Marked Hawkes process modeling of price dynamics and volatility estimation

Kyungsub Lee[*]and Byoung Ki Seo[†]

August 31, 2016

Abstract

A simple Hawkes model have been developed for the price tick structure dynamics incorporating market microstructure noise and trade clustering. In this paper, the model is extended with random mark to deal with more realistic price tick structures of equities. We examine the impact of jump in price dynamics to the future movements and dependency between the jump sizes and ground intensities. We also derive the volatility formula based on stochastic and statistical methods and compare with realized volatility in simulation and empirical studies. The marked Hawkes model is useful to estimate the intraday volatility similarly in the case of simple Hawkes model.

1 Introduction

In this paper, the tick dynamics of stock prices observed at ultra-high-frequency level are modeled based on the symmetric marked Hawkes process and the empirical properties of the price dynamics are examined. The simple self and mutually excited Hawkes model for the price dynamics with the unit jump size incorporates the stylized facts of the ultra-high-frequency financial data, such as market microstructure noise and order clustering. On the other hand, random size jumps, i.e., not a constant jump, as in the simple Hawkes model, are observed in the tick structure of equity markets, particularly when there is a high ratio between the stock price and minimum tick size. By combining the Hawkes model with a mark structure, which has additional information for each event, more realistic model of the tick price dynamics is proposed to deal with the random size jumps.

Recent studies on (ultra)-high-frequency data and the market microstructure have been developed in several ways. The volume of literature on the financial theory of the market microstructure and limit order book (Roşu, 2009), and the role of algorithmic trading at a high-frequency rate (Foucault, 2012; Chaboud et al., 2014; Hoffmann, 2014) is increasing. A number of studies focused on the reduced form or stochastic modeling of the limit order dynamics and order executions; the reader may refer to Lo et al. (2002), Cont et al. (2010), Malo and Pennanen (2012), Cont and De Larrard (2013), Abergel and Jedidi (2013).

The statistical property of the ultra-high-frequency data is also an important subject because they exhibit the distinctive characteristics from the macro price dynamics. For example, care should be taken when applying the typical statistical methods to ultra-high-frequency data, and when computing the realized volatility (Andersen et al., 2003) due to microstructure noise, which refers to the mean reverting properties of the price processes at high frequency level. Previous studies (Aït-Sahalia et al. (2005), Zhang et al. (2005), Hansen and Lunde (2006) and Aït-Sahalia et al. (2011)) measured the volatility of the return in the presence of market microstructure noise. Huth and Abergel (2014) examined the lead/lag relationship between asset prices and showed that there are significant cross correlations in the futures/stock at the high-frequency contrast with the daily data cases where cross correlations are negligible. The lead/lag relationship among the international index futures of different countries were also observed by Alsayed and McGroarty (2014). For the semi-Markov model with price jumps to explain the micorstructure noise, consult Fodra and Pham (2015).

The financial asset price time series at the ultra-high-frequency level exhibits several autocorrelations that are not observed on a daily basis. Under the tick structure with a minimum tick size of price

---

[*]Department of Statistics, Yeungnam University, Gyeongsan, Gyeongbuk 38541, Korea

[†]Corresponding author, School of Management Engineering, UNIST(Ulsan National Institute of Science and Technology), Ulsan 44919, Korea

variation, the price dynamics is a pure jump process that consists of up jumps and down jumps. First, the frequency of up movements tends to increase with increasing frequency of the past down movements and vice versa. This causes a mean reverting property in the price dynamics, even though the correlations last for less than a few seconds. Second, there are also autocorrelations between the movements of the same direction. This causes volatility clustering that is different from the clustering on the macro level, which is typically modeled by GARCH (Bollerslev, 1986) or the stochastic volatility model (Heston, 1993), because the clustering properties in the tick structure last for only a few seconds. The durations of the autocorrelations are much shorter than those of the autocorrelation observed on a daily level.

These properties are well incorporated into the Hawkes model, which belongs to the class of point processes and is introduced by Hawkes (1971). Therefore, there has been an increase in the related work of modeling price dynamics based on the Hawkes process. The bivariate Hawkes process was introduced to model the buy and sell order arrivals and the impact of the orders on future prices was examined (Hewlett, 2006). The generalized Hawkes models was used to study the dependence between the occurrence of time trades and changes to the mid quote as well as the dependences between trading days (Bowsher, 2007). Large (2007) examined the market resilience after trades using the limit order book data and mutually excited Hawkes models.

Bacry et al. (2012) explained the non-parametric estimation method for the symmetric Hawkes process based on high-frequency futures data. Based on the mutually exciting Hawkes process, Bacry et al. (2013) suggested the mathematical framework that incorporates the market microstructure noise and the Epps effect, which is the correlation between the returns of two different assets at high sampling frequency. The trade clustering properties of the price dynamics on the micro level was well incorporated by the self-excited Hawkes process (Da Fonseca and Zaatour, 2014). The formulas for the moments and correlation function for the self and mutually excited Hawkes process were derived by Da Fonseca and Zaatour (2015). A multivariate Hawkes process was introduced for the up and down price movements and buy and sell orders to explain the stylized facts of the market impact and microstructure (Bacry and Muzy, 2014). Zheng et al. (2014) suggested a multivariate constrained Hawkes process to describe the dynamics of the bid and ask prices. Lee and Seo (2014) focused on the daily and intraday volatility estimation based on the symmetric Hawkes process and compared the result with the realized volatility. For more about the kernel estimation in the Hawkes model, consult Bacry et al. (2016) and for the correlation and lead-lag relationship in a multi-asset model using the Hawkes process, consult Da Fonseca and Zaatour (2016).

The previous studies focused mainly on the simple point process model, where the jump size is constant. In the present study, the existing simple Hawkes model was extended to the marked Hawkes model to handle more realistic price movements in stock markets where the random size jumps (marks) are introduced. A marked point process was introduced based on the ACM-ACD model, where points are the transaction time and the marks are information on the transaction (Russell and Engle, 2005). The Hawkes process was adopted to explain the aftermath effects of the marks, which is more convenient for calculating the useful formula. The future effects of the marks depend on the absolute sizes of the marks and hence a linear impact function is introduced to deal with the future impact of the mark. Our empirical study shows that the estimates of the slope parameter of the impact function are significant positive values in stock markets. This suggests that the larger marks tend to magnify the future intensities more than the smaller marks. For the distribution of the mark, a specific distribution is not assumed in this paper but the empirical distribution is used for estimations and volatility calculation. Our model is not limited to the independent mark as the empirical studies show the intensity dependent mark distribution.

The remainder of the paper is organized as follows. In Section 2, the marked Hawkes model is proposed to describe the tick price dynamics of equities. Section 3 and Section 4 present the simulation results and empirical studies, respectively. Section 5 concludes the paper. The mathematical proofs are reported in the Appendix.

2 Symmetric marked Hawkes model

2.1 Marked point process

This subsection introduces the basic concepts of marked point processes. The mathematical framework is in line with Daley and Vere-Jones (2003). With the given complete separable metric state space, $\mathcal{X}$, a point process is a measure to count the number of random events that occur in an open set that belongs

to the σ-field of $\mathcal{X}$'s Borel set, $\mathcal{B}_\mathcal{X}$. To deal with random events, a probability space $(\Omega, \mathbb{P})$ is introduced. A point process $N(\omega)$ for $\omega \in \Omega$, or simply N, is the counting measure on $\mathcal{B}_\mathcal{X}$, i.e., $N(A) = N(\omega, A)$ is a random non-negative finite integer for any bounded $A \in \mathcal{B}_\mathcal{X}$. In other words, $N(A)$ is a random variable that counts the number of events in A. From now on, the term ω is omitted for the notational simplicity.

A marked point process is a more complex model that is introduced to describe not only the location of random events but also additional information, called the mark, attached to each event. A marked point process is a point process $\{(t_\ell, k(t_\ell))\}$ with locations $t_\ell \in \mathbb{R}$ and marks $k(t_\ell)$ in a mark space $\mathcal{K}$. The location space is not necessarily $\mathbb{R}$, but in this paper, the space is defined as a real line to model the price movements over time. In addition, the mark space $\mathcal{K} = \mathbb{Z}^+$ is the space of price jump sizes. This suggests that the absolute jump sizes of the price movements are represented by positive integer multiples of a minimum jump size, δ.

For price dynamics modeling, there are two marked point processes, N_1 and N_2, which represent the up and down movements, respectively. This study assumes that the probabilistic properties of k_1 and k_2, the mark size of N_1 and N_2, respectively, are the same. For each $i = 1, 2$, the ground process $N_{gi}(\cdot) = N_i(\cdot, \mathbb{Z}^+)$, which is the marginal process for locations, is itself a point process. Each ground process, N_{gi}, describes the arrival times of up or down price movements.

The price process is assumed to be represented by a two dimensional marked Hawkes point process, which belongs to the marked point processes. Let $\lambda_{gi}(t)$ be the conditional intensity of N_{gi} upon filtration $\mathcal{F}_{t-}$ which is the minimal σ-algebra generated by the history of the marked point processes of N_1 and N_2 on $(-\infty, t)$. The conditional intensities are $\mathcal{F}_t$-adapted stochastic processes, that heuristically speaking, satisfy

$$\lambda_{gi}(t)\mathrm{d}t = \mathbb{E}[N_{gi}(t + \mathrm{d}t) - N_{gi}(t)|\mathcal{F}_{t-}].$$

Owing to the discontinuity, the intensities may not be unique at the discontinuities and $\lambda_{gi}(t)$ are considered the left continuous modifications in general, but if the intensities are used as integrators of some stochastic integrations, then the right continuous modification of λ_{gi} is used as the integrators. The marked Hawkes model for the price dynamics does not need to be symmetric in general. However, for computational ease, it is convenient to assumed that the joint distributions of (k_1, λ_{g1}) and (k_2, λ_{g2}) are the same, for example, $\mathbb{E}[k_1] = \mathbb{E}[k_2]$.

The marked Hawkes process assumption implies that, for each $i = 1, 2$, the intensities of the ground processes satisfy

$$\lambda_{gi}(t) = \mu + \sum_{j=1,2} q_{ij} \int_{(-\infty, t) \times \mathbb{Z}^+} g_{ij}(k_j)\phi_{ij}(t - u)N_j(\mathrm{d}u \times \mathrm{d}k_j) \tag{1}$$

$$= \mu + \sum_{j=1,2} q_{ij} \sum_{-\infty < u_\ell < t} g_{ij}(k_j(u_\ell))\phi_{ij}(t - u_\ell) \tag{2}$$

where u_ℓ are the event times. The mark size is represented by k_j in the integration form with the counting measure, and is represented by $k_j(u_\ell)$ in the summation form with the associated event time u_ℓ. With the counting measure, inside the integration, k_j can be considered to be function on $\mathbb{R} \times \mathbb{Z}^+$ such that $(u, k_j) \to k_j$. If the occurrence time of k_j needs to be specified, then the mark is written as $k_j(u)$ as in Eq. (2) with some time notation, u, indicating that $k_j(u)$ takes place at time u.

The Hawkes processes generated by the above intensities are defined by the ancestor-offspring argument (as long as the intensities are finite). The immigrant ancestor of type i with mark k arrives at the system from outside in a Poisson process at rate μ. These ancestors generate offspring and the generated offspring become the new ancestor to generate new offspring. Owing to the ancestor type j born at time u with mark k, whether immigrant or not, the type i offspring are generated by a Poisson process with a rate $q_{ij}g_{ij}(k)\phi_{ij}(t - u)$ at time t. The Poisson rate is emphasized by $g_{ij}(k)$ at time u and decreases with $\phi_{ij}(t - u)$ as t increases. A normalization method to determine $q_{ij}, g_{ij}(k)$ and $\phi_{ij}(t - u)$ is used in general, since the combination of $q_{ij}g_{ij}(k)\phi_{ij}(t - u)$ is not unique.

We assume the exponential decay kernel for $\phi_{ij}(t - u) = \phi(t - u) = \beta e^{-\beta(t-u)}, \beta > 0$, which is normalized such that

$$\int_0^\infty \phi(\tau)\mathrm{d}\tau = 1.$$

The impact of mark g_{ij} is also normalized in the sense that $\mathbb{E}[g_{ij}(k)] = 1$. In addition, $q_{ij} > 0$ are called the branching coefficients and $\mathbf{Q} := \{q_{ij}\}_{i,j=1,2}$ is called the branching matrix. With an exponential decay kernel, $(\lambda_{g1}, \lambda_{g2})$ is Markovian and the calculations in this paper depend largely on this property.

This paper considers the case that the distribution of mark of type i may depend on λ_{gi} and hence the conditional distribution of the mark is represented by $f(k_i|\lambda_{gi}(t))$. On the other hand, this paper does not assume the specific parametric distribution for the mark size except when paths are generated under the simulation study. The estimation procedures and volatility analysis are performed without specifying the mark distribution.

The counting measure N_i can be interpreted as a stochastic jump process. To apply the stochastic integration theory later, without notational confusion, the associated jump processes is defined as

$$N_i(t) = \int_{(0,t]\times\mathbb{Z}^+} k_i N_i(\mathrm{d}u \times \mathrm{d}k_i)$$

where in the l.h.s., the stochastic process N_i is represented by the sole parameter t and in the r.h.s., the measure N_i is represented by both the location t and mark size k. In the stochastic process representation, $N_i(t)$ counts the number of events over $(0,t]$ with weight k_i. The jump processes $N_i(t)$ are considered to be right continuous with left limits. Similarly, the ground processes also have the stochastic process representations:

$$N_{gi}(t) = \int_{(0,t]\times\mathbb{Z}^+} N_i(\mathrm{d}u \times \mathrm{d}k_i)$$

which counts the number of events over $(0,t]$ without considering the jump size k_i. As a jump process, the ground process is also regarded to be right continuous with left limits.

2.2 Linear impact function

A linear impact function of the mark is introduced. The impact function g_{ij} and the distribution of the mark should satisfy some additional criteria so that the marked Hawkes process does not blow up.

Assumption 1. *(i) The ground intensities λ_{gi} are assumed to be stationary.*
(ii) The impact functions have the same formula for all $i,j = 1,2$ and are linear with a slope parameter η:

$$g(k) := g_{ii}(k) = g_{ji}(k) = \frac{1+(k-1)\eta}{\mathbb{E}[1+(k-1)\eta]}.$$

(iii) The branching matrix is symmetric with

$$q_s := q_{11} = q_{22} = \frac{\alpha_s}{\beta}\mathbb{E}[1+(k-1)\eta], \quad q_c := q_{12} = q_{21} = \frac{\alpha_c}{\beta}\mathbb{E}[1+(k-1)\eta].$$

(iv) For $i = 1,2$, we assume

$$\mathbb{E}[k_i\lambda_{gi}(t)] = K_{i\lambda_{gi}}\mathbb{E}[\lambda_{gi}(t)] \tag{3}$$

for some constant $K_{i\lambda_{gi}} > 0$ and

$$\{1+(K_{i\lambda_{gi}}-1)\eta\}\left(\frac{\alpha_s}{\beta}+\frac{\alpha_c}{\beta}\right) < 1. \tag{4}$$

In addition, the joint distributions of (k_1,λ_{g1}) and (k_2,λ_{g2}) are the same and we have $K_{1\lambda_{g1}} = K_{2\lambda_{g2}}$.

The condition of Eq. (4) is similar to the existence condition of the simple symmetric Hawkes process except for the additional term $\{1+(K_{i\lambda_{gi}}-1)\eta\}$. Indeed, Assumption 1 (ii)$\sim$(iv) leads to the weak stationarity of λ_{gi}. Under Eq. (1) with the exponential decay function, we have

$$\lambda_{g1}(t) = \mu + q_s \int_{(-\infty,t)\times\mathbb{Z}^+} g(k_1)\beta e^{-\beta(t-u)} N_1(\mathrm{d}u \times \mathrm{d}k_1)$$

$$+ q_c \int_{(-\infty,t)\times\mathbb{Z}^+} g(k_2)\beta e^{-\beta(t-u)} N_2(\mathrm{d}u \times \mathrm{d}k_2), \tag{5}$$

$$\lambda_{g2}(t) = \mu + q_c \int_{(-\infty,t)\times\mathbb{Z}^+} g(k_1)\beta e^{-\beta(t-u)} N_1(\mathrm{d}u \times \mathrm{d}k_1)$$

$$+ q_s \int_{(-\infty,t)\times\mathbb{Z}^+} g(k_2)\beta e^{-\beta(t-u)} N_2(\mathrm{d}u \times \mathrm{d}k_2). \tag{6}$$

4

The point process (N_1, N_2) defined under the above ground intensity is then a two dimensional marked self and mutually excited Hawkes process with a linear impact function.

