CIRJE-F-1045

The Simultaneous Multivariate Hawkes-type Point Processes and Their Application to Financial Markets

Naoto Kunitomo
Meiji University
Daisuke Kurisu
Graduate School of Economics, The University of Tokyo
Yusuke Amano
Graduate School of Economics, The University of Tokyo
Naoki Awaya
Graduate School of Economics, The University of Tokyo

April 2017

The Simultaneous Multivariate Hawkes-type Point Processes and their application to Financial Markets *

Naoto Kunitomo†

Daisuke Kurisu ‡

Yusuke Amano §

and

Naoki Awaya ¶

March 20, 2017

Abstract

In economic and financial time series we sometimes observe sudden and large jumps. Although these events are relatively rare, they would have significant influence not only on a financial market but also several different markets and macro economies. By using the simultaneous Hawkes-type multivariate point processes (SHPP) models, it is possible to analyze the causal effects of large events in the sense of the Granger-non-causality (GNC) and the instantaneous Granger-non-causality (IGNC). We investigate the financial market of Tokyo and other markets, and apply the Granger non-causality tests. We have found several important empirical findings among financial markets and macro economies.

Key Words

Simultaneous Hawkes-type marked point (SHPP) processes, Granger non-causality,

Instantaneous Granger non-causality, Causality tests, Financial markets

---

*A preliminary draft (2017-3-20). This work was supported by JSPS KAKENHI Grant Number JP25245033.

†Meiji University, School of Political Sicence and Economics

‡Graduate student, Graduate School of Economics, University of Tokyo

§Graduate student, Graduate School of Economics, University of Tokyo

¶Graduate student, Graduate School of Economics, University of Tokyo

1 Introduction

In economic and financial time series, we sometimes observe sudden and large jumps. Although they are relatively rare event, they often have significant influence not only on a single financial market but also several different markets and macro-economies. There have been several recent events occurred in European and Asian countries including the financial crisis of 2007-2008 (sometimes called Lehman shock).

The standard statistical method for investigating economic and financial time series is the statistical analysis of discrete time series in econometrics. In this statistical method, we often assume that the observed time series data are equally spaced realizations of stochastic processes and the state space is $\mathbf{R}^p$ in the multivariate time series analysis. Many statistical procedures of discrete time series analysis have been developed and applied to economic and financial time series in the last several decades. When we do not observe events frequently, however, the traditional use of discrete time series modeling with continuous state space may have some limitations.

In this paper we will propose to use an alternative way of investigating economic and financial events with time series data in macro-economies, that is, the statistical analysis of the marked point process approach to investigate time series events. Although it has not been a standard approach in time series econometrics, there have been statistical applications in statistical seismology (see Ogata (1978, 2015) and its related literature, for instance). We will show that this approach would be an alternative useful way to investigate multivariate economic and financial markets and shed some new light on some aspects sometimes ignored. In particular, we shall propose to use the simultaneous Hawkes-type multivariate point process models and their applications in this study. It seems that they are not standard statistical models in the past econometric analyses, but there are some reasons that they are useful in economic and financial time series analysis. By using the simultaneous multivariate Hawkes-type point process (SHPP) model, which is a new multivariate point process, it is possible to investigate the causal effects of sudden and large events of their magnitude in the sense of the Granger-non-causality (GNC) and the instantaneous Granger-non-causality (IGNC) through the stochastic intensity modeling. In the econometric time series analysis, the concept of Granger-Causality has been one of important tools to investigate the relationships among time series variables since Granger (1969). In econometric literature, Florens and Fougere (1996) have investigated several Granger-causality concepts in the framework of continuous time stochastic processes, but we argue that their formulation of the problem was incomplete because they had excluded the possibility of co-jumps, which means the simultaneous jumps in multivariate times series. In this paper, we shall investigate the use of co-jumps and will develop the tests of the Granger-non-causality and the instantaneous Granger-non-causality, which may give some new light on the econometric time series modeling.

As empirical examples, we will investigate the interactions among Tokyo- -NY (New York)-London financial markets and Tokyo-HK (Hong Kong) financial markets, and apply the Granger non-causality tests. We have found several important empirical findings among major financial markets.

In Section 2 we present a general formulation of the simultaneous multivariate Hawkes-type point process (SHPP) model in this study. Then in Section 3, we will discuss the estimation method and develop the non-causality tests in the sense of Granger (1969). In Section 4, we will discuss some simulation results and some empirical applications will be given in Section 5. Concluding remarks will be presented in Section 6. Some mathematical details will be given in Appendix.

2 Simultaneous Hawkes-type Point Processes

We divide the observation period $[0, T]$ to the discrete observation periods $I_i^n = (t_{i-1}^n, t_i^n]$ $(i = 1, \cdots, n)$. The initial time is $t_0^n = 0$ and we interpret I_i^n as the $i-$th day, but it is possible to use finer observations than daily data. Let the observable $d-$dimension stochastic process be $P_j(t)$ $(j = 1, \cdots, d\ ; t_{i-1}^n < t \leq t_i^n, i = 1, \cdots, n)$ and in $s \in I_i^n$ we consider the (negative) log-returns of prices $Y_j^n(s)$ $(t_{i-1}^n < s \leq t_i^n)$ be

$$(2.1)\quad Y_j^n(s) = -\log[P_j(s)/P_j(t_{i-1}^n)]\ (j = 1, \cdots, d; i = 1, \cdots, n)\ .$$

Let the first stopping time when $Y_j^n(s)$ exceeds the threshold u_j in $s \in I_i$ be $\tau_j^n(i,1)$. We define $X_j^n(s) = Y_j^n(s)$ for $s \in t_{i-1}^n \leq s \leq \tau_j^n(i,1)$ and $X_j^n(s) = X_j^n(\tau_j^n(i,1))$ for $s \in [\tau_j^n(i,1), t_i^n]$.

We consider the simple counting processes $N_j^{n*}(s, u_k)$ by the number of stopping times that $X_j^n(s)$ exceed u_j $(j = 1, \cdots, d)$ for a particular j but not for other $k \neq j$ by the time s. For the resulting simplicity, we assume that the jumps of the counting process $N_j^{n*}(s, u_k)$ can occur at t_i^n, the end of each intervals $(t_{i-1}^n, t_i^n]$, because the number of jumps over a threshold in a finite interval are finite with probability one and $u_j = u$ $(j = 1, \cdots, d)$. We notice that the interval length goes to zero, that is, $\max_{i=1,\cdots,n} |t_i^n - t_{i-1}^n| \longrightarrow 0$ as $n \longrightarrow \infty$ and the simple counting process $N_j^{n*}(s, u)$ converges to $N_j^*(s, u)$. The resulting counting process can be interpreted as the limiting process in the high frequency asymptotics, which is not a diffusion type process. (Ait-Sahalia and Jacod (2014), for instance.)

The simple point processes we consider $N_j^*(t)$ $(j = 1, \cdots, d)$ satisfy the standard condition for point processes that as $\Delta t \to 0$ we have

$$P(N_j^{n*}(t + \Delta t, u) - N_j^{n*}(t, u) = 1 | \mathcal{F}_t^n) = \lambda_j^{n*}(t, u)\Delta t + o_p(\Delta t) ,$$
$$P(N_j^{n*}(t + \Delta t, u) - N_j^{n*}(t, u) > 1 | \mathcal{F}_t^n) = o_p(\Delta t) ,$$
$$P(N_k^{n*}(t + \Delta t, u) - N_j^{n*}(t, u) \geq 1 | \mathcal{F}_t^n) = o_p(\Delta t) \text{ for } k \neq j ,$$

where $\mathcal{F}_t^n$ is the $\sigma-$field generated by the information at t, the (conditional) intensity functions are given by

$$\lambda_j^{n*}(t, u) = \lim_{\Delta t \to 0} \mathbf{E}[\frac{N_j^{n*}(t + \Delta t, u) - N_j^{n*}(t, u)}{\Delta t} | \mathcal{F}_{t-}^n] , \tag{2.2}$$

and we denote $\mathcal{F}_t$ for $\mathcal{F}^n_{t-}$ whenever there is no confusion on the notation.

Next, we define the simple proint processes $N^{n*}_{jk}(s,u)$ by the number of stopping times that $X^n_j(s)$ exceed u $(j = 1, \cdots, d)$ for a particular j, $X^n_k(s)$ exceed u_k $(k = 1, \cdots, d; k \neq j)$ for a particular k and other $X^n_l(s)$ $(l \neq j, k)$ do not exceed u_l by the time s in the interval I^n_i. By this construction, we can introduce the point processes $N^{n*}_{jk}(t,u)$ with co-jumps of N_j and N_k by

$$P(N^{n*}_j(t+\Delta t, u) - N^{n*}_j(t,u) = N^{n*}_k(t+\Delta t, u) - N^{n*}_k(t,u) = 1|\mathcal{F}_t)$$
$$= \lambda^{n*}_{jk}(t,u)\Delta t + o_p(\Delta t) ,$$
$$P(N^{n*}_j(t+\Delta t, u) - N^{n*}_j(t,u) > 1|\mathcal{F}_t) = o_p(\Delta t) ,$$
$$P(N^{n*}_k(t+\Delta t, u) - N^{n*}_j(t,u) > |\mathcal{F}_t) = o_p(\Delta t) \text{ for } k \neq j ,$$

where $\lambda^{n*}_{j,k}(t,u_j)$ are the conditional intensity functions of co-jumps. When we have co-jumps of two point processes, we can define the point processes

$$N^n_j(s,u) = N^{n*}_j(s,u) + \sum_{k \neq j} N^{n*}_{j,k}(s,u) \; (j,k = 1, \cdots, d)$$

and the corresponding conditional intensity functions are given by

$$\lambda^n_j(t,u) = \lambda^{n*}_j(t,u) + \sum_{k \neq j} \lambda^{n*}_{j,k}(t,u) \; . \tag{2.3}$$

The resulting point processes can be interpreted as some marginal point processes of the j-th component of the vector point process $\mathbf{N}^n(s,u)$.

It is straightforward to extend this formulation to have more complicated co-jumps. We define

$$N_j^n(s,u) = \sum_{J_j \in (1,\cdots,d)} N_{j_1,\cdots,j_l}^{n*}(s,u) \; (j = 1,\cdots,p), \tag{2.4}$$

where the index set $J_j = \{j_1, \cdots, j_l\} \in \{1, \cdots, d\}$ is a subset of $(1, \cdots, d)$.
The index sets are defined as $J_i = \{i\}$ for $(i = 1, \cdots, d)$, $J_i = \{1, 1 + (i - d)\}$ for $(i = d + 1, \cdots, 2d - 1), \cdots$, and $J_p = \{1, \cdots, d\}$. That is, we sequentially define $N_i^{n*}(s,u) = N_i^{n*}(s,u)$ $(i = 1, \cdots, d)$, $N_{d+1}^{n*}(s,u) = N_{1,2}^{n*}(s,u), \cdots$, and $N_p^{n*}(s,u) = N_{1,\cdots,d}^{n*}(s,u)$.

