Marked Hawkes process modeling of price dynamics and volatility estimation

Kyungsub Lee[a], Byoung Ki Seo[a,*]

[a]*School of Business Administration, UNIST(Ulsan National Institute of Science and Technology), Ulsan 689-798, Korea*

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.

Keywords: Tick price dynamics, Marked Hawkes process, Volatility, Ultra-high-frequency data, Impact of mark
JEL: C13, C51, C61, G17

1. Introduction

In this paper, we model the tick dynamics of stock prices observed at ultra-high-frequency level based on the symmetric marked Hawkes process and examine the empirical properties of the price dynamics. It is known that the simple self and mutually excited Hawkes model for the price dynamics with unit jump size incorporates the stylized facts of the ultra-high-frequency financial data such as market microstructure noise and order clustering. However, we observe random size jumps, i.e., not a constant jump as in the simple Hawkes model, in the tick structure of equity markets, especially when there is a high ratio between the stock price and minimum tick size. Combining the Hawkes model with a mark structure, additional information to each event, we propose more realistic model of the tick price dynamics to deal with random size jumps.

The recent studies on (ultra)-high-frequency data and the market microstructure have developed in several ways. There is growing literature on the financial theory of market microstructure and limit order book (Roşu, 2009), and the role of algorithmic trading at high-frequency rate (Foucault, 2012; Chaboud et al., 2014; Hoffmann, 2014). A number of researches focused on the reduced form or stochastic modeling of the limit order dynamics and order executions, see 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 ultra-high-frequency data is also an important subject as they exhibit the distinctive characteristics from the macro price dynamics. For example, one need to be careful with applying the typical statistical methods to ultra-high-frequency data, when computing the realized volatility (Andersen et al., 2003) due to the microstructure noise which refers to the mean reverting properties of price processes at high frequency level. Aït-Sahalia et al. (2005), Zhang et al. (2005), Hansen and Lunde (2006) and Aït-Sahalia et al. (2011) devoted to measure the volatility of the return in the presence of the 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 high-frequency contrast with the daily data cases where the cross correlations are negligible. The lead/lag relationship among international index futures of different countries are also observed by Alsayed and McGroarty (2014).

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

*Corresponding author. Tel.: +82 52 217 3150; fax: +82 52 217 3109.
Email addresses: klee@unist.ac.kr (Kyungsub Lee), bkseo@unist.ac.kr (Byoung Ki Seo)

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 as the frequency of the past down movements increase and vice versa. This causes the mean reverting property in price dynamics although the correlations last less than a few seconds. Second, there are also autocorrelations between the movements of the same direction. This cause the volatility clustering however different from the clustering in macro level, typically modeled by GARCH (Bollerslev, 1986) or the stochastic volatility model (Heston, 1993), as the clustering properties only last a few seconds. The durations of the autocorrelations are much shorter than that of the autocorrelation observed in daily level.

These properties are well incorporated into the Hawkes model, which belongs to the class of point process and is introduced by Hawkes (1971), and hence there has been growing related work of modeling price dynamics based on the Hawkes process. Hewlett (2006) introduced the bivariate Hawkes process to model the buy and sell oder arrivals and examined the impact of orders to future prices. Bacry et al. (2012) explained the non-parametric estimation method for the symmetric Hawkes process based on high-frequency futures data. Based on mutually exciting Hawkes process, Bacry et al. (2013) suggested the mathematical framework that incorporate the market microstructure noise and the Epps effect, the correlation between the returns of two different assets at high sampling frequency. Trade clustering properties of the price dynamics in micro level is well incorporated by the self excited Hawkes process (Da Fonseca and Zaatour, 2014b). Da Fonseca and Zaatour (2014a) derived the formulas for the moments and correlation function for the self and mutually excited Hawkes process. Bacry and Muzy (2014) introduced a multivariate Hawkes process for the up and down price movements and buy and sell orders to explain the stylized facts of the market impact and microstructure. 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 with the realized volatility.

The previous work mainly focused on the simple point process model where the jump size is constant. We extend the existing simple Hawkes model to the marked Hawkes model to handle more realistic price movements in stock markets where we observe the random size jumps (marks). The future effects of marks depend on the absolute sizes of the marks and hence to deal with the future impact of the mark, a linear impact function is introduced. Our empirical study shows that the estimates of the slope parameter of the impact function are significant positive values in stock markets. This implies that the larger marks tends to magnify the future intensities more than the smaller marks. For the distribution of the mark, we do not assume a specific distribution in this paper but we use the empirical distribution for estimations and volatility calculation. Our model is not limited to the independent mark as the empirical studies shows the intensity dependent mark distribution.

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

2. Symmetric marked Hawkes model

2.1. Marked point process

In this subsection, we introduce the basic concepts of marked point processes. Our mathematical framework is in line with Daley and Vere-Jones (2003). With given complete separable metric state space $\mathcal{X}$, a point process is a measure to count a number of random events that occur in an open set which belongs to the σ-field of $\mathcal{X}$'s Borel set, $\mathcal{B}_{\mathcal{X}}$. To deal with random events, we introduce a probability space $(\Omega, \mathbb{P})$. Then 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, we omit the term ω for the notational simplicity.

A marked point process is a more complex model and is introduced to describe not only the location of random events but also additional information, called 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, to model the price movements over time, the space is defined as a real line. In addition, the mark space $\mathcal{K} = \mathbb{Z}^+$, the space of price jump sizes. This implies that the absolute jump sizes of the price movements are represented by positive integer multiples of a minimum jump size, δ.

For the price dynamics modeling, there are two marked point processes N_1 and N_2 which represent up and down movements, respectively. We assume 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 $N_{gi}(\cdot) = N_i(\cdot, \mathbb{Z}^+)$, 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.

We assume that the price process is 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 a 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 and, heuristically speaking, satisfy

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

Due to the discontinuity, the intensities may not be unique at the discontinuities and we consider $\lambda_{gi}(t)$ are the left continuous modifications in general, but if the intensities are used as integrators of some stochastic integrations, then we use the right continuous modification of λ_{gi} as the integrators.

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

$$\begin{aligned}\lambda_{gi}(t) &= \mu + \sum_{j=1,2} q_{ij} \int_{(-\infty,t)\times\mathbb{Z}^+} g_{ij}(k_j)\phi_{ij}(t-u)N_j(\mathrm{d}u \times \mathrm{d}k) \\ &= \mu + \sum_{j=1,2} q_{ij} \sum_{-\infty<u_\ell<t} g_{ij}(k_j(u_\ell))\phi_{ij}(t-u_\ell).\end{aligned}$$

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 the system from outside in a Poisson process at rate μ. These ancestor generate offspring and generated offspring become new ancestor to generate new offspring. Due to the ancestor type j born at time u with mark k, whether immigrant or not, type i offspring are generated by a Poisson process with 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 diminishes with $\phi_{ij}(t-u)$ as t increase. Since the combination of $q_{ij}g_{ij}(k(u))\phi_{ij}(t-u)$ is not unique, a normalization method to determine $q_{ij}, g_{ij}(k)$ and $\phi_{ij}(t-u)$ is used in general.

We assume the exponential decay kernel for $\phi_{ij}(t-u) = \phi(t-u) = \beta \mathrm{e}^{-\beta(t-u)}, \beta > 0$, 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 branching coefficients and $\mathbf{Q} := \{q_{ij}\}_{i,j=1,2}$ is called branching matrix. With an exponential decay kernel, $(\lambda_{g1}, \lambda_{g2})$ is Markovian and the computations in this paper largely depends on this property.

In this paper, We consider 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 $f(k_i(t)|\lambda_{gi}(t))$. If the marks are independent given the ground processes, then the analysis is rather straightforward with given analysis on non-marked Hawkes processes from Lee and Seo (2014). Indeed, the dependency assumption is consistent with the empirical study in Section 4. However, throughout the paper, we do not assume the specific parametric distribution for the mark except in simulation studies. We perform the estimation procedures and volatility analysis 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, we define the associated jump processes by

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

where in the l.h.s., the stochastic process N_i is represented with the sole parameter t and in the r.h.s., the measure N_i is represented with both location t and mark size k. In the stochastic process representation, $N_i(t)$ counts the the number of events over $(0, t]$ with weight k. The jump processes $N_i(t)$ are considered as right continuous with left limit. 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. As a jump process, the ground process is also regarded as right continuous with left limit.

