Hawkes Process Presentation 1

Dong Shaocong,
Wang Zexin

Hawkes Process in Finance

Quantitative Finance
National University of Singapore

July 12, 2018

Today's Agenda

Introduction

- Hawkes processes are a particularly interesting class of stochastic process that have been applied in diverse areas, from earthquake modelling to financial analysis.
- Events that are observed in time frequently cluster naturally.
- “self-exciting”: Event arrival can excite the process in the sense that the chance of a subsequent arrival is increased for some time period after the initial arrival.
- Simple and flexible for high frequency finance.

Counting Process / Point Process

Counting process is the number of arrivals until a certain point of time. Arrivals are following certain prespecified distributions.

- Event times: T_i time of the i-th event
- $N_t = \sum_{i \geq 1} \mathbb{1}_{\{t \geq T_i\}}$
- $N_0 = 0$
- jump size $= 1$ at $t = T_i$ $\forall i$

Poisson Process

Arrival time are distributed as exponential random variables.

- $\tau_i \sim Exp(\lambda)$
- pdf of τ : $f_\tau(t) = \lambda e^{-\lambda t}, \forall t \geq 0$
- expectation of τ : $\mathbb{E}(\tau) = \lambda^{-1}$
- $T_n = \sum_{j=1}^{n} \tau_j$
- $N_t = \sum_{i \geq 1} \mathbb{1}_{\{t \geq T_i\}}$
- Memoryless property: $\mathbb{P}(\tau > t + m \mid \tau > m) = \mathbb{P}(\tau > t)$
- Homogeneous: all arrival times are distributed as exponential random variables with the same λ
- Non-Homogeneous : intensity varies with time, more formally defined as: $\lambda(t \mid H_t) = \lim_{h \to 0} \frac{\mathbb{P}\{N_{t+h} - N_t = 1 | H_t\}}{h}$

Hawkes Process

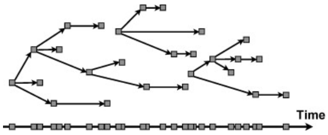
Formulation:

$\lambda(t \mid F_t) = \lambda_0(t) + \sum_{i:t>T_i} \phi(t - T_i)$

- F_t should be recognized as a filtration, a welter of all the information known at time t
- Self-exciting property: Arrivals of events increase the likelihood of future observations. Intensity at the current time depends on how many and how far away are the most recent arrivals in the past.
- Non-Homogeneous, as the intensity of the exponential random variables changes with time.
- Deterministic base intensity function: $\lambda_0(t)$
- Memory Kernel: $\phi(t - T_i)$ which links to past arrivals times
- Event Decay: monotonically decreasing kernel

Branching Structure

- We can divide the events into two categories, immigrants and the offsprings.
- Immigrants : the events directly caused by the base intensity
- Offsprings : the events '**excited**' by immigrants or another offspring
- Self-exciting is the property which can be observed in the financial markets as momentum, in which investors following the trend in the prices tend to cause trend in the same direction.
- In econometrics, this property of markets is observed as reverse casuality, as price movements drive demands, and demands drive prices as well.



Deterministic Intensity Functions

Base intensity function $\lambda_0(t) > 0$, describes externally triggered events (immigrants)

- Independence: previous events within the process
- Base (or background) intensity
- In the multivariate/bivariate case, the intensity function can be a constant.
- For example, $\lambda_i, 0(t) = \mu_i$ for a formulation of $\lambda_i(t) = \mu_i + \sum_{i:t>T_i} \phi(t - T_i)$

Cluster of offspring

- The offsprings of one immigrant events, which are the immediate offsprings of it, and their immediate offsprings, and their immediate offsprings, . . . , can be grouped as one cluster.
- Branching factor is defined as the expected number of events generated by a parent event : $\mid \Phi \mid = \int_0^\infty \phi(\tau) d\tau$
- Sub-Critical phase if $\mid \Phi \mid < 1$, meaning the number of events in one cluster is bounded.
- Super-Critical if $\mid \Phi \mid > 1$, meaning the number of events in one cluster is unbounded.
- The properties of Sub-Critical and Super-Critical mimics the stationarity property of stochastic processes.
- Expected number of events in one cluster $= \frac{1}{1-|\Phi|}$

Memory Kernels

Memory Kernels can be in any form, with two popular forms:

- Exponential decay kernel: $\phi(t) = \alpha e^{-\delta t}$
- Power law kernel: $\phi(t) = \frac{\alpha}{(1+\beta t)^{1+\gamma}}$
- By the self-exciting property, the kernel shoud possess decay property since we expect the intensity to be higher when there are more arrivals in the near past.
- The multiplier β to α is left out because a variable α is able to capture the changes by itself.

