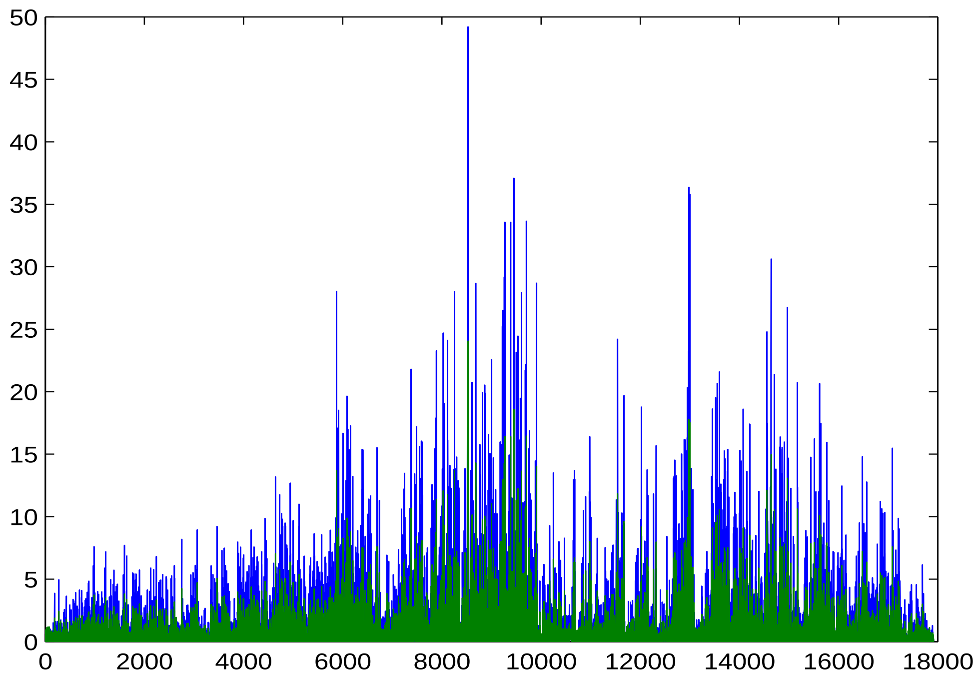

Figure 6. $x_i = t_i - t_{i-1}$ in blue and $\{\Lambda(t_{i-1}, t_i) \colon P = 1\}$ in green

