Hawkes processes with hidden marks

Yuzhi Cai* , Swansea University, UK

Abstract

We develop a novel Hawkes process (HP) model with hidden marks for financial event data, where the hidden marks are used to take account the effect of some extra random errors (ERE) caused by data collection mechanisms and some data cleaning procedures. We further propose a Bayesian method for parameter estimation. We use simulation studies and two data applications to evaluate the performance of the estimation method and the impact of ERE on the intensity of an underlying financial process and explain how to use the proposed model in practice. Our results show that the proposed estimation method works well, and they also confirm that when ERE cause information about the underlying process to be lost, the intensity function may be underestimated. We further find that the proposed model performs better in the presence of ERE compared with the standard HP model.

Key words: Hawkes process model; intensity function; quantile function; parameter estimation, hidden marks.

JEL classification: C58, G17

*Corresponding author. School of Management, Swansea University, Swansea, SA1 8EN, UK. Email: y.cai@swansea.ac.uk

1 Introduction

Hawkes process (HP) models, first proposed by Hawkes (1971a, 1971b), have become more and more popular in finance recently, see Hawkes (2018) for an excellent review. They are popular mainly because they have the flexibility and advantages of using the class of counting processes that can be specified by a conditional intensity function for financial market event data.

For example, Bowsher (2007) proposed a generalized HP model by using a vector conditional intensity to incorporate inhibitory events and dependence between trading days. Large (2007) extended the work of Bowsher (2007) to model the actions of liquidity providers in the construction process of the order book. Chavez-Demoulin and McGill (2012) used HP models to make inferences on instantaneous conditional Value-at-Risk. Bacry and Muzy (2014) introduced a HP model that accounts for the dynamics of market prices through the impact of market order arrivals at the microstructural level. Bauwens and Hautsch (2006) proposed a new type of stochastic intensity model in order to capture a common latent component in point processes. They assumed that the latent component follows a log-normal distribution, whose mean follows an autoregressive process, and the intensity function is driven not only by the observed process history but also by the dynamic latent component. They further applied their model to estimate the price intensities based on NYSE trading. Many other HP models have also been developed for financial data. See Bacry et al. (2015) and the references therein.

It is worth noting that the work discussed above focuses on how to capture important components (observable or latent) of an underlying financial process through its intensity function in the presence of randomness in realized jumps, and whether the formula designed for the intensity function improves the goodness-of-fit and allows researchers to test the existence of some unobserved features in the underlying process by using the data provided.

However, in addition to the randomness in realized jumps, the data to be analysed also

contain other sources of randomness. For example, for data with timestamps of one second or one millisecond precision, any durations shorter than the required threshold may not be observable; we may observe a trade with a missing price; we may observe a bid but without a corresponding ask (or vice-versa); and we may observe several events at the same time, etc. Hence, data collection mechanisms can introduce some extra random errors into the data to be analysed.

Some data cleaning techniques could be used to deal with some of the missing information. For example, we may take the missing value as the last observed value; we may delete a trade with missing price; we may also transform the data by adding some appropriate random noise to some event times, thus ensuring that no multiple events at the same time, see, e.g. Lorenzen (2012), Bowsher (2007) and Filimonov and Sornette (2012). It is seen that data cleaning techniques can also introduce some extra random errors into the data. We will refer to all of these errors caused by data collection mechanisms and/or data-cleaning procedures as extra random errors (ERE). Clearly, if such data are used, these ERE may seriously affect the estimation results of a financial model used. In this paper, we focus on HP models.

The HP models discussed above allow us to deal with the randomness in realized jumps of an underlying process but ignore the effect of ERE when analyzing the data. It is worth noting that it is impossible to remove all the ERE from the data to be analysed. Hence, it is important to assess the effect of ERE on the analysis, but currently, it is unclear how to evaluate the effect of ERE and how to deal with their effect and the randomness in realized jumps of an underlying financial process simultaneously. This literature gap motivated the work presented in this paper.

Since it is not possible to remove all the ERE from the data to be analysed, we develop a novel HP model with hidden marks in this paper, where the hidden marks are used to take account the effect of ERE. As the deviations from a pure Hawkes model caused by ERE can be very different, we further propose to use the generalized Lambda distribution (GLD) for the mark distribution due to its high degree of flexibility. As it turns out, existing HP

estimation methods (see, for example, Veen and Schoenberg, 2008, Rasmussen, 2013, and the references therein) are hard to apply to the novel structure of the proposed model, we further propose a Bayesian method for parameter estimation.

We focus our discussions on a simple self-exciting HP model because many other HP models in the literature are based on this type of models. We conduct extensive simulation studies to check the performance of our estimation method and evaluate the effect of ERE. More specifically, we consider six different scenarios that mimic real situations and in each scenario, we introduce some ERE into the data to be analysed. Our results confirm that when ERE cause information about the underlying process to be lost, the intensity function may be underestimated. Our results also show that the proposed estimation method works well and the proposed model performs better in the presence of ERE compared with the standard HP model. We further apply our method to some high frequency data to illustrate the use of the proposed model in practice.

In summary, the main contributions of this paper to the finance literature are given below. Firstly, we develop a novel HP model by using hidden marks to take account the effect of ERE, which provides a more flexible model for financial analysis and results in more reliable and meaningful statistical inferences about the underlying financial process. Secondly, we propose a Bayesian method for the estimation of the proposed model and show that the method works well. Hence, the method provides a useful tool for people in finance to use when they need.

In the following, we introduce the proposed model in Section 2 and discuss a Markov chain Monte Carlo (MCMC) estimation method in Section 3. Section 4 presents two simulation studies. The first one is used to check the performance of our estimation method, and the second one is used to assess the impact of the ERE. Section 5 considers a generalization of the proposed model briefly. Section 6 presents the results obtained by applying our method to some high frequency data. Finally, some comments and conclusions are given in Section 7.

2 Proposed Model

2.1 The self-exciting HP model

We start from the self-exciting HP model proposed by Hawkes (1971a, 1971b). Let $\{N(t)\}_{t\geq 0}$ be a counting process, defined by

$$P(N(t+h)-N(t)=m|N(s)(s<t))=\begin{cases}\lambda(t)h+o(h), & \text{if } m=1\,,\\ o(h), & \text{if } m>1\,,\\ 1-\lambda(t)h+o(h), & \text{if } m=0\,,\end{cases} \tag{1}$$

where $\lambda(t)$ is the conditional intensity function of the point process, i.e. conditional on the history of the process up to time t. Moreover, $h>0$ and $o(h)$ is a function so that $\lim_{h\downarrow 0}\frac{o(h)}{h}=0$. In other words, the probability of observing an event during the infinitesimal interval of time t and $t+h$ when $h\downarrow 0$ is $\lambda(t)h$, and the probability of observing more than one event during the same interval is negligible.

The self-exciting HP model is a point process with $\lambda(t)=v(t|\boldsymbol{\alpha}_0)+\int_0^t\gamma(t-s|\boldsymbol{\alpha}_1)dN(s)$, where $t\geq 0$, $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_1$ are vectors of model parameters, $v(t|\boldsymbol{\alpha}_0)>0$ is a deterministic function of t, representing the baseline intensity. $v(t|\boldsymbol{\alpha}_0)$ describes the arrival of events triggered by external sources, and their arrival does not depend on the previous events within the process. $\gamma(u|\boldsymbol{\alpha}_1)\geq 0$, called the memory kernel of the point process, is a non-negative function of u for $u\geq 0$ and satisfies $0<\int_0^\infty\gamma(u|\boldsymbol{\alpha}_1)du<1$.

If we let $v(t|\boldsymbol{\alpha}_0)=\alpha_0>0$ and let $t_0=0<t_1<\ldots<t_n=T$ be the observed event times between 0 and T, then the conditional intensity function can be expressed by

$$\lambda(t)=\alpha_0+\sum_{t_i<t}\gamma(t-t_i|\boldsymbol{\alpha}_1). \tag{2}$$

A special case of (2) is obtained by letting $\gamma(u|\boldsymbol{\alpha}_1)=\alpha_1 e^{-\alpha_2 u}$ in (2). It follows from $0<\int_0^\infty\gamma(u|\boldsymbol{\alpha}_1)du<1$ that $0<\alpha_1<\alpha_2<\infty$. Hence the intensity function of the

self-exciting HP model becomes

$$\lambda(t) = \alpha_0 + \sum_{t_i<t} \alpha_1 e^{-\alpha_2(t-t_i)}. \tag{3}$$

Model (3) shows that all the events having occurred before current time t contribute to the event intensity at time t. Moreover, the kernel $\gamma(t - t_i|\boldsymbol{\alpha}_1) = \alpha_1 e^{-\alpha_2(t-t_i)}$ modulates the change that an event at time t_i has on the intensity function at time t, showing that more recent events have higher influence on the current event intensity, compared to events having occurred further away in time. In this paper we focus our discussions on (2) with $\gamma(u|\boldsymbol{\alpha}_1) = \alpha_1 e^{-\alpha_2 u}$, and we refer to the HP model defined by (3) as Model A throughout the paper.

Given the observed event times, Ogata (1981) derived an expression for the likelihood function of the parameters, in which some components can be calculated recursively. A more general result can be found in, e.g. Daley and Vere-Jones (2003, Proposition 7.2III). More specifically, the probability density function of time t after time a is given by $f(t) = \lambda(t)e^{-\int_a^t \lambda(s)ds}$. Hence, the likelihood function of $\alpha_0, \alpha_1, \alpha_2$ is given by

$$L(\alpha_0, \alpha_1, \alpha_2|\mathbf{t}_n) = \prod_{j=1}^n f(t_j|\mathbf{t}_{j-1}) = \prod_{j=1}^n \lambda(t_j) e^{-\int_0^T \lambda(t)dt},$$

where $\mathbf{t}_j = (t_j, \ldots, t_0)$ for $j = 1, \ldots, n$. (Note that we use bold letters for vectors in this paper.) By taking logs, we have $\ln L = \sum_{j=1}^n \ln \lambda(t_j) - \int_0^T \lambda(t)dt$, where

$$\int_0^T \lambda(t)dt = \int_0^T \left\{\alpha_0 + \int_0^t \alpha_1 e^{-\alpha_2(t-s)} dN(s)\right\} dt = \alpha_0 T - \tfrac{\alpha_1}{\alpha_2} \textstyle\sum_{t_i<T} \left\{e^{-\alpha_2(T-t_i)} - 1\right\}.$$

Therefore, the estimates of $\alpha_0, \alpha_1, \alpha_2$ may be obtained by using a numerical procedure to maximize the log-likelihood function.

2.2 The HP model with hidden marks

It is worth noting that due to the effect of ERE, the event intensity at time t should be a random variable, while the event intensity at time t defined by model A is not. Since it is unclear how the HP model defined by (2) explains the role of ERE in the data to be analysed, we provide all points t_i with i.i.d. marks ξ_i such that $\xi_i + \alpha_0 > 0$ a.s. Furthermore, we let the conditional intensity depend on the marks only via the mark of the last observed point, which gives the following model:

$$\lambda(t \mid \boldsymbol{\alpha}) = \alpha_0 + \sum_{t_i<t} \gamma(t - t_i|\boldsymbol{\alpha}_1) + \xi_{t_{i_0}}\,, \quad t \geq 0, \quad \text{where } i_0 := \max\{i : t_i < t\}, \tag{4}$$

where $\boldsymbol{\alpha}$ is a vector containing all parameters of the model. Model (4) says that the event intensity $\lambda(t|\boldsymbol{\alpha})$ at time t is a random variable, whose distribution depends both on the observed history of the process and on the distribution of the mark $\xi_{t_{i_0}}$. Note that due to the nature of the ERE that we consider, it is reasonable to assume that ξ_i are i.i.d. random variables.

So, we need to specify a distribution for $\xi_{t_{i_0}}$ so that $\lambda(t \mid \boldsymbol{\alpha})$ defined by (4) is positive for any $t \geq 0$ in a parameter space of the model. We also hope that the distribution of $\xi_{t_{i_0}}$ is flexible enough to capture different characteristics of the distribution, including the location, scale, skewness and tail shape etc., which will make model (4) more robust to model specification errors.

We first consider which distribution could be a good candidate for $\xi_{t_{i_0}}$. Many standard distributions have been used in financial modelling, including normal, lognormal, Weibull, exponential, t-, skewed t- and F- distributions, as well as many others. However, as the deviations (caused by ERE) from a pure Hawkes model can be very different, we need a mark distribution that has a high degree of flexibility. The work of Fournier et al. (2007) suggests that the generalized lambda distribution (GLD) could be a better candidate for $\xi_{t_{i_0}}$ because it defines the most flexible class of distributions and it is also able to provide a very accurate approximation to many commonly used distributions, see Figure 1 for some

examples. Hence, we use the GLD for $\xi_{t_{i_0}}$ in this paper.

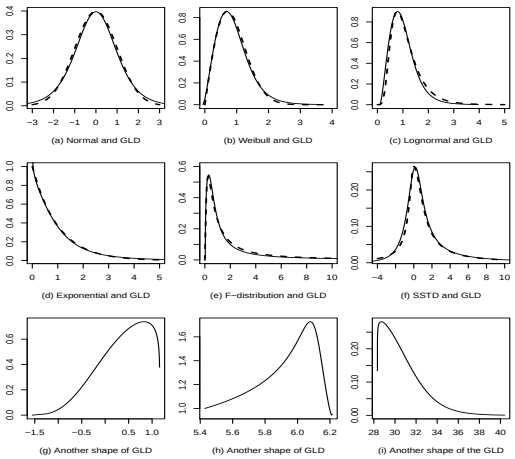

Figure 1: Density function plots of GLD (continuous curves) and Normal, Weibull, lognormal, exponential, F-, skewed t- distributions (dashed curves).

Note that GLD has already been used in the finance literature. For example, it has been used to study option pricing (Corrado, 2001), spot exchange rates (Lee, 2003) and income data (Tarsitano, 2004). More recently, Corlu and Corlu (2015) studied the performance of GLD in capturing the leptokurtic and skewed behaviour of exchange rate returns.

