Extremes (2019) 22:705–728
https://doi.org/10.1007/s10687-019-00352-4

Modeling extreme negative returns using marked renewal Hawkes processes

Tom Stindl[1] · **Feng Chen**[1]

Received: 15 March 2018 / Revised: 7 March 2019 / Accepted: 3 June 2019 /
Published online: 20 June 2019

Abstract
Extreme return financial time series are often challenging to model due to the presence of heavy temporal clustering of extremes and strong bursts of return volatility. One approach to model both these phenomena in extreme financial returns is the marked Hawkes self-exciting process. However, the Hawkes process restricts the arrival times of exogenously driven returns to follow a Poisson process and may fail to provide an adequate fit to data. In this work, we introduce a model for extreme financial returns, which provides added flexibility in the specification of the background arrival rate. Our model is a marked version of the recently proposed renewal Hawkes process, in which exogenously driven extreme returns arrive according to a renewal process rather than a Poisson process. We develop a procedure to evaluate the likelihood of the model, which can be optimized to obtain estimates of model parameters and their standard errors. We provide a method to assess the goodness-of-fit of the model based on the Rosenblatt residuals, as well as a procedure to simulate the model. We apply the proposed model to extreme negative returns for five stocks traded on the Australian Stock Exchange. The models identified for the stocks using in-sample data were found to be able to successfully forecast the out-of-sample risk measures such as the value at risk and provide a better quality of fit than the competing Hawkes model.

Keywords Goodness-of-fit · Marked point process · Finance · Forecast · Risk measures · Self-exciting

This research includes computations using the Linux computational cluster Katana supported by the Faculty of Science, UNSW Sydney. Stindl was supported by an Australian Government Research Training Program Scholarship. Chen was partly supported by a UNSW SFRGP grant.

✉ Tom Stindl
t.stindl@unsw.edu.au

Feng Chen
feng.chen@unsw.edu.au

[1] Department of Statistics, UNSW Sydney, Kensington, Australia

AMS 2000 Subject Classifications 62M09 · 60G70

1 Introduction

Modeling extreme financial returns has important applications, such as in the estimation of risk measures. However, like many other financial time series, the series of extreme returns are challenging to model due to the presence of heavy temporal clustering of extremes and strong bursts of return volatility. To address this challenge, Chavez-Demoulin et al. (2005) and Chavez-Demoulin and McGill (2012) proposed the marked Hawkes process model, and reported sufficient fits to extreme negative return data on a share price and on the Dow Jones Industrial Average index, while the traditional peaks over threshold (POT) model was shown not to be suitable for the data considered. The use of marked Hawkes processes for the purpose of forecasting market risk measures has also been applied in the work of McNeil et al. (2005, pp 306–311) and Herrera and Schipp (2009). Furthermore, Embrechts et al. (2011) considered multi-type event sequence data in which the multivariate version of the marked Hawkes process was used to model the interaction between positive and negative extreme returns for the Dow Jones Industrial Average index.

Extreme returns are fundamental to the risk management of financial institutions such as investment banks, insurers and pension funds as they are often required to demonstrate their financial stability under extreme market conditions. A useful measure of risk for extreme loss outcomes is given by the quantile of the loss distribution of a given asset or portfolio over a predefined period of time and this is known as the value at risk (VaR). Many approaches to estimate the unconditional VaR assume that the return distribution is normally distributed and then forecast volatility using the exponential-weighted moving average method as in Mina and Xiao (2001). Other unconditional approaches often rely on generalized autoregressive conditional heteroskedasticity (GARCH) models with either normal or t innovations.

In other approaches to estimate the VaR, the conditional return distribution, which takes into account the current financial environment in which the asset is traded, are often used. McNeil and Frey (2000) introduced a conditional approach using a two-stage procedure by combining GARCH models to forecast volatility and then applying techniques from extreme value theory (EVT) to the residuals from the GARCH analysis. Although this method circumvents the use of unconditional return distributions, it introduces a new problem that relates to the sensitivity of the EVT analysis on the GARCH model fit. Another conditional approach was developed by Chavez-Demoulin et al. (2005), in which they apply the POT model from EVT to the excesses (return above a given threshold), which are treated as independent and identically distributed (*i.i.d.*) observations and model the temporal patterns of exceedances (days when an excess occurs), using a marked Hawkes self-exciting process (Hawkes 1971) self-exciting process. This approach models the serial dependence present in returns and provides a convenient method to estimate the conditional VaR and other risk measures of interest, such as the expected shortfall (ES).

However, despite their success, marked Hawkes processes are not always able to provide an adequate fit to data. On these occasions, added flexibility in the specification of the background arrival rate may be required. For instance, the background arrival rate may be allowed to depend on some covariates, but this approach requires appropriate external covariates to be available. We propose that the marked renewal Hawkes (RHawkes) process model can provide this flexibility without the need to find suitable covariates. The RHawkes process is an adaptation of the classical Hawkes process, recently proposed by Wheatley et al. (2016), in which the Hawkes process is modified so that the immigrant arrival times are described by a renewal process rather than a Poisson process with a constant arrival rate. Chen and Stindl (2018) developed an algorithm to evaluate the likelihood of the RHawkes process, in quadratic time, which enables statistical inferences such as estimation, convenient to implement.

Unlike the marked Hawkes process in which the likelihood function is easy to evaluate using the direct likelihood formula for point process models (see e.g. Daley and Vere-Jones 2003, Proposition 7.2.III.), the likelihood function for the marked RHawkes process poses some additional computational challenges as the intensity function relative to the natural filtration is non-trivial to evaluate. To circumvent this difficulty, we develop a readily implementable recursive algorithm to evaluate the likelihood of the model in linear storage space and quadratic computational time, which can be optimized to obtain estimates of model parameters and their standard errors. We develop a procedure to assess the goodness-of-fit for both aspects of the model, the temporal patterns of exceedances and the distribution of excesses, by calculating the Rosenblatt residuals (see Rosenblatt 1952) and testing the residuals for uniformity and independence. As by-products of the direct likelihood evaluation algorithm, we obtain estimates of the two risk measures, conditional VaR and conditional ES. Furthermore, we provide methods to make predictions about future extreme negative returns, and in particular, the waiting time until the next exceedance and compare these predictions with actual observations.

In the next section, we introduce the dataset. Section 3 introduces the marked RHawkes process model, which includes its estimation and inferential methods such as goodness-of-fit assessment, prediction and estimation of risk measures. Numerical illustrations will follow in Section 4 with a simulation study. The focus of Section 5 is on applying the proposed methods to extreme negative returns for each of the five ASX stocks.

2 ASX stock data

This section introduces five commonly traded stocks on the ASX that are used in the analysis conducted in Section 5. The data was obtained from the Yahoo! Finance database and consists of the date, open, high, low and close price for the following stocks traded on the ASX from 1 January 2006 to 31 December 2016; JB Hi-Fi Limited (JBH), Adelaide Brighton Limited (ABC), Computershare Limited (CPU), Downer EDI Limited (DOW) and James Hardie Industries plc (JHX).

To assess the RHawkes model's performance at forecasting market risk and in particular estimating the conditional VaR, we shall perform backtesting. Therefore, the time period under consideration is split into two non-overlapping periods, which we call the in-sample and out-of-sample period. Data in the in-sample period are used to estimate the model parameters, and then the estimated model is used to make forecast of market risk measures during the out-of-sample period and the actual data in the out-of-sample period is used to assess the forecasted risk measures. The period from 1 January 2006 to 31 December 2015 is the in-sample period while the following year from 1 January 2016 to 31 December 2016 is used as the out-of-sample period. Descriptive statistics for the daily log-losses for the in-sample period are reported in Table 1, which contains the number of observations, minimum, maximum, mean, standard deviation and kurtosis. We observe that the kurtosis for all stocks are larger than three, which suggest that the return distributions are leptokurtic rather than normal.

Denote the percentage log-loss from day $t-1$ to day t by $r_t = -100 \times \log(s_t/s_{t-1})$ where s_t is the closing price. Our analysis is concerned with extreme negative returns which exceed a high threshold denoted by u. If the loss on day t exceeds the threshold value u, we say that an exceedance has occurred and provided that an exceedance has occurred, the excess is given by $w_t = r_t - u$. The choice of the threshold value u requires special attention. In our analysis, a threshold equal to the 90% quantile of the log-losses in the in-sample period was used for each stock, so that the 10% largest losses are considered as extreme negative returns. The choice of threshold value u is to some extent rather arbitrary, but we follow the convention used in the work of Chavez-Demoulin et al. (2005) and argue that using a lower threshold would question the validity of EVT while using a higher threshold would reduce the sample size considerably. With the current choice, the mean-excess plots (not shown) do not indicate that a violation of the assumptions is apparent (see e.g., Embrechts et al. 1997, p. 355). The threshold value u for all five stocks are shown in Table 2 as well as some descriptive statistics, including the number of exceedances, mean excess and median excess for the in-sample period.

