Supplementary Materials for 'Score Test for Marks in Hawkes Processes'

Kylie-Anne Richards[1*], William T.M. Dunsmuir[2,3†] and Gareth W. Peters[1,2†]

[1*]UTS Business School, University of Technology Sydney, 14/28 Ultimo Rd, Ultimo, 2007, NSW, Australia.
[2]School of Mathematics and Statistics, University of NSW, Sydney, 2052, NSW, Australia.
[3]Statistics & Applied Probability, University of California, Santa Barbara, CA, 93106-3110, State, USA.

*Corresponding author. E-mail: kylie-anne.richards@uts.edu.au;
†These authors contributed equally to this work.

Abstract

Supplementary Materials for 'Score Test for Marks in Hawkes Processes'. A score statistic for detecting the impact of marks in a linear Hawkes self-exciting point process is proposed, with its asymptotic properties, finite sample performance, and power properties using simulation and application to real data presented. A major advantage of the proposed inference procedure is that the Hawkes process can be fit under the null hypothesis that marks do not impact the intensity process. Therefore, for a given record of a point process, the intensity process is estimated once only and then assessed against any number of potential marks without refining the joint likelihood each time. Marks can be multivariate and serially dependent. The score function for any given set of marks is easily constructed as the covariance of functions of future intensities fits the unmarked process with functions of the marks under assessment. The asymptotic distribution of the score statistic is a chi-square distribution, with degrees of freedom equal to the number of parameters required to specify the boost function. Model-based or non-parametric estimation of required features of the marks marginal moments and serial dependence can be used. The use of sample moments of the marks in the test statistic construction does not impact size and power properties.

Keywords: Marked Hawkes point process; Score test statistic; High frequency financial data

Contents

A Score and Information Matrix Derivation Details

Recall $\nu = (\theta, \phi, \psi) \in \Theta \times \Phi \times \Psi$ be the collection of all parameters for the marked process with intensity function

$$\lambda_g(t;\theta,\phi,\psi) = \eta + \vartheta \int_{[0,t)\times\mathbb{X}} w(t-s;\alpha) g(\mathbf{x};\phi,\psi) N_g(ds \times d\mathbf{x}), \quad (1)$$

where $w : \mathbb{R}_+ \to \mathbb{R}_+$ is a non-negative decay function. We denote the true value of the parameters by $(\theta^*, \phi^*, \psi^*)$. When the null hypothesis holds, the true parameter vector is denoted $\nu^* = (\theta^*, \phi^*, 0)$.

The log-likelihood is given by

$$l_g(\nu) = \int_{[0,T]\times\mathbb{X}} \ln \lambda_g(t;\nu) N_g(dt \times d\mathbf{x}) - \Lambda_g(T;\nu) + \int_{[0,T]\times\mathbb{X}} \ln f_t(\mathbf{x};\phi) N_g(dt \times d\mathbf{x}), \quad (2)$$

where $\Lambda_g(T,\nu) = \int_0^T \lambda_g(t;\nu)dt$ is the compensator at T.

The derivatives of the log-likelihood (2) with respect to ν are

$$\partial_\nu l_g(\nu) = \int_{[0,T]\times\mathbb{X}} \frac{1}{\lambda_g(t;\nu)} \partial_\nu \lambda_g(t;\nu) N_g(dt \times d\mathbf{x}) - \partial_\nu \Lambda(T;\nu) + \frac{1}{f(\mathbf{x};\phi)} \partial_\nu f(\mathbf{x};\phi), \quad (3)$$

where $f(\mathbf{x};\phi)$ is the joint density of $\mathbf{x}_1, \ldots, \mathbf{x}_n$ conditional on the event times. Consider the components of these partial derivatives. First, with respect to $\theta = (\eta, \vartheta, \alpha)$, since $\partial_\theta f(\mathbf{x};\phi) = 0$ and $\partial_\psi f(\mathbf{x};\phi) = 0$, the last term in (3) does not contribute to $\partial_\theta l_g(\theta)$ or $\partial_\psi l_g(\theta)$. Now

$$\partial_\phi \lambda_g(t;\nu)|_{\nu^*} = \vartheta^* \int_{[0,t)\times\mathbb{X}} w(t-s;\alpha^*) \frac{-h(\mathbf{x};\phi^*,0)}{\mathbb{E}_{\phi^*}[h(\mathbf{X};\phi^*,0)]^2} \partial_\phi \mathbb{E}_{\phi^*}[h(\mathbf{X};\phi,\psi)]|_{\nu^*} N_g(ds \times d\mathbf{x}) = 0,$$

since at ν^*, $\psi = 0$, $h(\mathbf{x};\phi^*,0) = 1$, $\mathbb{E}_{\phi^*}([h(\mathbf{X};,\phi^*,0)] = 1$ and Condition 1(iv) assumes that $\partial_\phi \mathbb{E}_\phi[h(\mathbf{X};\phi,\psi)]|_{\nu^*} = 0$ in , hence $\partial_\phi \Lambda_g(T;\nu)|_{\psi=0} = 0$ and consequently

$$\partial_\phi l_g(\nu^*) = \partial_\phi \ln f(\mathbf{x};\phi^*).$$

Now

$$\partial_\psi l_g(\nu^*) = \int_{[0,T]} \lambda(t;\theta^*)^{-1} \partial_\psi \lambda_g(t;\nu^*) \tilde{N}(dt),$$

where $\tilde{N}(dt) = N(dt) - \lambda(t;\theta^*)dt$ and

$$\partial_\psi \lambda_g(t;\nu)|_{\nu^*} = \vartheta^* \int_{[0,t)\times\mathbb{X}} w(t-s;\alpha^*) \partial_\psi g(\mathbf{x};\phi^*) N_g(ds \times d\mathbf{x}) = \vartheta^* \int_{[0,t)\times\mathbb{X}} w(t-s;\alpha^*) G(\mathbf{x};\phi^*) N_g(ds \times d\mathbf{x}). \quad (4)$$

Hence, conditioning on the marks in an iterated expectation, we have

$$\mathbb{E}_{\phi^*}[\partial_\psi l_g(\nu^*) \partial_\phi l_g(\nu^*)^\mathsf{T}] = \mathbb{E}_{\nu^*}[\mathbb{E}_{\nu^*}[\partial_\psi l_g(\nu^*)|\mathbf{X}] \partial_\phi l_g(\nu^*)^\mathsf{T}] = 0$$

since $\mathbb{E}_{\nu^*}[\partial_\psi l_g(\nu^*)|\mathbf{X}] = 0$. We can also arrive at this result by showing that the off diagonal block of the second derivative of the likelihood with respect to ϕ and ψ satisfies $\mathbb{E}_{\nu^*}[\partial^2_{\phi\psi} l_g(\nu)|_{\nu^*}] = 0$, which again uses Condition 1(iv).

Also, it is straightforward to show that $\partial_\theta l_g(\nu)|_{\nu^*}$ does not depend on the marks, so $\mathbb{E}_{\nu^*}[\partial_\theta l_g(\nu^*)\partial_\psi l_g(\nu^*)^\mathsf{T}] = 0$, similarly. Furthermore, $\partial_\phi l_g(\nu^*) = \partial_\phi \ln f(\mathbf{x};\phi^*)$ results in $\mathbb{E}_{\nu^*}[\partial_\theta l_g(\nu^*)\partial_\phi l_g(\nu^*)^\mathsf{T}] = 0$.

We have shown that the information matrix is block diagonal with blocks corresponding to the partition of ν into θ, ϕ, and ψ.

The second derivative matrix of the log-likelihood with respect to ψ is

$$\begin{aligned}\partial^2_{\psi\psi} l_g(\nu) = &\int_{[0,T]\times\mathbb{X}} \partial_\psi(\lambda_g(t;\theta)^{-1}\partial_\psi\lambda_g(t;\nu))\\ &N_g(ds\times d\mathbf{x}) - \int_{[0,T]} \partial^2_{\psi\psi}\lambda_g(t;\nu)dt\\ =&\int_{[0,T]\times\mathbb{X}} \partial_\psi(\lambda_g(t;\theta)^{-1}\partial_\psi\lambda_g(t;\nu))\\ &(N_g(ds\times d\mathbf{x}) - \lambda_g(t;\nu^*)dt)\\ &-\int_{[0,T]} \lambda_g(t;\nu)^{-1}\partial^2_{\psi\psi}\lambda_g(t;\nu)\{\lambda_g(t;\nu)\\ &-\lambda_g(t;\nu^*)\}dt\\ &-\int_{[0,T]} \lambda_g(t;\nu)^{-2}\partial_\psi\lambda_g(t;\nu)^{\otimes 2}\lambda_g(t;\nu^*)dt,\end{aligned} \tag{5}$$

where

$$\begin{aligned}\partial^2_{\psi\psi}\lambda_g(t;\nu) = \vartheta\int_{[0,t)\times\mathbb{X}} &w(t-s;\alpha)\partial^2_{\psi\psi}\\ &g(\mathbf{x};\phi,\psi)N_g(ds\times d\mathbf{x}).\end{aligned} \tag{6}$$

