

Simultaneous multivariate Hawkes-type point processes and their application to financial markets

メタデータ	言語: eng 出版者: Springer 公開日: 2020-03-30 キーワード (Ja): キーワード (En): 作成者: Naoto, Kunitomo, Daisuke, Kurisu, Naoki, Awaya メールアドレス: 所属:
URL	http://hdl.handle.net/10291/20760



Simultaneous multivariate Hawkes-type point processes and their application to financial markets

Naoto Kunitomo¹ · Daisuke Kurisu² · Naoki Awaya³

Received: 29 November 2017 / Accepted: 22 July 2018 / Published online: 10 August 2018
© The Author(s) 2018

Abstract

In economic and financial time series we sometimes observe sudden and large price jumps. Although these events are relatively rare, they have significant impacts on not only a given financial market but also several different markets and wider macro economies. Using simultaneous Hawkes-type multivariate point process (SHPP) models, it is possible to analyze the causal effects of large events in the sense of the Granger-non-causality (GNC) from one market to other markets as well as the instantaneous Granger-non-causality (IGNC). We investigate the financial market of Tokyo and other major markets, and apply GNC tests to investigate the interdependence of large events among markets. Several important empirical findings emerge among financial markets and wider macro economies.

Keywords Simultaneous Hawkes-type marked point process (SHPP) · Granger-non-causality · Instantaneous non-causality · Non-causality tests · International financial markets

Mathematics Subject Classification MSC62P20 · 91B84

A revised version, July 2018.

✉ Naoto Kunitomo
naoto.kunitomo@gmail.com

¹ School of Political Science and Economics (Sarugakucho 3rd Building C-106), Meiji University, 1-1 Kanda-Surugadai, Chiyoda-ku, Tokyo 101-8301, Japan

² School of Engineering, Tokyo Institute of Technology, Meguro-ku, Ookayama 2-12-1 W9-74, Tokyo 152-8552, Japan

³ Graduate School of Economics, University of Tokyo, Bunkyo-ku, Hongo 7-3-1, Tokyo 103-0033, Japan

1 Introduction

In economic and financial time series, we sometimes observe sudden and large price jumps. Although they are relatively rare events, when they occur they often have significant impact on not only a single financial market but also several different markets and wider macro-economies. Several recent notable events in European and Asian countries with large jumps include the global 2008 crisis.

The standard econometric method for investigating economic and financial time series has been the statistical analysis of discrete time series. In statistical time series analysis, we often assume that the observed time series data are equally spaced realizations of stochastic processes and that the state space is \mathbf{R}^d ($d \geq 1$) in multivariate cases. Many statistical procedures in discrete time series analysis have been developed and applied to economic and financial time series in recent decades. When we do not observe events frequently, however, traditional discrete time series modeling with continuous state space may present important limitations. For instance, it may be difficult to distinguish major contagious events from small contagious events among different financial markets across international borders.

In this paper, we propose an alternative way of investigating economic and financial events with time series data in macro-economies, i.e., the statistical analysis of marked point processes to identify and explore multivariate time series events. Although this is not a standard approach in time series econometrics, there have been applications of this methodology in statistical seismology [see Ogata (1978, 2015) and the related literature, for instance]. We will show that this approach is a useful alternative way of investigating multivariate economic and financial markets to shed new light on issues that have hitherto sometimes neglected. In particular, we propose using simultaneous Hawkes-type multivariate point process models and their applications in this study. We argue that using the simultaneous multivariate Hawkes-type point process (SHPP) models, which are a new class of multivariate point processes, it is possible to investigate the causal effects of sudden and large jumps with their magnitude. We develop a new way to measure the Granger non-causality (GNC) and instantaneous Granger non-causality (IGNC) through stochastic intensity modeling of point processes.

In econometric time series analysis, the concept of Granger causality (after Granger 1969) has become an important, well-established tool for investigating relationships among multivariate time series variables (See Hosoya et al. 2017 for the recent developments). In the econometric literature, Florens and Fougere (1996) investigated several Granger causality concepts in the framework of continuous time stochastic processes, but their formulation of the problem was incomplete because they excluded the possibility of co-jumps therein, which means the simultaneous jumps that can be observed in multivariate times series data are excluded from the outset. The problem of co-jumps is important because we often analyze economic time series data in discrete time (monthly, weekly, daily, hourly, and/or minute), but continuous stochastic processes are also salient recently in financial econometrics. We need to coherently unify discrete time

series analyses and continuous stochastic processes. In this paper, we investigate the possible use of co-jumps in a systematic way and develop new GNC and IGNC tests, which may provide important insight for advancing the development and application of econometric time series modeling.

Previously, Kunitomo et al. (2017) have used the traditional multivariate Hawkes-type point process (THPP) models without co-jumps and the SHPP models are the extension of their models (that is, the THPP models are special cases of SHPP models). There are important cases when we need the SHPP models as we will illustrate in Sect. 6. In statistical seismology, researcher often uses the earthquake data of large magnitudes (greater than 3, say) and the time scale of measurement is short due to physical laws. In financial markets, however, the impacts of shocks occur in actual trades of financial commodities and digestion of bad or good news among market participants often needs time (1, 2 h, a day and days). Therefore, it may not be fruitful to use the method developed for seismology to financial problem mechanically and we need to consider the issues of time scales, discretization of statistical models and their measurements carefully.

Several recent studies in financial econometrics have utilized point processes and conditional intensity modeling (Ait-Sahalia and Jacod 2014; Ait-Sahalia et al. 2015; Embrechts et al. 2011; Grothe et al. 2014; and others). In a survey of these and other works, Bacry et al. (2015) noted that the focus therein is mostly on studies of financial micro-market structures. The approach developed in this paper is related to these works, but the main purpose is quite different because we develop a new point process approach to assess the relationships among different (international financial) markets. In this respect, there have also been studies on international linkage in financial markets (e.g., Hamao et al. 1990), but our proposed methodology is notably different because those studies utilized standard discrete time series modeling. In terms of empirical examples, we will investigate the interactions among Tokyo–New York, Tokyo–London, and Tokyo–Hong Kong financial markets and apply the GNC tests developed herein to those contexts. This yields several important findings among major financial markets.

The remainder of the paper is organized as follows. In Sect. 2 we present a general formulation of simultaneous multivariate Hawkes-type point process (SHPP) models. In Sect. 3, we describe the estimation method and develop non-causality tests in the sense of Granger (1969). In Sect. 4, we explore simulation results and the empirical applications are offered in Sect. 5. Concluding remarks are presented in Sect. 6 and the mathematical details are provided in the Appendix.

2 Simultaneous Hawkes-type point processes

We divide the observation period $[0, T]$ into discrete periods $I_i^n = (t_{i-1}^n, t_i^n]$ ($i = 1, \dots, n$) and set the initial time as $t_0^n = 0$. We may interpret I_i^n as the i -th day, but it is possible to use higher resolution of observation periods (e.g. hourly or per minute) and we allow the irregularly-spaced time series modeling in principle.

Let the observable price processes of Itô-semimartingale with the state space of \mathbf{R}^d (see Ikeda and Watanabe 1989) be $P_j(t)$ ($j = 1, \dots, d; t_{i-1}^n < t \leq t_i^n, i = 1, \dots, n$), and in $s \in I_i^n$ we denote the (negative) log-return of prices $X_j^n(s)$ ($s \in I_i^n$) as

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

Let the first stopping time when $X_j^n(s)$ exceeds the threshold $u_j (> 0)$ in $s \in I_i^n$ be

$\tau^n(i, j, 1)$. When $\tau^n(i, j, 1) < t_i^n$, we re-define the return process $X_j^n(s) = -\log[P_j(s)/P_j(\tau^n(i, j, 1))]$ ($s \in I_i^n, s \geq \tau^n(i, j, 1)$) and let the first stopping

time when $X_j^n(s)$ exceeds the threshold u_j crossing from below in $s \in (\tau^n(i, j, 1), t_i^n]$

be $\tau^n(i, j, 2)$. In this way, we define the sequence of $\tau^n(i, j, k)$ ($k \geq 1$) successively.

Let also a sequence of numbers of jumps in an interval be $J_j(i) = \#\{k : \tau^n(i, j, k) \in I_i^n, k \geq 1\}$ and then define

$$N_j^{n*}(t, u_j) = \sum_{1 \leq l \leq i, J_j(l) > 0} \frac{1}{J_j(l)} N_j(I_l^n, t, u_j) \quad (t \in I_i^n), \tag{2}$$

where $N_j(I_l^n, t, u_j)$ is the number of counts that the resulting return process $X_j^n(s)$ ($s \leq t$) exceeds u_j as the threshold crossing from below in $s \in I_l^n$ and $s \leq t \in I_i^n$.

The stochastic process N_j^{n*} varies at most one in every discrete observation period.

