Point Process Models for Multivariate High-Frequency Irregularly Spaced Data

11-29-2012

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) and a hybrid model combining the ACD and the Hawkes models. Data from the symbol SPY on three different electronic markets is used to estimate model parameters and generate illustrative plots.

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 $\theta(t) = \begin{cases} 0 & t < 0 \\ 1 & t \geqslant 0 \end{cases}$

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

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

$$\lambda^k(t|F_t) = \lim_{\Delta t \to 0} \frac{\Pr(N^k_{t+\Delta t} - N^k_t > 0 | F_t)}{\Delta t} \tag{2}$$

For small values of Δt we have

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

so that

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

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

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

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

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

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 $x_k = t^k_i - t^k_{i-1}$ 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

$$S_k(x^k_i) = P_k(T^k_i > x^k_i | F_{t_{i-1}+\tau}) \tag{8}$$

Let

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

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

$$S_k(x^k_i) = e^{-\int_{t_{i-1}}^{t_i} \lambda^k(t|F_t)\mathrm{d}t} = e^{-\tilde{\mathcal{E}}^k_{N(t)}} \tag{9}$$

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

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

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*[2, 2.1] 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

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

1.2. The Autoregressive Conditional Duration(ACD) Model.

Letting p_i be the family of conditional probability density functions for arrival time t_i, the log likelihood of the ACD model can be expressed in terms of the conditional densities or intensities as

$$\begin{aligned} \ln\mathcal{L} &= \sum_{i=1}^{N_t} \log p_i(t_i|t_0,\ldots,t_{i-1}) \\ &= \sum_{i=1}^{N_t} \log\lambda(t_i|i-1,t_0,\ldots,t_{i-1}) - \int_{t_0}^{T} \lambda(u|N_u,t_0,\ldots,t_{N_u})\mathrm{d}u \end{aligned} \tag{13}$$

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

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

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

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

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 $\omega \geqslant 0, \alpha_j \geqslant 0, \beta_j \geqslant 0$ so the conditional intensity is then written

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

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

$$\ln \mathcal{L}(\{x_i\}_{i=1,\ldots,n}) = \sum_{i=\max(m,q)}^{n} \ln\left(\frac{e^{-\frac{x_i}{\psi_i}}}{\psi_i}\right) \tag{22}$$

Note that we start the summation at max $(m,\ q)$ because of the required "burn in" steps, an alternative method if one wished to be pedantic would be to back-cast these "missing values", which would probably not result in much gain if a reasonable number of observations are gathered, that is, n is high enough. 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

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

1.3. The Autoregressive Conditional Intensity Model.

1.3.1. The ACI(1,1) Model.

Let the conditional intensity function for process k be given by the non-negative function

$$\lambda^k(t|F_t) = \omega_k e^{\phi^k_{N(t)}} \tag{25}$$

where $\omega_k > 0$ and $\phi^k_{N(t)}$ is a measurable function of the bivariate filtration of all past arrival times. [1, 4.2] Since $\phi^k_{N(t)}$ is time-invariant between arrivals in the pooled process it is therefore indexed by the associated counting process. Define the vector

$$\phi_{N(t)} = \begin{pmatrix} \phi^a_{N(t)} \\ \phi^b_{N(t)} \end{pmatrix} \tag{26}$$

In this bivariate setting, each arrival can be one of two types. Let y_i be the indicator variable

$$y_i = \begin{cases} 0 & i-\text{th event is of type } a \\ 1 & i-\text{th event is of type } b \end{cases} \tag{27}$$

The parameterization proposed by [11] is

$$\phi_{N(t)} = \begin{cases} \alpha_a \mathcal{E}^a_{N(t)-1} + B\phi_{N(t)-1} & \text{if } y_{N(t)-1} = 0 \\ \alpha_b \mathcal{E}^b_{N(t)-1} + B\phi_{N(t)-1} & \text{if } y_{N(t)-1} = 1 \end{cases} \tag{28}$$

or equivalently

$$\phi_{N(t)} = (\alpha_a + (\alpha_b - \alpha_a)\, y_{N(t)-1})\mathcal{E}_{N(t)-1} + B\phi_{N(t)-1} \tag{29}$$

where ω, α_a and α_b are 2-dimensional parameter vectors, B is a 2×2 matrix, and $\mathcal{E}_{N(t)}$ is an i.i.d. unit exponential random variable given by

$$\mathcal{E}_{N(t)} = \begin{cases} \mathcal{E}^a_{N(t)} & \text{if } y_{N(t)} = 1 \\ \mathcal{E}^b_{N(t)} & \text{if } y_{N(t)} = 0 \end{cases} \tag{30}$$

where the generalized residuals are

$$\begin{aligned}\mathcal{E}_i^a &= 1 - \int_{t_{i-1}^a}^{t_i^a} \lambda^a(t|F_t)\mathrm{d}t \\ &= 1 - \int_{t_{i-1}^a}^{t_i^a} \omega_a\, e^{\phi_{N(t)}^a}\mathrm{d}t \\ &= 1 - \int_{t_{i-1}^a}^{t_i^a} \omega_a\, e^{\alpha_a \mathcal{E}_{N(t)-1}^a + B\phi_{N(t)-1}}\mathrm{d}t\end{aligned} \tag{31}$$

and

$$\mathcal{E}_i^b = 1 - \int_{t_{i-1}^b}^{t_i^b} \lambda^b(t|F_t)\mathrm{d}t = 1 - \int_{t_{i-1}^b}^{t_i^b} \omega_b\, e^{\phi_{N(t)}^b}\mathrm{d}t \tag{32}$$