Assuming the integrand of the following formula is integrable, a predictable finite variation process

$$\int_{.}^{t} \mathbb{E}[g(k_i)|\lambda_{gi}(u)]\lambda_{gi}(u)\beta e^{-\beta(t-u)}du$$

is a compensator for

$$\int_{(\cdot,t)\times\mathbb{Z}^+} g(k_i)\beta e^{-\beta(t-u)} N_i(du \times dk_i),$$

and hence

$$\int_{(\cdot,t)\times\mathbb{Z}^+} g(k_i)\beta e^{-\beta(t-u)} N_i(du \times dk_1) - \int_{.}^{t} \mathbb{E}[g(k_i)|\lambda_{gi}(u)]\lambda_{gi}(u)\beta e^{-\beta(t-u)}du$$

is a martingale. Therefore, by taking the unconditional expectation for the ground intensity formulas in Eqs. (5) and (6), we have

$$\mathbb{E}[\lambda_{g1}(t)] = \mu + q_s \int_{-\infty}^{t} \mathbb{E}[\mathbb{E}[g(k_1)|\lambda_{g1}(u)]\lambda_{g1}(u)]\beta e^{-\beta(t-u)}du$$

$$+ q_c \int_{-\infty}^{t} \mathbb{E}[\mathbb{E}[g(k_2)|\lambda_{g2}(u)]\lambda_{g2}(u)]\beta e^{-\beta(t-u)}du,$$

$$\mathbb{E}[\lambda_{g2}(t)] = \mu + q_c \int_{-\infty}^{t} \mathbb{E}[\mathbb{E}[g(k_1)|\lambda_{g1}(u)]\lambda_{g1}(u)]\beta e^{-\beta(t-u)}du$$

$$+ q_s \int_{-\infty}^{t} \mathbb{E}[\mathbb{E}[g(k_2)|\lambda_{g2}(u)]\lambda_{g2}(u)]\beta e^{-\beta(t-u)}du$$

and by Eq. (3),

$$\mathbb{E}[\mathbb{E}[g(k_i)|\lambda_{gi}(u)]\lambda_{gi}(u)] = \mathbb{E}[g(k_i)\lambda_{gi}(u)] = \frac{\{1 + (K_{i\lambda_{gi}} - 1)\eta\}\mathbb{E}[\lambda_{gi}(u)]}{1 + (\mathbb{E}[k] - 1)\eta}$$

where $\mathbb{E}[k] = \mathbb{E}[k_i]$ since k_1 and k_2 have the same distributional property. We write

$$\mathbb{E}[\lambda_{g1}(t)] = \mu + \frac{\alpha_s}{\beta} \int_{-\infty}^{t} \{1 + (K_{1\lambda_{g1}} - 1)\eta\}\mathbb{E}[\lambda_{g1}(u)]\beta e^{-\beta(t-u)}du$$

$$+ \frac{\alpha_c}{\beta} \int_{-\infty}^{t} \{1 + (K_{2\lambda_{g2}} - 1)\eta\}\mathbb{E}[\lambda_{g2}(u)]\beta e^{-\beta(t-u)}du, \tag{7}$$

$$\mathbb{E}[\lambda_{g2}(t)] = \mu + \frac{\alpha_c}{\beta} \int_{-\infty}^{t} \{1 + (K_{1\lambda_{g1}} - 1)\eta\}\mathbb{E}[\lambda_{g1}(u)]\beta e^{-\beta(t-u)}du$$

$$+ \frac{\alpha_s}{\beta} \int_{-\infty}^{t} \{1 + (K_{2\lambda_{g2}} - 1)\eta\}\mathbb{E}[\lambda_{g2}(u)]\beta e^{-\beta(t-u)}du \tag{8}$$

or, in a system of linear differential equation,

$$\begin{bmatrix} \frac{d\mathbb{E}[\lambda_{g1}(t)]}{dt} \\ \frac{d\mathbb{E}[\lambda_{g2}(t)]}{dt} \end{bmatrix} = \begin{bmatrix} \alpha_s\{1 + (K_{1\lambda_{g1}} - 1)\eta\} - \beta & \alpha_c\{1 + (K_{1\lambda_{g1}} - 1)\eta\} \\ \alpha_c\{1 + (K_{1\lambda_{g1}} - 1)\eta\} & \alpha_s\{1 + (K_{1\lambda_{g1}} - 1)\eta\} - \beta \end{bmatrix} \begin{bmatrix} \mathbb{E}[\lambda_{g1}(t)] \\ \mathbb{E}[\lambda_{g2}(t)] \end{bmatrix} + \begin{bmatrix} \beta\mu \\ \beta\mu \end{bmatrix}$$

where $K_{1\lambda_{g1}} = K_{2\lambda_{g2}}$ is used. The eigenvalues of the system are

$$(\xi_1, \xi_2) = \left(-\beta + (\alpha_s - \alpha_c)\{1 + (K_{1\lambda_{g1}} - 1)\eta\}, -\beta + (\alpha_s + \alpha_c)\{1 + (K_{1\lambda_{g1}} - 1)\eta\}\right)$$

and the solution of the above system is

$$\begin{bmatrix} \mathbb{E}[\lambda_{g1}(t)] \\ \mathbb{E}[\lambda_{g2}(t)] \end{bmatrix} = \frac{-\lambda_{g1}(0) + \lambda_{g2}(0)}{2} e^{\xi_1 t} \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \frac{\lambda_{g1}(0) + \lambda_{g2}(0)}{2} e^{\xi_2 t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \frac{\mu\beta}{\xi_2}\left(1 - e^{\xi_2 t}\right) \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

If Eq. (4) holds, then $\xi_1, \xi_2 < 0$ and the solution converges to a constant as $t \to \infty$. A similar argument is applied to the second moments of $\lambda_{g1}(t)$ and $\lambda_{g2}(t)$ and $\lambda_{gi}(t)$ are weakly stationary at least in the long-run or by assuming that $\lambda_{gi}(0)$ are equal to their long-run expectations.

Note that by the symmetry and the stationarity of λ_{gi}, Eqs (7) and (8) lead directly to

$$\mathbb{E}[\lambda_{g1}(t)] = \mu + \frac{\alpha_s}{\beta}\{1 + (K_{1\lambda_{g1}} - 1)\eta\}\mathbb{E}[\lambda_{g1}(t)] + \frac{\alpha_c}{\beta}\{1 + (K_{1\lambda_{g1}} - 1)\eta\}\mathbb{E}[\lambda_{g2}(t)],$$

$$\mathbb{E}[\lambda_{g2}(t)] = \mu + \frac{\alpha_c}{\beta}\{1 + (K_{1\lambda_{g1}} - 1)\eta\}\mathbb{E}[\lambda_{g1}(t)] + \frac{\alpha_s}{\beta}\{1 + (K_{1\lambda_{g1}} - 1)\eta\}\mathbb{E}[\lambda_{g2}(t)].$$

and

$$\left(\mathbf{I} - \{1 + (K_{1\lambda_{g1}} - 1)\eta\} \begin{bmatrix} \dfrac{\alpha_s}{\beta} & \dfrac{\alpha_c}{\beta} \\ \dfrac{\alpha_c}{\beta} & \dfrac{\alpha_s}{\beta} \end{bmatrix}\right) \begin{bmatrix} \mathbb{E}[\lambda_{g1}(t)] \\ \mathbb{E}[\lambda_{g2}(t)] \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \end{bmatrix}.$$

By the symmetry between λ_{g1} and λ_{g2},

$$\mathbb{E}[\lambda_{g1}(t)] = \mathbb{E}[\lambda_{g2}(t)] = \frac{\mu\beta}{\beta - (\alpha_s + \alpha_c)\{1 + (K_{1\lambda_{g1}} - 1)\eta\}}. \tag{9}$$

Therefore, if condition (4) is satisfied, then the ground processes are well defined, i.e., the expectation of the ground intensities are positive and finite.

2.3 Second moment property

In this subsection, the volatility formula of the asset return generated by the marked Hawkes processes under the symmetric model is calculated. The symmetry of the model implies that i and j are interchangeable which makes the formula simple. In the following notation, various Ks and αs are defined in a similar manner to those in Eq. (3), which simplify the notations.

Notation 1. *For a jump process X such as λ_{gi} and N_i, let*

$$\mathbb{E}[k_i X(t)] = K_{iX}\mathbb{E}[X(t)], \quad \mathbb{E}[k_i^2 X(t)] = K_{iX}^{(2)}\mathbb{E}[X(t)].$$

(In the previous, when $X = \lambda_{gi}$, subscript is omitted for simplicity as in Eq. (3), i.e., $K = K_{i\lambda_{gi}}$.) Furthermore,

$$\bar{\bar{K}} = 1 + 2(K_{1\lambda_{g1}} - 1)\eta + (K_{1\lambda_{g1}}^{(2)} - 2K_{1\lambda_{g1}} + 1)\eta^2 = 1 + 2(K_{2\lambda_{g2}} - 1)\eta + (K_{2\lambda_{g2}}^{(2)} - 2K_{2\lambda_{g2}} + 1)\eta^2,$$

$$\bar{K} = K_{1\lambda_{g1}} + (K_{1\lambda_{g1}}^{(2)} - K_{1\lambda_{g1}})\eta = K_{2\lambda_{g2}} + (K_{2\lambda_{g2}}^{(2)} - K_{2\lambda_{g2}})\eta,$$

$$\breve{\alpha} = \alpha\{1 + (K_{1\lambda_{g1}^2} - 1)\eta\} = \alpha\{1 + (K_{2\lambda_{g2}^2} - 1)\eta\},$$

$$\tilde{\alpha} = \alpha\{1 + (K_{2\lambda_{g1}\lambda_{g2}} - 1)\eta\} = \alpha\{1 + (K_{1\lambda_{g2}\lambda_{g1}} - 1)\eta\},$$

$$\acute{\alpha} = \alpha\{1 + (K_{1\lambda_{g1}N_1} - 1)\eta\} = \alpha\{1 + (K_{2\lambda_{g2}N_2} - 1)\eta\},$$

$$\grave{\alpha} = \alpha\{1 + (K_{2\lambda_{g2}N_1} - 1)\eta\} = \alpha\{1 + (K_{1\lambda_{g1}N_2} - 1)\eta\},$$

and

$$\mathbf{M} = \begin{bmatrix} \breve{\alpha}_s - \beta & \tilde{\alpha}_c \\ \breve{\alpha}_c & \tilde{\alpha}_s - \beta \end{bmatrix}, \quad \mathbf{M}_2 = \begin{bmatrix} \acute{\alpha}_s - \beta & \grave{\alpha}_c \\ \acute{\alpha}_c & \grave{\alpha}_s - \beta \end{bmatrix},$$

$$\mathbf{K} = \begin{bmatrix} K_{1\lambda_{g1}} & 0 \\ 0 & K_{1\lambda_{g1}} \end{bmatrix}, \quad \mathbf{K}_2 = \begin{bmatrix} K_{1\lambda_{g1}^2} & 0 \\ 0 & K_{2\lambda_{g1}\lambda_{g2}} \end{bmatrix}, \mathbf{K}_3 = \begin{bmatrix} K_{1\lambda_{g1}N_1} & 0 \\ 0 & K_{1\lambda_{g1}N_2} \end{bmatrix}.$$

Theorem 1. *Let (N_1, N_2) be a two dimensional marked self and mutually excited Hawkes process with a linear impact function under Assumption 1 with ground intensities of Eqs. (5) and (6). If the price process S_t follows*

$$S_t = S_0 + \delta(N_1(t) - N_2(t))$$

with a minimum jump size δ, then the unconditional variance of the return over $[0, t]$ is

$$\mathrm{Var}\left(\frac{S_t - S_0}{S_0}\right) = \frac{\delta^2}{S_0^2}\mathbb{E}[(N_1(t) - N_2(t))^2]$$

with

$$\begin{bmatrix} \mathbb{E}[N_1^2(t)] \\ \mathbb{E}[N_1(t)N_2(t)] \end{bmatrix} = -\mathbb{E}[\lambda_{g1}(t)]\mathbf{K}_3 \left\{ \beta\mu\mathbf{K}\mathbf{M}_2^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} t^2 \right.$$
$$\left. + \left(2\mathbf{M}_2^{-1} \begin{bmatrix} \alpha_s\bar{K} \\ \alpha_c\bar{K} \end{bmatrix} - \mathbf{M}_2^{-1}\mathbf{K}_2\mathbf{M}^{-1} \begin{bmatrix} (\alpha_s^2 + \alpha_c^2)\bar{\bar{K}} \\ 2\alpha_s\alpha_c\bar{\bar{K}} \end{bmatrix} + 2\mathbf{M}_2^{-1}(\mathbf{K}\mathbf{M}_2^{-1} - \mathbf{K}_2\mathbf{M}^{-1}) \begin{bmatrix} \beta\mu \\ \beta\mu \end{bmatrix} - \begin{bmatrix} K_{1\lambda_g1}^{(2)}/K_{1\lambda_{g1}N_1} \\ 0 \end{bmatrix} \right) t \right\}.$$

Proof. See A. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

The result is slightly complicated but the following remark is useful for the practice.

Remark 2. *By assuming* $K_{1\lambda_{g1}N_1} \approx K_{1\lambda_{g1}N_2}$ *and* $K_{1\lambda_{g1}^2} \approx K_{1\lambda_{g1}\lambda_{g2}}$, *we have*

$$\mathbb{E}[(N_1(t) - N_2(t))^2] = 2(\mathbb{E}[N_1^2(t)] - \mathbb{E}[N_1(t)N_2(t)])$$
$$\approx 2K_{1\lambda_{g1}N_1}\mathbb{E}[\lambda_{g1}(t)] \left(\frac{K_{1\lambda_{g1}^2}\bar{\bar{K}}(\alpha_s - \alpha_c)^2}{(\beta - \breve{\alpha}_s + \breve{\alpha}_c)(\beta - \acute{\alpha}_s + \acute{\alpha}_c)} + \frac{2(\alpha_s - \alpha_c)\bar{K}}{\beta - \acute{\alpha}_s + \acute{\alpha}_c} + \frac{K_{1\lambda_{g1}}^{(2)}}{K_{1\lambda_{g1}N_1}} \right) t.$$
$$\tag{10}$$

Under the assumption, $\mathbf{M}, \mathbf{M}_2$ *and* $\mathbf{K}_2$ *are symmetric and the variance formula becomes simple. In addition, if all* K*s are equal to 1, then the variance formula is reduced to*

$$\mathbb{E}[(N_1(t) - N_2(t))^2] = 2\mathbb{E}[\lambda_{g1}(t)]\frac{\beta^2}{(\beta - \alpha_s - \alpha_c)^2}t.$$

which is the same formula of the variance in the simple Hawkes model.

Corollary 3. *Under the assumption in Theorem 1, if the marks* k_i *are i.i.d., then* $K_{iX} = K := \mathbb{E}[k_i]$ *and* $K_{iX}^{(2)} = K^{(2)} := \mathbb{E}[k_i^2]$ *for all* $X \in \{\lambda_{gi}, \lambda_{gi}\lambda_{gj}, \lambda_{gi}N_j : i, j = 1, 2\}$ *and* $\breve{\alpha} = \tilde{\alpha} = \acute{\alpha} = \grave{\alpha}$. *Therefore,*

$$\mathbb{E}[(N_1(t) - N_2(t))^2] = 2K\mathbb{E}[\lambda_{g1}(t)] \left(\frac{K\bar{\bar{K}}(\alpha_s - \alpha_c)^2}{(\beta - \breve{\alpha}_s + \breve{\alpha}_c)^2} + \frac{2(\alpha_s - \alpha_c)\bar{K}}{\beta - \breve{\alpha}_s + \breve{\alpha}_c} + \frac{K^{(2)}}{K} \right) t$$

where $\bar{\bar{K}}$ *and* $\bar{K}$ *are now represented by*

$$\bar{\bar{K}} = 1 + 2(K - 1)\eta + (K^{(2)} - 2K + 1)\eta^2, \qquad \bar{K} = K + (K^{(2)} - K)\eta.$$

2.4 Likelihood function

To estimate the parameters in the intensity processes, such as $\mu, \alpha_s, \alpha_c, \beta$ and η, the log-likelihood needs to be computed. The joint log-likelihood function of the realized interarrival of the jumps and marks of (N_1, N_2) over the period $[0, T]$ is represented by

$$\left(\int_{(0,T]} \log \lambda_{g1}(u) N_{g1}(\mathrm{d}u) + \int_{(0,T]} \log \lambda_{g2}(u) N_{g2}(\mathrm{d}u) - \int_0^T (\lambda_{g1}(u) + \lambda_{g2}(u))\mathrm{d}u \right)$$
$$+ \left(\int_{(0,T]\times\mathbb{Z}^+} \log f(k_1|\lambda_{g1}(u)) N_1(\mathrm{d}u \times \mathrm{d}k_1) + \int_{(0,T]\times\mathbb{Z}^+} \log f(k_2|\lambda_{g2}(u)) N_2(\mathrm{d}u \times \mathrm{d}k_2) \right)$$
$$=: \log L_g + \log L_m \tag{11}$$

where f denotes the conditional distribution of the mark k_i with a given λ_{gi}.