Let $\nu = (\theta,\phi,0)$ (that is at any values of θ and ϕ when $\psi = 0$) then (5) and (6) evaluate at ν, as

$$\begin{aligned}\partial^2_{\psi\psi} l_g(\nu) = &\int_{[0,T]} \{\lambda(t;\theta)^{-1}\partial^2_{\psi\psi}\lambda_g(t;\nu)\\ &-\lambda(t;\theta)^{-2}\partial_\psi\lambda_g(t;\nu)^{\otimes 2}\}\tilde{N}(ds)\\ &-\int_{[0,T]} \lambda(t;\theta)^{-1}\partial^2_{\psi\psi}\lambda_g(t;\nu)\{\lambda(t;\theta)\\ &-\lambda(t;\theta^*)\}dt\\ &-\int_{[0,T]} \lambda(t;\theta)^{-2}\partial_\psi\lambda_g(t;\nu)^{\otimes 2}\lambda(t;\theta^*)dt,\end{aligned} \tag{7}$$

and

$$\begin{aligned}\partial^2_{\psi\psi}\lambda_g(t;\nu) = \vartheta\int_{[0,t)\times\mathbb{X}} &w(t-s;\alpha)\partial^2_{\psi\psi} g(\mathbf{x};\phi,\psi)\\ &N_g(ds\times d\mathbf{x}),\end{aligned} \tag{8}$$

respectively.

Under the null hypothesis, the score with respect to ψ is

$$\partial_\psi l_g(\nu) = \vartheta\int_{[0,T]\times\mathbb{X}} A_T(t;\theta)G(\mathbf{x};\phi)N_g(dt\times d\mathbf{x}), \tag{9}$$

where

$$A_T(t;\theta) = \int_{(t,T]} \lambda(s;\theta)^{-1}w(s-t;\alpha)N(ds) - W(T-t;\alpha),$$

where $W(s;\alpha) = \int_0^s w(u;\alpha)du$ is the cumulative decay function.

To derive the information matrix at ν^* (the true parameter under H_0) using (9) we obtain

$$\begin{aligned}\mathcal{I}_\psi(\nu^*) &= \mathbb{E}_{\nu^*}[\partial_\psi l_g(\nu^*)\partial_\nu l_g(\nu^*)^\mathsf{T}]\\ &= \mathbb{E}[(\int_{[0,T]} \lambda(t;\theta^*)^{-1}\partial_\psi\lambda_g(t;\nu^*)\tilde{N}(dt))^{\otimes 2}]\\ &= \mathbb{E}[\int_{[0,T]} \lambda(t;\theta^*)^{-1}\partial_\psi\lambda_g(t;\nu^*)^{\otimes 2}dt].\end{aligned} \tag{10}$$

Alternatively, using (7) we obtain

$$\begin{aligned}\mathcal{I}_\psi(\nu^*) &= -\mathbb{E}_{\nu^*}[\partial^2_{\psi\psi} l_g(\nu^*)]\\ &= -\mathbb{E}[\int_{[0,T]} \lambda(t;\theta^*)^{-1}\partial^2_{\psi\psi}\lambda_g(t;\nu^*)\tilde{N}(dt)]\\ &\quad+\mathbb{E}[\int_{[0,T]} \lambda(t;\theta^*)^{-2}\partial_\psi\lambda_g(t;\nu^*)^{\otimes 2}N(dt)].\end{aligned} \tag{11}$$

Now

$$\begin{aligned}\partial^2_{\psi\psi} g(X;\phi,\psi)|_{\nu^*} &= H'(X) - \mathbb{E}_{\phi^*}[H'(X)]\\ &-2\mathbb{E}_{\phi^*}[H(X)](H(X) - \mathbb{E}_{\phi^*}[H(X)]),\end{aligned} \tag{12}$$

which has expectation of zero, hence

$$\mathbb{E}[\partial^2_{\psi\psi}\lambda_g(t;\nu^*)|\mathcal{F}_{t-}] = 0,$$

so that the first term in (11) is zero. Hence

$$\mathcal{I}_\psi(\nu^*) = \mathbb{E}[\int_{[0,T]} \lambda(t;\theta^*)^{-2}(\partial_\psi\lambda_g(t;\nu^*))^{\otimes 2}N(dt)], \tag{13}$$

which is obviously equal to (10) by using $N(dt) = \tilde{N}(dt) + \lambda(t;\theta^*)dt$ in (13).

B Adjusting the score test for stationary serially dependent marks

Simplification of the covariance matrix of the normalized score Ω_n

$$\begin{aligned}\Omega_n :&= \mathrm{cov}(S_n(\nu^*)) \\ &= \mathbb{E}_{\theta^*}\left[\frac{1}{n}\sum_{l=1}^{n-1}\sum_{m=1}^{n-1} A_{n,l}(\theta^*)A_{n,m}(\theta^*)\right. \\ &\left.\mathbb{E}_{\phi^*}[G(X_l;\phi^*)G(X_m;\phi^*)|t_1,\ldots,t_n],\right. \end{aligned} \tag{14}$$

can be made for stationary marks processes as follows. First, consider the situation where the marks are observations $\mathbf{x}_i = y_{t_i}$ on a continuous time process with covariance function $\mathbb{E}[G_k G_l^T] = \gamma(t_k - t_l;\phi)$, which is a function of the time between events. Then (14) becomes

$$\begin{aligned}\Omega_n =&\mathbb{E}\left[\frac{1}{n}\sum_{l=1}^{n-1} A^2_{n,l}\right]\Omega^*_G + \\ &\frac{1}{n}\sum_{m=1}^{n-1}\sum_{l=1}^{n-1}\mathbb{I}(l\neq m)\mathbb{E}\left[A_{n,k}A_{n,l}\right] \\ &\{\gamma_G(t_l - t_m;\phi) + \gamma_G(t_m - t_l;\phi)\}.\end{aligned}$$

A simpler alternative is to assume that the marks form a stationary time series indexed by m with the lag h autocovariances of $G(\mathbf{X}_m,\phi)$, denoted by $\gamma_G(h;\phi)$. Here the actual inter-event times don't matter in this case. Then the variance of the score vector simplifies to

$$\begin{aligned}\Omega_n = &\mathbb{E}\left[\frac{1}{n}\sum_{l=1}^{n-1} A^2_{n,l}\right]\Omega_G \\ &+\sum_{h=1}^{n-1}\mathbb{E}\left[\frac{1}{n}\sum_{l=1}^{n-h} A_{n,l}A_{n,l+h}\right]\{\gamma_G(h;\phi)^T \\ &+\gamma_G(h;\phi)\}.\end{aligned} \tag{15}$$

Obviously, if there is serial dependence in the marks and this is ignored in calculating the variance of the normalized score vector, the resulting score statistic will not converge to the chi-squared distribution but will converge to a constant times the chi-squared. If there is positive serial dependence, the constant will be greater than 1, in which case the (incorrect) test statistic will be over-sized under the null hypothesis, which will lead to falsely inflated power. This is illustrated simulation experiments. If the series dependence is predominantly negative, then the constant will be less than 1, and the size will be falsely low. Negative correlation can arise, but it is less common than positive dependence in our application. In either case, it is critical to correct for serial dependence by using a consistent estimate of Ω_n in (15).

There are several options available for estimating the covariances $\gamma_G(h;\phi)$. Firstly, one could estimate the parameters ϕ needed to specify the full joint conditional density $f(x_1,\ldots,x_n|t_1,\ldots,t_n;\phi)$ and use these in theoretical formulae for the needed covariances. For many of the series we have encountered, derivation of the theoretical autocovariances may not be a straightforward task, so this method may not be easily implemented. An alternative is to use nonparametric estimation of the required covariances. This is in principle, possible without requiring a model from which expressions for theoretical autocovariances would need to be derived. However, it will not always be possible to consistently estimate all the $n-1$ terms required for estimation of (15), the main reason being the lack of data to obtain consistent estimation of higher lag autocovariances of the $\{G(\mathbf{X}_m)\}$ process. There are two sets of serial dependence to take into account in forming (15), serial dependence associated with the $A_{n,m}$ series and serial dependence associated with the $G(\mathbf{X}_m)$ series. For the exponential decay

case, the autocorrelation in the $A_{n,l}$ appears, at least empirically, to decay geometrically fast to zero as h in increase. In that case, the rate of decay of $\gamma_G(l)$ is not as important for a truncation method to give an accurate calculation. Typically when quantities such as (15) need to be calculated from a single realization of a time series, various tapering methods are employed.