2.2. Linear impact function

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 and linear with 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 some constant $K > 0$, we assume

$$\mathbb{E}[k_i \lambda_{gi}(t)] = K\mathbb{E}[\lambda_{gi}(t)] \tag{1}$$

and

$$\{1 + (K-1)\eta\}\left(\frac{\alpha_s}{\beta} + \frac{\alpha_c}{\beta}\right) < 1. \tag{2}$$

The condition (2) is similar to the existence condition of the simple symmetric Hawkes process except that we have additional term $\{1 + (K-1)\eta\}$. We additionally assume that the ground intensity process is represented by

$$\begin{aligned}
\lambda_{g1}(t) = \mu &+ q_s \int_{(-\infty,t)\times\mathbb{Z}^+} g(k_1)\beta \mathrm{e}^{-\beta(t-u)} N_1(\mathrm{d}u \times \mathrm{d}k_1) \\
&+ q_c \int_{(-\infty,t)\times\mathbb{Z}^+} g(k_2)\beta \mathrm{e}^{-\beta(t-u)} N_2(\mathrm{d}u \times \mathrm{d}k_2), \qquad (3) \\
\lambda_{g2}(t) = \mu &+ q_c \int_{(-\infty,t)\times\mathbb{Z}^+} g(k_1)\beta \mathrm{e}^{-\beta(t-u)} N_1(\mathrm{d}u \times \mathrm{d}k_1) \\
&+ q_s \int_{(-\infty,t)\times\mathbb{Z}^+} g(k_2)\beta \mathrm{e}^{-\beta(t-u)} N_2(\mathrm{d}u \times \mathrm{d}k_2). \qquad (4)
\end{aligned}$$

The point process (N_1, N_2) defined under the above ground intensity is then 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

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

is a compensator for

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

and hence

$$\int_{(\cdot,t)\times\mathbb{Z}^+} g(k_i)\beta \mathrm{e}^{-\beta(t-u)} N_i(\mathrm{d}u \times \mathrm{d}k_1) - \int_{\cdot}^{t} \mathbb{E}[g(k_i)|\lambda_{gi}(u)]\lambda_{gi}(u)\beta \mathrm{e}^{-\beta(t-u)}\mathrm{d}u$$

is a martingale. Thus, by taking the unconditional expectation for the ground intensity formulas in Eqs. (3) and (4), we have

$$\begin{aligned}
\mathbb{E}[\lambda_{g1}(t)] &= \mu + q_s \int_{-\infty}^{t} \mathbb{E}[\mathbb{E}[g(k_1)|\lambda_{g1}(u)]\lambda_{g1}(u)]\beta \mathrm{e}^{-\beta(t-u)}\mathrm{d}u \\
&\quad + q_c \int_{-\infty}^{t} \mathbb{E}[\mathbb{E}[g(k_2)|\lambda_{g2}(u)]\lambda_{g2}(u)]\beta \mathrm{e}^{-\beta(t-u)}\mathrm{d}u, \\
\mathbb{E}[\lambda_{g2}(t)] &= \mu + q_c \int_{-\infty}^{t} \mathbb{E}[\mathbb{E}[g(k_1)|\lambda_{g1}(u)]\lambda_{g1}(u)]\beta \mathrm{e}^{-\beta(t-u)}\mathrm{d}u \\
&\quad + q_s \int_{-\infty}^{t} \mathbb{E}[\mathbb{E}[g(k_2)|\lambda_{g2}(u)]\lambda_{g2}(u)]\beta \mathrm{e}^{-\beta(t-u)}\mathrm{d}u
\end{aligned}$$

and by Eq. (1),

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

and we write

$$\begin{aligned}
\mathbb{E}[\lambda_{g1}(t)] &= \mu + \frac{\alpha_s}{\beta} \int_{-\infty}^{t} \{1+(K-1)\eta\}\mathbb{E}[\lambda_{g1}(u)]\beta \mathrm{e}^{-\beta(t-u)}\mathrm{d}u \\
&\quad + \frac{\alpha_c}{\beta} \int_{-\infty}^{t} \{1+(K-1)\eta\}\mathbb{E}[\lambda_{g2}(u)]\beta \mathrm{e}^{-\beta(t-u)}\mathrm{d}u, \\
\mathbb{E}[\lambda_{g2}(t)] &= \mu + \frac{\alpha_c}{\beta} \int_{-\infty}^{t} \{1+(K-1)\eta\}\mathbb{E}[\lambda_{g1}(u)]\beta \mathrm{e}^{-\beta(t-u)}\mathrm{d}u \\
&\quad + \frac{\alpha_c}{\beta} \int_{-\infty}^{t} \{1+(K-1)\eta\}\mathbb{E}[\lambda_{g2}(u)]\beta \mathrm{e}^{-\beta(t-u)}\mathrm{d}u.
\end{aligned}$$

By the stationarity of λ_{gi}, we have

$$\begin{aligned}
\mathbb{E}[\lambda_{g1}(t)] &= \mu + \frac{\alpha_s}{\beta}\{1+(K-1)\eta\}\mathbb{E}[\lambda_{g1}(t)] + \frac{\alpha_c}{\beta}\{1+(K-1)\mathbb{E}[\lambda_{g2}(t)]\eta\}, \\
\mathbb{E}[\lambda_{g2}(t)] &= \mu + \frac{\alpha_c}{\beta}\{1+(K-1)\eta\}\mathbb{E}[\lambda_{g1}(t)] + \frac{\alpha_s}{\beta}\{1+(K-1)\mathbb{E}[\lambda_{g2}(t)]\eta\}.
\end{aligned}$$

and

$$(I - \bar{\mathbf{Q}}) \begin{bmatrix} \mathbb{E}[\lambda_{g1}(t)] \\ \mathbb{E}[\lambda_{g2}(t)] \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \end{bmatrix}$$

where

$$\bar{\mathbf{Q}} = \{1+(K-1)\eta\} \begin{bmatrix} \frac{\alpha_s}{\beta} & \frac{\alpha_c}{\beta} \\ \frac{\alpha_c}{\beta} & \frac{\alpha_s}{\beta} \end{bmatrix}.$$

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

$$\mathbb{E}[\lambda_{g1}(t)] = \mathbb{E}[\lambda_{g2}(t)] = \frac{\mu\beta}{\beta - (\alpha_s+\alpha_c)\{1+(K-1)\eta\}}. \tag{5}$$

Thus, if condition (2) 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, we compute the volatility formula of the asset return generated by the marked Hawkes processes. In the following notation, we define various Ks as similarly in Eq. (1) which simplify the notations.

Notation 1. *For a jump process X, let*

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

where t is the arrival time associated with event of mark k_i. If $X = \lambda_{gi}$, then we omit subscript for simplicity as in Eq. (1), i.e., $K = K_{i\lambda_{gi}}$ and $K^{(2)} = K_{i\lambda_{gi}}^{(2)}$. Furthermore,

$$\begin{aligned}
\ddot{K} &= 1 + 2(K-1)\eta + (K^{(2)} - 2K + 1)\eta^2 \\
\dot{K} &= K + (K^{(2)} - K)\eta \\
\breve{\alpha} &= \alpha\{1 + (K_{1\lambda_{g1}^2} - 1)\eta\} \\
\tilde{\alpha} &= \alpha\{1 + (K_{1\lambda_{g1}\lambda_{g2}} - 1)\eta\}.
\end{aligned}$$

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. (3) and (4). If the price process S_t follows*

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

with 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

$$\begin{aligned}
&\begin{bmatrix} E[N_1^2(t)] \\ \mathbb{E}[N_1(t)N_2(t)] \end{bmatrix} = -\mathbb{E}[\lambda_{g1}(t)] \begin{bmatrix} K_{1\lambda_{g1}N_1} & 0 \\ 0 & K_{1\lambda_{g1}N_2} \end{bmatrix} \\
&\left\{ \beta\mu\mathbf{M}^{-1}\begin{bmatrix}1\\1\end{bmatrix} t^2 + \left(2\mathbf{M}^{-1}\begin{bmatrix}\alpha_s\dot{K}\\ \alpha_c\dot{K}\end{bmatrix} - \mathbf{M}^{-2}\begin{bmatrix} K_{1\lambda_{g1}^2}\{(\alpha_s^2+\alpha_c^2)\ddot{K} + 2\beta\mu\} \\ 2K_{1\lambda_{g1}\lambda_{g2}}(\alpha_s\alpha_c\ddot{K} + \beta\mu)\end{bmatrix} - \begin{bmatrix} K^{(2)}/K_{1\lambda_{g1}N_1} \\ 0 \end{bmatrix} \right) t \right\}
\end{aligned}$$

where

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

Proof. See Appendix Appendix A. □

Remark 2. *Assuming $K_{1\lambda_{g1}N_1} \approx K_{1\lambda_{g1}N_2}$ and $K_{1\lambda_{g1}^2} \approx K_{1\lambda_{g1}\lambda_{g2}}$, i.e., $\tilde{\alpha} \approx \breve{\alpha}$, we have*

$$\begin{aligned}
\mathbb{E}[(N_1(t) - N_2(t))^2] &= 2(\mathbb{E}[N_1^2(t)] - \mathbb{E}[N_1(t)N_2(t)]) \\
&\approx 2K_{1\lambda_{g1}N_1}\mathbb{E}[\lambda_{g1}(t)] \left(\frac{K_{1\lambda_{g1}^2}\ddot{K}(\alpha_s - \alpha_c)^2}{(\beta - \breve{\alpha}_s + \breve{\alpha}_c)^2} + \frac{2(\alpha_s - \alpha_c)\dot{K}}{\beta - \breve{\alpha}_s + \breve{\alpha}_c} + \frac{K^{(2)}}{K_{1\lambda_{g1}N_1}} \right) t. \qquad (6)
\end{aligned}$$

In addition, if all Ks 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.