This formulation of normalized counting in each intervals I_i^n ($i = 1, \dots, n$) allows us to measure the market impacts of financial price jumps in discrete intervals while we can use the standard statistical intensity modeling. We note that $J_j(i)$ are countable in $[0, T]$ (T is finite, n is sufficiently large and $u_j > 0$ are fixed constants) and the number of instantaneous jumps (i.e. $|P_j(s) - P_j(s-)| > u_j > 0$) is finite for the price processes of Itô-semimartingales. In financial risk management and regulation, the sudden and/or large downward movements of financial commodity prices in short time intervals are the most important subject because of their negative consequences to financial markets and economies.¹

In the following analysis, we consider the situation that there is a common threshold value u for u_j ($j = 1, \dots, d$) and the observations of the counting processes are available at day-start, day-minimum and day-end in each interval. In principle, however, we can allow the irregularly-spaced time series modeling, but we need an explicit discretization of observations. For these intervals of observation, we consider the point processes, $N_j^{n*}(t, u)$ ($j = 1, \dots, d$), which are simple. They satisfy the standard condition for point processes that as $\Delta t \rightarrow 0$, we have the conditions

¹ As a referee had pointed out, the present formulation may be complicated because we allow multiple jumps in a fixed interval. When the length of discretization becomes small, as a limit we can ignore the complication involved (see Kurisu 2018).

$$\begin{aligned}
 P(N_j^{n*}(t + \Delta t, u) - N_j^{n*}(t, u) = 1 | \mathcal{F}_{t-}^n) &= \lambda_j^{n*}(t, u)\Delta t + o_p(\Delta t), \\
 P(N_i^{n*}(t + \Delta t, u) - N_i^{n*}(t, u) > 1 \text{ for any } i | \mathcal{F}_{t-}^n) &= o_p(\Delta t), \\
 P(N_i^{n*}(t + \Delta t, u) - N_i^{n*}(t, u) = 1 \text{ for } i = j, k; j \neq k | \mathcal{F}_{t-}^n) &= o_p(\Delta t),
 \end{aligned}$$

where \mathcal{F}_{t-}^n is the σ -field generated by the latest information before t . The (conditional) intensity functions are given by

$$\lambda_j^{n*}(t, u) = \lim_{\Delta t \rightarrow 0} \mathbf{E} \left[\frac{N_j^{n*}(t + \Delta t, u) - N_j^{n*}(t, u)}{\Delta t} \mid \mathcal{F}_{t-}^n \right], \tag{3}$$

where we use the notation \mathcal{F}_{t-}^n as the latest information before t because we discretize the counting process and we have discrete observations. For convenience, we denote \mathcal{F}_{t-}^n as \mathcal{F}_t in the following analysis whenever such notation allows.

For expository purposes, in the following analysis, we interpret the increments of $N_j^{n*}(s, u_k)$ as if jumps of the counting process occur at t_i^n , the end of each interval I_i^n and we have set the threshold $u_j = u$ ($j = 1, \dots, d$). When we consider the situation when the interval length goes to zero, i.e., $\Delta_n t = \max_{i=1, \dots, n} |t_i^n - t_{i-1}^n| \rightarrow 0$ as $n \rightarrow \infty$ for a fixed T , the counting process, which is a simple point process, $N_j^{n*}(s, u)$ weakly converges to $N_j^*(s, u)$. The resulting counting process can be interpreted as a limiting continuous-time stochastic process in high-frequency asymptotics, which is not a diffusion type but a pure jump process (see Ikeda and Watanabe 1989; Ait-Sahalia and Jacod 2014; Kurisu 2018, for example).

Next, for $u_j = u$ ($j = 1, \dots, d$) we define the point processes $N_{jk}^{n*}(s, u)$ by the number of stopping times that $X_j^n(s)$ exceeds u ($j = 1, \dots, d$) for a particular j , $X_k^n(s)$ exceed u ($k = 1, \dots, d; k \neq j$) for another k , and other $X_l^n(s)$ ($l \neq j, k$) do not exceed u crossing from below by time s in the interval I_i^n . By this construction, we can introduce the point processes $N_{jk}^{n*}(t, u)$ with co-jumps of N_j and N_k by

$$\begin{aligned}
 P(N_j^{n*}(t + \Delta t, u) - N_j^{n*}(t, u) = N_k^{n*}(t + \Delta t, u) - N_k^{n*}(t, u) = 1 | \mathcal{F}_t) \\
 = \lambda_{jk}^{n*}(t, u)\Delta t + o_p(\Delta t), \\
 P(N_i^{n*}(t + \Delta t, u) - N_i^{n*}(t, u) > 1 \text{ for any } i | \mathcal{F}_t) = o_p(\Delta t), \\
 P(N_i^{n*}(t + \Delta t, u) - N_i^{n*}(t, u) = 1 \text{ for } i = j, k, l; j \neq k \neq l | \mathcal{F}_t) = o_p(\Delta t),
 \end{aligned}$$

where $\lambda_{jk}^{n*}(t, u)$ are the conditional intensity functions of co-jumps.

Then, when we have co-jumps of two point processes, we can define the point processes

$$N_j^n(s, u) = N_j^{n*}(s, u) + \sum_{k \neq j} N_{jk}^{n*}(s, u) \quad (j, k = 1, \dots, d) \tag{4}$$

and the corresponding conditional intensity functions are given by

$$\lambda_j^n(t, u) = \lambda_j^{n*}(t, u) + \sum_{k \neq j} \lambda_{jk}^{n*}(t, u). \tag{5}$$

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

By extending this formulation to more complex co-jumps, in general we define

$$N_j^n(s, u) = \sum_{J_j \in \{1, \dots, d\}} N_{J_j}^{n*}(s, u) \quad (j = 1, \dots, p), \tag{6}$$

where the index set $J_j = \{j_1, \dots, j_l\} \in \{1, \dots, d\}$ is a subset of $(1, \dots, d)$. The index sets are defined as $J_i = \{i\}$ for $(i = 1, \dots, d)$, $J_i = \{1, 1 + (i - d)\}$ for $(i = d + 1, \dots, 2d - 1), \dots$, and $J_p = \{1, \dots, d\}$.

Then we sequentially define $N_i^n(s, u) = N_i^{n*}(s, u)$ ($i = 1, \dots, d$); $N_{d+1}^n(s, u) = N_{1,2}^{n*}(s, u), \dots$, and $N_p^n(s, u) = N_{1, \dots, d}^{n*}(s, u)$.

We use the self-exciting form of conditional intensity functions $\lambda_{j_m^*}(\cdot)$ for co-jumps as $\lambda_{jk}^{n*}(t, x | \mathcal{F}_{t-}^n)$ in the same way and the marginal conditional intensity function for the j -th components as

$$\lambda_j^n(t, u) = \sum_{J_j \in \{1, \dots, d\}} \lambda_{J_j}^{n*}(t, u). \tag{7}$$

There is a one-to-one transformation between $N_j^n(s, u)$ for $j = 1, \dots, p$ and $N_{j_1, \dots, j_k}^{n*}(s, u)$ for $1 \leq k \leq d$ and between $\lambda_j^n(t, u)$ and $\lambda_{j_1, \dots, j_k}^{n*}(t, u)$ for $p = 2^d - 1$.

The self-exciting Hawkes-type conditional intensity functions for the marked point processes are given by

$$\lambda_j^n(t, x | \mathcal{F}_{t-}^n) = \lambda_{j,0} + \sum_{i=1}^p \int_{-\infty}^t c_{ji}(x) g_i(t - s) N_{J_i}^{*n}(ds \times dx) \tag{8}$$

for $j = 1, \dots, p$, where $N_{J_i}^{*n}(ds \times dx)$ are the marked point processes, $\lambda_{j,0}$ are the initial intensities, $g_i(t - s) = e^{-\gamma_i(t-s)}$ are the damping functions, and $C(X) = (c_{ji}(x))$ are the impact functions.

Since we are interested in sudden and large jumps of the underlying price processes (in the sense of negative returns), it is important to use the probability functions of the return process in the tail areas. In this respect, many empirical studies on the stock markets found that stock returns exhibit the non-Gaussianity and thick

tails. Hence it may be appropriate to use the generalized Pareto distributions (GPD) as tail probability functions for $x > u > 0$ ($j = 1, \dots, d$) as

$$\begin{aligned}
 P(X_j^n(s) > x | X_j^n(s) > u, \mathcal{F}_s) &= \frac{\left[1 + \frac{\xi_j}{\sigma_j} x\right]^{-1/\xi_j}}{\left[1 + \frac{\xi_j}{\sigma_j} u\right]^{-1/\xi_j}} \\
 &= \left[1 + \frac{\xi_j}{\sigma_j^*} (x - u)\right]^{-1/\xi_j},
 \end{aligned}
 \tag{9}$$

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

(See Resnick 2007 for the details of GPD in statistical extreme value theory (SEVT)).

Herein, we assume that given the return at s $X_j^n(s)$ ($j = 1, \dots, d$), the conditional density functions are given by

$$f_j(x, s) = \frac{1}{\sigma_j^*} \left[1 + \frac{\xi_j}{\sigma_j^*} (x - u)\right]^{-1/\xi_j - 1} \quad (x > u, \xi_j > 0).
 \tag{10}$$

In terms of impact functions and the intensity function of co-jumps, there can be many possible specifications. In our empirical study, we mainly investigate the form

$$c_{ij}(X) = \left(a_{ij} x_j^c\right) \quad (0 \leq c \leq 1; i = 1, \dots, p; j = 1, \dots, d)$$

and

$$c_{ij}(X) = \left(a_{ij} \max_{k \in J_i} x_k^c\right) \quad (0 \leq c \leq 1; i = 1, \dots, p; j = d + 1, \dots, p),$$

where a_{ij} and c are some constants.

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

Let $p \times 1$ vector point process $\mathbf{N}^n(t, u)$ be partitioned as $(d + (p - d)) \times 1$ processes as

$$\mathbf{N}^n(t, \mathbf{u}) = \begin{bmatrix} \mathbf{N}_1^n(t, u) \\ \mathbf{N}_2^n(t, u) \end{bmatrix} = \begin{bmatrix} N_1^n(t, u) \\ \vdots \\ N_d^n(t, u) \\ N_{1,2}^n(t, u) \\ \vdots \\ N_{1,2,\dots,d}^n(t, u) \end{bmatrix},
 \tag{11}$$

where $\mathbf{N}_1^n(t, u)$ is the $d \times 1$ vector of marginal point processes with $p = 2^d - 1$ and $\mathbf{N}_2^n(t, u)$ is the $(p - d)$ vector of co-jump point processes. The corresponding conditional intensity functions as

$$\lambda^n(t, \mathbf{u}) = \begin{bmatrix} \lambda_1^n(t, u) \\ \lambda_2^n(t, u) \\ \vdots \\ \lambda_d^n(t, u) \\ \lambda_{1,2}^n(t, u) \\ \vdots \\ \lambda_{1,2,\dots,d}^n(t, u) \end{bmatrix}, \tag{12}$$

and $p \times p$ matrices

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

We use notation such that $\lambda_i^n(t, u)$ is the vector process of conditional intensities of marginal jumps, $\text{diag}(\cdot)$ for diagonal matrices and we often omit n for $\lambda_j^n(s)$ ($i = 1, \dots, p$) and $N_{j_i}^n$ whenever their meanings are clear in the following

analysis.