Marked Process

Each event has a corresponding mark / magnitude m_i at event time T_i

- Event lies in domain $S \times M$
- power-law kernel $\phi_m(\tau) = \kappa m^{\beta}(\tau + c)^{-(1+\theta)}$
- Event marks m_i can be i.i.d. samples from a power law distribution $P(m) = (\alpha - 1)m^{-\alpha}$
- Four parameters $\theta = \{\kappa, \beta, c, \theta\}$
  1. κ: "event quality", scales the subsequent event rate
  2. $c > 0$: temporal shift to keep $\phi_m(\tau)$ bounded
  3. θ: power-law exponent, describing how fast an event is forgotten
- κm^{β} accounts for magnitude of influences
- $(\tau + c)^{-(1+\theta)}$ models the memory over time

Simulating events from Hawkes prosses

- **Goal**: simulate inter-arrival times τ_i, $i = 1, 2, \ldots,$ according to intensity function λ_t
- **Poisson Process**: $f_\tau(t) = \lambda e^{-\lambda t}, F_\tau(t) = 1 - e^{-\lambda t}, t > 0$
- *inverse transform sampling*: $Y = F_X(X) \sim U(0,1) \Rightarrow$ Sample $u \sim U(0,1)$, then compute $\tau = \frac{-\ln u}{\lambda}$

Thinning Algorithm - rejection sampling

Applies to all non-homogeneous Poisson processes.

- *thinning property*: Poisson process with intensity λ can be split to two independent processes with intensities λ_1 and λ_2, where $\lambda = \lambda_1 + \lambda_2$
- Monotonically decreasing kernel: $[T_i, T_{i+1})$, $\lambda(T_i)$ is the upper bound of event intensity.
- Sampling procedure:
  1. $T = T_i$, sample τ (inverse transform)
  2. $\lambda^\star = \lambda(T)$, update $T = T + \tau$ (thinning property)
  3. $s \sim U(0,1)$, $T_i = T$ if $s < \frac{\lambda(T)}{\lambda^\star}$, otherwise repeat process
  4. Repeat until react to $i = N$

Decomposition Algorithm

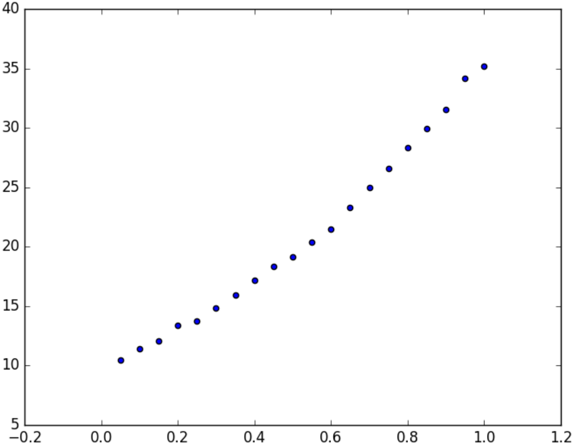
Efficient sampling for Hawkes process with exponential kernel

$$\lambda(t) = \underbrace{a + (\lambda_0 - a)e^{-\delta t}}_{\text{immigrant rate}} + \underbrace{\sum_{T_i < t} \gamma e^{-\delta(t - T_i)}}_{\text{jump by event}}, t > 0$$

- Markov (process) decomposition when Φ is exponential
- Split to two independent parts:
  - Part 1: sample $s_0 = -\frac{1}{a} \ln u_0$ (inverse transform)
  - Part 2: $s_1 = -\frac{1}{\delta} \ln d$, $d = 1 + \frac{\delta \ln u_1}{\lambda(T_{i-1}^{+}) - a}$ (Markov property)
- Inter Arrival time $= \min\{s_0, s_1\}$ to get the first event occurring time