It is also worth noting that GLD defines a class of distributions. The distributional range of GLD depends on its parameters (see e.g. Gilchrist, 2000, Chapter 7). As we need to ensure that the intensity function is positive for any t in a parameter space of the model, in this paper we use the GLD defined by the quantile function $Q(\tau) = \mu + \eta_0 Q(\tau, \eta_1, \eta_2)$, where τ is uniformly distributed between 0 and 1, μ and $\eta_0 > 0$ are the location and scale of the distribution respectively, and $Q(\tau, \eta_1, \eta_2)$ is given by

$$Q(\tau, \eta_1, \eta_2) = \frac{\tau^{\eta_1} - 1}{\eta_1} - \frac{(1-\tau)^{\eta_2} - 1}{\eta_2}, \;\; 0 < \eta_1 < 1, \; \eta_2 < 0, \tag{5}$$

in which η_1 and η_2 not only determine the skewness of the distribution but also determine the relative weights of the tails. In other words, the skewness of the distribution is modelled as a result of tail shape and not as an independence feature (see Gilchrist, 2000). It is worth noting that $Q(\tau, \eta_1, \eta_2)$ defined by (5) is a special case of GLD with location 0 and scale 1,

for which we have the following result.

Proposition 1 *Consider the GLD defined by (5). Then for any* $\tau \in (0,1)$*, we have* $-1/\eta_1 < Q(\tau,\eta_1,\eta_2) < \infty$*, where* $0 < \eta_1 < 1,\ \eta_2 < 0$.

See the Proof of Proposition 1 in Appendix I. Proposition 1 says that the support of the GLD defined by (5) is given by $(-1/\eta_1, \infty)$, which will be useful when we define our new model.

Since the GLD is defined by its quantile function because its density function and other equivalent functions do not have an explicit mathematical expression, we need to rewrite model (4) in a quantile function form so that the GLD defined by (5) could be used.

It is worth noting that if we let ξ_i follow the distribution defined by $\eta_0 Q(\tau,\eta_1,\eta_2)$, where $\tau \in (0,1)$, $\eta_0 > 0$ and $Q(\tau,\eta_1,\eta_2)$ is defined by (5), then we have $\xi_i = \eta_0 Q(\tau_i,\eta_1,\eta_2)$, where τ_i are i.i.d. random variables uniformly distributed between 0 and 1. Using this relation, model (4) can be expressed by

$$\lambda(t \mid \boldsymbol{\alpha}) = \alpha_0 + \sum_{t_i<t} \gamma(t-t_i|\boldsymbol{\alpha}_1) + \eta_0 Q(\tau_{t_{i_0}},\eta_1,\eta_2)\,, \quad t \geq 0, \quad \text{where } i_0 := \max\{i : t_i < t\},$$

according to which, we have the following proposition.

Proposition 2 *Let* $\xi_{t_{i_0}}$ *follow the distribution defined by* $\eta_0 Q(\tau,\eta_1,\eta_2)$*, where* $\tau \in (0,1)$*,* $\eta_0 > 0$ *and* $Q(\tau,\eta_1,\eta_2)$ *is defined by (5). Then the quantile function of the event intensity at time* t*, denoted by* $Q_\lambda(t,\tau|\boldsymbol{\alpha})$*, is given by*

$$\begin{aligned} Q_\lambda(t,\tau|\boldsymbol{\alpha}) &= \alpha_0 + \textstyle\sum_{t_i<t} \gamma(t-t_i|\boldsymbol{\alpha}_1) + \eta_0 Q(\tau,\eta_1,\eta_2) \\ &= \alpha_0 + \textstyle\sum_{t_i<t} \alpha_1 e^{-\alpha_2(t-t_i)} + \eta_0 Q(\tau,\eta_1,\eta_2), \end{aligned} \tag{6}$$

where $\boldsymbol{\alpha} = (\alpha_0, \boldsymbol{\alpha}_1, \boldsymbol{\eta})$*, in which* $\boldsymbol{\alpha}_1 = (\alpha_0,\alpha_1,\alpha_2)$ *and* $\boldsymbol{\eta} = (\eta_0,\eta_1,\eta_2)$.

See the Proof of Proposition 2 in Appendix I. It is worth noting that as ξ_i are i.i.d. random variables, we have dropped t_{i_0} from (6). Hence, the distribution of the event intensity at

time t is the GLD, whose location and scale are given by $\alpha_0 + \sum_{t_i<t} \alpha_1 e^{-\alpha_2(t-t_i)}$ and η_0 respectively, and whose tails are controlled by η_1 and η_2. We refer to the HP model defined by (6) as Model B throughout the paper.

Proposition 3 *Let* $\bar{\Omega}_1 = \{\boldsymbol{\alpha} | 0 < \alpha_1 < \alpha_2 < \infty,\ \alpha_0 > \eta_0/\eta_1,\ \eta_0 > 0,\ 0 < \eta_1 < 1,\ \eta_2 < 0\}$. *Then* $Q_\lambda(t, \tau|\boldsymbol{\alpha}) > 0$ *on* $\bar{\Omega}$ *for all* $\tau \in (0, 1)$ *and* $t \geq 0$.

See the Proof of Proposition 3 in Appendix I. Proposition 3 ensures that $Q_\lambda(t, \tau|\boldsymbol{\alpha})$ is positive on $\bar{\Omega}$ for any t and τ. That is, the intensity $\lambda(t|\boldsymbol{\alpha}) > 0$ for all t on the parameter space. Hence, the proposed Model B is well-defined on $\bar{\Omega}_1$.

Model B also tells us that the magnitude of the intensity jump right after an event occurrence is given by α_1, the rate at which the intensity decreases exponentially is given by α_2, while the baseline intensity of the process is a random variable given by $\alpha_0 + \eta_0 Q(\tau, \eta_1, \eta_2)$, which is positive on $\bar{\Omega}_1$. Therefore, the parameters of Model A and Model B are interpreted similarly, but Model B is more flexible than Model A because it involves hidden marks.

Since Model B defines the entire conditional distribution of event intensity at time t, it is easy to obtain statistical inferences about any features of event intensity at time t, see Proposition 4 given below for some examples.

Proposition 4 *Consider Model B.*

(i) The conditional median of the intensity function is given by $Q_\lambda(t, 0.5|\boldsymbol{\alpha})$ *for* $t \geq 0$.

(ii) The conditional expectation of the intensity function is given by

$$\mu_\lambda(t) = \alpha_0 + \sum_{t_i<t} \alpha_1 e^{-\alpha_2(t-t_i)} + \eta_0 \left(\frac{1}{\eta_2 + 1} - \frac{1}{\eta_1 + 1} \right), \quad t \geq 0.$$

(iii) The conditional variance of the intensity function is given by

$$\begin{aligned}\sigma_\lambda^2 &= \frac{\eta_0^2}{\eta_1^2}\left(\frac{1}{2\eta_1+1}-\frac{1}{(\eta_1+1)^2}\right)+\frac{\eta_0^2}{\eta_2^2}\left(\frac{1}{2\eta_2+1}-\frac{1}{(\eta_2+1)^2}\right)\\ &\quad -\frac{2\eta_0^2}{\eta_1\eta_2}\left(B(\eta_1+1,\eta_2+1)-\frac{1}{(\eta_1+1)(\eta_2+1)}\right),\quad t\geq 0,\end{aligned}$$

where $B(\eta_1+1,\eta_2+1)=\int_0^1 \tau^{\eta_1}(1-\tau)^{\eta_2}d\tau$*, which is in fact the Beta function.*

(iv) The 100$(2\tau-1)$*% probability interval for* $\lambda(t|\boldsymbol{\alpha})$ *is given by* $[Q_\lambda(t,1-\tau'|\boldsymbol{\alpha}),\ Q_\lambda(t,\tau'|\boldsymbol{\alpha})]$*, where* $t\geq 0$*,* $\tau'=\tau$ *if* $\tau>0.5$ *and* $\tau'=1-\tau$ *otherwise.*

See the Proof of Proposition 4 in Appendix I. In practice, we may use the conditional median or conditional expectation to approximate event intensity at time t. In this paper we use the conditional median rather than the conditional expectation because median is a more robust measure for the central location of a distribution. Moreover, a probability interval for event intensity at time t also gives us an interval forecast for the intensity. For example, a 95% interval forecast for event intensity at time t is given by $[Q_\lambda(t,0.025|\boldsymbol{\alpha}),\ Q_\lambda(t,0.975|\boldsymbol{\alpha})]$, which can be very useful in practice. In finance, we are also interested in probability forecasting for future events. Proposition 5 shows how these forecasts can be obtained.

Proposition 5 *Consider Model B. Let* N *be the number of events occurring in the interval* $(T,T+h]$*, where* $h>0$*. Then, given observed history up to time* T*, we have*

(i) $P(N\geq 1)=1-e^{\Lambda}$ *can be approximated by* $P_\tau(N\geq 1)|_{\tau=0.5}$*, where* $P_\tau(N\geq 1)=1-e^{-\Lambda_\tau}$*,* $\Lambda=\int_T^{T+h}\lambda(t|\boldsymbol{\alpha})dt$ *and*

$$\begin{aligned}\Lambda_\tau &= \int_T^{T+h} Q_\lambda(t,\tau|\boldsymbol{\alpha})dt\\ &= \{\alpha_0+\eta_0 Q(\tau,\eta_1,\eta_2)\}h+\textstyle\sum_{t_i<T+h}\frac{\alpha_1}{\alpha_2}\left\{e^{-\alpha_2(T-t_i)}-e^{-\alpha_2(T+h-t_i)}\right\}.\end{aligned}$$

(ii) The 100$(2\tau-1)$*% probability interval for* $P(N\geq 1)$ *is given by* $[P_{1-\tau'}(N\geq 1),\ P_{\tau'}(N\geq 1)]$*, where* $\tau'=\tau$ *if* $\tau>0.5$ *and* $\tau'=1-\tau$ *otherwise.*

See the Proof of Proposition 5 in Appendix I. For example, if $\tau = 0.975$, then a 95% interval forecast for $P(N \geq 1)$ is given by $[P_{0.025}(N \geq 1),\ P_{0.975}(N \geq 1)]$. It is worth noting that the Proof of Proposition 5 in Appendix I also shows that Λ_τ is the quantile function of Λ and $P_\tau(N \geq 1)$ is the quantile function of $P(N \geq 1)$. This means that any quantity of interest about $P(N \geq 1)$ and Λ can also be obtained easily.

Note that $P_\tau(N = k) = (\Lambda_\tau^k/k!)exp(-\Lambda_\tau)$ and $P(N = k) = (\Lambda^k/k!)exp(-\Lambda)$, where $k = 1, 2, ...$ However, $P_\tau(N = k)$ is not a monotone function of Λ_τ, which suggests that the τth quantile of $P(N = k)$ is not necessarily given by $P_\tau(N = k)$. Hence, $P_{0.5}(N = k)$ may not be a good approximation of $P(N = k)$. To deal with this problem, we propose the following method to estimate $P(N = k)$.

Let τ_j $(j = 1, \ldots, J)$ be an i.i.d. random sample of size J between 0 and 1. Then $\{P_{\tau_j}(N = k), j = 1, \ldots, J\}$ is a random sample of $P(N = k)$. Hence $P(N = k)$ can be estimated by the median of this sample. Moreover, the top and bottom e.g. 5% quantiles of this sample give a 90% interval forecast for $P(N = k)$.

It is seen that by using our modelling approach, we can analyze financial event data from a distribution perspective, which shows another advantage of the proposed model.

3 Parameter Estimation

3.1 The posterior density function

To use the proposed model, we need to estimate the model parameters. First note that Model B says that the conditional quantile function of event intensity at time t_j is given by $Q_\lambda(t_j, \tau|\boldsymbol{\alpha})$. Hence, a random sample of the intensity at time t_j is given by $Q_\lambda(t_j, \tau_j|\boldsymbol{\alpha})$, where τ_j is a random sample from $(0, 1)$. Therefore, the likelihood of an event at time t_j after time t_{j-1} is given by $f(t_j|\tau_j, t_{j-1}) = Q_\lambda(t_j, \tau_j|\boldsymbol{\alpha})\ \exp\left\{-\int_{t_{j-1}}^{t_j} Q_\lambda(s, \tau_j|\boldsymbol{\alpha})ds\right\}$, in which τ_j is a latent variable. As the latent variables make it difficult to use the MLE

method, we develop a Bayesian method for parameter estimation, for which we need to derive the posterior density function of the parameters. Note that a posterior density function is proportional to the product of a likelihood function and a prior density function of the parameters.