In the above formula, the log-likelihood is separated into two parts, $\log L_g$ and $\log L_m$. The first part $\log L_g$ is the log-likelihood function of the ground intensity processes, or more precisely, the joint log-likelihood function of the jump interarrival times. The second part $\log L_m$ is the log-likelihood function of the conditional mark distribution. When $\log L_g$ is used solely for the maximum likelihood estimation, then $\log L_g$ is indeed the conditional log-likelihood function of the jump interarrival times with the conditions on the realized marks, i.e., the estimation is the maximum likelihood estimation based on the interarrival times of jumps with given realized marks.

In the estimation procedure of the simulation and empirical studies later, no specific form of the joint distribution of the ground intensities and mark sizes is assumed but the empirical mark distribution is used to conduct the maximum likelihood procedure to maximize $\log L_g$, the conditional log-likelihood function of the jump interarrival times. Because the empirical mark distribution is used, the $\log L_m$ part does not affect the estimations of $\boldsymbol{\theta} = (\mu, \alpha_s, \alpha_c, \beta, \eta)$. On the other hand, if one assume a specific parametric modeling on the mark distribution by specifying a conditional distribution $f(k_i|\lambda_{gi})$ as in Subsection 3.1 and also want to estimate the parameters in f, then $\log L_m$ is also affected by $\boldsymbol{\theta}$ since when λ_{gi}s are inferred, it is calculated using $\boldsymbol{\theta}$.

In another aspect, the estimation procedure only on $\log L_g$ is possible since the marks k_i are observable. If the mark sizes are unobservable, it is then inevitable to assume a parametric modeling on the joint or conditional distribution between λ_{gi} and k_i to infer the mark size k_i. Owing to the parametric assumption on $f(k_i|\lambda_{gi}(u))$, as mentioned before, the parameter family of $\boldsymbol{\theta}$ also appears in the formula of $\log L_m$, and the changing values of $\boldsymbol{\theta}$ change not only the value $\log L_g$ but also $\log L_m$. Therefore, the estimator that maximizes $\log L_g$ may not converge to the estimator that maximizes $(\log L_g + \log L_m)$ as the sample size increases.

Fortunately, since all k_i are observable in our empirical study, the observed realized mark sizes are used to compute $\log L_g$, not the inferred k_i based on some parametric assumption. Hence, the estimator of $\boldsymbol{\theta}$ that maximizes $\log L_g$ are the maximum likelihood estimator associated with the conditional joint distribution of the interarrival times of jumps with the given marks. Note that $\log L_g$ and $\log L_m$ are separated not because the marks k_i are independent from the ground intensities but because $\log L_g$ can be represented as the log-likelihood function of the conditional joint distribution of the interarrival times with the given marks.

In practice, $\log L_g$ is computed as follows. With the presumed parameter values of $\alpha_s, \alpha_c, \beta$ and η, we compute the inferred ground intensity processes λ_{gi}s based on the realized N_is of the stock price process and Eqs. (5) and (6). With the inferred ground intensity processes, the realized values of the stochastic integration part are calculated by

$$\int_{(0,T]} \log \lambda_{gi}(u) N_{gi}(\mathrm{d}u) = \sum_{n=1}^{N_{gi}(T)} \log \lambda_{gi}(u_n)$$

where u_n denotes the realized jump time by N_{gi}. In addition, $\int_0^T (\lambda_{g1}(u) + \lambda_{g2}(u))\mathrm{d}u$ is computed using the Riemann integral which has a closed form formula. By repeating the above procedure with the changing presumed parameters, the numerical solver for the optimization try to find the global maximum of $\log L_g$.

Consider the concavity of the log-likelihood function. If the Hessian of the log-likelihood function is negative semi-definite for all parameters with a given realization, then the log-likelihood function is concave. However, the formula of the Hessian of the marked Hawkes model is complicated, we instead examine the conditional concavity when β and η are fixed. Note that, with the given realized jump times t_i, we have

$$\begin{aligned}
\log L_{g1}(T) &:= \int_{(0,T]} \log \lambda_{g1}(u) N_{g1}(\mathrm{d}u) - \int_0^T \lambda_{g1}(u)\mathrm{d}u \\
&= \sum_{t_i < T} \left(\log \lambda_{g1}(t_i) - \int_{t_{i-1}}^{t_i} \lambda_{g1}(u)\mathrm{d}u \right) - \int_{t_N}^T \lambda_{g1}(u)\mathrm{d}u \\
&= \sum_{t_i < T} \left(\log \lambda_{g1}(t_i) - \frac{1 - \mathrm{e}^{-\beta \tau_i}}{\beta} \lambda_{g1}(t_i) \right) - \int_{t_N}^T \lambda_{g1}(u)\mathrm{d}u
\end{aligned}$$

where t_N is the last jump time. Using the definition of λ_{g1}, each term in the summation can be rewritten

as follows:

$$\log \lambda_{g1}(t_i) - \frac{1 - e^{-\beta \tau_i}}{\beta} \lambda_{g1}(t_i)$$

$$= \log \left[\mu + (\lambda_{g1}(0) - \mu)e^{-\beta t_i} + \int_{(0,t_i] \times \mathbb{Z}_+} \alpha_s \{1 + (k_1 - 1)\eta\}e^{-\beta(t_i - u)} N_1(\mathrm{d}u \times \mathrm{d}k_1) \right.$$

$$\left. + \int_{(0,t_i] \times \mathbb{Z}_+} \alpha_c \{1 + (k_2 - 1)\eta\}e^{-\beta(t_i - u)} N_2(\mathrm{d}u \times \mathrm{d}k_2) \right]$$

$$- \frac{1 - e^{\beta \tau_i}}{\beta} \left[\mu + (\lambda_{g1}(0) - \mu)e^{-\beta t_i} + \int_{(0,t_i] \times \mathbb{Z}_+} \alpha_s \{1 + (k_1 - 1)\eta\}e^{-\beta(t_i - u)} N_1(\mathrm{d}u \times \mathrm{d}k_1) \right.$$

$$\left. + \int_{(0,t_i] \times \mathbb{Z}_+} \alpha_c \{1 + (k_2 - 1)\eta\}e^{-\beta(t_i - u)} N_2(\mathrm{d}u \times \mathrm{d}k_2) \right].$$

With fixed β and η, the term is represented as

$$\log(m\mu + a_s\alpha_s + a_c\alpha_c + C) - \frac{1 - e^{\beta \tau_i}}{\beta}(m\mu + a_s\alpha_s + a_c\alpha_c + C)$$

for some constants m, a_s, a_c, and C. Therefore, for any fixed β and η, the Hessian matrix of the term with respect to μ, α_s and α_c is

$$\frac{1}{\lambda_{g1}^2(t_i)} \begin{bmatrix} -m^2 & -ma_s & -ma_c \\ -ma_s & -a_s^2 & -a_s a_c \\ -ma_c & -a_s a_c & -a_c^2 \end{bmatrix}$$

which is negative semidefinite. The conditional Hessian of the log-likelihood function $\log L_g$ with fixed β and η is represented by

$$\sum_{t_i < T} \left(\frac{1}{\lambda_{g1}^2(t_i)} + \frac{1}{\lambda_{g2}^2(t_i)} \right) \begin{bmatrix} -m^2 & -ma_s & -ma_c \\ -ma_s & -a_s^2 & -a_s a_c \\ -ma_c & -a_s a_c & -a_c^2 \end{bmatrix}$$

which is also negative semidefinite.

Therefore, at least if the parameter values of β and η are fixed, the concavity of the log-likelihood $\log L_g(\mu, \alpha_s, \alpha_c | \beta, \eta)$ as a function of μ, α_s, α_c can be guaranteed, and we can assume that a numerical solver for the optimization will find the global maximum. Consider the points of β and η over a sufficiently large and dense grid. For each β and η, we can find the conditional global maximum of $\log L_g(\mu, \alpha_s, \alpha_c | \beta, \eta)$ over the parameter space of μ, α_s, α_c due to the conditional concavity. Now the interest is the shape of the conditional global maximums over the grid of β and η. If the conditional global maximum is still concave and the maximum of the conditional maximums can be found, then the numerical optimizer can be checked to determine if it finds the overall global maximum. In the numerical procedure later, the shape of the conditional log-likelihood function over a grid of β and η will be examined.

3 Simulation example

3.1 Symmetric model

In this paper, the specific distribution of the mark is generally not assumed. For the simulation study, however, it is necessary to assume a specific conditional distribution of the mark sizes to generate paths. Suppose that the mark k_i follows the conditional geometric distribution with

$$p(\lambda_{gi}(u)) = \frac{1}{\min(d + c\lambda_{gi}(u), U)}$$

for some constants c, d and U, i.e.,

$$\mathbb{P}(k_i = n | \lambda_{gi}(u)) = p(\lambda_{gi}(u))(1 - p(\lambda_{gi}(u)))^{n-1}.$$

9

Table 1: Simulation study for the marked Hawkes model with 500 sample paths

	μ	α_s	α_c	β	η	TSRV	H.Vol.
True	0.1000	0.9500	0.8200	2.2500	0.1900		
mean	0.0999	0.9496	0.8199	2.2487	0.1882	0.2807	0.2798
std.	0.0021	0.0194	0.0190	0.0326	0.0170	0.0333	0.0126
True	0.1500	0.6200	0.5000	1.9000	0.2200		
mean	0.1499	0.6193	0.5008	1.8999	0.2177	0.1317	0.1312
std.	0.0027	0.0172	0.0149	0.0419	0.0502	0.0101	0.0018
True	0.3000	1.0500	0.9200	2.3000	0.0100		
mean	0.3003	1.0508	0.9206	2.3014	0.0094	0.6391	0.6291
std.	0.0051	0.0136	0.0135	0.0229	0.0051	0.0562	0.0177
True	0.2000	1.1000	1.2600	2.5700	0.0100		
mean	0.2002	1.0997	1.2610	2.5702	0.0099	4.4299	4.3628
std.	0.0039	0.0154	0.0164	0.0238	0.0007	1.6973	0.6811

This suggests that the conditional expectation of the mark size k_i with a given ground intensity λ_{gi} is

$$\mathbb{E}[k_i|\lambda_{gi}(u)] = \min(d + c\lambda_{gi}(u), U)$$

for some slope c, intercept d, and upper bound U. It is needed to set the upper bound for the conditional mean of the mark size to prevent a blow up of the marked Hawkes process. With this setting, the conditional expectation of the impact depends on the current intensity:

$$\mathbb{E}[g(k_i)|\lambda_{gi}(u)] = \frac{1 + \{\min(d + c\lambda_{gi}(u), U) - 1\}\eta}{\mathbb{E}[1 + (k_i - 1)\eta]}.$$

With each differently presumed conditional distribution and parameter setting, 500 sample paths of the two dimensional marked Hawkes process and corresponding ground intensities are generated. The time horizon for the path is set to be 5.5 hours, which equals the time horizon used in empirical studies later. The simulation mechanism is similar to the simple Hawkes models but it needs to incorporate the mark size and its future impacts.

With the realized interarrival times of the generated path and realized mark sizes, the maximum likelihood estimation is performed on the maximized $\log L_g$ in Eq. (11) and the results are listed in Table 1. The table consists of three panels with different parameter settings, which are presented in 'True' rows. For the first panel, $c = 0.15, d = 1.0, U = 2.0$, for the second panel, $c = 0.18, d = 1.0, U = 2.2$; for the third panel, $c = 0.18, d = 1.0, U = 3.5$; and for the fourth panel, $c = 0.25, d = 1.0, U = 9$.

Because the likelihood for the ground processes was calculated, the estimates of $\mu, \alpha_s, \alpha_c, \beta$ and η were computed but not for c, d and U. The sample mean of the estimates with 500 sample paths are reported in the row 'mean'. The row 'std.' presents the sample standard deviations of each estimate with 500 samples. The table shows that the estimates are consistent with the true values.

Figure 1 presents the global maximums of the conditional log-likelihood functions with various β and η under the first simulation setting as explained in the previous section. The numerically computed conditional maximum points shows the concavity and it is expected that the numerical optimizer will find the global maximum in the procedure.

To calculate the volatility, we need to compute Ks in Notation 1 which involves several unconditional expectations with the mark size, intensities and counting processes. In the absence of the exact formula of the expectations due to the complicated relationship between the mark and the intensities, the following
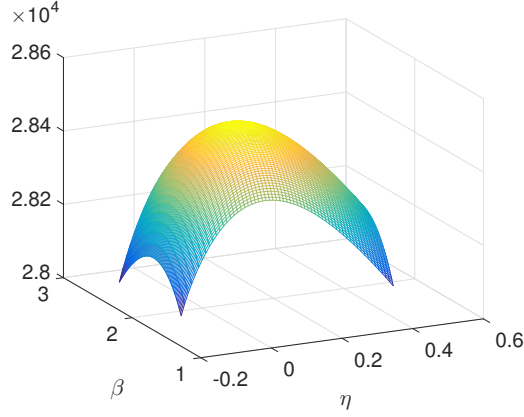
Figure 1: Maximums of conditional log-likelihood function over β and η

statistics are used for the expectations instead:

$$\mathbb{E}[\lambda_{gi}(t)] \approx \frac{1}{T} N_{gi}(T) \tag{12}$$

$$\mathbb{E}[k_i \lambda_{gi}(t)] \approx \frac{1}{T} N_i(T) \tag{13}$$

$$\mathbb{E}[k_i^2 \lambda_{gi}(t)] \approx \frac{1}{T} \int_{(0,T] \times \mathbb{Z}^+} k_i^2 N_i(\mathrm{d}u \times \mathrm{d}k_i) \tag{14}$$

$$\mathbb{E}[\lambda_{gi}^2(t)] \approx \frac{1}{T} \int_{(0,T] \times \mathbb{Z}^+} \lambda_{gi}(u) N_i(\mathrm{d}u \times \mathrm{d}k_i) \tag{15}$$

$$\mathbb{E}[k_i \lambda_{gi}^2(t)] \approx \frac{1}{T} \int_{(0,T] \times \mathbb{Z}^+} k_i \lambda_{gi}(u) N_i(\mathrm{d}u \times \mathrm{d}k) \tag{16}$$

$$\frac{1}{t} \mathbb{E}[\lambda_{gi}(t) N_i(t)] \approx \frac{2}{T^2} \int_{(0,T] \times \mathbb{Z}^+} N_i(u-) N_i(\mathrm{d}u \times \mathrm{d}k_i) \tag{17}$$

$$\frac{1}{t} \mathbb{E}[k_i \lambda_{gi}(t) N_i(t)] \approx \frac{2}{T^2} \int_{(0,T] \times \mathbb{Z}^+} k_i N_i(u-) N_i(\mathrm{d}u \times \mathrm{d}k_i) \tag{18}$$

where $[0, T]$ is an observation time interval. To calculate the right hand sides, the realized k_i, N_i and N_{gi} of the generated paths and inferred λ_{gi} from the estimates of $\mu, \alpha_s, \alpha_c, \beta$ and η are used. The inferred intensities λ_{gi} are computed using Eqs. (5) (6), once $\mu, \alpha_s, \alpha_c, \beta$ and η are estimated.

The expectations of the ground intensities are approximated by the sample average of the total number of corresponding up or down moves per unit time in Eq. (12). Similarly for $\mathbb{E}[k_i \lambda_{gi}(t)]$ where the counting process N_i is used instead to compute the sample average.