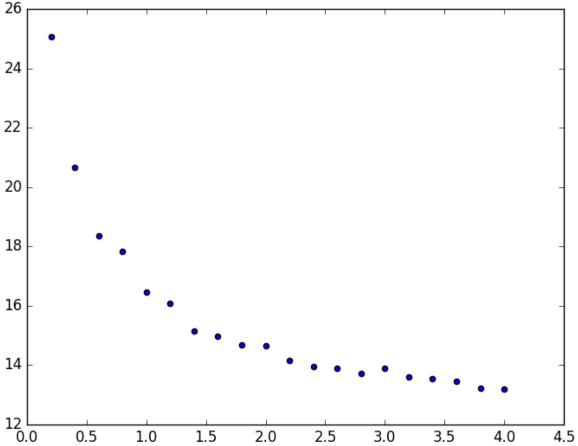
Univariate Exponential Kernel Simulation

Expected number of events with α varying from 0.05 to 1.
$\beta = 5$, $T = 10, \mu = 1$



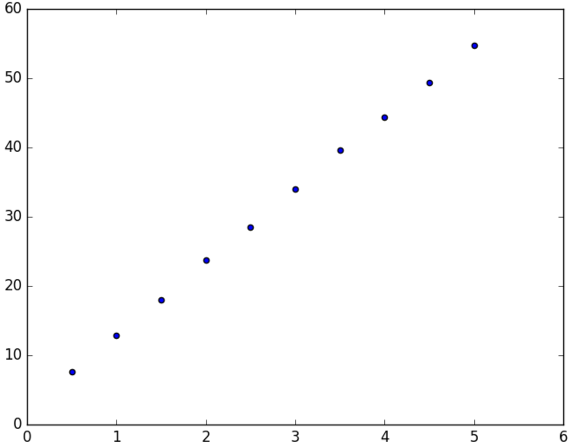
Univariate Exponential Kernel Simulation

Expected number of events with β varying from 0.2 to 4.
$\alpha = 0.2$, $T = 10, \mu = 1$



Univariate Exponential Kernel Simulation

Expected number of events with μ varying from 0.5 to 5.
$\alpha = 0.2, \beta = 5, T = 10$



Likelihood functions for Exponential Kernels

In this section we put some efforts into writing out the likelihood function of Univariate Hawkes Process, and solving for its parameters if possible and not too computationally costly.

- Ideally the likehood function should be a function of the probability density functions at the event times, and the cumulative density functions between event times.
- $L(\Theta) = \prod_{i=1}^{n} f(T_i) \prod_{i=1}^{n}(1 - F_{T_{i-1}}(T_i))(1 - F_{T_n}(T))$
- Cumulative density function for non-homogeneous Poisson Process can written as : $F(t) = 1 - \exp(-\int_0^t \lambda(s)ds)$, where the intensity $\lambda(s)$ is a function of time
- Probability density function for Hawkes Process, as we described in the previous slides, can be written as : $f(t) = \lambda(t)\exp(t\lambda(t))$, however when we let it participate in the likelihood function, $f(t) = \lambda(t)$ since the duration of event is negligible.

Likelihood functions for Exponential Kernels

$$
\begin{aligned}
L(\Theta) &= \prod_{i=1}^{n} f(T_i) \prod_{i=1}^{n}(1 - F_{T_{i-1}}(T_i))(1 - F_{T_n}(T)) \\
&=[\prod_{i=1}^{n} \lambda(T_i)][\prod_{i=1}^{n} \exp(-\int_{T_{i-1}}^{T_i} \lambda(s)ds)](\exp(-\int_{T_n}^{T} \lambda(s)ds)) \\
&= \exp(-\int_0^T \lambda(s)ds) \prod_{i=1}^{n} \lambda(T_i)
\end{aligned}
$$