We also use the self-exciting form of conditional intensity functions for co-jumps as $\lambda_{j,k}(t, x|\mathcal{F}_{t-}^n)$ in the same way and

$$\lambda_j^n(t,u) = \sum_{J_j \in (1,\cdots,d)} \lambda_{j,k}^{n*}(t,u) \; . \tag{2.5}$$

There are one-to-one transformation between $N_j^n(s,u)$ and $N_j^{n*}(s,u)$, and $\lambda_j^n(t,u)$ and $\lambda_j^{n*}(t,u)$, respectively.

We will also consider the self-exciting Hawkes-type conditional intensity function for the marked point processes as

$$\begin{aligned} \lambda_j^{n*}(t, x|\mathcal{F}_{t-}^n) &= \left[\lambda_{j,0} + \textstyle\sum_{i=1}^p \int_{-\infty}^t c_{ji}^*(x) g_{ji}^*(t-s) N_{J_i}^{*n}(ds \times dx)\right] \\ &\quad \times \left[- \frac{\partial}{\partial x_j} F_\theta(x_1, \cdots, x_d|\mathcal{F}_{t-}^n)\Big|_{x_1=u, x_d=u}\right] . \end{aligned} \tag{2.6}$$

for $j = 1, \cdots, p$, where $N_{J_i}^{*n}(ds \times dx)$ are the marked point precess representations and $F_\theta(x_1, \cdots, x_d)$ are the joint distribution function of the underlying processes $Y_j^n(s)$, the damping functions $g_{ji}(t - s) = e^{-\gamma_{ji}(t-s)}$ and the impact functions $C(X) = (c_{ji}(x))$.

Since we are interested in sudden and large jumps of the underlying price processes, it is important to use their probability functions in the tails. We use the tail probability functions for $y > u$ $(j = 1, \cdots, d)$ as

$$
\begin{aligned}
(2.7)\ P(Y_j^n(s) > y | Y_j^n(s) > u, \mathcal{F}_s) &= \frac{\left[1 + \frac{\xi_j}{\sigma_j(s)} y\right]^{-1/\xi_j}}{\left[1 + \frac{\xi_j}{\sigma_j(s)} u\right]^{-1/\xi_j}} \\
&= \left[1 + \frac{\xi_j}{\sigma_j^*(s)} (y - u_j)\right]^{-1/\xi_j} ,
\end{aligned}
$$

and we set $\sigma_j^*(s) = \xi_j u_j + \sigma_j(s)$ $(\sigma_j(s) > 0)$.

We assume that given the return at s $Y_j^n(s)$ the conditional density functions are given by

$$
(2.8)\ f_j(y, s) = \frac{1}{\sigma_j^*(s)} \left[1 + \frac{\xi_j}{\sigma_j^*(s)} (y - u)\right]^{-1/\xi_j - 1} \quad (x > u\ , \xi_j > 0)\ .
$$

(See Resnick (2007) for the generalized Pareto distribution as the statistical extreme value theory.)

As the specific form for (2.5) for our empirical study, we use the simple situation that $Y_j^n(s) = X_j^n(s)$, $\sigma_j^*(s) = \sigma_j$ and the conditional intensity functions as

$$
(2.9) \quad \lambda_{J_i}^{*n}(t, u) = \lambda_{j0}^n + \sum_{i=1}^{p} \int_0^t [A_{ji}(X_i^n)^c(s-)] g_i(t - s) dN_{J_i}^{*n}(s, u)
$$

where $N_{d+1}^{*n}(s, u) = N_{12}^n(s, u), \cdots, N_p^{*n}(s, u) = N_{1\cdots d}^n(s, u)$ and the parameters λ_{j0} and γ_i are constants. We denote the impact fucntion $C_{ji}(X) = (A_{ji} \max_{j \in J_i} x_j^c)$ $(0 \leq c \leq 1; i, j = 1, \cdots, p)$.

In particular when $p = d$ and $C_{ji} = \delta(j, i)$ (indicator functions), they correspond to the multivariate marked Hawkes-type processes, which are the simple point processes without co-jumps.

Let $p \times p$ transformation matrix as

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 1 & 1 & \cdots & 0 & \cdots & 1 \\ 0 & 1 & \cdots & 0 & 1 & 0 & \cdots & 0 & \cdots & 1 \\ \vdots & & & 1 & 0 & & & \vdots & \cdots & 1 \\ \vdots & & & 0 & 1 & & & \vdots & \cdots & 0 \\ 0 & \cdots & 0 & 0 & & \cdots & & 0 & \cdots & 1 \end{bmatrix} ,$$

where $\mathbf{D}_1$ is a $d \times p$ matrix, $\mathbf{D}_2$ is a $(p-d) \times p$ matrix and $p = 2^d - 1$. Also let $p \times 1$ vectors

$$\boldsymbol{\lambda}^n(t, \mathbf{u}) = \begin{bmatrix} \lambda_1^n(t,u) \\ \vdots \\ \lambda_d^n(t,u) \\ \lambda_{1,2}^n(t,u) \\ \vdots \\ \lambda_{1,2,\cdots,d}^n(t,u) \end{bmatrix} , \ \mathbf{N}(t, \mathbf{u}) = \begin{bmatrix} N_1(t,u) \\ \vdots \\ N_d(t,u) \\ N_{1,2}(t,u) \\ \vdots \\ N_{1,2,\cdots,d}(t,u) \end{bmatrix} ,$$

and $p \times p$ matrices

$$\mathbf{C}(X(s-)) = [c_{ij}(X_{s-})] \ , \ \mathbf{G}(t-s) = [\mathrm{diag}(g_j(t-s))] \ .$$

(We use the notation of diag$(\cdot)$ for diagonal matrices and we often omit n for and $\lambda_{J_i}^n(s)$ $(i = 1, \cdots, p)$ and $N_{J_i}^n$ whenever their meanings are clear.)

In this paper we call the above Hawkes-type conditional intensity models as the simultaneous multivariate Hawkes-type point process

(SHPP) models because the resulting market point processes are not necessarily simple [1]. The classical Hawkes point processes have been useful in applications because they are simple point processes, but they exclude the possibility of simultaneous jumps or co-jumps in consideration. The constructions of our marked point processes can be regarded as an extension of Solo (2007).

3 Stationarity and Decomposition of Bartrett Spectrum

In our applications, we will use the stationary self-exciting Hawkes-type (marked) point processes. We take the expectation of the intensity function of (2.9) in $(-\infty, t]$ as

$$(3.10) \quad \mathbf{E}[\boldsymbol{\lambda}^n(t,\mathbf{u})] = \boldsymbol{\lambda}_0 + \mathbf{E}[\int_{-\infty}^{t} \boldsymbol{C}(\mathbf{X}(s-)\mathbf{G}(t-s)d\mathbf{N}(s,\mathbf{u})] \ .$$

We take the non-negative intensity functions and then a set of sufficient conditions for the existence of stationary point processes are that $\mathbf{E}[\boldsymbol{C}(\mathbf{X}(s-))$ are bounded for any s and the dpectral radius

$$(3.11) \qquad \sup_{t} \max_{1\leq i\leq p} |\lambda_i(\mathbf{F}_t)| < 1 \ ,$$

where $\lambda_i(\mathbf{F}_t)$ is the characteristic roots of

$$(3.12) \qquad \mathbf{F}_t = \int_{-\infty}^{t} \mathbf{E}[\mathbf{C}(\mathbf{X}(s-)]\mathbf{G}(t-s)\mathbf{F}_s ds \ .$$

For instance, if we have a constant matrix $\mathbf{C} = \mathbf{E}[\mathbf{C}(\mathbf{X}(s-)]$ and $\boldsymbol{\Gamma} = (\mathrm{diag}(\gamma_i)), g_i(t) = e^{-\gamma_i t}$ $(\gamma_i > 0; i = 1,\cdots,p)$, then we have

[1] The definition of simple, basic terminologies of point processes and their mathematical details are given in Dalay and Vere-Jones (2003), for instance.

$\mathbf{F}_t = \mathbf{F} = \mathbf{C}\mathbf{\Gamma}^{-1}$ and $\mathbf{\Gamma} = \text{diag}(\gamma_j)$. When $d = p = 1$ (one-dimensional Hawkes process) in particular, $\mathbf{C} = \alpha$ and and $\mathbf{\Gamma} = \gamma$ (> 0), then $\mathbf{F} = \alpha/\gamma$.

Hawkes (1971) inrtoduced the spectral density for the stationary vector point process $\mathbf{N}(t) = (N_i(t))$, which was developed by Bartlett (1963), followed by

$$(3.13) \qquad \boldsymbol{\lambda}(t) = \boldsymbol{\lambda}_0 + \int_{-\infty}^{t} \boldsymbol{\gamma}(t-u)d\mathbf{N}(u) \ ,$$

where $\boldsymbol{\gamma}(u) = (\gamma_{ij}(u))$ is a $p \times p$ matrix and $\boldsymbol{\gamma}(u) = (0)$ (zero-matrix) for $u < 0$.

Let the Fourier transform of $\boldsymbol{\gamma}(tau)$ be

$$(3.14) \qquad \mathbf{\Gamma}^*(\omega) = \int_{-\infty}^{\infty} e^{-i\omega\tau}\boldsymbol{\gamma}(\tau)d\tau \ ,$$

where $i^2 = -1$.

Then when there are no co-jumps, the Bartrett spectral matrix for frequency ω is given by

$$(3.15) \qquad \mathbf{f}(\omega) = \frac{1}{2\pi}[\mathbf{I}_d - \mathbf{\Gamma}^*(\omega)]^{-1}\mathbf{\Sigma}[\mathbf{I}_d - \mathbf{\Gamma}^{*'}(-\omega)]^{-1} \ ,$$

where $p = d$ and $\mathbf{\Gamma}^*$ in (3.13) is a $d \times d$ matrix for the $d-$dimensional vector point process. When there can be co-jumps, the Bartrett spectral matrix is given by

$$(3.16) \mathbf{f}(\omega) = \frac{1}{2\pi}[\mathbf{I}_d, \mathbf{O}][\mathbf{D} - \mathbf{D}\mathbf{\Gamma}^*(\omega)]^{-1}\mathbf{\Sigma}[\mathbf{D}' - \mathbf{D}'\mathbf{\Gamma}^{*'}(\omega)]^{-1}\left[\begin{array}{c}\mathbf{I}_d \\ \mathbf{O}\end{array}\right],$$

where $\mathbf{f}(\omega) = (f_{ij}(\omega))$ is the $p \times p$ spectral density matrix and $\mathbf{\Sigma} = (\sigma_{ii})$ is the diagonal matrix with the variances σ_{ii} $(i = 1, \cdots, p)$ for the orthogonal point processes N_i^* $(i = 1, \cdots, p)$.