The right hand side of Eq. (14) is the sample average of the total number of jumps per unit time with weight k_i^2 for each jump and this approximates the left hand side. For Eqs. (15) and (16), consider

$$\mathbb{E}\left[\int_{(0,T] \times \mathbb{Z}^+} \lambda_{gi}(u) N_i(\mathrm{d}u \times \mathrm{d}k)\right] = \int_0^T \mathbb{E}[\lambda_{gi}^2(t)] \mathrm{d}t = T\mathbb{E}[\lambda_{gi}^2(t)]$$

$$\mathbb{E}\left[\int_{(0,T] \times \mathbb{Z}^+} k \lambda_{gi}(u) N_i(\mathrm{d}u \times \mathrm{d}k)\right] = \int_0^T \mathbb{E}[k_i \lambda_{gi}^2(t)] \mathrm{d}t = T\mathbb{E}[k_i \lambda_{gi}^2(t)].$$

Figure 2 presents the convergences of the computed K and $K^{(2)}$ with the above method as the sample size increases with the first parameter set in the simulation.

Furthermore, $\mathbb{E}[\lambda_{gi}(t) N_i(t)] = c_1 t + c_2$ for some constants c_1 and c_2, according to A and $\mathbb{E}[\lambda_{gi}(t) N_i(t)]/t$
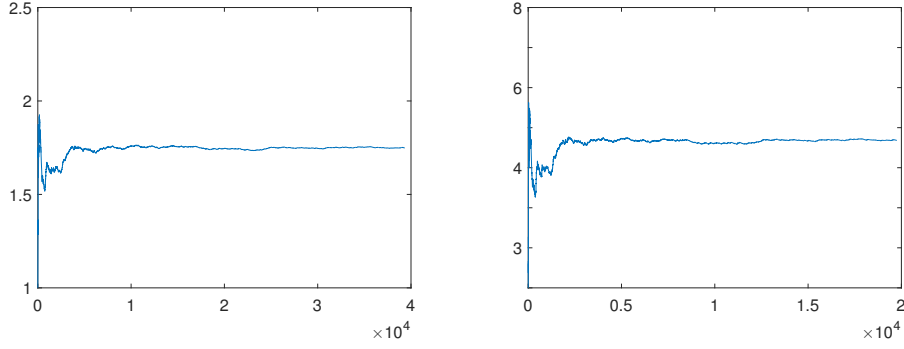
Figure 2: Convergences of $K_{1\lambda_{g1}}$ and $K^{(2)}_{1\lambda_{g1}}$ as sample size increases

converges to c_1 as t increases. Note that

$$\frac{2}{T^2}\mathbb{E}\left[\int_{(0,T]\times\mathbb{Z}^+} N_i(u-)N_{gi}(\mathrm{d}u\times\mathrm{d}k)\right] = \frac{2}{T^2}\int_0^T \mathbb{E}[\lambda_{gi}(t)N_i(t)]\mathrm{d}t$$

$$= c_1 + \frac{2c_2}{T} \approx c_1 \approx \frac{1}{2t}\mathbb{E}[\lambda_{gi}(t)N_i(t)]$$

with approximations for a large enough t and T. A similar argument was applied to Eq. (18).

The column 'H. vol' is for the mean of the volatility estimates computed by the likelihood estimates of $\mu, \alpha_s, \alpha_c, \beta, \eta$ and Ks using Remark 2. This was compared with the two scale realized volatility (TSRV) in the column 'TSRV' proposed by Zhang et al. (2005). The small time scale was set to be one second and the large time scale was set to be five minutes for the TSRV computation. The results show that the Hawkes volatility and TSRV are similar. The standard deviations of the Hawkes volatility are smaller than those of the TSRV for all simulation cases.

3.2 Other examples

This subsection examines the cases where there is a discrepancy between the Hawkes volatility and the realized volatility. First, the fully characterized Hawkes model is examined, i.e., the coefficients of the branching matrix is represented by

$$q_{ij} = \frac{\alpha_{ij}}{\beta_{ij}}\mathbb{E}[1 + (k_j - 1)\eta]$$

with the linear impact function of Assumption 1 (ii). Under this setting, no symmetry is guaranteed. Recall that in the symmetric model, $\alpha_s = \alpha_{11} = \alpha_{22}$, $\alpha_c = \alpha_{12} = \alpha_{21}$, and $\beta = \beta_{11} = \beta_{12} = \beta_{21} = \beta_{22}$.

Table 2 lists the estimation results of the fully characterized Hawkes model with simulated paths with the presumed parameters. The presumed parameters are presented in the 'true' column and 500 sample paths are generated over a one day time horizon, more precisely, 5.5 hours as in the previous example. The columns 'full' report the means and standard deviations of the estimates under the maximum likelihood estimation with the fully characterized Hawkes model. The likelihood estimations were also performed under the symmetric Hawkes model, even though the paths are generated by the fully characterized Hawkes model. The results are presented in the columns 'symmetric' at the centers of the rows of corresponding parameters. For example, μ is presented at the center of two rows of μ_1 and μ_2, α_s is presented at the center of two rows of α_{11} and α_{22}, etc..

The 'S.Vol.' represents the sample volatility of the return computed by the sample standard deviation of the closing stock prices generated by the 500 sample paths. The TSRV and marked Hawkes volatility with corresponding standard deviations are reported in column 'TSRV' and 'H.Vol', respectively. The Hawkes volatility is calculated using the estimates of the symmetric Hawkes model. Two volatilities are biased around 4% compared to the sample volatility. The TSRV are larger than the sample volatilities and the Hawkes volatilities are smaller in these cases.

Second, the symmetric marked Hawkes models were examined, where the model parameters change during the sample period. Table 3 lists the estimation results with the symmetric Hawkes models of the

Table 2: Fully characterized Hawkes model with 500 sample paths

	full			symmetric			full			symmetric	
	true	mean	std.	mean	std.		true	mean	std.	mean	std.
μ_1	0.1461	0.1467	0.0038	0.1345	0.0026		0.1130	0.1131	0.0030	0.1152	0.0024
μ_2	0.1155	0.1159	0.0032				0.1149	0.1153	0.0033		
α_{11}	0.3185	0.3204	0.0150	0.4102	0.0148		0.4994	0.5031	0.0242	0.5252	0.0159
α_{22}	0.3821	0.3865	0.0219				0.4682	0.4799	0.0210		
α_{12}	0.9812	0.9848	0.0282	1.1512	0.0223		0.5937	0.5992	0.0232	0.7012	0.0199
α_{21}	1.4949	1.5000	0.0334				0.9754	0.9854	0.0368		
β_{11}	1.1799	1.1893	0.0567				1.8305	1.8512	0.0948		
β_{22}	1.9553	1.9840	0.1195	2.0547	0.0315		1.4706	1.5142	0.0666	1.8744	0.0364
β_{12}	2.0952	2.1077	0.0697				1.5963	1.6110	0.0624		
β_{21}	2.5030	2.5132	0.0587				2.7850	2.8064	0.1036		
η	0.1488	0.1501	0.0235	0.1424	0.0255		0.1761	0.1768	0.0216	0.1756	0.0225
	$c = 0.1, d = 1.0, U = 2.0$						$c = 0.08, d = 1.5, U = 3.0$				
	S.Vol.	TSRV	std.	H.Vol.	std.		S.Vol.	TSRV	std.	H.Vol.	std.
	0.1405	0.1463	0.0146	0.1346	0.0051		0.1853	0.1897	0.0161	0.1795	0.0044

Table 3: Simulation study for the marked Hawkes model with 500 sample paths with time varying parameters

	μ	α_s	α_c	β	η	c	d	U	S.Vol.	TSRV	H.Vol.
True 1	0.1000	1.1000	1.2600	2.5700	0.0100	0.2500	1	7			
True 2	0.1000	1.1000	1.2600	2.5700	0.0100	0.2500	1	1.5			
mean	0.1017	1.1017	1.2662	2.5667	0.0085				0.6431	0.5801	0.6288
std.	0.0022	0.0205	0.0265	0.0319	0.0039					0.1990	0.1381
True 1	0.1000	1.1000	1.2600	2.5700	0.1000	0.1000	1	7			
True 2	0.0500	0.5000	0.5000	2.0000	0.1000	0.1000	1	1.5			
mean	0.0375	1.0162	1.1190	2.3567	0.0233				0.2496	0.2204	0.2300
std.	0.0010	0.0302	0.0347	0.0438	0.0142					0.0378	0.0205

5.5 hour's time horizon but the model parameter of the first one hour of the period is according to the row of 'True 1' and in the rest of the period, the model follows 'True 2'. In the first panel, the varying part is the upper bound of the conditional mean of the mark distribution. In other words, during the first part of the sample period, the price process is quite volatile due to the possible large size of the mark, and the remaining part is rather stable. This mimics the case of the 2010 Flash Crash and the empirical analysis will be performed later. The result shows the discrepancy between the TSRV and the Hawkes volatility which are both less than the sample volatility and the TSRV is even less than the Hawkes volatility.

4 Empirical study

4.1 Data

The empirical studies used the ultra high-frequency tick-by-tick data of some major stock prices consisting of several years with the best bid and ask quotes reported in the New York Stock Exchange (NYSE). The time horizon of the sample for each day is set to be from 10:00 to 15:30. The data of 30 minutes immediately after the opening and before the closing time were not used to reduce the seasonality effects. The price movement patterns are usually different at the near opening and closing from the rest of the day.

The jump sizes of the price movements of equities in the S&P 500 are not constant over time particularly when the price of the equity is high and hence the ratio between the price and the minimum tick size in transaction on the NYSE, $0.01, is high. The tick size of the NYSE was reduced from $1/8 to $1/16 in 1997 and from $1/16 to $0.01 in 2001. In this paper, the mid-price movements is considered for the marked Hawkes modeling to remove the bid-ask bounce and hence the minimum jump size is the

Table 4: Mark size distribution (%) of IBM (left) and GE (center) and CVX (right) from 2008 to 2011

mark	2008	2009	2010	2011	2008	2009	2010	2011	2008	2009	2010	2011
1	51.80	59.98	80.88	57.04	89.81	98.39	99.61	99.68	60.69	68.30	91.55	77.21
2	21.53	20.57	13.96	18.41	7.61	1.51	0.34	0.30	20.02	19.75	7.32	16.20
3	11.22	10.89	3.53	9.54	1.55	0.00	0.00	0.00	9.25	8.43	0.77	4.77
4	6.36	5.50	1.01	5.80	0.52	0.00	0.00	0.00	4.78	2.70	0.15	1.28
5	3.61	1.98	0.35	3.68	0.19	0.00	0.00	0.00	2.49	0.62	0.06	0.32
6	2.04	0.66	0.10	2.22	0.00	0.00	0.00	0.00	1.29	0.13	0.03	0.12
7	1.17	0.24	0.00	1.37	0.00	0.00	0.00	0.00	0.66	0.04	0.02	0.05
8	0.71	0.01	0.00	0.81	0.00	0.00	0.00	0.00	0.35	0.01	0.01	0.02

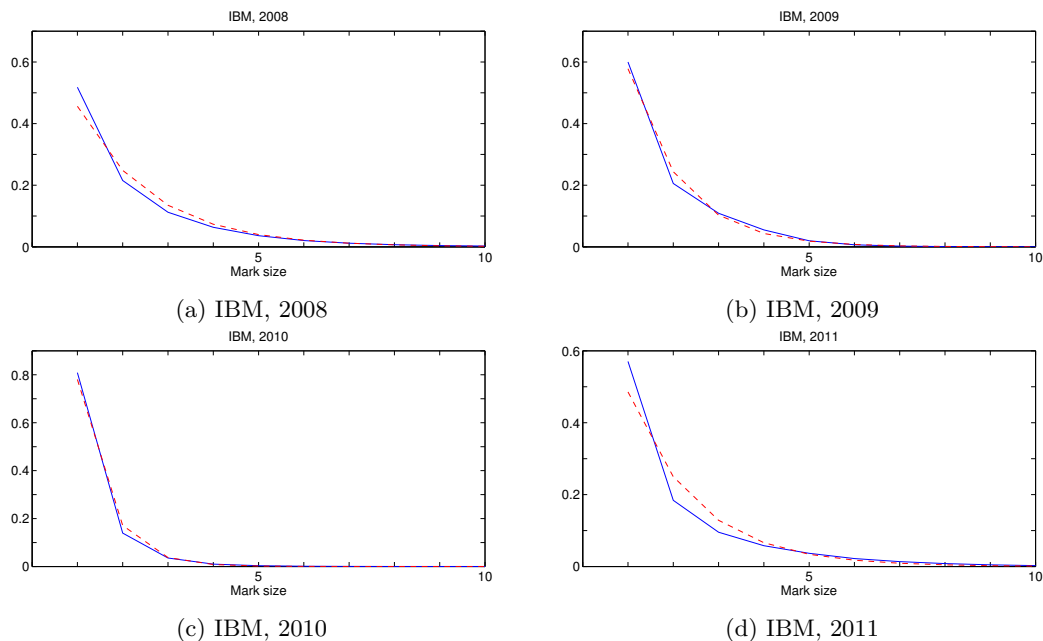(a) IBM, 2008

(b) IBM, 2009

(c) IBM, 2010

(d) IBM, 2011

Figure 3: Empirical unconditional distribution of mark

half tick size, $0.005.

In the original data, the time resolution of the record is one second. If more than one timestamps of the price changes are reported for one second, then the reported events are distributed over a one second interval to equidistant finer partitions.

4.2 Unconditional distribution of mark

Table 4 compares the percentage of the mark size of IBM, GE and CVX from 2008 to 2011, i.e., the unconditional distribution of mark sizes are reported in the table. IBM and CVX have a range of mark sizes over the years but GE's mark size distributions concentrates on the minimum mark size. This is because the price of IBM and CVX is relatively high (IBM is around $150 and CVX is around $100), whereas the price of GE is around $25. The unconditional distributions of the marks have exponentially decreasing shapes which are similar to the geometric distributions. The empirical distribution of the marks of IBM, 2010 and 2011, was compared with the geometric distributions in Figure 3. The solid lines are for the empirical distribution and the dashed lines are for the geometric distribution fitted by matching the first moments of the empirical and geometric distributions.

4.3 Mark size and intensity

This subsection examines the dependence between the mark size and the ground intensity, and the number of expected events over unit interval. The empirical evidence shows that the mark size and the current ground intensity are significantly related to each other. First, the empirical conditional

Table 5: Relationship between the mark size and the mean of proxy intensity (10 seconds), IBM

mark size	2011			2010		
	up	down	total	up	down	total
6	3.4784	3.6400	7.1184	6.4059	6.5618	12.9676
	(0.0145)	(0.0149)	(0.0292)	(0.2306)	(0.1171)	(0.1150)
5	3.1191	3.2461	6.3652	4.3289	4.5448	8.8738
	(0.0105)	(0.0106)	(0.0209)	(0.0890)	(0.0449)	(0.0448)
4	2.8206	2.9237	5.7443	3.4963	3.6881	7.1844
	(0.0078)	(0.0078)	(0.0155)	(0.0439)	(0.0221)	(0.0221)
3	2.5683	2.6551	5.2233	2.7119	2.8717	5.5836
	(0.0061)	(0.0060)	(0.0120)	(0.0182)	(0.0092)	(0.0092)
2	2.3051	2.3795	4.6846	2.1799	2.2892	4.4691
	(0.0042)	(0.0041)	(0.0083)	(0.0071)	(0.0036)	(0.0036)
1	2.3131	2.3501	4.6632	1.7682	1.8481	3.6163
	(0.0028)	(0.0028)	(0.0056)	(0.0024)	(0.0012)	(0.0012)
−1	2.2403	2.4140	4.6543	1.7488	1.8883	3.6371
	(0.0028)	(0.0028)	(0.0056)	(0.0024)	(0.0012)	(0.0012)
−2	2.3004	2.4376	4.7381	2.1655	2.3022	4.4687
	(0.0042)	(0.0042)	(0.0084)	(0.0072)	(0.0036)	(0.0036)
−3	2.5924	2.7167	5.3090	2.7264	2.8552	5.5816
	(0.0062)	(0.0062)	(0.0123)	(0.0185)	(0.0094)	(0.0093)
−4	2.8174	2.9552	5.7726	3.4806	3.6163	7.0969
	(0.0078)	(0.0079)	(0.0156)	(0.0432)	(0.0218)	(0.0217)
−5	3.1078	3.2399	6.3477	4.4487	4.5968	9.0454
	(0.0104)	(0.0106)	(0.0209)	(0.0902)	(0.0466)	(0.0469)
−6	3.4606	3.5845	7.0451	6.1019	6.2334	12.3352
	(0.0145)	(0.0149)	(0.0292)	(0.2141)	(0.1083)	(0.1074)

expectation of the intensities with given mark size, $\mathbb{E}[\lambda_{gi}(t)|k_i]$, were calculated. The proxy intensities are introduced because the ground intensities are unobservable. The proxy intensities for the up, down and total jumps are defined by the numbers of up, down and total jumps, respectively, over a fixed time period, which is ended just before the time of the jump, divided by the length of the period. The period for the proxy intensities was chosen as ten seconds. Mathematically, the up proxy intensity for the mark k_1, which takes place at time t is represented by $N_{g1}(t+\tau) - N_{g1}(t)$, where τ is the length of the period.