Likelihood functions for Exponential Kernels

Suppose we are going to use exponential decay kernel representation and constant base intensity function:
$\phi(t) = \alpha e^{\delta t}$, $\lambda_0(t) = \mu \rightarrow \lambda(t \mid F_t) = \mu + \sum_{i:t>T_i} \alpha e^{\delta(t-T_i)}$

$$\begin{aligned} L(\Theta) &= \exp\left(-\int_0^T \lambda(t \mid F_t) dt\right) \prod_{i=1}^{n} \lambda(T_i \mid F_{T_i}) \\ &= \exp\left(-\int_0^T \mu + \sum_{i:t>T_i} \alpha e^{\delta(t-T_i)} dt\right) \prod_{i=1}^{n} \left(\mu + \sum_{j=0}^{i} \alpha e^{\delta(T_i - T_j)}\right) \end{aligned}$$

$$\ln L(\Theta) = \sum_{i=1}^{n} \ln\left(\mu + \alpha \sum_{j=0}^{i} e^{\delta(T_i - T_j)}\right) - \mu T - \alpha \int_0^T \sum_{i:t>T_i} e^{\delta(t-T_i)} dt$$

Maximum Likelihood Estimates for Exponential Kernels

Differentiate with respect to μ:

$$\ln L(\Theta) = \sum_{i=1}^{n} \ln(\mu + \alpha \sum_{j=0}^{i} e^{\delta(T_i - T_j)}) - \mu T - \alpha \int_0^T \sum_{i:t>T_i} e^{\delta(t-T_i)} dt$$

$$\frac{\partial}{\partial \mu} \ln L(\Theta) = \sum_{i=1}^{n} \frac{1}{\mu + \alpha \sum_{j=0}^{i} e^{\delta(T_i - T_j)}} - T = 0$$

$$\rightarrow \sum_{i=1}^{n} \frac{1}{\mu + \alpha \sum_{j=0}^{i} e^{\delta(T_i - T_j)}} = T$$

Maximum Likelihood Estimates for Exponential Kernels

Differentiate with respect to α:

$$\ln L(\Theta) = \sum_{i=1}^{n} \ln(\mu + \alpha \sum_{j=0}^{i} e^{\delta(T_i - T_j)}) - \mu T - \alpha \int_0^T \sum_{i:t>T_i} e^{\delta(t-T_i)} dt$$

$$\frac{\partial}{\partial \alpha} \ln L(\Theta) = \sum_{i=1}^{n} \frac{e^{\delta(T_i - T_j)}}{\mu + \alpha \sum_{j=0}^{i} e^{\delta(T_i - T_j)}} - \int_0^T \sum_{i:t>T_i} e^{\delta(t-T_i)} dt = 0$$

$$\rightarrow \sum_{i=1}^{n} \frac{e^{\delta(T_i - T_j)}}{\mu + \alpha \sum_{j=0}^{i} e^{\delta(T_i - T_j)}} = \int_0^T \sum_{i:t>T_i} e^{\delta(t-T_i)} dt$$

Maximum Likelihood Estimates for Exponential Kernels

Differentiate with respect to δ:

$$\ln L(\Theta) = \sum_{i=1}^{n} \ln(\mu + \alpha \sum_{j=0}^{i} e^{\delta(T_i - T_j)}) - \mu T - \alpha \int_0^T \sum_{i:t>T_i} e^{\delta(t-T_i)} dt$$

$$\frac{\partial}{\partial \delta} \ln L(\Theta) = \sum_{i=1}^{n} \frac{(T_i - T_j)e^{\delta(T_i - T_j)}}{\mu + \alpha \sum_{j=0}^{i} e^{\delta(T_i - T_j)}} - \int_0^T \sum_{i:t>T_i} (t - T_i)e^{\delta(t-T_i)} dt = 0$$

$$\rightarrow \sum_{i=1}^{n} \frac{(T_i - T_j)e^{\delta(T_i - T_j)}}{\mu + \alpha \sum_{j=0}^{i} e^{\delta(T_i - T_j)}} = \int_0^T \sum_{i:t>T_i} (t - T_i)e^{\delta(t-T_i)} dt$$

