Hawkes-Based Models for High Frequency Financial Data

Kaj Nyström[a] and Changyong Zhang[b]

[a] Department of Mathematics, Uppsala University, Uppsala, Sweden
E-mail: kaj.nystrom@math.uu.se;
[b] Department of Finance and Banking, Faculty of Business, Curtin University, Miri, Malaysia
E-mail: changyong.zhang@curtin.edu.my

ARTICLE HISTORY
Compiled February 13, 2021

ABSTRACT
Compared with low frequency data, high frequency data exhibit distinct empirical properties, including, for instance, essentially discontinuous evolution paths, time-varying intensities, and self-exciting features. All these make it more challenging to model appropriately the dynamics associated with high frequency data such as order arrival and price formation. To capture more accurately the microscopic structures and properties pertaining to the limit order books, this paper focuses on modeling high frequency data using Hawkes processes. Two models, one with exponential kernels and the other with power-law kernels, are introduced systematically, algorithmized precisely, and compared with each other extensively from various perspectives, including the goodness of fit to the original data and the computational time in searching for the maximum likelihood estimator, with search algorithm being taken into consideration as well. To measure the goodness of fit, a number of quantities are proposed. Studies based on both multiple-trading-day data of one stock and multiple-stock data on one trading day indicate that Hawkes processes with slowly-decaying kernels are able to reproduce the intensity of jumps in the price processes more accurately. The results suggest that Hawkes processes with power-law kernels and their implied long memory nature of self-excitation phenomena could, on the level of microstructure, serve as a realistic model for high frequency data.

KEYWORDS
Hawkes processes; high frequency financial data; intensity kernel; maximum likelihood estimation; sample-path simulation

1. Introduction

In financial markets, the evolution of prices is driven by the interaction of buy and sell orders. Nowadays, more and more equity exchanges have been organized as order-driven markets, where orders are aggregated in a limit order book, which states quantities posted at each price level and is available to market participants. This new system makes it possible for high-frequency trading, a program trading platform with powerful computers using complex algorithms to analyze multiple markets and to execute orders based on market conditions at very fast speeds.

For high-frequency trading, it requires modeling the microstructure of the target market, based on which sophisticated techniques may be implemented. Intuitively, by modeling the dynamics of a limit order book, information on the current state of the book may be utilized to predict its short-term behavior. Fundamentally, models of

order
evolution. Practically, investors and trading desks may employ such models to design trading strategies and optimize trade execution (Alfonsi et al., 2010; Obizhaeva and Wang, 2013).

Due to the complexity involved, it has been a challenge to formulate the dynamics of a limit order book, particularly with models both statistically realistic and quantitatively analytical, in which case practical measures such as goodness of fit may naturally be more readily defined and evaluated. Equilibrium models, which use game theory to incorporate strategic interaction between traders, shed light on the price formation process, but with unobservable parameters indicating agent preferences (Foucault et al., 2005; Parlour, 1998; Rosu, 2009). This makes them difficult to be applied in reality. Empirical studies on properties of limit order books highlight statistical features of order book dynamics, which somehow are usually unrealistic to be described in a single model (Bouchaud et al., 2002; Doyne Farmer et al., 2004; Hollifield et al., 2004). Meanwhile, most existing stochastic models assume steady-state distributions, which are not necessarily verified by real high frequency data (Bouchaud et al., 2009; Cont et al., 2010; Luckock, 2003; Maslov and Mills, 2001; Smith et al., 2003).

Point processes, in particular Hawkes processes, on the other hand, can be adopted to describe directly event times of an underlying system, without pre-conditions assumed in the aforementioned models. This makes them more natural and suitable to model high frequency data, for example, order arrival and microstructure noise (Abergel et al., 2016; Bacry et al., 2015; Bauwens and Hautsch, 2009; Hawkes, 2018).

Hawkes processes with different kernels exhibit distinct behaviors, as partially illustrated by the preliminary results (Zhang, 2016). This paper focuses on a relatively complete and systematic study of the differences between exponential kernels and power-law kernels in modeling high frequency financial data, based on which an appropriate model can be identified. The results may then be adopted and applied readily by practitioners directly or indirectly as well as other researchers for possible further studies in high frequency trading as well as other related fields.

It is hoped that the study can make Hawkes-based models, which are analytically tractable and possess flexible statistical properties, more visible to the community, so that it may be applied in an even broader range of fields. In particular, although Hawkes-based models for high frequency financial data have been more favorably addressed recently, so far it has not been seen intensively explicit results on real data. The main focus in the literature has been more on Hawkes processes with exponential kernels. Power-law kernels have been referred to but have not been systematically looked into, including the distinction from the exponential kernels from the practical and computational perspectives. Compared with most of the existing literature, this is one of the very first to present complete first-hand results on real data, particularly the model with power-law kernels. In addition, since in the literature it still lacks theoretical results on the differences between the two models for high frequency financial data, a further hope is then to call for deeper analytical studies behind the empirical results demonstrated in this paper.

In the following Section 2, Hawkes processes and the associated properties are discussed, with models based on Hawkes processes being introduced in Section 3. A detailed study is then carried out in Section 4 on the models for describing high frequency financial data to investigate the differences between them from a variety of perspectives. Section 5 outlines directions for possible future work.

2. Point Processes to Hawkes Processes

Hawkes processes share a number of fine properties of the more generic point processes, which are hence briefly introduced first.

2.1. Point Processes

A point process is a random collection of points falling in a certain space, in general with each point representing the time (location) of an event such as a lightning strike or earthquake. To describe event-time based phenomena whose underlying systems move essentially by jumps, for example, the evolution of stock prices, point processes have been proved to be powerful tools (Daley and Vere-Jones, 2003, 2008).

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space with a filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\in\mathbb{R}}$ of σ-algebras satisfying the usual conditions. Recall that a point process N is orderly if

$$\lim_{\Delta t\to 0} \frac{P(N_{(t,t+\Delta t]} > 1)}{\Delta t} = 0, \forall t,$$

where $N_{(t,t+\Delta t]}$ is the number of points occurring in $(t, t+\Delta t]$ (Khinchine, 1969). The process $\{N_t\}_{t\in\mathbb{R}}$, where N_t is the number of points occurring before or at t, is called the associated counting process. The complete intensity function of N is defined by

$$\lambda_t = \lim_{\Delta t\to 0} \frac{P(N_{(t,t+\Delta t]} > 0|\mathcal{F}_t)}{\Delta t}. \tag{1}$$

If $\mathcal{F}_0$ has a specified distribution, then for a given function $\lambda = \{\lambda_t\}_{t\geq 0}$, there exists at most one orderly point process in $[0, \infty)$ satisfying (1) and λ determines the joint interval distribution (Hawkes and Oakes, 1974). That is, for the event times $\{t_k\}_{k\geq 1}$ with $0 \leq t_1 < t_2 < \cdots$,

$$P(t_1 \leq x_1|\mathcal{F}_0) = 1 - \exp\big(-\int_0^{x_1} \lambda_s ds\big) \tag{2}$$

and

$$P(t_k \leq x_k|\mathcal{F}_0, t_1 = x_1, \cdots, t_{k-1} = x_{k-1}) = 1 - \exp\big(-\int_{x_{k-1}}^{x_k} \lambda_s ds\big). \tag{3}$$

If $\mathrm{E}[\lambda_t]$ exists, then

$$\mathrm{E}[N_{(a,b]}] = \int_a^b \mathrm{E}[\lambda_s] ds$$

and for a stationary point process N, $\mathrm{E}[\lambda_t]$ is constant.

2.2. Hawkes Processes

Among others, a particularly interesting class of point processes is Hawkes processes, which are mathematically tractable and integrate two linear representations in one

model, one for clustering and the other for conditional intensity (Hawkes, 1971; Vere-Jones, 1970). Since it was introduced, Hawkes process has been widely applied in, for instance, seismology, shot noises, biology, and criminology (Brémaud and Massoulié, 2002; Coleman and Gastwirth, 1969; Mohler et al., 2011; Ogata, 1999; Reynaud-Bouret and Schbath, 2010; Zhuang et al., 2002). More recently, it has gradually become a popular tool in finance, from estimating VaR and valuing credit derivatives to modeling market event data, microstructure noise, and clusters of extremes (Bacry and Muzy, 2014; Chavez-Demoulin et al., 2005; Chavez-Demoulin and McNeil, 2012; Embrechts et al., 2011; Errais et al., 2010; Giesecke et al., 2011; Zheng et al., 2014).

A Hawkes process $N = \{N_t\}_{t\in\mathbb{R}}$ in one dimension is a counting process defined by the intensity, or the rate of arrival of events, λ_t,

$$\lambda_t = \mu + \int_{-\infty}^{t} \phi_{t-s} dN_s = \mu + \sum_{t_i<t} \phi_{t-t_i}, \tag{4}$$

where μ is a deterministic base intensity and the decay kernel ϕ represents the positive influence of past events t_i on the current value of the intensity process. In (4), the range of the integral is $(-\infty, t)$ instead of $(-\infty, t]$. This guarantees the predicability of λ (Ogata, 1978). For stationarity, it is assumed that $\mu > 0$, $\phi_s \geq 0, \forall s \geq 0$, and

$$\int_0^\infty \phi_s ds < 1. \tag{5}$$

By the assumptions and the condition that $\mathrm{E}[\lambda_t]$ is constant, it follows that

$$\mathrm{E}[\lambda_t] = \mathrm{E}[\mu + \int_{-\infty}^{t} \phi_{t-s} dN_s] = \mu + \mathrm{E}[\lambda_t] \int_0^\infty \phi_s ds,$$

which gives

$$\mathrm{E}[\lambda_t] = \frac{\mu}{1 - \int_0^\infty \phi_s ds}.$$

In addition, in this case, the interval distributions introduced in (2) and (3) can be expressed explicitly and can in principle be used for predicting future events (Hawkes and Oakes, 1974). Regarding existence and uniqueness, it has been proved that if $\mu > 0$, $\phi_s \geq 0, \forall s \geq 0$, and (5) is satisfied, then there exists a unique stationary orderly process of finite rate whose complete intensity function is given by (4). Similar results hold for mutually-exciting processes.

A stationary Hawkes process is thus essentially an immigrant-birth process, composed of a homogeneous Poisson immigrant with rate μ and nonhomogeneous Poisson descendants with rate ϕ_s. Accordingly, a Hawkes-based model exhibits self-exciting behavior. That is, the arrival of one event increases the probability of occurrence of new ones, or $\mathrm{Cov}(N_{(s,t]}, N_{(t,u]}) > 0$, $s < t < u$, where $N_{(s,t]}$ is the number of points occurring in $(s, t]$ of the underlying process. If $\mathrm{Cov}(N_{(s,t]}, N_{(t,u]}) < 0$, it is called a self-correcting model, which is used in fields such as ecology and forestry to capture occurrences that are well-dispersed (Isham and Westcott, 1979; Ogata and Vere-Jones, 1984).

For a d-dimensional multivariate mutually-exciting Hawkes process, the intensity

function is given by

$$\lambda_t^i = \mu^i + \sum_{j=1}^{d} \int_{-\infty}^{t} \phi_{t-s}^{ij} dN_s^j, i = 1, \ldots, d, \tag{6}$$

where $\phi_s^{ij} \geq 0, \forall s \geq 0, i, j = 1, \ldots, d$. For stationarity, it is assumed that

$$\text{the moduli of all eigenvalues of the matrix } \Phi = \left(\phi^{ij}\right)_{i,j=1}^{d} \text{ are less than one,} \tag{7}$$

where $\phi^{ij} = \int_0^\infty \phi_s^{ij} ds, i, j = 1, \ldots, d$.