Figure 1 visualizes the transformation from raw price data into exceedance data with threshold $u = 2.425$ for the stock JBH. The top panel shows a time series plot of the daily closing asset price s_t. The daily closing asset prices are then transformed to daily losses on the log scale r_t excluding weekends and non-trading weekdays aggregated together and the time series plot is shown in the middle panel. The plot

Table 1 Descriptive statistics for the in-sample daily log-losses in percent for each of the five ASX stocks

Stock	JBH	ABC	CPU	DOW	JHX
n	2528	2527	2528	2524	2528
Min	–16.03	–13.31	–14.51	–13.35	–20.15
Max	16.57	14.61	11.20	36.38	12.84
Mean	–0.0624	–0.0323	–0.0213	0.0260	–0.0262
Std. dev.	2.316	2.049	1.893	2.611	2.359
Kurtosis	8.125	7.430	8.011	30.212	8.320

Table 2 Descriptive statistics for the in-sample loss excesses for each of the five ASX stocks with a threshold value u chosen as the 90% quantile

Stock	JBH	ABC	CPU	DOW	JHX
Threshold	2.425	2.281	2.082	2.662	2.645
No. of exceedances	253	252	253	253	253
Mean excess	1.608	1.456	1.255	1.873	1.401
Median excess	1.074	0.898	0.743	0.935	0.887

shows clear signs of strong bursts of loss volatility. Next, we compute the excess above the threshold u given that the log-loss r_t exceeds the threshold. The bottom panel shows the times of exceedances and size of excesses w_t. The presence of heavy temporal clustering of extremes is evident and typically occur near periods of large losses.

3 Model and methodologies

Let the arrival times of exceedances be denoted by $\{\tau_i\}_{i\geq 1} \subset \mathbb{R}_{>0}$, $\tau_i < \tau_{i+1}$ and denote the associated excesses by $\{w_i\}_{i\geq 1}$. Let $N(t)$ be a simple point process on $\mathbb{R}_{>0}$ that counts the number of exceedances by time t. There are two types of exceedance events, namely exogenously and endogenously driven ones. The type is identified by a further (unobservable) mark $M_i \in \{0, 1\}$, where

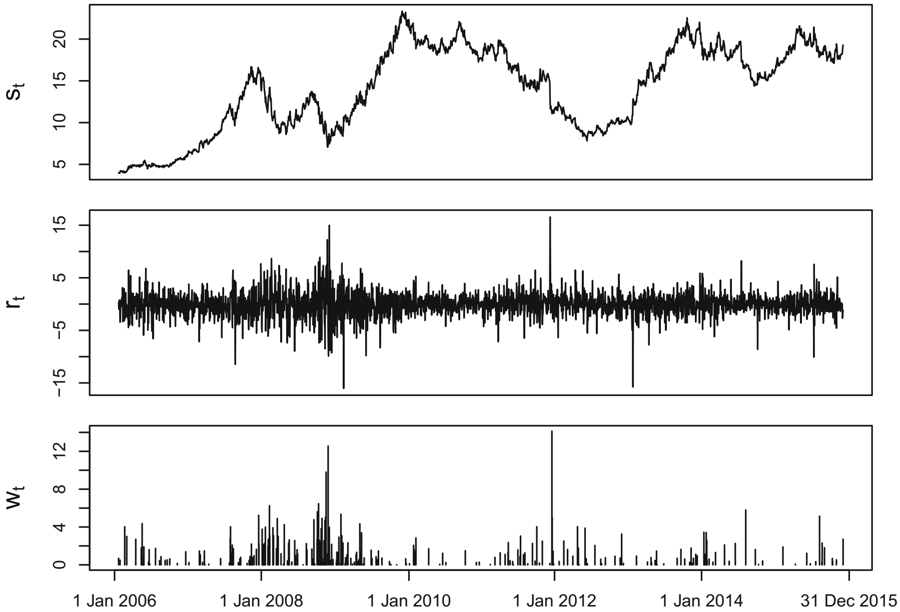

Fig. 1 JB Hi-Fi Limited (JBH) stock data from 1 January 2006 to 31 of December 2015. Top panel: time series plot of the daily closing stock price. Middle panel: time series plot of the negative daily log returns. Bottom panel: time of exceedances and size of excesses over the threshold $u = 2.425$

$M_i = 0$ indicates the arrival of an immigrant (exogenously driven exceedance) and $M_i = 1$ indicates an offspring event (endogenously driven exceedance). Furthermore, let $I(t) := \max\{i;\ \tau_i < t, M_i = 0\}$ denote the index of the most recent immigrant, with the convention that $I(t) := 0$ when $t < \tau_1$ and $\tau_0 := 0$.

To model the exceedance times and excesses we propose the marked RHawkes process, where the waiting times between successive immigrants are assumed to be *i.i.d.* so that the immigrants arrive according to a renewal process, and the exciting mechanism among the events is the same as in the classical marked Hawkes process model. That is, we assume the ground intensity process $\lambda(t), t \geq 0$ relative to the enlarged filtration $\tilde{\mathcal{F}}_t = \sigma\left\{N(s), w_{1:N(s)}, I(s); s \leq t\right\}, t \geq 0$ is given by

$$\begin{aligned}\lambda(t) = \frac{\mathbb{E}\left[\mathrm{d}N(t)|\tilde{\mathcal{F}}_{t-}\right]}{\mathrm{d}t} &= \mu(t - \tau_{I(t)}) + \sum_{j=1}^{N(t-)} \eta h(t - \tau_j) g(w_j) \\ &=: \mu(t - \tau_{I(t)}) + \phi(t).\end{aligned} \tag{1}$$

The function $\mu : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is the hazard rate function of the waiting times between successive exogenously driven exceedances (immigrants). For the stability of the process, we require that $\int_0^\infty e^{-\int_0^t \mu(s)\mathrm{d}s}\mathrm{d}t < \infty$, which ensures the expected waiting time between successive immigrants is finite. The mark's influence on the conditional intensity is governed by the impact function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$. The constant $\eta \geq 0$ is a normalizing constant, and for stability we require that $\eta\mathbb{E}\left[g(w_i)\right] < 1$ so that the expected number of children of an event is less than one. If the impact function is normalized so that $\mathbb{E}\left[g(w_i)\right] = 1$, the parameter $\eta \in [0, 1)$ has the interpretation of a branching ratio. The function $h : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is the offspring density function. The process $\phi(t)$ describes the total excitation effect of past events on current event intensity.

From the model specification (1), we can see that the background intensity depends on when the most recent immigrant arrives, and it resets to the function $\mu(\cdot)$ upon the arrival of an immigrant. When the hazard function $\mu(\cdot)$ is a constant, the waiting times are exponentially distributed and the immigrants arrive according to a Poisson process, and therefore the model reduces to the marked Hawkes process in Chavez-Demoulin et al. (2005) and Chavez-Demoulin and McGill (2012). However, in general, the marked RHawkes process is substantially more flexible than the marked Hawkes process because the event counts of the marked RHawkes process in regular time intervals can be over- or under-dispersed relative to the Poisson process, while the counts in a marked Hawkes process can only be over-dispersed.

The conditional intensity function in Eq. 1 is the time-intensity and only describes the dynamics of the ground process. It does not take into account the distribution of the marks. For full specification of the intensity process of the marked point process, we also need to specify the distribution of the event mark given an event happens at a certain time t and all the information before time t. In this work, we impose a conditional independence assumption, so that the mark w_i is independent of the event time τ_i conditional on the previous exceedance times $\tau_{1:i-1} := (\tau_1, \ldots, \tau_{i-1})$ and excesses $w_{1:i-1} := (w_1, \ldots, w_{i-1})$. This conditional independence assumption makes parametric methods for modeling marks simple to implement. In this instance, optimization of the log-likelihood function can be split into two separate optimization problems, and therefore the maximum likelihood estimators (MLEs) for parameters in the ground process model and in the mark distribution can be obtained separately.

3.1 Likelihood evaluation algorithm

This section develops an algorithm to compute the likelihood function of the marked RHawkes process. The likelihood function can be expressed as a product of conditional joint densities of the event time and mark, conditional on all previous times and marks as follows

$$L(\theta|\tau_{1:n}, w_{1:n}) = p_\theta(\tau_1, w_1)\left\{\prod_{i=2}^{n} p_\theta(\tau_i, w_i|\tau_{1:i-1}, w_{1:i-1})\right\}\mathbb{P}_\theta\left(\tau_{n+1} > T|\tau_{1:n}, w_{1:n}\right), \quad (2)$$

where T is the censoring time and $\mathbb{P}_\theta(\tau_{n+1} > T|\tau_{1:n}, w_{1:n})$ is the probability that no event occurs in the interval $(\tau_n, T]$. In what follows, we drop the subscript θ in p_θ and $\mathbb{P}_\theta$ for notational convenience, while the dependence of the relevant densities and probabilities on the parameter θ is silently understood. The conditional independence assumption allows the log-likelihood function in Eq. 2 to be split into two separate components and inferences can be performed separately for the temporal patterns of exceedances and the loss excesses. The log-likelihood function is then given by

$$\begin{aligned} l(\theta|\tau_{1:n}, w_{1:n}) = &\left[\log p(\tau_1) + \sum_{i=2}^{n}\log p(\tau_i|\tau_{1:i-1}, w_{1:i-1}) + \log\mathbb{P}\left(\tau_{n+1} > T|\tau_{1:n}, w_{1:n}\right)\right] \quad (3)\\ &+\left[\log p(w_1) + \sum_{i=2}^{n}\log p(w_i|\tau_{1:i-1}, w_{1:i-1})\right]\\ &=: l_\tau + l_w, \quad (4)\end{aligned}$$

with l_τ and l_w denoting the temporal component and the mark component of the log-likelihood respectively. We will treat the modeling of the marks and the evaluation of l_w in the next section. The remainder of this section is devoted to the evaluation of l_τ.

