Multi-kernel property in high-frequency price dynamics under Hawkes model

Kyungsub Lee[*]

Abstract

This study investigates and uses multi-kernel Hawkes models to describe a high-frequency mid-price process. Each kernel represents a different responsive speed of market participants. Using the conditional Hessian, we examine whether the numerical optimizer effectively finds the global maximum of the log-likelihood function under complicated modeling. Empirical studies that use stock prices in the US equity market show the existence of multi-kernels classified as ultra-high-frequency (UHF), very-high-frequency (VHF), and high-frequency (HF). We estimate the conditional expectations of arrival times and the degree of contribution to the high-frequency activities for each kernel.

1 Introduction

In recent years, interest in financial activities under the high-frequency level has increased as time records for financial market transactions, and quotes are achieved at higher time resolutions. Intraday events in major stock markets were recorded in milliseconds previously, and then the resolution gradually increases to microseconds and nanoseconds. High-frequency financial data comprise the actions of ultra-high-frequency and high-frequency traders as well as low-frequency participants in financial markets.

The Hawkes process (Hawkes, 1971), which has been applied to the natural and social sciences, captures the jump increment, called excitement, in intensities originating from past events and decays of the effect. Activities in financial markets under a tick structure, where prices vary tick-by-tick, such as transactions, quote revisions, and cancellations are activated by past events; their aftermath effects diminish over time. Due to this commonality, the Hawkes model has been actively used as a financial model in recent years. For the earliest studies, Bowsher (2007) proposed a generalized Hawkes model for multivariate financial market events, which incorporates various dependencies in market features, and Large (2007) utilized a mutually exciting Hawkes model to describe the resiliency in the limit order book.

Although the Hawkes model can be used in various fields of finance, such as credit risk analysis (Errais et al., 2010; Aït-Sahalia et al., 2015; Ma and Xu, 2016; Ketelbuters and Hainaut, 2022) and optimal execution (Choi et al., 2021; Da Fonseca and Malevergne, 2021; Jusselin, 2021; Gašperov and Kostanjčar, 2022), studies focusing on price dynamics, such as our study, are also abundant. To mention a few, Bacry et al. (2013a) used the multivariate Hawkes model for the tick dynamics of asset prices with the applications of microstructure noise analysis. Under settings similar to the standard Hawkes models, studies by Bacry and

---
[*]Department of Statistics, Yeungnam University, Gyeongsan, Gyeongbuk 38541, Republic of Korea, Corresponding author, Email: ksublee@yu.ac.kr

Muzy (2014); Da Fonseca and Zaatour (2014b, 2017); Arouri et al. (2019); Ji et al. (2020) contributed to the literature in aspects such as stylized facts of market impact, moment analysis, lead-lag relationship, capturing correlated jumps, extreme risk, and so on.

In addition, Hawkes models of a modified and extended form suitable for more complex structures have also been proposed. For example, to describe the bid and ask price dynamics, Zheng et al. (2014) introduced a multivariate constraint Hawkes-like point process and Lee and Seo (2022) proposed an extended Hawkes model to capture the spread dependent intensities. Jang et al. (2020) suggested the Hawkes flocking model to assess the systemic risk in crude oil and gasoline futures markets. Morariu-Patrichi and Pakkanen (2022) examined state-dependent Hawkes processes with an application to limit order book dynamics. Swishchuk (2021) proposed general regime-switching Hawkes models to describe the price processes in the limit order books. Zhang et al. (2021) suggested a new model by combining a Hawkes process and a finite-range contact process for stock price movements.

For reviews of other interesting studies on the Hawkes model with financial applications that are been mentioned here, please refer to Law and Viens (2015), Bacry et al. (2015), and Hawkes (2018).

We focus on the multi-kernel property in the high-frequency mid-price process in stock and futures markets. The mid-price, the mean of the best bid and ask prices, moves in multiples of the minimum tick in market microstructure. As mentioned before, many studies have modeled price movement in the tick structure using the Hawkes model. Among these, a single kernel is mainly used, as in Bacry et al. (2013b); Da Fonseca and Zaatour (2014a); Lee and Seo (2017b) or power-law kernel, as in Hardiman et al. (2013); Filimonov and Sornette (2015); Bacry et al. (2016). However, relatively few studies have been conducted on multi-kernel exponential models.

The intraday movements of the stock price are significantly influenced by ultra-high-frequency traders. However, there are also numerous activities that are performed in relatively lower frequencies. The responsiveness to previous events differs depending on their computing power, network speed, and available services such as colocation. The multi-kernel Hawkes model can capture the different responsive speeds and categorize the different types of activities in financial activities. We examine the differences in responsiveness among traders verified by multiple kernels.

As a basic property, we derive the first and second moments under the bi-variate multi-kernel Hawkes model, which is an extension of the results in a single kernel model, such as in Da Fonseca and Zaatour (2014b), Lee and Seo (2017b), and Cui et al. (2020). The moment expression is a closed-form formula and it helps verify the accuracy of the simulation of the multi-kernel Hawkes model.

We carefully examine the estimation procedure because the number of parameters increases as the number of kernels increases. We discuss the property of the global maximum of the log-likelihood based on a conditional Hessian to examine the reliability of the multivariate and multi-kernel model estimation. Various simulation examples are also presented.

The remainder of this paper is organized as follows. Section 2 proposes the multi-kernel Hawkes model for mid-price processes and discusses the global maximum property and moments. Section 3 present an empirical study based on high-frequency stock prices and the filtering effect. Finally, Section 4 concludes the paper.

2 Model

2.1 Multi-kernel model in high-frequency data

The mid-price processes of financial assets exhibit typical signs of high-frequency trading. The price reacts rapidly to previous events and the impact dissipates rapidly. However, market participants may have different processing speeds. We separate the kernels based on the frequency spectrum by introducing multi-kernels in the Hawkes model. It has often been argued that ultra-high-frequency financial raw data recorded at a high-resolution are not adequately suited to the single-kernel exponential Hawkes model. Note that stock movements have been recorded in nanoseconds in major stock exchanges since approximately 2015. Given that there are few in-depth studies on the financial application of the multi-kernel exponential Hawkes model, through this study, we intend to identify the advantages of the multi-kernel model as well as its disadvantages.

Consider a bi-variate counting process and its intensities:

$$\boldsymbol{N}_t = \begin{bmatrix} N_1(t) \\ N_2(t) \end{bmatrix}, \quad \boldsymbol{\lambda}_t = \begin{bmatrix} \lambda_1(t) \\ \lambda_2(t) \end{bmatrix}$$

where $N_1(t)$ and $N_2(t)$ count the number of the up and down movements of a mid-price process up to t, respectively. Thus, the mid-price process can be represented as the difference between the two counting processes. In this study, bold face letters indicate vectors or matrices, as we deal with the bi-variate model. Let the vector of the intensity process be

$$\boldsymbol{\lambda}_t = \boldsymbol{\mu} + \int_{-\infty}^{t} \boldsymbol{h}(t-u)\mathrm{d}\boldsymbol{N}_u \tag{1}$$

where $\boldsymbol{\mu} = [\mu_1, \mu_2]^\top$ is constant.

This study focuses on the multi-kernel property, which is different from other studies on the single kernel model; that is

$$\boldsymbol{h} = \sum_{k=1}^{K} \boldsymbol{h}_k$$

where k denotes the kernel number. Each kernel is

$$\boldsymbol{h}_k(t) = \boldsymbol{\alpha}_k \circ \begin{bmatrix} \mathrm{e}^{-\beta_{k11}t} & \mathrm{e}^{-\beta_{k12}t} \\ \mathrm{e}^{-\beta_{k21}t} & \mathrm{e}^{-\beta_{k22}t} \end{bmatrix}.$$

where $\boldsymbol{\alpha}_k$ is a 2×2 matrix whose element in i-th row and j-th column are represented by α_{ijk}. The smaller the k, the higher the kernel frequency, hence the larger the β_k.

As the number of parameters increases proportionally with the number of kernels, for the sake of model parsimony, we can assume that βs are the same for each kernel or for each raw of a kernel depending on the context. Multiple kernels are designed to capture different responsiveness speeds for high-frequency activities in the financial market.