If the $N(t)$-th arrival was of type a then $\mathcal{E}_{N(t)} = \mathcal{E}_{N^a(t)}^a$. We see that $\phi_{N(t)}$ is a weighted-average of its most recent value $\phi_{N(t)-1}$ and the error term $\mathcal{E}_{N(t)-1}$ and in this way the model has Kalman-filter like properties. If B is restricted to be diagonal then the model is called a Diagonal Autoregressive Conditional Intensity model. By rearranging terms (29) can be rewritten as

$$(I - BL)\phi_{N(t)} = (\alpha_a + (\alpha_b - \alpha_a)y_{N(t)-1})\mathcal{E}_{N(t)-1} \tag{33}$$

If the eigenvalues of B lie inside the unit circle then (29) can be written as infinite moving average

$$\phi_{N(t)} = \sum_{j=1}^{\infty} B^{j-1}(\alpha_a + \alpha_b^* y_{N(t)-j})\mathcal{E}_{N(t)-j} \tag{34}$$

The compensator for this parametization is given by

$$\begin{aligned}\Lambda^k(s_0, s_1) &= \int_{s_0}^{s_1} \lambda^k(t|F_t)\mathrm{d}t \\ &= \int_{s_0}^{s_1} \omega_k\, e^{\phi_{N(t)}^k}\mathrm{d}t\end{aligned} \tag{35}$$

1.3.2. Maximum Likelihood Estimation.

For a bivariate model that requires joint estimation of both processes the likelihood is expressed as

$$L = e^{-(\Lambda^a(0,T) + \Lambda^b(0,T))} \prod_{i=1}^{N^a(t)} \lambda^a(t_i^a|F_t) \prod_{i=1}^{N^b(t)} \lambda^b(t_i^b|F_t) \tag{36}$$

For a general K-variate model the likelihood is expressed as

$$L = e^{-\sum_{k=1}^{K}\Lambda^k(0,T)} \prod_{k=1}^{K} \prod_{i=1}^{N^k(t)} \lambda^k(t_i^k|F_t) \tag{37}$$

Due to the necessity of numerical integration, likelihood astimation for ACI processes tends to be complicated and laborious to implement in code. Diagnostic testing for this model is discussed in [7].

1.4. The Hawkes Process.

1.4.1. Linear Self-Exciting Processes.

A (univariate) linear self-exciting (counting) process N_t is one that can be expressed as [12]

$$\begin{aligned}\lambda(t) &= \lambda_0(t) + \int_{-\infty}^{t} \nu(t-s)\mathrm{d}N_s \\ &= \lambda_0(t) + \sum_{t_i < t} \nu(t - t_i)\end{aligned} \tag{38}$$

where $\lambda_0(t)$ is a deterministic base intensity and $\nu\colon \mathbb{R}_+ \to \mathbb{R}_+$ expresses the positive influence of past events t_i on the current value of the intensity process. The Hawkes process of order P is a linear self-exciting process defined by the exponential kernel

$$\nu(t) = \sum_{j=1}^{P} \alpha_j e^{-\beta_j t} \tag{39}$$

so that the intensity is written as

$$\begin{aligned} \lambda(t) &= \lambda_0(t) + \int_0^t \sum_{j=1}^{P} \alpha_j e^{-\beta_j (t-s)} \mathrm{d}N_s \\ &= \lambda_0(t) + \sum_{i=0}^{N_t} \sum_{j=1}^{P} \alpha_j e^{-\beta_j (t-t_i)} \end{aligned} \tag{40}$$

A univariate Hawkes process is stationary if

$$\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} < 1 \tag{41}$$

If a Hawkes process is stationary then the unconditional mean is

$$\begin{aligned} \mu = E[\lambda(t)] &= \frac{\lambda_0}{1 - \int_0^\infty \nu(t)\mathrm{d}t} \\ &= \frac{\lambda_0}{1 - \int_0^\infty \sum_{j=1}^{P} \alpha_j e^{-\beta_j t}\mathrm{d}t} \\ &= \frac{\lambda_0}{1 - \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}} \end{aligned} \tag{42}$$

For consecutive events, we have the compensator (7)

$$\begin{aligned} \Lambda(t_{i-1}, t_i) &= \int_{t_{i-1}}^{t_i} \lambda(t)\mathrm{d}t \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \sum_{k=0}^{i-1} \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} \big(e^{-\beta_j (t_{i-1}-t_k)} - e^{-\beta_j (t_i - t_k)}\big) \\ &= \int_{t_{i-1}}^{t_i} \lambda_0(s)\mathrm{d}s + \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} \big(1 - e^{-\beta_j (t_i - t_{i-1})}\big) A_j(i-1) \end{aligned} \tag{43}$$

where there is the recursion

$$\begin{aligned} A_j(i-1) &= \sum_{t_k \leqslant t_{i-1}} e^{-\beta_j (t_{i-1} - t_k)} \\ &= \sum_{k=0}^{i-2} e^{-\beta_j (t_{i-1} - t_k)} \\ &= 1 + e^{-\beta_j (t_{i-1} - t_{i-2})} A_j(i-2) \end{aligned} \tag{44}$$

with $A_j(0) = 0$. If $\lambda_0(t) = \lambda_0$ then (43) simplifies to

$$\begin{aligned} \Lambda(t_{i-1}, t_i) &= (t_i - t_{i-1})\lambda_0 + \sum_{k=0}^{i-1} \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} \big(e^{-\beta_j (t_{i-1}-t_k)} - e^{-\beta_j (t_i - t_k)}\big) \\ &= (t_i - t_{i-1})\lambda_0 + \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} \big(1 - e^{-\beta_j (t_i - t_{i-1})}\big) A_j(i-1) \end{aligned} \tag{45}$$