Next, we rewrite (6) and (7) as

$$\mathbf{N}_1^n(t, u) = \mathbf{D}_1 \mathbf{N}^n(t, u), \tag{13}$$

and

$$\mathbf{N}_2^n(t, u) = \mathbf{D}_2 \mathbf{N}^n(t, u), \tag{14}$$

where \mathbf{D}_1 is a $d \times p$ matrix as

$$\mathbf{D}_1 = \begin{bmatrix} 1 & 0 & \dots & 0 & 1 & 1 & \dots & 0 & \dots & 1 \\ 0 & 1 & \dots & 0 & 1 & 0 & \dots & 0 & \dots & 1 \\ \vdots & & & 1 & 0 & 0 & & \vdots & \dots & 1 \\ 0 & \dots & \dots & 1 & 0 & \dots & \dots & \dots & & 1 & 1 \end{bmatrix}$$

and \mathbf{D}_2 is a $(p - d) \times p$ matrix as $\mathbf{D}_2 = [\mathbf{0}, \mathbf{I}_{p-d}]$ ($p \geq d$).

We call the above Hawkes-type conditional intensity models as the simultaneous multivariate Hawkes-type point process (SHPP) models because the resulting marked point processes are not necessarily simple.² The classical Hawkes-type point processes have been useful in applications because they are simple point processes. However, they exclude the possibility of simultaneous jumps or co-jumps and they are not fit for our purposes here. The foregoing constructions of marked point processes can be regarded as an extension of Solo (2007).

² Definition and mathematical details of “simple-point process” and other basic point processes are given in Daley and Vere-Jones (Vol-I, 2003), for instance.

3 Stationarity and Bartlett spectrum decomposition

3.1 Stationarity of Hawkes-type processes

In our applications, we use stationary self-exciting Hawkes-type (marked) point processes. We take the expectation of the intensity function of (11) and (12) in $(-\infty, t]$ as

$$\mathbf{E}[\lambda^n(t, \mathbf{u})] = \lambda_0 + \mathbf{E} \left[\int_{-\infty}^t C(\mathbf{X}(s-))\mathbf{G}(t-s)d\mathbf{N}^n(s, \mathbf{u}) \right], \tag{15}$$

where $\lambda_0 = (\lambda_{j,0})$.

Let a $p \times 1$ vector of functions $\mathbf{v}(t) = (v_j(t))$ be $\mathbf{E}[\lambda^n(t, \mathbf{u})]$. For simplicity, we take $\mathbf{G}(t-s) = [\text{diag}(g_j(t-s))]$ with $g_j(t-s) = e^{-\gamma_j(t-s)}$ and $\mathbf{G}(0) = \mathbf{I}_p$, $\Gamma = [\text{diag}(\gamma_j)]$ ($\gamma_j > 0, j = 1, \dots, p$).

The stationarity implies $\mathbf{C} = \mathbf{E}[C(\mathbf{X}(s-))]$ and we can use the identity relation $\mathbf{v}(t) - \lambda_0 = \int_{-\infty}^t \mathbf{C}\mathbf{G}(t-s)\mathbf{v}(s)ds$. Then by a direct calculation, we have a set of differential equations

$$\frac{d\mathbf{v}(t)}{dt} = [\mathbf{C} - \mathbf{C}\Gamma\mathbf{C}^{-1}]\mathbf{v}(t) + \mathbf{C}\Gamma\mathbf{C}^{-1}\lambda_0, \tag{16}$$

provided the initial condition $\mathbf{v}(0)$ and the non-degeneracy condition $|\mathbf{C}| \neq 0$.

We need a condition for the convergence of $\mathbf{v}(t)$ as $t \rightarrow \infty$. Therefore, the condition for the existence of stationary point processes is that the spectral radius

$$\max_{1 \leq i \leq p} |\mu_i(\mathbf{F})| < 1, \tag{17}$$

where $\mu_i(\mathbf{F})$ are the characteristic roots of $\mathbf{F} = \mathbf{C}\Gamma^{-1}$. (See Theorem 2 of Kunitomo et al. 2017 as a special case.) When $d = p = 1$ (one-dimensional Hawkes process), $\mathbf{C} = \alpha$ and $\Gamma = \gamma$ (> 0), then $\mathbf{F} = \alpha/\gamma$.

3.2 Applying the Bartlett spectrum

Hawkes (1971) introduced the spectral density for the stationary vector point process $\mathbf{N}(t) = (N_i(t))$, which was originally developed by Bartlett (1963) without co-jumps ($p = d$); it is defined for the conditional intensity vector in the form of

$$\lambda(t) = \lambda_0 + \int_{-\infty}^t \gamma(t-u)d\mathbf{N}(u), \tag{18}$$

where $\gamma(u) = (\gamma_{ij}(u))$ is a $d \times d$ matrix and $\gamma(u) = (0)$ (zero-matrix) for $u < 0$. Let the Fourier transform of $\gamma(\tau)$ be

$$\Gamma^*(\omega) = \int_{-\infty}^{\infty} e^{-i\omega\tau} \gamma(\tau) d\tau, \tag{19}$$

where $i^2 = -1$.

Then, when $p = d$ (there are no co-jumps), the Bartlett spectral matrix for frequency $\omega \in \mathbf{R}$ is given by

$$\mathbf{g}(\omega) = \frac{1}{2\pi} [\mathbf{I}_d - \Gamma^*(\omega)]^{-1} \Sigma [\mathbf{I}_d - \Gamma^{*'}(-\omega)]^{-1}, \tag{20}$$

where Γ^* in (20) is a $d \times d$ matrix for the d -dimensional vector point process and $\Gamma^{*'}$ is the transposed matrix of Γ^* .

To permit co-jumps, Bartlett spectral matrix for the d -dimensional marginal point process vector can be defined by

$$\mathbf{g}(\omega) = \frac{1}{2\pi} [\mathbf{I}_d, \mathbf{O}] [\mathbf{D} - \mathbf{D}\Gamma^*(\omega)]^{-1} \Sigma [\mathbf{D}' - \mathbf{D}'\Gamma^{*'}(-\omega)]^{-1} \begin{bmatrix} \mathbf{I}_d \\ \mathbf{O} \end{bmatrix}, \tag{21}$$

where $\mathbf{g}(\omega) = (g_{ij}(\omega))$ is the $d \times d$ spectral density matrix, $\Gamma^*(\omega)$ is a $p \times p$ Fourier transform as (19), $\mathbf{D} = (\mathbf{D}'_1, \mathbf{D}'_2)'$ is a $p \times p$ choice matrix, and $\Sigma = (\sigma_{ii})$ is the diagonal matrix with diagonal elements of the variances σ_{ii} ($i = 1, \dots, p$).

Then we define the relative power contribution (RPC) of the marginal spectral density function $g_{ii}(\omega)$ ($i = 1, \dots, d$) where the frequency ω can be defined using the joint spectral density matrix $\mathbf{g}(\omega)$. The (i,i)-component of $\mathbf{g}(\omega)$ can be represented as

$$g_{ii}(\omega) = \sum_{k=1}^p |a_{ik}(\omega)|^2 \sigma_{kk} \tag{22}$$

and

$$\mathbf{RPC}_{k \rightarrow i}(\omega) = \frac{|a_{ik}(\omega)|^2 \sigma_{kk}}{g_{kk}(\omega)} \quad (i = 1, \dots, p; k = 1, \dots, d), \tag{23}$$

where $a_{ij}(\omega)$ ($i = 1, \dots, d; j = 1, \dots, p$) are the functions of complex variables. In addition, the instantaneous RPC ($\mathbf{IRPC}_{j \rightarrow i}$) can be defined by

$$\mathbf{IRPC}_{j \rightarrow i}(\omega) = \frac{|a_{ij}(\omega)|^2 \sigma_{jj}}{g_{ii}(\omega)} \quad (j = d + 1, \dots, p). \tag{24}$$

In this way, we can measure the RPCs for any frequency ω , which corresponds to the Granger-causality measures in the frequency domain. One important aspect of the above formulation is that we have a natural definition of instantaneous Granger

causality in the frequency domain, that is different from discrete time series modeling.

3.3 Conditional probability prediction

An important application of conditional intensity modeling involves assessing the conditional probability of rare events in the future from past observations. Let $\tau(j)$ ($j = 1, \dots, d$) be the first arrival time of an event in the j -th variable. Then we can write the probability of the random variable $\tau(j)$ as

$$Pr(\tau(j) \geq T' | \mathcal{F}_T^N) = \exp\left(-\int_T^{T'} \lambda_j^n(t, u | \mathcal{F}_T^N) dt\right), \tag{25}$$

where \mathcal{F}_T^N is the σ -field of information available at time $T < T'$ and $\lambda_j^n(t, u | \mathcal{F}_T^N)$ is the conditional intensity of the j -th variable.

Kunitomo et al. (2017) conducted some experiments suggesting that useful information on the conditional probability of future events can be extracted from past observations. For instance, they provided an important example vis-à-vis the conditional probability prediction of the Lehman Shock occurred in 2008 as a global crisis given past information available before that event. This illustrates the potential value of our approach.

4 Estimation and non-causality tests

4.1 Likelihood function

When the point process is simple and there is no co-jump, the log-likelihood function of the (d -dimensional) multivariate point process is known (see Daley 2003; Kunitomo et al. 2017) and it is given by

$$\sum_{i=1}^d \left\{ -\int_0^T \lambda_i^n(s) ds + \int_0^T \log(\lambda_i^n(s)) dN_i^n(s) \right\}. \tag{26}$$

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

$$L_T = L_{1T} + L_{T2}, \tag{27}$$

where

$$L_{1T} = \sum_{i=1}^d \left\{ - \int_0^T \lambda_i^n(s) ds + \int_0^T \log(\lambda_i^n(s)) dN_i^n(s) \right\},$$

$$L_{2T} = \sum_{i=1}^d \left\{ \int_0^T \log f_i(x_i^n(s-)) dN_i^n(s) \right\}$$

and the density function for the tail probability is given by

$$f_i(x) = \frac{1}{\sigma_i^*} \left(1 + \xi_i \frac{x_i - u_i}{\sigma_i^*} \right)^{-\frac{1}{\xi_i} - 1} \quad (i = 1, \dots, d). \tag{28}$$

Then we can apply the maximum likelihood (ML) method to L_{1T} and L_{2T} separately. In this formulation we use the GPD for the marginal distribution of return process.