For continuous time version (**??**), non-parametric estimation of the $\gamma_G(t_l - t_m; \phi)$ is considerably more challenging. One could use variogram based estimates as discussed in Diggle et al. [2], from which non-parametric estimates can be obtained. Typically these methods require very large amounts of data to be effective, and the lack of consistency of the estimated covariances for high order time separation remains an issue, so some form of tapering would be required in this case also.

Another option that avoids all of the above issues is to use an empirical version of $-\mathcal{I}_\psi(\nu^*)/(n\vartheta^2)$ from (13) to give

$$\hat{\Omega}_n = \frac{1}{n}\sum_{m=1}^{n} \lambda(t_i;\hat{\theta})^{-2} \left(\sum_{t_j<t_m} w(t_m - t_j;\hat{\alpha})\hat{G}_m\right)^{\otimes 2}. \quad (16)$$

This method only requires likelihood estimation of the parameters of the unmarked process parameters θ, α, and parametric or non-parametric estimation of the expected values of $H(\mathbf{X})$, similarly to above.

C Estimation of the unmarked Hawkes process

Under H_0, the observed event times are those of an unmarked Hawkes process N, with intensity

$$\lambda(t;\nu) = \eta + \vartheta \int_{[0,t)} w(t-s;\alpha)N(ds), \quad (17)$$

and the marks observed at the event times of this process do not impact its intensity.

For the description of the algorithms for the Hawkes process, we rely considerably on Liniger [6] with modifications for extensions. We assume that the initial value of the intensity is $\lambda(t_0) = C$ for some specified value of C. A sensible selection is $C = \mathbb{E}[\lambda(t)] = \eta/(1-\vartheta)$, which is the theoretical long-run average for a stationary Hawkes process [5]. We define the null hypothesis log-likelihood in (18) in summation form with $T = t_n$. The discrete summation form of the score statistic follows from this representation also. The log-likelihood function expresses the integrals in

$$l(\nu) = \int_{[0,T]} \ln\lambda(s;\nu)N(ds) - \Lambda(T;\nu), \quad (18)$$

by summation over the events as

$$\hat{l}(\nu) = \sum_{i=1}^{n} \ln\hat{\lambda}(t_i;\nu) - \hat{\Lambda}(T;\nu). \quad (19)$$

To evaluate (19), we require the calculation of the intensity process defined in (17), which in summation form is

$$\hat{\lambda}(t_i;\nu) = \eta + \vartheta\sum_{i=1}^{i-1} w(t_i - t_m). \quad (20)$$

The compensator $\Lambda_g(T,\nu) = \int_0^T \lambda_g(t;\nu)dt$ at time T, required for the unmarked process in (19) can be estimated by

$$\hat{\Lambda}(T,\nu) = \eta T + \vartheta\sum_{i=1}^{n} W(t_n - t_i;\alpha), \quad (21)$$

where $t_n = T$.

D Simulation algorithm for the univariate Hawkes process

To simulate a series of random events according to the specification of a given Hawkes process, we utilize Ogata's modified thinning algorithm [7], which can be applied to any choice of decay function. For the general framework and notation, we rely on the algorithm description by Liniger [6, Algorithm 1.21]; however, extensions are made for the inclusion of joint dependence structures for the marks. See Section F for further detail on copulas used within this research.

The intensity process λ is left continuous with right-hand side limits. For the purposes of simulation, we define λ^+ as the right-hand side limit after time t, that is, after the intensity is boosted. We denote λ^- as the left-hand-side limit, also known as the *hazard rate* [6].

The algorithm is comprised of an outer and inner loop. The outer loop iteration variable is $i \in \{2, \ldots, n\}$, and the inner loop is indexed by $r \geq 1$. For brevity, we present only the algorithm for the recursive method:

S.1 Define an end point for the iteration variable r. Set $i = 2$ and $\lambda^+(1) = \eta/(1-\vartheta)$.

S.2 Use a Monte-Carlo method for the case of jointly dependent marks. For $Y_{i,j} \sim F_{i,j}$, where $j \in \{1, \ldots, d\}$ are jointly dependent random variables with marginal distributions $F_{i,1}, \ldots, F_{i,d}$, let $U_{i,1}, \ldots, U_{i,d}$ have joint distribution C. For a very long time series $i = 1 \ldots n$ (i.e. $n = 20,000$): simulate $u_{i,1}$ from the random variable $U_{i,1} \sim U[0,1]$; simulate $u_{i,2}$ from the random variable $U_{i,2} \sim C_{i,2}(\cdot|u_{i,1})$; continue simulating, such that you simulate $u_{i,d}$ from the random variable $U_{i,d} \sim C_{i,d}(\cdot|u_{i,d-1})$; and evaluate $(y_{i,1}, \ldots, y_{i,d})$ from $F^{-1}(u_{i,1}), \ldots, F^{-1}(u_{i,d})$. Estimate $\mathbb{E}[Y_j]$ and $Var[Y_j]$ using long-run empirical moments. When $d = 2$, which is considered in this research, estimate the correlation in the event of joint dependence, $\rho(Y_1, Y_2)$ in (28).

S.3 Initialize the outer loop, such that, $\tau_1 := t_{i-1}$ and $\lambda_r^- := \lambda_{i-1}^+$, then set $r := 1$.

S.4 Enter the inner loop and calculate a new time τ_r

$$\tau_r = \begin{cases} t(i-1) + E/\lambda^+(i-1), & \text{if } r = 1; \\ \tau_{r-1} + E/\lambda^+(i-1), & \text{otherwise,} \end{cases}$$

where $E \sim \text{Exp}(1)$. Calculate the left hand limit intensity λ_r^-. Assuming an exponential decay function, we utilize the more efficient recursive expression for the simulation, $\lambda_r^- = \eta + e^{-\alpha(\tau_r - t_{i-1})}\left[\lambda_{i-1}^+ - \eta\right]$.

S.5 Sample a standard uniform random variable $U_r \sim U(0,1)$ and let $u_r := U_r \lambda_{i-1}^+$ and check the condition $u_r \leq \hat{\lambda}_r^-$. If this condition is true, exit the inner loop. Failing this condition being met, we start the inner loop again, sampling a new τ_r where $r = r + 1$.

S.6 Define $t_i := \tau_r$ and $\lambda(i) := \lambda^-(r)$. In the case of univariate marks or multi-dimensional marks that are independent, sample a random variable from $X_{i,j} \sim F_{i,j}$, where $j \in \{1, \ldots, d\}$ and define $x_{i,j} := X_{i,j}$.

S.7 In the case of $d \geq 2$ jointly dependent marks with distribution C and marginal mark distributions $F_{i,1}, \ldots, F_{i,d}$, let $U_{i,1}, \ldots, U_{i,d}$ have joint distribution C. Follow the method in Step S.2 to obtain samples $(X_{i,1}, \ldots, X_{i,d})$ from $F^{-1}(u_{i,1}), \ldots, F^{-1}(u_{i,d})$.

S.8 In the case of $d = 2$ considered in this research, calculate the boost function $g(\mathbf{x}; \phi, \psi)$ using the normalization adjustment, whereby $\mathbb{E}[X_1 X_2] \neq 0$ in the event of jointly dependent marks. If this cannot be calculated explicitly, utilize the long run empirical linear correlation $\rho(Y_1, Y_2)$ in (28) from Step S.2.

S.9 Define $\lambda^+(i) = \lambda_r^- + \vartheta\alpha g(\mathbf{x}; \phi; \psi)$. Let $i = i+1$ and continue with the outer iteration in Step S.3.

E Simulation extensions for a Hawkes process with serial dependent marks

To simulate a Hawkes process with serially dependent marks and non-Gaussian marginal distributions, we specify an unobserved (latent) parameter evaluation of an appropriate parametric distribution. To generate the time series of marks, we let δ_i be a latent stationary time series. Conditional on δ_i, the marks are independent with a specified parametric density $f(x_i|\theta_i)$, where for a selected component of θ_i we use δ_i and for the other components of θ_i we use fixed values.

For the single parameter distributions that we consider, they simply become $\text{Exp}(\lambda_i)$ and $\text{Pois}(\mu_i)$, where $\lambda_i = \delta_i$ and $\mu_i = \delta_i$, respectively. In the case of two parameter distributions, and for example, in the case of heavy-tailed replicates, the conditional generalized Pareto distribution has a shape parameter ζ and a scale parameter δ_i. For the negative binomial distribution, serial dependence is introduced into the process via success rate $r_i = \delta_i$.

For this paper, we use a latent process of the form

$$\delta_i = a_0 + \sum_{k=1}^{p} a_k \delta_{i-k} + \epsilon_i, \tag{22}$$

where ϵ_i is Gaussian white noise. Parameter link functions, $\Theta_i = g_i(\delta_i)$ are used to ensure the estimates do not conflict with the constraints of distribution parameters. We set the lag $p = 1$, $a_0 = 0$, and the coefficients term a_1 will vary.