Maximum Likelihood Estimates for Exponential Kernels

$$\sum_{i=1}^{n} \frac{1}{\mu + \alpha \sum_{j=0}^{i} e^{\delta(T_i - T_j)}} = T \tag{1}$$

$$\sum_{i=1}^{n} \frac{e^{\delta(T_i - T_j)}}{\mu + \alpha \sum_{j=0}^{i} e^{\delta(T_i - T_j)}} = \int_0^T \sum_{i:t>T_i} e^{\delta(t - T_i)} dt \tag{2}$$

$$\sum_{i=1}^{n} \frac{(T_i - T_j) e^{\delta(T_i - T_j)}}{\mu + \alpha \sum_{j=0}^{i} e^{\delta(T_i - T_j)}} = \int_0^T \sum_{i:t>T_i} (t - T_i) e^{\delta(t - T_i)} dt \tag{3}$$

Likelihood functions for Power Law Kernels

Substituting $\phi(t) = \frac{\alpha}{(1+\beta t)^{1+\gamma}}$ in:

$$
\begin{aligned}
L(\Theta) &= \exp(-\int_0^T \lambda(t \mid F_t)dt) \prod_{i=1}^{n} \lambda(T_i \mid F_{T_i}) \\
&= \exp(-\int_0^T \mu + \sum_{i:t>T_i} \frac{\alpha}{(1+\beta(t-T_i))^{1+\gamma}} dt) \\
&\quad \prod_{i=1}^{n} (\mu + \sum_{j=0}^{i} \frac{\alpha}{(1+\beta(T_i-T_j))^{1+\gamma}})
\end{aligned}
$$

$$
\begin{aligned}
\ln L(\Theta) &= \sum_{i=1}^{n} \ln(\mu + \alpha \sum_{j=0}^{i} \frac{1}{(1+\beta(T_i-T_j))^{1+\gamma}}) - \mu T \\
&\quad -\alpha \int_0^T \sum_{i:t>T_i} \frac{1}{(1+\beta(t-T_i))^{1+\gamma}} dt
\end{aligned}
$$

Maximum Likelihood Estimates for Power Law Kernels

$$\ln L(\Theta) = \sum_{i=1}^{n} \ln(\mu + \alpha \sum_{j=0}^{i} \frac{1}{(1+\beta(T_i - T_j))^{1+\gamma}}) - \mu T$$
$$-\alpha \int_0^T \sum_{i:t>T_i} \frac{1}{(1+\beta(t - T_i))^{1+\gamma}} dt$$

$$\frac{\partial}{\partial \mu} \ln L(\Theta) = \sum_{i=1}^{n} \frac{1}{\mu + \alpha \sum_{j=0}^{i} \frac{1}{(1+\beta(T_i - T_j))^{1+\gamma}}} - T = 0$$

$$\ln L(\Theta) = \sum_{i=1}^{n} \ln(\mu + \alpha \sum_{j=0}^{i} \frac{1}{(1+\beta(T_i - T_j))^{1+\gamma}}) - \mu T$$
$$-\alpha \int_0^T \sum_{i:t>T_i} \frac{1}{(1+\beta(t - T_i))^{1+\gamma}} dt$$

$$\frac{\partial}{\partial \alpha} \ln L(\Theta) = \sum_{i=1}^{n} \frac{\sum_{j=0}^{i} \frac{1}{(1+\beta(T_i - T_j))^{1+\gamma}}}{\mu + \alpha \sum_{j=0}^{i} \frac{1}{(1+\beta(T_i - T_j))^{1+\gamma}}}$$
$$-\int_0^T \sum_{i:t>T_i} \frac{1}{(1+\beta(t - T_i))^{1+\gamma}} dt = 0$$

Maximum Likelihood Estimates for Power Law Kernels

$$\ln L(\Theta) = \sum_{i=1}^{n} \ln(\mu + \alpha \sum_{j=0}^{i} \frac{1}{(1+\beta(T_i - T_j))^{1+\gamma}}) - \mu T$$
$$-\alpha \int_0^T \sum_{i:t>T_i} \frac{1}{(1+\beta(t - T_i))^{1+\gamma}} dt$$

$$\frac{\partial}{\partial \beta} \ln L(\Theta) = \sum_{i=1}^{n} \frac{\alpha \sum_{j=0}^{i} \frac{T_i - T_j}{(1+\beta(T_i - T_j))^{2+\gamma}}}{\mu + \alpha \sum_{j=0}^{i} \frac{1}{(1+\beta(T_i - T_j))^{1+\gamma}}}$$
$$-\int_0^T \sum_{i:t>T_i} \frac{t - T_i}{(1+\beta(t - T_i))^{2+\gamma}} dt = 0$$