2.4. Likelihood function

To estimate the parameters in the intensity processes such as $\alpha_s, \alpha_c, \beta$ and η, we need to compute the log-likelihood. The log-likelihood of the realized jumps and their marks of (N_1, N_2) over the period $[0, T]$ is represented by

$$\begin{aligned}
&\left(\int_{(0,T]} \log\lambda_{g1}(u) N_{g1}(\mathrm{d}u) + \int_{(0,T]} \log\lambda_{g2}(u) N_{g2}(\mathrm{d}u) - \int_0^T (\lambda_{g1}(u) + \lambda_{g2}(u))\mathrm{d}u \right) \\
&+ \left(\int_{(0,T]\times\mathbb{Z}^+} \log f(k_1|\lambda_{g1}(u)) N_1(\mathrm{d}u \times \mathrm{d}k_1) + \int_{(0,T]\times\mathbb{Z}^+} \log f(k_2|\lambda_{g2}(u)) N_2(\mathrm{d}u \times \mathrm{d}k_2) \right) \\
&=: \log L_g + \log L_m. \qquad (7)
\end{aligned}$$

The above formula implies that the log-likelihood is separated into two parts, $\log L_g$ of the ground processes and $\log L_m$ of the mark distribution. Note that $\log L_g$ dose not depend on the mark distribution f. Thus, even if we do not know the exact distribution of the mark k, we are able to compute the log-likelihood of the ground process part, $\log L_g$. In the empirical study later, we do not assume any specific form of f and just use the empirical distribution to conduct the maximum likelihood procedure to maximize $\log L_g$, the likelihood of the ground processes.

3. Simulation example

3.1. Symmetric model

In this paper, we generally do not assume the specific distribution of the mark. However, for the simulation study, we need 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.,

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

This implies that the conditional expectation of the mark size k_i with 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 the blow up of the marked Hawkes process. With this setting, the conditional expectation of the impact depends on the current intensity:

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

With each differently presumed conditional distribution and parameter setting, we generate 500 sample paths of the two dimensional marked Hawkes process and corresponding ground intensities. The time horizon for the path is set to be 5.5 hours which equals to the time horizon used in empirical studies later. The simulation mechanism is similar to the simple Hawkes models but additionally needed 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 to maximized $\log L_g$ in Eq. (7) and the results are presented 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, $c = 0.25, d = 1.0, U = 9$.

Since we only calculate the likelihood for the ground processes, the estimates of $\mu, \alpha_s, \alpha_c, \beta$ and η are 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 estimates with 500 samples. The table shows that the estimates are consistent with the true values.

To calculate the volatility, we additionally need to compute Ks in Notation 1 which involving 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

Table 1: Simulation study for 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	0.2778	0.2788
std.	0.0021	0.0194	0.0190	0.0326	0.0170	0.0332	0.0171
True	0.1500	0.6200	0.5000	1.9000	0.2200		
mean	0.1499	0.6193	0.5008	1.8999	0.2177	0.1316	0.1310
std.	0.0027	0.0172	0.0149	0.0419	0.0502	0.0101	0.0030
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

intensities, we use the following statistics for the expectations instead:

$$\mathbb{E}[\lambda_{gi}(t)] \approx \frac{1}{T} N_{gi}(T) \tag{8}$$

$$\mathbb{E}[k_i \lambda_{gi}(t)] \approx \frac{1}{T} N_i(T) \tag{9}$$

$$\mathbb{E}[k_i^2 \lambda_{gi}(t)] \approx \frac{1}{T} \int_{(0,T]\times\mathbb{Z}^+} k_i^2 N_{gi}(\mathrm{d}u \times \mathrm{d}k) \tag{10}$$

$$\mathbb{E}[\lambda_{gi}^2(t)] \approx \frac{1}{T} \int_{(0,T]\times\mathbb{Z}^+} \lambda_{gi}(u) N_{gi}(\mathrm{d}u \times \mathrm{d}k) \tag{11}$$

$$\mathbb{E}[k_i \lambda_{gi}^2(t)] \approx \frac{1}{T} \int_{(0,T]\times\mathbb{Z}^+} k_i \lambda_{gi}(u) N_{gi}(\mathrm{d}u \times \mathrm{d}k) \tag{12}$$

$$\frac{1}{t}\mathbb{E}[\lambda_{gi}(t) N_i(t)] \approx \frac{2}{T^2} \int_{(0,T]\times\mathbb{Z}^+} N_i(u-) N_{gi}(\mathrm{d}u \times \mathrm{d}k) \tag{13}$$

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

To compute the right hand sides, the realized k_i, N_i and N_{gi} of 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 by Eqs. (3) (4), once $\mu, \alpha_s, \alpha_c, \beta$ and η are estimated.

We approximate the expectation of the ground intensity by the sample average of the total number of corresponding up or down moves per unit time in Eq. (8). Similarly for $\mathbb{E}[k_i \lambda_{gi}(t)]$ where we use the counting process N_i instead to compute the sample average.

The right hand side of Eq. (10) 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. (11) and (12), consider

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

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

Furthermore, we can let $\mathbb{E}[\lambda_{gi}(t) N_i(t)] = c_1 t + c_2$ for some constants c_1 and c_2 according to Appendix Appendix A and $\mathbb{E}[\lambda_{gi}(t) N_i(t)]/t$ converges to c_1 as t increases. Note that

$$\begin{aligned}\frac{2}{T^2}\mathbb{E}\left[\int_{(0,T]\times\mathbb{Z}^+} N_i(u-) N_{gi}(\mathrm{d}u \times \mathrm{d}k)\right] &= \frac{2}{T^2}\int_0^T \mathbb{E}[\lambda_{gi}(t) N_i(t)]\mathrm{d}t \\ &= c_1 + \frac{2c_2}{T} \approx c_1 \approx \frac{1}{2t}\mathbb{E}[\lambda_{gi}(t) N_i(t)]\end{aligned}$$

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

with approximations for large enough t and T. Similar argument is applied to Eq. (14).

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 is compared with the two scale realized volatility (TSRV) in the column 'TSRV' proposed by Zhang et al. (2005). We set the small time scale to be one second and the large time scale to be five minutes for the TSRV computation. The results shows that the Hawkes volatility and TSRV are very close to each other. The standard deviations of the Hawkes volatility are smaller than the ones of the TSRV for all simulation cases.

3.2. Other examples

In this subsection, we examine the cases where there is a discrepancy between the Hawkes volatility and realized volatility. First, we examine the fully characterized Hawkes model, i.e., the coefficients of the branching matrix is represented by

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

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 $\beta = \beta_{11} = \beta_{12} = \beta_{21} = \beta_{22}$.

In Table 2, we show the estimation results of the fully characterized Hawkes model with simulated paths with presumed parameters. The presumed parameters are presented in 'true' column and 500 sample paths are generated over one day time horizon, more precisely, 5.5 hours as in the previous example. The columns 'full' report the means and standard deviations of estimates under the maximum likelihood estimation with the fully characterized Hawkes model. We also perform the likelihood estimations under the symmetric Hawkes model, although 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}, and so on.

The 'S.Vol.' represents the sample volatility of the return computed by the sample standard deviation of the closing stock prices generated by 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 computed using the estimates of the symmetric Hawkes model. Two volatilities are biased around 4% compared with the sample volatility. The TSRV are larger than the sample volatilities and the Hawkes volatilities are smaller in these cases.

Second, we examine the symmetric marked Hawkes models where the model parameters change during the sample period. In Table 3, we show the estimation results with the symmetric Hawkes models of 5.5 hours 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

Table 3: Simulation study for 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

of the sample period, the price process is very volatile due to the possible large size of mark, and the rest part is rather stable. This is mimicking the case of 2010 Flash Crash and we will examine the empirical analysis in the later. The result shows the discrepancy between the TSRV and the Hawkes volatility which are both less than the sample volatility and TSRV is even less than the Hawkes volatility.

4. Empirical study

4.1. Data

For empirical studies, we use ultra high-frequency tick-by-tick data of some major stock prices consisting of several years with best bid and ask quotes reported in New York Stock Exchange (NYSE). The time horizon of sample for each day is set to be from 10:00 to 15:30. We do not use the data of 30 minutes right after the opening and right before the closing time to reduce the seasonality effects. The price movements patterns are usually different at 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 especially 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 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 get rid of the bid-ask bounce and hence the minimum jump size is the half of the tick size, \$0.005.