For a stationary Hawkes process, the following statement holds (Hawkes, 1971).

Proposition 2.1. *Let $N = \{N_t\}_{t\in\mathbb{R}}$ be a stationary Hawkes process whose intensity is defined by* (6)*. Then*

$$\mathrm{E}[N_t] = (I - \Phi)^{-1} \mu t,$$

where I is the identity matrix, Φ is as defined in (7)*, and $\mu = (\mu^1, \ldots, \mu^d)^T$.*

2.2.1. Maximum Likelihood Estimation

To model an underlying system with a Hawkes process, a key step is to estimate the parameters, which requires the calculation of the log-likelihood function.

Let $T \in (0, \infty)$. For a point process with conditional intensity $\lambda_t = \lambda_t(\theta)$, if the law of the process is absolutely continuous w.r.t. that of a stationary standard Poisson process, then the log-likelihood function on $[0, T]$ is given by

$$L^T(\theta) = \int_0^T \ln \lambda_t(\theta) dN_t - \int_0^T \lambda_t(\theta) dt, \tag{8}$$

which is used for maximum likelihood estimation of $\theta = (\theta^i)_i$. The estimator is obtained as

$$\hat{\theta}^T = \arg\max_{\theta} L^T(\theta).$$

As for the asymptotic properties, under certain assumptions, the maximum likelihood estimator for sufficiently large T is consistent, asymptotically normal, and efficient (Lewis, 1969; Ogata, 1978). Specifically, as $T \to \infty$, $\hat{\theta}^T$ converges in probability to the true value θ^*, that is,

$$\lim_{T\to\infty} P(|\hat{\theta}^T - \theta^*| > \varepsilon) = 0, \forall \varepsilon > 0,$$

$$\sqrt{T}(\hat{\theta}^T - \theta^*) \sim \mathcal{N}\left(0, I(\theta^*)^{-1}\right),$$

where $I(\theta^*) = \left(\mathrm{E}[\frac{1}{\lambda} \frac{\partial \lambda}{\partial \theta^i} \frac{\partial \lambda}{\partial \theta^j}]_{\theta=\theta^*}\right)_{ij}$, and $\hat{\theta}^T$ asymptotically reaches the lower bound of the variance.

2.2.2. Sample-Path Simulation

For modeling and prediction purpose, an efficient algorithm is in demand to simulate sample paths. For nonhomogeneous Poisson processes, there have existed a number of simulating methods (Lewis and Shedler, 1976, 1979a,b). For a Hawkes process, it can be simulated by a thinning algorithm based on Proposition 2.2 (Ogata, 1981).

Proposition 2.2. *Let $T \in (0,\infty)$ and $(\Omega, \mathcal{F}, P)$ be a complete probability space with a filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\in[0,T]}$ of σ-algebras satisfying the usual conditions. Consider a multivariate point process $N = \{N_t^i\}_{t\geq 0}^{i=1,\dots,d}$ on the interval $(0,T]$ with joint intensity $\lambda = \{\lambda_t^i\}_{t\geq 0}^{i=1,\dots,d}$. Suppose it can be found a one-dimensional $\mathbb{F}$-predictable point process $\tilde{N} = \{\tilde{N}_t\}_{t\geq 0}$ with intensity function $\tilde{\lambda} = \{\tilde{\lambda}_t\}_{t\geq 0}$ such that*

$$\sum_{i=1}^{d} \lambda_t^i \leq \tilde{\lambda}_t, \forall t \in (0,T], P\text{-almost surely.}$$

Let $\tilde{t}_1, \tilde{t}_2, \dots, \tilde{t}_{n_T}$ be the points of $\tilde{N}$ on the interval $(0,T]$. For $k = 1, 2, \dots, n_T$, mark $\tilde{t}_k$ as i with probability $\frac{\lambda_{\tilde{t}_k}^i}{\tilde{\lambda}_{\tilde{t}_k}}, i = 1, \dots, d$. Then the points with marks $i = 1, \dots, d$ provide a copy of N.

In Proposition 2.2, $\tilde{\lambda}$ is usually a piecewise-constant process changing its rate according to the past history. By the predictability of $\tilde{\lambda}$, the three processes, $\tilde{\lambda}$, $\{\tilde{t}_k\}$, and $\{t_i\}$, are constructed sequentially. In particular, in the case of a one-dimensional Poisson process, the steps are as follows:

- Suppose that the last point t_i before time t has been obtained. Then construct $\tilde{\lambda}_t$, which is $\mathcal{F}_{t_i}$-measurable, piecewise constant, and satisfies $\tilde{\lambda}_t \geq \lambda_t, \forall t \geq t_i$;
- Simulate homogeneous Poisson points $\tilde{t}_k (> t_i)$ according to the intensity $\tilde{\lambda}_t$;
- For each of the points $\{\tilde{t}_k\}$, the probability $\frac{\lambda_{\tilde{t}_k}}{\tilde{\lambda}_{\tilde{t}_k}}$ is given conditionally independent of $\tilde{t}_k$ under the past history $\mathcal{F}_{\tilde{t}_k}$;
- t_{i+1} is the first accepted point among $\tilde{t}_k (> t_i)$.

3. Hawkes-Based Models

It has been empirically observed that order flow exhibits those properties such as time-varying intensity, autocorrelations in event arrivals, cross-correlations in arrival rates of different order types, conditional properties, and self-exciting features. Most of them are not appropriately captured by Poisson point processes (Chakraborti et al., 2011).

Hawkes processes possess flexible statistical properties allowing to incorporate autocorrelations, self-exciting and mutually-exciting features. Unlike time-series models such as ACD-GARCH, they remain analytically tractable. Likelihood functions, conditional distributions, moments, Laplace transforms, and characteristic functions can all be computed analytically or by solving ODEs. For this, recently they have been more favorably applied to model high frequency financial data (Abergel and Jedidi, 2011; Bacry et al., 2012, 2016; Bormetti et al., 2015; Cont, 2011; Filimonov et al., 2014; Filimonov and Sornette, 2012; Lee and Seo, 2017). For example, order arrival in

a single market can be modeled as a mutually-exciting multivariate point process, by categorizing orders and cancelations into several types (Bowsher, 2007; Large, 2007).

As the Epps effect (Epps, 1979), microstructure noise is a stylized fact in high frequency financial data (Aït-Sahalia et al., 2005, 2011; Barndorff-Nielsen et al., 2008; Robert and Rosenbaum, 2011; Rosenbaum, 2011). It captures the property that on the microscopic scales, an upward jump in price is more likely to be followed by a downward jump and vice versa. This provides a perspective for microeconomic analysis of price manipulation. On the other hand, the Epps effect states that the empirical correlation between two assets vanishes on fine scales. When the scale goes coarser, both effects become less significant and a diffusion starts to dominate (Bacry et al., 2013b).

For the price evolution of an asset, assume that the asset is traded in a single market and that the price moves by one tick at most, then it can be modeled using a Hawkes process (Bacry et al., 2013a). Let P_0 be the price at time 0. Then

$$P_t = P_0 + N_t^1 - N_t^2, \tag{9}$$

where N_t^1 is the number of upward jumps and N_t^2 is that of downward jumps between 0 and t. The intensities of N_t^1 and N_t^2 are λ_t^1 and λ_t^2, respectively, with

$$\lambda_t^i = \mu^i + \sum_{j=1}^{2} \int_0^t \phi_{t-s}^{ij} dN_s^j, i = 1, 2, \tag{10}$$

where $\phi_s^{ij} \geq 0, \forall s \geq 0, i, j = 1, 2$.

Hawkes processes with different kernels exhibit distinct behaviors. This paper focuses on studying the differences between the two models whose intensities are defined below by (11) and (12), respectively.

3.1. Model with Exponential Kernels

As a parameterization of the kernels in (6), the case that $\phi_s^{ij} = \alpha^{ij} e^{-\beta^{ij} s}$ with $\alpha^{ij} > 0$, $\beta^{ij} > 0$, and $\frac{\alpha^{ij}}{\beta^{ij}} < 1$, $i, j = 1, \ldots, d$ is the most studied one (Ozaki, 1979). This is partially due to the unique properties of exponential functions which lead to the possibility of closed-form expressions and recursive computation of quantities of interest.

In particular, in (10) if $\phi_s^{ij} = \alpha^{ij} e^{-\beta^{ij} s}$, then the intensities of N_t^1 and N_t^2 are given explicitly by

$$\lambda_t^i = \mu^i + \sum_{j=1}^{2} \int_0^t \frac{\alpha^{ij}}{e^{\beta^{ij}(t-s)}} dN_s^j, i = 1, 2, \tag{11}$$

where $\alpha^{ij} > 0$, $\beta^{ij} > 0$, and $\frac{\alpha^{ij}}{\beta^{ij}} < 1$, $i, j = 1, 2$.

Denote

$$\Phi = \begin{bmatrix} \frac{\alpha^{11}}{\beta^{11}} & \frac{\alpha^{12}}{\beta^{12}} \\ \frac{\alpha^{21}}{\beta^{21}} & \frac{\alpha^{22}}{\beta^{22}} \end{bmatrix},$$

then by Proposition 2.1,

$$\mathrm{E}[N_t] = (I - \Phi)^{-1} \mu t,$$

where I is the identity matrix and $\mu = (\mu^1, \mu^2)^T$.

Let $\{t_k^1\}$ and $\{t_k^2\}$ be the event times of N^1 and N^2, respectively. Then by (8), the log-likelihood function on the time interval $[0, T]$ is

$$L^T(\theta) = \sum_{i=1}^{2} L^i(\theta),$$

where for $i = 1, 2$,

$$\begin{aligned} L^i(\theta) &= \sum_{t_k^i<T} \ln\Big[\mu^i + \sum_{j=1}^{2}\sum_{t_l^j<t_k^i} \alpha^{ij} e^{-\beta^{ij}(t_k^i - t_l^j)}\Big] - \mu^i T - \sum_{j=1}^{2}\sum_{t_k^j<T} \frac{\alpha^{ij}}{\beta^{ij}}(1 - e^{-\beta^{ij}(T-t_k^j)}) \\ &= \sum_{t_k^i<T} \ln\Big[\mu^i + \sum_{j=1}^{2} \alpha^{ij} R^{ij}(k)\Big] - \mu^i T - \sum_{j=1}^{2}\sum_{t_k^j<T} \frac{\alpha^{ij}}{\beta^{ij}}(1 - e^{-\beta^{ij}(T-t_k^j)}) \end{aligned}$$

with

$$R^{ij}(k) = \sum_{t_l^j<t_k^i} e^{-\beta^{ij}(t_k^i - t_l^j)} = \begin{cases} e^{-\beta^{ij}(t_k^i - t_{k-1}^i)} R^{ij}(k-1) + \sum\limits_{t_{k-1}^i \le t_l^j < t_k^i} e^{-\beta^{ij}(t_k^i - t_l^j)}, & i \neq j, \\ e^{-\beta^{ij}(t_k^i - t_{k-1}^i)}(1 + R^{ij}(k-1)), & i = j. \end{cases}$$

The maximum likelihood estimator of $\theta = \big((\mu^i), (\alpha^{ij}), (\beta^{ij}), i, j = 1, 2\big)$ is then obtained as

$$\hat{\theta}^T = \arg\max_{\theta} L^T(\theta).$$

By Proposition 2.2, a Hawkes process with exponential kernels can be readily simulated. In one dimension where

$$\lambda_t = \mu + \int_0^t \frac{\alpha}{e^{\beta(t-s)}} dN_s,$$

a standard algorithm is as follows:

- Initialization: Set $k = 1$ and $\lambda^* = \mu$
- First Event: Generate $U \sim \mathcal{U}_{[0,1]}$ and set $s = -\frac{1}{\lambda^*} \ln U$
  - If $s \le T$, set $t_1 = s$. Otherwise, go to Output
- General Routine: Set $k = k + 1$
  - Update Maximum Intensity: Set $\lambda^* = \lambda_{t_{k-1}} + \alpha$
  - New Event: Generate $U \sim \mathcal{U}_{[0,1]}$ and set $s = s - \frac{1}{\lambda^*} \ln U$; If $s > T$, go to Output
  - Rejection Test: Generate $V \sim \mathcal{U}_{[0,1]}$; If $V \le \frac{\lambda_s}{\lambda^*}$, set $t_k = s$ and go to General Routine, else update $\lambda^* = \lambda_s$ and go to New Event
- Output: Retrieve the simulated process $\{t_k\}$ on $[0, T]$

Multi-dimensional processes can be simulated by following similar steps.