Univariate Model for market activity

As tested on the 10 years Euro-Bond future front contract by Bacry, when γ is closed to 0, the empirical kernel is well described by the power-law kernel.

$$\sum_{i=1}^{n} \frac{1}{\mu + \alpha \sum_{j=0}^{i} \frac{1}{1+\beta(T_i - T_j)}} = T$$

$$\sum_{i=1}^{n} \frac{\sum_{j=0}^{i} \frac{1}{(1+\beta(T_i - T_j))}}{\mu + \alpha \sum_{j=0}^{i} \frac{1}{1+\beta(T_i - T_j)}} = \int_0^T \sum_{i:t>T_i} \frac{1}{1+\beta(t - T_i)} dt$$

$$\sum_{i=1}^{n} \frac{\alpha \sum_{j=0}^{i} \frac{T_i - T_j}{(1+\beta(T_i - T_j))^2}}{\mu + \alpha \sum_{j=0}^{i} \frac{1}{1+\beta(T_i - T_j)}} = \int_0^T \sum_{i:t>T_i} \frac{t - T_i}{(1+\beta(t - T_i))^2} dt$$

Bivariate / Multivariate Hawkes Process

Formulation in our original paper:
$\lambda_t^i = \mu^i + \sum_{j=1}^{D} \int dN_{t'}^j \phi^{i,j}(t - t')$

- D : the number of variables following Hawkes Process
- $\phi^{i,j}$: the effect of variable j's arrival on variable i's intensity
- μ^i : the constant base intensity function of variable i
- λ_t^i : the intensity of variable i at time t
- $T_{j,k}$: the kth event time of variable j

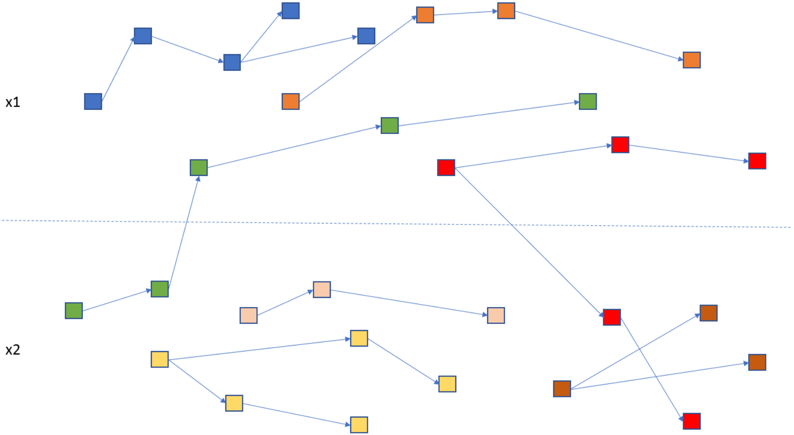
Formulation after reconcilation with notation in univariate case:
$\lambda_t^i = \mu^i + \sum_{j=1}^{2} \sum_{k:t>T_{j,k}} \phi^{i,j}(t - T_{j,k})$
$\lambda_1(t \mid F_t) = \lambda_{1,0}(t) + \sum_{i:t>T_{1,i}} \phi^{1,1}(t - T_{1,i}) + \sum_{i:t>T_{2,i}} \phi^{1,2}(t - T_{2,i})$
$\lambda_2(t \mid F_t) = \lambda_{2,0}(t) + \sum_{i:t>T_{1,i}} \phi^{2,1}(t - T_{1,i}) + \sum_{i:t>T_{2,i}} \phi^{2,2}(t - T_{2,i})$

Bivariate / Multivariate Hawkes Process

- Base intensity function $\lambda_{i,0}(t) > 0$ describes externally triggered events (immigrants) of variable i
- The memory kernels will appear in matrix of functions, in which the functions are positive and causal
- $\phi^{i,j}(t) = 0 \; \forall t < 0$
- $\phi^{i,j}(t) >= 0 \; \forall t$