The marks enter the Hawkes process via a normalized boost function. For i.i.d. marks, this boost function is normalized via the estimated theoretical moments of the marginal distribution. However, in the case of serially dependent marks, it is not obvious how to proceed with specifying the normalization, conditional generalized Pareto distribution. No theory has been developed with respect to the normalization of a boost function in the presence of marks with conditional distributions. The normalization for every time point, for example, an estimate of δ_i from an observed mark, is a non-trivial filtering problem.

There are many ways we could define a data-generating mechanism. Two such ways are:

1. Normalize the boost function locally using δ_i within the theoretical moment expression $\mathbb{E}[X] = \delta_i/1 - \zeta$;
2. Generate a very long series of serially dependent marks $X \sim GPD(\zeta, \delta_i)$. Calculate the empirical mean of this series and use the global estimate to normalize the boost function.

We will proceed with the second data-generating mechanism proposed. The algorithms below outline the steps that need to be replaced in Section D to simulate a series of random events according to a Hawkes process.

Replacement of Step S.2 in Section D:

S.1 Specify the autoregressive structure, which will be used to invoke serial dependence. In this setting, a reasonable choice is

$$\delta_i = 0.9\delta_{i-1} + \epsilon_i, \tag{23}$$

where ϵ_i is Gaussian white noise. The support of the parameter needs to be considered, for example, the generalized Pareto distribution requires $\delta > 0$; therefore, we take the exponential of the autoregressive structure.

S.2 Specify the marginal distributions $F_{i,j}$ of the mark, where the parameter of the distribution is specified by the autoregressive structure defined above. Some sensible choices for δ_i in (23) are: $Y_i \sim Exp(\lambda_i)$, where $\lambda_i = \delta_i$; $Y_i \sim Pois(\mu_i)$, where $\mu_i = \delta_i$; $Y_i \sim GPD(\delta_i, \zeta)$; and $Y_i \sim NB(r_i, p)$, where $r_i = \delta_i$.

S.3 In the case that the simulated marks are independent, sample a very long time series (i.e. $n = 20,000$) of random variables from $Y_{i,j} \sim F_{i,j}$, where $j \in \{1, \ldots, d\}$ and where the appropriate parameter of the distribution is specified by the autoregressive structure defined above.

S.4 In the case of $d \geq 2$ jointly dependent marks with copula C and marginal mark distributions $F_{i,1}, \ldots, F_{i,d}$, let $U_{i,1}, \ldots, U_{i,d}$ have joint distribution C. For a very long time series $i = 1 \ldots n$: simulate $u_{i,1}$ from the random variable $U_{i,1} \sim U[0,1]$; continue simulating, such that you simulate $u_{i,d}$ from the random variable $U_{i,d} \sim C_{i,d}(\cdot|u_{i,d-1})$; and evaluate $(y_{i,1}, \ldots, y_{i,d})$ from $F^{-1}(u_{i,1}), \ldots, F^{-1}(u_{i,d})$.

S.5 Calculate the long run empirical moments and correlation in the event of joint dependence which we will consider below and as required for the specification of the boost function, for example, $\mathbb{E}[Y_j]$, $Var[Y_j]$ and $\rho(Y_1, \ldots, Y_d)$.

Replacement of Step S.6 in Section D:

S.1 Define $t_i := \tau_r$ and $\lambda(i) := \lambda^-(r)$. In the case of univariate marks or multi-dimensional marks that are independent, sample a random variable from $X_{i,j} \sim F_{i,j}$, where $j \in \{1, \ldots, d\}$.

S.2 In the case of $d \geq 2$ jointly dependent marks with copula C and marginal mark distributions $F_{i,1}, \ldots, F_{i,d}$, let $U_{i,1}, \ldots, U_{i,d}$ have joint distribution C: simulate a random variable $U_{i,1}$ from $U[0,1]$ and let $u_{i,1} := U_{i,1}$; continue simulating a random variable $U_{i,d}$ from $C_{i,d}(\cdot|u_{i,d-1})$ and let $u_{i,d} := U_{i,d}$; and sample $(X_{i,1}, \ldots, X_{i,d})$ from $F^{-1}(u_{i,1}), \ldots, F^{-1}(u_{i,d})$. Define $x_{i,j} := X_{i,j}$

Replacement of Step S.8 in Section D:

S.1 Calculate the boost function $g(\mathbf{x}; \phi, \psi)$. However, for the normalization of the boost function we do not use the theoretical moments based on the marginal distribution,

we instead make use of the long run empirical moments above, $\mathbb{E}[Y_j]$, $Var[Y_j]$.

S.2 In the case of $d \geq 2$ jointly dependent marks recall that we have to make an adjustment to the normalization of the boost function for the evaluation of $\mathbb{E}[X_1 X_2] \neq 0$. Rather than calculating the linear correlation (explicitly or otherwise depending on the copula model specified), we utilize the long run empirical linear correlation $\rho(Y_1, \ldots, Y_d)$ from Step S.5.

F Copula models

For a definition of the Gaussian copula models that we consider in this work, see Cruz et al. [1]. Under such a framework, we consider developing the joint distribution of the marks random vector based on Sklars theorem, and where we may express the joint multivariate distribution as follows

$$F_{\boldsymbol{X}}(x_1, \ldots, x_d) = C\left(F_1(x_1), F_2(x_2), \ldots, F_d(x_d); \Upsilon\right), \tag{24}$$

for some choice of model for the copula C, parametrized by Υ and marginal distributions.

Note, we use

$$\begin{aligned}
\rho_S(X_1, X_2) &= 12 \int_{[0,1]} \int_{[0,1]} u_1 u_2 dC(u_1, u_2) - 3 \\
&= 12\mathbb{E}[U_1 U_2] - 3 \\
&= \frac{\mathbb{E}[U_1 U_2] - 1/4}{1/12} \\
&= \frac{\mathrm{Cov}\left[U_1, U_2\right]}{\sqrt{\mathrm{Var}[U_1]\mathrm{Var}[U_2]}} \\
&= \rho\left(F_1(X_1), F_2(X_2)\right),
\end{aligned} \tag{25}$$

hence, we need to introduce the mapping $\rho = \varphi(\rho_S)$ to adjust for the rank correlation versus the linear correlation.

In the case of the copula C, the off diagonal terms given when utilizing the bivariate Spearman's rank correlation for any copula C on the mark random vector are given by

$$\mathbb{E}\left[(X_1 - \mu_{1,1})(X_2 - \mu_{1,2})\right] = \sqrt{\mu_{2,1}\mu_{2,2}}$$

$$\varphi\left(\left[12 \int_{[0,1]} \int_{[0,1]} u_1 u_2 dC\left(u_1, u_2\right) - 3\right]\right), \tag{26}$$

for $\varphi(\rho_s) = \rho$, where ρ denotes the linear correlation coefficient between X_1 and X_2, $\varphi(\cdot)$ is the mapping and is known in closed form for many copula families.

The above central cross moments used the Spearman's rank correlation in terms of a copula density in a bivariate setting, given by

$$\rho_S(X_1, X_2) = 12 \int_{[0,1]} \int_{[0,1]} u_1 u_2 dC\left(u_1, u_2\right) - 3. \tag{27}$$

We utilize the rank transformation of the marks to obtain an expression for any copula for the Spearman's rank correlation from (27), to obtain the off diagonal terms explicitly.

For implementing the score statistic, the resulting asymptotic variance under the null of the linearly boosted covariance Σ_G^{-1} is given in general by

$$\Sigma_G = \begin{bmatrix} \mu_{2,1} & \mathbb{E}\left[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])\right] \\ & \mathbb{E}\left[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])\right] \quad \mu_{2,2} \end{bmatrix},$$

where $\mu_{2,1} = \mathbb{E}\left[(X_1 - \mathbb{E}[X_1])^2\right]$ and $\mu_{2,2} = \mathbb{E}\left[(X_2 - \mathbb{E}[X_2])^2\right]$. The off diagonal terms are evaluated using

$$\mathbb{E}[(X_1 - \mu_{1,1})(X_2 - \mu_{1,2})] = \rho\sqrt{\mu_{2,1}\mu_{2,2}},$$

where ρ is the pairwise linear correlation between X_1 and X_2. In the case of bivariate Gaussian copula with correlation parameter ρ, and using Cruz et al. [1, equation 10.51], the following relation is true

$$\rho\left(F_1(X_1), F_2(X_2)\right) = \varphi(\rho_S(X_1, X_2)) = 2\sin\left(\frac{\pi}{6}\rho_S(X_1, X_2)\right). \tag{28}$$

For copula models that do not have a closed form solution for calculating the linear correlation, such as the Gumbel and Clayton, we use a Monte-Carlo method as presented in Step S.2 of Section D. Alternative methods may include quadrature, or the use of Kendall's correlation

ρ_τ as a robust (but biased) estimator of the linear correlation. The linear correlation coefficient based on the covariance of the two variables is not preserved by copulas. However, the Kendall's correlation is a constant of the copula, see [1, Chapter 10].