Let $\boldsymbol{\tau}_j = (\tau_j, \tau_{j-1}, \ldots, \tau_0)$ and $\mathbf{t}_j = (t_j, t_{j-1}, \ldots, t_0)$, where $j = 0, \ldots, n$. Moreover, let the posterior density function of $(\boldsymbol{\tau}_n, \boldsymbol{\alpha})$ be $\pi(\boldsymbol{\tau}_n, \boldsymbol{\alpha}|\mathbf{t}_n)$, let the likelihood function be $\pi(\mathbf{t}_n|\boldsymbol{\tau}_n, \boldsymbol{\alpha})$, and let the prior density function be $\pi(\boldsymbol{\tau}_n, \boldsymbol{\alpha})$. Then the posterior density function of the parameters can be expressed by $\pi(\boldsymbol{\tau}_n, \boldsymbol{\alpha}|\mathbf{t}_n) \propto \pi(\mathbf{t}_n|\boldsymbol{\tau}_n, \boldsymbol{\alpha})\pi(\boldsymbol{\tau}_n, \boldsymbol{\alpha})$. However, as the likelihood function $\pi(\mathbf{t}_n|\boldsymbol{\tau}_n, \boldsymbol{\alpha})$ is difficult to evaluate due to the latent variables, we derive another expression for the posterior density function. It is worth noting that $\pi(\boldsymbol{\tau}_n, \boldsymbol{\alpha}|\mathbf{t}_n) \propto \pi(\mathbf{t}_n, \boldsymbol{\tau}_n|\boldsymbol{\alpha})\pi(\boldsymbol{\alpha})$, where $\pi(\boldsymbol{\alpha})$ is the prior density function of $\boldsymbol{\alpha}$, and $\pi(\mathbf{t}_n, \boldsymbol{\tau}_n|\boldsymbol{\alpha}) = \prod_{j=1}^{n} \pi(t_j|\tau_j, \mathbf{t}_{j-1}, \boldsymbol{\tau}_{j-1}, \boldsymbol{\alpha})\pi(\tau_j|\mathbf{t}_{j-1}, \boldsymbol{\tau}_{j-1}, \boldsymbol{\alpha})$, in which $\pi(\tau_j|\mathbf{t}_{j-1}, \boldsymbol{\tau}_{j-1}, \boldsymbol{\alpha})$ $= 1$ as τ_js are i.i.d. U(0, 1). Hence for Model B, the posterior density function is given by

$$\begin{aligned}\pi(\boldsymbol{\tau}_n, \boldsymbol{\alpha}|\mathbf{t}_n) \propto & \prod_{j=1}^{n} \left\{\alpha_0 + \sum_{t_i<t_j} \alpha_1 e^{-\alpha_2(t_j-t_i)} + \eta_0 \left(\frac{\tau_j^{\eta_1}-1}{\eta_1} - \frac{(1-\tau_j)^{\eta_2}-1}{\eta_2}\right)\right\} \\ & \times \exp\left[-\int_0^T \left\{\alpha_0 + \sum_{t_i<s} \alpha_1 e^{-\alpha_2(s-t_i)}\right\} ds \right. \\ & \left. - \sum_{j=1}^{n} \eta_0 \left(\frac{\tau_j^{\eta_1}-1}{\eta_1} - \frac{(1-\tau_j)^{\eta_2}-1}{\eta_2}\right)(t_j - t_{j-1})\right] \pi(\boldsymbol{\alpha}),\end{aligned} \tag{7}$$

where $(\boldsymbol{\alpha}, \boldsymbol{\tau}_n) \in \widetilde{\Omega} = \bar{\Omega}_1 \times \bar{\Omega}_2$ and $\bar{\Omega}_2 = (0, 1)^n$.

Clearly, the posterior distribution of the parameters is not a standard one. Hence, we will use a MCMC method for parameter estimation, which requires that the posterior distribution is well-defined in a parameter space Ω in the sense that $\int_\Omega \pi(\boldsymbol{\tau}_n, \boldsymbol{\alpha}|\mathbf{t}_n) d\boldsymbol{\tau}_n d\boldsymbol{\alpha} < \infty$.

A simple way to define Ω is to slightly modify $\widetilde{\Omega}$. Specifically, we let the modified parameter space be $\Omega = \Omega_1 \times \Omega_2$, where $\Omega_1 = \{\boldsymbol{\alpha}|0 < \alpha_1 < \alpha_2 \leq M,\, \eta_0/\eta_1 < \alpha_0 \leq M,\, 0 < \eta_0 \leq M,\, 0 < \eta_1 < 1,\, -M \leq \eta_2 \leq -\epsilon\}$, $\Omega_2 = (0,\, 1-\epsilon]^n$. In this paper, we let $\epsilon = 10^{-30}$ and $M = 10^{30}$, which ensures that the difference between $\widetilde{\Omega}$ and Ω can be safely ignored from a practical perspective.

Proposition 6 *If $\pi(\boldsymbol{\alpha})$ is a well-defined probability density function of $\boldsymbol{\alpha}$ on Ω_1, then the*

posterior density function $\pi(\boldsymbol{\tau}_n, \boldsymbol{\alpha}|\mathbf{t}_n)$ defined by (7) is well-defined on Ω.

See the Proof of Proposition 6 in Appendix I. Hence, we now need to define a proper prior density function for $\pi(\boldsymbol{\alpha})$.

3.2 The prior density function

To make use of the prior knowledge about the parameter space and the dependence structure between the parameters, we let $\pi(\boldsymbol{\alpha}) = \pi(\alpha_1)\pi(\alpha_2|\alpha_1)\pi(\eta_0)\pi(\eta_1)\pi(\eta_2)\pi(\alpha_0|\eta_0, \eta_1)$, where α_1, η_0 and $-\eta_2$ follow a log-normal distribution, and all others follow a truncated log-normal distribution. The Appendix II gives a detailed expression of the prior density function $\pi(\boldsymbol{\alpha})$, from which we see that the scale parameters of the prior distributions for α_i and η_i are denoted by u_i and v_i respectively, where $i = 0, 1, 2$. Clearly, $\pi(\boldsymbol{\alpha})$ is well-defined on Ω_1. Note that these scale parameters measure the knowledge we have about individual parameters before analysing the data. As in reality we usually do not have any prior knowledge at hand, in this paper we take $u_i = v_i = 2$ to reflect this fact. This is because, e.g. $u_1 = 2$ implies that the standard deviation of α_1 is given by $\sqrt{(e^{u_1^2} - 1)e^{u_1^2}} \approx 54$, which is very large, suggesting that the posterior distribution contains almost no prior information on α_1. Similar explanations hold for other parameters. So our method is also robust to the specification of priors.

3.3 The MCMC estimation method

One of the MCMC methods is the Metropolis-Hastings method. The basic idea of this method is to generate a sequence of model parameters that form a Markov chain in the parameter space, such that its equilibrium distribution is the posterior distribution of the parameters. This can be achieved by simulating a candidate parameter value from a chosen distribution and accepting this proposed value as the next in the sequence with a known probability; see, for example, Brooks (1998) and Geyer (2011) for details.

In our case, we let $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}_n$ represent the current values of the parameters, and $\boldsymbol{\alpha}'$ and $\boldsymbol{\tau}'_n$ represent the proposed values. The main steps of our MCMC method are also given in Appendix II. The Markov chain theory (see, e.g. O'Hagan and Forster, 2004) guarantees that the equilibrium distribution of the Markov chain generated by the above Metropolis-Hastings algorithm is the posterior distribution defined by (7).

Therefore, after a burn-in period, posterior samples of the parameters can be collected from the Markov chain, and these samples can be regarded as parameter samples drawn from the posterior distribution. The Bayes estimate of the unknown parameter is simply the mean of the posterior distribution, denoted by $\hat{\boldsymbol{\alpha}}$. Lehmann and Casella (1998, Section 6.8) shows that the Bayes estimator defined above is asymptotically unbiased and is asymptotically efficient.

3.4 Model evaluation

Let $\lambda(t)$ be the intensity function of a point process, $\{t_i\}$ a set of event times and $a_i = \int_0^{t_i} \lambda(s)ds$. Then it is well-known that $\{a_i\}$ follows a stationary unit rate Poisson process. Hence, the durations defined by $b_i = a_i - a_{i-1} = \int_{t_{i-1}}^{t_i} \lambda(s)ds$ should be i.i.d. random variables following a unit rate exponential distribution. Moreover, $U_i = 1 - e^{-b_i}$ are i.i.d. uniform random variables on $[0,\, 1)$. See e.g. Daley and Vere-Jones (2003, Proposition 7.4.IV) for details.

Let $\hat{a}_i$, $\hat{b}_i$ and $\hat{U}_i$ be the values calculated using the estimated intensity function $\hat{\lambda}(t)$. Then the above results suggest that, if $\hat{\lambda}(t)$ fits data well, then $\{\hat{a}_i\}$ should behave like a unit rate Poisson process, $\{\hat{b}_i\}$ should follow a unit rate exponential distribution, and $\{\hat{U}_i\}$ should follow a uniform distribution on $[0,\, 1)$.

Therefore, the goodness-of-fit of a point process can be checked by, e.g. graphical methods. For example, points on the plot of $\hat{a}_i$ against the cumulative number of events should be roughly along the reference line $y = x$. Similarly, points on the plot of the quantiles of $\hat{b}_i$ against those of a unit rate exponential distribution should also be roughly along

the reference line. Significant departures of the points from this line suggest a weakness in the model.

The above diagnostic methods for the goodness-of-fit of a point process can also be used for our model. If our intensity function is estimated by $Q_\lambda(t, 0.5|\hat{\boldsymbol{\alpha}})$, then we have

$$\begin{aligned}\hat{a}_i &= \int_0^{t_i} Q_\lambda(s, 0.5|\hat{\boldsymbol{\alpha}})ds \\ &= \{\hat{\alpha}_0 + \hat{\eta}_0 Q(0.5, \hat{\eta}_1, \hat{\eta}_2)\}(t_i - t_{i-1}) - \sum_{k=0}^{i-1} \frac{\hat{\alpha}_1}{\hat{\alpha}_2} \left\{e^{-\hat{\alpha}_2(t_i - t_k)} - e^{-\hat{\alpha}_2(t_{i-1} - t_k)}\right\}.\end{aligned}$$

Hence $\hat{b}_i = \hat{a}_i - \hat{a}_{i-1}$ and $\hat{U}_i = 1 - e^{-\hat{b}_i}$ can be calculated easily.

4 Simulation Studies

In this section we present results of two simulation studies: The first one assesses our estimation method, and the second one evaluates the effect of ERE on Models A and B respectively. Our results show that the estimation method works well, and they also confirm that when ERE cause information about the underlying process to be lost, the intensity function may be underestimated. Moreover, they show that Model B performs better in the presence of ERE compared with Model A. These findings are further supported by our first application in Section 6.

4.1 Simulation study 1: Assess the performance of the estimation method

We check the performance of the proposed estimation method from the following several aspects: the estimated parameters, the goodness-of-fit of the estimated models, and the differences between the estimated and the true probabilities $P_\tau(N \geq 1)$ for a range of $\tau \in (0, 1)$.

4.1.1 Simulated data

We consider the following two models

$$Q_\lambda(t,\tau|\boldsymbol{\alpha}) = 0.75 + \sum_{t_i<t} 0.6e^{-1.8(t-t_i)} + 0.2\left\{\frac{\tau^{0.8}-1}{0.8} - \frac{(1-\tau)^{-0.8}-1}{-0.8}\right\} \tag{8}$$

and

$$\begin{aligned} &Q_\lambda(t,\tau|\boldsymbol{\alpha}) \\ &= \alpha_0 + \textstyle\sum_{t_i<t} \alpha_1 e^{-\gamma_1(t-t_i)} + \sum_{t_i<t} \alpha_2 e^{-\gamma_2(t-t_i)} + \eta_0\left\{\frac{\tau^{\eta_1}-1}{\eta_1} - \frac{(1-\tau)^{\eta_2}-1}{\eta_2}\right\} \\ &= 0.533 + \textstyle\sum_{t_i<t} 2e^{-5(t-t_i)} + \sum_{t_i<t} 0.5e^{-4.33(t-t_i)} + 0.2\left\{\frac{\tau^{0.6}-1}{0.6} - \frac{(1-\tau)^{-0.5}-1}{-0.5}\right\}, \end{aligned} \tag{9}$$

where $t \geq 0$, $\tau \in (0,1)$ and the parameter values were arbitrarily chosen from the parameter space. It is seen that model (9) is a minor generalization of Model B as it contains an extra term. In fact, a further generalization in this direction is to include k terms into the model.

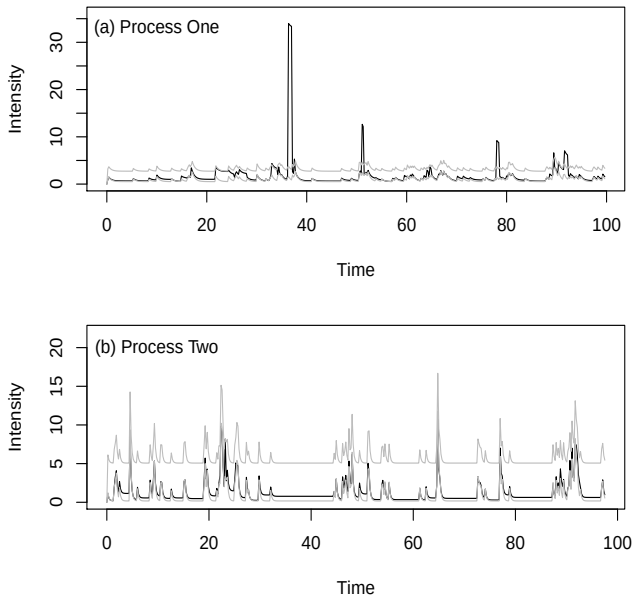

Figure 2: Observed intensity functions (darker curves) and the 90% interval forecasts (grey curves) for the true intensity function $\lambda(t|\boldsymbol{\alpha})$ of Processes One and Two respectively.

Table 1: 95% credible intervals for model parameters in 200 simulation studies

Model (8)	α_0	α_1	α_2	η_0	η_1	η_2
True values	0.75	0.60	1.80	0.2	0.8	-0.8
2.5%	0.407	0.304	0.972	0.009	0.255	-4.891
97.5%	1.09	1.96	11.80	0.586	0.981	-0.016

Model (9)	α_0	α_1	α_2	γ_1	γ_2	η_0	η_1	η_2
True values	0.533	2.000	0.500	5.00	4.333	0.200	0.600	-0.500
2.5% quantile	0.284	0.143	0.158	1.28	1.420	0.058	0.252	-2.483
97.5% quantile	0.661	3.325	3.429	12.23	18.38	0.405	0.981	-0.116

We simulated 200 independent point processes from models (8) and (9) respectively on the time interval between 0 and 100 with $\tau = 0.5$. Hence, in this simulation study $T = 100$. For illustration purposes, the first simulated process from model (8) and model (9) is selected, which is called Process One and Process Two, respectively. The black curves in Figure 2 show the intensity function plots for Processes One and Two respectively. It is worth noting that, when τ is fixed, the simulated data from these models can be obtained by using the method of Ogata (1981) easily.

4.1.2 Results about the estimated models

We applied our method to each simulated dataset, where the initial parameter values required by the MCMC method were obtained randomly from the parameter space. The prior information about the parameters used in the estimation is very weak as discussed in Section 3.2. Our results show that the convergence of the estimation method does not depend on the initial values. Based on the posterior samples collected, we constructed a 95% credible interval for each parameter of the model, where the lower and upper limits of the 95% credible interval are the lower and upper 2.5% quantiles of the posterior samples of the parameter, respectively. By repeating the above calculations for each simulated dataset, we obtained 200 credible intervals for each parameter, the average of which can be found in Table 1. It is seen that all the true parameter values are well within the respective credible intervals, suggesting that the performance of the estimation method is satisfactory.