When co-jumps are permitted, the log-likelihood function of the (d -dimensional) marginal point process is not as per the above form; instead, it should be given by

$$L_T^* = L_{1T}^* + L_{2T}^*, \tag{29}$$

where

$$L_{1T}^* = \sum_{i=1}^d \left\{ - \int_0^T \lambda_i^n(s) ds + \int_0^T \log(\lambda_i^n(s)) dN_i^n(s) \right\}$$

$$+ \sum_{i \neq j=1}^d \left\{ - \int_0^T \lambda_{ij}^n(s) ds + \int_0^T \log(\lambda_{ij}^n(s)) dN_{ij}^n(s) \right\}$$

$$+ \dots + \left\{ - \int_0^T \lambda_{i\dots d}^n(s) ds + \int_0^T \log(\lambda_{i\dots d}^n(s)) dN_{i\dots d}^n(s) \right\}.$$

and $L_{2T}^* = L_{2T}$.

In our applications, we are principally concerned with the case in which $d = 2$; thus, there is only one extra term in the likelihood function because $p = 2^d - 1$.

We assume the stationarity condition (17) 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 without and with co-jumps. Further, we take $\lambda(\mathbf{u})$ as the stationary conditional intensity and some $q \times p$ predictable processes $\xi(t)$ having second order moments. (Here $q \geq 1$ and we utilize the notation $\mathbf{g}_T(t)$ in Appendix, for instance.)

Then, because of the resulting martingale property given the information available at each time, it is straightforward to confirm the asymptotic properties as we have

$$\frac{1}{T} \int_0^T \xi(t) [\mathbf{N}(t, u) - \lambda(t, \mathbf{u})] dt \longrightarrow 0 \quad (a.s.) \tag{30}$$

and

$$\frac{1}{T} \int_0^T \xi(t)[\lambda(t, u) - \lambda(\mathbf{u})]dt \xrightarrow{p} 0 \tag{31}$$

as $T \rightarrow \infty$.

For the one-dimensional point processes with the stationary intensity function with $p = q = 1$, Ogata (1978) gave a set of sufficient conditions for the consistency and asymptotic normality of the 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 case. For the sake of completeness, we provide details of our approach based on a new MCLT in the Appendix, which may be more general than the standard literature. In the next subsection, we develop new non-causality tests in the sense of Granger, which are explored in the context of our empirical applications.

4.2 Non-causality tests

We develop and use novel GNC tests based on the likelihood ratio principle for the Hawkes-type point processes. In particular, our results in this subsection, whose derivations are given in the Appendix, include not only the multivariate extension of existing results, but also cases in which the resultant limiting Fisher information matrix can be random variables. We first state our results for the case of no co-jumps under a set of regularity conditions, which will be extended to the more general case. The proof is lengthy, but often along the standard line of asymptotic arguments and we only give its outline in Appendix.

Theorem 1 *Let the log-likelihood function of the Hawkes-type point processes with true parameters be $L_T(\theta_0)$ in (26) and (27), the log-likelihood function with the ML estimator $\hat{\theta}_{ML}$ be $L_T(\hat{\theta}_{ML})$ under $\theta \in \Theta$ and the log-likelihood function with the restricted maximum likelihood estimator $\hat{\theta}_{RML}$ be $L_T(\hat{\theta}_{RML})$ under $\theta \in \Theta_1$ ($\Theta_1 \subset \Theta$). We assume the sufficient condition for stationarity, the existence of the second-order moment condition of $\mathbf{C}(\mathbf{X})$, and we assume that the parameter spaces $\theta \in \Theta$ in \mathbf{R}^r and $\theta \in \Theta_1$ in \mathbf{R}^{r_1} ($0 \leq r_1 < r$) are compact sets. Under a set of regularity conditions (see Theorem A-3 in the Appendix), as $T \rightarrow \infty$,*

$$2\{L_T(\hat{\theta}_{ML}) - L_T(\hat{\theta}_{RML})\} \xrightarrow{d} \chi^2(r - r_1), \tag{32}$$

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

The details of a set of regularity conditions are discussed in the Appendix. When co-jumps are permitted in the Hawkes-type processes, we cannot apply Theorem 1, but it is important to obtain the corresponding results in such cases for econometric applications. When we use discrete versions of point processes, which would be often the case in econometric applications, we need to consider the existence of co-jumps. We

then develop non-causality tests based on the likelihood ratio principle. In this respect, note that in our setting discussed in Sect. 2, although we permit co-jumps, it is possible to apply the martingale central limit (MCLT) theorem for point processes. Our results, in consideration of co-jumps, are an extension of Theorem 1. The proof is lengthy, but often along the standard line of asymptotic arguments and we only give its outline in Appendix.

Theorem 2 *Let the log-likelihood function of the Hawkes-type point processes with true parameters be $L_T^*(\theta_0)$ in (29), the log-likelihood function with the ML estimator $\hat{\theta}_{ML}$ be $L_T^*(\hat{\theta}_{ML})$ under $\theta \in \Theta$ and the log-likelihood function with the restricted ML estimator $\hat{\theta}_{RML}$ be $L_T^*(\hat{\theta}_{RML})$ under $\theta \in \Theta_1$ ($\Theta_1 \subset \Theta$). We assume sufficient conditions for stationarity, and the existence of the second-order moment condition of $\mathbf{C}(\mathbf{X})$, and we assume that the parameter spaces $\Theta \in \theta$ in \mathbf{R}^r and $\Theta_1 \in \theta$ 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 $T \rightarrow \infty$,*

$$2\{L_T^*(\hat{\theta}_{ML}) - L_T^*(\hat{\theta}_{RML})\} \xrightarrow{d} \chi^2(r - r_1), \tag{33}$$

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

5 Simulations

To examine the relevance of the estimation and testing procedure proposed in this paper, a set of simulations are executed. The model used in these simulations is a simultaneous Hawkes-type model with two dimension and the intensity functions are given by

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

Table 1 Simulation results

	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

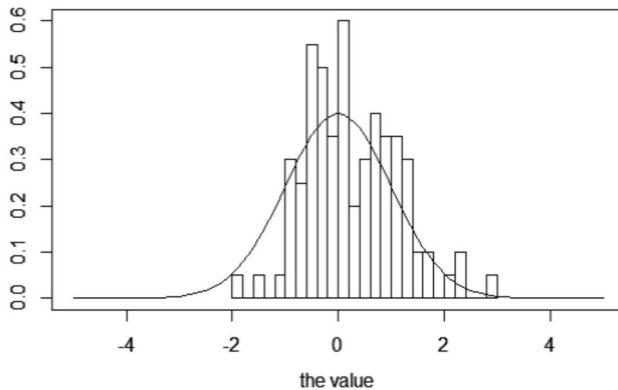


Fig. 1 α_{12}^*

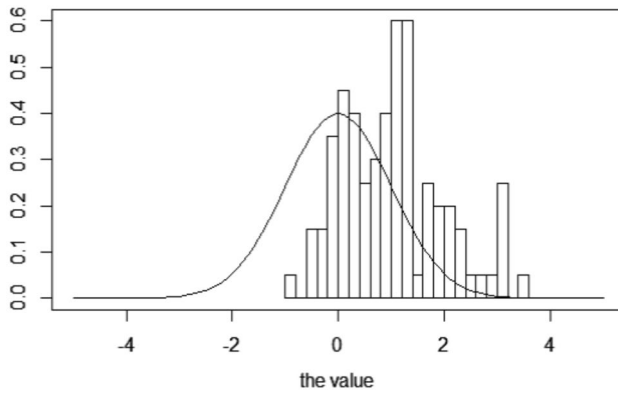
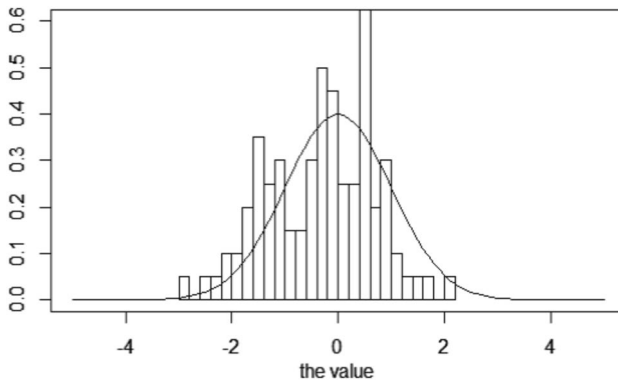
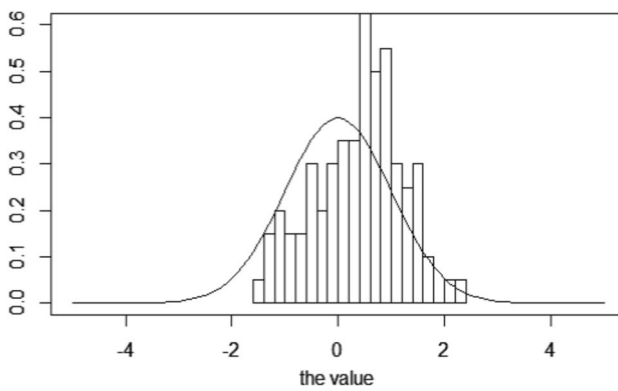
where $\lambda_{1,0}^n > 0, \lambda_{2,0}^n > 0, \lambda_{12,0}^n > 0$ and $\gamma > 0$.