G Simulation Experiments

G.1 Simulation methodology

As ψ increases the impact the mark has on the intensity is greater, and therefore we expect the power of the score test to approach 1. In the cases of a quadratic boost, the increase in ψ_1 leads to an equivalent increase in ψ_2. This facilitates the graphical representation of the power curves as we increase both boost function parameters. It is worth noting that the power curves will have some small degree of simulation variability, which is minimal, and smoothers are not applied to the power curves.

The empirical tail probabilities are calculated by estimating the proportion of simulated score statistics that exceed the nominal 1%, 5%, 10% upper tail probabilities from the $\chi^2_{(r)}$ distribution with r degrees of freedom. The size of the score test is evaluated as the probability of falsely rejecting the null hypothesis, and this should conform to the asymptotic $\chi^2(r)$ distribution. The 95% ranges expected around the nominal values are calculated as $p \pm 1.96\sqrt{p(1-p)/1,000}$, where p is the empirical exceedance probability. If we consider a single mark that is linearly boosted and set $\psi = 0$, the proportion of the 1,000 score statistic stimulants that exceed the $\alpha = 5\%$ upper tail probability for the $\chi^2_{(r=1)}$ distribution should be $\hat{\alpha} \approx 5\%$, implying that the empirical type I error matches the asymptotic chi-squared distribution.

When the dimension r of ψ is greater than 1, simulation of power against all combinations of the values $\psi_1, \ldots, \psi_r$ becomes computationally burdensome and graphical or tabular display of the results challenging. In the simulations presented here, we display the power of the test against alternatives on the 45^0 line $\psi_1 = \psi_2 = \psi$ with ψ increasing away from zero. While the choice of alternatives along a 45^0 line is acknowledged to be somewhat arbitrary, it treats both parameters equally in the examination of power. Of course, other restrictions expressing $\psi_1, \ldots, \psi_2$ in terms of a single ψ could be made, and the methods we present applied.

The algorithm to evaluate the size and power characteristics of the score test is:

- S.1 Specify the sample size as either an end time T or count of events n.
- S.2 Specify the dimension of the marks, form of the boost function, the marks distribution and associated parameters. In the event of serial dependence within the marks, specify the structure of serial dependence; see Section E for guidance. In the event of joint dependence between the marks, specify the copula model and parameters.
- S.3 Simulate the marked Hawkes process as per Section D and with adjustments in Section E for the serially dependent case under the alternative hypothesis. 1,000 replicates are simulated for each ψ starting with the null $\psi = 0$. Under the null case, the boost function will be equal to 1.
- S.4 For each replicate estimate the decay parameter α under the null using Section C. Calculate the score statistic as outlined in [9, Section 5].
- S.5 For each $\psi = [0, 0.01, 0.02, \ldots, 1]$ we will have 1,000 score statistics. For each ψ count, the proportion of the score statistics that exceed the nominal 1%, 5%, 10% upper tail probabilities from the $\chi^2_{(r)}$ distribution with r degrees of freedom.
- S.6 To create the power curves, plot the proportions against the ψ. The size properties are assessed at $\psi = 0$ under the null, and the power properties are presented in the power curve.

G.2 Size and power of the score test

G.2.1 Objective

To study the size and power of the score test, we consider different combinations of marks dimensions $d \in \{1, 2, 4\}$, boost function, and marks density. This will provide an assessment of the accuracy of the asymptotic chi-squared distribution for determining size of the score test under the null hypothesis that $\psi = 0$, for a variety of data generating mechanisms (DGM).

G.2.2 DGM

We consider both a linear boost $1 + \sum_1^r \psi_j X_j$ where $r = d$, and quadratic boost $1+\psi_1 X_1^2+\psi_2 X_2^2$ where $d = 1, r = 2$. We consider different combinations of marks dimensions $d \in \{1, 2, 4\}$, which are combined multiplicatively, and estimated theoretical moments for the evaluation of $G(\mathbf{X}; \phi)$.

Concerning the existence of moments of the generalized Pareto distribution, for the score test against a linear boost function, we use $\zeta = 0.25$, which guarantees that $\mathbb{E}[X^r] < \infty$ for $r = 1/\zeta < 4$ but the fourth moment is not finite. For a linear boost, the existence of the second moment of the marks distribution is required for the score test statistic to be correctly defined and for Theorem 1 [9, Section 4] to hold. For quadratically boosted generalized Pareto distributed marks, the fourth moment is required to exist for Theorem 1 [9, Section 4] to be valid. For this first study we use $\zeta = 0.24$, which is slightly smaller than the required fourth moment to exist.

The marks distributions considered and associated simulation parameters are:

- Poisson distribution, denoted Pois($\mu = 1.00$) with μ average counts;
- Exponential distribution with density, $f(x; \lambda)$ and denoted Exp($\lambda = 1.00$);
- Generalized Pareto distribution with location parameter 0, $f(x; \zeta, \delta)$ with shape parameter ζ and scale parameter δ, denoted GPD($\zeta = 0.25, \delta = 1.00$) for a linear boost and GPD($\zeta = 0.24, \delta = 1.00$) for a quadratic boost.

We also consider a pair of linearly boosted dependent marks, $X_1 \sim \text{GPD}(\zeta = 0.24, \delta = 1.00)$ and $X_2 \sim \text{GPD}(\zeta = 0.24, \delta = 1.00)$, with dependence between them modeled using a Gaussian copula with Spearman's correlation, $\rho_s = 0.8$. The final simulation study considers four independent, linearly boosted marks, with $X_j \sim \text{GPD}(\zeta_j = 0.24, \delta_j = 1.00)$ for $j \in \{1, \ldots, 4\}$. All three sample sizes are considered in this study, $T = 130$, $T = 300$ and $T = 1,000$.

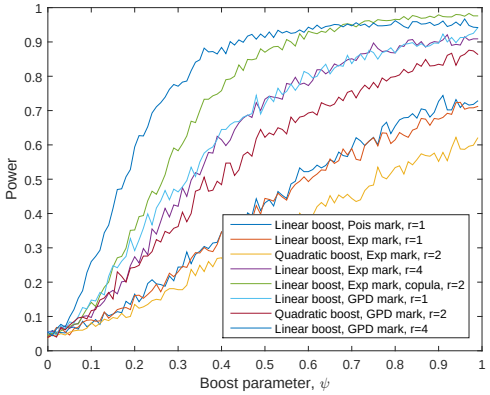

Fig. 1 $n = 500$: Power curves for sample size of $n = 500$ for a selection of cases: $X_i \sim \text{Pois}(\mu = 1.00)$, with linear boost; $X_i \sim \text{Exp}(\lambda = 1.00)$, with linear boost for i.i.d. marks and jointly dependent marks (Gaussian copula model, $\rho_s = 0.8$) and quadratic boost; $X_i \sim \text{GPD}(\zeta = 0.25, \delta = 1.00)$, with linear boost; and $X_i \sim \text{GPD}(\zeta = 0.24, \delta = 1.00)$, with quadratic boost.

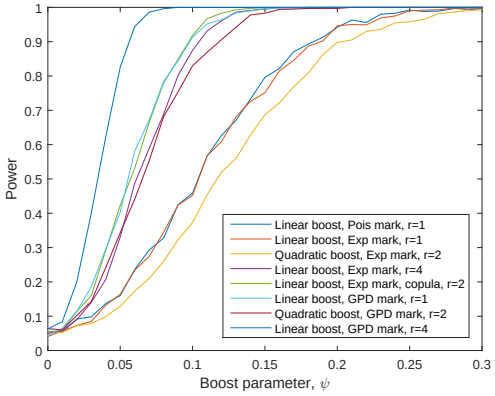

Fig. 2 $n = 4,000$: Power curves for sample size of $n = 4,000$ for a selection of cases: $X_i \sim \text{Pois}(\mu = 1.00)$, with linear boost; $X_i \sim \text{Exp}(\lambda = 1.00)$, with linear boost for i.i.d. marks and jointly dependent marks (Gaussian copula model, $\rho_s = 0.8$) and quadratic boost; $X_i \sim \text{GPD}(\zeta = 0.25, \delta = 1.00)$, with linear boost; and $X_i \sim \text{GPD}(\zeta = 0.24, \delta = 1.00)$, with quadratic boost.

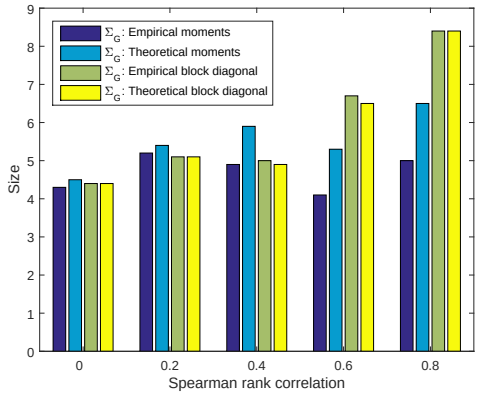

Fig. 5 Simulated upper tail probabilities of the score test using theoretical and empirical moments, with calculation of cross terms and imposed block diagonalize of the covariance matrix. We consider the score test for a linear boost, marks with a GPD and joint dependence modeled by a Gaussian copula, with Spearman's rank correlation $\rho_s = \{0, 0.2, 0.6, 0.8\}$. The sample size is $n = 4,000$.