1.4.2. The Hawkes(1) Model.

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

$$\lambda(t)=\lambda_0+\sum_{t_i<t}\alpha e^{-\beta(t-t_i)} \tag{46}$$

which has the unconditional mean

$$E[\lambda(t)]=\frac{\lambda_0}{1-\frac{\alpha}{\beta}} \tag{47}$$

1.4.3. Maximum Likelihood Estimation.

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

$$\ln\mathcal{L}(N(t)_{t\in[0,T]})=\int_0^T(1-\lambda(s))\mathrm{d}s+\int_0^T\ln\lambda(s)\mathrm{d}N_s \tag{48}$$

which in the case of the Hawkes model of order P can be explicitly written [10] as

$$\begin{aligned}\ln\mathcal{L}(\{t_i\}_{i=1\ldots n}) &=T-\Lambda(0,T)+\sum_{i=1}^n\ln\lambda(t_i)\\ &=T-\Lambda(0,T)+\sum_{i=1}^n\ln\left(\lambda_0(t_i)+\sum_{j=1}^P\sum_{k=1}^{i-1}\alpha_j e^{-\beta_j(t_i-t_k)}\right)\\ &=T-\Lambda(0,T)+\sum_{i=1}^n\ln\left(\lambda_0(t_i)+\sum_{j=1}^P\alpha_j R_j(i)\right)\\ &=T-\int_0^T\lambda_0(s)\mathrm{d}s-\sum_{i=1}^n\sum_{j=1}^P\frac{\alpha_j}{\beta_j}\left(1-e^{-\beta_j(t_n-t_i)}\right)\\ &+\sum_{i=1}^n\ln\left(\lambda_0(t_i)+\sum_{j=1}^P\alpha_j R_j(i)\right)\end{aligned} \tag{49}$$

where $T=t_n$ and we have the recursion[9]

$$\begin{aligned}R_j(i) &=\sum_{k=1}^{i-1}e^{-\beta_j(t_i-t_k)}\\ &=e^{-\beta_j(t_i-t_{i-1})}(1+R_j(i-1))\end{aligned} \tag{50}$$

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

$$\begin{aligned}\ln\mathcal{L}(\{t_i\}_{i=1\ldots n}) &=T-\lambda_0T-\sum_{i=1}^n\sum_{j=1}^P\frac{\alpha_j}{\beta_j}\left(1-e^{-\beta_j(t_n-t_i)}\right)\\ &+\sum_{i=1}^n\ln\left(\lambda_0+\sum_{j=1}^P\alpha_j R_j(i)\right)\end{aligned} \tag{51}$$

Note that it was necessary to shift each t_i by t_1 so that $t_1=0$ and $t_n=T$.

1.5. Combining the ACD and Hawkes Models.

The ACD and Hawkes models can be combined to provide a model for intraday volatility. [3] Let

$$\lambda(t) =\frac{1}{\psi_{N_t}}+\int_0^t\nu(t-s)\mathrm{d}N_s \tag{52}$$

where the ACD(20) part is

$$\psi_i=\omega+\sum_{j=1}^m\alpha_j x_{i-j}+\sum_{j=1}^q\beta_j\psi_{i-j} \tag{53}$$

and the Hawkes part has the exponential kernel(39)

$$\begin{aligned}\int_0^t \nu(t-s)\mathrm{d}N_s &= \int_0^t \sum_{j=1}^{P} \gamma_j e^{-\varphi_j(t-s)}\mathrm{d}N_s \\ &= \sum_{k=0}^{N_t}\sum_{j=1}^{P} \gamma_j e^{-\varphi_j(t-t_k)}\end{aligned} \tag{54}$$

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. Then we have

$$\begin{aligned}\lambda(t) &= \frac{1}{\omega + \sum_{j=1}^{m} \alpha_j x_{N_t-j} + \sum_{j=1}^{q} \beta_j \psi_{N_t-j}} + \int_0^t \sum_{j=1}^{P} \gamma_j e^{-\varphi_j(t-s)}\mathrm{d}N_s \\ &= \frac{1}{\omega + \sum_{j=1}^{m} \alpha_j x_{N_t-j} + \sum_{j=1}^{q} \beta_j \psi_{N_t-j}} + \sum_{k=0}^{N_t}\sum_{j=1}^{P} \gamma_j e^{-\varphi_j(t-t_k)}\end{aligned} \tag{55}$$

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

$$\begin{aligned}\ln \mathcal{L}(\{t_i\}_{i=1,..,n}) &= \sum_{i=1}^{n} \left(\ln \lambda(t_i) - \int_{t_{i-1}}^{t_i} \lambda(t)\mathrm{d}t \right) \\ &= \sum_{i=1}^{n} (\ln \lambda(t_i) - \Lambda(t_{i-1}, t_i))\end{aligned} \tag{56}$$