We first generate stock price returns using the GPD as marginal and the two-dimensional Gaussian copura. Then we employ ML method to obtain estimates of the underlying parameters. We provide a set of visualization (Figs. 1, 2, 3 and 4) to illustrate the key results on the finite sample distributions of the ML estimator. All histograms are standardized for the comparison of the standard normal distributions as

$$\mathbf{I}_n^{1/2}(\hat{\theta} - \theta), \tag{34}$$

where $\theta = (\theta_i)$ is a vector of parameters and $\hat{\theta}$ is the ML estimator.

In our numerical evaluations, the values of estimate sometimes hit the boundaries of the non-negativity of intensity functions with finite samples, resulting in instabilities. To mitigate against this, we thus set non-negativity restrictions on parameters in our simulations. The ensuring results are reasonable, but sometimes we observe that the ML estimators of coefficients exhibit biases, although they are not very large (Fig. 2 is a typical example of this). The sample size in

Fig. 2 α_{21}^* Fig. 3 α_{23}^* Fig. 4 γ^*

our experiments was around 1000 because it may be similar to the data size in the empirical examples and it seems that we need large number of data for reducing these biases. We summarize the configuration of our numerical experiments: the number of replication of simulations as 100, and for GPD(σ_j, ξ_j) we set $(\sigma_1, \xi_1) = (0.007, 0.22)$, and $(\sigma_2, \xi_2) = (0.008, 0.15)$. These numerical values give reasonable results, and they are based on the preliminary estimates from our empirical studies.

Among many simulations we illustrate key results in Table 1 and Figures. Note that because we have taken $\alpha_{12}^* = 0$, we have a sampling distribution around zero, and the resulting estimate is not significant as illustrated in Fig. 1. Other estimates of α_{ij} , which are around their true values, take reasonable values on average in the sense that they are not significantly different from the true values, and the sampling distributions are illustrated in Figs. 2, 3 and 4. We observe some positive biases on the estimates of α_{ij} and negative biases on the estimates of initial intensities, which may be due to the results of the non-negative constraints of the parameter restrictions and the number of sample size we employed. We have imposed the non-negativities of the intensities of variables directly in the ML computation.

In the ML estimation, there can be some effects of initial conditions and we have investigated this problem in the SHPP models, where such sensitivity is also apparent, albeit minor in overall simulations.

We also use the χ^2 -distributions as the limiting distributions of the likelihood ratio statistics for hypothesis testing in our empirical study. We confirm that the χ^2 -approximations with finite samples are basically appropriate.

6 Empirical applications

In this section, we report the empirical results on two empirical examples using the SHPP and THPP models. The first concerns the three major stock markets, namely, Tokyo, New York, and London. Since time differences exist when each market is open and closed, it is reasonable to assume that there are no co-jumps. In terms of the second example, we focus on analyzing the simultaneous interaction among Tokyo and Hong Kong financial markets. In this latter case, since the time zone differences are small (just a 1 h difference) compared to the first empirical example, it may be natural to use SHPP, which is the extended Hawkes-type point process model with co-jumps. Because of the limitations of data available to us, we have ignored the possibilities of crossing the threshold from below except the first one in a day, or between the day-start to day-minimum.

In the first example, daily data of day-start to day-minimum data are employed covering Nikkei225, S&P500 and FTSE100 during January 2, 1990–August 26, 2015. We choose $u = 2\%$ based on the earlier study of Kunitomo et al. (2017), which used the formulation of discrete process of returns and analyzed daily data of day-start to day-end for this case. Their empirical results were quite similar to those in the following analysis, but the numerical values are different. All computations were carried out by the original programs written in *R*. Example 2, which concerns the Tokyo and Hong-Kong markets, is entirely new and is the principal driver of the SHPP models developed in

Table 2 Tail distributions

	Log likelihood	σ_1^*	ξ_1
Tokyo	- 1919.31	0.00757	0.22778
SD		0.00051	0.05552
	Log likelihood	σ_2^*	ξ_2
New York	- 1346.55	0.00874	0.13978
SD		0.00064	0.05648
	Log likelihood	σ_3^*	ξ_3
London	- 1479.35	0.0079	0.18246
SD		0.00058	0.05808

this study. We will report the results for Example 2 using this type of data. Nonetheless, we have done robustness check of our results on the estimation of conditional intensity modeling and non-causality tests. We omit reporting the details of some results using the day-start to day-end data because they are basically quite similar.

6.1 Example 1 (Tokyo–NY–London)

We first maximize the likelihood L_{2T} to estimate the marginal distributions of financial market returns. As shown in Table 2, we confirmed that the marginal distributions of market returns (i.e., log-returns) have thicker tails than the normal distribution. It is because the estimates of ξ_i ($i = 1, 2, 3$) are positive and it is appropriate to use the GDP in our estimation procedure. The result means that the Frechet-type tail distribution is appropriate since the domain of attraction of Gaussian distribution is Gumbel (see Embrechts et al. 1998 Chapter 3 for instance). The standard deviations (SD) (or the standard errors of the estimates) in Tables are estimated by the numerical evaluation of the Fisher information matrix.

For the estimated models with two dimensions ($d = p = 2$), we take the impact functions $c(x)$ as Model 1($c(x) = 1$, Model 2($c(x) = x$, and Model 3($c(x) = x^c$ ($0 < c < 1$). The estimated values of the log-likelihood and Akaike information criterion (AIC) are those with the marginal distribution L_{1T} . The full likelihood can be calculated using L_{1T} and L_{2T} . The standard deviations (SD) or standard errors of the estimated coefficients are also evaluated numerically using the inverse of the estimated Fisher information matrix.

Model 1

We estimated the intensity function as

$$\lambda_1^n(t) = \lambda_{10}^n + \int_0^t \alpha_{11} e^{-\gamma_{11}(t-s)} dN_1^n(s) + \int_0^t \alpha_{12} e^{-\gamma_{12}(t-s)} dN_2^n(s),$$

$$\lambda_2^n(t) = \lambda_{20}^n + \int_0^t \alpha_{21} e^{-\gamma_{21}(t-s)} dN_1^n(s) + \int_0^t \alpha_{22} e^{-\gamma_{22}(t-s)} dN_2^n(s).$$

Table 3 Tokyo–New York

	Log likelihood	AIC	α_{11}	α_{12}
Tokyo–NY	– 3570.69	7155.37	0.0199	0.0073
SD			0.0024	0.0019
α_{21}	α_{22}	γ	λ_{10}	λ_{20}
0.0000	0.0267	0.0326	0.0094	0.0052
0.0009	0.0030	0.0034	0.0017	0.0011

Table 4 Tokyo–London

	Log likelihood	AIC	α_{11}	α_{12}
Tokyo–London	– 3670.86	7355.73	0.0209	0.0052
SD			0.0023	0.0016
α_{21}	α_{22}	γ	λ_{10}	λ_{20}
0.0000	0.0279	0.0333	0.0103	0.0051
0.0010	0.0030	0.0033	0.0017	0.0011

Table 5 Tokyo–New York

	Log likelihood	AIC	α_{11}	α_{12}
Tokyo–NY	– 3562.02	7138.03	0.6508	0.2193
SD			0.0784	0.0586
α_{21}	α_{22}	γ	λ_{10}	λ_{20}
0.0000	0.8582	0.0324	0.0105	0.0058
0.0287	0.0985	0.0035	0.0017	0.0011

Since the ML estimates can be numerically sometimes unstable without any restrictions on the parameter space, we have set restrictions that the discounted parameters γ_{ij} ($i, j = 1, 2$) have the same value γ in the following estimation. The estimation results for Case 1 are presented in Tables 3, 4.

In Table 3, N_1 of Model 1 corresponds to Tokyo and N_2 corresponds to New York in Tokyo–New York markets. In terms of Tokyo–London, in Table 4 N_1 of Model 1 corresponds to Tokyo while N_2 corresponds to London.

The most important finding here (and in Tables 5, 6 below), is that the coefficient α_{12} is statistically significant while the coefficient α_{21} is not statistically significant. This represents a kind of non-causality test, but we will discuss this more formally below. We found reasonable values for other parameters in their magnitudes and significance, and they are significant for both Tokyo–New York and Tokyo–London markets.

Table 6 Tokyo–London

	Log likelihood	AIC	α_{11}	α_{12}
Tokyo–London	– 3660.22	7334.43	0.6866	0.1423
SD			0.0779	0.0506
α_{21}	α_{22}	γ	λ_{10}	λ_{20}
0.0000	0.9089	0.0329	0.0114	0.0056
0.0310	0.0990	0.0033	0.0017	0.0011

Table 7 Tokyo–New York

	Log likelihood	AIC	α_{11}	α_{12}	α_{21}	α_{22}
Tokyo–NY	– 3562.23	7140.47	0.1702	0.0575	0.0000	0.2234
SD			0.0625	0.0242	0.0084	0.0791
γ	λ_{10}	λ_{20}	$c = c_{11} = c_{12} = c_{21} = c_{22}$			
0.0321	0.0100	0.0055	0.6146			
0.0035	0.0017	0.0011	0.1049			

Model 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 the estimation results are presented in Tables 5, 6.

In the present case, we have similar values for the estimated coefficients as Case 1 except α_{21} . The significance of coefficient is more pronounced here compared to Case 1, which corresponds to the likelihood values and their AIC.

Model 3

We estimated the intensity function as 0

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

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

Table 8 Tokyo–London

	Log likelihood	AIC	α_{11}	α_{12}	α_{21}	α_{22}
Tokyo–London	– 3659.49	7335.88	0.4244	0.0898	0.0000	0.5614
SD			0.6608	0.0994	0.0397	0.6023
γ	λ_{10}		λ_{20}		$c = c_{1,1} = c_{1,2} = c_{2,1} = c_{2,2}$	
0.0329	0.0111		0.0055		0.8593	
0.0033	0.0018		0.0014		0.3140	

Although we used the ML estimation in this case, the estimates of ML are often unstable numerically. In particular, we often found numerical difficulty to calculate the standard errors of estimates (Tables 7, 8). It was probably because the optimization computations with R often were unstable without any restrictions on the parameter space due to the near-singularity of the estimated Fisher information. Then we have tried to set some restrictions that the discounted parameters γ_j ($j = 1, 2$) have the same value γ and we set $c_{11} = c_{12}, c_{21} = c_{22}$ for instance. The results of estimation with this restriction have been given in Kunitomo et al. (2017) with the datasets of day-start to day-end. Here we report the estimation results of Model 3 with further restriction as $c = c_{11} = c_{12} = c_{21} = c_{22}$.