The relative power contribution (RPC) of the marginal spectral density function $f_{ii}(\omega)$ $(i = 1, \cdots, d)$ with the frequency ω can be defined by using the joint spectral density matrix $\mathbf{f}(\omega)$. The (i,i)-component of $\mathbf{f}(\omega)$ can be represented as

$$f_{ii}(\omega) = \sum_{k=1}^{p} |a_{ik}(\omega)|^2 \sigma_{kk} \tag{3.17}$$

and

$$\mathbf{RPC}_{k \to i}(\omega) = \frac{|a_{ik}(\omega)|^2 \sigma_{kk}}{f_{kk}(\omega)} \quad (i = 1, \cdots, p; k = 1, \cdots, d) , \tag{3.18}$$

where $a_{ij}(\omega)$ $(i = 1, \cdots, d; j = 1, \cdots, p)$ are the functions of complex variables. Also the instantaneous RPC ($\mathrm{IRPC}_{j \to i}$) can be defined by

$$\mathbf{IRPC}_{j \to i}(\omega) = \frac{|a_{ij}(\omega)|^2 \sigma_{jj}}{f_{ii}(\omega)} \ (j = d + 1, \cdots, p) . \tag{3.19}$$

In this way, we can measure the relative power contributions for any frequency ω, which corresponds to the Granger-causality measures in the frequency domain. One important aspect of the above formulation is the fact that we have a natural definition of Instantaneous Granger-causality in the frequency domain, which is different from the discrete time series modelling.

4 Estimation and Non-causality Tests

When the point process is simple and there is no co-jump, the log-likelihood function of (d-dimensional) multivariate point process has been known (see Daley and Jones (2003)) and it is given by

$$\sum_{i=1}^{d}\{-\int_0^T \lambda_i^n(s)ds+\int_0^T \log(\lambda_i^n(s))dN_i^n(s)\}\ . \tag{4.20}$$

The log-likelihood function of the marked multivariate point process with the density function $f_i(x)$ is given by

$$\begin{aligned}\log L_T &= \sum_{i=1}^{d}\{-\int_0^T \lambda_i^n(s)ds+\int_0^T \log(\lambda_i^n(s))dN_i^n(s)\} \\ &\quad +\sum_{i=1}^{d}\{\int_0^T \log f_i(X_i^n(s-))dN_i^n(s)\} \\ &= L_1+L_2\ ,\end{aligned} \tag{4.21}$$

where

$$\begin{aligned}L_1 &= \sum_{i=1}^{d}\{-\int_0^T \lambda_i^n(s)ds+\int_0^T \log(\lambda_i^n(s))dN_i^n(s)\}\ , \\ L_2 &= \sum_{i=1}^{d}\{\int_0^T \log f_i(x_i^n(s-))dN_i^n(s)\}\end{aligned}$$

and the density function

$$f_i(x)=\frac{1}{\sigma_i^*}(1+\xi_i\frac{x_i-u_i}{\sigma_i^*})^{-\frac{1}{\xi_i}-1}\ (i=1,\cdots,d)\ . \tag{4.22}$$

Then we can apply the maximum likelihood method to L_1 and L_2 separately. In this formulation we use the GPD (generalized Pareto distribution) for the marginal distributions.

When there can be co-jumps, the log-likelihood function of (d-dimensional) multivariate point process is not the above form and it should be given by

$$\log L_T^*=L_1^*+L_2^*\ , \tag{4.23}$$

where

$$\begin{aligned} L_1^* &= \sum_{i=1}^{d}\{-\int_0^T \lambda_i^n(s)ds + \int_0^T \log(\lambda_i^n(s))dN_i^n(s)\} \\ &+ \sum_{i\neq j=1}^{d}\{-\int_0^T \lambda_{ij}^n(s)ds + \int_0^T \log(\lambda_{ij}^n(s))dN_{ij}^n(s)\} \\ &+ \cdots + \{-\int_0^T \lambda_{i\cdots d}^n(s)ds + \int_0^T \log(\lambda_{i\cdots d}^n(s))dN_{i\ldots d}^n(s)\}\ . \end{aligned}$$

and $L_2^* = L_2$.

In our applications we mainly deal with the case when $d = 2$ and then there is only one extra term in the likelihood function because $p = 2^d - 1$.

We assume the stationarity condition (3.11) and the existence of second order moments of $\mathbf{C}(\mathbf{X}) = c_{ij}(\mathbf{X}(s))$ in the statistical inference of Hawkes-type point processes. Also we take $\boldsymbol{\lambda}(\mathbf{u})$ as the stationary conditional intensity and some $q \times p$ predictable processes $\boldsymbol{\xi}(t)$ having the second order moments. Then, because of the martingale property, it is straightforward to show the asymptotic properties as we have

$$\frac{1}{T}\int_0^T \boldsymbol{\xi}(t)[\mathbf{N}(t,u) - \boldsymbol{\lambda}(t,\mathbf{u})]dt \longrightarrow 0 \ \ (a.s.) \tag{4.24}$$

and

$$\frac{1}{T}\int_0^T \boldsymbol{\xi}(t)[\boldsymbol{\lambda}(t,u) - \boldsymbol{\lambda}(\mathbf{u})]dt \xrightarrow{p} 0 \tag{4.25}$$

as $T \to \infty$.

For the one-dimensional point processes with the stationary intensity function, Ogata (1978) have given a set of sufficient conditions for the consistency and asymptotic normality of the maximum likelihood

(ML) estimation. His derivations are based on a martingale central limit theorem (MCLT) and it is straightforward to extend his arguments to the multi-dimensional cases. For the sake of completeness, we have given some detail of our arguments based on a new MCLT in Mathematical Appendix, which may be more general than the standard literature as the ones given by Ogata (1978). We will also give the outline of our proofs of Theorem 6.1 and Theorem 6.2 in Appendix, which are used as the non-causality tests in our empirical study.

5 Simulations

To examine the relevance of our estimation procedure proposed in this paper we have done a set of simulations. The model we have used in our simulations are the simultaneous Hawkes-type model with two dimension and the intensity functions are given by

$$\begin{aligned}
\lambda_1^n(t) &= \lambda_{10}^n + \int_0^t \alpha_{11} e^{-\gamma(t-s)} X_1 dN_1^n(s) + \int_0^t \alpha_{12} e^{-\gamma(t-s)} X_2 dN_2^n(s) \\
&\quad + \int_0^t \alpha_{13} e^{-\gamma(t-s)} [\max_i X_i] dN_{12}^n(s) \;, \\
\lambda_2^n(t) &= \lambda_{20}^n + \int_0^t \alpha_{21} e^{-\gamma(t-s)} X_1 dN_1^n(s) + \int_0^t \alpha_{22} e^{-\gamma(t-s)} X_2 dN_2^n(s) \\
&\quad + \int_0^t \alpha_{23} e^{-\gamma(t-s)} [\max_i X_i] dN_{1,2}^n(s) \;, \\
\lambda_{12}^n(t) &= \lambda_{12,0}^n + \int_0^t \alpha_{31} e^{-\gamma(t-s)} X_1 dN_1^n(s) + \int_0^t \alpha_{32} e^{-\gamma(t-s)} X_2 dN_2^n(s) \\
&\quad + \int_0^t \alpha_{33} e^{-\gamma(t-s)} [\max_i X_i] dN_{1,2}^n(s) \;.
\end{aligned}$$

We first generate the stock price returns by using the generalized Pareto distribution as marginal and the two-dimensional Gaussian copura. Then we use the maximum likelihood (ML) method to obtain the estimates of the underlying parameters. We give several figures (Figures 5.1-5.6) among many results and all figures of the finite sample distributions of the ML estimator are standardized as

$$\mathbf{I}_n^{1/2}(\hat{\theta} - \theta) \ , \tag{5.26}$$

where $\boldsymbol{\theta} = (\theta_i)$ is the vector of parameters and and $\hat{\theta}$ is the ML estimator. This makes possible to compare them to the standard normal distributions.

In our numerical evaluations we sometimes hit the boundaries of the non-negativity of intensity functions with finite samples, which make the simulation some instabilities. Thus we have set non-negativity restrictions on parameters in our simulations. Then we have reasonable results, but then sometimes we observe that the maximum likelihood estimators of coefficients have the resulting biases, which are basically not very large. We summarize the setting of our numerical experiments : the simulation size $N = 100$, and for GPD$(\sigma_j,\ \xi_j)$ we set $(\sigma_1, \xi_1) = (0.007, 0.22)$, and $(\sigma_2, \xi_2) = (0.008, 0.15)$. These numerical values are based on our preliminary the empirical study, which give reasonable estimates.

