Point Process Models for Multivariate High-Frequency Irregularly Spaced Data

11-16-2012

Stephen Crowley
stephen.crowley@hushmail.com

Abstract. Definitions from the theory of point processes are recalled. Models of intensity function paramaterization and maximum likelihood estimation from data are explored. Closed-form log-likelihood expressions are given for the Hawkes process, Autoregressive Conditional Duration(ACD), and Log-ACD models. The Autoregressive Conditional Intensity model is also discussed. Data from the symbol SPY on the Nasdaq stock market on Oct 22nd, 2012 is used to estimate model parameters and generate illustrative plots.

Table of contents

1. Defintions

1.1. Point Processses and Intensities.

Consider a K dimensional multivariate point process. Let $N^k(t)$ 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 $t_{N^k(t)}^k$. 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 $\theta(t) = \begin{cases} 0 & t < 0 \\ 1 & t \geqslant 0 \end{cases}$

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

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

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

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

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

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 $\tau_k = t_i^k - t_{i-1}^k$ 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

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

Let

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

then provided the survivor function is absolutely continuous with respect to Lebesgue measure we have

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

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

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

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

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

1.2. The Autoregressive Conditional Duration Model.

Let $x_i = t_i - t_{i-1}$ be the interval between two 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

$$E(x_i|x_{i-1}, \ldots, x_1) = \psi_i(x_{i-1}, \ldots, x_1; \theta) = \psi_i \tag{13}$$

Then the ACD models are those that consist of the assumption

$$x_i = \psi_i \varepsilon_i \tag{14}$$

where ε_i is independently and identically distributed with density $p(\varepsilon;\ \phi)$ where θ and ϕ are variation free. These models are interesting but suffering from the drawback of being limited to the univariate setting. [5]

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

$$\lambda^k(t|F_t) = \omega_k e^{\phi_{N(t)}^k} \tag{15}$$

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

$$\phi_{N(t)} = \begin{pmatrix} \phi_{N(t)}^a \\ \phi_{N(t)}^b \end{pmatrix} \tag{16}$$

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

$$y_i = \begin{cases} 0 & i-\text{th event is of type } a \\ 1 & i-\text{th event is of type } b \end{cases} \tag{17}$$

The parameterization proposed by [8] is

$$\phi_{N(t)} = \begin{cases} \alpha_a \mathcal{E}_{N(t)-1}^a + B\phi_{N(t)-1} & \text{if } y_{N(t)-1} = 0 \\ \alpha_b \mathcal{E}_{N(t)-1}^b + B\phi_{N(t)-1} & \text{if } y_{N(t)-1} = 1 \end{cases} \tag{18}$$

or equivalently

$$\phi_{N(t)} = (\alpha_a + (\alpha_b - \alpha_a)\, y_{N(t)-1})\mathcal{E}_{N(t)-1} + B\phi_{N(t)-1} \tag{19}$$

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

$$\mathcal{E}_{N(t)} = \begin{cases} \mathcal{E}_{N(t)}^a & \text{if } y_{N(t)} = 1 \\ \mathcal{E}_{N(t)}^b & \text{if } y_{N(t)} = 0 \end{cases} \tag{20}$$

where the generalized residuals are

$$\begin{aligned}\mathcal{E}_i^a &= 1-\int_{t_{i-1}^a}^{t_i^a} \lambda^a(t|F_t)\mathrm{d}t\\ &= 1-\int_{t_{i-1}^a}^{t_i^a} \omega_a\, e^{\phi_{N(t)}^a}\mathrm{d}t\\ &= 1-\int_{t_{i-1}^a}^{t_i^a} \omega_a\, e^{\alpha_a\,\mathcal{E}_{N(t)-1}^a+B\phi_{N(t)-1}}\mathrm{d}t\end{aligned} \tag{21}$$

and

$$\mathcal{E}_i^b = 1-\int_{t_{i-1}^b}^{t_i^b} \lambda^b(t|F_t)\mathrm{d}t = 1-\int_{t_{i-1}^b}^{t_i^b} \omega_b\, e^{\phi_{N(t)}^b}\mathrm{d}t \tag{22}$$

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 (19) can be rewritten as

$$(I-BL)\phi_{N(t)} = (\alpha_a + (\alpha_b-\alpha_a)y_{N(t)-1})\mathcal{E}_{N(t)-1} \tag{23}$$