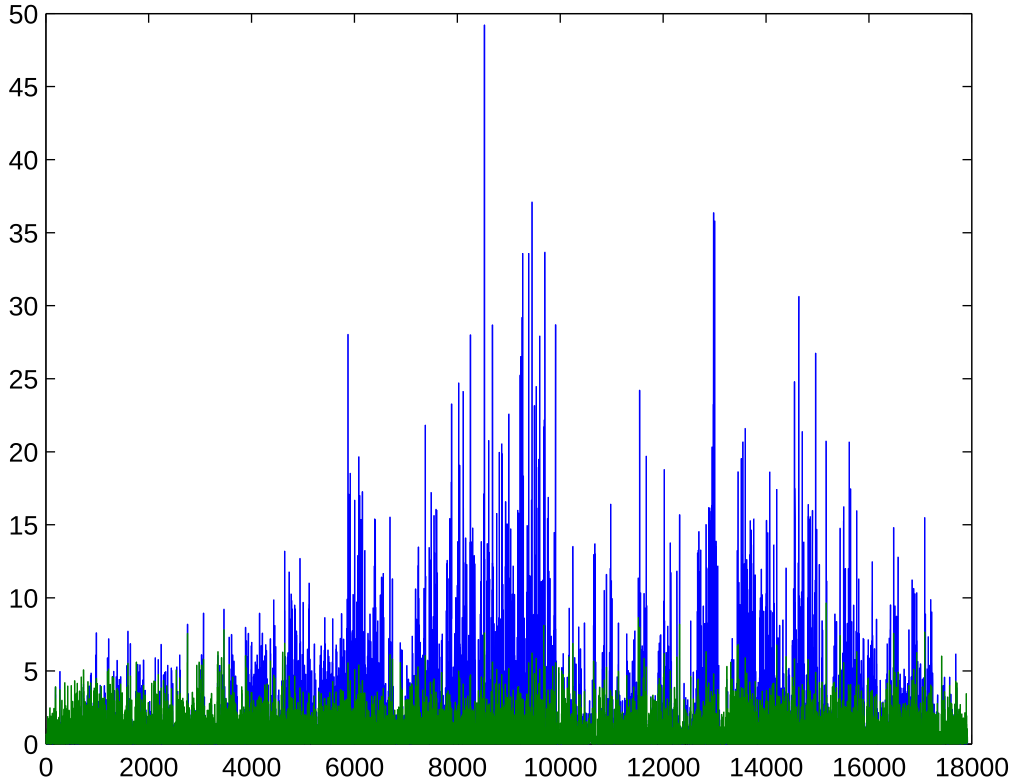

Figure 7. $x_i = t_i - t_{i-1}$ in blue and $\{\Lambda(t_{i-1}, t_i) \colon P = 6\}$ in green

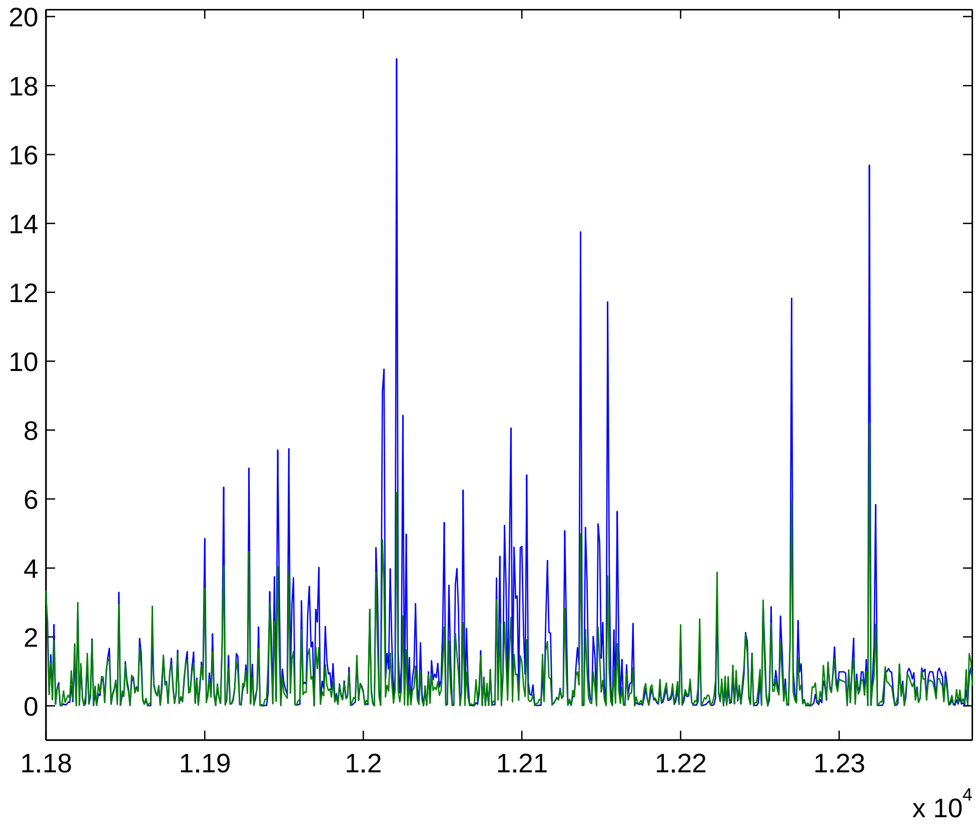

Figure 8. Zoomed in view of $x_i = t_i - t_{i-1}$ in blue and $\{\Lambda(t_{i-1}, t_i) : P = 6\}$ in green

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 seperately. The 5 dimensions are organized as follows:

BATS Buys	BATS Sells	INET Buys	INET Sells	ARCA Trades

(78)

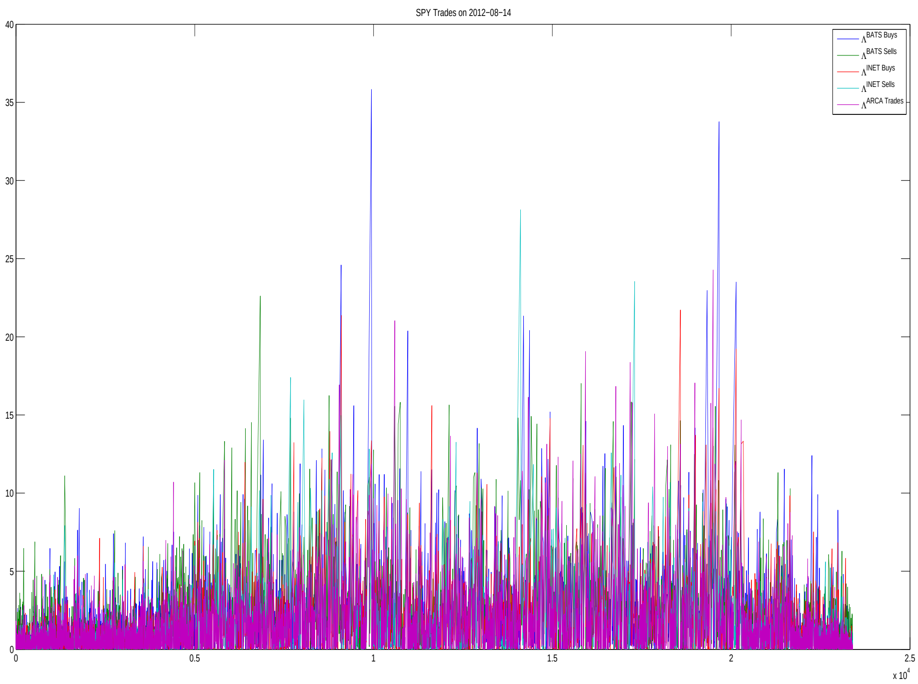

Figure 9.

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 α (>0.1) are highlighted in bold.

$$\lambda = \begin{pmatrix} 0.25380789517348 \\ 0.269289236349466 \\ 0.221292886522613 \\ 0.158954542395839 \\ 0.371572853723448 \end{pmatrix} \tag{79}$$

$$\alpha = \begin{pmatrix} 4.3514 \times 10^{-9} & 0.011879 & \mathbf{0.2648} & 1.917 \times 10^{-8} & \mathbf{0.10771} \\ 0.021881 & 2.6164 \times 10^{-8} & 2.5725 \times 10^{-8} & 0.024946 & \mathbf{0.25138} \\ \mathbf{0.29092} & \mathbf{0.51715} & 1.1254 \times 10^{-8} & 0.0029919 & 0.004607 \\ 0.0041449 & \mathbf{0.52852} & 0.018077 & 3.2535 \times 10^{-9} & 0.0237 \\ 0.021501 & \mathbf{0.71358} & \mathbf{1.0954} & \mathbf{0.15264} & 4.1222 \times 10^{-9} \end{pmatrix} \tag{80}$$

$$\beta = \begin{pmatrix} 1.0954 & 10.803 & 16.665 & 20.188 & 9.6059 \\ 5.6238 & 11.558 & 16.721 & 18.304 & 7.9016 \\ 7.8125 & 15.299 & 16.431 & 14.702 & 6.6458 \\ 8.3083 & 15.758 & 17.749 & 12.953 & 3.1621 \\ 9.4264 & 16.369 & 19.303 & 11.071 & 2.8302 \end{pmatrix} \tag{81}$$

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. α_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 β coefficients where it takes on some approximate stair-step pattern ranging from 2 to 22. This might be indicitative of some fixed-frequency algorithms operating across the different exchanges at approximate 1-second intervals.