Branching Structure and Clustering Representation

- The different variables share the clusters, as each variable have the ability to excite other variables and own events which parent the offsprings events in other variables



Bivariate Exponential Kernels

$$\Phi(t) = \begin{pmatrix} \phi^s(t) & \phi^c(t) \\ \phi^c(t) & \phi^s(t) \end{pmatrix}.$$

- $\phi^s(t) = \alpha^s \beta^s \exp(-\beta^s t), \forall t > 0$
- $\phi^c(t) = \alpha^c \beta^c \exp(-\beta^c t), \forall t > 0$
- This is one special case, in which two variables share the same function / parameter which determine the effect of them exciting themselves or exciting the other variable.
- $\phi^{1,2}(t) = \phi^{2,1}(t) = \phi^s(t)$: **s**elf-effect
- $\phi^{1,1}(t) = \phi^{2,2}(t) = \phi^c(t)$: **c**ross-effect

Bivariate Price Model

- The upward and downward price movements can be modelled separately using two variables following Hawkes Process.
- $P_t = P_0 + N_t^1 - N_t^2$
- In the two dimensional model, if we assume that upward and downward movements are the same, $\phi^s(t) = \phi^c(t)$, i.e. share the same parameters
- With mean-reverting assumption, $\phi^s(t) = 0$ and $\phi^c(t)$ follows exponential shape

Since price models using two variables following Hawkes Process with exponential kernels, we present its likelihood functions here:

$$
\begin{aligned}
L(\Theta) &= \prod_{i=1}^{2}\left[\prod_{j=1}^{n_i} f(T_{i,j}) \prod_{j=1}^{n_i}(1 - F_{T_{i,j-1}}(T_{i,j}))(1 - F_{T_{i,n_i}}(T))\right] \\
&= \prod_{i=1}^{2}\left[\prod_{j=1}^{n_i} \lambda_i(T_{i,j}) \left(\prod_{j=1}^{n_i} \exp(-\int_{T_{i,j-1}}^{T_{i,j}} \lambda_i(s)ds)\right) \exp\left(-\int_{T_{i,n}}^{T} \lambda_i(s)ds\right)\right] \\
&= \prod_{i=1}^{2}\left[\exp\left(-\int_{0}^{T} \lambda_i(s)ds\right) \prod_{j=1}^{n_i} \lambda_i(T_{i,j})\right]
\end{aligned}
$$

$$
\ln L(\Theta) = \sum_{i=1}^{2}\left[\sum_{j=1}^{n_i} \ln \lambda_i(T_{i,j}) - \int_{0}^{T} \lambda_i(s)ds\right]
$$

Likelihood functions for Bivariate Exponential Kernels

$\phi^{1,1}(t) = \phi^{2,2}(t) = \phi^{s}(t)$, $\phi^{1,2}(t) = \phi^{2,1}(t) = \phi^{c}(t)$, $\lambda_{i,0}(t) = \mu_i$,
$\rightarrow \lambda_i(t \mid F_t) = \mu_i + \sum_{j=1}^{2} \sum_{k:t>T_{j,k}} \phi^{i,j}(t)$

$$\begin{aligned}
\ln L(\Theta) &= \sum_{i=1}^{2} \left[\sum_{j=1}^{n_i} \ln \lambda_i(T_{i,j}) - \int_0^T \lambda_i(s) ds \right] \\
&= \sum_{j=1}^{n_1} \ln \lambda_1(T_{1,j}) - \int_0^T \lambda_1(t) dt + \sum_{j=1}^{n_2} \ln \lambda_2(T_{2,j}) - \int_0^T \lambda_2(t) dt
\end{aligned}$$