If the eigenvalues of B lie inside the unit circle then (19) can be written as infinite moving average

$$\phi_{N(t)} = \sum_{j=1}^{\infty} B^{j-1}(\alpha_a + \alpha_b^* y_{N(t)-j})\mathcal{E}_{N(t)-j} \tag{24}$$

The compensator for this parametization is given by

$$\begin{aligned}\Lambda^k(s_0,s_1) &= \int_{s_0}^{s_1} \lambda^k(t|F_t)\mathrm{d}t\\ &= \int_{s_0}^{s_1} \omega_k\, e^{\phi_{N(t)}^k}\mathrm{d}t\end{aligned} \tag{25}$$

1.3.2. Maximum Likelihood Estimation.

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

$$L = e^{-(\Lambda^a(0,T)+\Lambda^b(0,T))} \prod_{i=1}^{N^a(t)} \lambda^a(t_i^a|F_t) \prod_{i=1}^{N^b(t)} \lambda^b(t_i^b|F_t) \tag{26}$$

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

$$L = e^{-\sum_{k=1}^K \Lambda^k(0,T)} \prod_{k=1}^{K} \prod_{i=1}^{N^k(t)} \lambda^k(t_i^k|F_t) \tag{27}$$

Due to the necessity of numerical integration, likelihood astimation for ACI processes tends to be complicated and laborious to implement in code.

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 [9]

$$\begin{aligned}\lambda(t) &= \lambda_0(t) + \int_{-\infty}^{t} \nu(t-s)\mathrm{d}N(s)\\ &= \lambda_0(t) + \sum_{t_i<t} \nu(t-t_i)\end{aligned} \tag{28}$$

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

$$\nu(t) = \sum_{j=1}^{P} \alpha_j e^{-\beta_j t} \tag{29}$$

so that the intensity is written as

$$\begin{aligned} \lambda(t) \; &= \lambda_0(t) + \int_0^t \sum_{j=1}^{P} \alpha_j e^{-\beta_j (t-s)} \mathrm{d}N(s) \\ &= \lambda_0(t) + \sum_{t_i<t} \sum_{j=1}^{P} \alpha_j e^{-\beta_j (t-t_i)} \end{aligned} \tag{30}$$

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

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

For consecutive events, we have the compensator

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

where there is the recursion

$$\begin{aligned} A_j(i-1) \; &= \sum_{t_k \leqslant t_{i-1}} e^{-\beta_j (t_{i-1} - t_k)} \\ &= \sum_{k=0}^{i-2} e^{-\beta_j (t_{i-1} - t_k)} \\ &= 1 + e^{-\beta_j (t_{i-1} - t_{i-2})} A_j(i-2) \end{aligned} \tag{34}$$

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

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

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

$$\lambda(t)=\lambda_0+\sum_{t_i<t}\alpha e^{-\beta(t-t_i)} \tag{36}$$

which has the unconditional mean

$$E[\lambda(t)]=\frac{\lambda_0}{1-\frac{\alpha}{\beta}} \tag{37}$$

1.4.3. Maximum Likelihood Estimation.

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

$$\ln\mathcal{L}(N(t)_{t\in[0,T]})=\int_0^T(1-\lambda(s))\mathrm{d}s+\int_0^T\ln\lambda(s)\mathrm{d}N(s) \tag{38}$$

which in the case of the Hawkes(P) model can be explicitly written [7] as

$$\begin{aligned}\ln\mathcal{L}(\{t_i\}_{i=1\ldots n}) &=-\Lambda(0,t_n)+\sum_{i=1}^{n}\ln\lambda(t_i)\\ &=-\Lambda(0,t_n)+\sum_{i=1}^{n}\ln\left(\lambda_0(t_i)+\sum_{j=1}^{P}\sum_{k=1}^{i-1}\alpha_j e^{-\beta_j(t_i-t_k)}\right)\\ &=-\Lambda(0,t_n)+\sum_{i=1}^{n}\ln\left(\lambda_0(t_i)+\sum_{j=1}^{P}\alpha_j R_j(i)\right)\\ &=-\int_0^{t_n}\lambda_0(s)\mathrm{d}s-\sum_{i=1}^{n}\sum_{j=1}^{P}\frac{\alpha_j}{\beta_j}\big(1-e^{-\beta_j(t_n-t_i)}\big)\\ &+\sum_{i=1}^{n}\ln\left(\lambda_0(t_i)+\sum_{j=1}^{P}\alpha_j R_j(i)\right)\end{aligned} \tag{39}$$

where we have the recursion[6]

$$\begin{aligned}R_j(i) &=\sum_{k=1}^{i-1}e^{-\beta_j(t_i-t_k)}\\ &=e^{-\beta_j(t_i-t_{i-1})}(1+R_j(i-1))\end{aligned} \tag{40}$$