By direct calculation, combining (19) and (43) we have the compensator

$$\begin{aligned}\Lambda(t_{i-1}, t_i) &= \int_{t_{i-1}}^{t_i} \lambda(t)\mathrm{d}t \\ &= \int_{t_{i-1}}^{t_i} \frac{1}{\psi_{N_t}} + \int_0^t \nu(t-s)\mathrm{d}N_s\mathrm{d}t \\ &= \frac{t_i - t_{i-1}}{\psi_i} + \int_{t_{i-1}}^{t_i}\int_0^t \nu(t-s)\mathrm{d}N_s\mathrm{d}t \\ &= \frac{t_i - t_{i-1}}{\psi_i} + \sum_{k=0}^{i-1}\sum_{j=1}^{P} \frac{\gamma_j}{\varphi_j}(e^{-\varphi_j(t_{i-1}-t_k)} - e^{-\varphi_j(t_i-t_k)}) \\ &= \frac{t_i - t_{i-1}}{\psi_i} + \sum_{j=1}^{P} \frac{\gamma_j}{\varphi_j}(1 - e^{-\varphi_j(t_i - t_{i-1})})A_j(i-1)\end{aligned} \tag{57}$$

where ψ_i is defined by (53) and by construction $E\left[\frac{t_i - t_{i-1}}{\psi_i}\right] = 1$ and $A_j(i)$ is given by (44) so that (56) can be wriitten as

$$\begin{aligned}\ln \mathcal{L}(\{t_i\}_{i=1,..,n}) &= \sum_{i=1}^{n} (\ln \lambda(t_i) - \Lambda(t_{i-1}, t_i)) \\ &= \sum_{i=1}^{n} \left(\log \lambda(t_i) - \left(\frac{t_i - t_{i-1}}{\psi_i} + \sum_{j=1}^{P} \frac{\gamma_j}{\varphi_j}(1 - e^{-\varphi_j(t_i - t_{i-1})})A_j(i-1) \right) \right) \\ &= \sum_{i=1}^{n} \log\left(\frac{1}{\psi_i} + \sum_{k=1}^{N_{t_i}}\sum_{j=1}^{P} \gamma_j e^{-\varphi_j(t_i - t_k)} \right) - \left(\frac{t_i - t_{i-1}}{\psi_i} + \sum_{j=1}^{P} \frac{\gamma_j}{\varphi_j}(1 - e^{-\varphi_j(t_i - t_{i-1})})A_j(i-1) \right)\end{aligned} \tag{58}$$

1.6. Multivariate Hawkes Models.

Let $M \in \mathbb{N}^*$ and $\{(t_i^m)\}_{m=1,\ldots,M}$ be an M-dimensional point process. The associated counting process will be denoted $N_t = (N_t^1, \ldots, N_t^M)$. A multivariate Hawkes process[5][8] is defined with intensities $\lambda^m, m = 1\ldots M$ given by

$$\lambda^m(t) = \lambda_0^m(t) + \sum_{n=1}^{M} \int_0^t \sum_{j=1}^{P} \alpha_j^{m,n} e^{-\beta_j^{m,n}(t-s)} \mathrm{d}N_s^n \tag{59}$$

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

$$\begin{aligned} \lambda^m(t) &= \lambda_0^m + \sum_{n=1}^{M} \int_0^t \alpha^{m,n} e^{-\beta^{m,n}(t-s)} \mathrm{d}N_s^n \\ &= \lambda_0^m + \sum_{n=1}^{M} \sum_{t_i^n < t} \alpha^{m,n} e^{-\beta^{m,n}(t-t_i^n)} \end{aligned} \tag{60}$$

Rewriting (60) in vectorial notion, we have

$$\lambda(t) = \lambda_0 + \int_0^t G(t-s) \mathrm{d}N_s \tag{61}$$

where

$$G(t) = (\alpha^{m,n} e^{-\beta^{m,n}(t-s)})_{m,n=1\ldots M} \tag{62}$$

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

$$\begin{aligned} \mu &= \frac{\lambda_0}{I - \int_0^\infty G(u)\mathrm{d}u} \\ &= \frac{\lambda_0}{I - (\frac{\alpha^{m,n}}{\beta^{m,n}})} \\ &= \frac{\lambda_0}{I - \Gamma} \end{aligned} \tag{63}$$

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

$$\Gamma = \int_0^\infty G(s)\mathrm{d}s = \frac{\alpha^{m,n}}{\beta^{m,n}} \tag{64}$$

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

$$\rho(G) = \max_{a \in \mathcal{S}(G)} |a| \tag{65}$$

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

$$\begin{aligned} \Lambda^m(t_{i-1}^m, t_i^m) &= \int_{t_{i-1}^m}^{t_i^m} \lambda^m(s)\mathrm{d}s \\ &= \int_{t_{i-1}^m}^{t_i^m} \lambda_0^m(s)\mathrm{d}s \\ &+ \sum_{n=1}^{M} \sum_{j=1}^{P} \sum_{t_k^n < t_{i-1}^m} \frac{\alpha_j^{m,n}}{\beta_j^{m,n}} [e^{-\beta_j^{m,n}(t_{i-1}^m - t_k^n)} - e^{-\beta_j^{m,n}(t_i^m - t_k^n)}] \\ &+ \sum_{n=1}^{M} \sum_{j=1}^{P} \sum_{t_{i-1}^m \leqslant t_k^n < t_i^m} \frac{\alpha_j^{m,n}}{\beta_j^{m,n}} [1 - e^{-\beta_j^{m,n}(t_i^m - t_k^n)}] \end{aligned} \tag{66}$$

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

$$\begin{aligned} A_j^{m,n}(i-1) &= \sum_{t_k^n < t_{i-1}^m} e^{-\beta_j^{m,n}(t_{i-1}^m - t_k^n)} \\ &= e^{-\beta_j^{m,n}(t_{i-1}^m - t_{i-2}^m)} A_j^{m,n}(i-2) + \sum_{t_{i-2}^m \leqslant t_k^n < t_{i-1}^m} e^{-\beta_j^{m,n}(t_{i-1}^m - t_k^n)} \end{aligned} \tag{67}$$

and rewrite (66) as

$$\begin{aligned} \Lambda^m(t_{i-1}^m, t_i^m) &= \int_{t_{i-1}^m}^{t_i^m} \lambda_0^m(s)\mathrm{d}s \\ &+ \sum_{n=1}^{M} \sum_{j=1}^{P} \frac{\alpha_j^{m,n}}{\beta_j^{m,n}} \left[(1 - e^{-\beta_j^{m,n}(t_i^m - t_{i-1}^m)}) \times A_j^{m,n}(i-1) + \sum_{t_{i-1}^m \leqslant t_k^n < t_i^m} (1 - e^{-\beta_j^{m,n}(t_i^m - t_k^n)}) \right] \end{aligned} \tag{68}$$