Table 5 presents the calculated sample mean and standard error of the proxy intensities for each mark size and for each year of 2010 and 2011, IBM. For example, for the mark size 6, there are 86,738 up jumps reported and the sample mean of the up proxy intensity is 3.4784 and the sample standard error is 0.0145. Note that the intensity of 3.4784 implies that the expected number of movements over unit time, which was set to 1 second, is approximately 3.4784. The table shows that the proxy intensities increase with increasing given mark size. The negative integers in the column of the mark size represent the down jump of the price. The proxy intensities were also calculated for five seconds time horizon. The results are similar to the previous case of 10 seconds time horizon so the results are not shown.

Second, Table 6 presents the relationship between the mark sizes and the inferred ground intensities with the linear impact function using the IBM tick data. Prior to calculating the inferred ground process, the parameters $\omega, \alpha_s, \alpha_c, \beta$, and η were estimated by maximizing $\log L_g$ defined in Eq. (11). The estimations were performed on a daily basis and the detailed estimation results will be demonstrated later. Subsequently, the inferred ground intensities were computed with the estimates of $\omega, \alpha_s, \alpha_c, \beta$, and η using the definition of the ground intensities in Eqs. (5) and (6). The sample mean and sample standard errors of the inferred ground intensities for each mark size is reported. Similarly with the case of the proxy intensities, the inferred ground intensities increase with increasing given mark size. This implies that if a large size of the mark is observed, it is probably based on the large ground intensities.

Third, Figure 4 illustrates the empirical expectations of the mark size conditionally upon given inferred ground intensities, $\mathbb{E}[k_i|\lambda_{gi}(t)]$, using the tick data of IBM, 2008-2011. For each year, the empirical conditional expectation with a given $\lambda_{gi} = n$ for an integer n was computed using the sample mean of the mark sizes, whose associated inferred ground process is falling into $(n-1, n]$. The conditional

Table 6: Relationship between the mark size and the mean of inferred ground intensity with the linear impact function, IBM

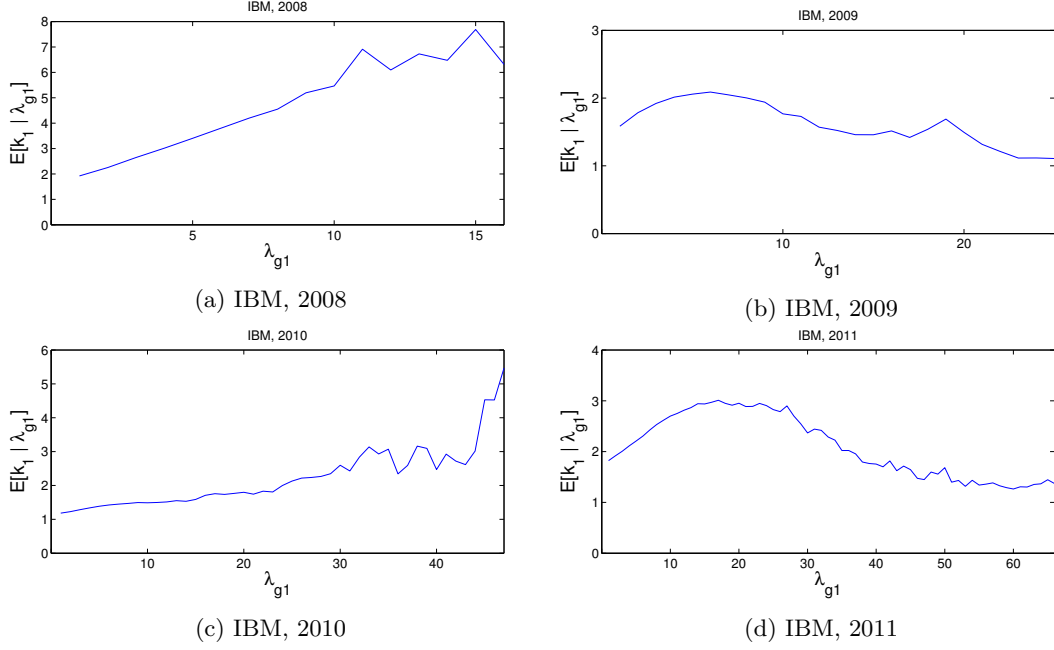| mark size	2011			2010		
	λ_{g1}	λ_{g2}	λ_g	λ_{g1}	λ_{g2}	λ_g
6	5.1666	5.1108	10.2774	8.1033	8.0330	16.1364
	(0.0222)	(0.0221)	(0.0442)	(0.1366)	(0.1370)	(0.2734)
5	4.7423	4.6834	9.4257	6.5586	6.4855	13.0443
	(0.0165)	(0.0164)	(0.0328)	(0.0621)	(0.0620)	(0.1240)
4	4.4092	4.3565	8.7656	5.4497	5.3975	10.8472
	(0.0126)	(0.0125)	(0.0251)	(0.0338)	(0.0337)	(0.0674)
3	4.0264	3.9842	8.0106	4.2182	4.1806	8.3988
	(0.0094)	(0.0094)	(0.0188)	(0.0145)	(0.0145)	(0.0290)
2	3.5870	3.5505	7.1375	3.4705	3.4476	6.9181
	(0.0066)	(0.0066)	(0.0131)	(0.0064)	(0.0064)	(0.0128)
1	3.5924	3.5580	7.1504	2.5428	2.5321	5.0749
	(0.0057)	(0.0057)	(0.0114)	(0.0021)	(0.0021)	(0.0041)
−1	3.5568	3.5942	7.1509	2.5702	2.5836	5.1538
	(0.0057)	(0.0057)	(0.0113)	(0.0021)	(0.0021)	(0.0041)
−2	3.5858	3.6270	7.2128	3.4730	3.4978	6.9708
	(0.0066)	(0.0066)	(0.0132)	(0.0064)	(0.0064)	(0.0129)
−3	4.0451	4.0927	8.1378	4.2369	4.2781	8.5150
	(0.0096)	(0.0096)	(0.0192)	(0.0148)	(0.0148)	(0.0296)
−4	4.3580	4.4177	8.7757	5.3775	5.4380	10.8154
	(0.0126)	(0.0127)	(0.0253)	(0.0333)	(0.0334)	(0.0667)
−5	4.6872	4.7587	9.4459	6.4356	6.5047	12.9402
	(0.0163)	(0.0164)	(0.0327)	(0.0624)	(0.0626)	(0.1250)
−6	5.0360	5.0884	10.1244	7.9426	8.0193	15.9620
	(0.0216)	(0.0217)	(0.0433)	(0.1343)	(0.1344)	(0.2685)

Figure 4: Conditional expectation of k_1 on λ_{g1}, IBM, 2008-2011

expectations of the marks were plotted where the total observed numbers of the mark were larger than 100 for each year, i.e., the samples with a small number of observations are dropped out. In the figure, the ground intensities vary by more as the years pass, suggesting that the overall number of activities increases. Most of intensities were less than 15 in 2008 but the inferred intensity in 2011 were more widely distributed as one can see that the large portion of the observed intensities were larger than 15.

The shape of the conditional expectation of the mark changes over time. The conditional expectation tends to increase with increasing intensity in 2008 and 2010. In 2009 and 2011, the conditional expectations showed humped shapes. Figure 5 shows the empirical conditional expectations of the mark given ground intensity computed monthly basis from January to June, 2011, of IBM. In the monthly basis empirical conditional expectation, irregular patterns were observed over time. The changing shape of the conditional distribution of marks over time is the reason why the mark distribution is not specified, and the estimation was performed in a non-parametric manner for the part of the mark distribution.

4.4 Estimation result

Table 7 lists the likelihood estimation results of the marked Hawkes model with the tick data of IBM, January 2011, where $\log L_g$ of Eq. (11) is maximized. The numerically computed standard errors are reported in the parentheses. The estimations were performed on a daily basis, i.e., the estimates were recalculated in every business day. The behaviors of μ, α_s, α_c, and β in Figure 6 are similar to those estimated in the simple Hawkes model, see Lee and Seo (2014).

The dynamics of η are illustrated in Figure 8 where η was estimated around 0.2 in general from 2008 to 2011. The slope parameter for the impact function, η, is positive in general, which means that large mark tends to have a large impact for the future intensities. On the other hand, η is less than 1 and this implies that the impact of mark size 2 is generally less than the total impact of the two consecutive unit size jumps that occur over a very short time interval. Note that few negative η are also observed. Figure 9 shows the dynamics of η estimated from CVX. The overall behaviors of η of IBM and CVX are similar but the η of CVX was more volatile.

Figure 10 compares the Hawkes volatility computed by Remark 2 and TSRV of IBM, 2008-2011. The trends of the Hawkes volatility and TSRV are similar but the Hawkes volatility is generally larger than the TSRV, especially when the volatility is high. This tendency was also found in the simple symmetric Hawkes model. This discrepancy might be because of the restriction that the parameter settings need be symmetric and Markovian. On the other hand, the precise reason for this is unclear at this point. Note that in the previous simulation study, the Hawkes volatility and TSRV converges.
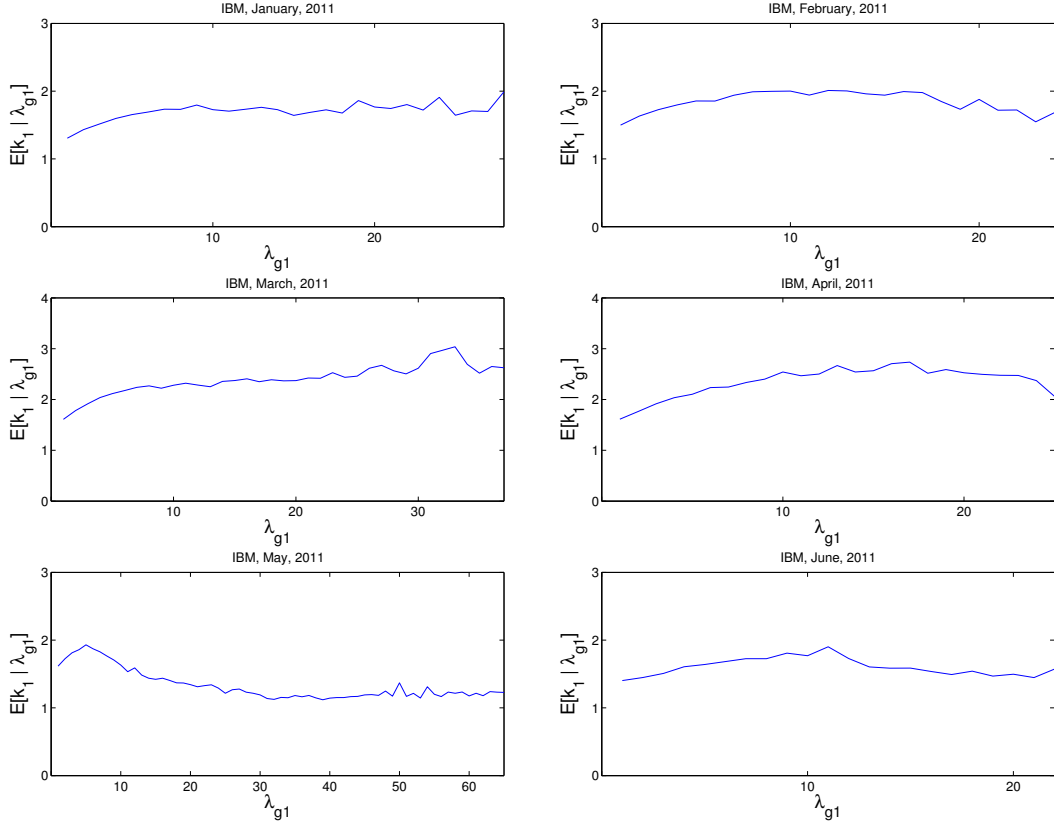
17

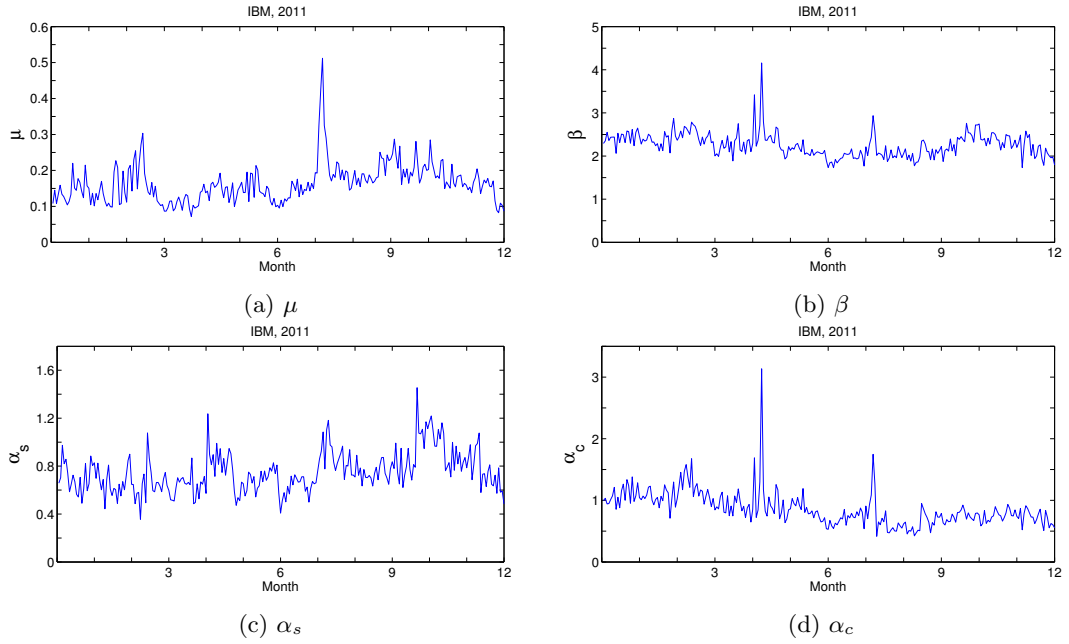Figure 5: Conditional expectation of k_1 on λ_{g1}, IBM, 2011

Figure 6: Marked Hawkes estimation result, IBM, 2011

Table 7: Estimates of IBM, 2011 with linear impact function

date	μ	α_s	α_c	β	η	$\log L_g$
0103	0.1080	0.6577	0.9956	2.2921	0.1241	-17185.5
	(0.0021)	(0.0175)	(0.0218)	(0.0339)	(0.0187)	
0104	0.1450	0.7033	1.0334	2.3527	0.2266	-17371.5
	(0.0026)	(0.0157)	(0.0188)	(0.0272)	(0.0233)	
0105	0.1079	0.9736	0.9414	2.550	0.1654	-13160.3
	(0.0021)	(0.0207)	(0.0203)	(0.0334)	(0.0180)	
0106	0.1335	0.8198	0.9475	2.3615	0.1357	-16603.4
	(0.0025)	(0.0173)	(0.0191)	(0.0296)	(0.0171)	
0107	0.1588	0.8574	1.0145	2.435	0.1560	-15956.4
	(0.0029)	(0.0164)	(0.0184)	(0.0284)	(0.0176)	
0110	0.1338	0.7423	1.0724	2.4388	0.1927	-16141.3
	(0.0025)	(0.0160)	(0.0191)	(0.0266)	(0.0163)	
0111	0.1271	0.5855	1.2108	2.4403	0.1570	-16923.1
	(0.0024)	(0.0159)	(0.0223)	(0.0314)	(0.0155)	
0112	0.1160	0.6517	0.8552	2.0639	0.3561	-19492.9
	(0.0023)	(0.0158)	(0.0185)	(0.0307)	(0.0481)	
0113	0.1042	0.7245	1.0502	2.5284	0.2372	-15508.4
	(0.0020)	(0.0192)	(0.0240)	(0.0369)	(0.0261)	
0114	0.1138	0.6702	0.8798	2.3142	0.2380	-16589.3
	(0.0022)	(0.0175)	(0.0202)	(0.0341)	(0.0183)	
0118	0.1330	0.5642	1.1548	2.5082	0.1651	-16374.7
	(0.0024)	(0.0169)	(0.0239)	(0.0354)	(0.0147)	
0119	0.2198	0.5423	1.2631	2.4964	0.1323	-11223.0
	(0.0036)	(0.0133)	(0.0207)	(0.0255)	(0.0093)	
0120	0.1509	0.7060	1.0017	2.3114	0.1824	-15709.4
	(0.0028)	(0.0158)	(0.0211)	(0.0332)	(0.0154)	
0121	0.1447	0.4901	1.3356	2.5806	0.1545	-16524.6
	(0.0026)	(0.0152)	(0.0247)	(0.0333)	(0.0132)	
0124	0.1771	0.6095	1.2424	2.5658	0.1649	-14711.6
	(0.0030)	(0.0151)	(0.0205)	(0.0290)	(0.0132)	
0125	0.1669	0.8214	0.9528	2.2982	0.1954	-13602.1
	(0.0030)	(0.0147)	(0.0167)	(0.0244)	(0.0170)	
0126	0.1421	0.5939	1.3809	2.6146	0.1217	-9239.8
	(0.0026)	(0.0154)	(0.0231)	(0.0295)	(0.0100)	