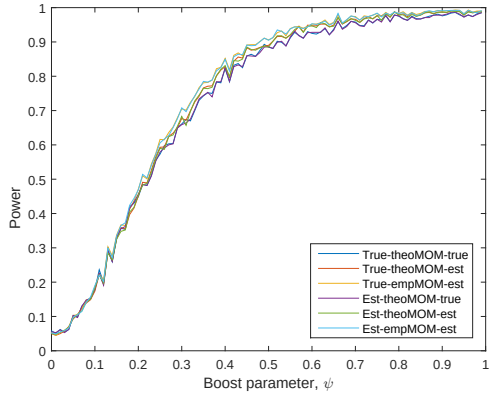

Fig. 3 $n = 500$: Power of the score test statistic for a linear boost and marks with an exponential distribution $X \sim \text{Exp}(\lambda = 1.00)$. This compares the use of theoretical moments with empirical moments in the score statistic, as well as true parameters and estimated parameters in the intensity function and the marks distribution.

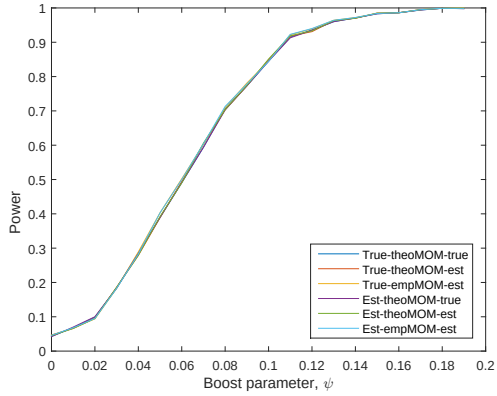

Fig. 4 $n = 4,000$: Power of the score test statistic for a linear boost and marks with an exponential distribution $X \sim \text{Exp}(\lambda = 1.00)$. This compares the use of theoretical moments with empirical moments in the score statistic, as well as true parameters and estimated parameters in the intensity function and the marks distribution.

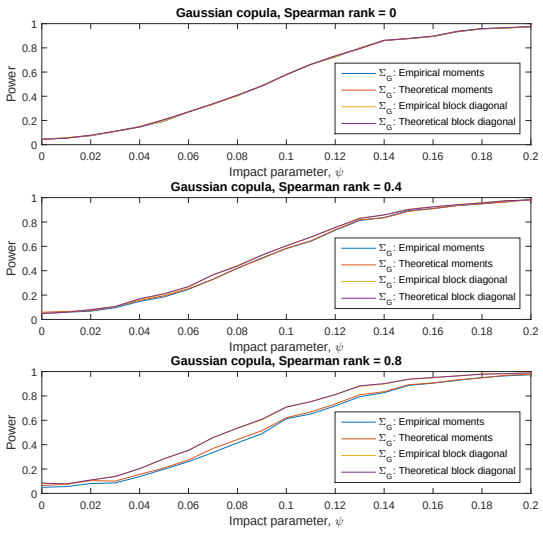

Fig. 6 Power of the score test for a linear boost, marks with a GPD and joint dependence modeled by a Gaussian copula, with Spearman's rank correlation, $\rho_s = \{0, 0.4, 0.8\}$. Comparing the use of theoretical moments and empirical moments, with calculation of cross terms and imposed block diagonalize of the covariance matrix in the score statistic. The sample size is $n = 4,000$.

G.3 Simulation Experiments with Extensions

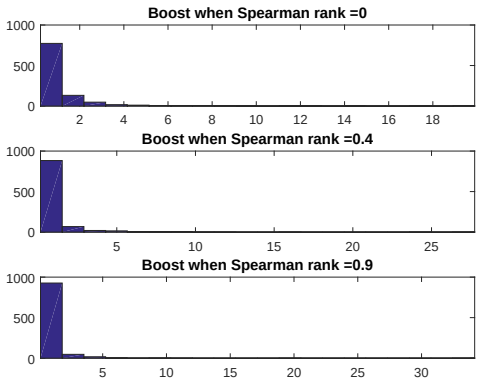

Fig. 7 Histograms of the linear boost function for a single simulant, jointly dependent marks with Spearman's rank $\rho_s \in \{0, 0.4, 0.9\}$. The joint dependence is modeled by a Gaussian copula. The marks are serially dependent, with a conditional GPD $X \sim \text{GPD}(\zeta = 0.10, \delta_i)$, where $\delta_i = 0.9\delta_{i-1} + \epsilon_i$. The sample size is $n = 4,000$.

H Score test of higher dimensional marks

H.1 Defining the mark vector

The endogenous mark vector that we consider for this model is derived from the information that is present in the reported limited order book, presented in Richards [8, Chapter 3]. We keep this section brief and refer the interested reader to the aforementioned paper for details.

The reported limit order book contains financial market exchanged traded limit order book data with a minimum event time granularity of one millisecond. The event process was defined on this time granularity, with an event e being any order type, market order (MO), limit order (LO) or cancellation (C) ($e \in \{MO, LO, C\}$) on any level of the limit order book, on either the bid (B) or ask (A) side.

Key to the formation of the random mark vector is the underlying event process under consideration. In the context of this research, we can define the event process based on three key factors price level $l \in \{1, \ldots, 10\}$, side $s \in \{B, A\}$ and event type $e \in \{LO, MO, C\}$.

Table 1 Marks constructed from matched LOB data [8], with numerical references for Figure 8.

Depth based:	**Centred depth/price based:**	**Price based:**
B.depth (1)	Imb. (9)	Rel.Price.LO (12)
B.opp.depth (2)	Rel.Imb (10)	Rel.Price.C (13)
Volume based:	Midprice (11)	**Count based:**
B.LOMOCV (3)	MidpriceRet (14)	B.CountMOLOC (19)
B.MOLOV (4)	Spread (15)	B.CountLO (20)
B.MOV (5)	MOprice (16)	B.CountC (21)
B.LOV (6)	Vol.midprice (17)	
B.CV (7)	VolmidpriceRet (18)	
Inside.MOV (8)		

The content below provides a clear description of each of the marks.

H.1.1 Volume based marks

B.depth. The volume depth of the LOB mark is the sum of the volume at t_i across all levels of the bid side.
B.opp.depth. The volume depth of the opposite side of the book is the sum of the volume on the opposing side to what is being modelled at t_i. If we consider a bid side process, this mark reflects the depth volume on the ask side at t_i, and conversely for the ask side process.

A number of marks are constructed from the volume associated with event types $e \in \{MO, LO, C\}$ and combinations thereof, across specified levels $l \in \{1, \ldots, 10\}$, for each side $s \in \{B, A\}$. For these marks, we are not calculating the sum of volumes on the actual LOB, but rather the volume associated with an event type. For this research, the volumes for each event type are aggregated across the price levels, which are specified when defining the event process. The specific event type combinations we consider are listed below.

- *B.LOMOCV.* The volume of all events $e \in \{MO, LO, C\}$ at t_i
- *B.MOLOV.* The volume of events $e \in \{MO, LO\}$ at t_i.
- *B.MOV.* The volume of event $e \in \{MO\}$ at t_i.
- *B.LOV.* The volume of event $e \in \{LO\}$ at t_i.
- *B.CV.* The volume of event $e \in \{C\}$ at t_i.

Inside.MOV . The volume that has transacted within the spread, not originating on either the bid or ask.

H.1.2 Volume imbalance marks

Imb. The first volume imbalance we consider is defined in Gould and Bonart [3] and described as queue imbalance in a LOB.
Rel.Imb. The relative depth profile is defined in [4], whereby level l is used to determine the normalized weight.

H.1.3 Price based marks

Midprice. The mid-price is definied as the mid-point between the bid and ask at level l at t_i
MidpriceRet. The return of the mid-price from t_{i-1} to t_i.
Spread. The spread is the difference between the ask and the bid at $l = 1$ at t_i.
MOprice. Is simply the traded (event type MO) at t_i.
Vol.midprice. The volatility of the mid-price across the previous 100 events.
VolmidpriceRet. The volatility of the mid-price returns across the previous 100 events.

The relative price of an event, refers to the price of the event on the bid or ask side relative to the best bid or ask, respectively. If there are say two LO events at level 3 and level 2 at event time, and for the purposes of constructing a mark, we need to summarize this information into a single relative price. To address this, we have taken the mean across the levels when more than one event occurs at an event time. For further details, refer to Richards [8, Chapter 3].
Rel.Price.LO. The relative price of event $e \in \{LO\}$.
Rel.Price.C. The relative price of event $e \in \{C\}$.