3.2. Model with Power-Law Kernels

It has been observed that Hawkes processes with power-law kernels describe certain phenomena more accurately than exponential kernels (Vere-Jones, 1970). Take $\phi_s^{ij} = \frac{\alpha^{ij}}{(s+\gamma^{ij})^{\beta^{ij}}}$ with $\alpha^{ij} > 0$, $\beta^{ij} > 1$, and $\gamma^{ij} > 0$, $i, j = 1, 2$, then explicitly (10) becomes

$$\lambda_t^i = \mu^i + \sum_{j=1}^{2} \int_0^t \frac{\alpha^{ij}}{(t-s+\gamma^{ij})^{\beta^{ij}}} dN_s^j, i = 1, 2. \tag{12}$$

Assume that the moduli of all eigenvalues of the matrix

$$\Phi = \begin{bmatrix} \frac{\alpha^{11}}{(\beta^{11}-1)\gamma^{11\beta^{11}-1}} & \frac{\alpha^{12}}{(\beta^{12}-1)\gamma^{12\beta^{12}-1}} \\ \frac{\alpha^{21}}{(\beta^{21}-1)\gamma^{21\beta^{21}-1}} & \frac{\alpha^{22}}{(\beta^{22}-1)\gamma^{22\beta^{22}-1}} \end{bmatrix}$$

are less than one. Then by (7) and Proposition 2.1,

$$\mathrm{E}[N_t] = (I - \Phi)^{-1}\mu t,$$

where I is the identity matrix and $\mu = (\mu^1, \mu^2)^T$.

Let $\{t_k^1\}$ and $\{t_k^2\}$ be the event times of N^1 and N^2, respectively. Then by (8) and (12), the closed-form log-likelihood function is

$$L^T(\theta) = \sum_{i=1}^{2} L^i(\theta),$$

where for $i = 1, 2$,

$$\begin{aligned} L^i(\theta) &= \sum_{t_k^i < T} \ln\Big[\mu^i + \sum_{j=1}^{2} \sum_{t_l^j < t_k^i} \frac{\alpha^{ij}}{(t_k^i - t_l^j + \gamma^{ij})^{\beta^{ij}}}\Big] - \mu^i T \\ &\quad - \sum_{j=1}^{2} \sum_{t_k^j < T} \frac{\alpha^{ij}}{\beta^{ij}-1} \Big(\frac{1}{\gamma^{ij\beta^{ij}-1}} - \frac{1}{(T - t_k^j + \gamma^{ij})^{\beta^{ij}-1}}\Big). \end{aligned}$$

The maximum likelihood estimator of $\theta = \big((\mu^i), (\alpha^{ij}), (\beta^{ij}), (\gamma^{ij}), i, j = 1, 2\big)$ is given by

$$\hat{\theta}^T = \arg\max_{\theta} L^T(\theta).$$

By Proposition 2.2, the process given in (12) can be simulated as follows:

- Initialization: Set $k = 1, k^i = 1, i = 1, 2$, and $\lambda^* = \lambda_0 = \sum_{i=1}^2 \mu^i$
- First Event: Generate $U \sim \mathcal{U}_{[0,1]}$ and set $s = -\frac{1}{\lambda^*} \ln U$
  - If $s \le T$, generate $V \sim \mathcal{U}_{[0,1]}$ and set $t_1^j = t_1 = s$ where $j = 1$ if $V \le \frac{\mu^1}{\lambda^*}$ and $j = 2$ if $\frac{\mu^1}{\lambda^*} < V \le \frac{\mu^1+\mu^2}{\lambda^*} = 1$. Otherwise, go to Output
- General Routine: Set $k = k + 1$ and $k^j = k^j + 1$
  - Update Maximum Intensity: Set $\lambda^* = \sum_{i=1}^2 \lambda_{t_{k-1}}^i + \sum_{i=1}^2 \frac{\alpha^{ij}}{\gamma^{ij}\beta^{ij}}$

- New Event: Generate $U \sim \mathcal{U}_{[0,1]}$ and set $s = s - \frac{1}{\lambda^*} \ln U$; If $s > T$, go to Output
- Attribution-Rejection Test: Generate $V \sim \mathcal{U}_{[0,1]}$; If $V \leq \frac{\sum_{i=1}^{2} \lambda_s^i}{\lambda^*}$, set $t_{k^j}^j = t_k = s$ where $j = 1$ if $V \leq \frac{\lambda_s^1}{\lambda^*}$ and $j = 2$ if $\frac{\lambda_s^1}{\lambda^*} < V \leq \frac{\lambda_s^1 + \lambda_s^2}{\lambda^*}$, and go to General Routine, else update $\lambda^* = \sum_{i=1}^{2} \lambda_s^i$ and go to New Event

- Output: Retrieve the simulated process $\{(t_{k^i}^i)\}_{i=1,2}$ on $[0, T]$

4. Empirical Study

The two models (11) and (12) are carefully and rigorously computerized in MATLAB, which is probably among the first to see in the literature. Since the focus is on the distinction between the two types of kernels, the underlying price process does not affect the result and the studies are hence based on the best bid. For consistency, the same price process (9) is assumed, where $P = \{P_t\}_{t \geq 0}$ denotes the best bid, $N^1 = \{N_t^1\}_{t \geq 0}$ the number of upward jumps with intensity λ^1, and $N^2 = \{N_t^2\}_{t \geq 0}$ the number of downward jumps with intensity λ^2. To verify the assumption, the jump sizes of the best bid of ERICB (Ericsson Telephone Company) are examined for the trading day (9:00-17:30) of September 7th, 2012 on five exchanges, namely BATS Europe (BATE), Nasdaq OMX Stockholm (BOOK), Burgundy (BURG), Chi-X (CHIX), and Turquoise (TRQX). As shown in Table 1, overall more than 97% ($\frac{7272}{7452} = 0.9758$) and on the exchange CHIX over 98% ($\frac{1719}{1742} = 0.9868$) jumps are 1 tick only. It is thus reasonable to first assume that the best bid moves by 1 tick for each jump. This is further backed up by what is revealed in Table 2, which presents jump sizes of the best bid of ERICB on all trading days in September 2012.

	BATE	BOOK	BURG	CHIX	TRQX	Overall
1 tick	1334	1295	1391	1719	1533	7272
2 ticks	18	19	18	9	10	74
3 ticks	10	2	8	9	3	32
$>$ 3 ticks	12	14	36	5	7	74
Total	1374	1330	1453	1742	1553	7452

Table 1. Jumps of Best Bid of ERICB on Five Exchanges on 07/09/2012

	BATE	BOOK	BURG	CHIX	TRQX	Overall
1 tick	19283	15124	16047	26037	21385	97876
2 ticks	694	223	356	308	233	1814
3 ticks	216	62	86	134	66	564
$>$ 3 ticks	216	193	448	109	271	1237
Total	20409	15602	16937	26588	21955	101491

Table 2. Jumps of Best Bid of ERICB on Five Exchanges in 09/2012

To ensure that the data are in precise accordance with the models, in the following numerical studies, jumps larger than 1 tick are adjusted to be 1 tick. Since the models describe the event times of jumps, not the sizes of jumps, the adjustment does not affect the estimation. In addition, as the event times in the original data are truncated to be in milliseconds, upward and downward jumps occur within the same millisecond

are then canceled one and another to reduce the possibility of multiple jumps occurring at the same time.

Assume N^1 and N^2 are both self-exciting and mutually-exciting. For efficiency in estimating the parameters, let $\alpha^{11} = \alpha^{22}$, $\alpha^{12} = \alpha^{21}$, $\beta^{11} = \beta^{22}$, $\beta^{12} = \beta^{21}$, and $\gamma^{11} = \gamma^{12} = \gamma^{21} = \gamma^{22}$ in (11) and (12), respectively.

For a general picture of price evolution, Figure 1 and Figure 2, respectively, illustrate the original evolution path, the adjusted evolution path, a randomly-simulated sample path, and a randomly-resimulated sample path based on the estimated parameters using the two models with exponential kernels and power-law kernels for the best bid of ERICB on CHIX on 06/09/2012.

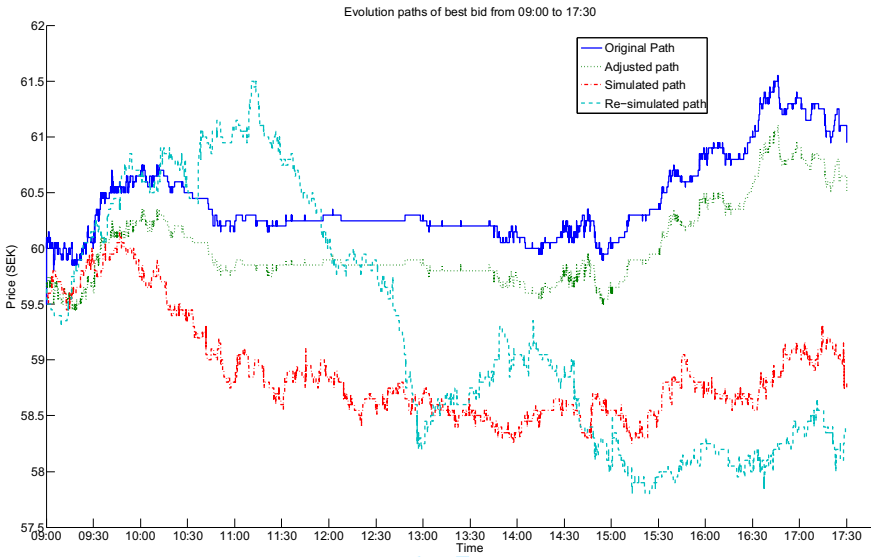

Figure 1. Price Evolution Paths Using a Model with Exponential Kernels for Best Bid of ERICB on CHIX on 06/09/2012

For a better resolution and an even clearer view, the corresponding sample paths for a two-hour period (10:00-12:00) are displayed in Figure 3 and Figure 4, respectively.

4.1. Goodness of Fit

For the goodness of fit of a model, residual analysis is often employed (Ogata, 1988). However, residual analysis is not consistently robust to reject the false model (Filimonov and Sornette, 2015). Meanwhile, the focus of this paper is on examining the two models to identify the more appropriate one for high frequency financial data. The analysis is hence based on a set of proposed measures instead, potentially to provide an alternative option for similar studies as well.