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

4.2. Unconditional distribution of mark

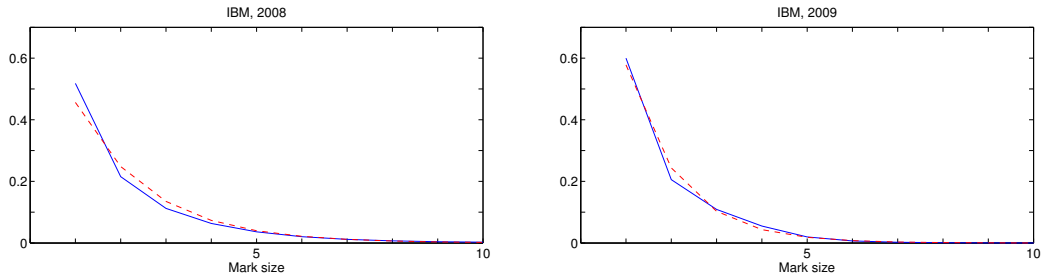
Table 4 compares the percentage of the mark size of IBM and GE from 2008 to 2011, i.e., the unconditional distribution of mark sizes are reported in the table. IBM has various mark sizes over the years but GE's mark size distributions rather concentrate on the minimum mark size. This is because that the price of IBM is relatively high and is around \$150, meanwhile the price of GE is around \$25. The unconditional distributions of the marks have exponentially decreasing shapes which are similar to the geometric distributions but do not seem to be identical to the distribution. We compare the empirical distribution of the marks of IBM, 2010 and 2011 with the geometric distributions in Figure 1. 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

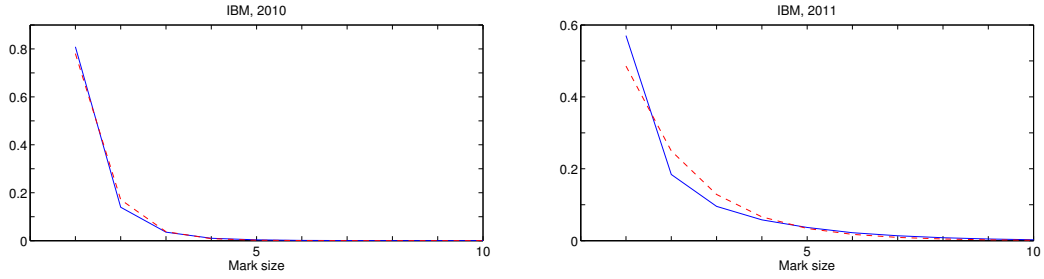
In this subsection, we examine the dependence between the mark size and the ground intensity, the number of expected events over unit interval. Our empirical evidences show that the mark size and the current ground intensity is significantly related to each other. First, we compute the empirical conditional expectation of the intensities with given mark size, $\mathbb{E}[\lambda_{gi}(t)|k_i]$. Since the ground intensities are unobservable, proxy intensities are introduced. The proxy intensities for 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

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

mark size	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
2	21.53	20.57	13.96	18.41	7.61	1.51	0.34	0.30
3	11.22	10.89	3.53	9.54	1.55	0.00	0.00	0.00
4	6.36	5.50	1.01	5.80	0.52	0.00	0.00	0.00
5	3.61	1.98	0.35	3.68	0.19	0.00	0.00	0.00
6	2.04	0.66	0.10	2.22	0.00	0.00	0.00	0.00
7	1.17	0.24	0.00	1.37	0.00	0.00	0.00	0.00
8	0.71	0.01	0.00	0.81	0.00	0.00	0.00	0.00



(a) IBM, 2008 (b) IBM, 2009



(c) IBM, 2010 (d) IBM, 2011

Figure 1: The empirical unconditional distribution of mark

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

	2011			2010		
mark size	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)

intensities is chosen as ten seconds. Mathematically, the up proxy intensity for the mark k_1 which takes place at time t is represented by

$$\frac{1}{\tau}\int_{(t-\tau,t]\times\mathbb{Z}^+} N_{g1}(\mathrm{d}u \times \mathrm{d}k_1)$$

where τ is the length of the period.

We compute the sample mean and standard error of the proxy intensities for each mark size in Table 5 for each year of 2010 and 2011, IBM. For example, for the mark size 6, there are 86,738 numbers of up jumps are reported and the sample mean of the up proxy intensity is 3.4784 and the sample standard error is 0.0145. The table shows that the proxy intensities increase as the given mark size increases. The negative integers in the column of mark size represent the down jump of the price. We also compute the proxy intensities for five seconds time horizon. The results are very similar to the previous case of 10 seconds time horizon and we do not report the result additionally.

Second, in Table 6, we present the relation between the mark sizes and the inferred ground intensities with the linear impact function using IBM tick data. Prior to calculate the inferred ground process, we estimates the parameters $\omega, \alpha_s, \alpha_c, \beta, \eta$ by maximizing $\log L_g$ defined in Eq. (7). The estimations are performed daily basis and the detailed estimation results are demonstrated later. After that the inferred ground intensities are computed with the estimates of $\omega, \alpha_s, \alpha_c, \beta, \eta$ using the definition of the ground intensities in Eqs. (3) and (4). 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 as the given mark size increases. This implies that if a large size of mark is observed, it is probably based on large ground intensities.

Third, we also illustrate the empirical expectations of the mark size conditionally upon given inferred ground intensities, $\mathbb{E}[k_i|\lambda_{gi}(t)]$, in Figure 2 using the tick data of IBM, 2008-2011. For each year, the empirical conditional expectation with given $\lambda_{gi} = n$ for an integer n is computed by the sample mean of the mark sizes whose associated inferred ground process is falling into $(n-1, n]$. We only plot the conditional expectations of the marks where total observed numbers of the mark are larger than 100 for each year, i.e., the samples with small numbers of observations are dropped out. In the figure, the

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

	2011			2010		
mark size	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)

ground intensities varies more as the years pass, implying that overall numbers of activities increase. We observe the most of intensities are less than 15 in 2008 but there is a huge spread in the inferred intensity in 2011 as a lot of observed intensities are larger than 15.

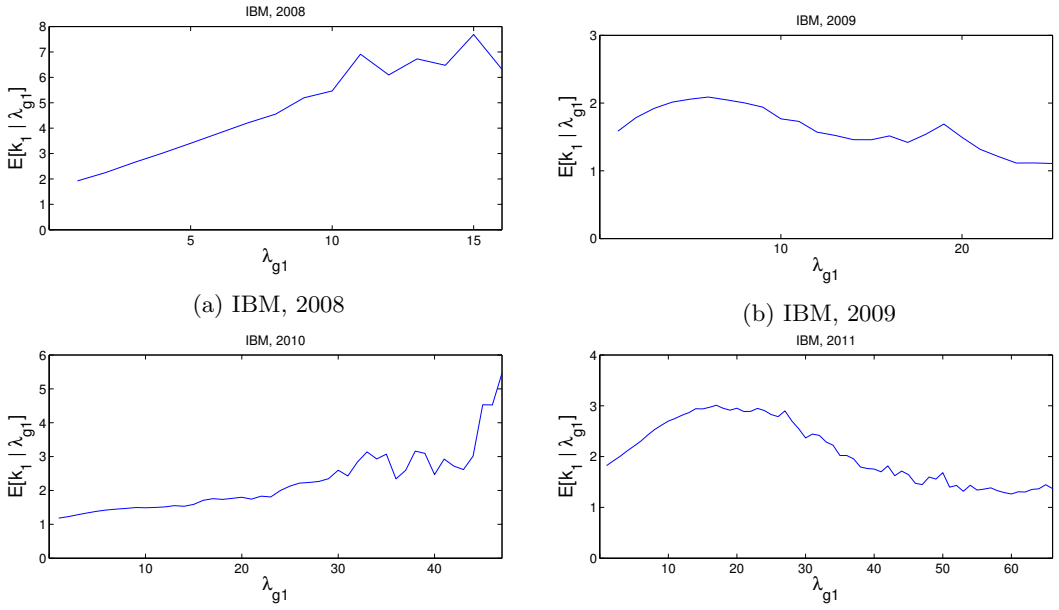
The shape of the conditional expectation of the mark changes over time. The conditional expectation tends to increase as the intensity increase in 2008 and 2010. In 2009 and 2011, the conditional expectations have humped shapes. The empirical conditional expectations of the mark given ground intensity computed monthly basis from January to June, 2011, of IBM are plotted in Figure 3. In the monthly basis empirical conditional expectation, we observe irregular patterns over time. The changing shape of the conditional distribution of marks over time is the reason why we do not specify the mark distribution, and perform the estimation in a non-parametric way for the part of the mark distribution.