Among many simulations we illustrate our results in Table 5-1 and Figures. Because we have taken $\alpha_{12}^* = 0$, the resulting estimate is not significant. Other estimates of α_{ij} take resonable values on average. There are some positive biases on α_{ij} and negative biases on the initial intensities, which are the results of the non-negativity of parameter

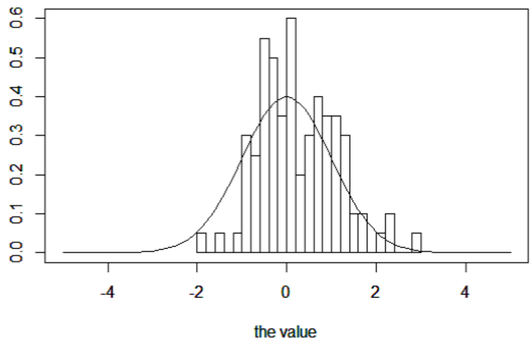

Figure 5-1 : α_{12}^{*}

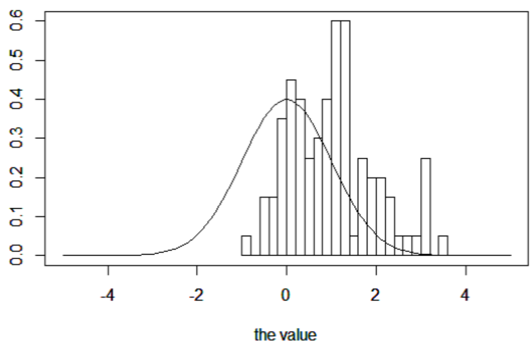

Figure 5-2 :α_{21}^{*}

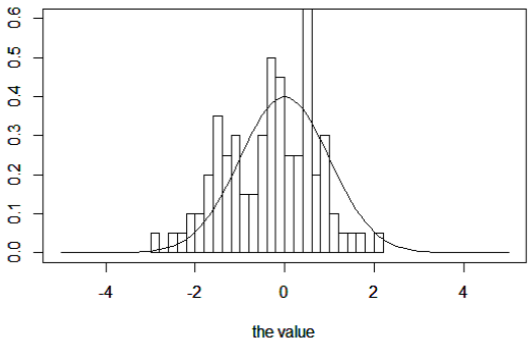

Figure 5-3 :α^*_{23}

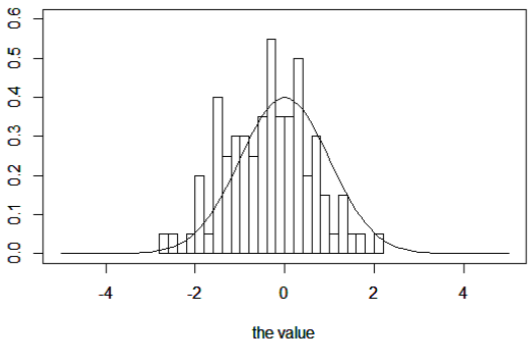

Figure 5-4 :α^*_{31}

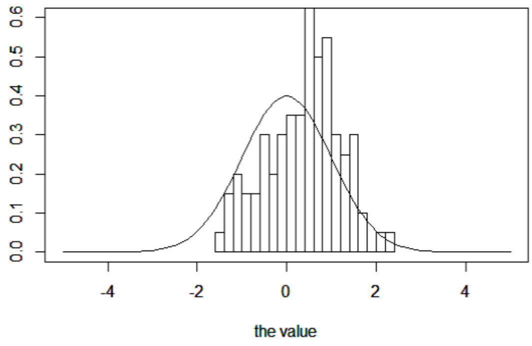

Figure 5-5 :γ^*

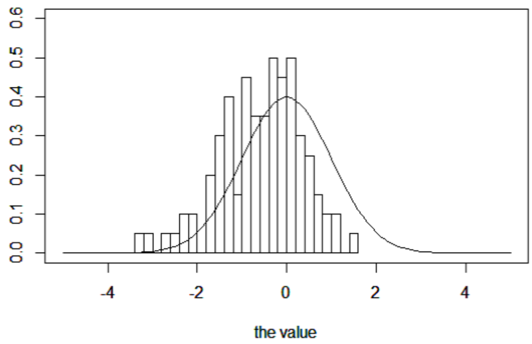

Figure 5-6 :$\lambda^*_{3,0}$

	α^*_{11}	α^*_{12}	α^*_{13}	α^*_{21}	α^*_{22}	α^*_{23}	
True	0.57000	0.00000	0.19000	0.00010	0.71000	0.09500	
Mean	0.63641	0.00259	0.12387	0.03994	0.76318	0.07905	
RMSE	0.01045	0.00426	0.00913	0.00568	0.01004	0.00557	
	α^*_{31}	α^*_{32}	α^*_{33}	γ^*	$\lambda^*_{1,0}$	$\lambda^*_{2,0}$	$\lambda^*_{3,0}$
True	0.05900	0.12000	0.20000	0.02700	0.00930	0.00530	0.00084
Mean	0.06748	0.13922	0.11315	0.02859	0.00853	0.00427	0.00107
RMSE	0.00272	0.00380	0.00963	0.00033	0.00019	0.00017	0.00007

Table 5-1 : Simulation results

restrictions.

We will also use the χ^2-distributions as the limiting distributions of the likelihood ratio statistics for hypotheses testing in our empirical study. We have confirmed that the χ^2- approximations with finite samples are often appropriate.

6 Empirical Applications

In this section we will report two empirical examples by using the SHPP models. The first one is the three major stock markets, namely, Tokyo, New-York, and London. Since there are some time differences when each markets are open and close, it is reasonable to assume that there is no co-jumps. As the second example, we will report the empirical analysis of Tokyo and HK (Hong Kong) markets. Since two markets are open and time difference is negligible, it may be reasonable to use the extended Hawkes models with co-jumps. Our data used

Table 6-1 : Tail Distributions

	Log Likelihood	σ_i^*	ξ_i
J	-1190.72	0.00806	0.16874
SD		0.00065	0.06431
	Log Likelihood	σ_i^*	ξ_i
NY	-797.385	0.00765	0.21538
SD		0.00076	0.08082
	Log Likelihood	σ_i^*	ξ_i
L	-775.779	0.00850	0.10799
SD		0.00084	0.07717

are daily data of Nikkei225, S&P500 and FTSE100 during 1990/1/2-2015/8/25. We have chosen $u = 2\%$ because Kunitomo, Ehara and Kurisu (2016) have analysed this case and obtained reasonable results.

6.1 Example 1 (Tokyo-NY-London)

We first maximize the likelihood L_2 to estimate the marginal distributions of financial market returns. As we have shown in Table 6.1, we have confirmed that the marginal distributions of market returns have thicker tails than the normal distribution. Hence, it may be appropriate to use the generalized Pareto distribution in our estimation. The standard deviations (SD) are estimated by the nimerical evaluation of Fisher Information matrix.

As the estimated models with two dimension ($d = p = 2$), we take the impact functions $c(x)$ as *Case* (1) $c(x) = 1$, *Case* (2) $c(x) = x$, and *Case* (3) $c(x) = x^c$ $(0 < c < 1)$. The estimated values of the log-likelihood and AIC are those with the marginal distributions L_1. The full likelihood can be calculated by using L_1 and L_2. The standard deviations of the estimated coefficients are evaluated numerically by using the inverse of the estimated Fisher information matrix.

Case 1

We estimated the intensity function as

$$\lambda_1^n(t) = \lambda_{10}^n + \int_0^t \alpha_{11} e^{-\gamma(t-s)} dN_1^n(s) + \int_0^t \alpha_{12} e^{-\gamma(t-s)} dN_2^n(s) \ ,$$
$$\lambda_2^n(t) = \lambda_{20}^n + \int_0^t \alpha_{21} e^{-\gamma_1(t-s)} dN_1^n(s) + \int_0^t \alpha_{22} e^{-\gamma(t-s)} dN_2^n(s) \ .$$

The maximum likelihood estimates can be sometimes unstable numerically, we set the restriction that the discounted parameters γ_i have the same value γ. We show the estimation results in Table 6.2.

Table 6-2 (1): Tokyo-NY

	Log Likelihood	AIC	α_{11}	α_{12}	α_{21}	α_{22}	γ	λ_{10}	λ_{20}
Tokyo-NY	-2444.14	4902.27	0.01490	0.00452	0.00000	0.01796	0.0234	0.00583	0.00390
SD			0.002102	0.00162	0.00070	0.00247	0.0030	0.00126	0.00093

In the above table N_1 corresponds to Tokyo and N_2 corresponds to NY in Tokyo-NY markets. In Tokyo-London, N_1 corresponds to

Table 6-2 (2) : Tokyo-London

	Log Likelihood	AIC	α_{11}	α_{12}	α_{21}	α_{22}	γ	λ_{10}	λ_{20}
Tokyo-London	-2421.02	4856.04	0.01692	0.00437	0.00062	0.02028	0.02729	0.00683	0.00361
SD			0.00235	0.00173	0.00073	0.00284	0.00341	0.00126	0.00087

Tokyo while N_2 correponds to London.

Case 2

We estimated the intensity function as

$$\lambda_1^n(t) = \lambda_{10}^n + \int_0^t \alpha_{11} e^{-\gamma(t-s)} X_1 dN_1^n(s) + \int_0^t \alpha_{12} e^{-\gamma(t-s)} X_2 dN_2^n(s) \ ,$$
$$\lambda_2^n(t) = \lambda_{20}^n + \int_0^t \alpha_{21} e^{-\gamma(t-s)} X_1 dN_1^n(s) + \int_0^t \alpha_{22} e^{-\gamma(t-s)} X_2 dN_2^n(s) \ ,$$

and we show our estimation results in Figure 6.3.

Table 6-3 (1) : Tokyo-NY

	Log Likelihood	AIC	α_{11}	α_{12}	α_{21}	α_{22}	γ	λ_{10}	λ_{20}
Tokyo-NY	-2441.59	4897.19	4.83076e-01	1.31388e-01	3.12087e-07	5.81280e-01	2.32784e-02	6.72085e-03	4.31518e-03
SD			0.07110	0.05232	0.02584	0.08305	0.00310	0.00131	0.00095

Case 3

Table 6-3 (2) : Tokyo-London

	Log Likelihood	AIC	α_{11}	α_{12}	α_{21}	α_{22}	γ	λ_{10}	λ_{20}
Tokyo-London	-2418.77	4851.55	0.57164	0.13416	0.02901	0.68684	0.02833	0.007645	0.00388
SD			0.08127	0.05768	0.02905	0.09947	0.0037	0.00130	0.00089

We estimated the intensity function as

$$\lambda_1^n(t) = \lambda_{10}^n + \int_0^t \alpha_{11} e^{-\gamma(t-s)} {X_1}^{c_{11}} dN_1^n(s) + \int_0^t \alpha_{12} e^{-\gamma(t-s)} {X_2}^{c_{12}} dN_2^n(s) \ ,$$
$$\lambda_2^n(t) = \lambda_{20}^n + \int_0^t \alpha_{21} e^{-\gamma(t-s)} {X_2}^{c_{21}} dN_1^n(s) + \int_0^t \alpha_{22} e^{-\gamma(t-s)} {X_2}^{c_{22}} dN_2^n(s) \ .$$

The maximum likelihood estimates are sometimes unstable numerically, we set the restriction that the discounted parameters γ_i have the same value γ and also we set the restriction $c_{11} = c_{12}, c_{21} = c_{22}$. We show the estimation results in Table 6.4.