To study the goodness of fit of the two models with the real data, a number of measures are introduced first.

Denote the time interval as $(0, T]$ after being shifted. Let N^1 and N^2 be the numbers of upward and downward jumps, respectively, and $P_{\max}$ and $P_{\min}$ be the maximum and minimum prices observed, respectively, of the price evolution path $P = \{P_t\}_{t\in(0,T]}$ of the real data. Accordingly, $\hat{N}^1$, $\hat{N}^2$, $\hat{P}_{\max}$, and $\hat{P}_{\min}$ are the counterparts of a sample path $\hat{P} = \{\hat{P}_t\}_{t\in(0,T]}$ generated from a model with the estimation of the parameters

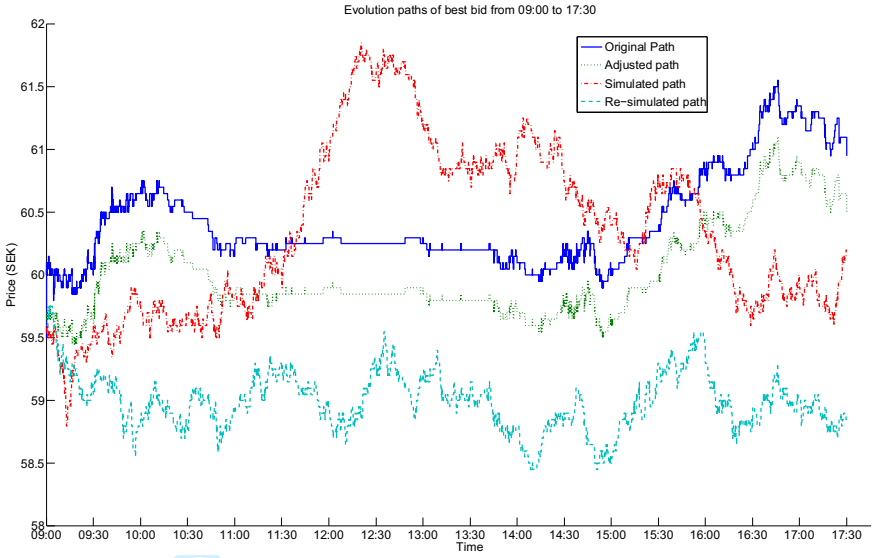

Figure 2. Price Evolution Paths Using a Model with Power-Law Kernels for Best Bid of ERICB on CHIX on 06/09/2012

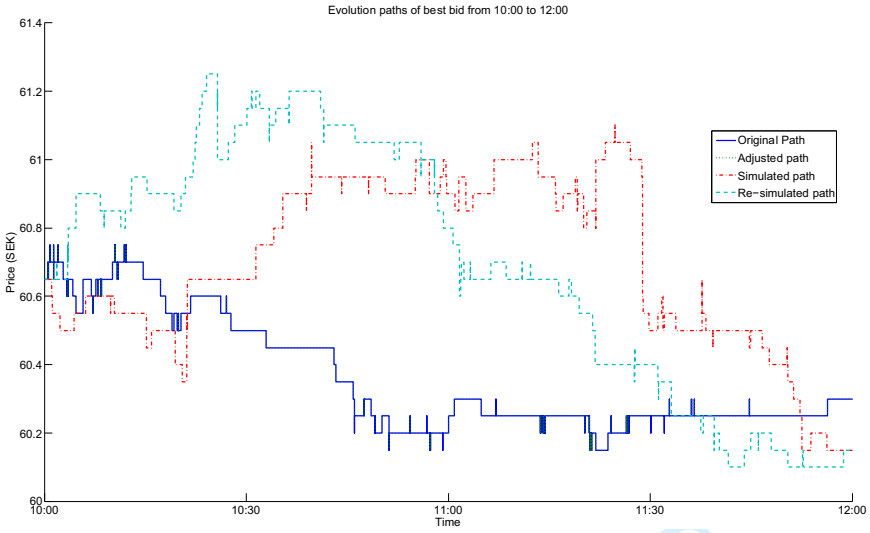

Figure 3. Price Evolution Paths Using a Model with Exponential Kernels for Best Bid of ERICB on CHIX in Two Hours on 06/09/2012

$(\hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\gamma})$. $\hat{N}_t$ is the estimated value of $\mathrm{E}[N_t]$ given in Proposition 2.1.

First, let

$$\hat{S}(L) = \frac{1}{L} \sum_{l=1}^{L} \frac{\hat{P}_{\max}(l) - \hat{P}_{\min}(l)}{P_{\max} - P_{\min}},$$

where L is the number of simulated paths, $\hat{P}_{\max}(l)$ and $\hat{P}_{\min}(l)$ are the highest and lowest prices observed from the lth simulated sample path, respectively. $\hat{S}$ thus mea-

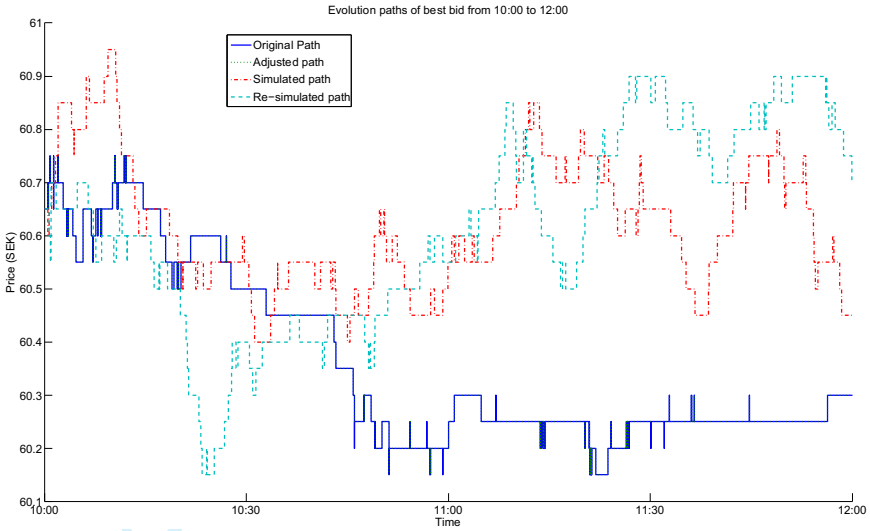

Figure 4. Price Evolution Paths Using a Model with Power-Law Kernels for Best Bid of ERICB on CHIX in Two Hours on 06/09/2012

sures the price fluctuation of the simulated paths relative to that of the real data. The closer $\hat{S}$ to 1 is, generally the more stable a model is to capture price movement.

For $i = 1, 2$, denote

$$\hat{R}^i(L) = \frac{1}{L}\sum_{l=1}^{L}\frac{|\hat{N}^i(l) - N^i|}{N^i}.$$

Here $\hat{R}^i$ indicates how far $\hat{N}^i$ diverges from N^i and so how well the underlying model fits the data in terms of number of jumps. The smaller $\hat{R}^i$ is, the less discrimination between $\hat{N}^i$ and N^i is and so the better the process models the intensity of jumps.

Next, define

$$\hat{U}^i(L) = \frac{1}{L}\sum_{l=1}^{L}\sum_{m=1}^{M}\frac{|\hat{N}^i_{(t_{m-1},t_m]}(l) - N^i_{(t_{m-1},t_m]}|}{N^i}, i = 1, 2$$

and

$$\hat{V}(L) = \frac{1}{L}\sum_{l=1}^{L}\int_0^T |\hat{P}_t(l) - P_t|dt,$$

where $\{t_m\}_{m=0,1,\ldots,M}$ is an even partition of the time interval $(0, T]$ with $t_0 = 0$ and $t_M = T$, $N^i_{(t_{m-1},t_m]}$ is the number of jumps within $(t_{m-1}, t_m]$ from the real data, and $\hat{N}^i_{(t_{m-1},t_m]}(l)$ is that of the lth simulated sample path. $\hat{U}^i, i = 1, 2$ hence measure the overall discrepancy between the simulated paths and the evolution path from the real data in both intensity and clustering of jumps. The smaller $\hat{U}^i, i = 1, 2$ are, the better a model fits the real data. In this paper, if without being specified explicitly, it is taken that $t_m - t_{m-1} = 1s$, $m = 1, \ldots, M$. From a different perspective, $\hat{V}$ measures

the total divergence of the simulated paths from the original evolution path. Hence in general, a model with a smaller $\hat{V}$ fits the real data better.

To assess the estimated results, a one-sample Student's t-test is undertaken for each $\hat{N}^i$, $i = 1, 2$ from each model. For each test, let $\bar{x}$ be the sample mean, s the sample standard deviation, and n the number of sample paths generated. Then, to verify the null hypothesis that the population mean associated with a model is equal to the corresponding value μ_0 of the real data, the t-statistic is obtained as

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}.$$

The degree of freedom is $n - 1$.

To investigate whether there is a discrepancy between the estimated results from the two models, i.e., the null hypothesis that the two population means of the two models are equal, a Welch's t-test is conducted for each of $\hat{N}^i$, $\hat{U}^i$, $\hat{S}$, and $\hat{V}$, $i = 1, 2$. Since the same number of sample paths n are generated for both models, for each test, the t-statistic is then

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\frac{\sqrt{s_1^2 + s_2^2}}{\sqrt{n}}}$$

and the degree of freedom is

$$\nu \approx \frac{(s_1^2 + s_2^2)^2}{\frac{s_1^4 + s_2^4}{n-1}} \geq n - 1,$$

$_i$, s_i^2, $i = 1, 2$ are the sample mean and sample variance estimated from the two models, respectively.

4.1.1. One Stock on One Trading Day

To evaluate the two models, 1000 sample paths are generated from each of them for the best bid of ERICB on the exchange CHIX in a two-hour time period (10:00-12:00) on September 6th, 2012. The estimated results, the Student's t-tests for $\hat{N}^i$, $i = 1, 2$, and the Welch's t-tests for the measures are provided in Tables 3, 4, 5, respectively. The Jarque-Bera tests given in Table 4 suggest that the numbers of jumps generally follow normal distributions, which is complementarily verified visually by the histograms of numbers of upward and downward jumps of the simulated paths from both models respectively shown in Figures 7, 8, 9, and 10.

As shown in Table 5, the p-values of both $\hat{N}^1$ and $\hat{N}^2$ are less than 0.01%, which indicates that there is a significant difference between the numbers of both upward and downward jumps estimated from the two models. This is further backed up by the results presented in Table 4. For the model with power-law kernels, the p-values of both $\hat{N}^1$ and $\hat{N}^2$ are greater than 0.1%, which means with a significance level 0.001, the null hypothesis that the mean numbers of both upward and downward jumps of the simulated paths equal the counterparts of the real data cannot be rejected. It is hence statistically verified that the intensity of jumps of the simulated path is in accordance with that of the evolution path of the data. Clearly, this is not the case for the model with exponential kernels, both p-values of which are far less than 0.01%, which tells

	Exponential Kernels		Power-Law Kernels	
N^1 N^2	60		67	
$P_{\max}$ $P_{\min}$	60.75		60.15	
$\hat{\mu}^1$ $\hat{\mu}^2$	0.00000524	0.00000513	0.00000835	0.00000920
$\hat{\alpha}^{11}$ $\hat{\alpha}^{12}$	0.00020516	0.00113956	0.0025147	0.00000001
$\hat{\beta}^{11}$ $\hat{\beta}^{12}$	0.00077190	0.00774702	1.36345147	1.92390056
$\hat{\gamma}^{11}$	–		0.99998405	
$\hat{N}_t^1$ $\hat{N}_t^2$	64.01972167	63.08666726	60.54599748	66.72305215

Table 3. MLE of Best Bid of ERICB on CHIX from 10:00 to 12:00 on 06/09/2012

	Exponential Kernels				Power-Law Kernels			
	t-test		Jarque-Bera Test		t-test		Jarque-Bera Test	
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
$\hat{N}^1$	9.437	–	15.021	0.0024	1.445	0.1489	5.136	0.0725
$\hat{N}^2$	-10.508	–	15.549	0.0021	-2.831	0.0047	1.034	0.5000

Table 4. Student's t-tests and Jarque-Bera Test for Numbers of Jumps of Best Bid of ERICB on CHIX from 10:00 to 12:00 on 06/09/2012

An entry "–" means that the p-value is less than 0.0001.