2.2 Moment property

In this section, we computes the moments using the bi-variate multi-kernel Hawkes model. The intensity process of the multi-kernel Hawkes process can be rewritten as

$$\boldsymbol{\lambda}_t = \boldsymbol{\mu} + \sum_{k=1}^{K} \boldsymbol{\lambda}_k(t) \tag{2}$$

where the components are defined by

$$\boldsymbol{\lambda}_k(t) = \begin{bmatrix} \lambda_{k1}(t) \\ \lambda_{k2}(t) \end{bmatrix} = \int_{-\infty}^{t} \boldsymbol{h}_k(t-u)\mathrm{d}\boldsymbol{N}_u.$$

To derive these formulas, we have additional restrictions on the Markov property:

$$\beta_{k1} := \beta_{k11} = \beta_{k12}, \quad \beta_{k2} := \beta_{k21} = \beta_{k22},$$

i.e.,

$$\boldsymbol{h}_k(t) = \boldsymbol{\alpha}_k \circ \begin{bmatrix} \mathrm{e}^{-\beta_{k1}t} & \mathrm{e}^{-\beta_{k1}t} \\ \mathrm{e}^{-\beta_{k2}t} & \mathrm{e}^{-\beta_{k2}t} \end{bmatrix}.$$

Following this, the differential form of the intensity process can be represented by

$$\mathrm{d}\boldsymbol{\lambda}_t = \sum_{k=1}^{K} \mathrm{d}\boldsymbol{\lambda}_k(t), \tag{3}$$

$$\mathrm{d}\boldsymbol{\lambda}_k(t) = -\boldsymbol{\beta}_k\boldsymbol{\lambda}_k(t)\mathrm{d}t + \boldsymbol{\alpha}_k\mathrm{d}\boldsymbol{N}_t \tag{4}$$

where

$$\boldsymbol{\beta}_k = \begin{bmatrix} \beta_{k1} & 0 \\ 0 & \beta_{k2} \end{bmatrix}.$$

For the simplicity of the formula, we assume that the processes are in a steady state at time 0, that is, by considering the intensity process to begin at $-\infty$, we always assume that the unconditional distribution of the intensity processes at time 0 are in a steady state. Under this assumption, the expectation of the intensity processes and their components are constant.

Proposition 1. Under the steady state assumption, for $t > 0$,

$$\mathbb{E}[\boldsymbol{\lambda}_t] = \left(\mathbf{I} - \sum_{k=1}^{K} \boldsymbol{\beta}_k^{-1}\boldsymbol{\alpha}_k \right)^{-1} \boldsymbol{\mu},$$

$$\mathbb{E}[\boldsymbol{\lambda}_k(t)] = \boldsymbol{\beta}_k^{-1}\boldsymbol{\alpha}_k\mathbb{E}[\boldsymbol{\lambda}_t].$$

To proceed further, define

$$\boldsymbol{\Lambda}_t = \begin{bmatrix} \boldsymbol{\lambda}_1(t) \\ \vdots \\ \boldsymbol{\lambda}_K(t) \end{bmatrix}, \quad \boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_K \end{bmatrix}, \quad \boldsymbol{\beta} = \mathrm{Dg}([\beta_{11} \ \beta_{12} \ \cdots \ \beta_{K1} \ \beta_{K2}])$$

where $\mathrm{Dg}(\cdot)$ denotes a diagonal matrix whose diagonal entries are elements in the argument. Thus, Λ_t is a $2K \times 1$ matrix, and $\boldsymbol{\alpha}$ is $2K \times 2$ and $\boldsymbol{\beta}$ is a diagonal $2K \times 2K$ matrix. Following this,

$$\mathrm{d}\boldsymbol{\Lambda}_t = \boldsymbol{\beta}\boldsymbol{\Lambda}_t\mathrm{d}t + \boldsymbol{\alpha}\mathrm{d}\boldsymbol{N}_t.$$

4

Lemma 2. Consider 2×1 vector processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ such that
$$\mathrm{d}\boldsymbol{X}_t = \boldsymbol{a}_t \mathrm{d}t + \boldsymbol{f}_x(t)\mathrm{d}\boldsymbol{N}_t, \quad \mathrm{d}\boldsymbol{Y}_t = \boldsymbol{b}_t \mathrm{d}t + \boldsymbol{f}_y(t)\mathrm{d}\boldsymbol{N}_t.$$

Subsequently, the differential form of the quadratic variation matrix process of $\boldsymbol{X}\boldsymbol{Y}^\top$ is represented by
$$\mathrm{d}[\boldsymbol{X}\boldsymbol{Y}^\top]_t = \boldsymbol{f}_x(t)\mathrm{Dg}(\mathrm{d}\boldsymbol{N}_t)\boldsymbol{f}_y^\top(t).$$

Thus,
$$\mathrm{d}(\boldsymbol{X}_t\boldsymbol{Y}_t^\top) = \boldsymbol{X}_{t-}\mathrm{d}\boldsymbol{Y}_t^\top + \mathrm{d}(\boldsymbol{X}_t)\boldsymbol{Y}_{t-}^\top + \boldsymbol{f}_x(t)\mathrm{Dg}(\mathrm{d}\boldsymbol{N}_t)\boldsymbol{f}_y^\top(t).$$

In addition,
$$\frac{\mathrm{d}\mathbb{E}[\boldsymbol{X}_t\boldsymbol{Y}_t^\top]}{\mathrm{d}t} = \mathbb{E}[\boldsymbol{X}_{t-}(\boldsymbol{b}_t^\top + \boldsymbol{\lambda}_t^\top \boldsymbol{f}_y^\top(t))] + \mathbb{E}[(\boldsymbol{a}_t + \boldsymbol{f}_x(t)\boldsymbol{\lambda}_t)\boldsymbol{Y}_{t-}^\top] + \boldsymbol{f}_x(t)\mathrm{Dg}(\mathbb{E}[\boldsymbol{\lambda}_t])\boldsymbol{f}_y^\top(t).$$

Proposition 3. Under the steady state assumption, $\mathbb{E}[\boldsymbol{\Lambda}_t\boldsymbol{\Lambda}_t^\top]$ satisfies the following Sylvester equation:
$$(\boldsymbol{\beta} - \boldsymbol{\alpha}\mathbf{J})\mathbb{E}[\boldsymbol{\Lambda}_t\boldsymbol{\Lambda}_t^\top] + \mathbb{E}[\boldsymbol{\Lambda}_t\boldsymbol{\Lambda}_t^\top](\boldsymbol{\beta} - \mathbf{J}^\top\boldsymbol{\alpha}^\top) = \boldsymbol{\alpha}\boldsymbol{\mu}\mathbb{E}[\boldsymbol{\Lambda}_t^\top] + \mathbb{E}[\boldsymbol{\Lambda}_t]\boldsymbol{\mu}^\top\boldsymbol{\alpha}^\top + \boldsymbol{\alpha}\mathrm{Dg}(\mathbb{E}[\boldsymbol{\lambda}_t])\boldsymbol{\alpha}^\top \quad (5)$$

where $\mathbf{J}$ is a $2 \times 2K$ matrix such that
$$\mathbf{J} = \begin{bmatrix} 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & \cdots & 0 & 1 \end{bmatrix}.$$

Remark 4. Solving Eq. (5) is equivalent to solving
$$(\mathbf{I}\otimes(\boldsymbol{\beta}-\boldsymbol{\alpha}\mathbf{J})+(\boldsymbol{\beta}-\boldsymbol{\alpha}\mathbf{J})\otimes\mathbf{I})\mathrm{vec}\left(\mathbb{E}[\boldsymbol{\Lambda}_t\boldsymbol{\Lambda}_t^\top]\right) = \mathrm{vec}\left(\boldsymbol{\alpha}\boldsymbol{\mu}\mathbb{E}[\boldsymbol{\Lambda}_t^\top] + \mathbb{E}[\boldsymbol{\Lambda}_t]\boldsymbol{\mu}^\top\boldsymbol{\alpha}^\top + \boldsymbol{\alpha}\mathrm{Dg}(\mathbb{E}[\boldsymbol{\lambda}_t])\boldsymbol{\alpha}^\top\right)$$

where $\otimes$ denotes the Kronecker product, and vec is the vectorization operator.