Table 6-4 (1) : Tokyo-NY

	Log Likelihood	AIC	α_{11}	α_{12}	α_{21}	α_{22}	γ	λ_{10}	λ_{20}	$c_{11} = c_{12}$	$c_{21} = c_{22}$
Tokyo-NY	-2440.769	4899.538	0.0789	0.02271	0.0000	0.0948	0.02320	0.006132	0.00403	0.4739	0.47605
SD			0.3553	0.1005	0.0053	0.1169	0.00305	0.00159	0.00096	1.28043	0.35940

From our estimated results, we find that Model 2 and Model 3 are

Table 6-4 (2) : Tokyo-NY

	Log Likelihood	AIC	α_{11}	α_{12}	α_{21}	α_{22}
Tokyo-London	-2417.74	4853.48	0.21233	0.05087	0.00656	0.17068
SD			0.27649	0.06779	0.01023	0.17829

	γ	λ_{10}	λ_{20}	$c_{1,1}$=$c_{1,2}$	$c_{2,1}$=$c_{2,2}$
	0.02800	0.00730	0.00373	0.71330	0.59962
	0.00363	0.00133	0.00089	0.37337	0.29613

better than Model 1. Also by using AIC Model 2 is better than Model 3. In other worlds, Model 3 has too many parameters and Model 2 is better than Model 3 as Tokyo-NY markets.

Granger-noncausality Tests

We use the non-causality tests based on the likelihood ratio principle. In particular, our results in Appendix include not only the multivariate cases, but also the limiting Fisher information matrix can be random variables. Under a regularity conditions including the conditions of no-cojumps, we summarize the basic results, which is called Wilks-property.

Theorem 6.1 : Let the log-likelihood function of the Hawkes-type point processes with true parameters be $L_T(\theta_0)$, the log-likelihood function with the maximum likelihood estimator $\hat{\theta}_{ML}$ be $L_T(\hat{\theta}_{ML})$ under $\mathbf{\Theta} \in \theta$ and the log-likelihood function with the restricted maximum likelihood estimator $\hat{\theta}_{RML}$ be $L_T(\hat{\theta}_{RML})$ under $\mathbf{\Theta}_1 \in \theta$ ($\mathbf{\Theta}_1 \subset \mathbf{\Theta}$). We assume that the sufficient conditions for the stationarity, the existence of the second order moment condition of $\mathbf{C}(\mathbf{X})$, and the parameter space $\mathbf{\Theta} \in \theta$ in $\mathbf{R}^r$ the parameter space and $\mathbf{\Theta}_1 \in \theta$ in $\mathbf{R}^{r_1}$ $(0 \leq r_1 < r)$ are compact sets. Under a set of regularity conditions (see Theorem A-3 in Appendix), as $T \to \infty$,

$$2\{L_T(\hat{\theta}_{ML}) - L_T(\hat{\theta}_{RML})\} \xrightarrow{d} \chi(r - r_1) \ , \tag{6.27}$$

where $r - r_1$ is the number of restrictions of $\theta = (\theta_k)$ and $\chi(r - r_1)$ is the χ^2- random variable with $r - r_1$ degrees of freedom.

Some details of the regularity conditions will be discussed in Appendix. When we apply the Granger-causality test procedure, we set the impact function as $c(x) = x$. We report our empirical results for the hypothesis $H_0 : \alpha_{ij} = 0$ by using the likelihood ratio test statistics. For the null-hypothesis $H_0 : \alpha_{21} = 0$, the likelihood ratio was $2 \times (-2441.594 + 2441.594) \sim 0$ and we could not reject the null-hypothesis. (The 95% upper-percentage point of $\chi^2(1)$ is 3.481 in Table 6-5(1).) This means that the change of the Japanese financial market has little impact on the U.S. financial market.

Next, for testing the null-hypothesis $H_0 : \alpha_{12} = 0$, the likelihood ratio test statistics was $2 \times (-2441.594 + 2446.297) = 9.406$, and then the null-hypothesis was rejected. This means that there is a significant effect from the U.S. market to the Japanese financial markets (see Table 6-5(2)).

Table 6-5(1) : Tokyo-NY

	Log Likelihood	AIC	α_{11}	α_{12}
Tokyo-NY	-2441.59	4895.19	0.48299	0.13130
SD			0.06967	0.05249

	α_{21}	α_{22}	γ	λ_{10}	λ_{20}
	null	0.58119	0.02327	0.00672	0.00432
	null	0.08278	0.00302	0.00131	0.00084

Table 6-5(2) : Tokyo-NY

	Log Likelihood	AIC	α_{11}	α_{12}
Tokyo-NY	-2446.30	4904.59	0.47794	null
SD			0.06887	null

	α_{21}	α_{22}	γ	λ_{10}	λ_{20}
	0.0000	0.53663	0.021209	0.008161	0.00414
	0.02415	0.07646	0.00279	0.00129	0.00095

Similarly, we have done empirical analysis on Tokyo-London markets. For the null-hypothesis $H_0 : \alpha_{21} = 0$, the likelihood ratio statistic was $2 \times (-2418.773 + 2419.359) = 1.172$ and the null-hypothesis was not rejected. This means that the effect of Japanese financial market on London is rather limited.

For the null-hypothesis $H_0 : \alpha_{12} = 0$, the likelihood ratio statistic was $2 \times (-2418.773 + 2422.848) = 8.15$ and the null-hypothesis was rejected. This means that there is an effect of London market on Tokyo (see Table 6-6(1),(2)).

Table 6-6(1) : Tokyo-London

	Log Likelihood	AIC	α_{11}	α_{12}	α_{21}	α_{22}	γ	λ_{10}	λ_{20}
Tokyo-London	-2419.36	4850.72	0.56271	0.13039	null	0.68456	0.02774	0.00759	0.00442
SD			0.07960	0.05631	null	0.09854	0.00356	0.00130	0.00079

Table 6-6(2) : Tokyo-London

	Log Likelihood	AIC	α_{11}	α_{12}	α_{21}	α_{22}	γ	λ_{10}	λ_{20}
Tokyo-London	-2422.85	4857.70	0.57022	null	0.02494	0.64452	0.02626	0.00874	0.00380
SD			0.08093	null	0.02690	0.09437	0.00346	0.00130	0.00090

To summarize our findings among three major financial markets, the effects of Japanese market on the U.S. and London are rather limited while we can find significant effects of U.S. financial market and London financial market on Tokyo market were rather significant.

6.2 Example 2 : Tokyo-HK markets

For the second example, we have used daily data of Nikkei-225 and Hansen Inde of Hong-Kong(HK) during 1990/1/2-2015/8/25. Since the trading periods in two financial markets are quite similar, there

Table 6-7 : Tail Distributions

	Log Likelihood	σ_i^*	ξ_i
J	-1919.307	0.00757	0.22778
SD		0.00051	0.05552
	Log Likelihood	σ_i^*	ξ_i
HK	-1888.716	0.00861	0.15773
SD		0.00055	0.05076

should be simultaneous movements in two markets. Because there can be many additional parameters in Case 3, the estimated results are often not statistically significant and we omit the results of Case 3.

We first maximize the likelihood L_2 to estimate the marginal distributions of financial market returns. As we have shown before, we have confirmed that the marginal distributions of market returns have thicker tails than the normal distribution as in Table 5.9. Hence, it may be appropriate to use the generalized Pareto distribution in our estimation.

The estimated models with two dimensions ($d = 2$ and $p = 3$), we take the impact functions $c(x)$ as $Case$ (1) $c(x) = 1$ and $Case$ (2) $c(x) = x$. The estimated values of the log-likelihood and AIC are those with the marginal distributions L_1. The full likelihood can be calculated by using L_1 and L_2^*. The standard deviations of the estimated coefficients are evaluated numerically by using the inverse of the estimated Fisher information matrix.

Case 1

We estimated the intensity function as

$$\begin{aligned}
\lambda_1^n(t) &= \lambda_{10}^n + \int_0^t \alpha_{11} e^{-\gamma(t-s)} dN_1^n(s) + \int_0^t \alpha_{12} e^{-\gamma(t-s)} dN_2^n(s) \\
&\quad + \int_0^t \alpha_{13} e^{-\gamma(t-s)} dN_{1,2}^n(s) \ , \\
\lambda_2^n(t) &= \lambda_{20}^n + \int_0^t \alpha_{21} e^{-\gamma(t-s)} dN_1^n(s) + \int_0^t \alpha_{22} e^{-\gamma(t-s)} dN_2^n(s) \\
&\quad + \int_0^t \alpha_{23} e^{-\gamma(t-s)} dN_{1,2}^n(s) \ , \\
\lambda_{12}^n(t) &= \lambda_{12,0}^n + \int_0^t \alpha_{31} e^{-\gamma(t-s)} dN_1^n(s) + \int_0^t \alpha_{32} e^{-\gamma(t-s)} dN_2^n(s) \\
&\quad + \int_0^t \alpha_{33} e^{-\gamma(t-s)} dN_{12}^n(s) \ .
\end{aligned}$$

The maximum likelihood estimates can be sometimes unstable numerically, we set the restriction that the discounted parameters γ_i have the same value γ. We show the estimation results in Table 6-8.

Table 6-8 : Tokyo-HK

	Log Likelihood	AIC	α_{11}	α_{12}	α_{13}
Tokyo-HK	-3954.73	7935.47	0.015	0.000	0.012
SD			0.002	0.001	0.0036

	α_{31}	α_{32}	α_{33}
Tokyo-HK	0.0015	0.0035	0.0086
SD	0.0007	0.0008	0.0022

	α_{21}	α_{22}	α_{23}	γ	λ_1	λ_2	λ_3
Tokyo-HK	0.000	0.020	0.007	0.0262	0.0090	0.0048	0.0008
SD	0.00074	0.0025	0.0033	0.0028	0.0016	0.0012	0.0007

In the above table N_1 corresponds to Tokyo and N_2 corresponds to NY in Tokyo-NY markets.

Case 2

We estimated the intensity function as

$$\begin{aligned}
\lambda_1^n(t) &= \lambda_{10}^n + \int_0^t \alpha_{11} e^{-\gamma(t-s)} X_1 dN_1^n(s) + \int_0^t \alpha_{12} e^{-\gamma(t-s)} X_2 dN_2^n(s) \\
&\quad + \int_0^t \alpha_{13} e^{-\gamma(t-s)} [\max_i X_i] dN_{1,2}^n(s)\ , \\
\lambda_2^n(t) &= \lambda_{20}^n + \int_0^t \alpha_{21} e^{-\gamma(t-s)} X_1 dN_1^n(s) + \int_0^t \alpha_{22} e^{-\gamma(t-s)} X_2 dN_2^n(s) \\
&\quad + \int_0^t \alpha_{23} e^{-\gamma(t-s)} [\max_i X_i] dN_{12}^n(s)\ , \\
\lambda_{12}^n(t) &= \lambda_{12,0}^n + \int_0^t \alpha_{31} e^{-\gamma(t-s)} X_1 dN_1^n(s) + \int_0^t \alpha_{32} e^{-\gamma(t-s)} X_2 dN_2^n(s) \\
&\quad + \int_0^t \alpha_{33} e^{-\gamma(t-s)} [\max_i X_i] dN_{1,2}^n(s)\ .
\end{aligned}$$

We show the estimation results in Figure 6.9.