4.4. Estimation result

Table 7 reports the likelihood estimation results of the marked Hawkes model with tick data of IBM, January 2011, where we maximize $\log L_g$ of Eq. (7). Numerically computed standard errors are reported in the parentheses. The estimations are performed on a daily basis, i.e., the estimates are recalculated in every business day. The behaviors of $\mu, \alpha_s, \alpha_c, \beta$ in Figure 4 are similar to those estimated in the simple Hawkes model, see Lee and Seo (2014).

The dynamics of η is illustrated in Figure 6 where η is 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 large impact for the future intensities. However, η is much less that one and this implies that the impact of mark size 2 is generally less than the total impact of two consecutive unit size jumps that occur over very short time interval. Note that few negative η are also observed. We also illustrate the dynamics of η estimated from CVX in Figure 7. The overall behaviors of η of IBM and CVX are similar but the η of CVX is more volatile.

We compare the Hawkes volatility computed by Remark 2 and TSRV of IBM, 2008-2011, in Figure 8. The trends of the Hawkes volatility and TSRV are similar to each other but the Hawkes volatility is generally larger than TSRV especially when the volatility is high. This tendency is also found in the simple symmetric Hawkes model (Lee and Seo, 2014). This discrepancy might be because of the



(a) IBM, 2008

(b) IBM, 2009

(c) IBM, 2010

(d) IBM, 2011

Figure 2: The conditional expectation of k_1 on λ_{g1}, IBM, 2008-2011

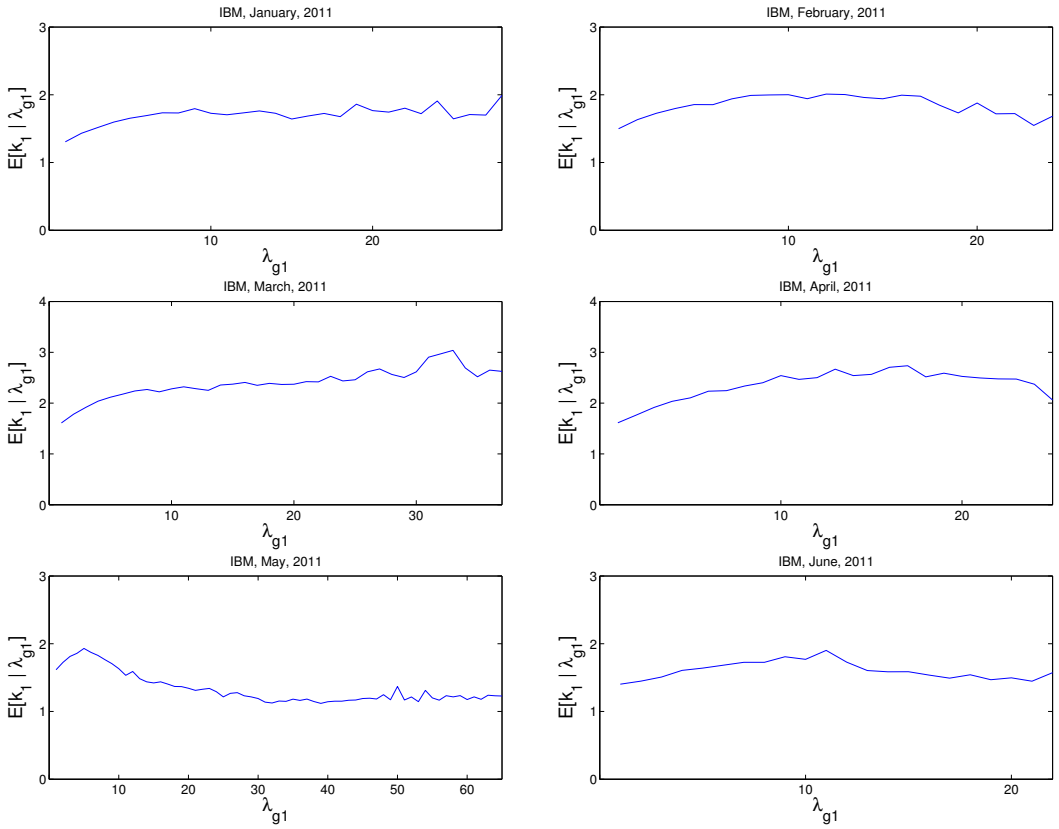
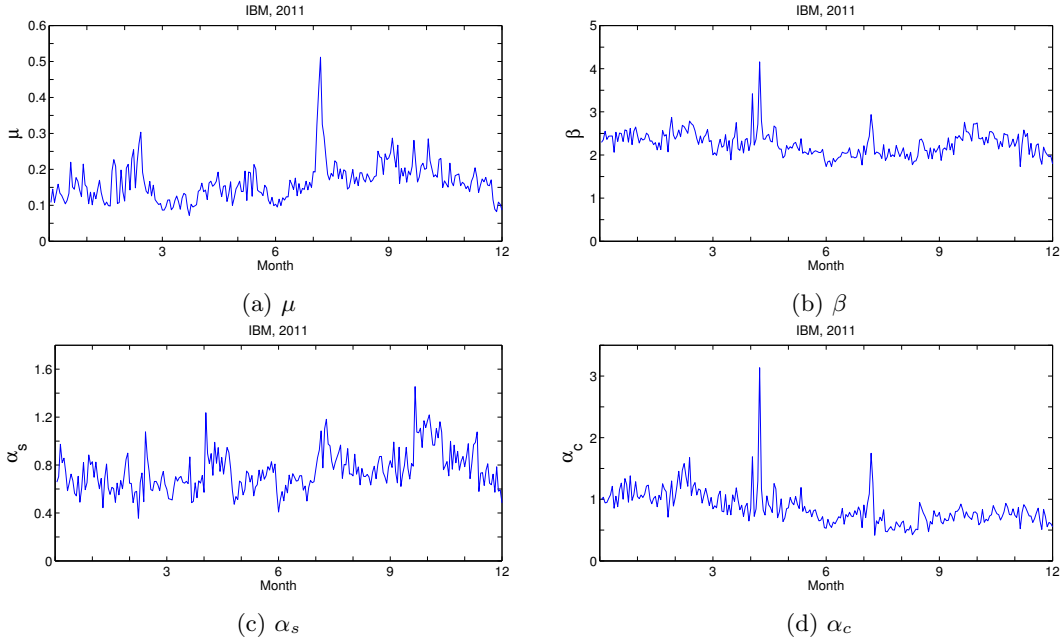


Figure 3: The conditional expectation of k_1 on λ_{g1}, IBM, 2011

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)	



(a) μ (b) β

(c) α_s (d) α_c

Figure 4: Marked Hawkes estimation result, IBM, 2011

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

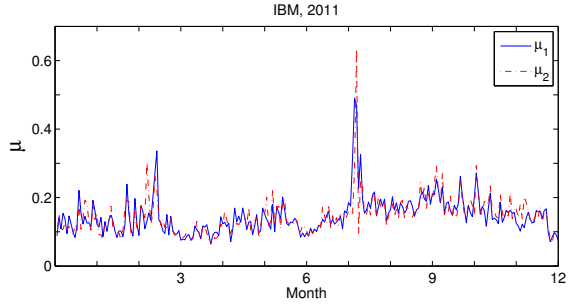
	μ_1	μ_2	α_{11}	α_{22}	α_{12}	α_{21}	β_{11}	β_{22}	β_{12}	β_{21}
IBM, 2011										
mean	0.1462	0.1470	0.5404	0.5621	0.6443	0.6611	1.5499	1.5793	1.7538	1.8056
std.	0.0543	0.0545	0.1962	0.2361	0.2144	0.2717	0.3845	0.4192	0.3373	0.4282
MAPE	0.1436		0.3924		0.3179		0.2881		0.2247	

restriction in the parameter settings to be symmetric and Markovian. However, the exact reason is unclear at this point. Note that in the previous simulation study, the Hawkes volatility and TSRV are converges to each other.

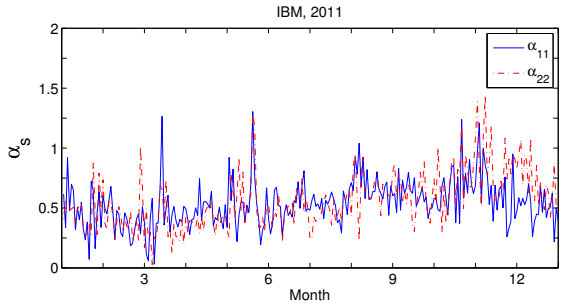
Indeed, empirical studies suggest that the parameter restrictions for symmetry does not perfectly meet with the real data. In Figure 5, we illustrate the dynamics of of all parameters of the fully characterized Hawkes model of Subsection 3.2 for IBM, 2011. The result shows that for each parameter pair (μ_1, μ_2), $(\alpha_{11}, \alpha_{22})$, $(\alpha_{12}, \alpha_{21})$, (β_{11}, β_{22}) and (β_{12}, β_{21}), we observe the similar trend over time but those are not exactly the same to each other. The summary statistics in Table 8 shows that the parameter pair is almost the same in the mean but the mean absolute percentage error (MAPE) also shows the 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. Similar observation is found in the simple Hawkes approach (Lee and Seo, 2014).