Proposition 5. We have
$$\mathbb{E}[\boldsymbol{\lambda}_t\boldsymbol{N}_t^\top] \approx \mathbf{A}t + \mathbf{B}$$

where
$$\mathbf{A} = \left(\mathbf{I} - \sum_{k=1}^K \boldsymbol{\beta}_k^{-1}\boldsymbol{\alpha}_k\right)^{-1}\boldsymbol{\mu}\mathbb{E}[\boldsymbol{\lambda}_t^\top] = \mathbb{E}[\boldsymbol{\lambda}_t]\mathbb{E}[\boldsymbol{\lambda}_t^\top] \quad (6)$$

$$\mathbf{B} = \left(\mathbf{I} - \sum_{k=1}^K \boldsymbol{\beta}_k^{-1}\boldsymbol{\alpha}_k\right)^{-1}\sum_{k=1}^K \boldsymbol{\beta}_k^{-1}\left(-\boldsymbol{\beta}_k^{-1}\boldsymbol{\alpha}_k\mathbf{A} + \mathbb{E}[\boldsymbol{\lambda}_k(t)\boldsymbol{\lambda}_t^\top] + \boldsymbol{\alpha}_k\mathrm{Dg}(\mathbb{E}[\boldsymbol{\lambda}_t])\right) \quad (7)$$

Remark 6. Note that
$$\boldsymbol{\lambda}_k(t)\boldsymbol{\lambda}_t^\top = \mathbf{J}_k\boldsymbol{\Lambda}_t(\boldsymbol{\mu} + \mathbf{J}\boldsymbol{\Lambda}_t)^\top$$

where $\mathbf{J}_k$ is a 2 by $2K$ matrix such that
$$[\mathbf{J}_k]_{ij} = \begin{cases} 1, & \text{if } i = 1 \text{ and } j = 2k-1 \text{ or } i = 2 \text{ and } j = 2k \\ 0, & \text{otherwise.} \end{cases}$$

Proposition 7. Under the steady state condition,
$$\mathbb{E}[\boldsymbol{N}_t\boldsymbol{N}_t^\top] = \mathbf{A}t^2 + \left(\mathbf{B} + \mathbf{B}^\top + \mathrm{Dg}(\mathbb{E}[\boldsymbol{\lambda}_t])\right)t$$

where $\mathbf{A}$ and $\mathbf{B}$ are defined by Eqs. (6) and (7), respectively.

Proposition 8. Under simple calculation, we have
$$\mathrm{Var}(N_1(t) - N_2(t)) = 2\begin{bmatrix} 1 & -1 \end{bmatrix}\mathbf{B}\begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} 1 & 1 \end{bmatrix}\mathbb{E}[\boldsymbol{\lambda}_t].$$

2.3 Conditional concavity

Multi-kernel Hawkes models are inferred by the maximum likelihood estimation (MLE). It is worthwhile to verify whether the numerical procedure of MLE is performed properly because of the complexity of the model. We discuss the issue on finding the global maximum of the log-likelihood function in terms of the conditional concavity introduced by Lee and Seo (2017a) in a single-kernel framework.

Consider a general case of the Hawkes model:

$$\boldsymbol{N}_t = \begin{bmatrix} N_1(t) \\ \vdots \\ N_m(t) \end{bmatrix}, \quad \boldsymbol{\lambda}_t = \begin{bmatrix} \lambda_1(t) \\ \vdots \\ \lambda_m(t) \end{bmatrix}$$

where the intensities follow the m-dimensional version of Eq. (1). The component of $\boldsymbol{h}$ is

$$h_{ij}(t) = \sum_{k=1}^{K} \alpha_{ijk} e^{-\beta_{ijk}t}.$$

The log-likelihood function is

$$L(\mu, \alpha, \beta) = \sum_{i=1}^{m} \left(\sum_{n=1}^{N_i(T)} \log \lambda_i(t_{i,n}) - \int_0^T \lambda_i(u) \mathrm{d}u \right) \tag{8}$$

where $\{t_{i,n}\}$ are the n-th jump times of N_i. We rewrite

$$\lambda_i(u) = \mu_i + \sum_{j=1}^{m} \sum_{k=1}^{K} \sum_{n=1}^{N_j(T)} \alpha_{ijk} e^{-\beta_{ijk}(u-t_{j,n})} \mathbb{1}_{\{u>t_{j,n}\}} + \varepsilon_i(u) \tag{9}$$

$$\approx \mu_i + \sum_{j=1}^{m} \sum_{k=1}^{K} \sum_{n=1}^{N_j(T)} \alpha_{ijk} e^{-\beta_{ijk}(u-t_{j,n})} \mathbb{1}_{\{u>t_{j,n}\}} \tag{10}$$

where

$$\varepsilon_i(u) = \sum_{j=1}^{n} \sum_{k=1}^{K} \int_{-\infty}^{0} \alpha_{ijk} e^{-\beta_{ijk}(u-s)} \mathrm{d}N_j(s)$$

which is the remaining impact on λ_i by the events which occur before time zero and this term can be ignored. If $\beta_i := \beta_{ijk}$ for all j and k, then,

$$\varepsilon_i(u) = (\lambda_i(0) - \mu_i) e^{-\beta_i u}.$$

If the system is Markovian, then the assumption that ε can be ignored is unnecessary.

By integrating Eq. (10),

$$\int_0^T \lambda_i(u) \mathrm{d}u \approx \mu_i T + \sum_{j=1}^{m} \sum_{k=1}^{K} \sum_{n=1}^{N_j(T)} b_{ijk,n} \alpha_{ijk}, \quad b_{ijk,n} := \frac{1 - e^{-\beta_{ijk}(T-t_{j,n})}}{\beta_{ijk}},$$

and by substituting the above in Eq. (8),

$$L(\mu,\alpha,\beta) \approx \sum_{i=1}^{m} \left\{ -\mu_i T + \sum_{n=1}^{N_i(T)} \left(\log \lambda_i(t_{i,n}) - \sum_{j=1}^{m}\sum_{k=1}^{K} b_{jik,n}\alpha_{jik} \right) \right\}$$

$$= \sum_{i=1}^{m} \left\{ -\mu_i T + \sum_{n=1}^{N_i(T)} \left(\log \left(\mu_i + \sum_{j=1}^{m}\sum_{k=1}^{K} a_{ijk,n}\alpha_{ijk} \right) - \sum_{j=1}^{m}\sum_{k=1}^{K} b_{jik,n}\alpha_{jik} \right) \right\}$$

where

$$a_{ijk,n} = \int_{-\infty}^{t_{i,n}} e^{-\beta_{ijk}(t_{i,n}-u)} dN_j(u)$$

and we use

$$\lambda_i(t_{i,n}) = \mu_i + \sum_{j=1}^{m}\sum_{k=1}^{K} \int_{-\infty}^{t_{i,n}} \alpha_{ijk} e^{-\beta_{ijk}(t_{i,n}-u)} dN_j(u).$$

We examine the conditional concavity of $L(\mu,\alpha|\beta)$ with fixed βs with the conditional Hessian. Let

$$L_{i,n}(\mu,\alpha|\beta) = \log \left(\mu_i + \sum_{j=1}^{m}\sum_{k=1}^{K} a_{ijk,n}\alpha_{ijk} \right) - \sum_{j=1}^{m}\sum_{k=1}^{K} b_{jik,n}\alpha_{jik},$$

then

$$L(\mu,\alpha|\beta) \approx \sum_{i=1}^{m} \left(-\mu_i T + \sum_{n=1}^{N_i(T)} L_{i,n}(\mu,\alpha|\beta) \right).$$

Since the second derivatives of $\mu_i T$ with respect to μ_is and α_{ij}s are all zeros, it is sufficient to check the second derivatives of $L_{i,n}(\mu,\alpha|\beta)$ to examine the Hessian of $L(\mu,\alpha|\beta)$. Note that

$$\frac{\partial^2 L_{i,n}}{\partial \mu_i^2} = -\frac{1}{\lambda_i^2(t_{i,n})}, \qquad \frac{\partial^2 L_{i,n}}{\partial \mu_i \partial \alpha_{ijk}} = -\frac{a_{ijk,n}}{\lambda_i^2(t_{i,n})},$$

$$\frac{\partial^2 L_{i,n}}{\partial \alpha_{ijk} \partial \alpha_{ij'k'}} = \begin{cases} -\dfrac{a_{ijk,n} a_{ij'k',n}}{\lambda_i^2(t_{i,n})}, & \text{if } j \neq j' \text{ and } k \neq k' \\ 0, & \text{otherwise.} \end{cases}$$

As the Hessian of $L_{i,n}(\mu,\alpha|\beta)$ is negative semidefinite, the Hessian of $L(\mu,\alpha|\beta)$ is and hence $L(\mu,\alpha|\beta)$ is concave. Due to the conditional concavity, a numerical optimizer can find

$$L^*(\beta) := \max_{\mu,\alpha} L(\mu,\alpha|\beta).$$

for each β (if it exists). Therefore, once we determine $L^*(\beta)$ for enough βs and if we can calculate the maximum of $L^*(\beta)$, it is possible to determine the global maximum of $L(\mu,\alpha,\beta)$. Given that this strategy takes a long time, it is only used in this study when examining the global maximum of the Hawkes log-likelihood. For the empirical studies, a typical quasi-Newton method is applied under the assumption that the numerical optimizer works well.