Table 6-9 : Tokyo-HK

	Log Likelihood	AIC	α_{11}	α_{12}	α_{13}
Tokyo-HK	-3944.79	7915.58	0.5675	0.000	0.1930
SD			0.764	0.0373	0.0738

	α_{31}	α_{32}	α_{33}
Tokyo-HK	0.0586	0.1242	0.0547
SD	0.0007	0.0008	0.0022

	α_{21}	α_{22}	α_{23}	γ	λ_1	λ_2	λ_3
Tokyo-HK	0.0001	0.7147	0.0950	0.0267	0.0094	0.0053	0.0008
SD	0.0241	0.0871	0.0701	0.0029	0.0016	0.0012	0.0007

From our estimated results, we find that Model 2 is better than Model 1.

Granger-noncausality Tests

We use the non-causality tests based on the likelihood ratio principle. Although we allow the possible co-jumps, it is possible to apply the martingale central limit (MCLT) theorem for point processes. For the sake of completeness, we re-state the Wilks-property.

Theorem 6.2 : Let the log-likelihood function of the Hawkes-type point processes with true parameters be $L_T(\theta_0)$, the log-likelihood function with the maximum likelihood estimator $\hat{\theta}_{ML}$ be $L_T(\hat{\theta}_{ML})$ under $\mathbf{\Theta} \in \theta$ and the log-likelihood function with the restricted maximum likelihood estimator $\hat{\theta}_{RML}$ be $L_T(\hat{\theta}_{RML})$ under $\mathbf{\Theta}_1 \in \theta$ ($\mathbf{\Theta}_1 \subset \mathbf{\Theta}$). We assume that the sufficient conditions for the stationarity, the existence of the second order moment condition of $\mathbf{C}(\mathbf{X})$, and the parameter space $\mathbf{\Theta} \in \theta$ in $\mathbf{R}^r$ the parameter space and $\mathbf{\Theta}_1 \in \theta$ in $\mathbf{R}^{r_1}$ $(0 \leq r_1 < r)$ are compact sets. Under a set of regularity conditions (see Theorem A-3 in Appendix), as $T \to \infty$,

(i) The maximum likelihood estimator $\hat{\theta}_{ML}$ under the misspecified likelihood function $L_T(\hat{\theta}_{ML})$ is not consistent when there are some co-jumps.

(ii) Under a set of regularity conditions even when co-jumps exist, as $T \to \infty$,

$$2\{L_T(\hat{\theta}_{ML}) - L_T(\hat{\theta}_{RML})\} \xrightarrow{d} \chi(r - r_1) \ , \tag{6.28}$$

where $r - r_1$ is the number of restrictions of $\theta = (\theta_k)$ and $\chi(r - r_1)$ is the χ^2- random variable with $r - r_1$ degrees of freedom.

Some details of the regularity conditions will be discussed in Appendix. When we apply the Granger-causality test procedure, we set the impact function as $c(x) = x$. We report our empirical results for the hypothesis $H_0 : \alpha_{ij} = 0$ by using the likelihood ratio test statistics.

For the null-hypothesis $H_0 : \alpha_{13} = 0$, the likelihood ratio was 11.14 and we reject the null-hypothesis. (The 95% upper-percentage point of $\chi^2(1)$ is 3.481 in Table 6-10(1)). This means that we have a significant instantaneous causal relation between the Japanese financial market and Hong-Kong financial market.

Table 6-10(1) : Tokyo-HK

	Log Likelihood	AIC	α_{11}	α_{12}	α_{13}
Tokyo-HK	-3950.36	7924.72	0.6163	0.0023	null
SD			0.077	0.028	null

	α_{31}	α_{32}	α_{33}
Tokyo-HK	0.057	0.124	0.02055
SD	0.027	0.030	0.054

	α_{21}	α_{22}	α_{23}	γ	λ_1	λ_2	λ_3
Tokyo-HK	0.000	0.701	0.092	0.0262	0.0103	0.0012	0.0007
SD	0.024	0.085	0.069	0.0029	0.0017	0.0012	0.0007

Next, for testing the null-hypothesis $H_0 : \alpha_{12} = 0$, the likelihood ratio test statistics was 0.0, and then the null-hpothesis was accepted.

Next, for testing the null-hypothesis $H_0 : \alpha_{12} = 0, \alpha_{13} = 0$ the likelihood ratio test statistics was 11.14, and then the null-hypothesis was rejected (see Table 6-10(3)).

Table 6-10(2) : Tokyo-HK

	Log Likelihood	AIC	α_{11}	α_{12}	α_{13}
Tokyo-HK	-3944.79	7913.58	0.5644	null	0.1932
SD			0.076	null	0.068

	α_{31}	α_{32}	α_{33}
Tokyo-HK	0.058	0.1229	0.2071
SD	0.0268	0.0297	0.0544

	α_{21}	α_{22}	α_{23}	γ	λ_1	λ_2	λ_3
Tokyo-HK	0.000	0.7099	0.0916	0.0265	0.0093	0.0053	0.0008
SD	0.0239	0.0895	0.0691	0.0029	0.0015	0.0012	0.0007

For the null-hypothesis $H_0 : \alpha_{21} = 0$, the likelihood ratio was 0.006 and we reject the null-hypothesis. (The 95% upper-percentage point of $\chi^2(1)$ is 3.481 in Table 6-10(4)).

Next, for testing the null-hypothesis $H_0 : \alpha_{23} = 0$, the likelihood ratio test statistics was 2.42, and then the null-hypothesis was accepted(see Table 6-10(5)).

Similarly, for testing the null-hypothesis $H_0 : \alpha_{21} = 0, \alpha_{23} = 0$ the likelihood ratio test statistics was 2.66, and then the null-hypothesis could not be rejected (see Table 10-6(6)).

To summarize our findings among Tokyo and Hong Kong financial markets, the simultaneous effects of two markets are significant.

6.3 A Further Empirical Analysis

We also use the spectral decomposition and the relative power contributions as we explained in Section 3. Three figures of US, UK and HK

Table 6-10(3) : Tokyo-HK

	Log Likelihood	AIC	α_{11}	α_{12}	α_{13}
Tokyo-HK	-3950.36	7922.72	0.6181	null	null
SD			0.0762	null	null

	α_{31}	α_{32}	α_{33}
Tokyo-HK	0.0576	0.1229	0.2077
SD	0.0267	0.0297	0.0541

	α_{21}	α_{22}	α_{23}	γ	λ_1	λ_2	λ_3
Tokyo-HK	0.00003	0.7047	0.09145	0.0263	0.0104	0.0053	0.0008
SD	0.0237	0.0854	0.0688	0.0029	0.0015	0.0012	0.0007

are given as Figure 6-1, 6-2 and 6-3. In the first two decompositions we assume that there are no co-jumps while in the last one we do have co-jumps terms. We have adopted the cases when $c_{ij}(x) = x$ because the resulting models have the minimum AIC.

For the relashionship between Tokyo-NY financial markets, the self contribution play major contribution while there is some contribution from NY to Tokkyo in the low frequency, which corresponds to the long-run relation. On the other hand, for the relashionship between Tokyo-HK financial markets, the instantaneous contribution plays a major contribution in all frequencies as well as the self contribution. This aspect reflects the fact that we have used the SHPP models.

Figure 6-1 : Relative Power Contributions

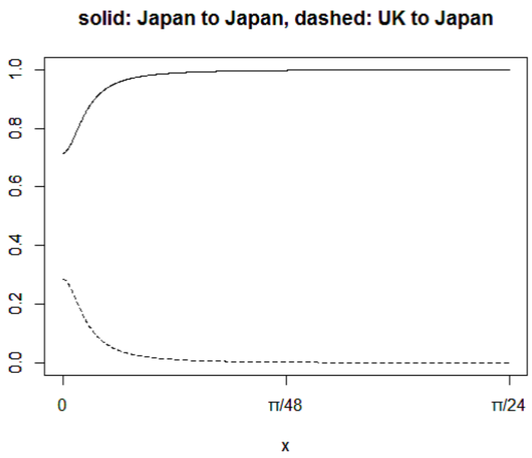

Figure 6-2 : Relative Power Contributions

Table 6-10(4) : Tokyo-HK

	Log Likelihood	AIC	α_{11}	α_{12}	α_{13}
Tokyo-HK	-3944.79	7913.58	0.5647	0.0000	0.1912
SD			0.0759	0.0371	0.0733

	α_{31}	α_{32}	α_{33}
Tokyo-HK	0.0595	0.1229	0.2084
SD	0.0271	0.0298	0.0547

	α_{21}	α_{22}	α_{23}	γ	λ_1	λ_2	λ_3
Tokyo-HK	null	0.7082	0.0972	0.0266	0.0094	0.0053	0.0008
SD	null	0.0859	0.0679	0.0029	0.0016	0.0010	0.0007

7 Conclusions

In this paper we developed a new method of econometric analysis of multivariate time series of events and proposed the simultaneous Hawkes-type point process modeling. Unlike some existing literatures, we can use statistical models for simultaneous sudden and large events and delayed events occurred explicitly. By using the simultaneous multivariate Hawkes-type point process approach and the SHPP models, we have investigated the Granger-causality and the instantaneous Granger causality on different financial markets and economies and developed the non-causality tests.

By applying the non-causality tests for both the Granger non-causality (GNC) and the Granger instantaneous non-causality (GINC), we have found the important relations among major financial markets and several empirical findings. In Tokyo-NY financial markets, there

Table 6-10(5) : Tokyo-HK

	Log Likelihood	AIC	α_{11}	α_{12}	α_{13}
Tokyo-HK	-3946.0	7916.0	0.5628	0.000	0.1911
SD			0.0754	0.0369	0.0732

	α_{31}	α_{32}	α_{33}
Tokyo-HK	0.0577	0.1230	0.2067
SD	0.0267	0.0297	0.0543

	α_{21}	α_{22}	α_{23}	γ	λ_1	λ_2	λ_3
Tokyo-HK	0.0112	0.737		0.0264	0.0093	0.0054	0.0008
SD	0.0239	0.0857		0.0029	0.0016	0.0012	0.0007

is a strong one way direction in causation while in Tokyo-HK financial markets the simultaneous effects are dominant.