It is interesting to note that 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 is around 0.02 which is much smaller than the annual average 0.16, when the market is very unstable with TSRV 0.8766 and the Hawkes volatility 1.4530. At the May 6, 2010 Flash Crash, the estimated η of IBM is around 0.01 (annual average of $\eta = 0.23$) when TSRV is 0.9656 and the Hawkes volatility is 1.3264. Due to the statement of Federal reserve, the stock market is highly volatile at August 9, 2011, and we observe η of IBM is around 0.01 which is much smaller than annual average 0.11. CVX also shows the similar pattern. The estimated η of CVX for the above cases are around 0.04, 0.01 and 0.07, respectively, meanwhile the annual averages in 2008, 2010 and 2011 are 0.14, 0.41 and 0.24, respectively. In a highly volatile market, much more numbers of large size marks are observed than in a stable market, and with those large size marks, the linear relation between the mark size and the future impact is weak.

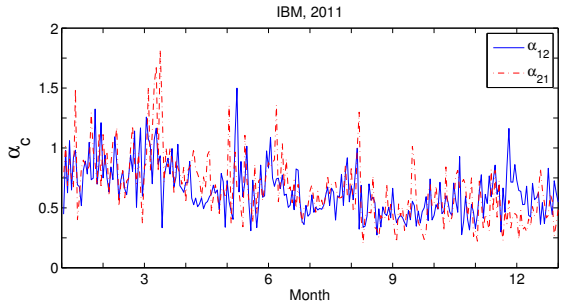
In the estimation of the Hawkes models, we use the data of all arrival times over the sample period. Even with ten minutes interval, usually more than a thousand data of arrival times are available, enough



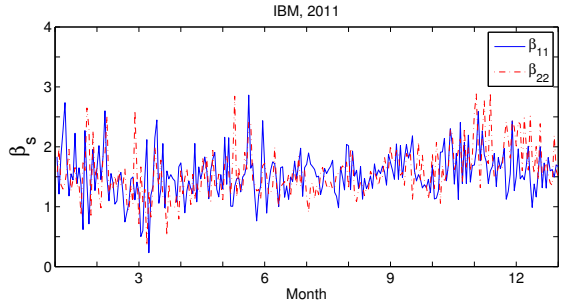
(a) μ_1 and μ_2



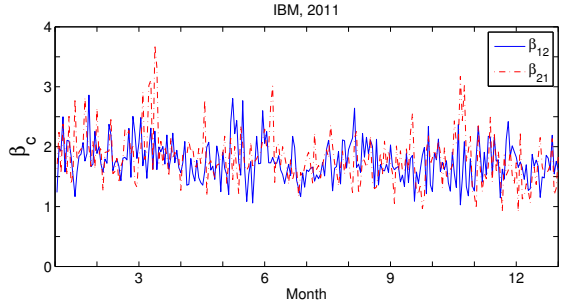
(b) α_{11} and α_{22}



(c) α_{12} and α_{21}



(d) β_{11} and β_{22}



(e) β_{12} and β_{21}

Figure 5: Estimation result with fully characterized marked Hawkes, IBM, 2011

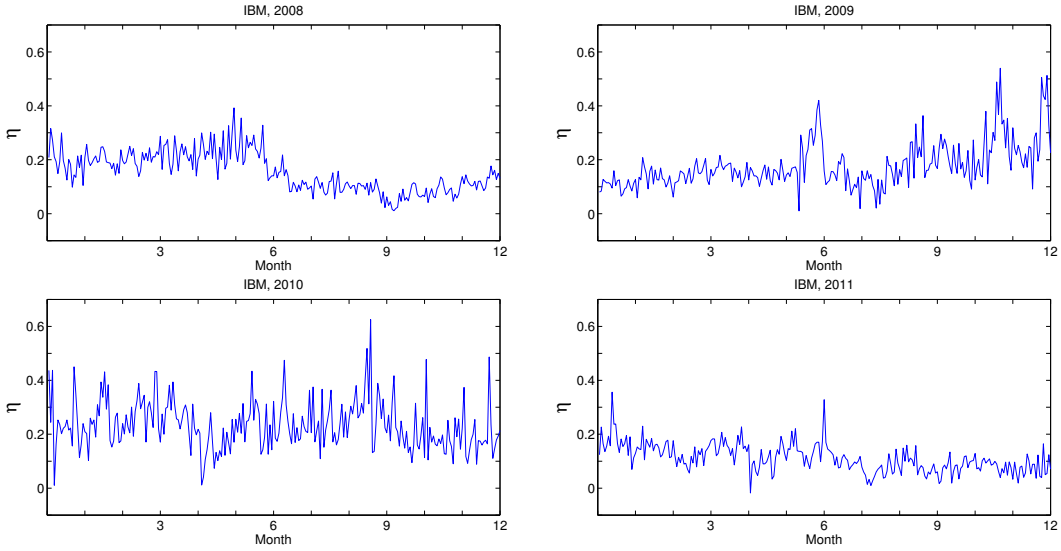


Figure 6: The estimation results of η, IBM, 2008-2011

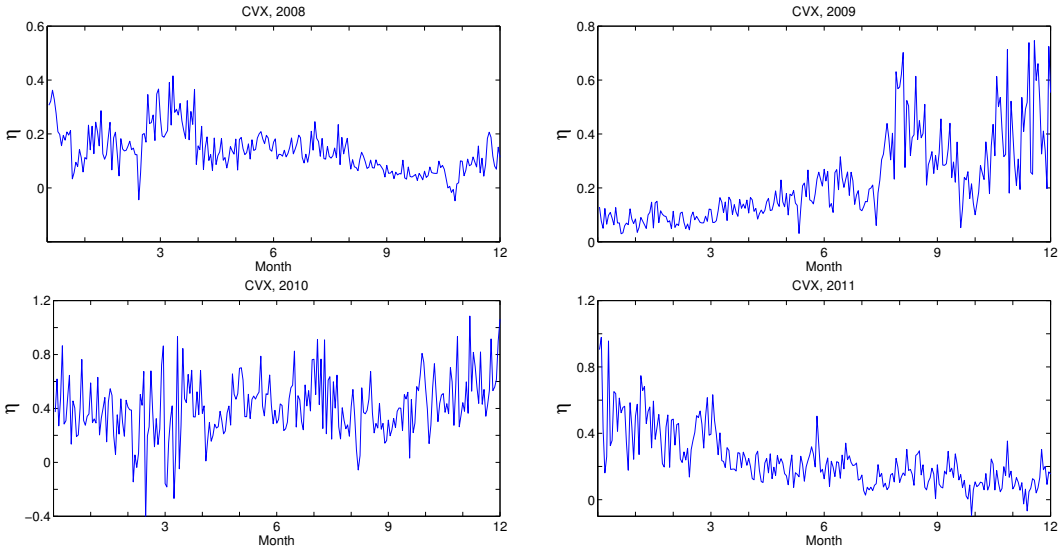
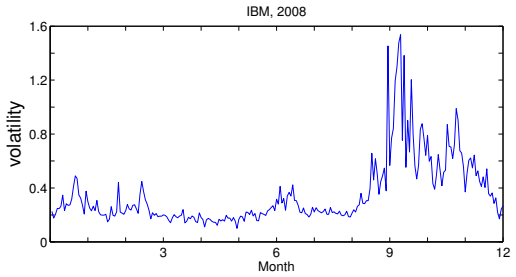
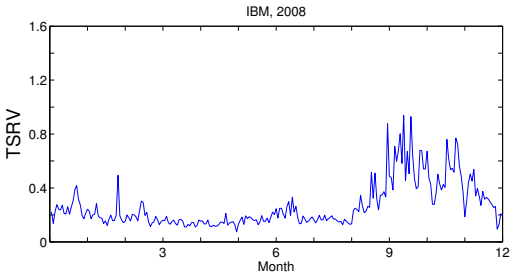


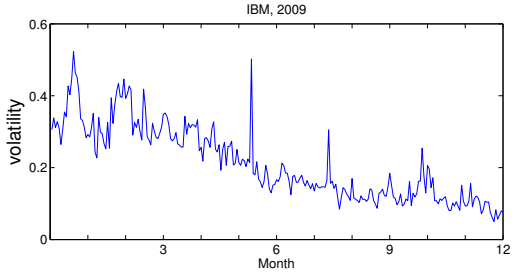
Figure 7: The estimation results of η, CVX, 2008-2011