It follows from Proposition 5 that a 90% probability interval for $\lambda(t|\boldsymbol{\alpha})$ can be estimated by $[Q_\lambda(t, 0.05|\hat{\boldsymbol{\alpha}}),\ Q_\lambda(t, 0.95|\hat{\boldsymbol{\alpha}})]$. The grey curves in Figure 2 show the 90% interval prediction of the true intensity function of Processes One and Two, respectively. So about 10% of the intensities are expected to be outside this interval. Our results show that, for Process One, 13.18% of the true intensities are outside this interval, and for Process Two, it is 8.05%, suggesting that the 90% intervals forecasts are also reasonably good.

We checked the goodness-of-fit of the estimated models using the method discussed in Section 3.4 and obtained satisfactory results. Figure 3 illustrates the results using Processes One and Two respectively, where the straight line is the reference line. Hence, we do not have any major concerns about the estimated models.

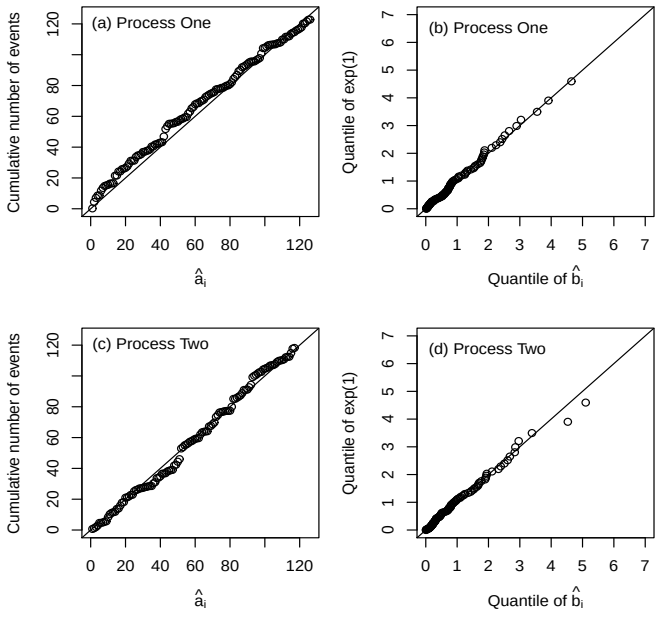

Figure 3: (a)(c): Plots of $\hat{a}_i$ versus the cumulative number of events, (b)(d): plots of the quantiles of the durations $\hat{b}_i$ against the corresponding quantiles of a unit rate exponential distribution, for Processes One and Two, respectively.

4.1.3 Results about the probability forecasts for $P_\tau(N \geq 1)$

By using Proposition 5, we calculated the true and estimated conditional probabilities of at least one event occurring on $(T, T + h]$, denoted by $P_\tau(N \geq 1)$ and $\hat{P}_\tau(N \geq 1)$ re-

spectively, where $h > 0$ and $T = 100$. More specifically, $P_\tau(N \geq 1)$ was calculated by using Proposition 5 with the true parameter values, while $\hat{P}_\tau(N \geq 1)$ is the median of $\{\hat{P}_\tau^{(\ell)}(N \geq 1), \ell = 1, \ldots, L\}$, in which L is the number of the posterior samples collected from the MCMC method and $\hat{P}_\tau^{(\ell)}(N \geq 1)$ was calculated by using Proposition 5 with the ℓth posterior sample of the parameters. We further let L-$\hat{P}_\tau(N \geq 1)$ and U-$\hat{P}_\tau(N \geq 1)$ be the lower and upper 5% quantiles of $\{\hat{P}_\tau^{(\ell)}(N \geq 1), \ell = 1, \ldots, L\}$ respectively.

For illustration purposes, we let $h = 1, 2, 3, 4, 5$ and $\tau = 0.01, 0.05, 0.25, 0.5, 0.75, 0.95$ and 0.99. We use the RMSE to measure the difference between$P_\tau(N \geq 1)$ and $\hat{P}_\tau(N \geq 1)$. A good performance of the method is expected if the RMSE values are small. Table 2 gives the summarized results obtained for Process One (similar results for Process Two are not shown to save space).

Table 2 shows that all RMSE values are small, suggesting that the estimated values $\hat{P}_\tau(N \geq 1)$ are very close to the theoretical ones $P_\tau(N \geq 1)$. Moreover, a 90% interval forecast for $P_\tau(N \geq 1)$ is given by $[\,\text{L-}\hat{P}_\tau(N \geq 1),\ \text{U-}\hat{P}_\tau(N \geq 1)\,]$. For example, if $h = 1$, then it is given by $[0.243,\ 0.645]$. Clearly, all the true values of $P_\tau(N \geq 1)$ are well within their respective 90% interval forecasts. It is worth emphasising that $P_\tau(N \geq 1)$ defines the quantile function of $P(N \geq 1)$. Therefore, any feature of the distribution of $P(N \geq 1)$ can be easily obtained. In summary, this simulation study shows that the estimation method performed satisfactorily.

4.2 Simulation study 2: Assess the effect of ERE

We want to evaluate the effect of ERE on models A and B through the estimated intensity function. To this end, we first obtained simulated data from Model A by using the method of Ogata (1981). We call the data obtained at this step the raw data. Then we created several scenarios that mimic some actual situations, and introduced some ERE into the raw data under each scenario. We call the data obtained at this step the processed data. We further fitted Models A and B to the processed data respectively. Finally, we compared the

Table 2: Probability forecasts for $P_\tau(N \geq 1)$ using process A

h	τ	0.010	0.050	0.250	0.500	0.750	0.950	0.990	RMSE
1	$P_\tau(N \geq 1)$	0.473	0.486	0.541	0.617	0.738	0.966	1.000	
	$\hat{P}_\tau(N \geq 1)$	0.487	0.504	0.550	0.622	0.725	0.889	0.978	0.032
	L-$\hat{P}_\tau(N \geq 1)$	0.186	0.243	0.408	0.544	0.586	0.634	0.666	
	U-$\hat{P}_\tau(N \geq 1)$	0.641	0.645	0.658	0.699	0.979	1.000	1.000	
2	$P_\tau(N \geq 1)$	0.690	0.705	0.765	0.837	0.923	0.999	1.000	
	$\hat{P}_\tau(N \geq 1)$	0.701	0.719	0.770	0.838	0.914	0.986	0.999	0.009
	L-$\hat{P}_\tau(N \geq 1)$	0.273	0.371	0.614	0.765	0.809	0.845	0.875	
	U-$\hat{P}_\tau(N \geq 1)$	0.849	0.853	0.867	0.899	0.999	1.000	1.000	
3	$P_\tau(N \geq 1)$	0.814	0.827	0.877	0.929	0.977	1.000	1.000	
	$\hat{P}_\tau(N \geq 1)$	0.816	0.836	0.878	0.927	0.972	0.998	1.000	0.004
	L-$\hat{P}_\tau(N \geq 1)$	0.337	0.480	0.740	0.873	0.908	0.932	0.951	
	U-$\hat{P}_\tau(N \geq 1)$	0.935	0.938	0.945	0.965	1.000	1.000	1.000	
4	$P_\tau(N \geq 1)$	0.888	0.899	0.936	0.969	0.993	1.000	1.000	
	$\hat{P}_\tau(N \geq 1)$	0.888	0.900	0.935	0.967	0.991	1.000	1.000	0.001
	L-$\hat{P}_\tau(N \geq 1)$	0.404	0.561	0.821	0.930	0.953	0.970	0.980	
	U-$\hat{P}_\tau(N \geq 1)$	0.972	0.974	0.978	0.988	1.000	1.000	1.000	
5	$P_\tau(N \geq 1)$	0.933	0.941	0.966	0.986	0.998	1.000	1.000	
	$\hat{P}_\tau(N \geq 1)$	0.930	0.940	0.965	0.985	0.997	1.000	1.000	0.001
	L-$\hat{P}_\tau(N \geq 1)$	0.459	0.632	0.877	0.961	0.977	0.987	0.992	
	U-$\hat{P}_\tau(N \geq 1)$	0.988	0.989	0.991	0.996	1.000	1.000	1.000	

h represents the width of the time interval on which forecasts were obtained, τ is the quantile level, $P_\tau(N \geq 1)$ represents the true probability; $\hat{P}_\tau(N \geq 1)$ represents the estimated probability; $(\hat{P}_{0.05}(N \geq 1), \hat{P}_{0.95}(N \geq 1))$ gives a 90% probability interval forecast for $P(N \geq 1)$; L-$\hat{P}_\tau(N \geq 1)$ (U-$\hat{P}_\tau(N \geq 1)$) represents the lower (upper) bound of a 90% probability interval for $P_\tau(N \geq 1)$; and RMSE is the square root of the mean squared errors between $P_\tau(N \geq 1)$ and $\hat{P}_\tau(N \geq 1)$.

estimated intensity functions with the true intensity function so that the effect of ERE can be evalueated.

4.2.1 Simulated data

The steps for obtaining processed data are given below.

(a) Simulate event times $s_0 < s_1 < \ldots < s_{n_1}$ from a HP model with intensity function

$$\lambda(t|\boldsymbol{\alpha}) = 0.75 + \sum_{s_i<t} 0.6e^{-1.8(t-s_i)}, \tag{10}$$

where $s_0 = 0$ and $s_{n_1} = T$ and $T = 500$. Hence, $s_0 < s_1 < \ldots < s_{n_1}$ are our raw data.

(b) Delete s_i as follows. Let $u_0 = 0$ and for $j > 0$, let $u_j = \max\{s_i | s_i - u_{j-1} < m_0\}$, resulting in $u_0 < u_1 < \ldots < u_{n_2}$, where $u_{n_2} = T$. This step mimics the situation in the data collection stage where events occurred in a very short time period, here shorter than m_0, are not recorded. In this study, we let $m_0 = 1$.

(c) Randomly delete m_1 event times from $u_1, \ldots, u_{n_2-1}$, resulting in $v_0 < v_1 < \cdots < v_{n_3}$, where $v_0 = 0$, $v_{n_3} = T$ and $n_3 = n_2 - m_1 + 1$. This step mimics the situation where some event times were not recorded for some unknown reasons. In this study, we let $m_1 = [w_1 * n_2]$, where $[x]$ represents the integer part of x and $w_1 = 5\%, 15\%, 25\%, 35\%, 45\%, 55\%$. Hence, w_1 and m_1 are the percentage and the number of events that are missing respectively.

(d) Randomly select m_2 event times from $v_1, \ldots, v_{n_3-1}$, denoted by v_{i_ℓ}, $\ell = 1, \ldots, m_2$. Let $w_{i_\ell j} = v_{i_\ell} + r_j$, where $j = 1, \ldots, m_3$, $r_1 = 0$ and r_j ($j \neq 1$) is a random sample between 0 and 1. Then we arrange the event times in the set $\{v_0, \ldots, v_{n_3}\} \backslash \{v_{i_\ell}, \ell = 1, \ldots, m_2\} \cup \{w_{i_\ell j}, \ell = 1, \ldots, m_2, j = 1, \ldots, m_3\}$ in an increasing order, resulting in the final processed data $t_0 < t_1 < \ldots < t_n$, where $t_0 = 0$, $t_n = T$ and $n = n_3 + 1 + m_2(m_3 - 1)$.

This final step mimics the situation where m_3 events occurred at time v_{i_ℓ}. In this case, we transform v_{i_ℓ} by adding a small amount that is uniformly distributed on (0, 1). In this

Table 3: Average number of events in 200 processed datasets

		$w_1 = 5\%$		$w_1 = 15\%$		$w_1 = 25\%$		$w_1 = 35\%$		$w_1 = 45\%$		$w_1 = 55\%$	
n_1	n_2	n_3	n	n_3	n	n_3	n	n_3	n	n_3	n	n_3	n
563	236	225	236	201	223	178	197	154	171	131	146	107	120

study, we let $m_3 = 3$ and $m_2 = [w_2 * n_3]$, where $w_2 = 5\%$, which corresponds to a situation where only 5% of the observed event times are affected by this type of random errors. Hence, this step allows us to assess the effect of ERE on the intensity of an underlying process when the percentage and the number of events that occurred simultaneously are fixed at a given level.

Therefore, for each raw dataset generated from model (10), 6 processed datasets have been obtained since we have six values of w_1, representing 6 different scenarios. Large values of w_1 imply that the quality of the processed data is low. Each processed dataset contains three different types of random errors introduced by the steps (b), (c) and (d) respectively, that mimic a possible scenario in reality. Clearly, the raw data $s_0 < s_1 < \ldots < s_{n_1}$ follow model (10), but the processed data $t_0 < \ldots < t_n$ are actually very different from the raw data due to the ERE introduced by steps (b)-(d). By repeating the steps (a)-(d) 200 times, we obtained 200 independent raw datasets, each of which introduced six processed datasets.

The average values of n_1, n_2, n_3 and n over 200 simulations are given in Table 3. It is seen that the number of event times that have been lost from the raw data in steps (b) and (c) is larger than the number of event times that have been added to the raw data in step (d). Therefore, we should expect that the estimated unconditional and conditional intensities of the processed data are lower than those of the raw data.

Table 4 shows the mean of the observed average intensity rates over 200 processed datasets in each scenario, confirming that estimated unconditional intensities are indeed much smaller than the true value, which is 1.125.

Table 4: Mean of the observed average intensity rates in 200 processed datasets

w_1	5%	15%	25%	35%	45%	55%
	0.476	0.451	0.399	0.346	0.294	0.243

4.2.2 Results about the estimated conditional intensity function

We evaluated the estimated conditional intensity functions at 2000 equally spaced time points on $[0, T]$ by using each processed dataset and the corresponding estimated parameters. Similarly, we also evaluated the true intensity functions at the same 2000 points on $[0, T]$ by using the raw data and the true parameter values. Then we calculated the average difference between: (i) the true intensity and the estimated intensity from Model B; (ii) the true intensity and the estimated intensity from Model A; and (iii) the estimated intensities from Model B and those from Model A.

By repeating the above calculations for each processed dataset, we obtained 200 values for each of the cases (i)-(iii) and for each value of w_1. If the ERE has no effect on the estimated intensities, then we should expect that these average values are distributed around 0. To avoid normality assumptions, the non-parametric singed rank test was used to test the null hypothesis that the median of the distribution of these average values is 0, against the alternative hypothesis that the median of the distribution is greater than 0.