Overall, the results suggest that Models 2 and 3 are better than Model 1. In addition, according to AIC, Model 2 is better than Model 3 mainly because the latter is over-parametrized for Tokyo–New York markets. Hence, we adopted Model 2 in the following non-causality tests.

6.2 Non-causality tests

In applying the GNC test procedure, we set the impact function as $c(x) = x$. We report our empirical results for the hypothesis $H_0 : \alpha_{ij} = 0$ using the likelihood ratio test (LRT) statistics based on the Tokyo–New York data. For the null-hypothesis $H_0 : \alpha_{21} = 0$, LRT statistic is $2 \times (-3562.015 + 3562.017) \sim 0$, and we could not reject the null-hypothesis. (The upper 95% critical point of $\chi^2(1)$ is 3.481 in Table 5.) This means that changes of the Japanese financial market have little impact on the U.S. financial market.

For testing the null-hypothesis $H_0 : \alpha_{12} = 0$, LRT statistic based on the Tokyo–New York data was $2 \times (-3562.017 + 3572.843) = 21.652$, and the null-hypothesis was rejected. Thus, there is a significant effect from U.S. financial markets to Tokyo financial market (see Table 5).

Similarly, in Tokyo–London markets, for the null-hypothesis $H_0 : \alpha_{21} = 0$, LRT statistic was $2 \times (-3660.215 + 3660.215) \sim 0.0$; that the null-hypothesis was not rejected. Thus, knock-on effects from Tokyo to London financial market are rather limited.

For the null-hypothesis $H_0 : \alpha_{12} = 0$, LRT statistic based on the Tokyo–London data was $2 \times (-3660.215 + 3665.593) \sim 10.756$, and the null-hypothesis was

Table 9 Tail distributions

	Log likelihood	σ_1^*	ξ_1
Tokyo	– 1919.307	0.00757	0.22778
SD		0.00051	0.05552
	Log likelihood	σ_2^*	ξ_2
HK	– 1888.716	0.00861	0.15773
SD		0.00055	0.05076

rejected. This means that the London market affects Tokyo market (see Tables 4 and 6).

To summarize our findings among three major financial markets, the effects of the Japanese market on the U.S. and London are rather limited, while we found significant effects of both of these markets on the Tokyo market. This finding agrees with several empirical findings obtained using different statistical methods as explained by Kunitomo et al. (2017).

6.3 Example 2: Tokyo–Hong Kong markets

For the second example, we have used daily data of day-start to day-minimum from the Nikkei-225 and the Hansen Index of Hong-Kong during January 2, 1990–August 26, 2015, which is the same sample period as Example 1. Since the trading periods in these two financial markets are quite similar, we expected simultaneous movements in the two markets. For the estimated models with two dimensions ($d = 2, p = 3$), we take the impact functions $c(x)$ as Model 1 $c(x) = 1$ and Model 2 $c(x) = x$. Because there can be many additional parameters in Model 3, which has the general form of impact functions, the estimated results are often not statistically significant and we omitted reporting our results thereof.

We first maximize the likelihood L_{2T}^* to estimate the marginal distributions of financial market returns. As we have shown before, we confirmed that the marginal distributions of market returns have thicker tails than the normal distribution in Table 9. Hence, it may be appropriate to use the GPD in our estimation. It is because the estimates of ξ_i ($i = 1, 2$) are positive and it means that the Frechet-type tail distribution is appropriate

The estimated model consists of two dimensions ($d = 2$ and $p = 3$), and 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_{1T}^* . The full likelihood can be calculated using L_{1T}^* and L_{2T}^* . The SD of the estimated coefficients are evaluated numerically using the inverse of the estimated Fisher information matrix.

Model 1

We estimated the intensity function as

Table 10 Tokyo–Hong Kong

	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

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

Again the ML estimates can sometimes be numerically unstable, we set restrictions so that the discounted parameters γ_{ij} ($i, j = 1, 2, 3$) have the same value γ . We show the estimation results in Table 10.

Note that in the above table N_1 of Model 1 corresponds to Tokyo and N_2 corresponds to Hong Kong in Tokyo–Hong Kong markets.

Model 2

We estimated the intensity function as

Table 11 Tokyo–Hong Kong

	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

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

We present our estimation results in Table 11. From our estimated results, we find that Model 2 is better than Model 1 as in Example 1. When comparing Tables 10 and 11, we see several interesting findings. The value of AIC in Model 2 is better than Model 1 as we observed in the Tokyo–New York and Tokyo–London datasets. The estimates of coefficients of past effects are often statistically insignificant in the estimated intensity functions (α_{12} and α_{21}), while the contemporaneous effects of the co-jump term are statistically significant (α_{13} and α_{23}). This aspect basically agrees with our motivations for developing the SHPP models.

6.4 Non-causality tests

In applying the Granger non-causality test procedure, we set the impact function as $c(x) = x$. We report our empirical results for the hypothesis $H_0 : \alpha_{ij} = 0$ using LRT statistics.

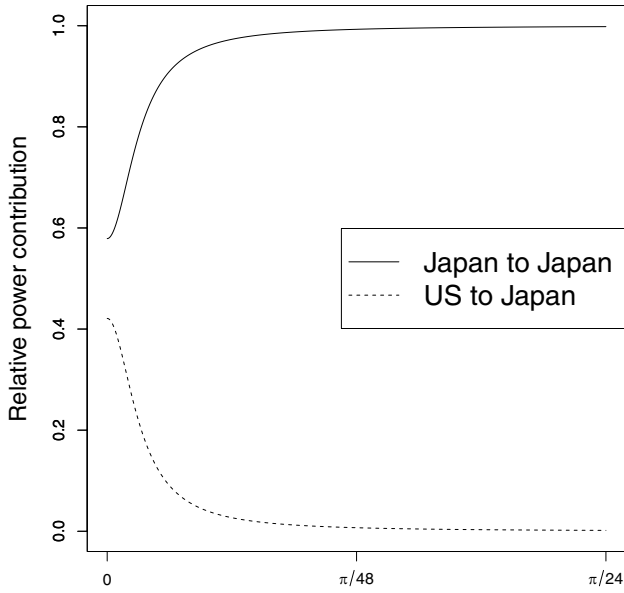


Fig. 5 Relative power contributions of Nikkei-225 (Tokyo–NY model)

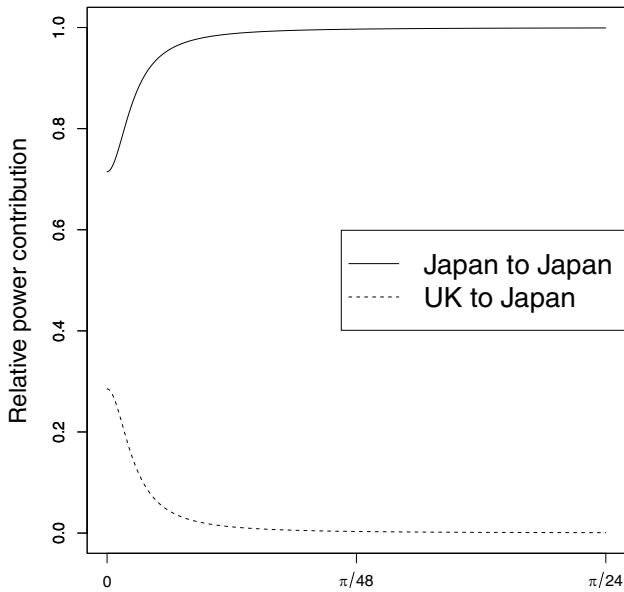


Fig. 6 Relative power contributions Nikkei-225 (Tokyo–London model)

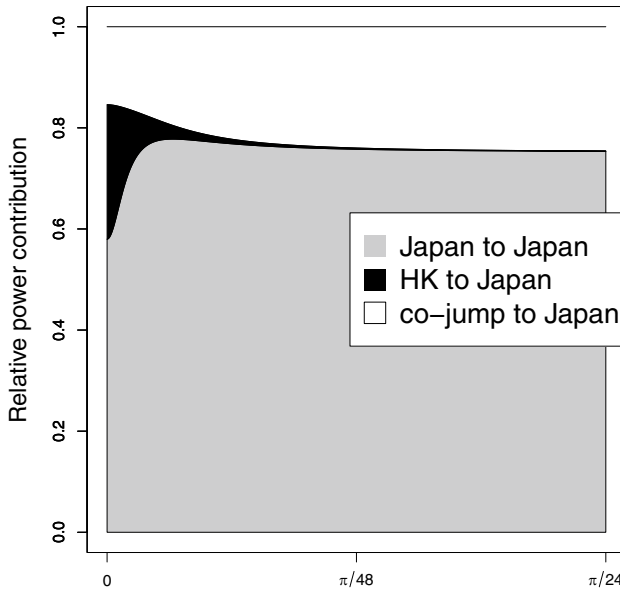


Fig. 7 Relative power contributions of Nikkei-225 (Tokyo–HK model)

For the null-hypothesis $H_0 : \alpha_{13} = 0$, LRT statistic based on Tokyo–Hong Kong data was 11.14 and we reject the null-hypothesis. (The upper 95% critical value of $\chi^2(1)$ is 3.481). Thus, we revealed a significant instantaneous causal relationship between the Japanese financial market and Hong-Kong financial markets.

For testing the null-hypothesis $H_0 : \alpha_{12} = 0$, LRT statistics was 0.0, and the null hypothesis was accepted. In addition, for testing the null-hypothesis $H_0 : \alpha_{12} = 0, \alpha_{13} = 0$, LRT statistic was 11.14; thus the null-hypothesis was rejected.