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

1.6.1. Log-Likelihood.

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

$$\ln \mathcal{L}(\{t_i\}_{i=1,\ldots,N_T^m}) = \sum_{m=1}^{M} \ln \mathcal{L}^m(\{t_i\}) \tag{69}$$

where each term is defined by

$$\ln\mathcal{L}^m(\{t_i\}) = \int_0^T (1 - \lambda^m(s))\mathrm{d}s + \int_0^T \ln\lambda^m(s)\mathrm{d}N_s^m \tag{70}$$

which in this case can be written as

$$\ln \mathcal{L}^m(\{t_i\}) = T - \Lambda^m(0,T) + \sum_{i=1}^{N} z_i^m \ln\left[\lambda_0^m(t) + \sum_{n=1}^{M} \sum_{j=1}^{P} \sum_{t_k^n < t_i} \alpha_j^{m,n} e^{-\beta_j^{m,n}(t_i - t_k^n)}\right. \tag{71}$$

where again $t_{N_T^m} = T$ and

$$z_i^m = \begin{cases} 1 & \text{event } t_i \text{ of type } m \\ 0 & \text{otherwise} \end{cases} \tag{72}$$

Similiar to to the one-dimensional case, we have the recursion

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

so that (71) can be rewritten as

$$\begin{aligned} \ln \mathcal{L}^m(\{t_i\}) &= T - \sum_{i=1}^{N_T^m} \sum_{n=1}^{M} \sum_{j=1}^{P} \frac{\alpha_j^{m,n}}{\beta_j^{m,n}} (1 - e^{-\beta_j^{m,n}(T - t_i)}) \\ &+ \sum_{i=1}^{N_T^m} \ln\left[\lambda_0^m(t_i^m) + \sum_{n=1}^{M} \sum_{j=1}^{P} \alpha_j^{m,n} R_j^{m,n}(i)\right] \end{aligned} \tag{74}$$

with initial conditions $R_j^{m,n}(0) = 0$ and 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.

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.

3.1.1. 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 Oct 22nd, 2012. The unconditional sample mean intensity for this symbol on this day on this exchange was 0.7655998283415355 trades per second where the number of samples is $n = 17916$. No deasonalization was attempted, which would surely benefit the results; this will be reserved for future work. 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. Estimation for $P = 7$ and greater was attempted but the optimizer kept settling on prior solutions by taking some α parameters to 0 thus essentially reducing the order of the model. Standard deviations are not provided, but presumably they could be estimated with derivative information.

P	λ_0	$\alpha_{1\ldots P}$	$\beta_{1\ldots P}$	$\ln \mathcal{L}(\{t_i\}_{i=1\ldots n}) - t_n$	$E[\lambda(t)]$
1	0.4888895840	5.4436229616	15.0588031220	−14606.0079680	0.76567384816
2	0.13718922357	7.2188754084 0.0782472258	25.399826568 0.1454607237	−12733.4619196	0.77131730144
3	0.13163151059	0.0000000003 7.5467174975 0.0677609554	28.852294270 23.166515568 0.1276584845	−12506.0576338	0.917666203197
4	0.13296929140	0.0723686778 1.8881451880 5.1594817028 0.2982510629	0.1349722452 16.637110622 30.626390900 32.490874482	−12716.5362393	0.769984967876
5*	0.06084821553	0.0000055317 7.6260052075 0.1866285010 0.0000939392 0.0101541140	0.5138236561 29.316263593 0.7694261263 0.0693359346 0.0241678794	−12505.9421508	0.802736706908
6*	0.04014430354	7.6812049064 0.0000040868 0.0282570213 0.1970449132 0.0314334590 0.0027981168	30.467204143 7.5984574690 0.1178289377 1.2119099089 4.7015553402 0.0096010396	−12478.0771035	0.847703217380

*=The exp/ln transform was used to ensure positivity of parameters of the estimate whereas absolute value was used for the others, this resulted in the search point getting over local minima to achieve better likelihood.