There are several questions remained to be answered. Although we have used the Hawkes-type point processes, there can be many possible non-linear point processes and it may be interesting to investigate the robustness of our empirical results. Also the choice of threshold parameters is an important one, which is related to the relevance of the generalized Pareto distribution (GPD) in the statistical extreme value theory and we need a more convincing statistical theory on the choice of theresholds. These questions are currently under investigation.

References

[1] Ait-Sahalia, Y. and J. Jacod (2014), *High-Frequency Financial Econometrics*, Princeton University Press.

Table 6-10(6) : Tokyo-HK

	Log Likelihood	AIC	α_{11}	α_{12}	α_{13}
Tokyo-HK	-3946.12	7914.24	0.5600	0.000	0.1929
SD			0.0750	0.0368	0.0733

	α_{31}	α_{32}	α_{33}
Tokyo-HK	0.05690	0.1227	0.2061
SD	0.0265	0.0297	0.0541

	α_{21}	α_{22}	α_{23}	γ	λ_1	λ_2	λ_3
Tokyo-HK	null	0.7361	null	0.0263	0.0093	0.0057	0.0008
SD	null	0.0855	null	0.0029	0.0016	0.0016	0.0007

[2] Bartrett, M.S. (1963) “The Spectral Analysis of Point Processes,” *Journal of Royal Statistical Society (B)*, Vol.25-2, 264-296.

[3] Daley, D.J. and D. Vere-Jones (2003), *An Introduction to the Theory of Point Processes*, Volume I, Volume II, 2nd Edition, Springer.

[4] Florens, J-P. and D. Fougere (1996) “Noncausality in Continuous Time,” *Econometrica*, 64, 1195-1212.

[5] Granger, C.W. (1969) “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods,” *Econometrica*, 37, 161-194.

[6] Grothe, O., V. Korniichuk and H. Mannera (2014) “Modeling multivariate extreme events using self-exciting point processes,” *Journal of Econometrics*, 182, 269-289.

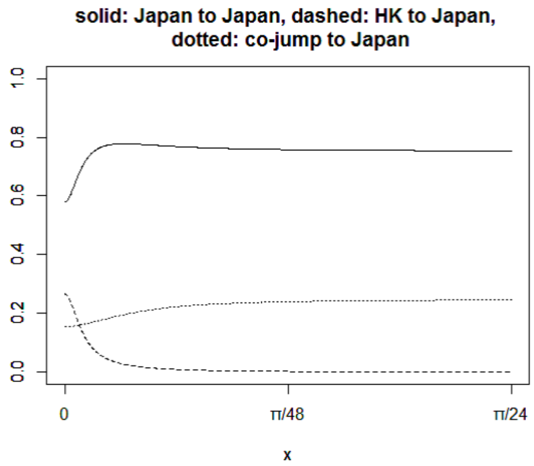

Figure 6-3 : Relative Power Contributions

[7] Hawkes, A. G. (1971) “Point Spectra of Some Mutually Exciting Point Processes,” *Journal of the Royal Statistical Society. Series B*, 33-3, 438-443.

[8] Kunitomo,N., A. Ehara and D. Kurisu (2016) ”A Causality Analysis of Financial Markets by Multivariate Hawkes-type Models,”(in Japanese), CIRJE, Discussion Paper J-278, CIRJE, University of Tokyo.

[9] Ogata, Y. (1978) “The Asymptotic Behavior of Maximum Likelihood Estimators of Stationary Point Processes,” *Annal of Institute of Statistical Mathematics*, 30, 243-261.

[10] Ogata, Y. (2015) “Studies of Probabilistic Prediction of Earthquakes: A Survey,” *Statistical Mathematics*, 63-11, 3-27 (in Japanese).

[11] Protter, P. (2003), *Stochastic Integration and Differential Equations*, Springer.

[12] Resnick, S. (2007), *Heavy-Tail Phenomena*, Springer.

[13] Solo, V. (2007) "Likelihood Functions for Multivariate Point Processes with Coincidences," *Proceedings of the 46th IEEE Conference on Decicon and Control*, 4245-4250.

APPENDIX : Mathematical Details

In this Appendix, we give some mathematical details we have used in the previous sections. In the statistical analysis of point processes, Ogata (1978) derived the asymptotic properties of consistency and asymptotic normality of the maximum likelihood estimation for one dimensional intensity models, which have been classical and often cited in the related studies. He obtained the results by using a martingale central limit (CLT) theorem for point processes, which has not been well-known for econometricians, and also the asymptotic normality holds under more general conditions often cited. Hence, we first discuss some properties of jump martingales with continuous time parameter. We omit the subscript n without any loss of generality in this Appendix.

(i) **A Martingale CLT**

We present a general martingale CLT for one-dimensional point processes and then we can apply to our situation as an application.

Theorem A.1 : Let an $\mathcal{F}-$adapted simple point process on $\mathbf{R}_+$ be N and the $\mathcal{F}-$(continuous)compensator be A. We assume that for any T (> 0) there exists an $\mathcal{F}-$adapted $g_T(t)$ and an $\mathcal{F}_0-$adapted (positive) random variable η, which satisfy the following conditions.
(i) $\mathbf{E}[\frac{1}{T}\int_0^T (g_T(x))^2 dA(x)] < \infty$,
(ii) For any δ (> 0),

(A.29) $$\frac{1}{T^{1+\delta}}A(T) \xrightarrow{p} 0,$$

(iii) As $T \longrightarrow \infty$

$$\frac{1}{T}\int_0^T (g_T(x))^2 dA(x) \xrightarrow{p} \eta^2, \tag{A.30}$$

(iv) For any $c > 0$, as $c \to \infty$

$$\mathbf{E}[\frac{1}{T}\int_0^T (g_T(x)I(|g_T(x)| > c))^2 dA(x)|\mathcal{F}_0] \xrightarrow{p} 0 \ . \tag{A.31}$$

Then

$$X_T = \frac{1}{\sqrt{T}}\int_0^T g_T(x)[dN(x) - dA(x)] \tag{A.32}$$

converges to $U\eta$ in the sense of $\mathcal{F}_0-$(stable convergence sense), where U is $N(0,1)$, which is independent of $\mathcal{F}_0$.

Note : The method of proof is basically a modification of the one given in Daley=Vere-Jones (2008, VolII), Theorem 14.5.I. They derived a martingale CLT under a Lyapunov condition. Our condition includes the speed of compensator, which may be a reasonable condition.

Proof : For any real number y and $f_T(u) = (1/\sqrt{T})g_T(u)$, we define

$$\zeta_T(t,y) = \exp\left(iy\int_0^t f_T(u)[dN(u) - dA(u)] + \frac{1}{2}y^2\int_0^t [f_T(u)]^2 dA(u)\right) \ . \tag{A.33}$$

By using **Lemma A-1**, when $A(t)$ and $N(t)$ are a continuous process and a pure jump process, respectively, we can represent

$$\begin{aligned} \zeta_T(t,y) &= \exp\left((\frac{1}{2}y^2[f_T(u)]^2 - iy\int_0^t f_T(u))dA(u)\right) \\ &\quad \times \prod_i [1 + (\exp(iy\int_0^t f_T(t_i) - 1)\Delta N(t_i)] \ , \end{aligned} \tag{A.34}$$

where t_i are jump times. By using the transformation of jump process, we have

$$
\begin{aligned}
&= \zeta_T(t,y) - 1 = \int_0^t \zeta_T(u-,y)\left[\frac{1}{2}y^2[f_T(u)]^2 - iyf_T(u)]dA(u)\right. \\
&\quad \left.+[\exp(iyf_T(u)) - 1]dN(u)\right] \\
&= \int_0^t \zeta_T(u-,y)(\exp(iyf_T(u)) - 1)(dN(u) - dA(u)) \\
&\quad + \int_0^t \zeta_T(u-,y)\left[\exp(iyf_T(u)) - 1 - iyf_T(u) + \frac{1}{2}y^2[f_T(u)]^2\right]dA(u) \; .
\end{aligned}
$$

We define the stopping time τ by $\tau = \inf\{t : \int_0^T [f_T(u)]^2 dA(u) \geq \eta^2\}$. Then for any $\mathcal{F}_0-$measureable and essentially bounded random variable Z, we set $t = T \wedge \tau$. By the martingale property we have

$$
\mathbf{E}\left[Z\int_0^{T\wedge\tau} \zeta_T(u-,y)(\exp(iyf_T(u)) - 1)(dN(u) - dA(u))|\mathcal{F}_0\right] = 0 \; .
$$

Hence

$$
|\mathbf{E}(Z\zeta_T(T\wedge\tau)|\mathcal{F}_0] - Z)| \leq \mathbf{E}[|Z|\int_0^{T\wedge} |\zeta(u-,y)R(f_T(u),y)|dA(u)|\mathcal{F}_0] \; ,
$$

where

$$
R(f_T(u),y) = \exp(iyf_T(u)) - 1iyf_T(u) + \frac{1}{2}y^2[f_T(u)]^2 \; .
$$

For $0 < u < T\wedge\tau$ we have

$$
|\zeta_T(T\wedge\tau)| \leq \exp(\frac{1}{2}y^2\int_0^{T\wedge\tau}[f_T(u)]^2 dA(u)) \leq \exp(\frac{1}{2}y^2\eta^2) \; .
$$

Also by the Taylor-expansion,

$$
|R(f_T(u),y)| \leq y^2|f_T(u)|^2 I[|f_T(u)| > c_T] + \frac{|\theta y|^3}{3!}|f_T(u)|^3 I[|f_T(u)| \leq c_T]
$$

and then

$$
\begin{aligned}
& |\mathbf{E}(Z\zeta_T(T\wedge\tau)|\mathcal{F}_0] - Z| \\
\leq\ & |Z|[y^2\int_0^{T\wedge\tau} |f_T(u)|^2 I[|f_T(u)| > c_T]dA(u) \\
& +y^3\int_0^{T\wedge\tau} |f_T(u)|^3 I[|f_T(u) \leq c_T]dA(u)\ ,
\end{aligned}
$$

where $|\theta| \leq 1$. Therefore the right-hand side multiplying $\exp(-1/2y^2\eta^2)$ is bounded by

$$
|\mathbf{E}(Z[\rho_T e^{iyX_T} - e^{-1/2y^2\eta^2}])|\ ,
$$

where

$$
\rho_T = \exp\left[iy\int_{T\wedge\tau}^T f_T(u)([dN(u)-dA(u)] - \frac{1}{2}(\eta^2 - \int_0^T [f_T(u)]^2 dA(u))_+\right]\ .
$$