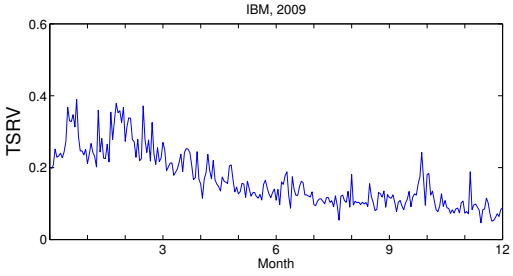
(a) Hawkes volatility, 2008



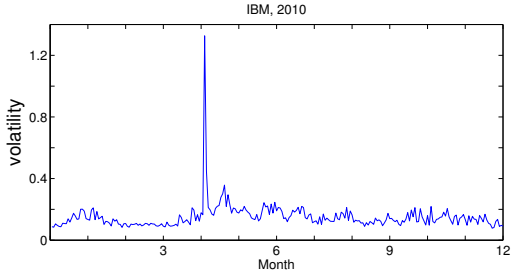
(b) TSRV, 2008



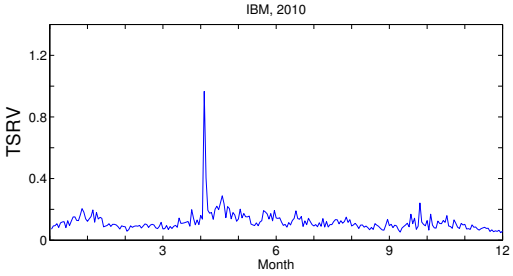
(c) Hawkes volatility, 2009



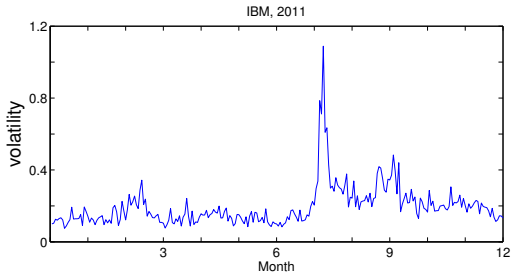
(d) TSRV, 2009



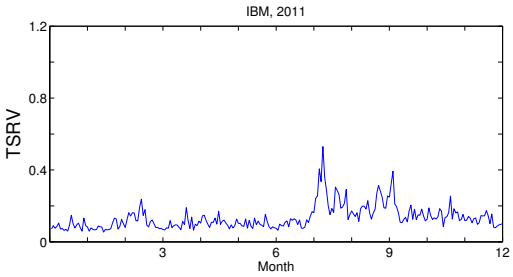
(e) Hawkes volatility, 2010



(f) TSRV, 2010



(g) Hawkes volatility, 2011



(h) TSRV, 2011

Figure 8: The estimation results of volatility, IBM, 2008-2011

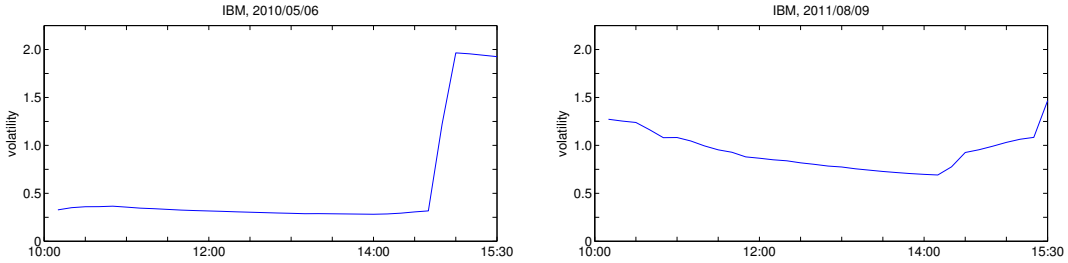


Figure 9: Cumulative intraday volatitliy estimated under the marked Hawkes model with every ten minutes update

to provide a reliable result for the estimation in the aspect of sample size and hence adequate intraday analyses are possible. The left of Figure 9 presents the intraday variation of the volatility of IBM at May 6, 2010, the day of the Flash Crash. The estimation is performed in ten minute basis for 10:00 to 15:30 and the cumulative volatility is plotted in the figure. The abrupt increase of the volatility is observed between 14:30 and 15:00 when the stock market crashed and the volatility is stabilized after 15:00. The right of the figure plots the intraday U-shape pattern of the volatility of IBM at August 9, 2011.

5. Concluding remark

We have developed a marked Hawkes model for price tick dynamics in equity markets. A linear impact function has been employed to describe the future effects of price jumps. We have not assumed a specific distribution for the jump size but used empirical distribution for estimation. Our 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 realize volatility are almost the same in the symmetric cases and the Hawkes volatility has less standard error. However, there are biases when the underlying path is not symmetric or the parameters are time varying.

We observed the significant positive linear impact, roughly around 0.2, and various type, linear or humped shape, of conditional mean structure of mark size in the empirical studies based on the equity prices reported in NYSE. The Hawkes model is useful to estimate the intraday volatility especially when the volatility is time varying. We observed the U shape seasonality of the changing volatility and also examined the interesting example of Flash Crash. As we discussed in the simulation and empirical studies, the discrepancy between the Hawkes volatility and realized volatility and the biases from sample volatility will be the important subject for the future study. In the presence of the asymmetry in the price dynamics and the time varying parameters, more robust estimation method is required for more exact volatility computation.

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

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. 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.

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. (2014a). Clustering and mean reversion in a Hawkes microstructure model. *Journal of Futures Markets*, pages n/a–n/a.

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

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

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.

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.

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.

Appendix 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 semimartingales X and Y for $t \geq 0$ is defined by

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

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

$$[X]_t := [X,X]_t = X_0^2 + \sum_{0 \leq s < t} (\Delta X_s)^2 \tag{A.1}$$

and

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

Appendix A.1. Step 1

In this subsection, we derive 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 a 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 r.h.s. of the above equation is represented by

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

or heuristically, we can write

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

Since the jump size of λ_{g1} is represented by $\alpha.(1+(k_1-1)\eta)$, using Eq. (A.1), we have

$$\begin{aligned}\mathbb{E}[[\lambda_{g1}]_t] &= \mathbb{E}[\lambda_{g1}^2(0)] + \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} \alpha_s^2(1+(k_1-1)\eta)^2 N_1(\mathrm{d}u\times\mathrm{d}k_1)\right] \\ &\quad + \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} \alpha_c^2(1+(k_2-1)\eta)^2 N_2(\mathrm{d}u\times\mathrm{d}k_2)\right] \\ &= \mathbb{E}[\lambda_{g1}^2(t)] + (\alpha_s^2+\alpha_c^2)\{1+2(K-1)\eta+(K^{(2)}-2K+1)\eta^2\}\mathbb{E}[\lambda_{g1}(t)]t\end{aligned}$$

where we use the stationarity of λ_{gi}. In addition,

$$\begin{aligned}&\mathbb{E}\left[\int_0^t \lambda_{g1}(u)\mathrm{d}\lambda_{g1}(u)\right] = \mathbb{E}\left[\int_0^t \lambda_{g1}(u)\beta(\mu-\lambda_{g1}(u))\mathrm{d}u\right] \\ &\quad + \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} q_s\beta g(k_1)\lambda_{g1}(u)N_1(\mathrm{d}u\times\mathrm{d}k_1) + \int_{(0,t)\times\mathbb{Z}^+} q_c\beta g(k_2)\lambda_{g1}(u)N_2(\mathrm{d}u\times\mathrm{d}k_2)\right] \\ &= (\beta\mu\mathbb{E}[\lambda_{g1}(t)] + [\alpha_s\{1+(K_{1\lambda_{g1}^2}-1)\eta\}-\beta]\mathbb{E}[\lambda_{g1}^2(t)] + \alpha_c\{1+(K_{1\lambda_{g1}\lambda_{g2}}-1)\eta\}\mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)])t \\ &=: (\beta\mu\mathbb{E}[\lambda_{g1}(t)] + (\breve{\alpha}_s-\beta)\mathbb{E}[\lambda_{g1}^2(t)] + \tilde{\alpha}_c\mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)])t.\end{aligned}$$