H.1.4 Count based marks

Similarly to the relative price calculations, to define the count based marks that follow, we use LOB data prior to aggregation. The formulation below, counts the number of events that contribute to the unique time-stamped event.
B.CountMOLOC . The count of all events $e \in \{MO, LO, C\}$ is the sum of events that make

up a single aggregated unique event time across specified levels $l \in \{1, \ldots, 10\}$.
B.CountLO. The count of events $e \in \{LO\}$.
B.Countc. The count of events $e \in \{LO\}$.

Richards [8] presents the strong pairwise correlation between the marks. This correlation could impact the effectiveness of the score test for detecting marks and needs to be accounted for in the score test. Figure 8 presents the proportion of significant bivariate marks across the time segments of 6 minutes for SILVER. The pairs of marks are indexed by i and j, where $i \in \{1, \ldots, 20\}$ and $j \in \{2, \ldots, 21\}$. Table 1 presents the names corresponding to the numerical values i and j can take. The combinations of pairs of marks considered are $\{(1,2), (1,3), \ldots, (2,3), (2,4), \ldots, (20,21)\}$. As with the heat charts in the case of single marks, a lighter color represents a higher proportion of times the bivariate marks are significantly different from the null hypothesis.

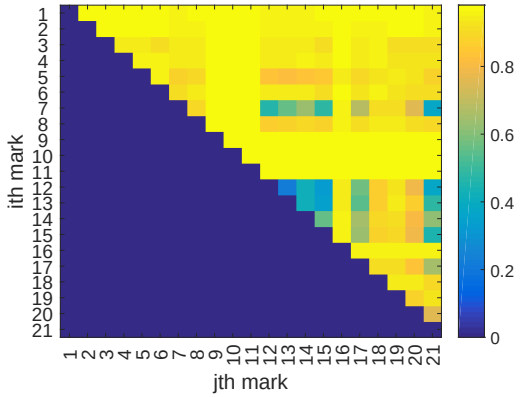

Fig. 8 *Pairwise marks.* The proportion of time segments, which pairs of marks are significant via the score test, for time segments of 6 minutes, for SILVER, across 10 trading days from 20-July-2015 to 31-July-2015. Combinations of $\{i, j\}$ marks are $i \in \{1, \ldots, 20\}$ and $j \in \{2, \ldots, 21\}$. Table 1 shows the names corresponding to the numerical values i and j. Marks are constructed using matched bid side LOB data, with event types $e \in \{LO, MO, C\}$ and levels $l \in \{1, \ldots, 5\}$.

It is not possible to plot all combinations of marks, however, to demonstrate the flexibility of the score test, we present the proportion of significant marks, which are assessed jointly and are sequentially introduced in an increasing number. A test of this kind would not be computationally feasible if the assessment was conducted with a model of the log-likelihood, again highlighting the significant practical application of the score test.

Figure 9 displays the proportion of significant marks of dimension $d \in \{2, \ldots, 21\}$, across the defined time segments of 6 minutes for SILVER. The marks are indexed by i and j with Table 1 presenting the names corresponding to the numerical values of i and j. The combinations of marks considered are $\{(1,2), (1,2,3), \ldots, (1, \ldots, 21), \ldots, (2,3), (2,3,4), \ldots, (2, \ldots, 21), \ldots, (20,21)\}$. When the mark, inside MO Volume, is introduced into the sequence of marks, the sequence reduces in significance from approximately 94.8% of time segments to 47.5% of time segments (Figure 9). Combinations of relative price of LO and C and mid-price returns have contributed to a reduction of the number of time segments that are significant for SILVER (Figure 9).

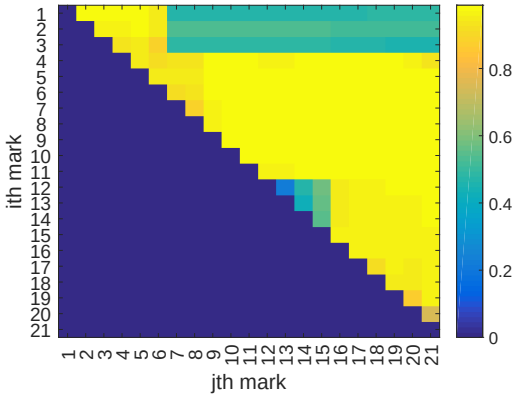

Fig. 9 *Multiple marks.* The proportion of significant marks via the score test, for time segment $n = 1,000$, for SILVER, across 10 trading days, 20-July-2015 to 31-July-2015, for combinations j:i of marks. Marks are constructed using matched bid side LOB data, with event types $e \in \{LO, MO, C\}$ and levels $l \in \{1, \ldots, 5\}$. Table 1 shows the names corresponding to the numbers in the chart.

Throughout the intensity, function is defined in terms of an exponential decay $w(t; \alpha) = \alpha \exp(-\alpha t)$ and the parameters $\theta = (\eta, \vartheta, \alpha)$ are estimated by maximizing the log-likelihood (18) for the unmarked Hawkes process. All simulations were carried out using MATLAB and further details of the simulation methods are provided in the Appendices. The samples sizes n used in the simulations were: $n = 300$; $n = 1,000$; and $n = 4,000$. The number of replications simulated at

each configuration of model parameters and marks distribution is 1,000. In Section H.3 we consider the size and power of the score test for different combinations of marks dimension $d \in \{1, 2, 4\}$, with linear and quadratic boosts. To evaluate the derivative of the boost function in Equation 12 [9, Section 3] we calculate the empirical or estimated theoretical moments, comparing both methods. We investigate the impact of serial dependence with a proposed correction to the score test and the impact of joint dependence between marks. Finally, in Section H.6 we investigate an extended simulation study that encapsulates features of the real LOB data that is subsequently used in the application of the score test.

H.2 Simulation Methodology

A brief summary of the simulation methodology used is reported here with more extensive detail available in the Appendices and [8]. For any given specification of the marked Hawkes process with intensity (1), simulation of it uses an extension to Ogata's modified thinning algorithm [7] to allow for the inclusion of marks $\mathbf{X}$ with joint dependence between components of via a copula specification serial dependence in the marks. Further details are presented in Section E. Note [6] also presents such a method for the case of i.i.d. vector marks.

For simulation of size ($\psi = 0$) and power against an alternative $\psi > 0$, fresh stimulants need to be generated for each value of ψ and the other parameters θ and ϕ because each combination, through the thinning algorithm, generates different event times. However, the marks process is reused across different levels of the boost. In order to control simulation error we used 1,000 simulation replicates per combination of parameters. This yielded sufficiently accurate results for our comparisons and assessments. For example, to simulate a single power curve for each $\psi \in \{0, 0.01, 0.02, \ldots, 1\}$, 1,000 replicates of n events were simulated. In the studies where two parameters vary (Section H.6), for example, $\psi \in \{0, 0.01, 0.02, \ldots, 1\}$ and $\rho_s \in \{0, 0.4, 0.8\}$, where ρ_s is the Spearman's rank correlation, each combination of the two parameters is simulated 1,000 times. See Section E for further details on the simulation methodology.

H.3 Size and Power of Score Test: i.i.d. marks

In this subsection, empirical sizes are calculated for the score test using i.i.d. scalar or vector marks with a variety of distributions and for three sample sizes $n \in \{300, 1000, 4000\}$. We also report on the power of the score test as ψ increases away from $\psi = 0$. The intensity parameters are set to $\eta = 0.8$, $\vartheta = 0.8$, and $\alpha = 2.1$, which represent reasonable parameter values that give balance between the immigration intensity and branching ratio.

We explore the use of linear boosts $h(X; \psi) = 1 + \psi X$ and quadratic boost $h(X; \psi_1, \psi_2) = 1 + \psi_1 X + \psi_2 X^2$ for scalar marks. For vector marks $\mathbf{X} = (X_1, \ldots, X_d)$, $d > 1$, linear boosts are combined multiplicatively as $h(\mathbf{X}) = \prod_1^d (1 + \psi_i X_i)$. The marks are i.i.d. with distribution selected from: the Poisson distribution Pois($\mu = 1.00$); the exponential distribution Exp($\lambda = 1.00$); or the generalized Pareto distribution GPD(ζ, δ) with location parameter 0, shape parameter ζ and scale parameter δ. Note that for the exponential and Poisson distributions, all moments are finite, so the score statistic and the sufficient conditions for Theorem 1 [9, Section 4] are satisfied for the linear and quadratic boost functions considered here. For the GPD, this depends on the value of ζ and the selection of boost since $\mathbb{E}[X^K] < \infty$ for $K < 1/\zeta$. Details are provided below.