We set $g_T(u) = f_T(u)/\sqrt{T}$ and $c = c_T/\sqrt{T}$. Then

$$
\int_0^{T\wedge\tau} |f_T(u)|^3 I[|f_T(u) \leq c_T]dA(u) \leq \frac{c^3}{T^{3/2}}A(T\wedge\tau)\ ,
$$

which converges to zero by our conditions. Here we have

$$
\begin{aligned}
\zeta_T(u-,y)e^{-y^2\eta^2/2} &= e^{iyX_T}\left[e^{iy\int_0^{T\wedge\tau} f_T(u)(dN-dA)+\frac{y^2}{2}\int_0^T f_T(u)^2 dA - iy\int_0^T f_T(u)(dN-dA) - \frac{y^2\eta^2}{2}}\right] \\
&= e^{iyX_T}\rho_T\ .
\end{aligned}
$$

Because $|\rho_T| \leq 1$ and $\rho \to 1$, we find that $\mathbf{E}[Z(\rho_T - 1)e^{itX_T})] \to 0$ and then

$$
\mathbf{E}[Z\exp(iyX_T)] \longrightarrow \mathbf{E}[Ze^{-\frac{1}{2}y^2\eta^2/2}]\ .
$$

Then by the use of weak-convergence and stable convergence (Dalay=Vere-Jones(2008), Jacod=Protter (2012)), we have that $X_T \longrightarrow X$ ($\mathcal{F}_0$)-stably). This means that for any bounded $\mathcal{F}_0-$measurable random variable Z, $\mathcal{E}[Ze^{iyX}] = \mathcal{E}[Ze^{-y^2\eta^2/2}]$, which implies $\mathcal{E}[e^{iyX_T/\eta}|\mathcal{F}_0] = e^{-y^2/2}$.

Q.E.D.

We give the integration-by-parts formula, which has been known in stochastic analysis (see Chapter II of Protter (2003), for instance).

Lemma A.1 : Let

$$\text{(A.35)}\; G_1(t) = \prod_i (1 + w(t_i))\Delta N(t_i), G_2(t) = \exp\left(\int_0^t v(u)dA(u)\right),$$

where $v(u) = (y^2/2)[f_T(u)]^2 - iyf_T(u)$ and $w(t_i)\exp(iyf_T(t_i) - 1)$. Then by the integration-by-parts formula,

$$\begin{aligned}
\text{(A.36)}\;\; & G_1(t)G_2(t) - G_1(0)G_2(0) \\
= \;\; & \int_0^t G_1(u)dG_2(u) + \int_0^t G_2(u)dG_1(t) \\
= \;\; & \int_0^t G_1(u-)G_2(u)v(u)dA(u) + \sum_i G_2(t_i)G_1(t_i-)w(t_i)\Delta N(t_i) \; .
\end{aligned}$$

By using *Theorem A.1*, it is straightforward to obtain a martingale convergence result under the same assumptions of *Theorem A.1.* That is, for any $\epsilon > 0$ we have

$$\text{(A.37)} \qquad Y_T = \frac{1}{T^{1/2+\epsilon}} \int_0^T g_T(x)[dN(x) - dA(x)] \xrightarrow{p} 0 \; .$$

Thus, we do not need to use the *Ergodic Thorem* for stationary stochastic processes, which was one of key arguments on the asymptotic results obtained by Ogata (1978).

It is also straightforward to extend *Theorem A.1* to the multivariate

cases. Let $\mathbf{N} = (N_i)$ be a $p \times 1$ vector $\mathcal{F}-$adapted simple point processes on $\mathbf{R}_+$ and $\mathbf{A} = (A_k)$ are the $\mathcal{F}-$(continuous)compensators. For any $T\ (> 0)$ we consider $q \times p$ $\mathcal{F}-$adapted and predictable processes $\mathbf{g}_T(t) = (g_T^{ij}(t))$ and a $q \times q$ $\mathcal{F}_0-$adapted (positive-definite) random matrix $\boldsymbol{\eta} = (\eta_{ij})$, we assume the following conditions.
$(i)'$ $\max_{1\le i,j\le q} \max_{1\le k\le p} \mathbf{E}[\frac{1}{T}\int_0^T |g_T^{ik}(t)||g_T^{ik}(t)|dA_k(t)] < \infty$,
$(ii)'$ For any $\delta\ (> 0)$,

$$\text{(A.38)} \qquad \frac{1}{T^{1+\delta}} \max_{1\le k\le p} A_k(T) \xrightarrow{p} 0,$$

$(iii)'$ As $T \longrightarrow \infty$

$$\text{(A.39)} \qquad \frac{1}{T}\int_0^T \sum_{k=1}^{p} g_T^{ik}(t) g_t^{jk}(x) dA_k(t) \xrightarrow{p} \eta_{ij}\ ,$$

where $\boldsymbol{\eta} = (\eta_{ij})$ is a $q \times q$ non-negative definite matrix.
$(iv)'$ For any $c > 0$, as $c \to \infty$

$$\text{(A.40)} \max_{1\le k\le p} \mathbf{E}[\frac{1}{T}\int_0^T \|\mathbf{g}_T^{\cdot,k}(t)\|^2 I(\|\mathbf{g}_T^{\cdot,k}(t)\| > c) dA_k(t)|\mathcal{F}_0] \xrightarrow{p} 0\ ,$$

where $\mathbf{g}_T^{\cdot k}(t) = (g_T^{1,k}, \cdots, g_T^{p,k})'$.
Then we have the result.

Theorem A.2 : For the proint processes $\mathbf{N} = (N_i)$ and their compensators $\mathbf{A} = (A_i)$ stated, we assume the conditions $(i)' - (iv)'$. Then a $q \times 1$ vector process

$$\text{(A.41)} \qquad \mathbf{X}_T = \frac{1}{\sqrt{T}}\int_0^T \sum_{i=1}^{p} \mathbf{g}_T^{\cdot,k}(t)[dN_k(t) - dA_k(t)]$$

converges to $\boldsymbol{\eta}^{1/2}\mathbf{U}$ in the sense of $\mathcal{F}_0-$(stable convergence sense), where $\mathbf{U}$ is $N_q(\mathbf{0}, \mathbf{I}_q)$, which is independent of $\mathcal{F}_0$ and we have used the notation $\boldsymbol{\eta}^{1/2}\boldsymbol{\eta}^{1/2} = \boldsymbol{\eta}$.

(ii) **A Wilks Property**

We consider the parametric point process models for the case when the intensity function is $\lambda_i(s,\theta)$ for the point processes $N_i(s,\theta)$ $(i = 1,\cdots,p)$ over the observation period $[0,T]$. We take $\boldsymbol{\theta} = (\theta_i) \in \boldsymbol{R}^r$. Then the log-likelihood function is given by

$$L_T(\theta) = \sum_{i=1}^{p} L_{iT}(\theta) \ , \tag{A.42}$$

where

$$L_{iT}(\theta) = \int_0^T \log \lambda_i(s,\theta) dN_i(s) - \int_0^T \lambda_i(s,\theta) ds \ , \tag{A.43}$$

and its derivatives are given by

$$\frac{\partial L_{iT}(\theta)}{\partial \theta} = \int_0^T \frac{\log \lambda_i(s,\theta)}{\partial \theta} [dN_i(s) - \lambda_i(s,\theta)ds] \ , \tag{A.44}$$

and

$$\frac{\partial^2 L_{iT}(\theta)}{\partial \theta \partial \theta'} = \int_0^T \frac{1}{\lambda_i} \frac{\partial^2 \lambda_i}{\partial \theta \partial \theta'} [dN_i(s) - \lambda_i(s,\theta)ds] - \int_0^T [\frac{\log \lambda_i(s,\theta)}{\partial \theta} \partial \theta^{'}] \lambda_i(s,\lambda) ds \ . \tag{A.45}$$

Theorem A.3 : Let the log-likelihood function be $L_T(\theta)$, the log-likelihood function under the true parameter θ_0 be $L_T(\theta_0)$, and the log-likelihood function under the maximum likelihood estimator $\hat{\theta}_{ML}$ be $L_T(\hat{\theta}_{ML})$. Then under the following regularity conditions as $T \to \infty$

$$2\{L_T(\hat{\theta}_{ML}) - L_T(\theta_0)\} \xrightarrow{d} \chi(r) \ , \tag{A.46}$$

where r is the dimension of $\theta = (\theta_k)$ and $\chi(r)$ is the χ^2−distribution with degrees of freedom r. The conditions are

$$\frac{1}{T}\sum_{i=1}^{p}\int_0^T [\frac{\partial \log \lambda_i}{\partial \theta}\frac{\partial \log \lambda_i}{\partial \theta'}]\lambda_i(s,\theta)ds \xrightarrow{p} I(\theta_0) \ > 0 \text{ (a positive definite matrix)}, \tag{A.47}$$

$$\frac{1}{\sqrt{T}}\sum_{i=1}^{p}\int_0^T [\frac{\partial \log \lambda_i}{\partial \theta}][dN_i(s) - \lambda_i(s,\theta)ds] \xrightarrow{w} N_r(0,\mathbf{I}(\theta_0)) \ , \tag{A.48}$$

$$\frac{1}{T}\sum_{i=1}^{p}\int_0^T [\frac{\partial^2 \lambda_i}{\partial \theta \partial \theta'}]\frac{1}{\lambda_i}[dN_i(s) - \lambda_i(s,\theta)ds] \xrightarrow{p} 0 \ , \tag{A.49}$$

and

$$\frac{1}{T}\sum_{i=1}^{p}\int_0^T [\frac{\partial \log \lambda_i}{\partial \theta}\frac{\partial \log \lambda_i}{\partial \theta'}][dN_i(s) - \lambda_i(s,\theta)ds] \xrightarrow{p} 0 \ , \tag{A.50}$$

where $\mathbf{I}(\theta_0)$ is the Fisher information matrix.

As Corollaries of *Theorem A.2*, it may be straightforward to give the proofs of Theorem 6.1 and Theorem 6.2 as the non-causality tests we have used in Section 6.

Finally, we should notice that while Ogata (1978) has discussed a set of sufficient conditions for the consistency and the asymptotic normality of the ML estimator in one-dimensional self-exciting point processes, we have extended his results significantly to the multivariate point processes under a set of weaker conditions. For instance, $\mathbf{I}(\theta_0)$ is not necessarily a constant matrix and our conditions means the mixed Gaussianity in our formulation. Then the limiting χ^2 property of the statistics is called the Wilks Property of statistics. As an example, Kunitomo, Ehara and Kurisu (2016) have used Theorem 6.1 for the non-causality test in the sense of Granger.