Likelihood functions for Bivariate Exponential Kernels

$$
\begin{aligned}
\ln L(\Theta) = & \sum_{j=1}^{n_1} \ln \left(\mu_1 + \sum_{k=1}^{j-1} \phi^s(T_{1,j} - T_{1,k}) + \sum_{k:T_{1,j} > T_{2,k}} \phi^c(T_{1,j} - T_{2,k}) \right) \\
& - \mu_1 T - \int_0^T \sum_{k:t > T_{1,k}} \phi^s(t - T_{1,k}) + \sum_{k:t > T_{2,k}} \phi^c(t - T_{2,k}) dt \\
& + \sum_{j=1}^{n_2} \ln \left(\mu_2 + \sum_{k=1}^{j-1} \phi^s(T_{2,j} - T_{2,k}) + \sum_{k:T_{2,j} > T_{1,k}} \phi^c(T_{2,j} - T_{1,k}) \right) \\
& - \mu_2 T - \int_0^T \sum_{k:t > T_{2,k}} \phi^s(t - T_{2,k}) + \sum_{k:t > T_{1,k}} \phi^c(t - T_{1,k}) dt
\end{aligned}
$$

Likelihood functions for Bivariate Exponential Kernels

Since it is too computationally costly to expand ϕ^s and ϕ^c, we are going to leave the derivation out in this presentation.

$$\frac{\partial}{\partial \mu_1} \ln L(\Theta) = \sum_{j=1}^{n_1} \left(\mu_1 + \sum_{k=1}^{j-1} \phi^s(T_{1,j} - T_{1,k}) + \sum_{k: T_{1,j} > T_{2,k}} \phi^c(T_{1,j} - T_{2,k}) \right)^{-1} - T = 0$$

$$\rightarrow T = \sum_{j=1}^{n_1} \left(\mu_1 + \sum_{k=1}^{j-1} \phi^s(T_{1,j} - T_{1,k}) + \sum_{k: T_{1,j} > T_{2,k}} \phi^c(T_{1,j} - T_{2,k}) \right)^{-1}$$

Maximum Likelihood Estimates

- It has been universally agreed that computation with MLE for Hawkes process is always going to be costly
- Computation of the likelihood is of order $O(M^2D)$ for a general Hawkes process, e.g. power law kernal
- where M is the number of events, and D is the number of Hawkes processes
- Computation of the likelihood is of order $O(MD)$ for a Hawkes process with exponential kernel
- Part of the reason for popularity of bivariate exponential price models

Application for market activity

A result of literature review shows that Hawkes Process models can be applied for:

- prices of equity index, bond futures, foreign exchange rates
- peculiar non-stationarities in market microstructures such as intraday seasonalities and overnight gaps
- volatility clustering phenomenon modelling (correlated nature of volatility fluctutations)

Application for market impact modelling

Hawkes Process models can also be applied to model the market impact which is:

- the impact on price levels caused by placing order with significant amount
- extra cost induced per transaction which needs to be added to the transaction costs charged by market
- mechanism enforcing efficiency of market, allowing prices to reflectg fundamental information

Application for market impact modelling

Hawkes Impact Model(HIM) proposed by Bacry and Muzy

- directly accounts for the joint dynamics of mid-prices and market order occurrences
- breaks down into:
    - self-excitement of market order flow
    - self-excitement of price movements
    - impact of the market orders on prices (market impact)
    - impact of price movements on market order flow intensity (feedback influence)
- was able to disentangle the self and cross excitation dynamics of price changes from market impacts
- was not able to model for either the volume of the market orders, nor the price movement sizes
- was not able to consider the bid and ask (offer) price movements

Application for market impact modelling

2-dimensional Hawkes Impact Model(HIM)

- $\lambda_t^1 = \mu + \phi^s dN_t^2 + \phi^I f(r_t)$
- $\lambda_t^2 = \mu + \phi^s dN_t^1 + \phi^x f(r_t)$
    - dN_t^1 : upward price movements
    - dN_t^2 : downward price movements
    - $\phi^s(t)$: self-excitement of price movements
    - $\phi^I(t)$: market impact - buy on upward / sell on downward
    - $\phi^x(t)$: cross market impact - buy on downward / sell on upward
    - $r(t)$: trading rate per unit time
    - $r(t)dt$: number of shares bought in dt time
- Impulsive-HIM: $\phi^x(t) = C\frac{\phi^s(t)}{||\phi^s||}$
    - $C = 0 \rightarrow$ no contrarian impact, strong permanent impact
    - $C = 1 \rightarrow$ strong contrarian impact, no permanent impact

Thank You

E0007424@u.nus.edu

E0012680@u.nus.edu