We found that all the p-values of the tests are virtually 0 for cases (i)-(iii). For cases (i) and (ii), a small p-value implies that the conditional intensity function has been underestimated; and for case (iii), a small p-value suggests that Model B performs better than Model A under the simulation scenarios considered here.

We further calculated the values of the Bayesian Information Criterion (BIC) for all models, resulting in 2400 BIC values, half of which are for Model A and another half for Model B, where BIC is defined by BIC $= k\ln(n) - 2\ln(\hat{L})$, in whihc n is the number of event times, k is the number of parameters in the model, and $\hat{L}$ is the likelihood evaluated at $\hat{\boldsymbol{\alpha}}$. Table 5 shows the average BIC values over 200 estimated models in each case. The

Table 5: Average BIC values over the 200 simulated datasets in each case

w_1	5%		15%		25%	
	Model B	Model A	Model B	Model A	Model B	Model A
Average BIC	-792.23	-752.079	-773.102	-768.958	-728.703	-718.43
w_1	35%		45%		55%	
	Model B	Model A	Model B	Model A	Model B	Model A
Average BIC	-677.555	-686.758	-618.93	-613.938	-551.636	-526.151

overall average BIC values for Models A and B are given by -678.317 and -690.359 respectively. Clearly, based on the BIC, the average performance of Model B is also better compared with Model A. Hence, this simulation study shows that the use of hidden marks and the GLD has made Model B more flexible in dealing with the effect of ERE.

It is worth noting that this simulation study is very limited as many other possible situations have not been considered here. For example, new scenarios could occur if we use a different value of m_0, or if we combine step (a) with any one of the steps (b)-(d) or any combinations of (b)-(d) or any others that have not been considered in this paper. It is clearly out of the scope of this paper if we consider all these possible situations. Hence, further research on the effect of ERE is certainly required in the future.

5 Generalization

Now we briefly discuss how to generalize Model B to a mutually-exciting HP model with hidden marks. Consider a collection of m counting processes $\{N_1(t), \ldots, N_m(t)\}$, $t \geq 0$. Let $t_{ij} \in [0, T)$ be the observed arrival times for each counting process, where $i = 1, \ldots, m$ and $j = 0, 1, \ldots, n_i$. Similar to one dimensional case, we define $\lambda_i(t, \tau|\boldsymbol{\alpha}_i)$ by

$$\lambda_i(t, \tau|\boldsymbol{\alpha}_i) = \mu_i + \sum_{j=1}^{m} \sum_{t_{jk}<t} \alpha_{ij} e^{-\beta_{ij}(t-t_{jk})} + \xi_{t_{ij_0^i}}, \quad j_0^i = \max\{j : t_{ij} < t\},$$

where the hidden marks $\{\xi_{ij}, j = 1, \ldots, n_i\}$ are i.i.d. and ξ_{ij} and $\xi_{k\ell}$ are independent for $i \neq k$. Moreover, let ξ_{ij} follow the distribution defined by the quantile function $Q_i(\tau, \boldsymbol{\eta}_i) = \eta_{i1}\left\{\frac{\tau^{\eta_{i2}}-1}{\eta_{i2}} - \frac{(1-\tau)^{\eta_{i3}}-1}{\eta_{i3}}\right\}$, where $\tau \in (0,1)$. Then the quantile function of the conditional intensity function of the mutually-exciting HP model with hidden marks is given by

$$Q_{\lambda_i}(t, \tau|\boldsymbol{\alpha}_i) = \mu_i + \sum_{j=1}^{m}\sum_{t_{jk}<t} \alpha_{ij} e^{-\beta_{ij}(t-t_{jk})} + Q_i(\tau, \boldsymbol{\eta}_i), \quad i = 1, \ldots, m,$$

where $t \geq 0$, $\tau \in (0,1)$, $\boldsymbol{\alpha}_i = (\mu_i, \alpha_{ij}, \beta_{ij}, \boldsymbol{\eta}_i, j = 1, \ldots, m)$, $\boldsymbol{\eta}_i = (\eta_{i1}, \eta_{i2}, \eta_{i3})$, $\mu_i > \eta_{i1}/\eta_{i2}$, $\eta_{i1} > 0$, $0 < \eta_{i2} < 1$, $\eta_{i3} < 0$, $\alpha_{ij} > 0, \beta_{ij} > 0$, $\sum_{j=1}^{m} \alpha_{ij}/\beta_{ij} < 1$. Specifically, if $m = 2$, we have

$$\begin{aligned} Q_{\lambda_1}(t, \tau|\boldsymbol{\alpha}_1) &= \mu_1 + \textstyle\sum_{t_{1k}<t} \alpha_{11} e^{-\beta_{11}(t-t_{1k})} + \sum_{t_{2k}<t} \alpha_{12} e^{-\beta_{12}(t-t_{2k})} + Q_1(\tau, \boldsymbol{\eta}_1), \\ Q_{\lambda_2}(t, \tau|\boldsymbol{\alpha}_2) &= \mu_2 + \textstyle\sum_{t_{1k}<t} \alpha_{21} e^{-\beta_{21}(t-t_{1k})} + \sum_{t_{2k}<t} \alpha_{22} e^{-\beta_{22}(t-t_{2k})} + Q_2(\tau, \boldsymbol{\eta}_2). \end{aligned} \tag{11}$$

Clearly, the intensity function of the ith process depends not only on its own history but also on the history of all other processes. Hence the interactions between financial processes can be taken into account. Moreover, we use different $Q_i(\tau, \boldsymbol{\eta}_i)$ to model the effect of ERE, which allows us to deal with the situations in which different data collection mechanisms and/or different data cleaning techniques have been used for different processes.

The estimation of the mutually-exciting HP model with hidden marks is also similar to that of Model B. Let $\mathbf{t}_{ij} = (t_{i0}, \ldots, t_{ij})$, $\boldsymbol{\tau}_{ij} = (\tau_{i0}, \ldots, \tau_{ij})$, $\tau_{ij} \in (0,1)$, $\mathbf{t}_i = \mathbf{t}_{in_i}$ and $\boldsymbol{\tau}_i = \boldsymbol{\tau}_{in_i}$, where $i = 1, \ldots, m$, $j = 0, \ldots, n_i$, and let $\boldsymbol{\tau} = (\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_m)$, $\mathbf{t} = (\mathbf{t}_1, \ldots, \mathbf{t}_m)$ and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m)$. Then it follows from Daley and Vere-Jones (2003, p251) that the likelihood function of the parameters is given by $L = L_1 \cdots L_m$, where L_i is the likelihood for the ith counting process, where $i = 1, \ldots, m$. Hence, for example, when $m = 2$, i.e. for model (11), the posterior density function of the model parameters is given by

$$\begin{aligned} &\pi(\boldsymbol{\tau}, \boldsymbol{\alpha}|\mathbf{t}) \\ &\propto \pi(\boldsymbol{\alpha}) \textstyle\prod_{i=1}^{n_1} Q_{\lambda_1}(t_{1i}, \tau_{1i}|\boldsymbol{\alpha}_1) e^{-\int_0^T Q_{\lambda_1}(t, \tau_{1i}|\boldsymbol{\alpha}_1)dt} \prod_{j=1}^{n_2} Q_{\lambda_2}(t_{2j}, \tau_{2j}|\boldsymbol{\alpha}_2) e^{-\int_0^T Q_{\lambda_2}(t, \tau_{2j}|\boldsymbol{\alpha}_2)dt}, \end{aligned}$$

where $\pi(\boldsymbol{\alpha})$ is the prior density function of $\boldsymbol{\alpha}$. Therefore, a MCMC method can be used for parameter estimation.

6 Data Applications

6.1 Transaction Data

This application shows that the effect of ERE can indeed cause the intensity of an underlying financial process to be underestimated in practice and Model B performs better compared with Model A in the presence of ERE. Hence, these results confirm what we have found in our simulation studies.

We consider a small set of transaction data with timestamps of one second precision[1]. The transaction data contain trading times within a time period of 56 minutes. We noticed that some trades that occurred simultaneously have also been recorded. Hence, we followed the work of Bowsher (2007) and adjusted the times of the trading events by adding some appropriate random noise. The black continuous curve in Figure 4 shows the empirical trade counts per minute for the adjusted data. It is seen that the adjusted data is the data to be analysed and they contain at least two types of random errors that are similar to those introduced by steps (b) and (d) in Section 4.2.1.

Models A and B were fitted to the adjusted data and the estimation results are given in Table 6. Hence, the estimated intensity function corresponding to Model A is given by

$$\lambda(t) = 0.1496 + \sum_{t_i<t} 1.2206\, e^{-2.7491(t-t_i)}, \tag{12}$$

[1]The dataset is available from `http://www.tickdatamarket.com/formats-donnees.php?id=1/&nom=tick`

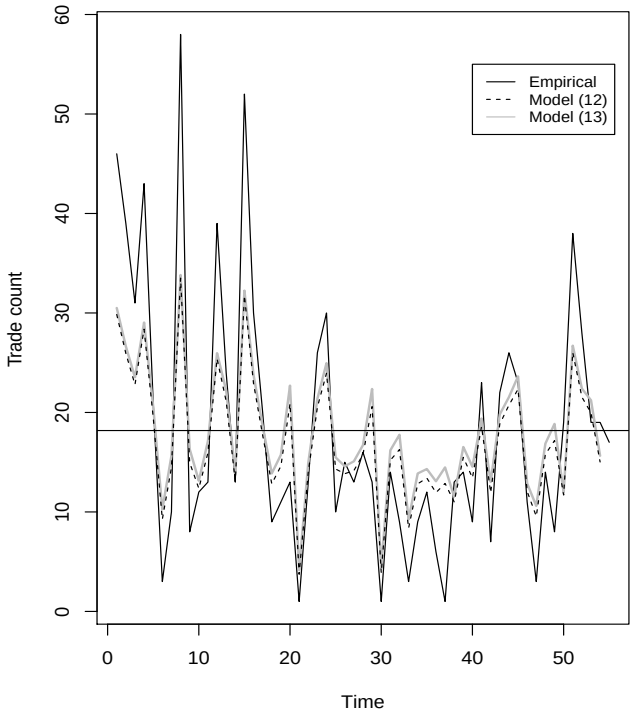

Figure 4: Empirical (black continuous) and estimated (dashed curve for model (12), grey curve for model (13)) trade counts per minute for the transaction data, where the horizontal line indicates the empirical average trade count per minute.

---

and corresponding to Model B it is given by

$$Q_\lambda(t,\tau|\boldsymbol{\alpha}) = 0.1612 + \sum_{t_i<t} 119.1\, e^{-276.0(t-t_i)} + 0.027\left\{\frac{\tau^{0.4524}-1}{0.4524} - \frac{(1-\tau)^{-0.704}-1}{-0.704}\right\}. \tag{13}$$

For model (12) the MLE method was used to obtain the estimated parameters and the corresponding 95% confidence intervals. For model (13), our Bayesian MCMC method was used. The parameters were estimated by the mean of the respectively posterior samples, and the lower and upper bounds of the 95% credible intervals correspond to the 2.5% and 97.5% quantiles of the posterior samples respectively.

Figure 5 checks the goodness-of-fit of the two models. It is seen that model (12) fits badly when $(x, y) < (0.4, 0.4)$. Note that in this application some observed event times were adjusted by adding some appropriate random noise to ensure that no events could

Table 6: Estimated model parameters

	Model (12)			Model (13)					
Parameters	α_0	α_1	α_2	α_0	α_1	α_2	η_0	η_1	η_2
Estimates	0.1496	1.2206	2.7491	0.161	119.1	276.0	0.027	0.452	-0.704
L-bound	0.1348	1.0062	2.2081	0.120	102.7	249.8	0.002	0.060	-3.960
U-bound	0.1645	1.4349	3.2902	0.181	136.4	304.2	0.104	0.963	-0.026

The row labeled by 'Estimates' gives the estimated parameter values. The rows labeled by 'L-bound' and 'U-bound' define the lower and upper bounds of a 95% confidence/credible interval for the corresponding parameters.

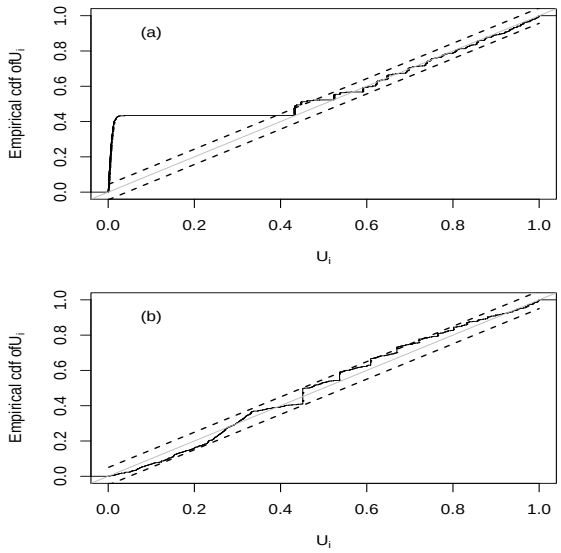

Figure 5: Goodness-of-fit of (a) model (12) and (b) model (13) for the transaction data. The dark continuous curve is the empirical cdf of $\hat{U}_i = 1 - e^{-\hat{b}_i}$. The grey line $y = x$ is the theoretical cdf of U(0, 1). The dashed lines represent 95% confidence bounds for U(0, 1).

occur at the same time, which results in a group of events in the adjusted data that are close to each other. Figure 5(a) reflects this fact correctly as the empirical cdf of U_i increases quickly when U_i is small. Note that small values of U_i correspond to small values of b_i, and hence small values of $t_i - t_{i-1}$. On the other hand, we used data with timestamps of one second precision, which means that short durations of trades were not recorded, leading to a gap on the observed event times. This gap is also reflected by the horizontal segment of the empirical cdf of U_i on Figure 5(a). Hence, the large deviation from the reference line shown in Figure 5(a) is caused by the lack of small durations in the data due to the data collection mechanism and the data cleaning procedure. However, Figure 5(b)

shows clearly that model (13) does not suffer from the same problem. This example further confirms that the effect of ERE should not be ignored in practice and the proposed model is able to provide much improved results in the presence of ERE. The results also suggest that statistical inferences on short time durations could be problematic if model (12) is used, and hence caution should be exercised in practice.