Table 8: Estimation result of fully characterized self and mutually excited Hawkes process, IBM, 2011

	μ_1	μ_2	α_{11}	α_{22}	α_{12}	α_{21}	β_{11}	β_{22}	β_{12}	β_{21}
IBM, 2011										
mean	0.1462	0.1470	0.5404	0.5621	0.6443	0.6611	1.5499	1.5793	1.7538	1.8056
std.	0.0543	0.0545	0.1962	0.2361	0.2144	0.2717	0.3845	0.4192	0.3373	0.4282
MAPE	0.1436		0.3924		0.3179		0.2881		0.2247	

The empirical studies suggest that the parameter restrictions for symmetry do not perfectly meet with the real data. Figure 7 presents the dynamics of all parameters of the fully characterized Hawkes model of Subsection 3.2 for IBM, 2011. The result shows that for each parameter pair (μ_1, μ_2), $(\alpha_{11}, \alpha_{22})$, $(\alpha_{12}, \alpha_{21})$, (β_{11}, β_{22}) and (β_{12}, β_{21}), a similar trend is observed over time but those were not exactly the same as each other. The summary statistics in Table 8 show that the parameter pair has a similar mean but the mean absolute percentage error (MAPE) also shows a difference between the parameters. The row 'MAPE' presents the error between two adjacent parameters. In addition, β_{11} and β_{12} are different, even in the mean. A similar observation is found in the simple Hawkes approach.

Interestingly, when the stock market is in a highly volatile state, the slope for the impact function, η is estimated to be relatively close to zero. For example, September 29, 2008, at the beginning of the financial crisis, the reported η of IBM was around 0.02, which was much smaller than the annual average, 0.16, when the market is very unstable with a TSRV of 0.8766 and a Hawkes volatility 1.4530. In the May 6, 2010 Flash Crash, the estimated η of IBM was around 0.01 (annual average of $\eta = 0.23$) when the TSRV was 0.9656 and the Hawkes volatility was 1.3264. Because of the statement of Federal Reserve, the stock market was highly volatile at August 9, 2011, and the η of IBM was around 0.01 which is much smaller than the annual average 0.11. CVX also showed a similar pattern. The estimated η of CVX for the above cases are around 0.04, 0.01, and 0.07, respectively, whereas the annual averages in 2008, 2010 and 2011 were 0.14, 0.41 and 0.24, respectively. In a highly volatile market, a much larger number of large size marks are observed than in a stable market, and with those large size marks, the linear relationship between the mark size and the future impact is weak.

In the estimation of the Hawkes models, the data of all arrival times over the sample period are used. Even with a ten minute interval, usually more than a thousand records of arrival times are available, which is sufficient to provide a reliable result for the estimation in the aspect of the sample size and hence adequate intraday analyses are possible. The left of Figure 11 presents the intraday variation of the volatility of IBM on May 6, 2010, the day of the Flash Crash. The estimation was performed on a ten minute basis from 10:00 to 15:30 and the cumulative volatility is plotted in the figure. An abrupt increase in volatility is observed between 14:30 and 15:00 when the stock market crashed and the volatility was stabilized after 15:00. The right of the figure plots the intraday U-shape pattern of the volatility of IBM on an ordinary day of August 9, 2011.

In the volatility estimation perspective, the volatility formula under the i.i.d. mark in Corollary 3 and the volatility formula without assuming the i.i.d. property are similar. The empirical result also showed that the two formulas have similar values over time, as shown in Figure 12. Therefore, for practical purposes, such as volatility computation, it may be sufficient to introduce an i.i.d. mark distribution.

5 Concluding remark

A marked Hawkes model was developed for price tick dynamics in equity markets. A linear impact function was employed to describe the future effects of price jumps. A specific distribution for the jump size was not assumed but the empirical distribution was used for the estimation. This model is not limited to the independent mark since the empirical studies showed that the jump size depends on the ground intensities. The volatility formula was derived based on stochastic calculus and statistical methods and the simulation studies showed that the Hawkes volatility and realized volatility are similar in the symmetric cases and the Hawkes volatility has less standard error. On the other hand, there are biases when the underlying path is not symmetrical or the parameters vary with time.

A significant positive linear impact was observed, approximately 0.2, and various types, linear or humped shape, of the conditional mean structure of mark size in the empirical studies based on the equity prices reported in NYSE. The Hawkes model is useful for estimating the intraday volatility particularly when the volatility is time varying. The U shape seasonality of the changing volatility was observed
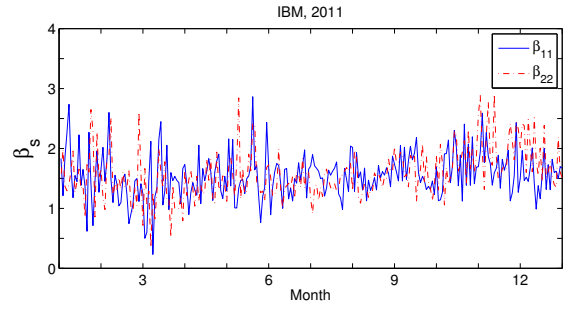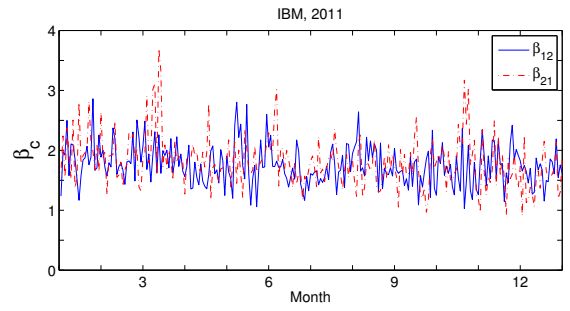
(a) μ_1 and μ_2

(b) α_{11} and α_{22}

(c) α_{12} and α_{21}

(d) β_{11} and β_{22}
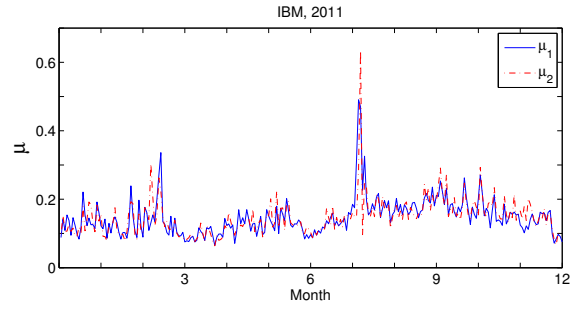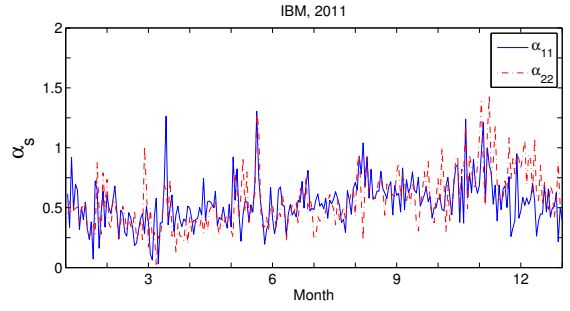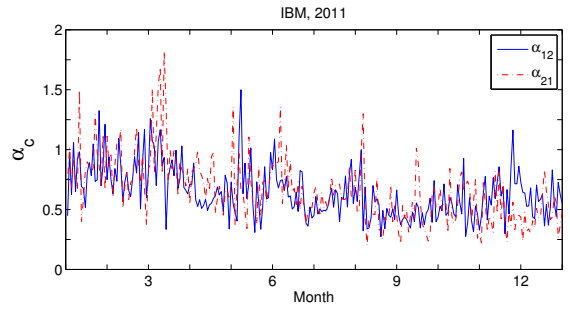
(e) β_{12} and β_{21}

Figure 7: Estimation result with the fully characterized marked Hawkes, IBM, 2011
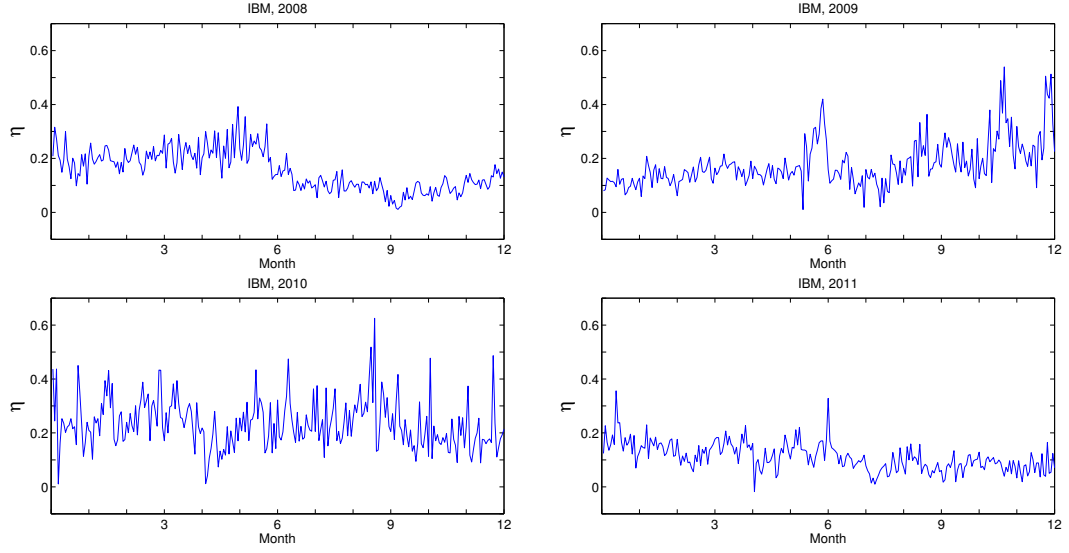
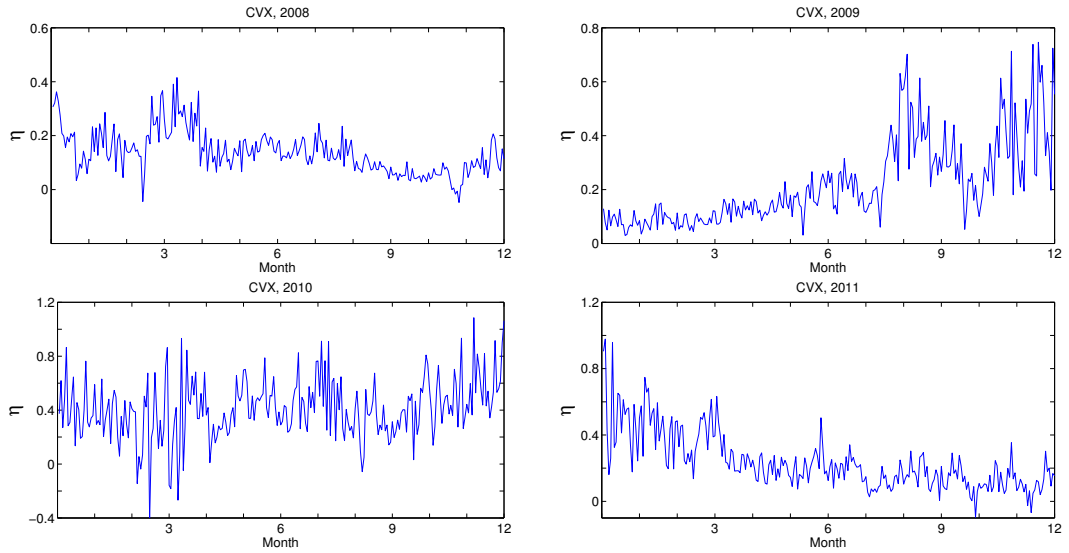Figure 8: Estimation results of η, IBM, 2008-2011

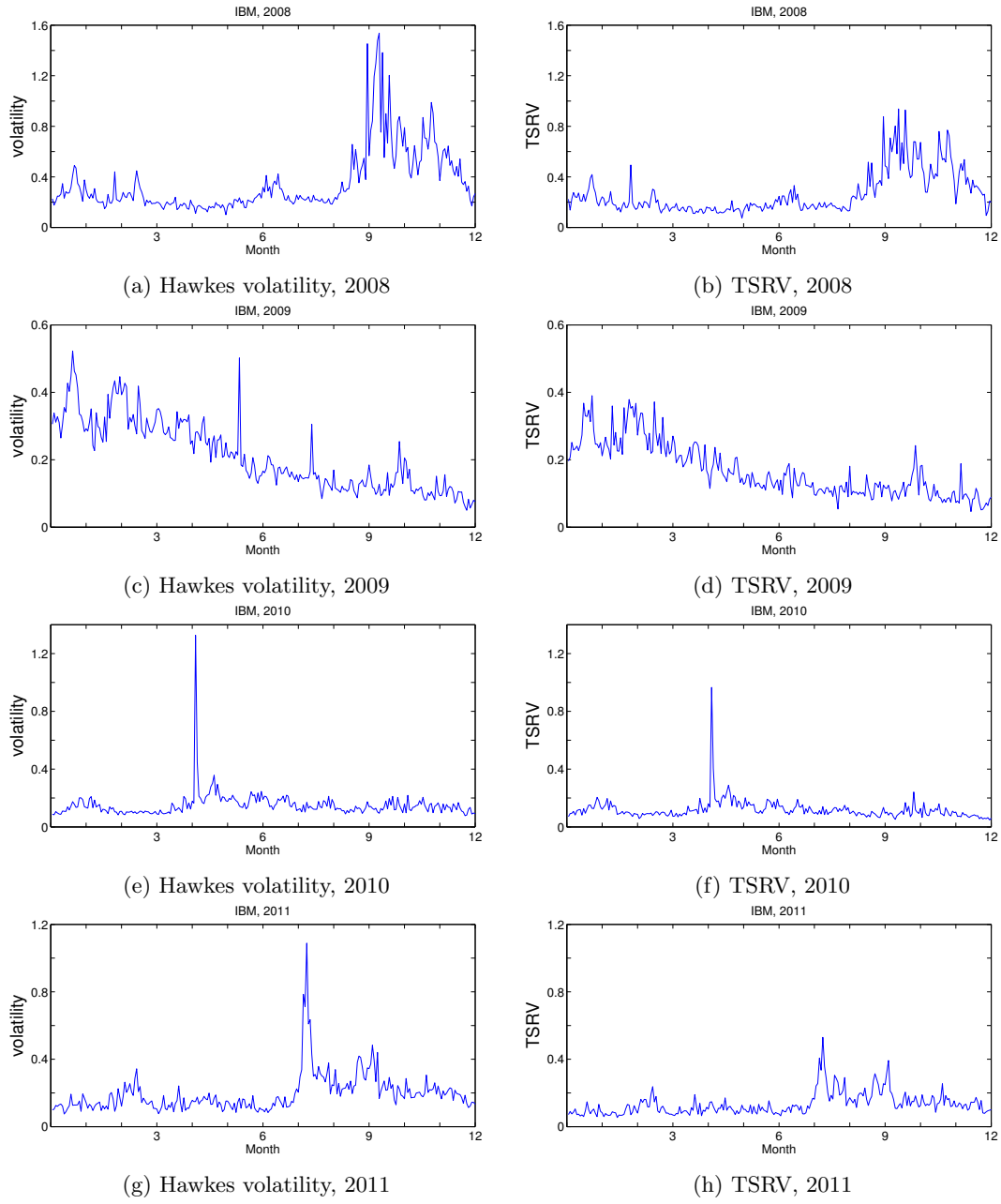Figure 9: Estimation results of η, CVX, 2008-2011