2.4 Simulation study

First, by comparing the theoretically calculated moment values derived in Subsection 2.2 with the sample mean obtained through simulation, we verify whether the simulation process is performed well. The simulation method is based on the thinning algorithm, which can generally be applied to nonhomogeneous Poisson processes (Lewis and Shedler, 1979).

Example 1. Under the three kernel model, with presumed parameters

$$\boldsymbol{\mu} = \begin{bmatrix} 0.0757 \\ 0.0757 \end{bmatrix}, \quad \boldsymbol{\alpha}_1 = \begin{bmatrix} 23.34 & 15.67 \\ 15.67 & 23.34 \end{bmatrix}, \quad \boldsymbol{\alpha}_2 = \begin{bmatrix} 6 & 9 \\ 9 & 6 \end{bmatrix}, \quad \boldsymbol{\alpha}_3 = \begin{bmatrix} 0.10 & 0.02 \\ 0.02 & 0.1 \end{bmatrix},$$

$$\boldsymbol{\beta}_1 = \begin{bmatrix} 140 & 0 \\ 0 & 140 \end{bmatrix}, \quad \boldsymbol{\beta}_2 = \begin{bmatrix} 30 & 0 \\ 0 & 30 \end{bmatrix}, \quad \boldsymbol{\beta}_3 = \begin{bmatrix} 0.8 & 0 \\ 0 & 0.8 \end{bmatrix},$$

and $t = 1,000$, by Propositions 1 and 7, we have

$$\mathbb{E}[\boldsymbol{N}_t] = \mathbb{E}[\boldsymbol{\lambda}_t]t = \begin{bmatrix} 1,059.8 \\ 1,059.8 \end{bmatrix}, \quad \mathbb{E}[\boldsymbol{N}_t \boldsymbol{N}_t^\top] = \begin{bmatrix} 1,227,649 & 1,226,463 \\ 1,226,463 & 1,227,649 \end{bmatrix}.$$

Our simulation with $10,000$ generated paths shows that the sample means

$$\overline{N_1(t)} = 1,058, \quad \overline{N_1^2(t)} = 1,218,060, \quad \overline{N_1(t)N_2(t)} = 1,215,964$$

which confirms that the simulation process works well.

Next, we examine the estimation performance of the model. The concavity of the log-likelihood depends on the sample size and the branching ratio. If the number of observations is insufficient and the branching ratio or α is close to zero, then the log-likelihood function is probably not concave. The log-likelihood function is likely concave with sufficient sample size and significant α, Note that the Hawkes process with insignificant α compared with β is very close to the homogeneous Poisson; it is evident that the model with insignificant α is not well estimated. The simulation study shows that the more significant α, the lower the sample size that is needed and if α is close to zero, the estimation could be incorrect.
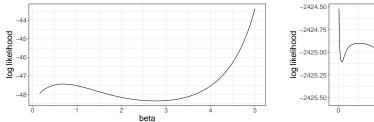
Figure 1 illustrates examples of non-concave log-likelihood functions in a univariate Hawkes model. The left represents the small sample size case with parameters $\alpha = 0.15, \beta = 1.0$ and 10 observations, while the right represents the case of insignificant $\alpha = 0.001$ with $\beta = 1.0$ and sample size 500. Using a numerical optimizer, we determine the $L^*(\beta)$ for each β on a sufficiently dense set. As non-concavities are observed in the figure, the numerical optimizer may fail to find the global maximum.

A small α value can cause an inaccurate estimation for β. In a multivariate Hawkes model, if βs are considered insignificantly different from each other, it would be preferable to assume that βs are equal in the model for parsimony. For example, consider the following model:

$$\begin{bmatrix} \lambda_1(t) \\ \lambda_2(t) \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.2 \end{bmatrix} + \int_{(0,t]} \begin{bmatrix} 0.5e^{-2.1(t-u)} & 0.001e^{-2.3(t-u)} \\ 0.9e^{-2.2(t-u)} & 0.5e^{-2.0(t-u)} \end{bmatrix} \begin{bmatrix} dN_1(u) \\ dN_2(u) \end{bmatrix},$$

where $\alpha_{12} = 0.001$ is close to zero, and β_{ij}s differ insignificantly. With a simulated path with 5,000 observations, the estimates of βs are

$$\begin{bmatrix} \hat{\beta}_{11} & \hat{\beta}_{12} \\ \hat{\beta}_{21} & \hat{\beta}_{22} \end{bmatrix} = \begin{bmatrix} 2.28 & 0.14 \\ 2.33 & 2.35 \end{bmatrix}$$

Figure 1: Example of numerically computed conditional maximum of log-likelihood function, L^*, for each β with simulated paths, small sample size case (left) and insignificant α case (right)

where $\hat{\beta}_{12}$ is inaccurate. With a constraint of all equal β_{ij}, the estimate of $\beta_{11} = \beta_{12} = \beta_{21} = \beta_{22}$ is 2.3048, which is much more reliable.

We conduct additional experiments to test the relationship between the sample size and the uniqueness of the local maximum. For computational complexity, the experiment is based on a one-dimensional Hawkes model. However, it is important to demonstrate that the sample size is crucial.

A path is simulated for each of the 0.1, 0.5, and 0.9 branching ratios. A numerical optimizer finds $L^*(\beta)$ for all βs over a sufficiently dense set. Examining computed $L^*(\beta)$ makes it possible to check the uniqueness of local maximum. We assume that the numerical optimizer would ultimately succeed in finding the global maximum of L and count it a success, if there is a unique local maximum in L^*. Otherwise, we cannot be sure of its success and count this as a failure. Figure 2 shows success rates after repeating 100 experiments. As mentioned earlier, the lowest success rate is observed when the branching ratio is low. However, with more than 150 samples, the rate is greater than 90%. We use more than 10,000 observations in our empirical study. Furthermore, as the branching ratios are generally significant, we assume that the numerical results are reliable.
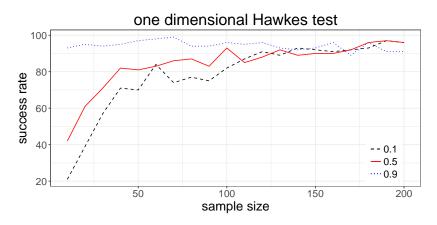

Figure 2: The success rate of estimation to find the global maximum depending on the branching ratio and sample size
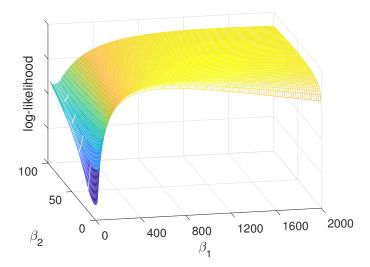
Figure 3: Example of the conditional log-likelihood of the two-kernel Hawkes model as a function of βs

3 Empirical result

3.1 Basic result

We conduct empirical studies using high-frequency stock price data under the US equity and futures market from 2016[1] to 2019. The data comprise of the timestamps of changes in the national best bid and offer (NBBO) prices, derived from the trade and quote (TAQ) data released by the New York Stock Exchange.

Before conducting the main analysis, we visualize $L^*(\beta_1, \beta_2)$ under the two-kernel model for IBM on January 3, 2017 as shown in Figure 3. The global maximum is located at $\beta_1 = 880$ and $\beta_2 = 12$. The other estimates at the maximum are $\hat{\mu}_1 = \hat{\mu}_2 = 0.089$, $\hat{\alpha}_{1s} = 258.8$, $\hat{\alpha}_{1c} = 102.6$, $\hat{\alpha}_{2s} = 0.926$ and $\hat{\alpha}_{2c} = 1.314$. The surface is not entirely concave (see the region where β_1 is close to zero in the figure). However, the function is concave and regular around the global maximum.

Now, we discuss the model-selection issue. The MLEs are performed on IBM stock's mid-price process with one-, two-, three-, and four-kernel Hawkes models. In Figure 4, based on IBM data on Jan 03, 2018 we present a Q-Q (quantile-quantile) plot of the residuals versus the unit exponential distribution. The set of residuals is defined by

$$R = \bigcup R_i, \quad R_i = \left\{ \int_{t_{i,j}}^{t_{i,j+1}} \hat{\lambda}_i(u) \mathrm{d}u \right\}$$

where $\hat{\lambda}_i$ denote inferred intensities. The points are closer to the straight line, as the number of kernels increases up to three. Meanwhile, the three-kernel and the four-kernel models differ insignificantly.