It follows from model (12) that the immigration intensity of the trading process is given by 0.1496 and the expected intensity is given by 0.2691. Hence, model (12) says that about 16.143 (or about 16) trades are expected to occur in one minute. For model (13), the intensity $\lambda(t|\boldsymbol{\alpha})$ is estimated by $Q_\lambda(t, 0.5|\boldsymbol{\alpha})$. Hence the immigration intensity is given by $0.1612 + 0.0081 = 0.1693$ and the expected intensity is given by 0.2978. So model (13) tells us that the average trade count per minute is 17.869, i.e. about 18 trades are expected in one minute. Table 7 summarizes these results. It is seen that the results obtained from model (13) are much more consistent with the observed data. Figure 4 also shows the estimated trade counts per minute for both models, where the horizontal line indicates the empirical average trade count per minute. In summary, the results obtained from this application are consistent with what we have found in our simulation studies.

Table 7: The empirical and the estimated intensity rate and average trade count per minute

	Average intensity rate	Average trade count per minute
Observed	0.3030	18.180
Model (12)	0.2691	16.143
Model (13)	0.2978	17.869

The average intensity rate is estimated by $\alpha_0/(1-\alpha_1/\alpha_2)$ for model (12) and by $(\alpha_0+\eta_0 Q(0.5, \eta_1, \eta_2))/(1-\alpha_1/\alpha_2)$ for model (13). The average trade count per minute is calculated by multiplying the average intensity rate by 60.

6.2 Futures Data

We consider a sample of futures data on trades and quotes that cover a 7.5-hour period[2]. This dataset contains information that allows us to study the interaction between trading ac-

[2]The data are available from `http://www.tickdatamarket.com/formats-donnees.php?id=3/&nom=trade-and-quotes`

tivity and price volatility. Hence, we illustrate the analysis by using the proposed mutually-exciting HP model with hidden marks, i.e. model (11).

The observed intensity in one minute window for both trade arrival times and mid-quote price change times are shown by the black curves in Figure 6, in which the red curves, obtained by using local polynomial regression fitting method (see e.g. Cleveland et al. Chapter 8, 1992), show the overall evolution patterns of two intensity functions respectively. It is seen that the evolution patterns of the two intensity functions are similar, which suggests that a trade might affect the waiting time to the next change in the mid-quote price, and a mid-quote price change might also lead to a change in the trading intensity.

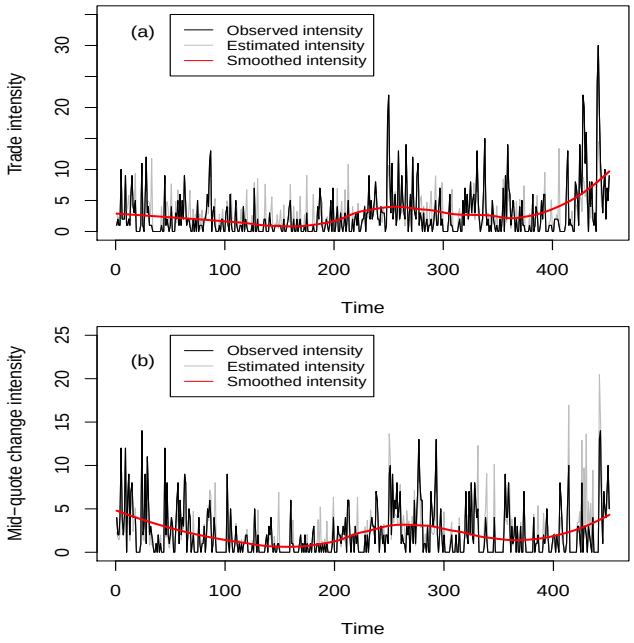

Figure 6: Estimated (grey) and observed (black) intensity functions in one minute window for (a) trade arrival times and (b) mid-quote change times respectively. Red curves are the smoothed intensities.

Table 8 gives the estimated parameter values and the associated 95% credible intervals, where the estimated parameter values are the mean of the respective posterior samples and the lower and upper bounds of the 95% credible intervals correspond to the 2.5% and 97.5% quantiles of the posterior samples respectively. Hence, the estimated model is given by

Table 8: Estimated parameters and the associated 95% credible intervals

Parameters	α_{11}	α_{12}	α_{21}	α_{22}	β_{11}	β_{12}	β_{21}	β_{22}
Estimates	2.212	0.011	16.399	0.040	7.468	0.026	45.223	0.103
L-bound	1.769	0.008	14.349	0.032	5.899	0.018	40.813	0.079
U-bound	2.795	0.015	18.791	0.052	9.721	0.040	50.267	0.138
Parameters	η_{11}	η_{12}	η_{13}	η_{21}	η_{22}	η_{23}	μ_1	μ_2
Estimates	0.004	0.560	-0.296	0.002	0.641	-0.394	0.016	0.005
L-bound	0.000	0.082	-3.560	0.000	0.158	-2.942	0.011	0.003
U-bound	0.012	0.980	-0.009	0.004	0.982	-0.013	0.020	0.006

$$
\begin{aligned}
Q_{\lambda_1}(t,\tau|\boldsymbol{\alpha}_1) &= 0.016 + \textstyle\sum_{t_{1k}<t} 2.212 e^{-7.468(t-t_{1k})} + \sum_{t_{2k}<t} 0.011 e^{-0.026(t-t_{2k})} \\
&\quad + 0.004 \left\{ \tfrac{\tau^{0.56}-1}{0.56} - \tfrac{(1-\tau)^{-0.296}-1}{-0.296} \right\}, \\
Q_{\lambda_2}(t,\tau|\boldsymbol{\alpha}_2) &= 0.005 + \textstyle\sum_{t_{1k}<t} 16.399 e^{-45.223(t-t_{1k})} + \sum_{t_{2k}<t} 0.040 e^{-0.103(t-t_{2k})} \\
&\quad + 0.002 \left\{ \tfrac{\tau^{0.641}-1}{0.641} - \tfrac{(1-\tau)^{-0.394}-1}{-0.394} \right\}.
\end{aligned}
\tag{14}
$$

The estimated model has the following interpretations. Let us consider $Q_{\lambda_1}(t,\tau|\boldsymbol{\alpha}_1)$, corresponding to the trading process. We see that each new trade in the system instantaneously increases the trading intensity by 2.212, then over time this arrival's influence decays at a rate of 7.468. Moreover, each new price change instantaneously increases the trading intensity by 0.011, then over time this arrival's influence also decays but at a rate of 0.026. Similar interpretations hold for $Q_{\lambda_2}(t,\tau|\boldsymbol{\alpha}_2)$, corresponding to the mid-quote price change process.

Therefore, the estimated model suggests that, an arrival of trade has a larger instantaneous impact on the two intensity functions but its impact lasts only for a short period of time, while an arrival of price change has a smaller instantaneous impact on the intensity functions but its impact can last for a much longer time period.

The estimated intensity functions were also shown in Figure 6 by using grey curves. As it is difficult to visually check the differences between the estimated and observed intensities in one minute window, we calculated the RMSE between the observed and estimated number of events per minute, the results of which are 2.66 and 1.74 for trades and mid-

quote price changes respectively. This means that, on average, the difference between the observed and estimated trade counts is less than 3 trades per minute and that for the mid-quote changes is less than 2 changes per minute.

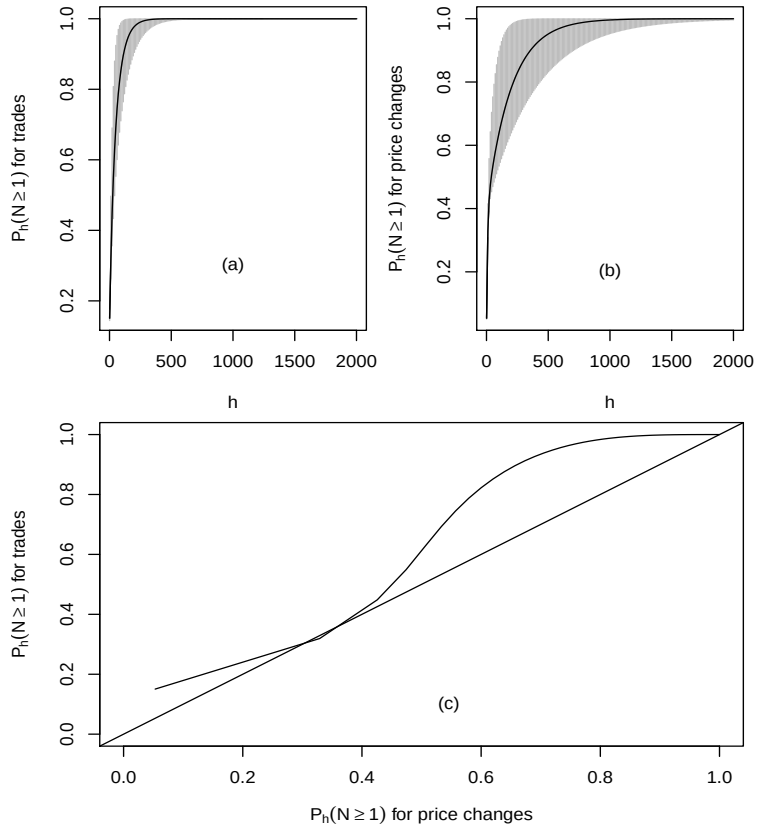

Figure 7: (a) (b): The plots of the probability $P_h(N_i \geq 1)$ against h for trades and price changes respectively, where the grey areas show the 95% interval forecasts for these probabilities. (c) Plot of $P_h(N_1 \geq 1)$ for trades against that for price changes.

Let N_1 and N_2 be the number of trades and mid-quote changes occurring in $(T_1, T_1+h]$ and $(T_2, T_2+h]$ respectively, where $h > 0$ and T_1 and T_2 are the final time for trades and mid-quote changes respectively. Then, probability forecasts for $P(N_i \geq 1)$, denoted by $P_h(N_i \geq 1)$, can be calculated for different values of h by using Proposition 5(i). Moreover, the lower and upper bounds of a 95% interval forecast for $P(N_i \geq 1)$ can be obtained by using Proposition 5(ii).

Figure 7 shows the results of probability forecasts for $h = 1, \ldots, 2000$ seconds, where Figure 7 (a) and (b) are for trade arrival times and mid-quote price change times respec-

tively, and the grey areas show the 95% interval forecasts for $P(N_i \geq 1)$ on $(T_i, T_i+h]$. It is seen that trades arrive much sooner than that for mid-quote price changes as the probability $P_h(N_1 \geq 1)$ for trades increases much faster than that for price changes. Figure 7 (c) is the plot of $P_h(N_1 \geq 1)$ for trades against $P_h(N_2 \geq 1)$ for mid-quote price changes. The positive association between them suggests that when the probability of having a price change increases, the probability of having a trade also increases. Figure 7 (c) further shows that the probability $P_h(N_1 \geq 1)$ for trades is larger than that for price changes over almost the entire range between 0 and 1, suggesting that price changes are more likely to trigger more trading activities.

Finally, to have a closer look at how the waiting time of one process is affected by another process, we fitted Model B to the trading times and the mid-quote changes separately, resulting in the following two models. For the trading process we have

$$Q_\lambda(t, \tau|\boldsymbol{\alpha}) = 0.028 + \sum_{t_i<t} 2.796\, e^{-10.991(t-t_i)} + 0.006 \left\{ \frac{\tau^{0.558} - 1}{0.558} - \frac{(1-\tau)^{-0.288} - 1}{-0.288} \right\}, \tag{15}$$

and for the mid-quote changes we have

$$Q_\lambda(t, \tau|\boldsymbol{\alpha}) = 0.025 + \sum_{t_i<t} 1.826\, e^{-16.598(t-t_i)} + 0.007 \left\{ \frac{\tau^{0.590} - 1}{0.590} - \frac{(1-\tau)^{-0.278} - 1}{-0.278} \right\}. \tag{16}$$

By using the estimated models (14), (15) and (16), we calculated the average intensity rates of the two processes separately, see Table 9. It is seen that the interaction between

Table 9: Estimated average intensity rate for trades and mid-quote changes

	Trades	Mid-quote changes
Model (14)	0.0591	0.020
Model (15)	0.0386	-
Model (16)	-	0.030

the two processes has shortened the average waiting time for the trading process, which is consistent with what we have observed from Figure 7 (c). On the other hand, the interaction

between the two processes has lengthened the average waiting time for the mid-quote price change process, which might suggest that unless necessary, traders may not want to change prices when the trading process is going smoothly.

7 Conclusions

We propose a novel HP model with hidden marks for financial event data. The hidden marks are used to take account the effect of ERE caused by data collection mechanisms and data cleaning procedures. We also develop a MCMC method for parameter estimation to allow the proposed model to be used in practice.

We conduct extensive simulation studies to assess the estimation method and the effect of ERE on the intensity function. We show that the proposed estimation method works well. We also show that when ERE cause information about the underlying process to be lost, the intensity function may be underestimated. Our results further show that the proposed model performs better in the presence of ERE.

It is worth noting that our simulation study for assessing the effect of ERE is very limited as many other possible situations that could occur in the data collection and data cleaning stage in practice have not been considered. Therefore, further work is definitely needed in this regard. This could be done by using, for example, a simulation method such as the one we exemplified in Section 4.

As pointed out by a referee, the hidden marks $\xi_{t_{i_0}}$ could also be included into Model A by using the following formulation:

$$\lambda(t|\boldsymbol{\alpha}) = \alpha_0 + \sum_{t_i<t} \xi_{t_{i_0}} \gamma(t - t_i|\boldsymbol{\alpha}_1).$$

It is worth noting that under this formulation, we need to choose a new distribution for $\xi_{t_{i_0}}$ in order to ensure that the intensity function is positive for all t. There is no doubt that

a comprehensive study of this new model is worthwhile, but we have to leave it to future work.