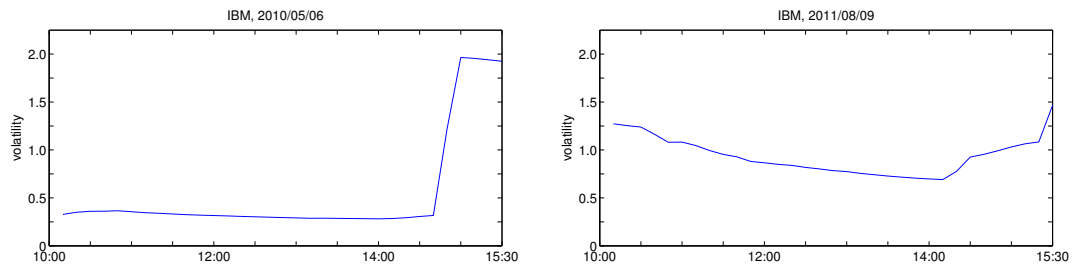Figure 10: Estimation results of the volatility, IBM, 2008-2011

Figure 11: Cumulative intraday volatitliy estimated under the marked Hawkes model with every ten minutes update
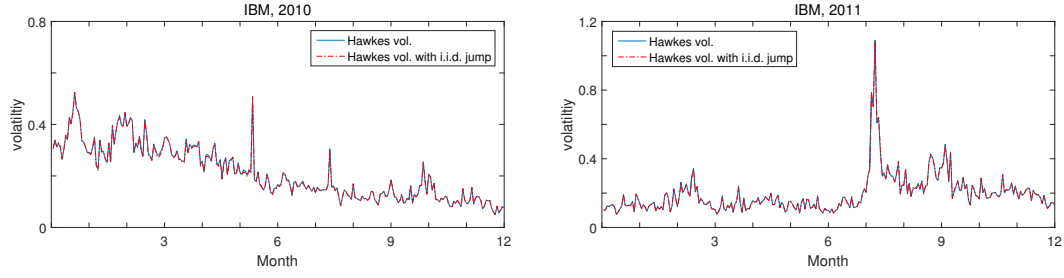
Figure 12: Comparison between the daily volatility under the i.i.d. assumption and without i.i.d. assumption

and the interesting example of Flash Crash was examined. As discussed in the simulation and empirical studies, the discrepancy between the Hawkes volatility and realized volatility and the biases from the sample volatility will be an important subject for a future study. In the presence of the asymmetry in the price dynamics and the time varying parameters, a more robust estimation method is required for a more exact volatility computation.

One of the financial application of the marked Hawkes model is intraday volatility measurements. There is a discrepancy between the realized volatility and marked Hawkes volatility but the trend and the overall movements of two volatilities are consistent. In addition, in the marked Hawkes volatility estimation, the all available events reported in the exchange is used and hence even using the data in a relatively small interval, one can compute the Hawkes volatility and recognize the intraday changes of the volatility movement. This feature will help the traders, portfolio managers or algorithmic machines for the decision making by checking the changes in the intraday volatility trends of concerned assets.

The return model of the mark size will be considered for future work. The distribution of the mark size depending on the current price of the underlying asset and it is worthwhile to examine the relationship between the mark size measured under the return process and future intensities. In addition, it would be interesting to compare the performance of the marked Hawkes model with the ACD-GARCH model with modification of the future intensity as a function the marks.

References

Abergel, F. and Jedidi, A. (2013). A mathematical approach to order book modeling. *International Journal of Theoretical and Applied Finance*, 16:1350025.

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial studies*, 18:351–416.

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, 160:160–175.

Alsayed, H. and McGroarty, F. (2014). Ultra-high-frequency algorithmic arbitrage across international index futures. *Journal of Forecasting*, 33:391–408.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71:579–625.

Bacry, E., Dayri, K., and Muzy, J.-F. (2012). Non-parametric kernel estimation for symmetric Hawkes processes. application to high frequency financial data. *The European Physical Journal B-Condensed Matter and Complex Systems*, 85:1–12.

Bacry, E., Delattre, S., Hoffmann, M., and Muzy, J.-F. (2013). Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13:65–77.

Bacry, E., Jaisson, T., and Muzy, J. (2016). Estimation of slowly decreasing hawkes kernels: application to high-frequency order book dynamics. *Quantitative Finance*, 16:1179–1201.

Bacry, E. and Muzy, J.-F. (2014). Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 14:1147–1166.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31:307–327.

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

Chaboud, A. P., Chiquoine, B., Hjalmarsson, E., and Vega, C. (2014). Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance*, 69:2045–2084.

Cont, R. and De Larrard, A. (2013). Price dynamics in a Markovian limit order market. *SIAM Journal on Financial Mathematics*, 4:1–25.

Cont, R., Stoikov, S., and Talreja, R. (2010). A stochastic model for order book dynamics. *Operations research*, 58:549–563.

Da Fonseca, J. and Zaatour, R. (2014). Hawkes process: Fast calibration, application to trade clustering, and diffusive limit. *Journal of Futures Markets*, 34:548–579.

Da Fonseca, J. and Zaatour, R. (2015). Clustering and mean reversion in a Hawkes microstructure model. *Journal of Futures Markets*, 35:813–838.

Da Fonseca, J. and Zaatour, R. (2016). Correlation and leadlag relationships in a hawkes microstructure model. *Journal of Futures Markets*.

Daley, D. J. and Vere-Jones, D. (2003). *An introduction to the theory of point processes*, volume 1. Springer.

Fodra, P. and Pham, H. (2015). Semi-markov model for market microstructure. *Applied Mathematical Finance*, 22:261–295.

Foucault, T. (2012). *Algorithmic Trading: Issues and Preliminary Evidence*, pages 1–40. John Wiley & Sons Ltd.

Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business and Economic Statistics*, 24:127–161.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58:83–90.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6:327–343.

Hewlett, P. (2006). Clustering of order arrivals, price impact and trade path optimisation. In *Workshop on Financial Modeling with Jump processes, Ecole Polytechnique*.

Hoffmann, P. (2014). A dynamic limit order market with fast and slow traders. *Journal of Financial Economics*, 113:156–169.

Huth, N. and Abergel, F. (2014). High frequency lead/lag relationships empirical facts. *Journal of Empirical Finance*, 26:41 – 58.

Large, J. (2007). Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10:1–25.

Lee, K. and Seo, B. K. (2014). Modeling microstructure price dynamics with symmetric Hawkes and diffusion model using ultra-high-frequency stock data. *Working paper*, 1:1–1.

Lo, A. W., MacKinlay, A. C., and Zhang, J. (2002). Econometric models of limit-order executions. *Journal of Financial Economics*, 65:31–71.

Malo, P. and Pennanen, T. (2012). Reduced form modeling of limit order markets. *Quantitative Finance*, 12:1025–1036.

Roşu, I. (2009). A dynamic model of the limit order book. *Review of Financial Studies*, 22:4601–4641.

Russell, J. R. and Engle, R. F. (2005). A discrete-state continuous-time model of financial transactions prices and times. *Journal of Business & Economic Statistics*, 23:166–180.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005). A tale of two time scales. *Journal of the American Statistical Association*, 100:1394–1411.

Zheng, B., Roueff, F., and Abergel, F. (2014). Modelling bid and ask prices using constrained Hawkes processes: Ergodicity and scaling limit. *SIAM Journal on Financial Mathematics*, 5:99–136.

A Proof of Theorem 1

We apply the stochastic integration theory to derive the second moment property of the two dimensional symmetric marked Hawkes process. The quadratic (co)variation process of the semimartingales X and Y for $t \geq 0$ is defined as

$$[X, Y]_t = X_t Y_t - \int_0^t X_{u-} \mathrm{d}Y_u - \int_0^t Y_{u-} \mathrm{d}X_u.$$

For the quadratic pure jump processes X and Y, such as $\lambda_{gi}(t)$, $N_{gi}(t)$ and $N_i(t)$ in the proposed model

$$[X]_t := [X, X]_t = X_0^2 + \sum_{0 \leq s < t} (\Delta X_s)^2 \tag{19}$$

and

$$[X, Y]_t = X_0 Y_0 + \sum_{0 \leq s < t} (\Delta X_s \Delta Y_s).$$

A.1 Step 1

This subsection derives the unconditional expectations of $\lambda_{gi}^2(t)$ and $\lambda_{gi}(t)\lambda_{gj}(t)$. By the definition of the quadratic variation process, we have

$$\mathbb{E}[\lambda_{g1}^2(t)] = \mathbb{E}[[\lambda_{g1}]_t] + 2\mathbb{E}\left[\int_0^t \lambda_{g1}(u)\mathrm{d}\lambda_{g1}(u)\right].$$

When the ground intensity λ_{gi} is used as an integrator, we consider λ_{gi} as the right continuous modification. The stochastic integration is indeed a path-by-path Lebesgue-Stieltjes integral. The integration part of the r.h.s. of the above equation is represented by

$$\int_0^t \lambda_{g1}(u)\mathrm{d}\lambda_{g1}(u) = \int_0^t \lambda_{g1}(u)\beta(\mu - \lambda_{g1}(u))\mathrm{d}u$$
$$+ \int_{(0,t)\times\mathbb{Z}^+} q_s\beta g(k_1)\lambda_{g1}(u)N_1(\mathrm{d}u\times\mathrm{d}k_1) + \int_{(0,t)\times\mathbb{Z}^+} q_c\beta g(k_2)\lambda_{g1}(u)N_2(\mathrm{d}u\times\mathrm{d}k_2)$$

or heuristically,

$$\mathrm{d}\lambda_{g1}(u) = \beta(\mu - \lambda_{g1}(u))\mathrm{d}u + \int_{(u,u+\mathrm{d}u)\times\mathbb{Z}^+} q_s\beta g(k_1)N_1(\mathrm{d}u\times\mathrm{d}k_1) + \int_{(u,u+\mathrm{d}u)\times\mathbb{Z}^+} q_c\beta g(k_2)N_2(\mathrm{d}u\times\mathrm{d}k_2)$$

$$= \beta(\mu - \lambda_{g1}(u))\mathrm{d}u + \int_{(u,u+\mathrm{d}u)\times\mathbb{Z}^+} \alpha_s(1 + (k_1 - 1)\eta)N_1(\mathrm{d}u\times\mathrm{d}k_1)$$

$$+ \int_{(u,u+\mathrm{d}u)\times\mathbb{Z}^+} \alpha_c(1 + (k_2 - 1)\eta)N_2(\mathrm{d}u\times\mathrm{d}k_2).$$

Since the jump size of λ_{g1} is represented by $\alpha_s(1 + (k_1 - 1)\eta)$ and $\alpha_c(1 + (k_2 - 1)\eta)$, using Eq. (19), we have

$$\mathbb{E}[[\lambda_{g1}]_t] = \mathbb{E}[\lambda_{g1}^2(0)] + \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} \alpha_s^2(1 + (k_1 - 1)\eta)^2 N_1(\mathrm{d}u\times\mathrm{d}k_1)\right]$$

$$+ \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} \alpha_c^2(1 + (k_2 - 1)\eta)^2 N_2(\mathrm{d}u\times\mathrm{d}k_2)\right]$$

$$= \mathbb{E}[\lambda_{g1}^2(t)] + \int_0^t \alpha_s^2\left(1 + \mathbb{E}[2(k_1 - 1)\eta\lambda_{g1}(u)] + \mathbb{E}[(k_1^2 - 2k_1 + 1)\eta^2\lambda_{g1}(u)]\right)\mathrm{d}u$$

$$+ \int_0^t \alpha_c^2\left(1 + \mathbb{E}[2(k_2 - 1)\eta\lambda_{g2}(u)] + \mathbb{E}[(k_2^2 - 2k_2 + 1)\eta^2\lambda_{g2}(u)]\right)\mathrm{d}u$$

$$= \mathbb{E}[\lambda_{g1}^2(t)] + (\alpha_s^2 + \alpha_c^2)\{1 + 2(K_{1\lambda_{g1}} - 1)\eta + (K_{1\lambda_{g1}}^{(2)} - 2K_{1\lambda_{g1}} + 1)\eta^2\}\mathbb{E}[\lambda_{g1}(t)]t$$

$$= \mathbb{E}[\lambda_{g1}^2(t)] + (\alpha_s^2 + \alpha_c^2)\bar{\bar{K}}\mathbb{E}[\lambda_{g1}(t)]t$$

where the stationarity of λ_{gi} and the symmetry are used. In addition,

$$\mathbb{E}\left[\int_0^t \lambda_{g1}(u)\mathrm{d}\lambda_{g1}(u)\right] = \mathbb{E}\left[\int_0^t \lambda_{g1}(u)\beta(\mu - \lambda_{g1}(u))\mathrm{d}u\right]$$

$$+ \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} q_s\beta g(k_1)\lambda_{g1}(u)N_1(\mathrm{d}u\times\mathrm{d}k_1) + \int_{(0,t)\times\mathbb{Z}^+} q_c\beta g(k_2)\lambda_{g1}(u)N_2(\mathrm{d}u\times\mathrm{d}k_2)\right]$$

$$= (\beta\mu\mathbb{E}[\lambda_{g1}(t)] + [\alpha_s\{1 + (K_{1\lambda_{g1}^2} - 1)\eta\} - \beta]\mathbb{E}[\lambda_{g1}^2(t)] + \alpha_c\{1 + (K_{2\lambda_{g1}\lambda_{g2}} - 1)\eta\}\mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)])t$$

$$= (\beta\mu\mathbb{E}[\lambda_{g1}(t)] + (\breve{\alpha}_s - \beta)\mathbb{E}[\lambda_{g1}^2(t)] + \tilde{\alpha}_c\mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)])t$$

Similarly,

$$\mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)] = \mathbb{E}[[\lambda_{g1}, \lambda_{g2}]_t] + \mathbb{E}\left[\int_0^t \lambda_{g1}(u)\mathrm{d}\lambda_{g2}(u)\right] + \mathbb{E}\left[\int_0^t \lambda_{g2}(u)\mathrm{d}\lambda_{g1}(u)\right],$$

and, for the covariation process $[\lambda_{g1}, \lambda_{g2}]_t$, we have

$$\mathbb{E}[[\lambda_{g1}, \lambda_{g2}]_t] = \mathbb{E}[\lambda_{g1}(0)\lambda_{g2}(0)] + \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} \alpha_s\alpha_c(1+(k_1-1)\eta)^2 N_1(\mathrm{d}u\times\mathrm{d}k_1)\right]$$

$$+ \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} \alpha_s\alpha_c(1+(k_2-1)\eta)^2 N_2(\mathrm{d}u\times\mathrm{d}k_2)\right]$$

$$= \mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)] + \int_0^t \alpha_s\alpha_c\left(1+\mathbb{E}[2(k_1-1)\eta\lambda_{g1}(u)] + \mathbb{E}[(k_1^2-2k_1+1)\eta^2\lambda_{g1}(u)]\right)\mathrm{d}u$$

$$+ \int_0^t \alpha_s\alpha_c\left(1+\mathbb{E}[2(k_2-1)\eta\lambda_{g2}(u)] + \mathbb{E}[(k_2^2-2k_2+1)\eta^2\lambda_{g2}(u)]\right)\mathrm{d}u$$

$$= \mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)] + \alpha_s\alpha_c\left\{1+2(K_{1\lambda_{g1}}-1)\eta + (K_{1\lambda_{g1}}^{(2)}-2K_{1\lambda_{g1}}+1)\eta^2\right\}\mathbb{E}[\lambda_{g1}(t)]t$$

$$+ \alpha_s\alpha_c\left\{1+2(K_{2\lambda_{g2}}-1)\eta + (K_{2\lambda_{g2}}^{(2)}-2K_{2\lambda_{g2}}+1)\eta^2\right\}\mathbb{E}[\lambda_{g2}(t)]t$$

$$= \mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)] + 2\alpha_s\alpha_c\{1+2(K_{1\lambda_{g1}}-1)\eta + (K_{1\lambda_{g1}}^{(2)}-2K_{1\lambda_{g1}}+1)\eta^2\}\mathbb{E}[\lambda_{g1}(t)]t$$

$$= \mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)] + 2\alpha_s\alpha_c\bar{\bar{K}}\mathbb{E}[\lambda_{g1}(t)]t$$

and

$$\mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} \lambda_{g1}(u)\mathrm{d}\lambda_{g2}(u)\right] = \mathbb{E}\left[\int_0^t \lambda_{g1}(u)\beta(\mu-\lambda_{g2}(u))\mathrm{d}u\right]$$

$$+ \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} q_c\beta g(k_1)\lambda_{g1}(u)N_1(\mathrm{d}u\times\mathrm{d}k_1) + \int_{(0,t)\times\mathbb{Z}^+} q_s\beta g(k_2)\lambda_{g1}(u)N_2(\mathrm{d}u\times\mathrm{d}k_2)\right]$$

$$= (\beta\mu\mathbb{E}[\lambda_{g1}(t)] + \alpha_c\{1+(K_{1\lambda_{g1}^2}-1)\eta\}\mathbb{E}[\lambda_{g1}^2(t)] + [\alpha_s\{1+(K_{2\lambda_{g1}\lambda_{g2}}-1)\eta\}-\beta]\mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)])t$$

$$= (\beta\mu\mathbb{E}[\lambda_{g1}(t)] + \breve{\alpha}_c\mathbb{E}[\lambda_{g1}^2(t)] + (\tilde{\alpha}_s-\beta)\mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)])t.$$