	Exponential Kernels		Power-Law Kernels		Welch's t Test	
	Mean	Variance	Mean	Variance	t-statistic	p-value
$\hat{N}^1$	63.43500000	132.49627127	60.34800000	58.03693293	7.072	–
$\hat{N}^2$	63.22500000	129.05943443	66.28000000	64.68628629	-6.941	–
$\hat{S}$	1.61083333	0.29481634	1.47408333	0.22615865	5.991	–
$\hat{U}^1$	2.04268333	0.03661336	1.99013333	0.01635233	7.220	–
$\hat{U}^2$	1.92888060	0.02885839	1.97250746	0.01431305	-6.640	–
$\hat{V}$	0.87027389	0.27629331	0.64937867	0.13496205	10.893	–

Table 5. Welch's t-tests for Best Bid of ERICB on CHIX from 10:00 to 12:00 on 06/09/2012

An entry "–" means that the p-value is less than 0.0001.

that the null hypothesis should be rejected under the same significance level.

In Table 5, the p-values of both $\hat{S}$ and $\hat{V}$ are less than 0.01%, which implies that the differences between the corresponding population means of the two models are significant. The sample means of both measures of the model with power-law kernels are less than the counterparts of the model with exponential kernels. This indicates that the model with power-law kernels is more stable in capturing the price movement and that on average the simulated paths diverge from the original price evolution path less than the one with exponential kernels. The only measure that does not differentiate the two models is $\hat{U}$. The model with power-law kernels outperforms the one with exponential kernels in $\hat{U}^1$ and vice versa in $\hat{U}^2$. A plausible explanation of this is that the two models result in different numbers of jumps of the simulated paths. Larger number of jumps potentially leads to larger value of $\hat{U}$, provided that jumps do not occur intensively.

From the statistical studies, it can thus be inferred that overall the model with power-law kernels fits the real data better than the one with exponential kernels. This is further confirmed by the implications concluded in Sections 4.1.2 and 4.1.3.

4.1.2. One Stock on Multiple Trading Days

For robustness test on different trading days, the two models are further assessed on data of all trading days in September 2012. For each trading day, 1000 sample paths are generated for the best bid of ERICB on the exchange CHIX in a six-hour time period (10:00-16:00). The Student's t-tests for $\hat{N}^i$, $i = 1, 2$ and the Welch's t-tests for the measures are reported in Tables 6 and 7, respectively.

Date	N^1	$\hat{N}^1$								N^2	$\hat{N}^2$							
		Exponential Kernels				Power-Law Kernels					Exponential Kernels				Power-Law Kernels			
		$\bar{x}$	s	t^*	p^*	$\bar{x}$	s	t^*	p^*		$\bar{x}$	s	t^*	p^*	$\bar{x}$	s	t^*	p^*
03/09	146	138.730	280	-13.728	–	146.334	143	0.884	0.3771	156	165.743	331	16.926	–	155.799	159	-0.504	0.6143
04/09	171	155.018	234	-33.004	–	171.320	184	0.745	0.4564	177	194.447	303	31.705	–	177.701	179	1.658	0.0976
05/09	328	323.807	961	-4.278	–	327.651	325	-0.612	0.5407	326	332.904	978	6.981	–	325.086	320	-1.615	0.1067
06/09	313	309.050	1067	-3.823	0.0001	311.990	305	-1.830	0.0676	309	315.863	1086	6.587	–	308.402	297	-1.096	0.2731
07/09	294	289.194	892	-5.089	–	293.689	288	-0.580	0.5622	323	325.662	888	2.825	0.0048	323.434	319	0.769	0.4423
10/09	236	209.105	414	-41.793	–	235.771	258	-0.451	0.6523	222	247.437	491	36.312	–	222.872	236	1.794	0.0731
11/09	281	247.828	585	-43.383	–	281.586	252	1.168	0.2433	267	301.312	649	42.608	–	267.522	267	1.010	0.3128
12/09	271	289.817	1397	15.922	–	272.087	258	2.139	0.0327	270	250.236	1202	-18.028	–	268.119	277	-3.576	0.0004
13/09	145	135.326	338	-16.630	–	145.017	149	0.044	0.9649	150	160.873	380	17.643	–	150.004	134	0.011	0.9912
14/09	206	187.393	452	-27.662	–	206.782	189	1.800	0.0721	187	207.505	457	30.317	–	187.890	183	2.079	0.0378
17/09	129	128.722	127	-0.779	0.4361	128.908	131	-0.254	0.7996	134	134.080	145	0.210	0.8334	133.974	132	-0.072	0.9429
18/09	195	181.518	458	-19.911	–	193.795	213	-2.613	0.0091	198	212.757	532	20.237	–	198.981	198	2.206	0.0276
19/09	181	160.914	272	-38.511	–	180.954	169	-0.112	0.9108	181	200.159	315	34.115	–	181.188	184	0.439	0.6609
20/09	124	117.751	195	-14.146	–	124.102	126	0.288	0.7737	117	123.278	209	13.735	–	116.279	113	-2.146	0.0321
21/09	254	244.183	595	-12.731	–	253.487	256	-1.014	0.3106	262	271.738	640	12.174	–	261.550	261	-0.881	0.3785
24/09	113	104.272	225	-18.410	–	112.958	109	-0.127	0.8987	122	127.745	248	11.546	–	121.734	121	-0.766	0.4440
25/09	138	137.068	395	-1.484	0.1382	137.665	138	-0.903	0.3670	149	148.897	428	-0.157	0.8749	148.674	154	-0.830	0.4065
26/09	214	197.781	313	-28.967	–	213.299	229	-1.464	0.1434	220	236.581	347	28.137	–	220.137	221	0.292	0.7706
27/09	185	173.736	442	-16.949	–	185.351	191	0.804	0.4217	197	208.626	512	16.251	–	197.108	202	0.240	0.8103
28/09	295	267.752	558	-36.480	–	294.481	310	-0.933	0.3512	297	326.710	611	38.013	–	297.256	311	0.459	0.6464

Table 6. Student's t-tests for Numbers of Jumps of Best Bid of ERICB on CHIX from 10:00 to 16:00 in 09/2012

An entry "–" means that the p-value is less than 0.0001.

Date	$\hat{N}^1$		$\hat{N}^2$		$\hat{S}$		$\hat{R}^1$		$\hat{R}^2$		$\hat{U}^1$		$\hat{U}^2$		$\hat{V}$	
	t^*	p^*	t^*	p^*	t^*	p^*	t^*	p^*	t^*	p^*	t^*	p^*	t^*	p^*	t^*	p^*
03/09	-11.687	–	14.201	–	13.243	–	12.332	–	13.403	–	-11.254	–	14.511	–	2.945	0.0033
04/09	-25.185	–	24.133	–	28.579	–	16.729	–	18.455	–	-24.696	–	24.527	–	25.783	–
05/09	-3.390	0.0007	6.860	–	3.247	0.0012	15.605	–	15.397	–	-2.456	0.0141	7.639	–	4.591	–
06/09	-2.510	0.0122	6.344	–	-10.268	–	17.261	–	17.486	–	-1.306	0.1919	7.344	–	-10.658	–
07/09	-4.139	–	2.028	0.0427	-2.851	0.0044	16.483	–	14.357	–	-3.306	0.0010	2.873	0.0041	0.220	0.8258
10/09	-32.524	–	28.813	–	15.754	–	26.193	–	23.717	–	-31.491	–	29.245	–	40.489	–
11/09	-36.909	–	35.311	–	33.095	–	31.032	–	29.177	–	-34.920	–	35.655	–	52.435	–
12/09	13.782	–	-14.708	–	26.767	–	23.831	–	24.453	–	14.134	–	-13.891	–	23.326	–
13/09	-13.884	–	15.158	–	14.124	–	15.773	–	17.370	–	-13.501	–	15.243	–	11.430	–
14/09	-24.214	–	24.506	–	-14.414	–	21.703	–	22.224	–	-23.838	–	24.372	–	35.609	–
17/09	-0.366	0.7145	0.202	0.8403	-0.118	0.9061	-0.526	0.5992	1.865	0.0623	-0.402	0.6878	0.190	0.8490	-0.112	0.9107
18/09	-14.986	–	16.129	–	12.797	–	15.800	–	17.479	–	-14.364	–	16.156	–	14.869	–
19/09	-30.189	–	26.857	–	30.349	–	22.566	–	20.159	–	-29.340	–	26.363	–	28.770	–
20/09	-11.211	–	12.337	–	-7.742	–	10.030	–	10.622	–	-10.874	–	12.492	–	12.220	–
21/09	-10.089	–	10.735	–	9.298	–	14.959	–	14.837	–	-9.014	–	11.540	–	5.766	–
24/09	-15.039	–	9.906	–	9.383	–	15.030	–	11.175	–	-15.029	–	9.968	–	-6.245	–
25/09	-0.818	0.4133	0.292	0.7701	-7.709	–	15.056	–	14.100	–	-0.372	0.7101	0.900	0.3680	-6.258	–
26/09	-21.066	–	21.822	–	19.980	–	15.369	–	14.801	–	-20.402	–	21.504	–	25.517	–
27/09	-14.607	–	13.630	–	16.733	–	15.890	–	15.829	–	-13.774	–	13.849	–	5.212	–
28/09	-28.697	–	30.674	–	38.936	–	23.300	–	24.349	–	-27.740	–	30.157	–	42.232	–

Table 7. Welch's t-tests for Best Bid of ERICB on CHIX from 10:00 to 16:00 in 09/2012

An entry "–" means that the p-value is less than 0.0001.

In Tables 6, 7 and the following Tables 8, 9, $\bar{x}$ and s respectively represent the sample mean and sample standard deviation of the corresponding $\hat{N}^1$ or $\hat{N}^2$, t^* the t-statistic, and p^* the p-value.

As shown in Table 6, all forty p-values except six resulting from the model with exponential kernels are less than 0.0001. In the meantime, the p-values of both $\hat{N}^1$ and $\hat{N}^2$ for all trading days from the model with power-law kernels are greater than 0.0001. This implies that with a significance level 0.0001, the null hypothesis that the population means associated with the simulated paths equal the counterparts observed from the real data should be rejected for almost all days for the model with

exponential kernels and should not be rejected for any trading day for the model with power-law kernels. This further restates what concluded from Table 4. Moreover, all the forty p-values, except the one on $\hat{N}^2$ of 25/09, from the model with exponential kernels are less than the corresponding ones from the model with power-law kernels. This is additionally illustrated by Figures 5, 6. That is, the curves resulting from the model with power-law kernels almost surely overlap with the corresponding ones from the real data, whereas those resulting from the model with exponential kernels clearly sway away to some degree. Moreover, for the model with power-law kernels, the null hypothesis that the population means associated with the simulated paths equal the counterparts estimated from the real data should not be rejected for all forty cases except two with a significance level 0.01 and for all forty cases except six with a significance level 0.05. All these indicate that the model with power-law kernels is consistently more stable than the one with exponential kernels in describing and hence predicting the intensity of jumps for high frequency financial data.