In this paper we combined our approach with a self-exciting HP model. However, it is worth reemphasizing that our method can also be used in conjunction with existing HP models, such as the model proposed by Bauwens and Hautsch (2006). If we do so, we will be able to deal with the effect of ERE, the randomness in realized jumps and some other components of an underlying financial process simultaneously. When more data is available, it is absolutely necessary to make a more thorough comparison between the two types of HP models. We leave them for future research.

Acknowledgements

We sincerely thank the referees for the very insightful comments and very constructive suggestions. We would also like to thank Professor Alan Hawkes for his detailed comments. All the comments and suggestions we received have led to substantial improvements in this paper.

Appendix I

Proof of Proposition 1.

Note that $\frac{dQ(\tau,\eta_1,\eta_2)}{d\tau} = \tau^{\eta_1-1} + (1-\tau)^{\eta_2-1} > 0$ for all $\tau \in (0,1)$. Hence, for all $\tau > 0$, we have $Q(\tau,\eta_1,\eta_2) > Q(0,\eta_1,\eta_2) = -1/\eta_1$. On the other hand, for all $\tau < 1$, we have $Q(\tau,\eta_1,\eta_2) < \lim_{\tau\to 1} Q(\tau,\eta_1,\eta_2) = \infty$. This completes the proof.

Proof of Proposition 2.

As $\eta_0 Q(\tau,\eta_1,\eta_2)$ is the quantile function of $\xi_{t_{i_0}}$, we have, for any $\tau \in (0,1)$, $P(\xi_{t_{i_0}} \leq$

$\eta_0 Q(\tau, \eta_1, \eta_2)) = \tau$. Hence,

$$P\left(\alpha_0 + \sum_{t_i<t} \gamma(t - t_i|\boldsymbol{\alpha}_1) + \xi_{t_{i_0}} \leq \alpha_0 + \sum_{t_i<t} \gamma(t - t_i|\boldsymbol{\alpha}_1) + \eta_0 Q(\tau, \eta_1, \eta_2)\right) = \tau.$$

Since $\lambda(t|\boldsymbol{\alpha}) = \alpha_0 + \sum_{t_i<t} \gamma(t - t_i|\boldsymbol{\alpha}_1) + \xi_{t_{i_0}}$, we see that

$$P\left(\lambda(t|\boldsymbol{\alpha}) \leq \alpha_0 + \sum_{t_i<t} \gamma(t - t_i|\boldsymbol{\alpha}_1) + \eta_0 Q(\tau, \eta_1, \eta_2)\right) = \tau.$$

Hence, by definition, the quantile function of the intensity function at time t is given by $Q_\lambda(t, \tau|\boldsymbol{\alpha})$ as required.

Proof of Proposition 3.

It follows from Proposition 1 that $\eta_0 Q(\tau, \eta_1, \eta_2) > -\eta_0/\eta_1$. Hence, $Q_\lambda(t, \tau|\boldsymbol{\alpha}) > \alpha_0 + \eta_0 Q(\tau, \eta_1, \eta_2) > \alpha_0 - \eta_0/\eta_1$. It follows from $\alpha_0 > \eta_0/\eta_1$ on $\bar{\Omega}$ that $Q_\lambda(t, \tau|\boldsymbol{\alpha}) > 0$ for all t and τ, as required.

Proof of Proposition 4.

(i) First note that it follows from the definition that $Q_\lambda(t, 0.5|\boldsymbol{\alpha})$ is the conditional median of the intensity function.

(ii) Consider the conditional expectation of the intensity function.

$$\begin{aligned}
\mu_\lambda(t) &= \int_0^1 \left[\alpha_0 + \textstyle\sum_{t_i<t} \alpha_1 e^{-\alpha_2(t-t_i)} + \eta_0 \left(\frac{\tau^{\eta_1}-1}{\eta_1} - \frac{(1-\tau)^{\eta_2}-1}{\eta_2}\right)\right] d\tau \\
&= \alpha_0 + \textstyle\sum_{t_i<t} \alpha_1 e^{-\alpha_2(t-t_i)} + \eta_0 \left[\frac{1}{\eta_1}\left(\frac{1}{\eta_1+1} - 1\right) - \frac{1}{\eta_2}\left(\frac{1}{\eta_2+1} - 1\right)\right] \\
&= \alpha_0 + \textstyle\sum_{t_i<t} \alpha_1 e^{-\alpha_2(t-t_i)} + \eta_0 \left[\frac{1}{\eta_2+1} - \frac{1}{\eta_1+1}\right]
\end{aligned}$$

as required.

(iii) Consider the conditional variance of the intensity function.

$$\begin{aligned}
\sigma_\lambda^2 &= \eta_0^2 \left[\int_0^1 \left(\tfrac{\tau^{\eta_1}-1}{\eta_1} - \tfrac{(1-\tau)^{\eta_2}-1}{\eta_2} \right)^2 d\tau - \mu_\lambda^2 \right] \\
&= \eta_0^2 \left[\tfrac{1}{\eta_1^2} \left(\tfrac{1}{2\eta_1+1} - \tfrac{2}{\eta_1+1} + 1 \right) + \tfrac{1}{\eta_2^2} \left(\tfrac{1}{2\eta_2+1} - \tfrac{2}{\eta_2+1} + 1 \right) \right. \\
&\quad \left. - \tfrac{2}{\eta_1\eta_2} \left(B(\eta_1+1, \eta_2+1) - \tfrac{1}{\eta_1+1} - \tfrac{1}{\eta_2+1} + 1 \right) - \left(\tfrac{1}{\eta_2+1} - \tfrac{1}{\eta_1+1} \right)^2 \right] \\
&= \eta_0^2 \left[\tfrac{1}{\eta_1^2} \left(\tfrac{1}{2\eta_1+1} - \tfrac{1}{(\eta_1+1)^2} \right) + \tfrac{1}{\eta_2^2} \left(\tfrac{1}{2\eta_2+1} - \tfrac{1}{(\eta_2+1)^2} \right) - \tfrac{2}{\eta_1\eta_2} \left(B(\eta_1+1, \eta_2+1) - \tfrac{1}{(\eta_1+1)(\eta_2+1)} \right) \right]
\end{aligned}$$

as required.

(iv) This is because $Q_\lambda(t, 1-\tau'|\boldsymbol{\alpha})$ and $Q_\lambda(t, \tau'|\boldsymbol{\alpha})$ define the $(1-\tau')$th and τ'th quantiles of $\lambda(t|\boldsymbol{\alpha})$ respectively. This completes the proof.

Proof of Proposition 5.

(i) To show this part of the proposition, we need to show (a) Λ_τ is the quantile function of Λ, (b) $P_\tau(N \geq 1)$ is the quantile function of $P(N \geq 1)$. Hence, $P(N \geq 1)$ can be estimated by $P_{\tau=0.5}(N \geq 1)$.

First, we consider (i)(a). First note that model (6) implies that $\xi_{t_{i_0}}$ follows the distribution defined by the quantile function $\eta_0 Q(\tau|\eta_1, \eta_2)$. Then it follows from

$$\begin{aligned}
& P(\Lambda \leq \Lambda_\tau) \\
&= P\left\{ \textstyle\int_T^{T+h} \lambda(t|\boldsymbol{\alpha}) dt \leq \int_T^{T+h} Q_\lambda(t, \tau|\boldsymbol{\alpha}) dt \right\} \\
&= P\left\{ \textstyle\int_T^{T+h} \left(\alpha_0 + \sum_{t_i<t} \gamma(t-t_i|\boldsymbol{\alpha}_1) + \xi_{t_{i_0}} \right) dt \right. \\
&\qquad \left. \leq \textstyle\int_T^{T+h} \left(\alpha_0 + \sum_{t_i<t} \gamma(t-t_i|\boldsymbol{\alpha}_1) + \eta_0 Q(\tau|\eta_1, \eta_2) \right) dt \right\} \\
&= P\left\{ h\xi_{t_{i_0}} \leq h\eta_0 Q(\tau|\eta_1, \eta_2) \right\} = P\left\{ \xi_{t_{i_0}} \leq \eta_0 Q(\tau|\eta_1, \eta_2) \right\} = \tau
\end{aligned}$$

that Λ_τ is the τth quantile of Λ as required. Moreover, for model (6), we have

$$
\begin{aligned}
&\Lambda_\tau = \textstyle\int_T^{T+h} \left\{\alpha_0 + \sum_{t_i<t} \alpha_1 e^{-\alpha_2(t-t_i)} + \eta_0 Q(\tau, \eta_1, \eta_2)\right\} dt \\
&= \alpha_0 h + \eta_0 Q(\tau, \eta_1, \eta_2) h + \textstyle\sum_{t_i<T+h} \frac{\alpha_1}{\alpha_2} \left\{1 - e^{-\alpha_2(T+h-t_i)}\right\} \\
&\quad - \textstyle\sum_{t_i<T} \frac{\alpha_1}{\alpha_2} \left\{1 - e^{-\alpha_2(T-t_i)}\right\} \\
&= \{\alpha_0 + \eta_0 Q(\tau, \eta_1, \eta_2)\} h + \textstyle\sum_{t_i<T+h} \frac{\alpha_1}{\alpha_2} \left\{e^{-\alpha_2(T-t_i)} - e^{-\alpha_2(T+h-t_i)}\right\}
\end{aligned}
$$

as required. Note that in the above expression we used the fact $\sum_{t_i<T+h} = \sum_{t_i<T}$ as we have no observed events between T and $T+h$.

Now we consider (i)(b). It follows from

$$
P\left\{P(N \geq 1) \leq P_\tau(N \geq 1)\right\} = P\left\{1 - e^{-\Lambda} \leq 1 - e^{-\Lambda_\tau}\right\} = P(\Lambda \leq \Lambda_\tau) = \tau
$$

that $P_\tau(N \geq 1)$ is the τth quantile of $P(N \geq 1)$. Hence, $P(N \geq 1)$ can be estimated by $P_{\tau=0.5}(N \geq 1)$ as required.

(ii) This is because $P_{1-\tau'}(N \geq 1)$ and $P_{\tau'}(N \geq 1)$ are the $(1-\tau')$th and τ'th quantiles of $P(N \geq 1)$ respectively. This completes the proof.

Proof of Proposition 6.

It follows from $-M \leq \eta_2 \leq -\epsilon$ and $0 < \tau_j \leq 1 - \epsilon$ that $\epsilon^{-M} \geq \epsilon^{\eta_2} \geq \epsilon^{-\epsilon}$ and $(1-\tau_j)^{\eta_2} \leq \epsilon^{\eta_2}$ respectively. Hence

$$
\begin{aligned}
&\pi(\boldsymbol{\tau}_n, \boldsymbol{\alpha}|\mathbf{t}_n) = M_1 \textstyle\prod_{j=1}^n \left\{\alpha_0 + \sum_{t_i<t_j} \alpha_1 e^{-\alpha_2(t_j-t_i)} + \eta_0 Q(\tau_j, \eta_1, \eta_2)\right\} \\
&\quad \times exp\left[-\textstyle\int_0^T \left\{\alpha_0 + \sum_{t_i<s} \alpha_1 e^{-\alpha_2(s-t_i)}\right\} ds - \eta_0 \textstyle\sum_{j=1}^n Q(\tau_j, \eta_1, \eta_2)(t_j - t_{j-1})\right] \pi(\boldsymbol{\alpha}) \\
&\leq M_1 \textstyle\prod_{j=1}^n \left\{\alpha_0 + n\alpha_1 + \eta_0 \frac{\epsilon^{\eta_2}-1}{-\eta_2}\right\} \; e^0 \; e^{-\eta_0 \sum_{j=1}^n Q(\tau_j, \eta_1, \eta_2)(t_j - t_{j-1})} \pi(\boldsymbol{\alpha}) \\
&\leq M_1 \textstyle\prod_{j=1}^n \left\{\alpha_0 + nM + \frac{M(\epsilon^{-M}-1)}{\epsilon}\right\} \; e^{(\eta_0/\eta_1) \sum_{j=1}^n (t_j - t_{j-1})} \pi(\boldsymbol{\alpha}) \\
&\leq M_1 \textstyle\prod_{j=1}^n \left\{\alpha_0 + nM + \frac{M(\epsilon^{-M}-1)}{\epsilon}\right\} \; e^{\eta_0 T/\eta_1} \pi(\boldsymbol{\alpha}) \\
&\leq M_1 \textstyle\prod_{j=1}^n \left\{M + nM + \frac{M(\epsilon^{-M}-1)}{\epsilon}\right\} \; e^{MT} \pi(\boldsymbol{\alpha}) \leq \bar{M} \pi(\boldsymbol{\alpha}),
\end{aligned}
$$

where M_1 is the normalizing constant and $\bar{M}$ is a constant such that

$$\bar{M} \geq M_1 \left\{ M + nM + \frac{M(\epsilon^{-M} - 1)}{\epsilon} \right\}^n \; e^{MT}.$$

Hence

$$\int_\Omega \pi(\boldsymbol{\tau}_n, \boldsymbol{\alpha}|\mathbf{t}_n) d\boldsymbol{\tau}_n d\boldsymbol{\alpha} \leq \bar{M} \int_{\Omega_1} \pi(\boldsymbol{\alpha}) d\boldsymbol{\alpha} \int_{\Omega_2} d\boldsymbol{\tau}_n < \infty$$

as required.