For the null-hypothesis $H_0 : \alpha_{21} = 0$, LRT statistic was 0.006 and we cannot reject the null-hypothesis. (The upper 95% critical value of $\chi^2(1)$ is 3.481). For testing the null-hypothesis $H_0 : \alpha_{23} = 0$, LRT statistic was 2.42, and the null-hypothesis was accepted. Similarly, for the null-hypothesis $H_0 : \alpha_{21} = 0, \alpha_{23} = 0$ LRT statistic was 2.66, and the null-hypothesis could not be rejected.

To summarize our findings in this subsection among Tokyo and Hong Kong financial markets, we found that the simultaneous effects of the two markets are significant, while the effects of past events are rather small.

6.5 Further empirical analysis

Here we employ spectral decomposition and RPC as explained in Sect. 3; see Figs. 5, 6 and 7 for the United States, United Kingdom, and Hong Kong, respectively. In the former (two) decompositions, we assume there are no co-jumps, while in the last one co-jumps are permitted. We adopted Models with $c_{ij}(x) = x$

because the resulting models minimize AIC. The graphs are depicted at each frequency, but truncated in x axis because it seems that our empirical data of point processes do not have much information in high frequencies.

In particular, Fig. 5 gives the spectral decomposition based on the estimated intensity model (Model 2) from Japan (Nikkei-225) data, which gives the relative contributions from Japan itself and from US (S&P 500). Figure 6 gives the spectral decomposition based on the estimated intensity model (Model 2) from Japan (Nikkei-225) data, which gives the relative contributions from Japan itself and from UK (FTSE). Then Fig. 7 gives the spectral decomposition based on the estimated intensity model with co-jumps (Model 2) from Japan (Nikkei-225) data, which gives the relative contributions from Japan itself, Hong Kong (Hansen) and the instantaneous relation.

From these figures, we found that for the relationship between the Tokyo–New York financial markets, self contribution of past events plays a major role, while there is some contribution from New York to Tokyo in the low frequency, which corresponds to the long-run relationship. On the other hand, for the relationship between the Tokyo–Hong Kong financial markets, the instantaneous contribution and the self contribution play major roles in all frequencies. This aspect reflects the fact that we used SHPP models.

7 Conclusions

In this paper, we developed a new method of econometric analysis of multivariate time series of events and proposed the simultaneous multivariate Hawkes-type point process (SHPP) modeling. Unlike some existing studies, we developed and used new statistical models for simultaneous sudden, large events and delayed events occurring explicitly. Using the SHPP models, we investigated Granger causality and instantaneous Granger causality on several financial markets and economies, and developed bespoke non-causality tests.

By applying GNC and IGNC tests, we revealed the important relationships among major financial markets and several empirical findings. In the Tokyo–New York financial markets, there is a strong unidirectional causation, while in the Tokyo–Hong Kong financial markets the simultaneous effects are dominant.

Several questions remain to be answered. First, although we used Hawkes-type marked point processes, there can be many possible non-linear point processes and Kurisu (2018) discussed one way to justify the use of SHPP models. In economic and financial econometrics, it is standard to handle discrete time series observations in yearly, monthly, weekly, daily, hourly, and per minute terms. Thus, we need a coherent way of investigating abrupt or sudden events, and we propose one way to deal with discrete time series events in this paper. In this respect, it should be interesting to investigate the robustness of our empirical results further. Second, the choice of threshold parameter is an important issue that is related to the relevance of the GPD in SEVT. Since we used a simple threshold parameter, we need a convincing justification on the choice of threshold. Finally, when $d > 2$ there can be many parameters to be estimated, and the estimated parameters would often be statistically

insignificant. This aspect is important when we have the possibility of co-jumps and we used the likelihood and its AIC as a criterion. We need to investigate the problem of choosing statistical point process models further.

These issues are currently under investigation, we shall report our progress in these respects on another occasion.

Acknowledgements We thank two referees of this journal for their detailed comments to the earlier version of this paper. We also thank Yusuke Amano for providing computational assistance and also Takaki Hayashi for useful comments to the earlier version. The research was supported by Grant-in-Aid for Scientific Research (JP25245033 and 15H01943) from the JSPS, and D. Kurisu and N. Awaya have been further supported by Grant-in-Aid as JSPS Research Fellows (16J06454 and 17J03957).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix: Mathematical details

In this Appendix, we provide some mathematical details, which are related to expositions in the main text, and the outline of proofs of Theorems 1 and 2 in Sect. 4.2. 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 since been classical and widely cited in later studies. He obtained the results using a martingale central limit (CLT) theorem for point processes, which is not widely known among econometricians; in addition, the asymptotic normality holds under more general conditions that are often cited. Hence, we first discuss some properties of jump martingales with a continuous time parameter and then apply the CLT for deriving the Wilks property of non-causality tests. We omit the subscript n without any loss of generality in this Appendix.

(i) A Martingale CLT

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space and $\{\mathcal{F}_t\}$ ($0 \leq t \leq T$) be the continuous-time filtration. We present a general martingale CLT for one-dimensional point processes, which is useful for our application.

Theorem A.1: *Let an \mathcal{F} -adapted simple point process on \mathbf{R}_+ be N and the \mathcal{F} -adapted (continuous) compensator be A . We assume that for any $T (> 0)$ there exist an \mathcal{F}_t -adapted function $g_T(t)$ ($0 \leq t \leq T$) and an \mathcal{F}_0 -adapted (positive) random variable η , which satisfy the following conditions.*

$$(i) \quad \mathbf{E} \left[\frac{1}{T} \int_0^T (g_T(x))^2 dA(x) \right] < \infty,$$

(ii) For any $\delta (> 0)$,

$$\frac{1}{T^{1+\delta}} A(T) \xrightarrow{p} 0, \tag{A.35}$$

(iii) As $T \rightarrow \infty$

$$\frac{1}{T} \int_0^T (g_T(x))^2 dA(x) \xrightarrow{p} \eta^2, \tag{A.36}$$

(iv) For any $c > 0$ and some positive ϵ ($0 < \epsilon < 1/6$), as $T \rightarrow \infty$

$$\mathbf{E} \left[\frac{1}{T} \int_0^T [g_T(x)I(|g_T(x)| > cT^\epsilon)]^2 dA(x) | \mathcal{F}_0 \right] \xrightarrow{p} 0. \tag{A.37}$$

Then

$$X_T = \frac{1}{\sqrt{T}} \int_0^T g_T(x)[dN(x) - dA(x)] \tag{A.38}$$

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

Remark A-1 The method of proof is basically a modification of that given by Daley and Vere–Jones (Vol-II, 2008) as their Theorem 14.5.I. They derived a martingale CLT under a Lyapunov condition. Our condition includes the speed of divergence of compensator, which may be a reasonable condition for applications.

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

$$\zeta_T(t, y) = \exp \left(iy \int_0^t f_T(u)[dN(u) - dA(u)] + \frac{1}{2}y^2 \int_0^t [f_T(u)]^2 dA(u) \right). \tag{A.39}$$

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

$$\begin{aligned} \zeta_T(t, y) = & \exp \left(\frac{1}{2}y^2 \int_0^t [f_T(u)]^2 - iy \int_0^t f_T(u) dA(u) \right) \\ & \times \prod_i [(1 + (\exp(iyf_T(t_i)) - 1))\Delta N(t_i)], \end{aligned} \tag{A.40}$$

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

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

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 -measurable 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

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

where

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

For $0 < u < T \wedge \tau$, from (A.39) we find that

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

In addition, we take positive c_T and using 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

$$\begin{aligned}
 & |\mathbf{E}(Z\zeta_T(T \wedge \tau)|\mathcal{F}_0) - Z| \\
 & \leq \exp\left(\frac{1}{2}y^2\eta^2\right)\mathbf{E}\left\{ |Z| \left[y^2 \int_0^{T \wedge \tau} |f_T(u)|^2 I[|f_T(u)| > c_T] dA(u) \right. \right. \\
 & \quad \left. \left. + |y\theta|^3 \int_0^{T \wedge \tau} |f_T(u)|^3 I[|f_T(u)| \leq c_T] dA(u) \right] \right\}, \tag{A.41}
 \end{aligned}$$

where $|\theta| \leq 1$.

We set $f_T(u) = g_T(u)/\sqrt{T}$ and $c_T = c/T^{\delta_1}$ for a positive c and $3\delta_1 > 1$. Then for the second term of the right-hand side of (A.41)

$$\int_0^{T \wedge \tau} |f_T(u)|^3 I[|f_T(u)| \leq c_T] dA(u) \leq \frac{c^3}{T^{3\delta_1}} A(T \wedge \tau),$$

which converges to zero by (A.35).

By setting $\varepsilon = 1/2 - \delta_1$ with $1/3 < \delta_1 < 1/2$, the first term of the right-hand side of (A.41) is

$$\int_0^{T \wedge \tau} |f_T(u)|^2 I[|f_T(u)| > c_T] dA(u) = \frac{1}{T} \int_0^{T \wedge \tau} |g_T(u)|^2 I[|g_T(u)| > cT^\varepsilon] dA(u),$$

which converges to zero by (A.37) with $0 < \varepsilon < 1/6$.