From a closely related while slightly different perspective, Table 7 tells that there is a significant discrepancy in the measures resulting from the two models. For $\hat{U}^1$ and $\hat{U}^2$, the same implication holds as that in Section 4.1.1. Meanwhile, the resulting measure from the model with power-law kernels, based on the t-statistics on the others, namely $\hat{R}^1$, $\hat{R}^2$, $\hat{S}$, and $\hat{V}$, apparently outperforms that from the model with exponential kernels in most cases. In particular, among the forty t-statistics related to $\hat{R}^i, i = 1, 2$, only one is nonpositive but still close to zero. The resulting t-statistics of all trading days, except 17/09, are greater than 10. This further indicates that the model with power-law kernels captures the intensity of jumps as well as the other features exhibited by real data significantly better than the model with exponential kernels, which is in consistence with what implied in Section 4.1.1.

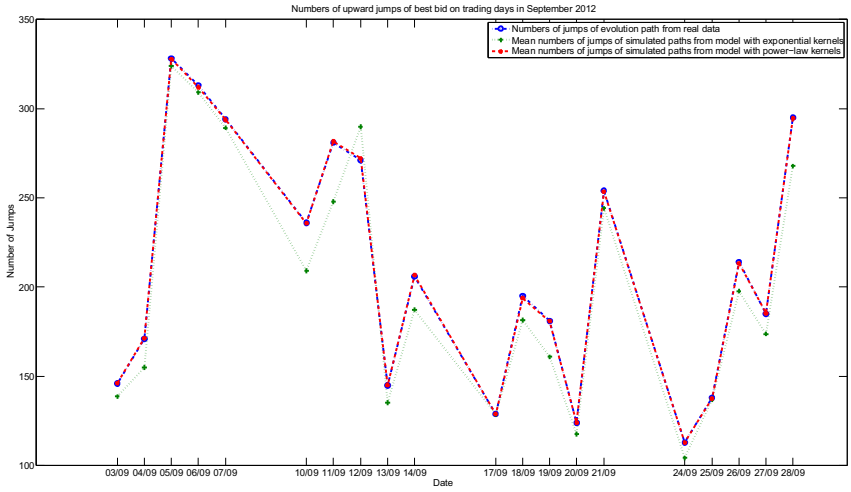

Figure 5. Numbers of Upward Jumps of Best Bid of ERICB on Trading Days in 09/2012

4.1.3. Multiple Stocks on One Trading Day

For robustness test on different stocks, the two models are examined further by using data on October 1st, 2012 of ten stocks among those most liquid ones on the exchange CHIX, including ATCOA (Atlas Copco), ELUXB (Electrolux), ERICB (Telefonak-

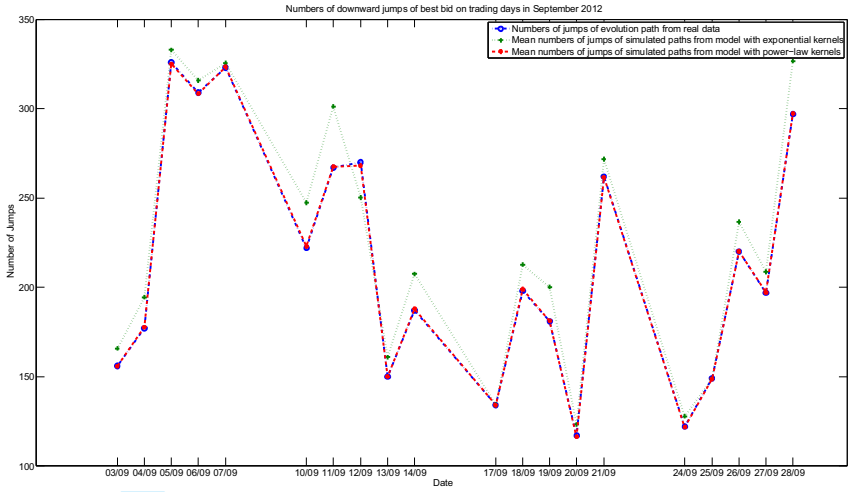

Figure 6. Numbers of Downward Jumps of Best Bid of ERICB on Trading Days in 09/2012

tiebolaget LM Ericsson), HMB (Hennes & Mauritz), NDA (Nordea Bank), SAND (Sandvik), SEBA (Skandinaviska Enskilda Banken), SKFA (SKF), SWEDA (Swedbank), and VOLVB (Volvo). The results are presented in Tables 8 and 9.

Stock	N^1	$\hat{N}^1$								N^2	$\hat{N}^2$							
		Exponential Kernels				Power-Law Kernels					Exponential Kernels				Power-Law Kernels			
		$\bar{x}$	s	t^*	p^*	$\bar{x}$	s	t^*	p^*		$\bar{x}$	s	t^*	p^*	$\bar{x}$	s	t^*	p^*
ATCOA	275	277.371	524	3.276	0.0011	273.259	257	-3.437	0.0006	260	258.492	479	-2.178	0.0296	259.534	235	-0.960	0.3371
ELUXB	241	242.553	454	2.305	0.0214	240.782	239	-0.446	0.6555	228	227.788	390	-0.339	0.7343	227.965	227	-0.073	0.9414
ERICB	144	148.147	261	8.113	–	142.884	137	-3.014	0.0026	150	146.185	257	-7.520	–	150.153	153	0.391	0.6958
HMB	383	417.958	1219	31.660	–	382.331	362	-1.112	0.2666	404	365.619	1033	-37.762	–	403.370	395	-1.003	0.3162
NDA	102	108.542	170	15.884	–	102.406	109	1.227	0.2201	101	93.740	141	-19.314	–	101.113	102	0.354	0.7238
SAND	331	344.830	665	16.958	–	329.040	341	-3.354	0.0008	328	314.909	604	-16.837	–	328.716	338	1.231	0.2186
SEBA	182	196.789	380	23.997	–	180.806	173	-2.868	0.0042	179	163.782	319	-26.935	–	177.365	175	-3.908	0.0001
SKFA	135	131.414	270	-6.907	–	133.941	127	-2.973	0.0030	129	132.619	264	7.038	–	128.287	125	-2.015	0.0441
SWEDA	160	175.256	366	25.216	–	159.653	160	-0.866	0.3866	150	135.384	285	-27.388	–	149.995	155	-0.013	0.9899
VOLVB	353	381.474	749	32.895	–	353.295	331	0.513	0.6082	365	336.849	626	-35.591	–	364.266	357	-1.228	0.2198

Table 8. Student's t-tests for Numbers of Jumps of Best Bid of Ten Stocks from 10:00 to 16:00 on 01/10/2012

An entry "–" means that the p-value is less than 0.0001.

Stock	$\hat{N}^1$		$\hat{N}^2$		$\hat{S}$		$\hat{R}^1$		$\hat{R}^2$		$\hat{U}^1$		$\hat{U}^2$		$\hat{V}$	
	t^*	p^*	t^*	p^*	t^*	p^*	t^*	p^*	t^*	p^*	t^*	p^*	t^*	p^*	t^*	p^*
ATCOA	4.655	–	-1.233	0.2179	5.524	–	10.020	–	10.542	–	5.360	–	-0.324	0.7460	1.961	0.0500
ELUXB	2.128	0.0335	-0.225	0.8217	4.255	–	9.049	–	8.038	–	3.068	0.0022	0.914	0.3607	2.290	0.0221
ERICB	8.338	–	-6.193	–	-6.640	–	10.183	–	9.185	–	8.698	–	-5.537	–	6.646	–
HMB	28.331	–	-31.594	–	22.006	–	25.712	–	28.449	–	29.987	–	-28.364	–	42.212	–
NDA	11.614	–	-14.943	–	12.911	–	8.937	–	10.388	–	11.692	–	-14.750	–	9.589	–
SAND	15.738	–	-14.220	–	15.832	–	13.641	–	12.061	–	16.470	–	-13.310	–	11.630	–
SEBA	21.491	–	27.533	–	-19.320	–	16.683	–	17.890	–	21.856	–	-18.684	–	20.180	–
SKFA	-4.013	–	6.941	–	-4.626	–	11.306	–	10.651	–	-3.567	0.0004	7.613	–	0.477	0.6333
SWEDA	21.502	–	-22.036	–	35.621	–	17.621	–	18.161	–	21.160	–	-21.639	–	17.786	–
VOLVB	27.112	–	-27.653	–	23.106	–	23.394	–	22.469	–	28.407	–	-25.194	–	29.110	–

Table 9. Welch's t-tests for Best Bid of Ten Stocks from 10:00 to 16:00 on 01/10/2012

An entry "–" means that the p-value is less than 0.0001.

From the study, it is supplementarily indicated again that the model with power-law kernels is a more appropriate one for high frequency data than the one with exponential kernels. As can be noted from Table 8, the intensities of jumps of the simulated paths from the model with power-law kernels consistently match those of the real data, which can be told directly from the fact that all the resulting p-values are no less than 0.0001,

while those from the model with exponential kernels do not, which is evidenced by the fact that the majority of the resulting p-values are significantly less than 0.0001. From Table 9, same observations and implications can be obtained as those from Table 5 and Table 7. In particular, all resulting t-statistics related to $\hat{R}^i, i = 1, 2$ are close to or greater than 10.

4.2. Computational Time

One possible disadvantage of power-law kernels, compared with exponential kernels, could lie in the efficiency to obtain the maximum likelihood estimator, due to the fact that the log-likelihood function of a model with exponential kernels can be calculated recursively (Ogata, 1981).

For a model with exponential kernels, at each iteration, the number of multiplications for evaluating the log-likelihood function is in the scale of $\mathcal{O}(1)$. For a model with power-law kernels, at each iteration, the log-likelihood function has to be reckoned explicitly and the number of multiplications is in the scale of $\mathcal{O}(n)$, where n is the number of jumps. The computational times (in seconds) of MATLAB R2012b using global search for 100 simulations, in a machine with CPU @2200MHz (AMD Opteron^{TM} Processor 6274) and RAM @128GB, to search for the maximum likelihood estimators of the best bid of ERICB on CHIX in a one-hour time period (10:00-11:00) on September 6th, 2012, are shown in Table 10.

	Exponential Kernels		Power-Law Kernels	
	Mean	Variance	Mean	Variance
Time (Seconds)	34.43669765	168.75639050	674.80850479	39826.69494600

Table 10. Computational Time in Seconds for Maximum Likelihood Estimator of Best Bid of ERICB on CHIX in One Hour on 06/09/2012

As indicated by Table 10, the evaluation is indeed relatively more efficient with exponential kernels than that with power-law kernels. The difference may potentially be reduced by looking into more advanced algorithms.

4.3. Search Algorithm

When an algorithm is employed to determine the maximum likelihood estimator, MATLAB generally adopts two options, local search and global search. The latter is used for all results presented so far. To assess the two options, the estimated parameters as well as computational times based on 100 simulations using a model with power-law kernels for the un-adjusted best bid of ERICB on CHIX in a one-hour time period (10:00-11:00) on September 6th, 2012 are reported in Table 11.