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 compute

$$\begin{aligned}\mathbb{E}[[\lambda_{g1},\lambda_{g2}]_t] &= \mathbb{E}[\lambda_{g1}(0)\lambda_{g2}(0)] + \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} \alpha_s\alpha_c(1+(k_1-1)\eta)^2 N_1(\mathrm{d}u\times\mathrm{d}k_1)\right] \\ &\quad + \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} \alpha_s\alpha_c(1+(k_2-1)\eta)^2 N_2(\mathrm{d}u\times\mathrm{d}k_2)\right] \\ &= \mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)] + 2\alpha_s\alpha_c\{1+2(K-1)\eta+(K^{(2)}-2K+1)\eta^2\}\mathbb{E}[\lambda_{g1}(t)]t\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} \lambda_{g1}(u)\mathrm{d}\lambda_{g2}(u)\right] = \mathbb{E}\left[\int_0^t \lambda_{g1}(u)\beta(\mu-\lambda_{g2}(u))\mathrm{d}u\right] \\ &\quad + \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} q_c\beta g(k_1)\lambda_{g1}(u)N_1(\mathrm{d}u\times\mathrm{d}k_1) + \int_{(0,t)\times\mathbb{Z}^+} q_s\beta g(k_2)\lambda_{g1}(u)N_2(\mathrm{d}u\times\mathrm{d}k_2)\right] \\ &= (\beta\mu\mathbb{E}[\lambda_{g1}(t)] + \alpha_c\{1+(K_{1\lambda_{g1}^2}-1)\eta\}\mathbb{E}[\lambda_{g1}^2(t)] + [\alpha_s\{1+(K_{1\lambda_{g1}\lambda_{g2}}-1)\eta\}-\beta]\mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)])t \\ &=: (\beta\mu\mathbb{E}[\lambda_{g1}(t)] + \breve{\alpha}_c\mathbb{E}[\lambda_{g1}^2(t)] + (\tilde{\alpha}_s-\beta)\mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)])t\end{aligned}$$

Thus,

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

where $\ddot{K} = 1+(K-1)\eta+(K^{(2)}-2K+1)\eta^2$ and

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

Appendix A.2. Step 2

We have

$$\mathbb{E}[\lambda_{g1}(t)N_1(t)] = \mathbb{E}[[\lambda_{g1},N_1]_t] + \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} \lambda_{g1}(u)k_1N_1(\mathrm{d}u\times\mathrm{d}k_1)\right] + \mathbb{E}\left[\int_0^t N_1(u-)\mathrm{d}\lambda_{g1}(u)\right]$$

and

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

and

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

and

$$\mathbb{E}\left[\int_0^t N_1(u-)\mathrm{d}\lambda_1(u)\right] = \int_0^t \{\beta\mu\mathbb{E}[\lambda_{g1}(u)]u + (\breve{\alpha}_s-\beta)\mathbb{E}[\lambda_{g1}(u)N_1(u)] + \breve{\alpha}_c\mathbb{E}[\lambda_{g2}(u)N_1(u)]\}\mathrm{d}u.$$

Similarly,

$$\mathbb{E}[\lambda_{g1}(t)N_2(t)] = \mathbb{E}[[\lambda_{g1}, N_2]_t] + \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} \lambda_{g1}(u)k_2 N_2(\mathrm{d}u\times\mathrm{d}k_2)\right] + \mathbb{E}\left[\int_0^t N_2(u-)\mathrm{d}\lambda_1(u)\right]$$

and

$$\begin{aligned}\mathbb{E}[[\lambda_{g1}, N_2]_t] &= \alpha_c(K+(K^{(2)}-K)\eta)\mathbb{E}[\lambda_{g1}(t)]t\\ &=: \alpha_c\dot{K}\mathbb{E}[\lambda_{g1}(t)]t\end{aligned}$$

and

$$\mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} \lambda_{g1}(u)k_2 N_2(\mathrm{d}u\times\mathrm{d}k)\right] = \int_0^t \mathbb{E}[k_2\lambda_{g1}(u)\lambda_{g2}(u)]\mathrm{d}u = K_{1\lambda_{g1}\lambda_{g2}}\mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)]t$$

and

$$\mathbb{E}\left[\int_0^t N_2(u-)\mathrm{d}\lambda_1(u)\right] = \int_0^t \{\beta\mu\mathbb{E}[\lambda_{g1}(u)]u + \tilde{\alpha}_c\mathbb{E}[\lambda_{g1}(u)N_1(u)] + (\tilde{\alpha}_s-\beta)\mathbb{E}[\lambda_{g1}(u)N_2(u)]\}\mathrm{d}u$$

Thus,

$$\begin{aligned}\begin{bmatrix}\dfrac{\mathrm{d}\mathbb{E}[\lambda_{g1}(t)N_1(t)]}{\mathrm{d}t}\\ \dfrac{\mathrm{d}\mathbb{E}[\lambda_{g1}(t)N_2(t)]}{\mathrm{d}t}\end{bmatrix} &= \mathbf{M}\begin{bmatrix}E[\lambda_{g1}(t)N_1(t)]\\ \mathbb{E}[\lambda_{g1}(t)N_2(t)]\end{bmatrix} + \begin{bmatrix}\alpha_s\dot{K}\mathbb{E}[\lambda_{g1}(t)] + K_{1\lambda_{g1}^2}\mathbb{E}[\lambda_{g1}^2(t)] + \beta\mu\mathbb{E}[\lambda_{g1}(t)]t\\ \alpha_c\dot{K}\mathbb{E}[\lambda_{g1}(t)] + K_{1\lambda_{g1}\lambda_{g2}}\mathbb{E}[\lambda_{g1}(t)\lambda_{g2}(t)] + \beta\mu\mathbb{E}[\lambda_{g1}(t)]t\end{bmatrix}\\ &= \mathbf{M}\begin{bmatrix}E[\lambda_{g1}(t)N_1(t)]\\ \mathbb{E}[\lambda_{g1}(t)N_2(t)]\end{bmatrix} + \mathbb{E}[\lambda_{g1}(t)]\left(\begin{bmatrix}\beta\mu\\ \beta\mu\end{bmatrix}t + \begin{bmatrix}\alpha_s\dot{K}\\ \alpha_c\dot{K}\end{bmatrix} - \frac{1}{2}\mathbf{M}^{-1}\begin{bmatrix}K_{1\lambda_{g1}^2}\{(\alpha_s^2+\alpha_c^2)\ddot{K}+2\beta\mu\}\\ 2K_{1\lambda_{g1}\lambda_{g2}}(\alpha_s\alpha_c\ddot{K}+\beta\mu)\end{bmatrix}\right)\end{aligned}$$

and

$$\begin{bmatrix}E[\lambda_{g1}(t)N_1(t)]\\ \mathbb{E}[\lambda_{g1}(t)N_2(t)]\end{bmatrix} = -\mathbb{E}[\lambda_{g1}(t)]\left(\beta\mu\mathbf{M}^{-1}\begin{bmatrix}1\\1\end{bmatrix}t + \mathbf{M}^{-1}\begin{bmatrix}\alpha_s\dot{K}\\ \alpha_c\dot{K}\end{bmatrix} - \frac{1}{2}\mathbf{M}^{-2}\begin{bmatrix}K_{1\lambda_{g1}^2}\{(\alpha_s^2+\alpha_c^2)\ddot{K}+2\beta\mu\}\\ 2K_{1\lambda_{g1}\lambda_{g2}}(\alpha_s\alpha_c\ddot{K}+\beta\mu)\end{bmatrix}\right).$$

Appendix A.3. Step 3

We have

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

and

$$\mathbb{E}[N_1(t)N_2(t)] = \mathbb{E}[[N_1,N_2]_t] + \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} N_1(u-)N_2(\mathrm{d}u\times\mathrm{d}k)\right] + \mathbb{E}\left[\int_{(0,t)\times\mathbb{Z}^+} N_2(u-)N_1(\mathrm{d}u\times\mathrm{d}k)\right]$$
$$= 2\int_0^t K_{1\lambda_{g1}N_2}\mathbb{E}[\lambda_{g1}(u)N_2(u)]\mathrm{d}u.$$

Then

$$\begin{bmatrix} E[N_1^2(t)] \\ \mathbb{E}[N_1(t)N_2(t)] \end{bmatrix} = -\mathbb{E}[\lambda_{g1}(t)] \begin{bmatrix} K_{1\lambda_{g1}N_1} & 0 \\ 0 & K_{1\lambda_{g1}N_2} \end{bmatrix}$$
$$\left\{\beta\mu\mathbf{M}^{-1}\begin{bmatrix}1\\1\end{bmatrix}t^2 + \left(2\mathbf{M}^{-1}\begin{bmatrix}\alpha_s\dot{K}\\ \alpha_c\dot{K}\end{bmatrix} - \mathbf{M}^{-2}\begin{bmatrix}K_{1\lambda_{g1}^2}\{(\alpha_s^2+\alpha_c^2)\ddot{K}+2\beta\mu\}\\ 2K_{1\lambda_{g1}\lambda_{g2}}(\alpha_s\alpha_c\ddot{K}+\beta\mu)\end{bmatrix} - \begin{bmatrix}K^{(2)}/K_{1\lambda_{g1}N_1}\\ 0\end{bmatrix}\right)t\right\}.$$