We obtain similar results based on the Akaike information criterion (AIC). Figure 5 plots the daily AICs for various kernel models based on IBM data for Jan 2018. The solid black

---

[1]From 2016, the data provided by the New York Stock Exchange is recorded in nanoseconds.
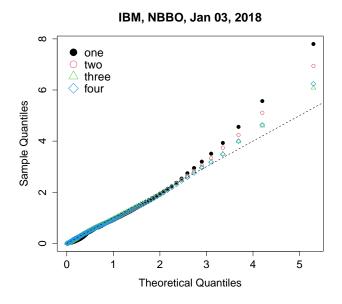
Figure 4: Q-Q plot of residual process of Hawkes models with various kernels

line, the blue dotted line and the red dash-dot line represent the one-kernel, two-kernel, and three-kernel models, respectively. The one-kernel model has a significant information loss compared to the others. The three-kernel model is slightly better than the two-kernel model. Although not depicted, the four-kernel model has very similar AICs to those of the three-kernel model. We observe similar patterns in the other cases. The model extensions are useful for up to two or three kernels but not for higher.
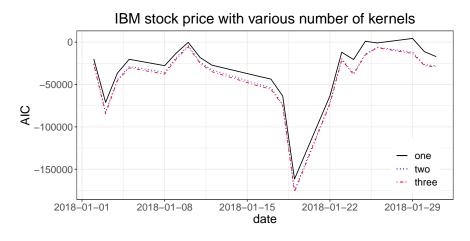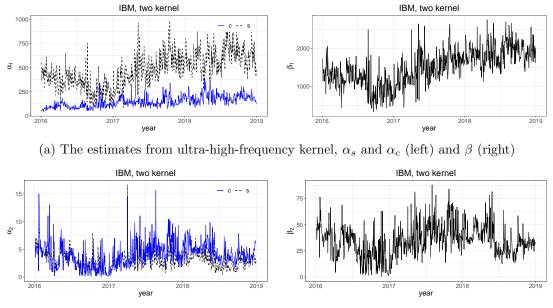
Figure 5: Akaike information criterion (AIC) for one, two and three-kernel models, mid price from NBBO of IBM, January, 2018

The estimates with a one second time unit under the two-kernel model are reported in Table 5, with standard errors in parentheses. The estimates of αs of the first kernel are several hundred; $\hat{\beta}$s range from approximately 1,000 to 2,000, $\hat{\alpha}$s of the second kernel are less than 10, and $\hat{\beta}$s range from approximately 30 to 80.

(a) The estimates from ultra-high-frequency kernel, α_s and α_c (left) and β (right)

(b) The estimates from high-frequency kernel, α_s and α_c (left) and β (right)

Figure 6: The dynamics of daily estimates of αs and βs under the two-kernel model for IBM stock price
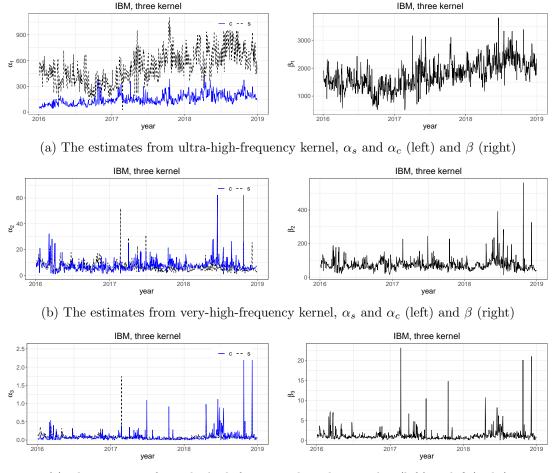
Figure 6 and 7 illustrate the dynamics of the daily estimates under the two-kernel and three-kernel models, respectively. We use the NBBO mid-price processes of IBM from 2016 to 2018. We term each kernel of the two-kernel model as ultra-high-frequency (UHF) and very-high-frequency (VHF) kernels. Figure 6a reveals $\hat{\alpha}_{1s}$, $\hat{\alpha}_{1c}$ and $\hat{\beta}_1$ for the UHF and $\hat{\alpha}_{1s}$s range from approximately 100 to 200 whereas $\hat{\alpha}_{1c}$s range from approximately 200 to 1,000.

This implies that, if the mid-price changes, the intensities increase instantly at several hundred. This can lead to numerous additional activities at frequencies of hundreds per second. Moreover, as shown in the figure, $\hat{\beta}_1$ is quite large and the effect dissipates quickly. This fast responsiveness is possibly due to automated high-frequency trading and quotes. Meanwhile, both $\hat{\alpha}_{2s}$ and $\hat{\alpha}_{2c}$ for the VHF kernel ranged from approximately 0 to 15, which is much less than $\hat{\alpha}$s in the UHF kernel. Furthermore, $\hat{\beta}_2$ is distributed between 0 and 80. This implies a longer persistence than in the UHF kernel.

For the three-kernel model, we term each kernel to be ultra-high-frequency (UHF), very-high-frequency (VHF), and high-frequency (HF). The $\hat{\alpha}$s in the UHF kernel are hundreds, $\hat{\alpha}$s in the VHF kernel are up to 60 and $\hat{\alpha}$s in the HF kernel are less than 2.5. Table 1 presents the summary statistics of the parameter estimates for various models.

3.2 Responsiveness

Under the Hawkes model, the intensities are instantly excited when an event occurs and decay to the level they would have been at if the excitement had not occurred. The increased intensity may or may not cause additional event during this period. If the excitement is large, and the decay is slow, then the occurrence of an event is highly possible. Based on this argument, we compute the conditional expectation of the arrival time of an event caused by a single excitement. This analysis helps us understand the responsive speed of each kernel.

(a) The estimates from ultra-high-frequency kernel, α_s and α_c (left) and β (right)

(b) The estimates from very-high-frequency kernel, α_s and α_c (left) and β (right)

(c) The estimates from the high-frequency kernel, α_s and α_c (left) and β (right)

Figure 7: The dynamics of daily estimates of αs and βs under the three-kernel model for IBM stock price

If the excitation parameter is α and the exponential decaying parameter is β, then the remaining amount of excitement in the intensity owing to α when time s has elapsed since the event occurred is $\alpha \exp(-\beta s)$. Based on the survival analysis with a hazard function $\alpha \exp(-\beta s)$, the conditional probability for the inter-arrival time, τ, of the next event owing to excitement α is

$$\mathbb{P}\{\tau < u < \infty\} = 1 - \exp\left(-\int_0^u \lambda(s)\mathrm{d}s\right) = 1 - \exp\left(-\alpha\frac{1 - \mathrm{e}^{-\beta u}}{\beta}\right) \tag{11}$$

Further, the conditional expectation of the next event owing to excitement α is

$$\mathbb{E}[\tau|\tau < \infty] = \int_0^\infty \alpha u \exp\left(-\alpha\frac{1 - \mathrm{e}^{-\beta u}}{\beta} - \beta u\right) \mathrm{d}u \tag{12}$$

which can be computed numerically. Using the above formula and the median estimates of αs and βs for IBM in 2018, Table 2 presents the estimated conditional expectation of the arrival time for each kernel.

Table 1: Summary statistics for estimates from the NBBO of IBM, 2018 under the one, two, and three-kernel Hawkes model for top to bottom

stock	statistic	μ	UHF			VHF			HF		
			α_s	α_c	β	α_s	α_c	β	α_s	α_c	β
	mean	0.2768	322.8	128.6	883.5						
IBM	median	0.2239	318.2	123.1	871.8						
	SD	0.1595	79.86	34.68	191.4						
	mean	0.2029	615.2	193.5	1915	2.909	4.605	35.75			
IBM	median	0.1728	619.8	188.2	1922	2.786	4.344	34.47			
	SD	0.1055	124.5	51.69	294.8	1.376	1.619	12.63			
	mean	0.1382	677.8	196.6	2191	6.010	8.397	78.13	0.0940	0.1271	1.449
IBM	median	0.1190	662.4	192.2	2132	4.862	7.457	66.77	0.0803	0.0661	1.180
	SD	0.0755	146.3	55.16	388.1	5.502	5.398	54.32	0.1065	0.2501	2.055

Table 2: Conditional expected time of the next arrival by each kernel, IBM, 2018

kernel	UHF		VHF		HF	
	self	cross				
one	319.4μs	145.7μs				
	self	cross	self	cross		
two	132.1μs	59.09μs	2.207ms	3.327ms		
	self	cross	self	cross	self	cross
three	115.8μs	55.61μs	1.032ms	1.530ms	72.74ms	62.12ms

In the three-kernel model, the conditional expectation of the arrival time originating from the UHF kernel ranges from approximately 50 to 100μs. The response time is very fast, and the corresponding activities may include automated actions such as flickering quotes and fleeting orders (Hasbrouck and Saar, 2009). The conditional expectation of arrival time by the VHF kernel ranges from approximately 1-2ms, while that for the HF kernel, ranges from approximately 60-70 ms. The expected arrival time by the base intensity is approximately 100-200ms.