The ground intensity function $\lambda(t)$ depends on the most recent immigrant arrival time and so to compute the conditional densities required in Eq. 3, the distribution of the most recent immigrant is needed. If we define the following terms

$$d_{ij} := p\left(\tau_i|\tau_{1:i-1}, w_{1:i-1}, I(\tau_i) = j\right), \quad (5)$$

$$S_{n+1,j} := \mathbb{P}\left(\tau_{n+1} > T|\tau_{1:n}, w_{1:n}, I(\tau_{n+1}) = j\right), \quad (6)$$

$$p_{ij} := \mathbb{P}\left(I(\tau_i) = j|\tau_{1:i-1}, w_{1:i-1}\right), \quad (7)$$

and by conditioning on the index of the most recent immigrant, we obtain the following

$$p\left(\tau_i|\tau_{1:i-1}, w_{1:i-1}\right) = \sum_{j=1}^{i-1} d_{ij}p_{ij}, \qquad i = 2, \ldots, n, \quad (8)$$

$$\mathbb{P}\left(\tau_{n+1} > T|\tau_{1:n}, w_{1:n}\right) = \sum_{j=1}^{n} S_{n+1,j}\,p_{n+1,j}. \quad (9)$$

We introduce some convenient notation to obtain computable expressions for Eqs. 8 and 9. Let $U(t) = \int_0^t \mu(s)\mathrm{d}s$ be the cumulative immigrant hazard function, $H(t) = \int_0^t h(s)\mathrm{d}s$ be the offspring distribution function and $\Phi(t) = \int_0^t \phi(s)\mathrm{d}s = \eta \sum_{j=1}^{N(t-)} H(t-\tau_j)g(w_j)$. By construction of the process, the first event is an immigrant with event time density $p(\tau_1) = e^{-U(\tau_1)}\mu(\tau_1)$. The conditional densities and survival probabilities in Eqs. 5 and 6 are computed using

$$d_{ij} = e^{-\{U(\tau_i-\tau_j)-U(\tau_{i-1}-\tau_j)\}-\{\Phi(\tau_i)-\Phi(\tau_{i-1})\}}\left(\mu(\tau_i-\tau_j)+\phi(\tau_i)\right), \quad (10)$$

$$S_{n+1,j} = e^{-\{U(T-\tau_j)-U(\tau_n-\tau_j)\}-\{\Phi(T)-\Phi(\tau_n)\}}, \quad (11)$$

see also Eq. 6 and Eq. 7 in Chen and Stindl (2018). Next, the conditional probabilities p_{ij} in Eq. 7 are computed using the following forward recursion with initial conditions $p_{21} = 1$ and $p(\tau_2|\tau_1, w_1) = d_{21}$,

$$p_{ij} = \begin{cases} \frac{\phi(\tau_{i-1})}{\mu(\tau_{i-1}-\tau_j)+\phi(\tau_{i-1})} \frac{d_{i-1,j}\,p_{i-1,j}}{p(\tau_{i-1}|\tau_{1:i-2}, w_{1:i-2})}, & j = 1, \dots, i-2, \\ 1 - \sum_{k=1}^{i-2} p_{ik}, & j = i-1, \end{cases} \quad (12)$$

for $i = 3, \dots, n+1$. The derivation of Eq. 12 is similar to the derivation of Eq. 8 found in the supplementary materials of Chen and Stindl (2018).

Direct evaluation of the likelihood is now possible given some parameter values. To evaluate the conditional densities $p(\tau_i|\tau_{1:i-1}, w_{1:i-1})$ and the most recent immigrant probabilities p_{ij}, we implement the bivariate recursion given in Eqs. 8 and 12 together with the d_{ij} given by Eq. 10. The survival probability $\mathbb{P}(\tau_{n+1} > T|\tau_{1:n}, w_{1:n})$ is computed using Eqs. 9, 11, and the $p_{n+1,j}$. The above terms are then substituted into the first pair of square brackets in Eq. 3, to calculate the part of the log-likelihood needed for the estimation of parameters of the ground process model, that is, l_τ.

3.2 Excess modeling

The generalized Pareto distribution (GPD) is frequently used in EVT and has been applied to model excesses of extreme negative returns (e.g. Chavez-Demoulin et al. 2005), and shall also be used in our work. The use of GPD is justified by the following result from the EVT (cf. Embrechts et al. 1997, Theorem 3.4.5): If the random variable X has a distribution function $F(\cdot)$ belonging to the maximum domain of attraction of the standard generalized extreme value distribution $H_\xi(\cdot)$,

$$H_\xi(x) = \begin{cases} 1 - \exp\left\{-(1+\xi x)^{-1/\xi}\right\}, & \xi \neq 0; \\ 1 - \exp\left\{-e^{-x}\right\}, & \xi = 0, \end{cases}$$

then there exists a positive function $a(\cdot)$ such that

$$\mathbb{P}(X - u \le x | X > u) \to G_{\xi,a(u)}(x), \quad \text{as } u \to x_F,$$

where $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$, and $G_{\xi,\sigma}$ is the *generalized Pareto distribution* function with shape parameter ξ and scale parameter σ, defined by

$$G_{\xi,\sigma}(x) = \begin{cases} 1 - (1+\xi x/\sigma)_+^{-1/\xi}, & \xi \neq 0; \\ 1 - \exp(-x/\sigma), & \xi = 0. \end{cases}$$

Following Chavez-Demoulin et al. (2005), we model the loss excesses using generalized Pareto distributions with a common shape parameter and scale parameters evolving according to a first order Markov process. More specifically,

$$w_i|w_{1:i-1} \sim G_{\xi,a+bw_{i-1}}, \tag{13}$$

with parameters $\xi > 0$, $a > 0$ and $b > 0$. For stability we also require that $\xi + b < 1$. It can then be shown using Theorem 3.2 in Cline and Pu (2002) that the Markov process (13) is geometrically ergodic. These parameters can be estimated by maximizing the part of the log-likelihood in the second pair of brackets in Eq. 3, that is,

$$l_w = \sum_{i=1}^{n} \left\{ -(\frac{1}{\xi} + 1) \log \left(1 + \frac{\xi w_i}{a + bw_{i-1}} \right) - \log(a + bw_{i-1}) \right\}, \tag{14}$$

where we set $w_0 := a/(1 - \xi - b)$ to be the mean of the excesses.

3.3 Goodness-of-fit assessment

Two aspects of the model should be considered in the goodness-of-fit assessment; the temporal patterns of exceedances and the distribution of excesses. For the former, we apply a similar procedure to Chen and Stindl (2018) using the Rosenblatt (1952) residuals. The basis of the method is to transform the exceedance times using the Rosenblatt (1952) transformation to produce residuals which should be independent and uniformly distributed on the unit interval when the model is correctly specified. The residuals are given by $U_1 = \hat{F}_1(\tau_1) = 1 - e^{-\hat{U}(\tau_1)}$ and

$$U_i = \hat{F}_i(\tau_i|\tau_{1:i-1}, w_{1:i-1}) = 1 - \sum_{j=1}^{i-1} \hat{p}_{ij} \hat{S}_{ij}, \quad i = 2, \ldots, n, \tag{15}$$

where $\hat{F}_i(t|\tau_{1:i-1}, w_{1:i-1})$ is the estimated conditional distribution function of τ_i conditional on $\tau_{1:i-1}$ and $w_{1:i-1}$, $\hat{p}_{ij}$ are the estimated most recent immigrant probabilities in Eq. 12 and $\hat{S}_{ij}$ are given by

$$\hat{S}_{ij} = e^{-\left\{\hat{U}(\tau_i-\tau_j)-\hat{U}(\tau_{i-1}-\tau_j)\right\}-\left\{\hat{\Phi}(\tau_i)-\hat{\Phi}(\tau_{i-1})\right\}}, \quad j = 1, \ldots, i-1.$$

Note that $\hat{U}(t)$ and $\hat{\Phi}(t)$ are the plug-in estimates of the cumulative hazard function $U(t)$ and cumulative excitation effect $\Phi(t)$ defined on Page 7 above (10).

For a consistent approach to the goodness-of-fit assessment of the model, we also apply the Rosenblatt (1952) transformation to the excesses. We only consider the case of the GPD model in Eq. 13, although the procedure is general enough to apply to most choices of marked distributions. In this instance, the residuals are given by

$$V_i = \hat{G}_i(w_i|\tau_{1:i-1}, w_{1:i-1}) = 1 - \left[1 + \frac{\hat{\xi} w_i}{\hat{a} + \hat{b} w_{i-1}} \right]^{-1/\hat{\xi}} \tag{16}$$

where $\hat{G}_i(w|\tau_{1:i-1}, w_{1:i-1})$ is the estimated conditional distribution function of w_i conditional on $\tau_{1:i-1}$ and $w_{1:i-1}$. However, for the GPD model considered in this paper, the conditional distribution function $\hat{G}_i(\cdot|\tau_{1:i-1}, w_{1:i-1})$ only depends on w_{i-1}.