$$\lambda = \begin{pmatrix} 0.113371928486215301 \\ 0.116069526955243113 \\ 0.120010488406567112 \\ 0.140864383337674315 \\ 0.236370243964866722 \end{pmatrix} \tag{82}$$

$$\alpha_1 = \begin{pmatrix} 0.000000400520039 & 0.000743243048280 & 0.0730760324025721 & 0.0235425002925593 & \mathbf{0.14994903109} \\ 0.000836306407254 & 0.000048087752871 & 0.0009983197029208 & \mathbf{0.36091325418001} & 0.0303494022034 \\ 0.000007657273830 & 0.008293393618634 & 0.0000346485386433 & \mathbf{0.55279157046563} & 0.0303324666473 \\ 0.000000051209296 & 0.044218944305554 & 0.0165858723488658 & 0.0002898699267899 & \mathbf{0.12041188377} \\ 0.000343063367497 & 0.019728025120072 & \mathbf{0.22664219457110} & \mathbf{0.20883023885464} & 0.0002187148763 \end{pmatrix} \tag{83}$$

$$\alpha_2 = \begin{pmatrix} 0.0247169438667 & 0.045938324942878493 & \mathbf{0.52035195378729} & 0.0015976654768 & 0.0219865625857849 \\ \mathbf{0.10369500283} & 0.000000961851428240 & 0.0058603752158104 & \mathbf{0.17159388407} & 0.0001956826269151 \\ 0.0619247685514 & 0.005680420895898976 & 0.0000041940337011 & 0.0009132788022 & 0.0161550464515489 \\ 0.0073308612563 & \mathbf{0.3760898786954499} & 0.0078995090167169 & 0.0000971358022 & 0.0022020712790430 \\ \mathbf{0.37860663035} & \mathbf{0.8648532461379836} & 0.0096939577784123 & \mathbf{0.23909856627} & 0.0000001318796171 \end{pmatrix} \tag{84}$$

$$\beta_1 = \begin{pmatrix} 2.02691486662775 & 4.58853278669795 & 9.21516653991608 & 14.2039223554899 & 17.7230908440328108 \\ 2.30228990848878 & 5.70815142794409 & 9.75920981324501 & 15.0047495693597 & 17.1640776964259771 \\ 2.71360844613891 & 6.97390906252072 & 10.9112224210093 & 16.3935104902520 & 17.3801721025480269 \\ 3.18861359927744 & 6.93702281997507 & 12.0261860231254 & 17.5228876305459 & 17.8876296984556440 \\ 3.95262799649030 & 7.76155541730819 & 13.5039942724633 & 17.3549525971848 & 18.0730780733303966 \end{pmatrix} \tag{85}$$

$$\beta_2 = \begin{pmatrix} 19.6811983441165 & 20.56326127197891 & 18.53440853276660 & 11.10183435325997 & 5.955287687038747 \\ 20.2253306600591 & 21.39051471260508 & 16.97184115533537 & 9.548598696946248 & 5.459761230875715 \\ 20.2208259457254 & 22.20704300748698 & 17.88989095276187 & 8.724870367131993 & 4.215302773261564 \\ 19.7356631996375 & 21.67330389603866 & 15.76838788843381 & 7.534795006501931 & 3.517163899772246 \\ 20.2972304557004 & 19.06667927692781 & 13.19618799557176 & 6.812943703872132 & 2.825437512911523 \end{pmatrix} \tag{86}$$

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] 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.

[5] P. Embrechts, T. Liniger and L. Lin. Multivariate hawkes processes: an application to financial data. *Journal of Applied Probability*, 48:367–378, 2011.

[6] R.F. Engle and J.R. Russell. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, :1127–1162, 1998.

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

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

[9] T.J. Liniger. *Multivariate Hawkes Processes*. PhD thesis, Swiss Federal Institute Of Technology Zurich, 2009.

[10] 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.

[11] 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.

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

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

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

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

[16] HM Wiseman. Quantum theory of continuous feedback. *Physical Review A*, 49(3):2133, 1994.