The left-hand side multiplying $\exp[(-1/2)y^2\eta^2]$ is equal of larger than

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

where

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

It is because

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

Since $|\rho_T| \leq 1$ and (A.36), we find $\rho \rightarrow 1$ as $T \rightarrow \infty$. Then we have that $\mathbf{E}[Z(\rho_T - 1)e^{itX_T}] \rightarrow 0$ as $T \rightarrow \infty$ and

$$\mathbf{E}[Z \exp(iyX_T)] \longrightarrow \mathbf{E}\left[Ze^{-\frac{1}{2}y^2\eta^2/2}\right]. \tag{A.42}$$

Hence, using weak-convergence and stable convergence (Daley and Vere-Jones (Vol-II, 2008), Jacod and Protter 2012), we have $X_T \rightarrow X$ (\mathcal{F}_0 -stably). This means that for any bounded \mathcal{F}_0 -measurable random variable Z , $\mathbf{E}[Ze^{iyX}] = \mathbf{E}[Ze^{-y^2\eta^2/2}]$, which implies $\mathbf{E}[e^{iyX_T/\eta} | \mathcal{F}_0] = e^{-y^2/2}$. \square

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*

$$G_1(t) = \prod_i (1 + w(t_i))\Delta N(t_i), G_2(t) = \exp\left(\int_0^t v(u)dA(u)\right), \tag{A.43}$$

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,

$$\begin{aligned} &G_1(t)G_2(t) - G_1(0)G_2(0) \\ &= \int_0^t G_1(u)dG_2(u) + \int_0^t G_2(u)dG_1(t) \\ &= \int_0^t G_1(u-)G_2(u)v(u)dA(u) + \sum_i G_2(t_i)G_1(t_i-)w(t_i)\Delta N(t_i). \end{aligned} \tag{A.44}$$

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 \mathcal{F} -adapted function $g_T(x)$ and any $\epsilon > 0$ we have

$$Y_T = \frac{1}{T^{1/2+\epsilon}} \int_0^T g_T(x)[dN(x) - dA(x)] \xrightarrow{p} 0. \tag{A.45}$$

Thus, we do not need to use the *Ergodic Theorem* 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 case. 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 $q \times p \mathcal{F}_t$ -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 $\eta = (\eta_{ij})$, we assume the following conditions.

- (i)' $\max_{1 \leq i, j \leq q} \max_{1 \leq k \leq p} \mathbf{E}[\frac{1}{T} \int_0^T |g_T^{ik}(t)| |g_T^{jk}(t)| dA_k(t)] < \infty,$
- (ii)' For any $\delta (> 0),$

$$\frac{1}{T^{1+\delta}} \max_{1 \leq k \leq p} A_k(T) \xrightarrow{p} 0, \tag{A.46}$$

- (iii)' As $T \rightarrow \infty$

$$\frac{1}{T} \int_0^T \sum_{k=1}^p g_T^{ik}(t) g_T^{jk}(t) dA_k(t) \xrightarrow{p} \eta_{ij}, \tag{A.47}$$

where $\eta = (\eta_{ij})$ is a $q \times q$ non-negative definite matrix.

- (iv)' For any $c > 0$ and some positive $\epsilon (0 < \epsilon < 1/6),$ as $T \rightarrow \infty$

$$\max_{1 \leq k \leq p} \mathbf{E} \left[\frac{1}{T} \int_0^T \|\mathbf{g}_T^{\cdot k}(t)\|^2 I(\|\mathbf{g}_T^{\cdot k}(t)\| > cT^\epsilon) dA_k(t) | \mathcal{F}_0 \right] \xrightarrow{p} 0, \tag{A.48}$$

where $\mathbf{g}_T^{\cdot k}(t) = (g_T^{1,k}, \dots, g_T^{p,k})'$.

Here we abuse the notation $N_i (i = 1, \dots, p)$ slightly, which may differ from that in the main text. Under the above conditions, we have the next result.

Theorem A.2: *For the point 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*

$$\mathbf{X}_T = \frac{1}{\sqrt{T}} \int_0^T \sum_{i=1}^p \mathbf{g}_T^{\cdot i}(t) [dN_i(t) - dA_i(t)] \tag{A.49}$$

converges to $\eta^{1/2} \mathbf{u}$ 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 $\eta^{1/2} \eta^{1/2} = \eta.$

(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 $N_i(s, \theta) (i = 1, \dots, p)$ over the observation period $[0, T].$ We take $\theta = (\theta_i) \in R^r.$ Then the log-likelihood function is given by

$$L_T(\theta) = \sum_{i=1}^p L_{iT}(\theta), \tag{A.50}$$

where

$$L_{iT}(\theta) = \int_0^T \log \lambda_i(s, \theta) dN_i(s) - \int_0^T \lambda_i(s, \theta) ds, \quad (\text{A.51})$$

and its derivatives are given by

$$\frac{\partial L_{iT}(\theta)}{\partial \theta} = \int_0^T \frac{\log \lambda_i(s, \theta)}{\partial \theta} [dN_i(s) - \lambda_i(s, \theta) ds], \quad (\text{A.52})$$

and

$$\frac{\partial^2 L_{iT}(\theta)}{\partial \theta \partial \theta'} = \int_0^T \frac{1}{\lambda_i(s, \theta)} \frac{\partial^2 \lambda_i(s, \theta)}{\partial \theta \partial \theta'} [dN_i(s) - \lambda_i(s, \theta) ds] - \int_0^T \left[\frac{\log \lambda_i(s, \theta)}{\partial \theta} \right] \left[\frac{\lambda_i(s, \theta)}{\partial \theta'} \right] ds. \quad (\text{A.53})$$

Theorem A.3: Let the log-likelihood function be $L_T(\theta)$, the log-likelihood function under the true parameter vector 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 $T \rightarrow \infty$

$$2\{L_T(\hat{\theta}_{ML}) - L_T(\theta_0)\} \xrightarrow{d} \chi(r), \quad (\text{A.54})$$

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

$$\frac{1}{T} \sum_{i=1}^p \int_0^T \left[\frac{\partial \log \lambda_i(s, \theta)}{\partial \theta} \frac{\partial \log \lambda_i(s, \theta)}{\partial \theta'} \right] \lambda_i(s, \theta) ds \xrightarrow{p} I(\theta_0) > 0 \text{ (positive definite)}, \quad (\text{A.55})$$

$$\frac{1}{\sqrt{T}} \sum_{i=1}^p \int_0^T \left[\frac{\partial \log \lambda_i(s, \theta)}{\partial \theta} \right] [dN_i(s) - \lambda_i(s, \theta) ds] \xrightarrow{w} N_r(0, \mathbf{I}(\theta_0)), \quad (\text{A.56})$$

$$\frac{1}{T} \sum_{i=1}^p \int_0^T \left[\frac{\partial^2 \lambda_i(s, \theta)}{\partial \theta \partial \theta'} \right] \frac{1}{\lambda_i(s, \theta)} [dN_i(s) - \lambda_i(s, \theta) ds] \xrightarrow{p} 0, \quad (\text{A.57})$$

and

$$\frac{1}{T} \sum_{i=1}^p \int_0^T \left[\frac{\partial \log \lambda_i(s, \theta)}{\partial \theta} \frac{\partial \log \lambda_i(s, \theta)}{\partial \theta'} \right] [dN_i(s) - \lambda_i(s, \theta) ds] \xrightarrow{p} 0, \quad (\text{A.58})$$

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

Remark A-2 As corollaries of Theorems A.2 and A.3, it is straightforward and standard, but lengthy to give the formal proofs of Theorems 1 and 2 as the non-causality tests we have developed and discussed in Sect. 4.

Remark A-3 As a final remark, we re-emphasize that while Ogata (1978) discussed a set of sufficient conditions for the consistency and asymptotic normality of the ML estimator in one-dimensional self-exciting point processes, we 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 mean the mixed Gaussian distribution in the present formulation of the Appendix. The limiting χ^2 property of the statistics is often called the Wilks Property.

References

- Ait-Sahalia, Y., & Jacod, J. (2014). *High-frequency financial econometrics*. Princeton: Princeton University Press.
- Ait-Sahalia, Y., Cacho-Diaz, J., & Laeven, L. (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, 117, 585–606.
- Bacry, E., Mastromatteo, I., & Muzy, J.-F. (2015). Hawkes processes in finance. *Market Microstructure and Liquidity*, 1, 1550005, World Scientific.
- Bartlett, M. S. (1963). The spectral analysis of point processes. *Journal of Royal Statistical Society (B)*, 25–2, 264–296.
- Daley, D. J., & Vere-Jones, D. (2003, 2008). *An Introduction to the Theory of Point Processes*, Volume I, Volume II, 2nd Edition, Springer, New York .
- Embrechts, P., Kluppelberg, C., & Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. New York: Springer.
- Embrechts, P., Liniger, T., & Lin, L. (2011). Multivariate Hawkes processes: An application to financial data. *Journal of Applied Probability, Special*, 48A, 367–378.
- Florens, J.-P., & Fougere, D. (1996). Noncausality in continuous time. *Econometrica*, 64, 1195–1212.
- Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37, 161–194.
- Grothe, O., Korniihuk, V., & Manner, H. (2014). Modeling multivariate extreme events using self-exciting point processes. *Journal of Econometrics*, 182, 269–289.
- Hamao, Y., Masulis, R. W., & Ng, V. (1990). Correlations in price changes and volatility across international stock markets. *Review of Financial Studies*, 3(1990), 281–308.
- Hawkes, A. G. (1971). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society, Series B*, 33–3, 438–443.
- Hosoya, Y., Oya, K., Takimoto, T., & Kinoshita, R. (2017). *Characterizing interdependencies of multiple time series: Theory and applications*. New York: Springer.
- Ikeda, N., & Watanabe, S. (1989). *Stochastic Differential Equations and Diffusion Processes* (2nd ed.). Amsterdam: North-Holland.
- Jacod, J., & Protter, P. (2012). *Discretization of Processes*. New York: Springer.

- Kunitomo, N., Ehara, A., & Kurisu, D. (2017). A causality analysis of financial markets by multivariate Hawkes-type models (in Japanese). *Journal of Japan Statistical Society*, 46–2, 137–171.
- Kurisu, D. (2018). Discretization of self-exciting peaks over threshold models, Tokyo Tech IEEE Working Paper 2018-3, Tokyo Institute of Technology. https://educ.titech.ac.jp/iee/eng/publications/file/pub_19457.pdf
- Ogata, Y. (1978). The asymptotic behavior of maximum likelihood estimators of stationary point processes. *Annals of Institute of Statistical Mathematics*, 30, 243–261.
- Ogata, Y. (2015). Studies of probabilistic prediction of earthquakes: A survey. *Statistical Mathematics*, 63–11, 3–27. (in Japanese).
- Protter, P. (2003). *Stochastic integration and differential equations*. New York: Springer.
- Resnick, S. (2007). *Heavy-tail phenomena*. New York: Springer.
- Solo, V. (2007) Likelihood functions for multivariate point processes with coincidences. *Proceedings of the 46th IEEE Conference on Decision and Control*, 4245–4250.