Combining the above results, we have

$$0 = (\alpha_s^2+\alpha_c^2)\bar{\bar{K}}\mathbb{E}[\lambda_{g1}(t)]t + 2\left\{\beta\mu\mathbb{E}[\lambda_{g1}(t)] + (\breve{\alpha}_s-\beta)\mathbb{E}[\lambda_{g1}^2(t)] + \tilde{\alpha}_c\mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)]\right\}t$$

$$0 = 2\alpha_s\alpha_c\bar{\bar{K}}\mathbb{E}[\lambda_{g1}(t)]t + 2\left\{\beta\mu\mathbb{E}[\lambda_{g1}(t)] + \breve{\alpha}_c\mathbb{E}[\lambda_{g1}^2(t)] + (\tilde{\alpha}_s-\beta)\mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)]\right\}t$$

and, in a matrix form,

$$\begin{bmatrix} \mathbb{E}[\lambda_{g1}^2(t)] \\ \mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)] \end{bmatrix} = -\frac{1}{2}\mathbb{E}[\lambda_{g1}(t)]\mathbf{M}^{-1}\begin{bmatrix} (\alpha_s^2+\alpha_c^2)\bar{\bar{K}}+2\beta\mu \\ 2(\alpha_s\alpha_c\bar{\bar{K}}+\beta\mu) \end{bmatrix}$$

where

$$\mathbf{M} = \begin{bmatrix} \breve{\alpha}_s-\beta & \tilde{\alpha}_c \\ \breve{\alpha}_c & \tilde{\alpha}_s-\beta \end{bmatrix}.$$

A.2 Step 2

In this step, the goal is to calculate $\mathbb{E}[\lambda_{g1}(t)N_1(t)]$ and $\mathbb{E}[\lambda_{g1}(t)N_2(t)]$. We have

$$\mathbb{E}[\lambda_{g1}(t)N_1(t)] = \mathbb{E}[[\lambda_{g1}, N_1]_t] + \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} \lambda_{g1}(u)k_1 N_1(\mathrm{d}u\times\mathrm{d}k_1)\right] + \mathbb{E}\left[\int_0^t N_1(u-)\mathrm{d}\lambda_{g1}(u)\right].$$

For each component in the r.h.s.,

$$\mathbb{E}[[\lambda_{g1}, N_1]_t] = \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} \alpha_s(1+(k_1-1)\eta)k_1 N_1(\mathrm{d}u\times\mathrm{d}k_1)\right]$$

$$= \alpha_s(K_{1\lambda_{g1}} + (K_{1\lambda_{g1}}^{(2)}-K_{1\lambda_{g1}})\eta)\mathbb{E}[\lambda_{g1}(t)]t$$

$$= \alpha_s\bar{K}\mathbb{E}[\lambda_{g1}(t)]t$$

28

and
$$\mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+}\lambda_{g1}(u)k_1 N_1(\mathrm{d}u\times\mathrm{d}k_1)\right]=\int_0^t\mathbb{E}[k_1\lambda_{g1}^2(u)]\mathrm{d}u=K_{1\lambda_{g1}^2}\mathbb{E}[\lambda_{g1}^2(t)]t$$

and
$$\mathbb{E}\left[\int_0^t N_1(u-)\mathrm{d}\lambda_{g1}(u)\right]=\mathbb{E}\left[\int_0^t N_1(u)\beta(\mu-\lambda_{g1}(u))\mathrm{d}u\right]$$
$$+\mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+}N_1(u-)\alpha_s(1+(k_1-1)\eta)N_1(\mathrm{d}u\times\mathrm{d}k_1)\right]$$
$$+\mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+}N_1(u-)\alpha_c(1+(k_2-1)\eta)N_2(\mathrm{d}u\times\mathrm{d}k_2)\right]$$
$$=\int_0^t\{\beta\mu K\mathbb{E}[\lambda_{g1}(u)]u+(\acute{\alpha}_s-\beta)\mathbb{E}[\lambda_{g1}(u)N_1(u)]+\grave{\alpha}_c\mathbb{E}[\lambda_{g2}(u)N_1(u)]\}\mathrm{d}u$$
$$=\int_0^t\{\beta\mu K\mathbb{E}[\lambda_{g1}(u)]u+(\acute{\alpha}_s-\beta)\mathbb{E}[\lambda_{g1}(u)N_1(u)]+\grave{\alpha}_c\mathbb{E}[\lambda_{g1}(u)N_2(u)]\}\mathrm{d}u.$$

Similarly,
$$\mathbb{E}[\lambda_{g1}(t)N_2(t)]=\mathbb{E}[[\lambda_{g1},N_2]_t]+\mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+}\lambda_{g1}(u)k_2 N_2(\mathrm{d}u\times\mathrm{d}k_2)\right]+\mathbb{E}\left[\int_0^t N_2(u-)\mathrm{d}\lambda_{g1}(u)\right]$$

and
$$\mathbb{E}[[\lambda_{g1},N_2]_t]=\mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+}\alpha_c(1+(k_2-1)\eta)k_2 N_2(\mathrm{d}u\times\mathrm{d}k_2)\right]$$
$$=\alpha_c(K_{1\lambda_{g1}}+(K^{(2)}-K_{1\lambda_{g1}})\eta)\mathbb{E}[\lambda_{g1}(t)]t$$
$$=\alpha_c\bar{K}\mathbb{E}[\lambda_{g1}(t)]t$$

and
$$\mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+}\lambda_{g1}(u)k_2 N_2(\mathrm{d}u\times\mathrm{d}k_2)\right]=\int_0^t\mathbb{E}[k_2\lambda_{g1}(u)\lambda_{g2}(u)]\mathrm{d}u=K_{2\lambda_{g1}\lambda_{g2}}\mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)]t$$

and
$$\mathbb{E}\left[\int_0^t N_2(u-)\mathrm{d}\lambda_{g1}(u)\right]=\mathbb{E}\left[\int_0^t N_2(u)\beta(\mu-\lambda_{g1}(u))\mathrm{d}u\right]$$
$$+\mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+}N_2(u-)\alpha_s(1+(k_1-1)\eta)N_1(\mathrm{d}u\times\mathrm{d}k_1)\right]$$
$$+\mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+}N_2(u-)\alpha_c(1+(k_2-1)\eta)N_2(\mathrm{d}u\times\mathrm{d}k_2)\right]$$
$$=\int_0^t\{\beta\mu K_{1\lambda_{g1}}\mathbb{E}[\lambda_{g1}(u)]u+(\grave{\alpha}_s-\beta)\mathbb{E}[\lambda_{g1}(u)N_2(u)]+\acute{\alpha}_c\mathbb{E}[\lambda_{g2}(u)N_2(u)]\}\mathrm{d}u$$
$$=\int_0^t\{\beta\mu K_{1\lambda_{g1}}\mathbb{E}[\lambda_{g1}(u)]u+(\grave{\alpha}_s-\beta)\mathbb{E}[\lambda_{g1}(u)N_2(u)]+\acute{\alpha}_c\mathbb{E}[\lambda_{g1}(u)N_1(u)]\}\mathrm{d}u.$$

Combing the above results, we have
$$\mathbb{E}[\lambda_{g1}(t)N_1(t)]=\alpha_s\bar{K}\mathbb{E}[\lambda_{g1}(t)]t+K_{1\lambda_{g1}^2}\mathbb{E}[\lambda_{g1}^2(t)]t$$
$$+\int_0^t\{\beta\mu K\mathbb{E}[\lambda_{g1}(u)]u+(\acute{\alpha}_s-\beta)\mathbb{E}[\lambda_{g1}(u)N_1(u)]+\grave{\alpha}_c\mathbb{E}[\lambda_{g1}(u)N_2(u)]\}\mathrm{d}u$$
$$\mathbb{E}[\lambda_{g1}(t)N_2(t)]=\alpha_c\bar{K}\mathbb{E}[\lambda_{g1}(t)]t+K_{2\lambda_{g1}\lambda_{g2}}\mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)]t$$
$$+\int_0^t\{\beta\mu K\mathbb{E}[\lambda_{g1}(u)]u+(\grave{\alpha}_s-\beta)\mathbb{E}[\lambda_{g1}(u)N_2(u)]+\acute{\alpha}_c\mathbb{E}[\lambda_{g1}(u)N_1(u)]\}\mathrm{d}u.$$

Let

$$\mathbf{M}_2 = \begin{bmatrix} \acute{\alpha}_s - \beta & \grave{\alpha}_c \\ \acute{\alpha}_c & \grave{\alpha}_s - \beta \end{bmatrix}, \quad \mathbf{K}_2 = \begin{bmatrix} K_{1\lambda_{g1}^2} & 0 \\ 0 & K_{2\lambda_{g1}\lambda_{g2}} \end{bmatrix}.$$

Then,

$$\begin{bmatrix} \dfrac{\mathrm{d}\mathbb{E}[\lambda_{g1}(t)N_1(t)]}{\mathrm{d}t} \\ \dfrac{\mathrm{d}\mathbb{E}[\lambda_{g1}(t)N_2(t)]}{\mathrm{d}t} \end{bmatrix} = \mathbf{M}_2 \begin{bmatrix} \mathbb{E}[\lambda_{g1}(t)N_1(t)] \\ \mathbb{E}[\lambda_{g1}(t)N_2(t)] \end{bmatrix} + \begin{bmatrix} \alpha_s \bar{K}\mathbb{E}[\lambda_{g1}(t)] + K_{1\lambda_{g1}^2}\mathbb{E}[\lambda_{g1}^2(t)] + \beta\mu K \mathbb{E}[\lambda_{g1}(t)]t \\ \alpha_c \bar{K}\mathbb{E}[\lambda_{g1}(t)] + K_{2\lambda_{g1}\lambda_{g2}}\mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)] + \beta\mu K \mathbb{E}[\lambda_{g1}(t)]t \end{bmatrix}$$

$$= \mathbf{M}_2 \begin{bmatrix} E[\lambda_{g1}(t)N_1(t)] \\ \mathbb{E}[\lambda_{g1}(t)N_2(t)] \end{bmatrix} + \mathbb{E}[\lambda_{g1}(t)] \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \beta\mu Kt + \begin{bmatrix} \alpha_s \bar{K} \\ \alpha_c \bar{K} \end{bmatrix} - \frac{1}{2}\mathbf{K}_2 \mathbf{M}^{-1} \begin{bmatrix} (\alpha_s^2 + \alpha_c^2)\bar{\bar{K}} + 2\beta\mu \\ 2(\alpha_s\alpha_c\bar{K} + \beta\mu) \end{bmatrix} \right)$$

and the particular solution is

$$\begin{bmatrix} \mathbb{E}[\lambda_{g1}(t)N_1(t)] \\ \mathbb{E}[\lambda_{g1}(t)N_2(t)] \end{bmatrix} = -\mathbb{E}[\lambda_{g1}(t)] \left\{ \beta\mu \mathbf{K}\mathbf{M}_2^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} t + \mathbf{M}_2^{-1} \begin{bmatrix} \alpha_s \bar{K} \\ \alpha_c \bar{K} \end{bmatrix} - \frac{1}{2}\mathbf{M}_2^{-1}\mathbf{K}_2\mathbf{M}^{-1} \begin{bmatrix} (\alpha_s^2 + \alpha_c^2)\bar{\bar{K}} \\ 2\alpha_s\alpha_c\bar{K} \end{bmatrix} \right.$$

$$\left. + \mathbf{M}_2^{-1}(\mathbf{K}\mathbf{M}_2^{-1} - \mathbf{K}_2\mathbf{M}^{-1}) \begin{bmatrix} \beta\mu \\ \beta\mu \end{bmatrix} \right\}.$$

Note that if $\mathbf{K}\mathbf{M}_2^{-1} - \mathbf{K}_2\mathbf{M}^{-1}$ is close to zero, then the last term of the above equation is also close to zero and the formula is similar to the one in the simple Hawkes model.

A.3   Step 3

In the final step, $\mathbb{E}[N_1^2(t)]$ and $\mathbb{E}[N_1(t)N_2(t)]$ are derived. We have

$$\mathbb{E}[N_1^2(t)] = \mathbb{E}[[N_1]_t] + 2\mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} k_1 N_1(u) N_1(\mathrm{d}u \times \mathrm{d}k_1) \right]$$

$$= \mathbb{E}[k_1^2 \lambda_{g1}(t)]t + 2\int_0^t \mathbb{E}[k_1 \lambda_{g1}(u) N_1(u)]\mathrm{d}u$$

$$= K^{(2)}\mathbb{E}[\lambda_{g1}(t)]t + 2\int_0^t K_{1\lambda_{g1}N_1}\mathbb{E}[\lambda_{g1}(u) N_1(u)]\mathrm{d}u$$

and

$$\mathbb{E}[N_1(t)N_2(t)] = \mathbb{E}[[N_1, N_2]_t] + \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} k_2 N_1(u-) N_2(\mathrm{d}u \times \mathrm{d}k_2) \right] + \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} k_1 N_2(u-) N_1(\mathrm{d}u \times \mathrm{d}k_1) \right]$$

$$= \int_0^t K_{2\lambda_{g2}N_1}\mathbb{E}[\lambda_{g2}(u) N_1(u)]\mathrm{d}u + \int_0^t K_{1\lambda_{g1}N_2}\mathbb{E}[\lambda_{g1}(u) N_2(u)]\mathrm{d}u$$

$$= 2\int_0^t K_{1\lambda_{g1}N_2}\mathbb{E}[\lambda_{g1}(u) N_2(u)]\mathrm{d}u.$$

Then

$$\begin{bmatrix} \dfrac{\mathrm{d}\mathbb{E}[N_1^2(t)]}{\mathrm{d}t} \\ \dfrac{\mathrm{d}\mathbb{E}[N_1(t)N_2(t)]}{\mathrm{d}t} \end{bmatrix} = 2\mathbf{K}_3 \begin{bmatrix} \mathbb{E}[\lambda_{g1}(t)N_1(t)] \\ \mathbb{E}[\lambda_{g1}(t)N_2(t)] \end{bmatrix} + \begin{bmatrix} K^{(2)}\mathbb{E}[\lambda_{g1}(t)] \\ 0 \end{bmatrix}$$

where

$$\mathbf{K}_3 = \begin{bmatrix} K_{1\lambda_{g1}N_1} & 0 \\ 0 & K_{1\lambda_{g1}N_2} \end{bmatrix}$$

and

$$\begin{bmatrix} \mathbb{E}[N_1^2(t)] \\ \mathbb{E}[N_1(t)N_2(t)] \end{bmatrix} = -\mathbb{E}[\lambda_{g1}(t)]\mathbf{K}_3 \left\{ \beta\mu \mathbf{K}\mathbf{M}_2^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} t^2 \right.$$

$$\left. + \left(2\mathbf{M}_2^{-1} \begin{bmatrix} \alpha_s \bar{K} \\ \alpha_c \bar{K} \end{bmatrix} - \mathbf{M}_2^{-1}\mathbf{K}_2\mathbf{M}^{-1} \begin{bmatrix} (\alpha_s^2 + \alpha_c^2)\bar{\bar{K}} \\ 2\alpha_s\alpha_c\bar{K} \end{bmatrix} + 2\mathbf{M}_2^{-1}(\mathbf{K}\mathbf{M}_2^{-1} - \mathbf{K}_2\mathbf{M}^{-1}) \begin{bmatrix} \beta\mu \\ \beta\mu \end{bmatrix} - \begin{bmatrix} K^{(2)}/K_{1\lambda_{g1}N_1} \\ 0 \end{bmatrix} \right) t \right\}.$$