The two residual series $\{U_i\}$ and $\{V_i\}$ then serve as the basis for assessing the model's ability to model the temporal patterns of exceedances and the distribution of excesses. Both residual series should be approximately *i.i.d.* uniformly on (0,1) if both aspects of the model are adequate. To check the uniformity and independence, we can use graphical techniques such as uniform Quantile-Quantile (Q-Q) plot and the Autocorrelation function (ACF) plot, or formal statistical tests, such as the Kolmogorov-Smirnov (K-S) test and the Ljung-Box (L-B) test respectively.

3.4 Predicting exceedances

Predictions using point process models often rely on simulations as explicit algorithms are generally not available (Daley and Vere-Jones 2003, pp. 274). The distribution of quantities of interests such as the time until the next exceedance or the number of exceedances in a given time interval can be extracted from predictive simulations. The algorithm works by sequentially simulating event times until the censoring time, as in Section 4.1. This procedure requires a computable expression for the hazard function (or cumulative hazard function) and so the index of the most recent immigrant before time T needs to be simulated using the conditional probabilities $p_{n+1,j} = \mathbb{P}(I(\tau_{n+1}) = j|\tau_{1:n}, w_{1:n})$ so that $\mu(t - \tau_{I(T)})$ can be computed for all $t > T$. For each realization of the future we can extract any quantity of interest that we want to predict, and use its empirical distribution obtained from a large number of realizations as the basis for prediction.

A particularly important prediction for financial stakeholders is the time until the next extreme loss. For this purpose, predictive simulations are not necessary, as we can directly compute the predictive density and hazard function for the waiting time until the next exceedance after the censoring time, which are given respectively by

$$p(\tau_{n+1}|\tau_{1:n}, w_{1:n}, \tau_{n+1} > T) = \frac{\sum_{j=1}^{n} p_{n+1,j} d_{n+1,j}}{\mathbb{P}(\tau_{n+1} > T|\tau_{1:n}, w_{1:n})}, \qquad \tau_{n+1} > T, \tag{17}$$

and

$$\text{haz}(\tau_{n+1}|\tau_{1:n}, w_{1:n}, \tau_{n+1} > T) = \frac{\sum_{j=1}^{n} p_{n+1,j} d_{n+1,j}}{\sum_{j=1}^{n} p_{n+1,j} \tilde{S}_{n+1,j}}, \qquad \tau_{n+1} > T, \tag{18}$$

where the $p_{n+1,j}$'s are calculated using (12), the denominator in Eq. 17 is computed using (9), the $d_{n+1,j}$'s are given as in Eq. 10 and

$$\tilde{S}_{n+1,j} = e^{-\{U(\tau_{n+1}-\tau_j)-U(\tau_n-\tau_j)\}-\{\Phi(\tau_{n+1})-\Phi(\tau_n)\}}.$$

The estimated parameters can then be substituted into Eqs. 17 and 18 to obtain the estimated predictive density and hazard function which is useful in making predictions about the time until the next exceedance.

3.5 Forecasting conditional risk measures

Conditional risk measures such as VaR and ES are important quantities used by many financial institutions and as such, a method to estimate their value is of particular importance. The proposed algorithm to evaluate the likelihood of the marked RHawkes process implies a procedure to compute the predictive distribution of loss excesses conditional on the history of the process by time t, $\mathcal{F}_t = \sigma\left\{N(t), \tau_{1:N(t)}, w_{1:N(t)}\right\}$, and therefore estimates of conditional VaR and ES can be easily obtained. For the remainder of this article, we will refer to the conditional VaR and conditional ES as VaR and ES, observing that these quantities are conditioned upon all priorly observed exceedance times and loss excesses. To forecast the VaR and ES for the out-of-sample period, the MLEs obtained from the in-sample period are used to obtain the plug-in predictive loss distribution.

Denote R_{t+1} as the daily log-loss on day $t+1$, then the VaR at level q on day $t+1$ is given by

$$\mathrm{VaR}_{t+1}^{q} = \inf\left\{r \in \mathbb{R} : \; F_{R_{t+1}|\mathcal{F}_t}(r) \geq q\right\},$$

which takes into account the observed data up to day t. In financial applications we are often concerned with extreme outcomes and so we typically take quantile levels $q = 0.95$ or $q = 0.99$. Now by conditioning on the index of the most recent immigrant before time $t+1$, the survival function for the daily log-loss R_{t+1} becomes

$$\mathbb{P}\left(R_{t+1} > r|\mathcal{F}_t\right) = \sum_{k=1}^{N(t)} \mathbb{P}\left(R_{t+1} > r|\mathcal{F}_t,\; I(t) = k\right)\mathbb{P}\left(I(t) = k|\mathcal{F}_t\right). \tag{19}$$

As we are only concerned with extreme returns, we condition upon the return being greater than the threshold u, and so for $r > u$ we obtain

$$\begin{aligned}\mathbb{P}\left(R_{t+1} > r|\mathcal{F}_t,\; I(t) = k\right) &= \mathbb{P}\left(R_{t+1} - u > r - u|R_{t+1} > u, \mathcal{F}_t, I(t) = k\right)\\ &\quad \times \mathbb{P}\left(R_{t+1} > u|\mathcal{F}_t, I(t) = k\right). \end{aligned} \tag{20}$$

To compute (20) we approximate a discrete time process using a continuous time process. The second term on the right of Eq. 20 can be approximated using the probability that at least one exceedance event occurs in the interval $(t, t+1]$ conditional on index k being the most recent immigrant. This type of approximation has previously been applied in the work of Chavez-Demoulin et al. (2005) and Chavez-Demoulin and McGill (2012). In this context, the approximation is given by

$$\mathbb{P}\left(R_{t+1} > u|\mathcal{F}_t, I(t) = k\right) \approx 1 - \exp\left(-\int_t^{t+1} \hat{\mu}(s - \tau_k) + \hat{\phi}(s)\mathrm{d}s\right). \tag{21}$$

The first term on the right hand side of Eq. 20 is computed using the fitted GPD for the excesses,

$$R_{t+1} - u|\mathcal{F}_t;\; R_{t+1} > u \sim G_{\hat{\xi}, \hat{a}+\hat{b}w_{N(t)}}.$$

A forecast of the VaR at level q can be obtained by solving

$$\mathbb{P}\left(R_{t+1} > \mathrm{VaR}_{t+1}^{q}|\mathcal{F}_t\right) = 1 - q,$$

and is given by

$$\widehat{\mathrm{VaR}}^{q}_{t+1} = \frac{\hat{a} + \hat{b} w_{N(t)}}{\hat{\xi}} \left[\left(\frac{C_{t+1}}{1-q} \right)^{\hat{\xi}} - 1 \right] + u, \tag{22}$$

where

$$C_{t+1} = \sum_{k=1}^{N(t)} \left[1 - \exp\left\{ - \int_{t}^{t+1} \hat{\mu}(s - \tau_k) + \hat{\phi}(s)\mathrm{d}s \right\} \right] \hat{p}_{N(t)+1,k},$$

is an approximation to the probability that the return on day $t+1$ is greater than the threshold u, i.e. $\mathbb{P}(R_t > u|\mathcal{F}_t)$. The expression in Eq. 22 is only valid when $C_{t+1}/(1-q) > 1$, or more elegantly put $q \geq \mathbb{P}(R_{t+1} < u|\mathcal{F}_t)$. When this does not hold, we apply a conservative approach and define our VaR estimate to equal the threshold value u.

Although the VaR is a very useful tool for measuring risk, it does not give an indication of the size of an extreme loss. This deficiency has led to alternative risk measures being considered. One such alternative is the ES, which is an attractive risk measure as it provides a measure of the size of the loss given that it exceeds the VaR level. The conditional ES for day $t+1$ at quantile level q is defined by

$$\mathrm{ES}^{q}_{t+1} = \frac{\int_q^1 \mathrm{VaR}^{\alpha}_{t+1} \mathrm{d}\alpha}{1-q}.$$

Based on this definition, a forecast of the conditional ES on day $t+1$ is given by (cf. Chavez-Demoulin and McGill 2012):

$$\widehat{\mathrm{ES}}^{q}_{t+1} = \frac{\mathrm{V\hat{a}R}^{q}_{t+1}}{1-\hat{\xi}} + \frac{\hat{a} + \hat{b} w_{N(t)} - u\hat{\xi}}{1-\hat{\xi}}. \tag{23}$$

4 Simulation study

4.1 Simulation algorithm

The algorithm to simulate the process requires a sequential approach as the event marks might be autocorrelated. The algorithm works as follows. First, simulate the initial immigrant arrival time according to the specified inter-renewal distribution. Then each event is simulated by first simulating the corresponding waiting time since the last event according to an appropriate hazard function, and then simulating the event type (immigrant or offspring) according to an appropriate Bernoulli distribution, and finally simulating the event mark according to previously simulated events marks (and event times, depending on model specification) and the specified dependence structure. Events are simulated sequentially until the next simulated event time is beyond the censoring time. The realization of the marked RHawkes process consists of the time-mark pairs corresponding to all the simulated events by the censoring time. Note that the event types are not retained.