Next, we examine the proportions of the causes of events among the kernels and the base intensity. According to the immigrant offspring argument, an event can be an external immigrant represented by the base intensity, or the descendant of an internally generated offspring described by the kernels. Based on the inferred intensities derived from the maximum

Table 3: The proportion of the causes of mid price changing events under two-kernel model and three-kernel model

	two-kernel model				
stock	year	base	UHF	VHF	HF
IBM	2016	38.40%	39.81%	21.79%	
	2017	37.03%	43.96%	19.01%	
	2018	35.52%	42.79%	21.69%	
	2019	37.35%	42.41%	20.24%	
AAPL	2016	35.45%	40.44%	24.11%	
	2017	32.39%	43.10%	24.50%	
	2018	31.34%	45.91%	22.76%	
	2019	33.64%	43.31%	23.04%	
	three-kernel model				
IBM	2016	26.54%	35.67%	20.29%	17.49%
	2017	24.54%	41.04%	18.51%	15.91%
	2018	24.22%	40.81%	19.70%	15.27%
	2019	25.17%	40.65%	16.91%	17.27%
AAPL	2016	26.24%	34.31%	16.53%	22.92%
	2017	18.70%	38.28%	20.90%	22.12%
	2018	16.30%	43.83%	20.61%	19.26%
	2019	16.12%	40.46%	20.42%	23.01%

likelihood estimates, we compute the probability that the parent of an event belongs to the source, that is, kernels or the base intensity. By comparing the estimated kernel components of the intensities at the time of an event, we calculate the empirical probability of the causes as the proportion of the intensity component. More precisely, in the three-kernel model, assuming that an upward move occurs at time t, the following

$$\frac{\hat{\mu}_1}{\hat{\lambda}_1(t)}, \quad \frac{\hat{\lambda}_{11}(t)}{\hat{\lambda}_1(t)}, \quad \frac{\hat{\lambda}_{21}(t)}{\hat{\lambda}_1(t)}, \quad \frac{\hat{\lambda}_{31}(t)}{\hat{\lambda}_1(t)}$$

represent the estimated probabilities that the cause of event belongs to base intensity, UHF, VHF, and UF kernels, respectively.

The annually aggregated empirical probabilities are presented in Table 3 from 2016 to 2019 for IBM and AAPL. Approximately 40% of the events are caused by the UHF kernel which is the largest proportion. The other kernels have similar percentages, ranging from approximately 15-25%. The percentages of the base intensity range from approximately 30-40% and 15-25% for the two-kernel and three-kernel models, respectively.

4 Conclusion

The multi-kernel Hawkes model is applied to the high-frequency price process and the basic moment properties are derived. Furthermore, the suitability of the MLE is examined in terms of conditional concavity. The empirical results reveal that high-frequency mid-price dynamics

derived directly from raw data follow multi-kernel Hawkes processes. The diagnostics test show that two- or three-kernel model are appropriate for ultra-high-frequency modeling.

In the two-kernel model, the kernel is composed of the ultra-high-frequency (50-100µs in expected response time) and very-high-frequency (2-3ms). In the three-kernel model, the kernel is composed of the ultra-high-frequency (50-100µs), very-high-frequency (1-1.5ms), and high-frequency (60-70ms). The percentage estimated by the ultra-high-frequency kernel is approximately 40%, which is the largest proportion in both the two and three-kernel models. This percentage is approximately 20% for the very-high-frequency and high-frequency kernels. The estimated percentages of external events are 15-25% and 30-40% for the three-kernel and two-kernel models, respectively.

Acknowledgements

References

Aït-Sahalia, Y., J. Cacho-Diaz, and R. J. Laeven (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics 117*, 585–606.

Arouri, M., O. M'saddek, D. K. Nguyen, and K. Pukthuanthong (2019). Cojumps and asset allocation in international equity markets. *Journal of Economic Dynamics and Control 98*, 1–22.

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

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

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

Bacry, E., I. Mastromatteo, and J.-F. Muzy (2015). Hawkes processes in finance. *Market Microstructure and Liquidity 1*, 1550005.

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

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

Choi, S. E., H. J. Jang, K. Lee, and H. Zheng (2021). Optimal market-making strategies under synchronised order arrivals with deep neural networks. *Journal of Economic Dynamics and Control 125*, 104098.

Cui, L., A. Hawkes, and H. Yi (2020). An elementary derivation of moments of Hawkes processes. *Advances in Applied Probability 52*, 102–137.

Da Fonseca, J. and Y. Malevergne (2021). A simple microstructure model based on the Cox-BESQ process with application to optimal execution policy. *Journal of Economic Dynamics and Control 128*, 104137.

Da Fonseca, J. and R. Zaatour (2014a). Clustering and mean reversion in a Hawkes microstructure model. *Journal of Futures Markets 35*, 813–838.

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

Da Fonseca, J. and R. Zaatour (2017). Correlation and lead–lag relationships in a Hawkes microstructure model. *Journal of Futures Markets 37*, 260–285.

Errais, E., K. Giesecke, and L. R. Goldberg (2010). Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics 1*, 642–665.

Filimonov, V. and D. Sornette (2015). Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data. *Quantitative Finance 15*, 1293–1314.

Gašperov, B. and Z. Kostanjčar (2022). Deep reinforcement learning for market making under a Hawkes process-based limit order book model. *IEEE Control Systems Letters 6*, 2485–2490.

Hardiman, S. J., N. Bercot, and J.-P. Bouchaud (2013). Critical reflexivity in financial markets: A Hawkes process analysis. *The European Physical Journal B 86*, 1–9.

Harris, L. (1999). Trading in pennies: a survey of the issues. Unpublished working paper, University of Southern California.

Hasbrouck, J. and G. Saar (2009). Technology and liquidity provision: The blurring of traditional definitions. *Journal of Financial Markets 12*, 143–172.

Hawkes, A. G. (1971). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological) 33*, 438–443.

Hawkes, A. G. (2018). Hawkes processes and their applications to finance: a review. *Quantitative Finance 18*, 193–198.

Jang, H. J., K. Lee, and K. Lee (2020). Systemic risk in market microstructure of crude oil and gasoline futures prices: A Hawkes flocking model approach. *Journal of Futures Markets 40*, 247–275.

Ji, J., D. Wang, D. Xu, and C. Xu (2020). Combining a self-exciting point process with the truncated generalized Pareto distribution: An extreme risk analysis under price limits. *Journal of Empirical Finance 57*, 52–70.

Jusselin, P. (2021). Optimal market making with persistent order flow. *SIAM Journal on Financial Mathematics 12*, 1150–1200.

Ketelbuters, J.-J. and D. Hainaut (2022). CDS pricing with fractional Hawkes processes. *European Journal of Operational Research 297*, 1139–1150.

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

Law, B. and F. Viens (2015). Hawkes processes and their applications to high-frequency data modeling. In I. Florescu, M. C. Mariani, H. E. Stanley, and F. G. Viens (Eds.), *Handbook of High-Frequency Trading and Modeling in Finance*, Chapter 6, pp. 183–219. John Wiley & Sons.

Lee, K. and B. K. Seo (2017a). Marked Hawkes process modeling of price dynamics and volatility estimation. *Journal of Empirical Finance 40*, 174–200.

Lee, K. and B. K. Seo (2017b). Modeling microstructure price dynamics with symmetric Hawkes and diffusion model using ultra-high-frequency stock data. *Journal of Economic Dynamics and Control 79*, 154–183.

Lee, K. and B. K. Seo (2022). Modeling bid and ask price dynamics with an extended Hawkes process and its empirical applications for high-frequency stock market data. *Journal of Financial Econometrics*, To appear.

Lewis, P. A. W. and G. S. Shedler (1979). Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly 26*, 403–413.

Ma, Y. and W. Xu (2016). Structural credit risk modelling with Hawkes jump diffusion processes. *Journal of Computational and Applied Mathematics 303*, 69–80.

Morariu-Patrichi, M. and M. S. Pakkanen (2022). State-dependent hawkes processes and their application to limit order book modelling. *Quantitative Finance 22*, 563–583.

Swishchuk, A. (2021). Modelling of limit order books by general compound Hawkes processes with implementations. *Methodology and Computing in Applied Probability 23*, 399–428.

Zhang, J., J. Wen, and J. Chen (2021). Modeling market fluctuations under investor sentiment with a Hawkes-contact process. *The European Journal of Finance*, To appear.