Appendix II

The prior density functions.

$$\begin{aligned}
\pi(\alpha_1) &= (\sqrt{2\pi}\, u_1 \alpha_1)^{-1} \; e^{-(\log^2 \alpha_1)/2u_1^2}, \\
\pi(\alpha_2) &= (\sqrt{2\pi}\, u_2 \alpha_2)^{-1} \; e^{-(\log^2 \alpha_2)/2u_2^2} \; [1 - \Phi\{(\log \alpha_1)/u_2\}]^{-1},
\end{aligned}$$

$$\begin{aligned}
&\pi(\alpha_0|\eta_0, \eta_1) \\
&\quad = (\sqrt{2\pi}\, u_0 \alpha_0)^{-1} e^{-(\log^2 \alpha_0)/2u_0^2} \left\{ \textstyle\int_{\eta_0/\eta_1}^{\infty} \frac{1}{\sqrt{2\pi}\, u_0 \alpha_0} e^{-(\log^2 \alpha_0)/2u_0^2} d\alpha_0 \right\}^{-1} \\
&\quad = e^{-(\log^2 \alpha_0)/2u_0^2} \left(\sqrt{2\pi}\, u_0 \alpha_0 \left[1 - \Phi \left\{ \tfrac{\log(\eta_0/\eta_1)}{u_0} \right\} \right] \right)^{-1}, \\
&\pi(\eta_1) = (\sqrt{2\pi}\, v_1 \eta_1)^{-1} e^{-(\log^2 \eta_1)/2v_1^2} \left\{ \textstyle\int_0^1 \frac{1}{\sqrt{2\pi}\, v_1 \eta_1} e^{-(\log^2 \eta_1)/2v_1^2} d\eta_1 \right\}^{-1} \\
&\qquad = 2(\sqrt{2\pi}\, v_1 \eta_1)^{-1} e^{-(\log^2 \eta_1)/2v_1^2}, \\
&\pi(\eta_2) = \{\sqrt{2\pi}\, v_2(-\eta_2)\}^{-1} e^{-\{\log^2(-\eta_2)\}/2v_2^2}, \\
&\pi(\eta_0) = \{\sqrt{2\pi}\, v_0(\eta_0)\}^{-1} e^{-\{\log^2(\eta_0)\}/2v_0^2},
\end{aligned}$$

where u_i is the scale parameter for α_i, v_i is the scale parameter for η_i and $i = 0, 1, 2$.

MCMC method.

Let $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}_n$ represent the current values of the parameters, and $\boldsymbol{\alpha}'$ and $\boldsymbol{\tau}_n'$ represent the proposed values. Then the main steps of our MCMC method are given below.

Step 1. Obtain the proposed value $\boldsymbol{\alpha}'$. Note that due to the dependence structure between the parameters, we need to obtain $\boldsymbol{\alpha}'$ in the following order of priority:

(a) Propose α_1' by simulating $\log \alpha_1' \sim N(\log \alpha_1, \sigma_1^2)$

(b) Propose α_2' by simulating $\log \alpha_2' \sim N(\log \alpha_2, \sigma_2^2)$ such that $\log \alpha_2' > \log \alpha_1'$.

(c) Propose η_0' by simulating $\log \eta_0' \sim N(\log \eta_0, s_0^2)$

(d) Propose η_2' by simulating $\log(-\eta_2)' \sim N(\log(-\eta_2), s_2^2)$

(e) Propose η_1' by simulating $\log \eta_1' \sim N(\log \eta_1, s_1^2)$ such that $\log \eta_1' < 0$.

(f) Propose α_0' by simulating $\log \alpha_0' \sim N(\log \alpha_0, \sigma_0^2)$ such that $\log \alpha_0' > \log \frac{\eta_0'}{\eta_1'}$.

Step 2. Obtain the proposed $\boldsymbol{\tau}_n'$ by letting $\tau_j' = 0.5$. So the proposed intensity is the median intensity.

Step 3. Accept the proposed values with probability $\min\{AB, 1\}$, where A and B are given below.

Step 4. If $\boldsymbol{\alpha}'$ and $\boldsymbol{\tau}_n'$ are accepted, let $\boldsymbol{\alpha} = \boldsymbol{\alpha}'$, $\boldsymbol{\tau}_n = \boldsymbol{\tau}_n'$ and goto Step 1. Otherwise, discard $\boldsymbol{\alpha}'$ and $\boldsymbol{\tau}_n'$ and goto Step 1.

Calculation of the acceptance probability:

$$
\begin{aligned}
A &= \pi(\boldsymbol{\tau}_n', \boldsymbol{\alpha}' | \mathbf{t}_n) / \pi(\boldsymbol{\tau}_n, \boldsymbol{\alpha} | \mathbf{t}_n) \\
&= \frac{\prod_{j=1}^n \{\alpha_0' + \sum_{t_i < t_j} \alpha_1' e^{-\alpha_2'(t_j - t_i)} + \eta_0' Q(\tau_j', \eta_1', \eta_2')\}}{\prod_{j=1}^n \{\alpha_0 + \sum_{t_i < t_j} \alpha_1 e^{-\alpha_2(t_j - t_i)} + \eta_0 Q(\tau_j, \eta_1, \eta_2)\}} \times \frac{exp\Big\{-\Big[\alpha_0' T - (\alpha_1'/\alpha_2') \sum_{t_i < T} \{e^{-\alpha_2'(T - t_i)} - 1\}\Big]\Big\}}{exp\Big\{-\Big[\alpha_0 T - (\alpha_1/\alpha_2) \sum_{t_i < T} \{e^{-\alpha_2(T - t_i)} - 1\}\Big]\Big\}} \\
&\quad \times \frac{exp\{-\eta_0' \sum_{j=1}^n Q(\tau_j', \eta_1', \eta_2')(t_j - t_{j-1})\}}{exp\{-\eta_0 \sum_{j=1}^n Q(\tau_j, \eta_1, \eta_2)(t_j - t_{j-1})\}} \times \frac{\pi(\boldsymbol{\alpha}')}{\pi(\boldsymbol{\alpha})};
\end{aligned}
$$

$$
\begin{aligned}
\frac{\pi(\boldsymbol{\alpha}')}{\pi(\boldsymbol{\alpha})} &= \frac{\alpha_0 \alpha_1 \alpha_2 \eta_0 \eta_1 \eta_2}{\alpha_0' \alpha_1' \alpha_2' \eta_0' \eta_1' \eta_2'} \; e^{-(\log^2 \alpha_0' - \log^2 \alpha_0)/2u_0^2} \; e^{-(\log^2 \alpha_1' - \log^2 \alpha_1)/2u_1^2} \\
&\quad \times e^{-(\log^2 \alpha_2' - \log^2 \alpha_2)/2u_2^2} \; e^{-(\log^2 \eta_0' - \log^2 \eta_0)/2v_0^2} \; e^{-(\log^2 \eta_1' - \log^2 \eta_1)/2v_1^2} \\
&\quad \times e^{-\{\log^2(-\eta_2') - \log^2(-\eta_2)\}/2v_2^2} \; \frac{1 - \Phi[\{\log(\eta_0/\eta_1)\}/u_0]}{1 - \Phi\Big[\{\log(\eta_0'/\eta_1')\}/u_0\Big]} \; \frac{1 - \Phi\{(\log \alpha_1)/u_2\}}{1 - \Phi\{(\log \alpha_1')/u_2\}}.
\end{aligned}
$$

$$B = \frac{q(\boldsymbol{\alpha}' \to \boldsymbol{\alpha})}{q(\boldsymbol{\alpha} \to \boldsymbol{\alpha}')} \frac{q(\boldsymbol{\tau}'_n \to \boldsymbol{\tau}_n)}{q(\boldsymbol{\tau}_n \to \boldsymbol{\tau}'_n)}$$
$$= \frac{\alpha'_0 \alpha'_1 \alpha'_2 \eta'_0 \eta'_1 \eta'_2}{\alpha_0 \alpha_1 \alpha_2 \eta_0 \eta_1 \eta_2} \times \frac{1-\Phi\left\{\frac{\log(\eta'_0/\eta'_1)-\log\alpha_0}{\sigma_0}\right\}}{1-\Phi\left\{\frac{\log(\eta_0/\eta_1)-\log\alpha'_0}{\sigma_0}\right\}} \times \frac{\Phi\{-(\log\eta_1)/s_1\}}{\Phi\{-(\log\eta'_1)/s_1\}} \times \frac{1-\Phi\left\{\frac{\log\alpha'_1-\log\alpha_2}{\sigma_2}\right\}}{1-\Phi\left\{\frac{\log\alpha_1-\log\alpha'_2}{\sigma_2}\right\}},$$

since $q(\boldsymbol{\tau}'_n \to \boldsymbol{\tau}_n)/q(\boldsymbol{\tau}_n \to \boldsymbol{\tau}'_n) = 1$, where $q(u \to v)$ denotes the density function of v conditional on u.

Combining all the above results, we have

$$AB = \frac{\prod_{j=1}^n \{\alpha'_0 + \sum_{t_i<t_j} \alpha'_1 e^{-\alpha'_2(t_j-t_i)} + \eta'_0 Q(\tau'_j,\eta'_1,\eta'_2)\}}{\prod_{j=1}^n \{\alpha_0 + \sum_{t_i<t_j} \alpha_1 e^{-\alpha_2(t_j-t_i)} + \eta_0 Q(\tau_j,\eta_1,\eta_2)\}} \times \frac{e^{-\left[\alpha'_0 T - (\alpha'_1/\alpha'_2)\sum_{t_i<T}\left\{e^{-\alpha'_2(T-t_i)}-1\right\}\right]}}{e^{-\left[\alpha_0 T - (\alpha_1/\alpha_2)\sum_{t_i<T}\left\{e^{-\alpha_2(T-t_i)}-1\right\}\right]}}$$

$$\times \frac{e^{-\eta'_0 \sum_{j=1}^n Q(\tau'_j,\eta'_1,\eta'_2)(t_j-t_{j-1})}}{e^{-\eta_0 \sum_{j=1}^n Q(\tau_j,\eta_1,\eta_2)(t_j-t_{j-1})}} \times e^{-\frac{\log^2\alpha'_0-\log^2\alpha_0}{2u_0^2}} e^{-\frac{\log^2\alpha'_1-\log^2\alpha_1}{2u_1^2}}$$

$$\times e^{-\frac{\log^2\alpha'_2-\log^2\alpha_2}{2u_2^2}} e^{-\frac{\log^2\eta'_0-\log^2\eta_0}{2v_0^2}} e^{-\frac{\log^2\eta'_1-\log^2\eta_1}{2v_1^2}} e^{-\frac{\log^2(-\eta_2)'-\log^2(-\eta_2)}{2v_2^2}}$$

$$\times \frac{1-\Phi\left\{\frac{\log(\eta_0/\eta_1)}{u_0}\right\}}{1-\Phi\left\{\frac{\log(\eta'_0/\eta'_1)}{u_0}\right\}} \frac{1-\Phi\left\{\frac{\log\alpha_1}{u_2}\right\}}{1-\Phi\left\{\frac{\log\alpha'_1}{u_2}\right\}} \frac{1-\Phi\left\{\frac{\log(\eta'_0/\eta'_1)-\log\alpha_0}{\sigma_0}\right\}}{1-\Phi\left\{\frac{\log(\eta_0/\eta_1)-\log\alpha'_0}{\sigma_0}\right\}} \frac{\Phi\left\{-\frac{\log\eta_1}{s_1}\right\}}{\Phi\left\{-\frac{\log\eta'_1}{s_1}\right\}} \frac{1-\Phi\left\{\frac{\log\alpha'_1-\log\alpha_2}{\sigma_2}\right\}}{1-\Phi\left\{\frac{\log\alpha_1-\log\alpha_2}{\sigma'_2}\right\}}.$$

References

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

[2] Bacry, E and Muzy, J.F. (2014). Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance* 14, 1-20.

[3] Bauwens, L. and Hautsch, N. (2006). Stochastic conditional intensity processes. *Journal of Financial Econometrics* 4, 450-493.

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

[5] Brooks, S.P. (1998). Markov chain Monte Carlo method and its application. *The Statistician* 47, 69–100.

[6] Chavez-Demoulin, V. and McGill, J.A. (2012). High-frequency financial data modelling using HPes. *Journal of Banking & Finance* 36, 3415-3426.

[7] Cleveland, W.S., Grosse, E. and Shyu, W.M. (1992). Local regression models. Chapter 8 of *Statistical Models in S*. Eds. J.M. Chambers and T.J. Hastie, Wadsworth & Brooks/Cole.

[8] Corlu, C.G. and Corlu, A. (2015). Modelling exchange rate returns: which flexible distribution to use? *Quantitative Finance* 15, 1851-1864.

[9] Corrado, C.J. (2001). Option pricing based on the generalized lambda distribution. *Journal of Futures Markets* 21, 213-236.

[10] Daley, D.J., Vere-Jones, D. (2003). *An introduction to the theory of point processes*. 2nd Edition. Volume I. Springer: Heidelberg.

[11] Filimonov, V. and Sornette, D. (2012). Quantifying refexivity in Financial markets: towards a prediction of flash crashes. *Physical Review E*, 85:056108.

[12] Fournier, B., Rupin, N., Bigerelle, M., Najjar, D., Iost, A. & Wilcox, R. (2007). Estimating the parameters of a generalized lambda distribution. *Computational Statistics and Data Analysis* 51, 2813-2835.

[13] Gilchrist, W.G. (2000). *Statistical Modelling with Quantile Functions*. Chapman & Hall/CRC.

[14] Geyer, C.J. (2011) Introduction to Markov Chain Monte Carlo. In *Handbook of Markov chain Monte Carlo*. (eds S. Brooks, A. Gelman, G. K. Jones and Meng X. L.) FL: CRC Press, 3–48.

[15] Hawkes, A.G. (1971a). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58, 83-90.

[16] Hawkes, A.G. (1971b). Point spectra of some mutually exciting point processes. *Journal of Royal Statistical Society B* 33, 438-443.

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

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

[19] Lee, P.P. (2003). The generalized lambda distribution applied to spot exchange rates. *PhD Dissertation*. Carnegie Mellon University, 2003.

[20] Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*. Springer, New York, NY.

[21] Lorenzen, F. (2012). Analysis of order clustering using high frequency data: A point process approach. Swiss Federal Institute of Technology Zurich (ETH Zurich). `http://arno.uvt.nl/show.cgi?fid=126950`.

[22] Ogata, Y. (1981). On Lewis' simulation method for point processes. *Information Theory, IEEE Transactions on* 27, 23-31.

[23] O'Hagan, A. and Forster, J.J. (2004). *Bayesian Inference*. Arnold, London.

[24] Rasmussen, J.G. (2013). Bayesian inference for HPes. *Methodology and Computing in Applied Probability* 15, 623-642.

[25] Tarsitano, A. (2004). Fitting the generalized lambda distribution to income data. *COMPSTAT 2004-Proceedings in Computational Statistics*, 16th Symposium Held in Prague, Czech Republic, 2004.

[26] Veen, A. and Schoenberg, F.P. (2008) Estimation of space-time branching process models in seismology using an EM-type algorithm. *Journal of the American Statistical Association* 103, 614-624