4.2 Simulation model

The rest of this section reports numerical evidence of the finite sample performance of the MLEs in a simulation study. The models chosen to perform the simulations are motivated by the model choices in Section 5 and consist of gamma inter-renewal waiting times with hazard function

$$\mu(t) = \frac{1}{\Gamma(t/\beta, \kappa)\beta^{\kappa}} t^{\kappa-1} e^{-t/\beta}, \tag{24}$$

where κ is the shape parameter, β is the scale parameter and $\Gamma(x, k) = \int_x^{\infty} s^{k-1} e^{-s} \mathrm{d}s$ is the upper incomplete gamma function. The offspring density is exponential $h(t) = e^{-t/\gamma}/\gamma$ with mean waiting time parameter γ. The event marks are conditionally generalised Pareto distributed, and follows the first order Markov process (13) with parameters ξ, a and b. The impact function $g(w)$ is the normalized version of the affine function $1 + \delta w$, that is,

$$g(w) = \frac{1 + \delta w}{\mathbb{E}\left[1 + \delta w_i\right]} = \frac{1 + \delta w}{1 + \delta a/(1 - \xi - b)}. \tag{25}$$

The simulations consist of 1000 realizations of the marked RHawkes process up to a predetermined censoring time T for a range of parameter values specified in Table 3. The censoring time T is chosen so that the expected numbers of events by T are approximately 500 and 1000 respectively. For each realization, the parameters of the mark distribution ξ, a and b are estimated by directly minimizing $-l_w$, and the parameters of the RHawkes process κ, β, γ, δ and η are estimated by directly minimizing $-l_{\tau}$.

4.3 Results

All computations were performed on Intel Xeon X5675 processors (12M cache, 3.06 GHz, 6.4GT/S QPI) using the R language (R Core Team 2017). Likelihood maximization was done by direct calls to the R function `optim`. As the log-likelihood function is rather flat along the parameters ξ and δ, we have used the reparametrization $\theta = e^{\theta'}$ to improve convergence speed and estimation accuracy. Table 3 presents the estimation results, which contains the true value for each parameter (True), the mean of the 1000 parameter estimates (Est), the empirical standard error of each estimator (SE), i.e. the standard deviation of the 1000 estimates, the average of the 1000 standard error estimates by inverting the approximate Hessian matrix ($\widehat{\text{SE}}$), the mean squared error (MSE), the censoring time (T), the mean number of events (ML) and the mean running time (RT) to perform the optimization procedure and compute the Hessian matrix. Due to the heavy-tailed nature of the generalized Pareto distribution, occasionally extremely large mark values occur. The finite and somewhat small sample size allows these extreme mark values to have a substantial influence on some parameter estimates, especially on the estimates of δ. Therefore in summarizing the estimation results in Table 3, we have trimmed the extreme estimates by removing a percentage (2.5% when the mean number of events is 500 and 1% when the mean number of events is 1000) of the smallest and largest estimates for each parameter.

Table 3 Results of the maximum likelihood estimation based on 1000 simulated data sets of the marked RHawkes process with gamma distributed inter-immigration waiting times, exponential offspring densities, normalized affine linear impact function, and event marks following the one-step Markov model in Eq. 13

	Immigration		Offspring		Mark			Impact
	κ	β	γ	η	ξ	a	b	δ
True	2	0.5	1	0.3	0.1	1	0.1	0.1
Est.	2.039	0.495	1.138	0.289	0.094	1.009	0.098	0.366
SE	0.328	0.059	0.627	0.069	0.044	0.076	0.045	2.434
$\hat{\text{SE}}$	0.359	0.067	0.588	0.074	0.050	0.087	0.049	4.695
MSE.	0.109	0.004	0.413	0.005	0.002	0.006	0.002	5.996
	RT = 325.89 secs	$T = 350$	$ML = 498.71$					
Est.	2.023	0.500	1.081	0.296	0.097	1.005	0.099	0.146
SE	0.241	0.048	0.440	0.051	0.034	0.058	0.033	0.162
$\hat{\text{SE}}$	0.254	0.049	0.509	0.055	0.035	0.061	0.035	0.195
MSE.	0.058	0.002	0.200	0.003	0.001	0.003	0.001	0.029
	RT = 1031.71 secs	$T = 700$	$ML = 996.75$					
True	2	0.5	1	0.7	0.1	1	0.1	0.1
Est.	2.263	0.475	1.030	0.674	0.092	1.005	0.097	0.155
SE	1.045	0.151	0.260	0.072	0.043	0.074	0.041	0.170
$\hat{\text{SE}}$	0.899	0.155	0.278	0.074	0.050	0.087	0.050	0.204
MSE.	1.161	0.024	0.068	0.006	0.002	0.006	0.002	0.032
	RT = 330.55 secs	$T = 150$	$ML = 494.77$					
Est.	2.099	0.496	1.008	0.683	0.098	1.003	0.099	0.123
SE	0.689	0.124	0.183	0.054	0.032	0.057	0.034	0.109
$\hat{\text{SE}}$	0.618	0.122	0.189	0.053	0.035	0.061	0.035	0.117
MSE.	0.484	0.015	0.034	0.003	0.001	0.003	0.001	0.012
	RT = 1055.37 secs	$T = 300$	$ML = 990.62$					
True	0.5	2	1	0.3	0.1	1	0.1	0.1
Est.	0.506	2.014	1.026	0.299	0.094	1.011	0.097	0.259
SE	0.038	0.298	0.352	0.061	0.043	0.073	0.041	0.438
$\hat{\text{SE}}$	0.042	0.337	0.378	0.070	0.050	0.087	0.049	0.599
MSE.	0.001	0.089	0.124	0.004	0.002	0.005	0.002	0.217
	RT = 369.06 secs	$T = 350$	$ML = 500.03$					
Est.	0.503	2.007	1.016	0.298	0.096	1.005	0.098	0.184
SE	0.029	0.219	0.253	0.047	0.033	0.058	0.033	0.246
$\hat{\text{SE}}$	0.029	0.234	0.268	0.049	0.035	0.061	0.035	0.280
MSE.	0.001	0.048	0.064	0.002	0.001	0.003	0.001	0.068
	RT = 1227.40 secs	$T = 700$	$ML = 998.89$					
True	0.5	2	1	0.7	0.1	1	0.1	0.1
Est.	0.533	1.858	0.997	0.671	0.093	1.014	0.095	0.157
SE	0.076	0.503	0.200	0.065	0.043	0.075	0.044	0.179

Table 3 (continued)

	Immigration		Offspring		Mark			Impact
	κ	β	γ	η	ξ	a	b	δ
$\hat{\text{SE}}$	0.083	0.578	0.223	0.074	0.051	0.089	0.050	0.230
MSE.	0.007	0.274	0.040	0.005	0.002	0.006	0.002	0.035
	RT = 348.17 secs $T = 150$ $ML = 490.36$							
Est.	0.516	1.941	1.001	0.687	0.095	1.006	0.099	0.129
SE	0.053	0.409	0.143	0.048	0.032	0.058	0.034	0.123
$\hat{\text{SE}}$	0.056	0.434	0.151	0.051	0.035	0.061	0.035	0.138
MSE.	0.003	0.171	0.021	0.002	0.001	0.003	0.001	0.016
	RT = 1140.52 secs $T = 300$ $ML = 996.05$							

In most cases the bias and standard errors are decreasing as the censoring time increases. The standard errors are also decreasing and approximately at a rate of $1/\sqrt{T}$ as the standard errors reduce by a factor of approximately $1/\sqrt{2}$ when the censoring time T is doubled. The mean of the standard errors are fairly comparable with the empirical standard errors as they agree in most cases and again this improves as the censoring time increases. For the majority of the parameters, the MSE is fairly close to zero with the only notable exceptions being γ and δ. However, the MSE drops quite substantially as we increase the censoring time, e.g. the MSE for δ decreases from 5.996 to 0.029 in the first simulation model considered while most decrease by a factor of two.

The estimated mean waiting time $\hat{\gamma}$ between an event and its direct offspring events and the estimated parameter $\hat{\delta}$ in the impact function tend to have rather large standard errors when compared to the standard errors of the other parameter estimates. However, the standard errors of $\hat{\gamma}$ and $\hat{\delta}$ are still shrinking as the censoring time gets larger. The relatively large standard errors for the mean offspring waiting time is also evident in the simulation study conducted by Chen and Stindl (2018) for the RHawkes process. The estimates for γ and impact function parameter δ have significantly less bias when the immigrant arrivals exhibit heavy clustering ($\kappa = 1/2$) compared to when the immigrants exhibit more evenly distributed arrival times ($\kappa = 2$). The parameter γ is well estimated when the branching ratio η is large, i.e. when the level of self-excitation is high, as the expected number of offspring events present in those realizations is larger. The large bias evident in the estimation of the parameter δ relates to the heavy-tailed nature inherent in the generalised Pareto distribution and a rather large sample would be required to reduce the bias to a reasonable level. The branching ratio parameter η is generally well estimated with little to no bias for the range of censoring times, level of self-excitation and variability of immigrant inter-event waiting times. This leads us to draw the conclusion that the MLE obtained by directly maximizing the log-likelihood function has satisfactory finite sample performances.

In Fig. 2 we present the normal Q-Q plots of the estimates of each parameter for the last simulation model considered in Table 3. For most parameters, the standard normal quantiles and the empirical quantiles of the estimator align fairly well. However, as the true value 0.1 for δ is very close to the lower bound 0, and because of the large variance of the estimator of δ, the empirical distribution for the estimator of δ is heavily skewed to the right, causing severe deviation of the quantile points from the Q-Q line in the Q-Q plot for $\hat{\delta}$.