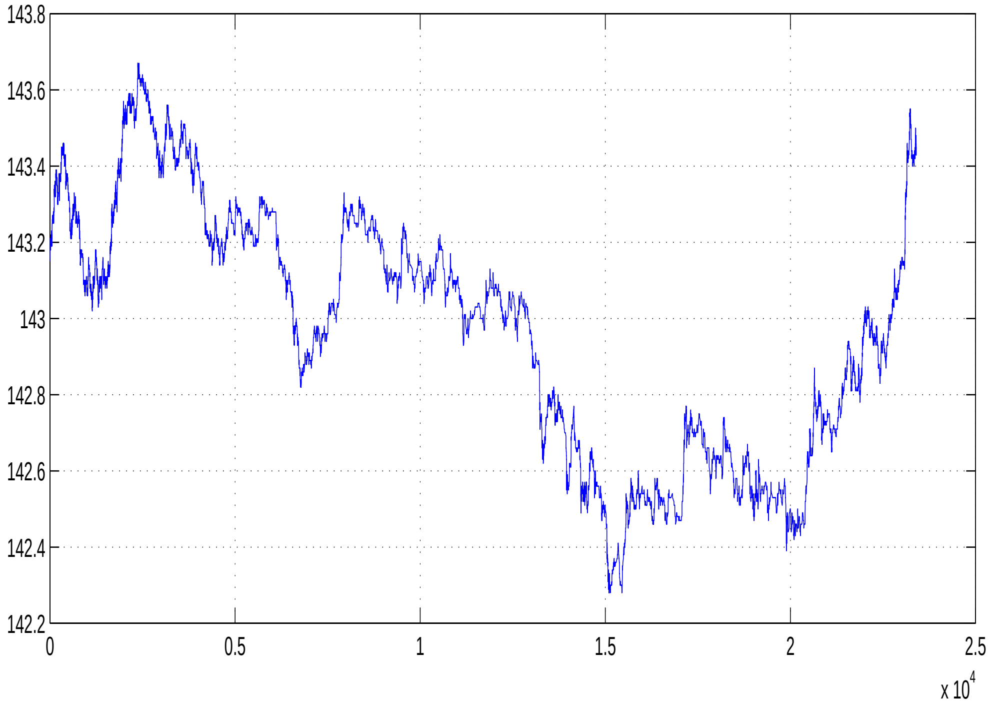

Figure 1. Price history for SPY traded on INET on Oct 22nd, 2012

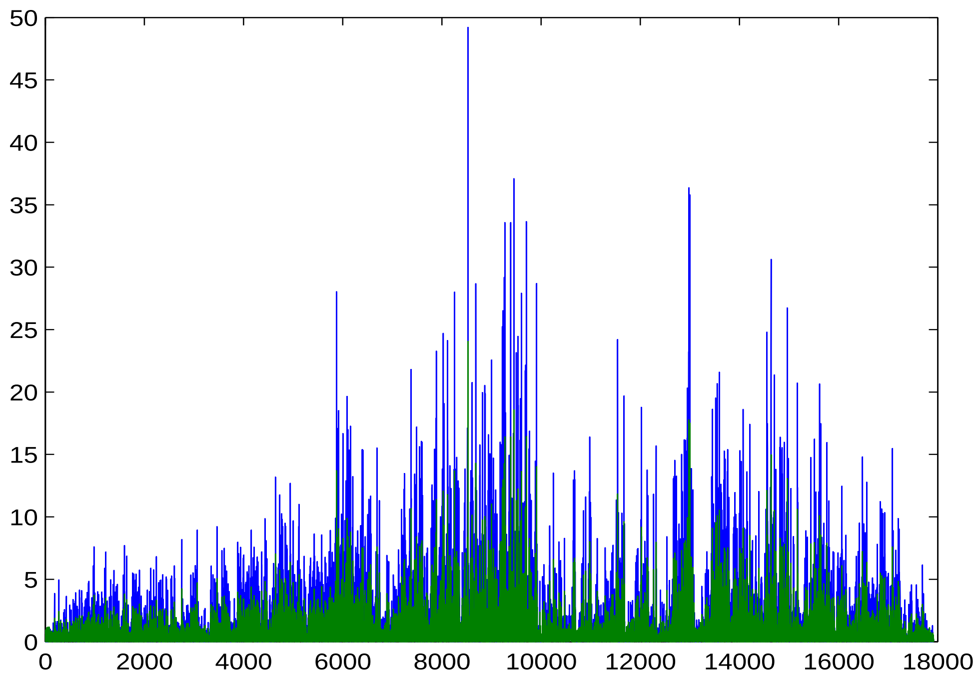

Figure 2. $x_i = t_i - t_{i-1}$ in blue and $\{\Lambda(t_{i-1}, t_i): P = 1\}$ in green

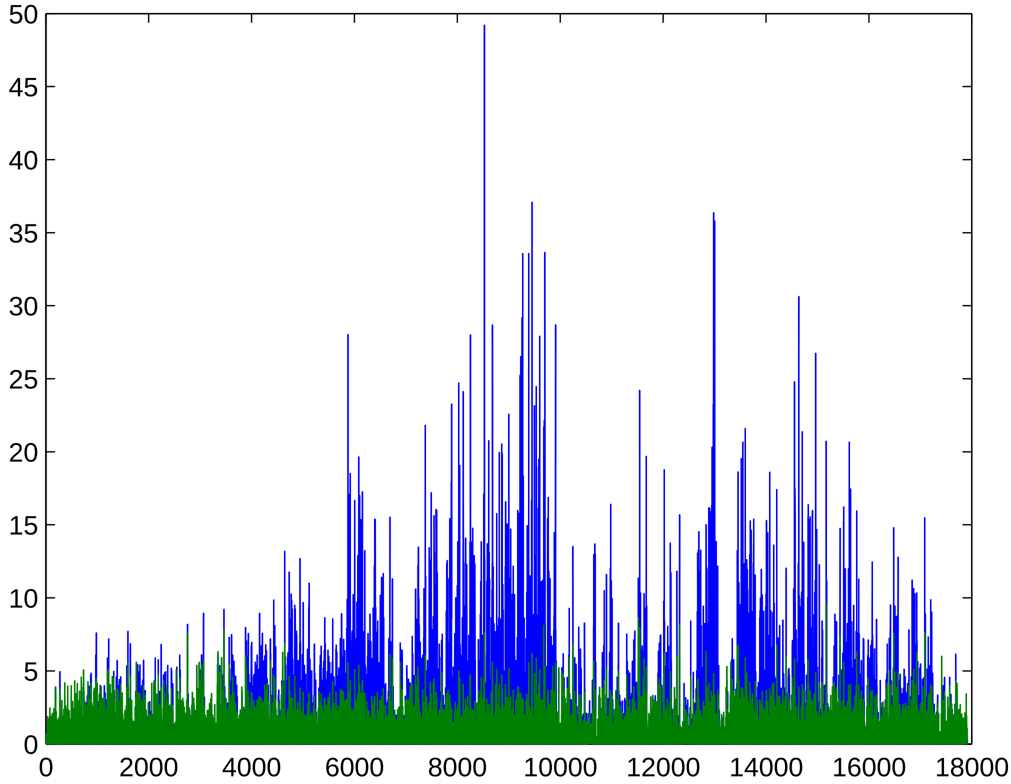

Figure 3. $x_i = t_i - t_{i-1}$ in blue and $\{\Lambda(t_{i-1}, t_i): P = 6\}$ in green