Table 11 shows that between the two options, while the resulting maximum likelihood estimators and properties of simulated paths are comparable, local search is significantly more efficient than global search.

In the case that there is significant difference in the estimated results, one option to accelerate the convergence of global search is to apply advanced algorithms such as parallel computing.

	Local Search		Global Search	
	Mean	Variance	Mean	Variance
N^1	78			
N^2	76			
$\hat{\mu}^1$	0.00002115	0.00000000	0.00002115	0.00000000
$\hat{\alpha}^{11}$	0.00135989	0.00000000	0.00131769	0.00000028
$\hat{\alpha}^{12}$	0.00135998	0.00000000	0.00140206	0.00000028
$\hat{\beta}^{11}$	1.26455344	0.00000000	1.30029038	0.01640706
$\hat{\beta}^{12}$	1.26455472	0.00000000	1.27752660	0.00704741
$\hat{\gamma}^{11}$	0.99993098	0.00000000	0.99999971	0.00000000
Time (Seconds)	9.03369287	0.02069485	674.80850479	39826.69494600
$\hat{N}^1$	77.07000000	86.51020202	76.92000000	66.78141414
$\hat{N}^2$	76.18000000	76.06828283	76.38000000	69.89454545

Table 11. Estimated Results and Computational Times Using a Model with Power-Law Kernels for Best Bid of ERICB on CHIX in One Hour on 06/09/2012

5. Conclusion

Hawkes processes with exponential kernels and with power-law kernels are studied from a variety of perspectives, including maximum likelihood estimation for statistical inference and sample-path generation for computer simulation. For consistence with the setting-up of the models, data are first pre-processed, which essentially does not affect the results. To examine the goodness of fit of the two models for high frequency financial data, a number of measures are introduced and statistical tests are proposed accordingly. It is robustly demonstrated by empirical studies based on one stock on one trading day, one stock on multiple trading days, and multiple stocks on one trading day that, although with the former it is relatively more efficient in determining the maximum likelihood estimator when using global search, the latter fits real data notably better than the former. It is also shown that there is no significant difference in terms of optimality between local search and global search for the maximum likelihood estimator, while it is significantly more efficient by using a local search algorithm. This suggests that a Hawkes-based model with power-law kernels shall be an appropriate choice for high frequency financial data and that a local search algorithm may be used for maximum likelihood estimation in the first instance. The proposed analysis integratingly from both statistical and computational perspectives may also provide an alternative for similar studies.

The studies are largely based on models for one stock with jumps of 1 tick at most on one exchange. There are several possible future directions for further work, including the cases that one stock in different time intervals, different stocks in the same time interval, and one stock in different markets. A few examples are outlined here.

For simplicity, it is assumed that $\alpha^{11} = \alpha^{22}$, $\alpha^{12} = \alpha^{21}$, $\beta^{11} = \beta^{22}$, $\beta^{12} = \beta^{21}$, and $\gamma^{11} = \gamma^{12} = \gamma^{21} = \gamma^{22}$ in (11) and (12). Certainly by intuition as well as empirical observations, the assumptions are not unreasonable. That is, the intensities of upward and downward jumps do not diverse from each other in general. Meanwhile, it is

a full model to be considered is to remove the assumptions. This may be particularly relevant in examining the robustness of the models during, for example, financial crises, when the market is more moving in one direction than the other. Meanwhile, in this

case, it is expected that the evaluation for the maximum likelihood estimator will be slowed down accordingly.

It is indicated by Figure 1 or Figure 2 as well as the price evolution of different stocks on different trading days that the intensities of jumps in different time intervals follow non-identical patterns. This hints that it is worth considering dividing a whole trading day into sub-intervals and modeling them separately.

In reality, the price may move by more than 1 tick, as shown in Tables 1, 2. Suppose that the price of an asset moves up to d ticks for a jump, then the price can be represented as a multivariate Hawkes model,

$$P_t = P_0 + \sum_{i=1}^{d} i \cdot N_t^{i,1} - \sum_{i=1}^{d} i \cdot N_t^{i,2},$$

where P_0 is the price at time 0, $N^{i,1}$ is the number of upward jumps with i ticks, and $N^{i,2}$ is that of downward jumps with i ticks between 0 and t, $i = 1, \ldots, d$. The intensities of $N^{i,1}$ and $N^{i,2}$ are $\lambda^{i,1}$ and $\lambda^{i,2}$, respectively,

$$\lambda_t^{i,k} = \mu^{i,k} + \sum_{j=1}^{d} \int_0^t \phi_{t-s}^{ij,k1} dN_s^{j,1} + \sum_{j=1}^{d} \int_0^t \phi_{t-s}^{ij,k2} dN_s^{j,2}, k = 1, 2. \tag{13}$$

It is not uncommon that an asset is traded on multiple exchanges. For example, ERICB is traded in five markets, as shown in Table 1. Assume that an asset is traded on d exchanges and that the price moves by 1 tick at most. Denote the price as

$$P_t = P_0 + \sum_{i=1}^{d} N_t^{i,1} - \sum_{i=1}^{d} N_t^{i,2},$$

where P_0 is the price at time 0, $N^{i,1}$ is the number of upward jumps on exchange i, and $N^{i,2}$ is that of downward jumps on exchange i between 0 and t, $i = 1, \ldots, d$. The intensities of $N^{i,1}$ and $N^{i,2}$ are $\lambda^{i,1}$ and $\lambda^{i,2}$, respectively, as defined in (13).

It has been widely recognized that the price evolution of a stock is heavily affected by large orders. Another direction is then to take into consideration the dimension of market impact (Almgren and Chriss, 2000; Bertsimas and Lo, 1998).

Acknowledgements

This research has been supported by a grant from Riksbankens Jubileumsfond (P10 - 0113:1) and in part by the Curtin Sarawak Start-Up Research Fund. The empirical analysis in this paper is carried out by using the high frequency data supplied by NasdaqOMX in Stockholm, which is greatly acknowledged. The authors thank Björn Blomqvist for helping sort out one set of data files. The authors gratefully appreciate the two anonymous referees and the associate editor for reviewing the manuscript and for providing valuable comments and suggestions. The authors would also like to express their gratitude to Professor Zhe George Zhang for all the communication in the course of handling the submission.

References

Abergel, F., Anane, M., Chakraborti, A., Jedidi, A., and Muni Toke, I. (2016). *Limit Order Books*. Physics of Society: Econophysics and Sociophysics. Cambridge University Press.

Abergel, F. and Jedidi, A. (2011). A mathematical approach to order book modelling. In Abergel, F., Chakrabarti, B. K., Chakraborti, A., and Mitra, M., editors, *Econophysics of Order-driven Markets*, New Economic Windows, pages 93–107. Springer Milan.

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(2):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.

Alfonsi, A., Fruth, A., and Schied, A. (2010). Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, 10(2):143–157.

Almgren, R. and Chriss, N. (2000). Optimal execution of portfolio transactions. *Journal of Risk*, 3(2):5–39.

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*, 85(5):1–12.

Bacry, E., Delattre, S., Hoffmann, M., and Muzy, J.-F. (2013a). Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13(1):65–77.

Bacry, E., Delattre, S., Hoffmann, M., and Muzy, J.-F. (2013b). Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, 123(7):2475–2499.

Bacry, E., Jaisson, T., and Muzy, J.-F. (2016). Estimation of slowly decreasing Hawkes kernels: application to high-frequency order book dynamics. *Quantitative Finance*, 16(8):1179–1201.

Bacry, E., Mastromatteo, I., and Muzy, J.-F. (2015). Hawkes processes in finance. *Market Microstructure and Liquidity*, 01(01):1550005.

Bacry, E. and Muzy, J.-F. (2014). Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 14(7):1147–1166.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536.

Bauwens, L. and Hautsch, N. (2009). Modelling financial high frequency data using point processes. In Mikosch, T., Kreiß, J.-P., Davis, R. A., and Andersen, T. G., editors, *Handbook of Financial Time Series*, pages 953–979. Springer Berlin Heidelberg.

Bertsimas, D. and Lo, A. W. (1998). Optimal control of execution costs. *Journal of Financial Markets*, 1(1):1–50.

Bormetti, G., Calcagnile, L. M., Treccani, M., Corsi, F., Marmi, S., and Lillo, F. (2015). Modelling systemic price cojumps with Hawkes factor models. *Quantitative Finance*, 15(7):1137–1156.

Bouchaud, J.-P., Farmer, J. D., and Lillo, F. (2009). How markets slowly digest changes in supply and demand. In Hens, T. and Schenk-Hoppe, K. R., editors, *Handbook of financial markets: dynamics and evolution*, pages 57–160. Elsevier, North-Holland.

Bouchaud, J.-P., Mézard, M., and Potters, M. (2002). Statistical properties of stock order books: empirical results and models. *Quantitative Finance*, 2(4):251–256.

Bowsher, C. G. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141(2):876–912.

Brémaud, P. and Massoulié, L. (2002). Power spectra of general shot noises and Hawkes point processes with a random excitation. *Advances in Applied Probability*, 34(1):205–222.

Chakraborti, A., Toke, I. M., Patriarca, M., and Abergel, F. (2011). Econophysics review: I. empirical facts. *Quantitative Finance*, 11(7):991–1012.

Chavez-Demoulin, V., Davison, A. C., and McNeil, A. J. (2005). Estimating value-at-risk: a point process approach. *Quantitative Finance*, 5(2):227–234.

Chavez-Demoulin, V. and McNeil, A. J. (2012). High-frequency financial data modeling using Hawkes processes. *Journal of Banking & Finance*, 36(12):3415–3426.

Coleman, R. and Gastwirth, J. (1969). Some models for interaction of renewal processes related to neuron firing. *Journal of Applied Probability*, 6:38–58.

Cont, R. (2011). Statistical modeling of high-frequency financial data. *IEEE Signal Processing Magazine*, 28(5):16–25.

Cont, R., Stoikov, S., and Talreja, R. (2010). A stochastic model for order book dynamics. *Operations Research*, 58(3):549–563.

Daley, D. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes*, volume I: Elementary Theory and Methods. Springer, 2nd edition.

Daley, D. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes*, volume II: General Theory and Structure. Springer, 2nd edition.

Doyne Farmer, J., Gillemot, L., Lillo, F., Mike, S., and Sen, A. (2004). What really causes large price changes? *Quantitative Finance*, 4(4):383–397.

Embrechts, P., Liniger, T., and Lin, L. (2011). Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability*, 48:367–378.

Epps, T. W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74(366):291–298.

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

Filimonov, V., Bicchetti, D., Maystre, N., and Sornette, D. (2014). Quantification of the high level of endogeneity and of structural regime shifts in commodity markets. *Journal of International Money and Finance*, 42(0):174–192.

Filimonov, V. and Sornette, D. (2012). Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E*, 85:056108.

Filimonov, V. and Sornette, D. (2015). Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data. *Quantitative Finance*, 15(8):1293–1314.

Foucault, T., Kadan, O., and Kandel, E. (2005). Limit order book as a market for liquidity. *The Review of Financial Studies*, 18(4):1171–1217.

Giesecke, K., Goldberg, L. R., and Ding, X. (2011). A top-down approach to multiname credit. *Operations Research*, 59(2):283–300.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90.

Hawkes, A. G. (2018). Hawkes processes and their applications to finance: a review. *Quantitative Finance*, 18(2):193–198.