5 Modeling extreme negative returns

In this section we demonstrate the versatility of the marked RHawkes process for modeling extreme negative returns for five stocks traded on the ASX introduced in Section 2. The marked RHawkes process model was fitted to the data with three different choices of inter-renewal distributions; exponential, gamma and Weibull. Recall that the exponential inter-renewal distribution is equivalent to the competitor approach based on the classical marked Hawkes process for which the marked RHawkes process will be backtested against. The gamma model has an inter-renewal hazard function given by Eq. 24 and the Weibull model with shape parameter κ and scale parameter β has hazard function given by

$$\mu(t) = \frac{\kappa}{\beta^{\kappa}} t^{\kappa-1}, \quad t \geq 0. \tag{26}$$

For all models, we use a normalized affine impact function as in Eq. 25, where δ reflects the strength of the excesses on the ground intensity process. The offspring

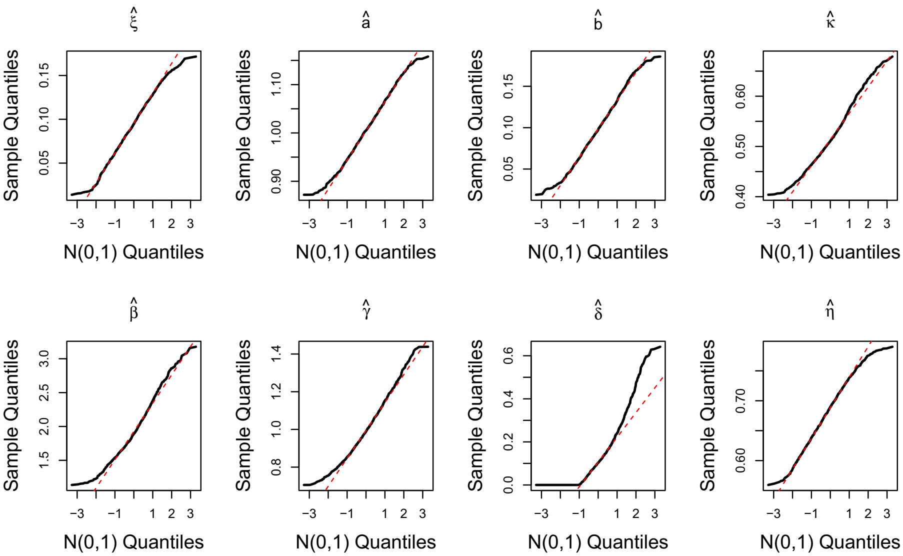

Fig. 2 Normal quantiles plots of the estimates of each parameter for the case where $\kappa = 0.5$, $\beta = 2$, $\gamma = 1$, $\delta = 0.1$ and $\eta = 0.7$ with a mean numbers of events 1000 approximately

density is chosen to be exponential $h(t) = e^{-t/\gamma}/\gamma$ where γ is the mean waiting time between an exceedance event and any exceedance events directly excited by it. The estimated parameters for each model on the ASX stocks were found by directly optimizing the log-likelihoods and the standard errors by inverting the Hessian matrix. The Rosenblatt residuals for goodness-of-fit assessment were also calculated.

To model the loss excesses, following Chavez-Demoulin et al. (2005) we assume the loss excesses follow the order-1 Markov process model (13). We optimized the contribution to the likelihood in Eq. 3 based only on the excesses. The results are reported in Table 4, which contains the estimates, the standard errors (in parentheses), and the p-values of the K–S tests and L–B tests on the Rosenblatt residuals. The large p-values suggest the model for the excesses is adequate for most of the stocks, except in the case of the JHX and CPU stocks, where there is still significant serial correlation among the residuals. A higher order Markov process might be considered for these stocks. The estimated parameter $\hat{b}$ is positive for all the stocks, suggesting that an excessively large loss is likely to be followed by another large loss, although the result is only significant for the stocks JBH and CPU. The estimated shape parameter $\hat{\xi}$ is positive for all the stocks, although it is statistically significant only for the stock DOW, suggesting the loss excesses on this stock is heavier-tailed than on the other stocks, which agrees with the kurtosis statistics shown in Table 1.

Table 5 contains the estimates of the model parameters for the temporal component, their standard errors (in parentheses), the mean waiting time between exogenously driven exceedances (WT) and the p-values of the K-S and L-B tests on the residuals calculated using the in-sample data. The marked RHawkes process with gamma or Weibull inter-renewal distributions clearly fits to the data better than the Hawkes process model (the exponential case), with uniformity of residuals for the RHawkes process models passing the K-S test at the 5% level on all five stocks while the residuals for the Hawkes process are always smaller and mostly fail at the 5% level, although the results on the L-B tests of independence of residuals for the Hawkes and RHawkes processes are very similar, except in the case of the DOW stock. The goodness-of-fit results of the gamma and Weibull RHawkes processes are similar, although the fit by the gamma RHawkes model is slightly better. Figure 3 shows the uniform Q-Q plots and the ACF plots of the Rosenblatt residuals U_i (15) for the gamma RHawkes model fitted to the different stocks. Graphically, the uniform Q-Q plots indicates good agreement between empirical and theoretical quantiles,

Table 4 Results of the maximum likelihood estimation of excesses using the GPD with common shape parameter ξ and scale parameters evolving according to a first order Markov process $\sigma_j = a + bw_{j-1}$, for each ASX stock

	$\hat{\xi}$	$\hat{a}$	$\hat{b}$	K–S	L–B
JBH	0.113 (0.0672)	1.213 (0.152)	0.133 (0.0675)	0.7720	0.3189
ABC	0.118 (0.0706)	1.124 (0.137)	0.0883 (0.0697)	0.7737	0.4681
CPU	0.129 (0.0687)	0.748 (0.108)	0.286 (0.0883)	0.8115	0.0058
DOW	0.337 (0.0755)	1.192 (0.134)	0.0071 (0.0351)	0.4579	0.0649
JHX	0.0976 (0.0720)	1.089 (0.128)	0.122 (0.0657)	0.9627	2.656×10^{-7}

Table 5 Results of the maximum likelihood estimation of the marked RHawkes process with exponential (mean μ), Weibull and gamma (shape κ and scale β) distributed inter-immigration waiting times, exponential offspring densities (mean γ) and linear impact function (δ describes the strength of the excesses) and branching ratio η, for each ASX stock

	$\hat{\kappa}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\delta}$	$\hat{\eta}$	WT	K-S	L-B
JBH								
Exponential	$\hat{\mu} = 22.67$ (5.73)		42.82 (12.67)	1.643 (3.118)	0.564 (0.115)	22.67	0.049	0.015
Weibull	1.544 (0.336)	28.07 (6.10)	29.30 (9.47)	0.694 (0.712)	0.609 (0.0906)	25.25	0.146	0.019
Gamma	1.983 (0.906)	12.44 (4.28)	29.87 (10.28)	0.772 (0.908)	0.599 (0.101)	24.67	0.189	0.019
ABC								
Exponential	$\hat{\mu} = 23.37$ (6.33)		51.84 (16.97)	0.722 (0.870)	0.575 (0.120)	23.37	0.008	0.185
Weibull	1.384 (0.220)	24.50 (4.85)	35.37 (12.71)	0.571 (0.548)	0.555 (0.0911)	22.37	0.044	0.187
Gamma	1.606 (0.374)	13.21 (3.26)	37.37 (13.96)	0.700 (0.733)	0.530 (0.0948)	21.21	0.064	0.177
CPU								
Exponential	$\hat{\mu} = 27.30$ (8.82)		41.77 (15.22)	0.170 (0.250)	0.639 (0.123)	27.30	0.067	0.340
Weibull	1.322 (0.242)	29.74 (8.19)	32.61 (10.86)	0.152 (0.201)	0.639 (0.103)	27.38	0.116	0.314
Gamma	1.704 (0.616)	15.66 (4.94)	32.13 (10.77)	0.160 (0.208)	0.630 (0.106)	26.68	0.131	0.314
DOW								
Exponential	$\hat{\mu} = 19.90$ (4.29)		20.44 (6.91)	0.110 (0.137)	0.498 (0.109)	19.90	0.074	0.287
Weibull	1.768 (0.578)	30.49 (8.57)	15.49 (4.46)	0.0844 (0.0787)	0.633 (0.108)	27.14	0.230	0.139
Gamma	7.129 (4.409)	4.95 (2.60)	17.51 (4.10)	0.0791 (0.0671)	0.719 (0.0656)	35.26	0.244	0.071
JHX								
Exponential	$\hat{\mu} = 21.82$ (6.40)		47.92 (15.61)	0.800 (1.168)	0.549 (0.139)	21.82	0.019	0.332
Weibull	1.374 (0.234)	25.39 (6.65)	36.19 (10.77)	0.552 (0.613)	0.575 (0.113)	23.21	0.065	0.345
Gamma	1.705 (0.513)	12.76 (2.98)	36.12 (11.10)	0.686 (0.856)	0.546 (0.122)	21.75	0.109	0.355