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

$$\begin{aligned}\ln\mathcal{L}(\{t_i\}_{i=1\ldots n}) &=-\lambda_0 t_n-\sum_{i=1}^{n}\sum_{j=1}^{P}\frac{\alpha_j}{\beta_j}\big(1-e^{-\beta_j(t_n-t_i)}\big)\\ &+\sum_{i=1}^{n}\ln\left(\lambda_0+\sum_{j=1}^{P}\alpha_j R_j(i)\right)\end{aligned} \tag{41}$$

1.5. Combining the ACD and Hawkes Models.

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

2. Numerical Methods

2.1. The Nelder-Mead Algorithm.

The Nelder-Mead simplex algorithm[4] was used to optimize the likelihoods derived above.

2.1.1. Starting Points for the Optimizer.

A starting point 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 two arrival times as in the ACD model (14). 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 $E[x_i]$ as a starting point of 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. SPY (SPDR S&P 500 ETF Trust).

Consider these parameter estimates for 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. As can be seen, $P = 6$ provides the best likelihood but a more rigorous method to choose P would be to use some information criteria perhaps. 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	α	β	$\ln \mathcal{L}(\{t_i\}_{i=1\ldots n})$	$E[\lambda(t)]$
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.

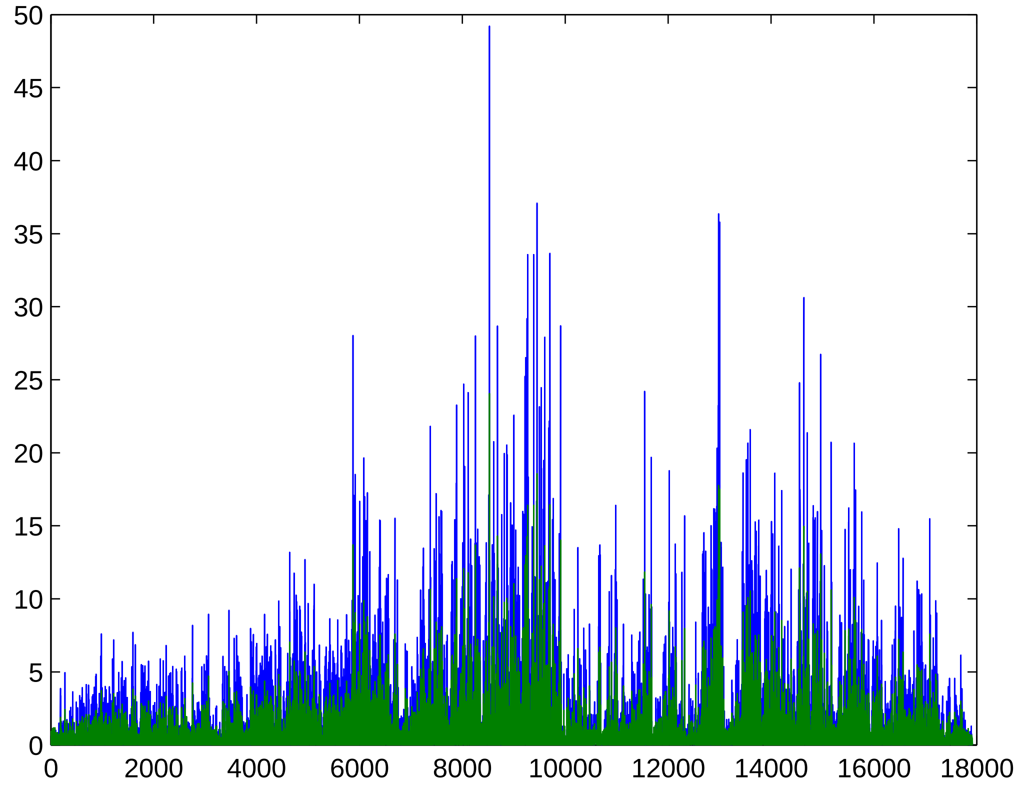

Figure 1. $x_i = t_i - t_{i-1}$ in blue and $\{\Lambda(t_{i-1}, t_i): P = 1\}$ in green

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

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

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

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

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