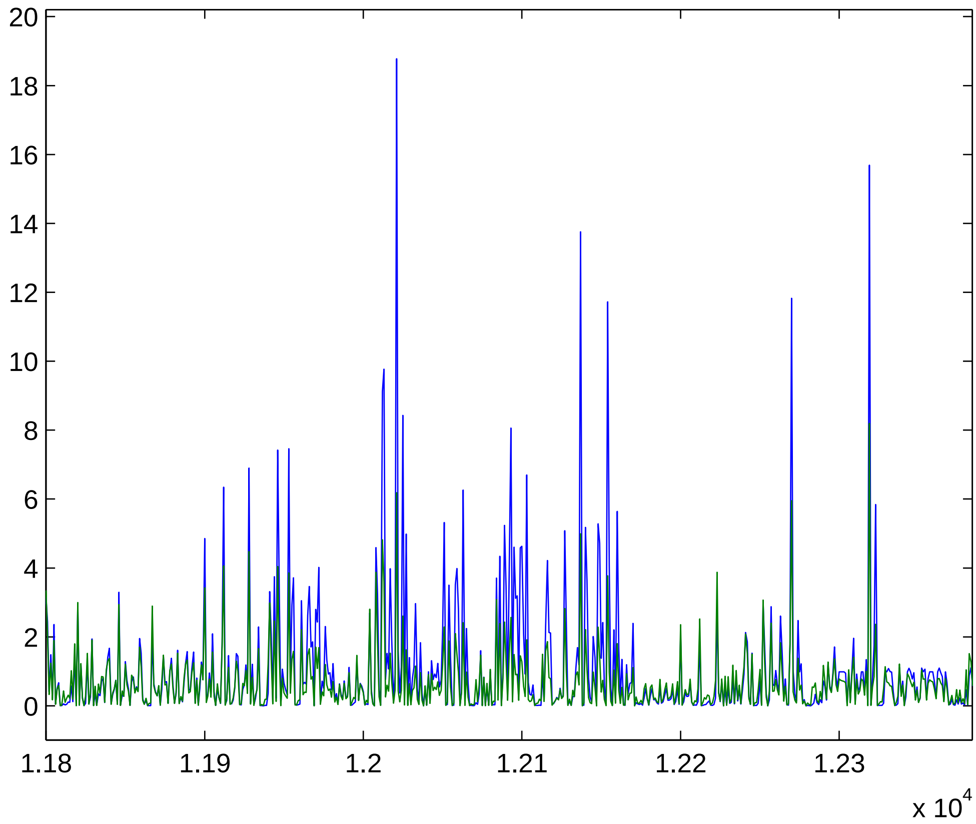

Figure 4. Zoomed in view of $x_i = t_i - t_{i-1}$ in blue and $\{\Lambda(t_{i-1}, t_i) \colon P = 6\}$ in green

3.1.2. 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

(75)

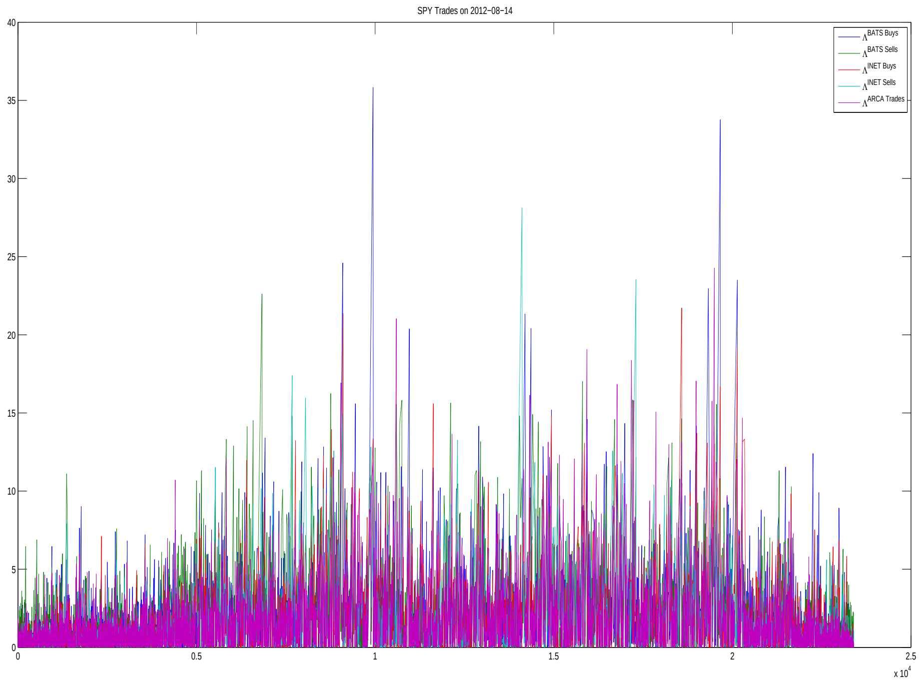

Figure 5.