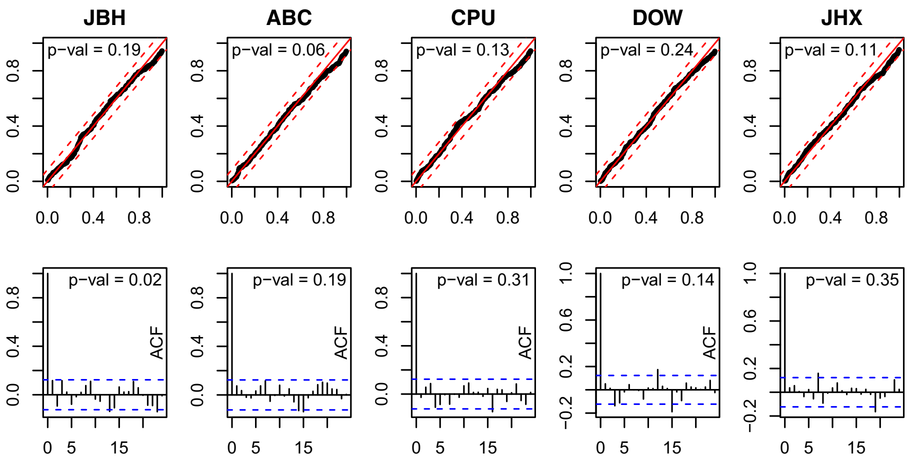

Fig. 3 Graphical goodness-of-fit test of the Rosenblatt residuals for the RHawkes process with gamma distributed inter-immigration waiting times for each ASX stock. The top panels are the uniform Q-Q plots and the lower panels are the ACF plots

although they tend to depart in the upper quantiles slightly but still remain within the 95% confidence intervals. The autocorrelations among the residuals are mostly negligible as seen in the ACF plots, and this is confirmed by the large p-values of the L-B tests for the majority of the stocks.

The added flexibility of the RHawkes process introduces an extra parameter and so to have an appropriate comparison between the RHawkes process and the less flexible Hawkes process, we compute the Akaike information criterion (AIC) for each model on each of the five stocks, as seen in Table 6. For each stock the Weibull and gamma renewal models outperform the Hawkes model. The AIC for the gamma and Weibull models are again rather comparable. Therefore, by taking into consideration the goodness-of-fit test results and the AIC values, to correctly model the temporal patterns of exceedances, the more flexible RHawkes model is to be preferred to the classical Hawkes process. In the sequel, when talking about RHawkes processes, we only consider the case with gamma distributed inter-renewal times.

In interpreting the model fit results, we consider only the JBH stock since the results on other stocks can be interpreted similarly. The estimated shape parameter of the gamma inter-renewal waiting time, $\hat{\kappa} = 1.983$, suggests that exogenously driven extreme losses occur more evenly through time than suggested by the classical Hawkes process (i.e. $\hat{\kappa} = 1$). The mean waiting time between successive immigrants is $\hat{\kappa}\hat{\beta} = 24.67$ days which is slightly larger than suggested by the Hawkes model (22.67 days). The mean waiting time from an extreme loss to an extreme loss directly generated by it is 29.87 days, while the Hawkes model suggests this mean waiting time is considerably longer at 42.82 days. Another interesting comparison is between the mean waiting time between exogenously and endogenously driven exceedances suggested by the RHawkes model. By comparing the WT and the estimated $\hat{\gamma}$ values, it is clear that exogenously driven exceedances occur more rapidly than endogenously driven ones and this is consistent with all but one of the stocks. The value $\hat{\delta} = 0.772$ reflects a moderate impact of the excess values on the propensity for future exceedances. For example, an excess of 2% leads to an increase in propensity contribution from that exceedance by 13.44% while an excess of 5% leads to an increase of 16.7%. The estimated δ value decreases when moving from exponential inter-renewals to gamma inter-renewals and this is consistent for all the stocks. Using gamma inter-renewals thus reduces the impact that the loss excesses has on the intensity for future exceedances as compared to the exponential inter-renewals. The relatively large branching ratio $\hat{\eta} = 0.599$ suggests a high degree of self-excitation,

Table 6 Akaike information criterion (AIC) for each ASX stock with exponential, Weibull and gamma distributed inter-renewals

Stock	JBH	ABC	CPU	DOW	JHX
Exponential	1613.36	1625.59	1634.97	1644.57	1627.29
Weibull	1609.37	1622.72	1634.06	1640.93	1623.64
Gamma	1609.60	1622.37	1633.00	1640.48	1622.41

with the model interpreting slightly more exceedances to be endogenous rather than exogenous.

Next we use the procedure developed in Section 3.5 to estimate the risk measures on the five stocks. We also assess the performance of the estimation using backtesting, where the number of VaR exceptions expected by the estimated risk measure is compared to the actual number of VaR exceptions. Here, a VaR exception occurs whenever the actual log-loss on a particular day exceeds the estimated value of the VaR. A large number of exceptions implies that the model is underestimating the risk. Figure 4 presents the time series plot of the log-losses and the estimated 95% and 99% VaR based on the RHawkes process and based on the classical Hawkes process for the ASX stocks over the period from 1 January 2012 to 31 December 2016. Here we have used a longer time period for backtesting which contains part of the in-sample period as well as the entire out-of-sample period, for the comparison between the expected and actual numbers of VaR exceptions to be meaningful.

One method to formally assess how well the VaR estimator performs is to test whether the observed proportion of VaR exceptions $\hat{p}$ agrees with the expected proportion of exceptions $p = 1 - q$, where q is the quantile level used in the VaR calculation. The null hypothesis states that the model correctly forecast the VaR, while the alternate hypothesis states that the model underestimates the VaR, since in financial applications we typically give importance to not underestimating risk. For the stock JBH, the percentage of actual exceptions based on the RHawkes model estimate of the 95% VaR is 4.49%, and is 0.528% based on the

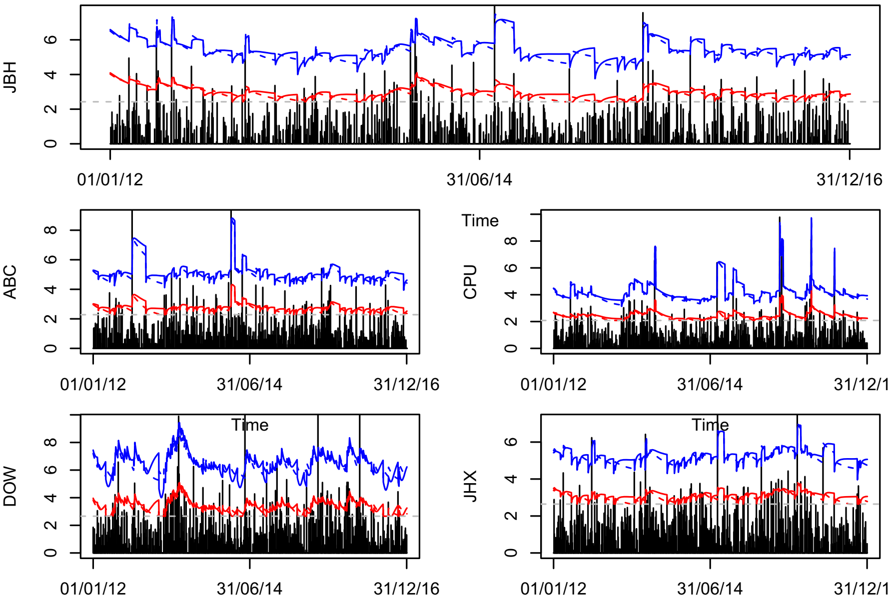

Fig. 4 Time series plot of the daily log-losses for each ASX stock for the period 1 January 2012 to 31 December 2016 together with the 95% and 99% estimates of VaR. The solid lines are based on estimates for the gamma RHawkes model and the dashed lines are based on the Hawkes model

99% VaR estimate. Both these values agree well with the respective expected proportions, with the p-values of the one-side exact binomial tests equal to 0.8345 and 0.9839 respectively. The corresponding p-values on the other four stocks are all much larger than 5%, suggesting the RHawkes model based VaR estimator has satisfactory performance. The Hawkes model based VaR estimator also passes the test on all five stocks, although typically with smaller p-values. Therefore, the VaR estimators based on the RHawkes and Hawkes models have similar performances.

However, Fig. 4 reveals that the VaR estimate based on the RHawkes model can drop following an exceedance, while the VaR estimate by the Hawkes model always jumps up following an exceedance. The reason for this is that the estimated shape parameter $\hat{\kappa}$ of the inter-renewal distribution is larger than 1, implying that the estimated hazard function μ given in Eq. 24 is monotonically increasing from 0, and therefore $\mu(t - \tau_k) + \phi(t)$ can drop following an immigrant event, which in turn causes the approximation (21) used in the calculation of the VaR in Eq. 22 to go down following an event with a small excess, but a high probability to be an immigrant. In contrast, the hazard function μ in the Hawkes model is a constant, so the intensity $\mu + \phi(t)$ always jumps up when the excitation effect enters $\phi(t)$ following each exceedance, which causes the VaR estimate in Eq. 22 to jump up. The estimated ES at the 95% level for the five stocks are shown in Fig. 5, from which we can see the estimates using both the RHawkes and the Hawkes models are again similar to each other. However, similar to the VaR estimate, the ES estimate by the RHawkes model