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

A   Proofs

Proof for Proposition 1. Eq. (4) implies

$$\boldsymbol{\lambda}_k(t) - \boldsymbol{\lambda}_k(0) = -\int_0^t \boldsymbol{\beta}_k \boldsymbol{\lambda}_k(s)\mathrm{d}s + \int_0^t \boldsymbol{\alpha}_k \mathrm{d}\boldsymbol{N}_s.$$

Under the steady state condition, by taking expectation to the both sides of the above equation, we have

$$\mathbb{E}[\boldsymbol{\lambda}_k(t)] - \mathbb{E}[\boldsymbol{\lambda}_k(0)] = -\boldsymbol{\beta}_k \mathbb{E}[\boldsymbol{\lambda}_k(t)]t + \boldsymbol{\alpha}_k \mathbb{E}[\boldsymbol{\lambda}_t]t$$

and using $\mathbb{E}[\boldsymbol{\lambda}_k(t)] = \mathbb{E}[\boldsymbol{\lambda}_k(0)]$,

$$\mathbb{E}[\boldsymbol{\lambda}_k(t)] = \boldsymbol{\beta}_k^{-1} \boldsymbol{\alpha}_k \mathbb{E}[\boldsymbol{\lambda}_t]$$

and substituting the above to

$$\mathbb{E}[\boldsymbol{\lambda}_t] = \boldsymbol{\mu} + \sum_{k=1}^K \mathbb{E}[\boldsymbol{\lambda}_k(t)],$$

which is derived by taking the expectation on Eq. (2), we have the desired result. □

Proof for Propositoin 3. By Lemma 2,

$$\boldsymbol{\beta}\mathbb{E}[\boldsymbol{\Lambda}_t\boldsymbol{\Lambda}_t^\top] + \mathbb{E}[\boldsymbol{\Lambda}_t\boldsymbol{\Lambda}_t^\top]\boldsymbol{\beta} = \mathbb{E}[\boldsymbol{\Lambda}_t\boldsymbol{\lambda}_t^\top]\boldsymbol{\alpha}^\top + \boldsymbol{\alpha}\mathbb{E}[\boldsymbol{\lambda}_t\boldsymbol{\Lambda}_t^\top] + \boldsymbol{\alpha}\mathrm{Dg}(\mathbb{E}[\boldsymbol{\lambda}_t])\boldsymbol{\alpha}^\top.$$

Since

$$\boldsymbol{\lambda}_t = \boldsymbol{\mu} + \mathbf{J}\boldsymbol{\Lambda}_t,$$

we have

$$(\boldsymbol{\beta} - \boldsymbol{\alpha}\mathbf{J})\mathbb{E}[\boldsymbol{\Lambda}_t\boldsymbol{\Lambda}_t^\top] + \mathbb{E}[\boldsymbol{\Lambda}_t\boldsymbol{\Lambda}_t^\top](\boldsymbol{\beta} - \mathbf{J}^\top\boldsymbol{\alpha}^\top) = \mathbb{E}[\boldsymbol{\Lambda}_t]\boldsymbol{\mu}^\top\boldsymbol{\alpha}^\top + \boldsymbol{\alpha}\boldsymbol{\mu}\mathbb{E}[\boldsymbol{\Lambda}_t^\top] + \boldsymbol{\alpha}\mathrm{Dg}(\mathbb{E}[\boldsymbol{\lambda}_t])\boldsymbol{\alpha}^\top.$$

□

Proof for Propositon 5. For simplicity, consider only the particular solution for $\frac{\mathrm{d}\mathbb{E}[\boldsymbol{\lambda}_t\boldsymbol{N}_t^\top]}{\mathrm{d}t}$. Assume that

$$\mathbb{E}[\boldsymbol{\lambda}_t\boldsymbol{N}_t^\top] \approx \mathbf{A}t + \mathbf{B}, \quad \mathbb{E}[\boldsymbol{\lambda}_k(t)\boldsymbol{N}_t^\top] \approx \mathbf{A}_k t + \mathbf{B}_k.$$

Since

$$\mathbb{E}[\boldsymbol{\lambda}_t\boldsymbol{N}_t^\top] = \mathbb{E}\left[\left(\boldsymbol{\mu} + \sum_{k=1}^K \boldsymbol{\lambda}_k(t)\right)\boldsymbol{N}_t^\top\right],$$

we have

$$\mathbf{A}t + \mathbf{B} = \left(\boldsymbol{\mu}\mathbb{E}[\boldsymbol{\lambda}_t^\top] + \sum_{k=1}^K \mathbf{A}_k\right)t + \sum_{k=1}^K \mathbf{B}_k$$

and

$$\mathbf{A} = \boldsymbol{\mu}\mathbb{E}[\boldsymbol{\lambda}_t^\top] + \sum_{k=1}^K \mathbf{A}_k, \quad \mathbf{B} = \sum_{k=1}^K \mathbf{B}_k. \tag{13}$$

Note that

$$d(\boldsymbol{\lambda}_k(t)\boldsymbol{N}_t^\top) = \boldsymbol{\lambda}_k(t)d\boldsymbol{N}_t^\top + (d\boldsymbol{\lambda}_k(t))\boldsymbol{N}_t^\top + d[\boldsymbol{\lambda}_k\boldsymbol{N}^\top]_t$$
$$= \boldsymbol{\lambda}_k(t)d\boldsymbol{N}_t^\top + (-\boldsymbol{\beta}_k\boldsymbol{\lambda}_k(t)dt + \boldsymbol{\alpha}_k d\boldsymbol{N}_t)\boldsymbol{N}_t^\top + \boldsymbol{\alpha}_k \mathrm{Dg}(\mathbb{E}[\boldsymbol{\lambda}_t])dt$$

and

$$\frac{d\mathbb{E}[\boldsymbol{\lambda}_{k,t}\boldsymbol{N}_t^\top]}{dt} = \mathbb{E}[\boldsymbol{\lambda}_{k,t}\boldsymbol{\lambda}_t^\top] - \boldsymbol{\beta}_k\mathbb{E}[\boldsymbol{\lambda}_{k,t}\boldsymbol{N}_t^\top] + \boldsymbol{\alpha}_k\mathbb{E}[\boldsymbol{\lambda}_t\boldsymbol{N}_t^\top] + \boldsymbol{\alpha}_k\mathrm{Dg}(\mathbb{E}[\boldsymbol{\lambda}_t])$$

and

$$\mathbf{A}_k = \mathbb{E}[\boldsymbol{\lambda}_k(t)\boldsymbol{\lambda}_t^\top] - \boldsymbol{\beta}_k(\mathbf{A}_k t + \mathbf{B}_k) + \boldsymbol{\alpha}_k(\mathbf{A}t + \mathbf{B}) + \boldsymbol{\alpha}_k\mathrm{Dg}(\mathbb{E}[\boldsymbol{\lambda}_t])$$

and

$$\mathbf{A}_k = \boldsymbol{\beta}_k^{-1}\boldsymbol{\alpha}_k\mathbf{A}$$
$$\mathbf{B}_k = \boldsymbol{\beta}_k^{-1}(-\mathbf{A}_k + \boldsymbol{\alpha}_k\mathbf{B} + \mathbb{E}[\boldsymbol{\lambda}_k(t)\boldsymbol{\lambda}_t^\top] + \boldsymbol{\alpha}_k\mathrm{Dg}(\mathbb{E}[\boldsymbol{\lambda}_t]))$$
$$= \boldsymbol{\beta}_k^{-1}(-\boldsymbol{\beta}_k^{-1}\boldsymbol{\alpha}_k\mathbf{A} + \boldsymbol{\alpha}_k\mathbf{B} + \mathbb{E}[\boldsymbol{\lambda}_k(t)\boldsymbol{\lambda}_t^\top] + \boldsymbol{\alpha}_k\mathrm{Dg}(\mathbb{E}[\boldsymbol{\lambda}_t])).$$

Substituting the above to Eq. (13), we obtain the desired result. $\square$

Proof for Propositon 7. Using Lemma 2 with $\boldsymbol{a} = \boldsymbol{b} = 0, \boldsymbol{f}_x = \boldsymbol{f}_y = \mathbf{I}$ and Proposition 5, we obtain the result. $\square$