Hawkes, A. G. and Oakes, D. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503.

Hollifield, B., Miller, R. A., and Sandås, P. (2004). Empirical analysis of limit order markets. *The Review of Economic Studies*, 71(4):1027–1063.

Isham, V. and Westcott, M. (1979). A self-correcting point process. *Stochastic Processes and their Applications*, 8(3):335–347.

Khinchine, A. Y. (1969). *Mathematical methods in the theory of queueing*. Hafner Publishing Co., New York. Translated from the Russian by D. M. Andrews and M. H. Quenouille, Second edition, with additional notes by Eric Wolman, Griffin's Statistical Monographs and Courses, No. 7.

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

Lee, K. and Seo, B. K. (2017). 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.

Lewis, P. A. W. (1969). Asymptotic properties and equilibrium conditions for branching Poisson processes. *Journal of Applied Probability*, 6:355–371.

Lewis, P. A. W. and Shedler, G. S. (1976). Simulation of nonhomogeneous Poisson processes

with log linear rate function. *Biometrika*, 63(3):501–505.
Lewis, P. A. W. and Shedler, G. S. (1979a). Simulation of Nonhomogeneous Poisson Processes by Thinning. *Naval Research Logistics Quarterly*, 26(3):403–413.
Lewis, P. A. W. and Shedler, G. S. (1979b). Simulation of nonhomogeneous Poisson processes with degree-two exponential polynomial rate function. *Operations Research. The Journal of the Operations Research Society of America*, 27(5):1026–1040.
Luckock, H. (2003). A steady-state model of the continuous double auction. *Quantitative Finance*, 3(5):385–404.
Maslov, S. and Mills, M. (2001). Price fluctuations from the order book perspective - Empirical facts and a simple model. *Physica A*, 299:234–246.
Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., and Tita, G. E. (2011). Self-Exciting Point Process Modeling of Crime. *Journal of the American Statistical Association*, 106(493):100–108.
Obizhaeva, A. A. and Wang, J. (2013). Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1):1–32.
Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 30(1):243–261.
Ogata, Y. (1981). On Lewis' Simulation Method for Point Processes. *IEEE Transactions on Information Theory*, 27(1):23–30.
Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27.
Ogata, Y. (1999). Seismicity analysis through point-process modeling: A review. *Pure and Applied Geophysics*, 155(2-4):471–507.
Ogata, Y. and Vere-Jones, D. (1984). Inference for earthquake models: a self-correcting model. *Stochastic Processes and their Applications*, 17(2):337–347.
Ozaki, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, 31(1):145–155.
Parlour, C. A. (1998). Price dynamics in limit order markets. *Review of Financial Studies*, 11(4):789–816.
Reynaud-Bouret, P. and Schbath, S. (2010). Adaptive estimation for Hawkes processes; application to genome analysis. *The Annals of Statistics*, 38:2781–2822.
Robert, C. Y. and Rosenbaum, M. (2011). A new approach for the dynamics of ultra-high-frequency data: The model with uncertainty zones. *Journal of Financial Econometrics*, 9(2):344–366.
Rosenbaum, M. (2011). A new microstructure noise index. *Quantitative Finance*, 11(6):883–899.
Rosu, I. (2009). A dynamic model of the limit order book. *Review of Financial Studies*, 22(11):4601–4641.
Smith, E., Farmer, D. J., Gillemot, L., and Krishnamurthy, S. (2003). Statistical theory of the continuous double auction. *Quantitative Finance*, 3:481–514.
Vere-Jones, D. (1970). Stochastic models for earthquake occurrence. *Journal of the Royal Statistical Society. Series B (Methodological)*, 32(1):1–62.
Zhang, C. (2016). Modeling High Frequency Data Using Hawkes Processes with Power-Law Kernels. *Procedia Computer Science*, 80:762–771.
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(1):99–136.
Zhuang, J., Ogata, Y., and Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97(458):369–380.

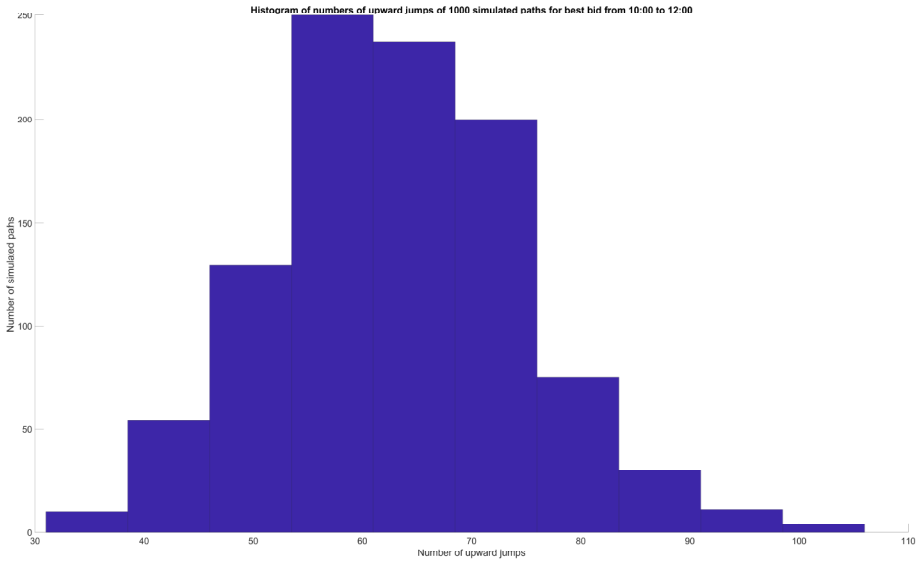

Figure 7. Histogram of Numbers of Upward Jumps of 1000 Sample Paths from a Model with Exponential Kernels for Best Bid of ERICB on CHIX from 10:00 to 12:00 on 06/09/2012

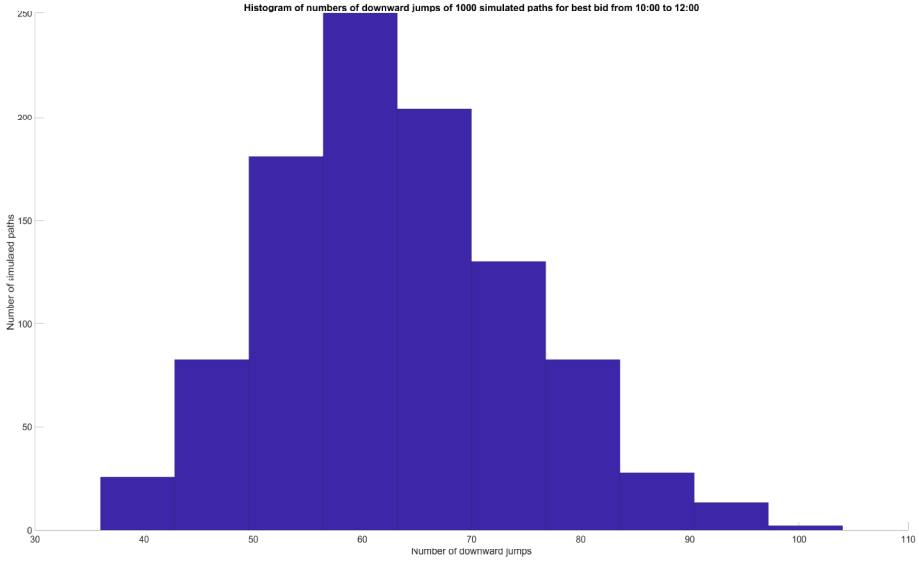

Figure 8. Histogram of Numbers of Downward Jumps of 1000 Sample Paths from a Model with Exponential Kernels for Best Bid of ERICB on CHIX from 10:00 to 12:00 on 06/09/2012

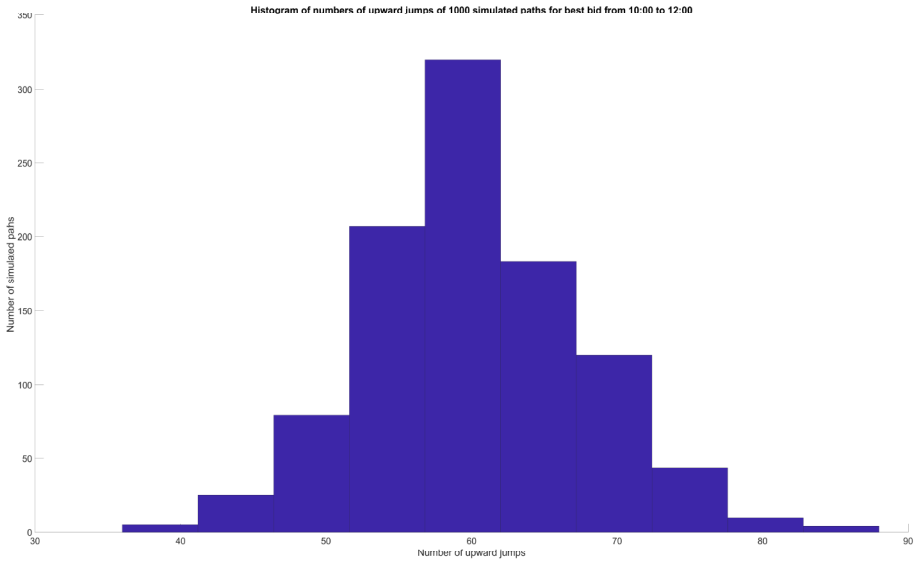

Figure 9. Histogram of Numbers of Upward Jumps of 1000 Sample Paths from a Model with Power-Law Kernels for Best Bid of ERICB on CHIX from 10:00 to 12:00 on 06/09/2012

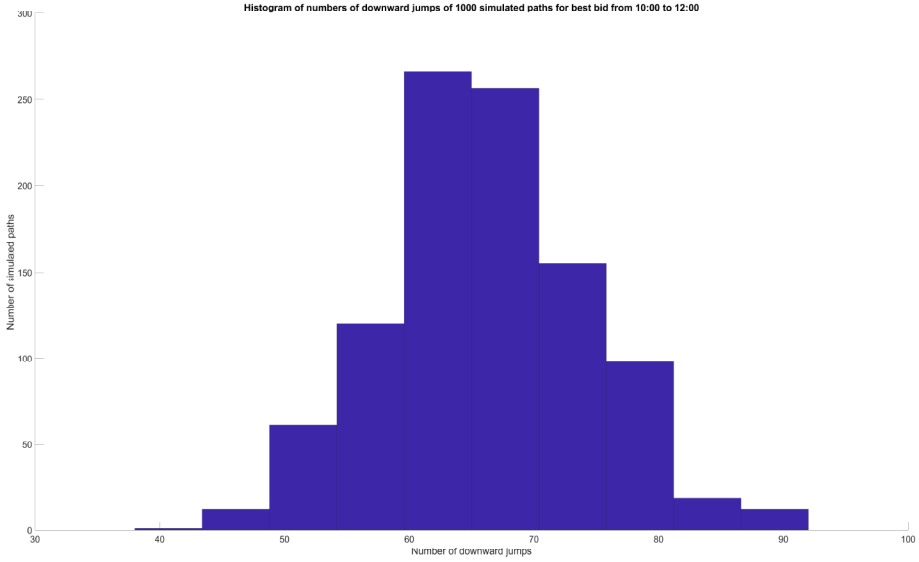

Figure 10. Histogram of Numbers of Downward Jumps of 1000 Sample Paths from a Model with Power-Law Kernels for Best Bid of ERICB on CHIX from 10:00 to 12:00 on 06/09/2012