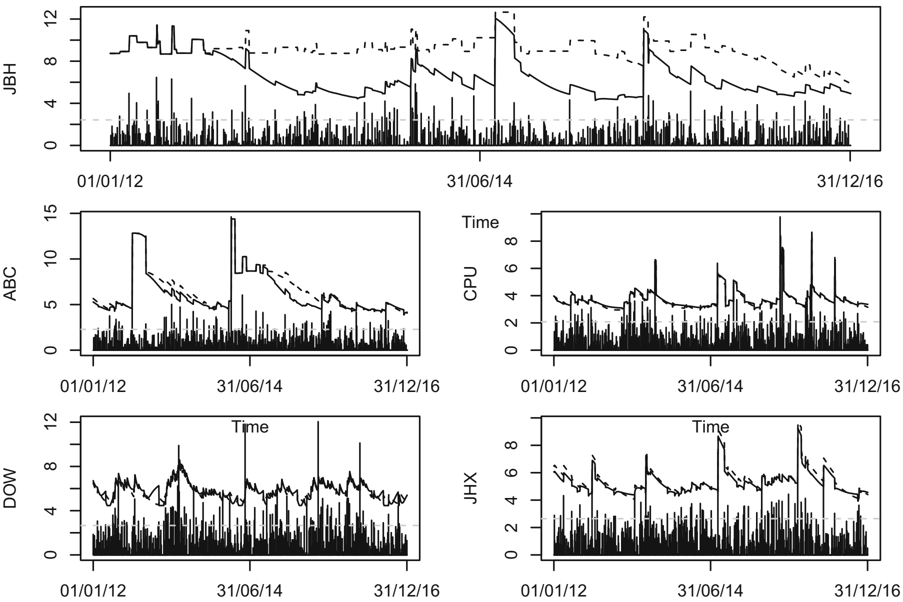

Fig. 5 Time series plot of the daily log-losses for the ASX stocks for the period 1 January 2012 to 31 December 2016 together with the 95% estimates of ES. The solid line is based on estimates from the gamma RHawkes model and the dashed line is the Hawkes model

can drop momentarily following an exceedance with a small excess, suggesting that the chance and the size of an extreme loss right after a small exceedance can both be smaller than before the exceedance.

Another interesting forecast to perform is the number of days until the next extreme loss. The developments in Section 3.4 are used to make this forecast for each stock and then compared with actual observations. To predict the waiting time until the next extreme loss, we condition on all available information at the censoring time, and compute the predictive density and hazard function using both the RHawkes and Hawkes models. In Fig. 6 we plot the predictive densities using solid lines while the dashed lines are used for the predictive hazard functions. The black lines are for the RHawkes model while the grey lines are used for the Hawkes model.

For the JBH stock, the probability that an exceedance occurs in the first, second, third and fourth 10-day period using the RHawkes model are 54.12%, 24.53%, 11.29% and 5.28% respectively, with the actual exceedance indeed occurring during the first ten-day period. These predicted probabilities by the Hawkes model are similar, although the predicted probability of having the first extreme loss in the first 10-day period is slightly smaller. The predictive hazard function gives us an estimate of the hazard, or conditional probability of having an extreme loss on a given day conditional on that it has not occurred by the previous day. We can see from Fig. 6 that by the RHawkes model, the hazard of seeing an extreme loss on 1 January 2016

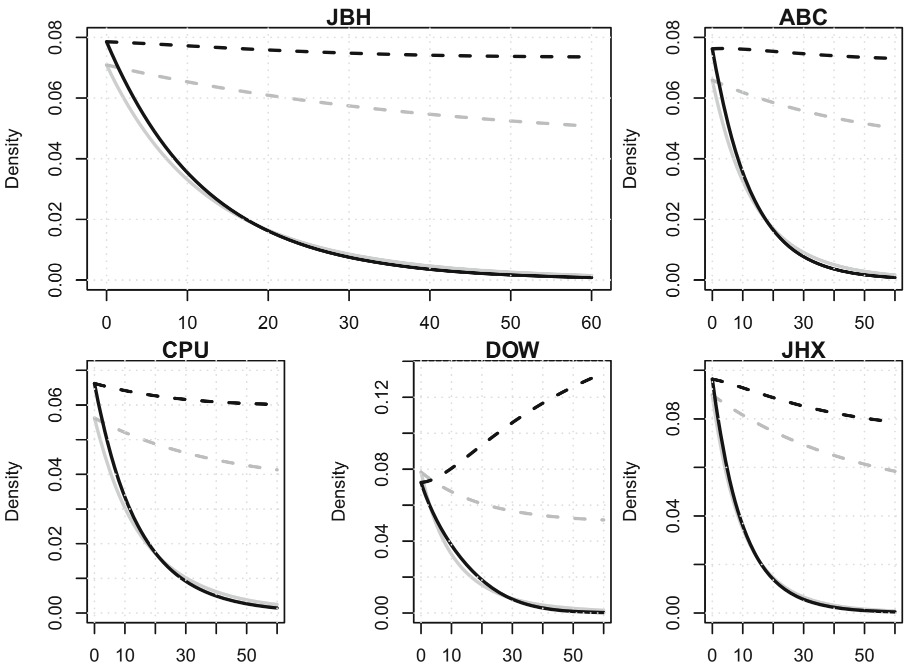

Fig. 6 Solid lines show the predictive densities and the dashed lines show the predictive hazard functions for the waiting time (in days) until the next exceedance after the censoring time. The black lines represent the RHawkes model while the shaded lines represent the Hawkes model

is 7.86%, while by the Hawkes model, it is only 7.09% The predictions for the other four stocks by the two models can be interpreted similarly. Note that the hazard function for the stock DOW increase over time and this is a result of the large estimated shape parameter $\hat{\kappa} = 7.129$ and a recent exceedance having a high probability of being exogenously driven.

6 Concluding remarks

In this work we have proposed an extension to the marked Hawkes process considered by Chavez-Demoulin et al. (2005) and Chavez-Demoulin and McGill (2012) to a marked renewal Hawkes process, where the immigrant (or spontaneous/exogenous) events can follow a general renewal process rather than a Poisson process. We developed a recursive algorithm to evaluate the likelihood of the proposed model that is similar to the algorithms in Chen and Stindl (2018) and Stindl and Chen (2018) for the evaluation of the likelihoods of the unmarked and multivariate versions of the renewal Hawkes process. Through simulation experiments, we demonstrated that the likelihood evaluation algorithm is effective for the purpose of fitting the marked renewal Hawkes process model to data using the maximum likelihood approach. Our application of the marked renewal Hawkes process to data on the extreme negative returns of several stocks showed better goodness-of-fit than the marked Hawkes process considered by Chavez-Demoulin et al..

It should be noted that, although we have used a continuous-time approach following the works of Chavez-Demoulin et al. in modelling the times and excess losses of extreme negative stock returns, marked versions of the discrete-time counterpart of the Hawkes process (Kirchner 2016) and the binary time-series model discussed e.g. in Cui and Lund (2009) might be more natural models for such data due to the measurement of event times in multiples of days.

Acknowledgements The authors appreciate and thank the editor, the associate editor, and two anonymous reviewers for their very insightful comments that have led to a more consistent, improved and well presented paper.

References

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

Chavez-Demoulin, V., McGill, J.: High-frequency financial data modeling using Hawkes processes. J. Bank. Financ. **36**(12), 3415–3426 (2012)

Chen, F., Stindl, T.: Direct likelihood evaluation for the renewal Hawkes process. J. Comput. Graph. Stat. **27**(1), 119–131 (2018)

Cline, D., Pu, H.: A note on a simple Markov bilinear stochastic process. Statist. Probab. Lett. **56**(3), 283–288 (2002)

Cui, Y., Lund, R.: A new look at time series of counts. Biometrika **96**(4), 781–792 (2009)

Daley, D.J., Vere-Jones, D. An Introduction to the Theory of Point Processes Volume I: Elementary Theory and Methods, 2nd edn. Springer, New York (2003)

Embrechts, P., Mikosch, T., Klüppelberg, C.: Modelling Extremal Events for Insurance and Finance. Springer, New York (1997)

Embrechts, P., Liniger, T., Lin, L.: Multivariate Hawkes processes: An application to financial data. J. Appl. Probab. **48**(A), 367–378 (2011)
Hawkes, A.G.: Spectra of some self-exciting and mutually exciting point processes. Biometrika **58**(1), 83–90 (1971)
Herrera, R., Schipp, B.: Self-exciting extreme value models for stock market crashes. In: Statistical Inference, Econometric Analysis and Matrix Algebra, pp. 209–231. Physica, Heidelberg (2009)
Kirchner, M.: Hawkes and INAR(∞) processes. Stoch. Processes Appl. **126**(8), 2494–2525 (2016)
McNeil, A.J., Frey, R.: Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. J. Emp. Fin. **7**(3), 271–300 (2000)
McNeil, A., Frey, R., Embrechts, P.: Quantitative Risk Management: Concepts, Techniques, and Tools. Princeton Series in Finance. Princeton University Press (2005)
Mina, J., Xiao, J.Y.: Return to riskmetrics: The evolution of a standard. RiskMetrics Group (2001)
R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna (2017)
Rosenblatt, M.: Remarks on a multivariate transformation. Ann. Math. Stat. **23**(3), 470–472 (1952)
Stindl, T., Chen, F.: Likelihood based inference for the multivariate renewal Hawkes process. Comput. Stat. Data Anal. **123**, 131–145 (2018)
Wheatley, S., Filimonov, V., Sornette, D.: The Hawkes process with renewal immigration & its estimation with an EM algorithm. Comput. Stat. Data Anal. **94**(C), 120–135 (2016)