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.

$$\lambda = \begin{pmatrix} 0.25380789517348 \\ 0.269289236349466 \\ 0.221292886522613 \\ 0.158954542395839 \\ 0.371572853723448 \end{pmatrix} \tag{76}$$

$$\alpha = \begin{pmatrix} 4.3514\times10^{-9} & 0.011879 & \mathbf{0.2648} & 1.917\times10^{-8} & \mathbf{0.10771} \\ 0.021881 & 2.6164\times10^{-8} & 2.5725\times10^{-8} & 0.024946 & \mathbf{0.25138} \\ \mathbf{0.29092} & \mathbf{0.51715} & 1.1254\times10^{-8} & 0.0029919 & 0.004607 \\ 0.0041449 & \mathbf{0.52852} & 0.018077 & 3.2535\times10^{-9} & 0.0237 \\ 0.021501 & \mathbf{0.71358} & \mathbf{1.0954} & \mathbf{0.15264} & 4.1222\times10^{-9} \end{pmatrix} \tag{77}$$

$$\beta = \begin{pmatrix} 1.0954 & 10.803 & 16.665 & 20.188 & 9.6059 \\ 5.6238 & 11.558 & 16.721 & 18.304 & 7.9016 \\ 7.8125 & 15.299 & 16.431 & 14.702 & 6.6458 \\ 8.3083 & 15.758 & 17.749 & 12.953 & 3.1621 \\ 9.4264 & 16.369 & 19.303 & 11.071 & 2.8302 \end{pmatrix} \tag{78}$$

with a log-likelihood score of 39714.1497.

3.1.3. Multivariate SPY Data for 2012-11-19.

Consider the same symbol, SPY, as a 5-dimensional Hawkes process as in 3.1.2, 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.

$$\lambda = \begin{pmatrix} 0.113371928486215301 \\ 0.116069526955243113 \\ 0.120010488406567112 \\ 0.140864383337674315 \\ 0.236370243964866722 \end{pmatrix} \tag{79}$$

$$\alpha_1 = \begin{pmatrix} 0.000000400520039 & 0.000743243048280 & 0.0730760324025721 & 0.0235425002925593 & \mathbf{0.14994903109} \\ 0.000836306407254 & 0.000048087752871 & 0.0009983197029208 & \mathbf{0.36091325418001} & 0.0303494022034 \\ 0.000007657273830 & 0.008293393618634 & 0.0000346485386433 & \mathbf{0.55279157046563} & 0.0303324666473 \\ 0.000000051209296 & 0.044218944305554 & 0.0165858723488658 & 0.0002898699267899 & \mathbf{0.12041188377} \\ 0.000343063367497 & 0.019728025120072 & \mathbf{0.22664219457110} & \mathbf{0.20883023885464} & 0.0002187148763 \end{pmatrix} \tag{80}$$

$$\alpha_2 = \begin{pmatrix} 0.0247169438667 & 0.045938324942878493 & \mathbf{0.52035195378729} & 0.0015976654768 & 0.0219865625857849 \\ \mathbf{0.10369500283} & 0.000000961851428240 & 0.0058603752158104 & \mathbf{0.17159388407} & 0.0001956826269151 \\ 0.0619247685514 & 0.005680420895898976 & 0.0000041940337011 & 0.0009132788022 & 0.0161550464515489 \\ 0.0073308612563 & \mathbf{0.3760898786954499} & 0.0078995090167169 & 0.0000971358022 & 0.0022020712790430 \\ \mathbf{0.37860663035} & \mathbf{0.8648532461379836} & 0.0096939577784123 & \mathbf{0.23909856627} & 0.0000001318796171 \end{pmatrix} \tag{81}$$

$$\beta_1 = \begin{pmatrix} 2.02691486662775 & 4.58853278669795 & 9.21516653991608 & 14.2039223554899 & 17.7230908440328108 \\ 2.30228990848878 & 5.70815142794409 & 9.75920981324501 & 15.0047495693597 & 17.1640776964259771 \\ 2.71360844613891 & 6.97390906252072 & 10.9112224210093 & 16.3935104902520 & 17.3801721025480269 \\ 3.18861359927744 & 6.93702281997507 & 12.0261860231254 & 17.5228876305459 & 17.8876296984556440 \\ 3.95262799649030 & 7.76155541730819 & 13.5039942724633 & 17.3549525971848 & 18.0730780733303966 \end{pmatrix} \tag{82}$$

$$\beta_2 = \begin{pmatrix} 19.6811983441165 & 20.56326127197891 & 18.53440853276660 & 11.10183435325997 & 5.955287687038747 \\ 20.2253306600591 & 21.39051471260508 & 16.97184115533537 & 9.548598696946248 & 5.459761230875715 \\ 20.2208259457254 & 22.20704300748698 & 17.88989095276187 & 8.724870367131993 & 4.215302773261564 \\ 19.7356631996375 & 21.67330389603866 & 15.76838788843381 & 7.534795006501931 & 3.517163899772246 \\ 20.2972304557004 & 19.06667927692781 & 13.19618799557176 & 6.812943703872132 & 2.825437512911523 \end{pmatrix} \tag{83}$$

BIBLIOGRAPHY

[1] L. Bauwens and N. Hautsch. Modelling financial high frequency data using point processes. *Handbook of Financial Time Series*, :953–979, 2009.

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

[3] Y. Cai, B. Kim, M. Leduc, K. Szczegot, Y. Yixiao and M. Zamfir. A model for intraday volatility. , , 2007.

[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] S.M.S. Kwok and W.K. Li. On diagnostic checking of the autoregressive conditional intensity model. *Canadian Journal of Statistics*, 36(4):561–576, 2008.

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

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

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

[11] J.R. Russell. Econometric modeling of multivariate irregularly-spaced high-frequency data. *Manuscript, GSB, University of Chicago*, , 1999.

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