B   Estimation results

Table 4: Estimates and standard errors under the two-kernel model for January 2018 based on the NBBO of IBM

date	μ	α_{1s}	α_{1c}	α_{2s}	α_{2c}	β_1	β_2
2018-01-02	0.1073	683.0	229.2	3.429	4.353	2041	46.34
	(0.0017)	(0.0596)	(0.1365)	(0.0347)	(0.0771)	(0.0500)	(0.1181)
2018-01-03	0.1697	637.0	215.3	4.773	6.473	1707	56.46
	(0.0021)	(0.0111)	(0.0392)	(0.0261)	(0.0464)	(0.0049)	(0.0352)
2018-01-04	0.1295	797.5	260.7	7.451	9.918	2423	81.54
	(0.0018)	(0.0850)	(0.0762)	(0.0421)	(0.0964)	(0.0386)	(0.1340)
2018-01-05	0.1193	804.3	200.2	4.314	5.815	2434	46.56
	(0.0018)	(0.1918)	(0.1414)	(0.2975)	(0.1434)	(0.2182)	(0.0872)
2018-01-08	0.1223	806.1	218.1	4.633	6.917	2430	51.03
	(0.0018)	(0.0605)	(0.0879)	(0.0394)	(0.0560)	(0.0153)	(0.0693)
2018-01-09	0.1129	707.7	179.5	3.519	3.869	2182	38.15
	(0.0017)	(0.0851)	(0.0237)	(0.1310)	(0.3077)	(0.0891)	(0.0997)
2018-01-10	0.0747	575.1	186.1	3.315	4.344	2006	40.49
	(0.0014)	(0.3245)	(0.3192)	(0.1709)	(0.2196)	(0.0404)	(0.1087)
2018-01-11	0.0918	737.8	200.5	3.537	5.074	2094	45.51
	(0.0015)	(0.0893)	(0.0980)	(0.0330)	(0.1472)	(0.0844)	(0.0306)
2018-01-12	0.1033	608.9	176.5	5.976	7.822	1755	64.47
	(0.0016)	(0.0746)	(0.1005)	(0.0481)	(0.2354)	(0.0312)	(0.0941)
2018-01-16	0.1778	484.1	186.6	4.066	6.259	1582	46.20
	(0.0021)	(0.0530)	(0.0377)	(0.0327)	(0.0576)	(0.0181)	(0.0488)
2018-01-17	0.1815	552.5	192.1	5.098	8.305	1777	60.78
	(0.0023)	(0.0381)	(0.1950)	(0.1334)	(0.1306)	(0.0661)	(0.1500)
2018-01-18	0.2074	660.1	207.4	5.144	5.520	1922	54.13
	(0.0023)	(0.0288)	(0.0242)	(0.0554)	(0.0470)	(0.0137)	(0.0209)
2018-01-19	0.2851	602.5	310.7	2.597	3.306	7.392	36.94
	(0.0027)	(0.2681)	(0.2784)	(0.2357)	(0.0352)	(0.2185)	(0.0155)
2018-01-22	0.1443	662.8	290.4	2.333	4.713	1791	44.51
	(0.0019)	(0.0343)	(0.0358)	(0.0424)	(0.0164)	(0.0210)	(0.0648)
2018-01-23	0.1250	474.6	172.4	2.835	5.142	1657	38.61
	(0.0018)	(0.1237)	(0.1455)	(0.2019)	(0.1499)	(0.0533)	(0.1033)
2018-01-24	0.2082	576.3	178.0	2.506	6.525	2183	33.79
	(0.0024)	(0.0528)	(0.0775)	(0.0588)	(0.0717)	(0.0404)	(0.1154)
2018-01-25	0.1656	485.5	116.1	2.658	8.710	2130	36.02
	(0.0021)	(0.0863)	(0.0879)	(0.0960)	(0.0717)	(0.0537)	(0.0372)
2018-01-26	0.0873	585.7	140.9	4.259	7.836	2093	54.11
	(0.0015)	(0.1618)	(0.1657)	(0.0982)	(0.2668)	(0.1149)	(0.1861)
2018-01-29	0.1543	522.4	105.8	1.974	8.045	2209	31.83
	(0.0021)	(0.0347)	(0.1528)	(0.1201)	(0.1694)	(0.2561)	(0.0945)
2018-01-30	0.1676	566.0	138.7	2.614	7.365	2085	35.06
	(0.0021)	(0.0369)	(0.0625)	(0.0331)	(0.0574)	(0.0223)	(0.0281)
2018-01-31	0.1274	651.3	141.3	3.326	6.150	2073	36.84
	(0.0018)	(0.0635)	(0.0241)	(0.0723)	(0.0443)	(0.0900)	(0.0579)

Table 5: Maximum likelihood estimates by the Hawkes models with different number of kernels based on NBBO of IBM for first four days of 2018

date	kernel	μ	α_{1s}	α_{1c}	α_{2s}	α_{2c}	α_{3s}	α_{3c}	α_{4s}	α_{4c}	β_1	β_2	β_3	β_4	llh
01-02	one	0.1306	323.8	130.4							854.9				9967
		(0.0018)	(0.0537)	(0.0777)							(0.0151)				
	two	0.1073	683.1	229.2	3.429	4.354					2041	46.34			12748
		(0.0017)	(0.0596)	(0.1365)	(0.0347)	(0.0771)					(0.0500)	(0.1181)			
	three	0.0757	639.1	206.5	4.861	6.501	0.0632	0.0324			1947	74.10	0.7097		13210
		(0.0018)	(0.1463)	(0.1246)	(0.1470)	(0.0827)	(0.0041)	(0.0040)			(0.0496)	(0.0863)	(0.0395)		
	four	0.0884	609.5	220.2	7.515	8.118	0.2603	0.6074	0.0794	0.0425	1982	121.9	17.76	1.405	13105
		(0.0028)	(0.1943)	(0.2937)	(0.1724)	(0.0567)	(0.0350)	(0.1174)	(0.0193)	(0.0127)	(0.1912)	(0.1162)	(0.0191)		
01-03	one	0.2127	301.2	127.0							699.3				35526
		(0.0026)	(0.1107)	(0.1127)							(0.0329)				
	two	0.1694	637.0	215.3	4.773	6.473					1707	56.46			41300
		(0.0021)	(0.0111)	(0.0392)	(0.0261)	(0.0464)					(0.0049)	(0.0352)			
	three	0.1217	804.4	252.4	8.694	10.99	0.0982	0.0407			2322	93.87	1.221		42155
		(0.0023)	(0.0309)	(0.0185)	(0.0282)	(0.0054)	(0.0045)	(0.0036)			(0.0197)	(0.0426)	(0.0038)		
	four	0.1279	1189	400.5	240.3	85.94	5.2708	8.0202	0.1150	0.0502	8220	1113	77.42	1.664	42500
		(0.0022)	(0.2202)	(0.1362)	(0.1831)	(0.0589)	(0.0684)	(0.1180)	(0.0056)	(0.0053)	(0.0405)	(0.1558)	(0.1408)	(0.0173)	
01-04	one	0.1539	298.6	138.9							749.4				18304
		(0.0020)	(0.1674)	(0.1570)							(0.1553)				
	two	0.1295	797.5	260.7	7.451	9.918					2423	81.54			21984
		(0.0018)	(0.0851)	(0.0762)	(0.0421)	(0.0964)					(0.3862)	(0.1340)			
	three	0.0882	722.7	235.5	8.443	12.17	0.1361	0.0239			2203	106.7	1.226		22706
		(0.0021)	(0.1882)	(0.1390)	(0.0154)	(0.0192)	(0.0063)	(0.0046)			(0.1251)	(0.0707)	(0.0740)		
	four	0.0909	1311	386.0	245.6	99.23	0.1477	0.0233	4.664	8.283	9772	1125	1.423	82.49	23060
		(0.0018)	(0.0522)	(0.1534)	(0.0258)	(0.0754)	(0.0062)	(0.0046)	(0.0477)	(0.0622)	(0.0377)	(0.0455)	(0.0336)	(0.0214)	
01-05	one	0.1516	291.2	104.9							747.0				10118
		(0.0024)	(0.6086)	(1.9505)							(0.3064)				
	two	0.1193	804.3	200.2	4.314	5.815					2434	46.56			14405
		(0.0018)	(0.1918)	(0.1414)	(0.2975)	(0.1434)					(0.2182)	(0.0872)			
	three	0.0827	831.1	199.4	5.459	7.532	0.1012	0.0353			2551	64.05	0.9769		15000
		(0.0023)	(0.1007)	(0.2396)	(0.0500)	(0.1007)	(0.0072)	(0.0048)			(0.0606)	(0.0920)	(0.0850)		
	four	0.0839	1023	176.2	131.1	51.38	3.087	5.564	0.1115	0.0393	4934	926.9	51.37	1.151	15129
		(0.0019)	(0.0808)	(0.0970)	(0.0991)	(0.1765)	(0.0813)	(0.0990)	(0.0057)	(0.0047)	(0.1392)	(0.0954)	(0.0802)	(0.0475)	(0.0117)