We consider the following scenarios:

Linear boost in a scalar mark X with the exponential, Poisson or GPD($\zeta = 0.25, \delta = 1.00$) distribution. Here degrees of freedom are $r = 1$. Note that for the GPD case $\mathbb{E}[X^2] < \infty$ so the score statistic is well defined but that $\mathbb{E}[X^4]$ just fails to be finite so the Condition 3 [9, Section 4], one of the sufficient conditions for Theorem 1 [9, Section 4], does not hold;

Quadratic boost in a scalar mark X with the exponential or GPD($\zeta = 0.24, \delta = 1.00$) distribution. Here degrees of freedom are $r = 2$. For the GPD case, $\mathbb{E}[X^4] < \infty$ and the score test is well defined but Condition 3 [9, Section 4] requires $\mathbb{E}[X^8]$ in this case which does not hold;

Bivariate mark, $d = 2$**,** $\mathbf{X} = (X_1, X_2)$**, linear boosts, combined multiplicatively** with X_1 and X_2 exponentially distributed and dependence between them modeled using a Gaussian copula with Spearman's correlation, $\rho_s = 0.8$

(details in Section F). Here degrees of freedom are $r = 2$;

Multivariate marks, $d = 4$, linear boosts, combined multiplicatively with $X_j \sim$ i.i.d. GPD($\zeta_j = 0.24, \delta_j = 1.00$) for $j \in \{1, \ldots, 4\}$. Here degrees of freedom are $r = 4$. Note that $\mathbb{E}[X_i^4] < \infty$ so that the conditions for Theorem 1 [9, Section 4] are satisfied.

The combinations selected are shown in Figure 10, along with the empirical tail probabilities obtained using the proportion of simulated score statistics exceeding the nominal $1\%, 5\%, 10\%$ upper tail probabilities from the $\chi^2_{(r)}$ distribution, with r degrees of freedom. The degrees of freedom depends on the combination of options selected, $r \in \{1, 2, 4\}$ and are listed in Figure 10. Almost all simulated probabilities are within the 95% ranges expected around the nominal values. There is a slight bias above the nominal values but no obvious pattern to the individual variations with a combination of sample size represented by $\log(n)$ on the horizontal axis. A selection of power curves corresponding to Figure 10 for sample sizes $n = 500$ and $n = 4,000$ are presented in Figures 11 and Figure 12, respectively. As expected, power increase monotonically as ψ increases and with larger sample size. Whilst there is some degradation in power for the smaller sample size, the power properties for the score test perform well for most cases. The light-tailed marks distributions present very similar power properties, whereas the marks with the heavy-tailed generalized Pareto distribution have stronger power properties. We see a reduction in power performance for the quadratic boost function compared with the linear boost. Marks that exhibit joint dependence and are modeled via a copula model present stronger power properties.

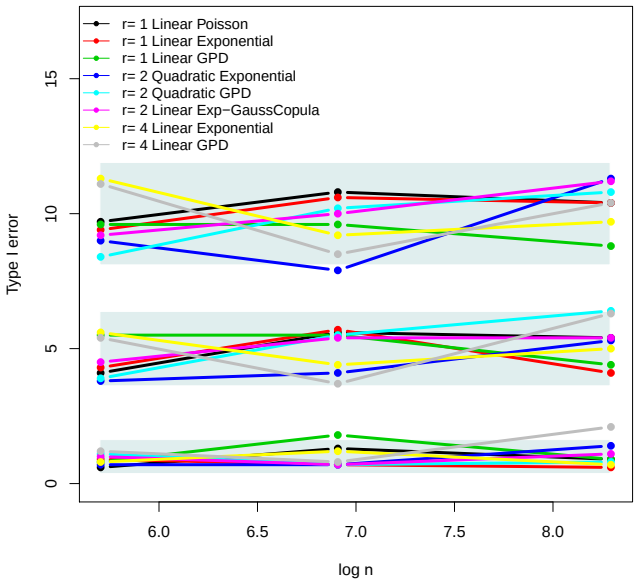

Fig. 10 Simulated upper tail probabilities of score test under a range of data generating mechanisms. The light blue bands reflect 95% ranges around the nominal chi-squared upper $1\%, 5\%, 10\%$ type I errors.

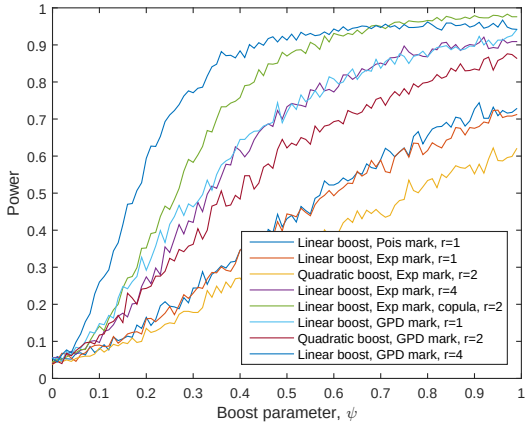

Fig. 11 Power curves for sample size of $T = 130$ for a selection of cases: $X_i \sim \text{Pois}(\mu = 1.00)$, with linear boost; $X_i \sim \text{Exp}(\lambda = 1.00)$, with linear boost for i.i.d. marks and jointly dependent marks (Gaussian copula model, $\rho_s = 0.8$) and quadratic boost; $X_i \sim \text{GPD}(\zeta = 0.25, \delta = 1.00)$, with linear boost; and $X_i \sim \text{GPD}(\zeta = 0.24, \delta = 1.00)$, with quadratic boost.

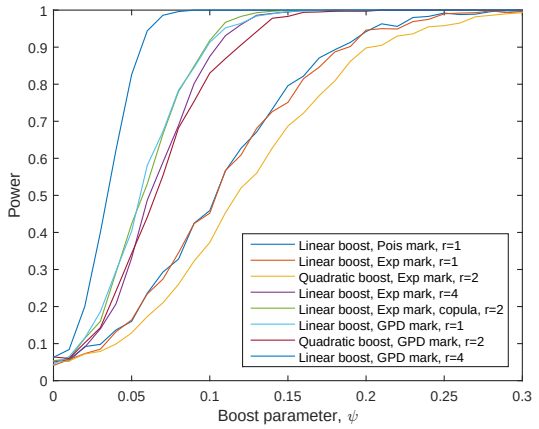

Fig. 12 Power curves for sample size of $T = 1,000$ for a selection of cases: $X_i \sim \text{Pois}(\mu = 1.00)$, with linear boost; $X_i \sim \text{Exp}(\lambda = 1.00)$, with linear boost for i.i.d. marks and jointly dependent marks (Gaussian copula model, $\rho_s = 0.8$) and quadratic boost; $X_i \sim \text{GPD}(\zeta = 0.25, \delta = 1.00)$, with linear boost; and $X_i \sim \text{GPD}(\zeta = 0.24, \delta = 1.00)$, with quadratic boost.

We investigated the impact on the score test using empirical moments versus estimated theoretical moments for the marks distribution by considering a linear boost with an exponentially distributed mark $X \sim \text{Exp}(\lambda = 1.00)$. The intensity function $\lambda(t; \nu)$ was calculated under the null using two options: the true parameters ν; and the estimated parameters $\hat{\nu}$ fitted under the null. For the estimation of moments used in calculating $G(\mathbf{X}, \phi)$ and estimating the covariance matrix Ω_n, three options were used: empirical moments; theoretical moments evaluated using the true values of ϕ; and using the maximum likelihood estimates $\hat{\phi}$. For this study, the resulting six combinations of methods were applied to the identical $1,000$ replications used for each parameter value in the power curves helping to further control simulation variability between methods.

For the six combinations described Section G and sample sizes $n = 300$ and $n = 4,000$, no discernible difference was detected between the power curves for the statistic calculated using sample moments and calculated using estimated theoretical moments. The results suggest, at least in the cases investigated here and in Richards [8, Section 6.6], that the use of empirical moments compared to estimated theoretical moments does not degrade the size or power of the score statistic, and this also holds for smaller sample sizes. Knowing the true parameter values does not appear to improve size and power characteristics compared with estimated parameters. The advantage of estimating the moments empirically allows one to avoid estimating a parametric model for the marks distribution assuming that the moments exist.

H.4 Impact of Ignored Serial Dependence in Marks

We briefly confirm that failure to account for positive serial dependence in the score statistic has the expected inflation of size. The linearly boosted scalar marks are conditionally GPD $X_i \sim \text{GPD}(\zeta = 0.10, \delta_i)$, with scale parameter from an autoregressive process $\delta_i = 0.9\delta_{i-1} + \epsilon_i$ and $\epsilon_i \sim$ i.i.d. $\text{N}(0, 1)$. Moments serial autocovariances are estimated empirically since, for this marks process, theoretical moments are not available in closed form. Figure 13 and Figure 14 shows the sampling distribution of the score statistic using the correct $\hat{\Omega}_{K,n}$ from Richards et al. [9, Equation 38] when the marks are serially dependent, compared with the score statistic using $\hat{\Omega}_{0,n}$ from Richards et al. [9, Equation 37], which is incorrect since it assumes marks are serially uncorrelated.

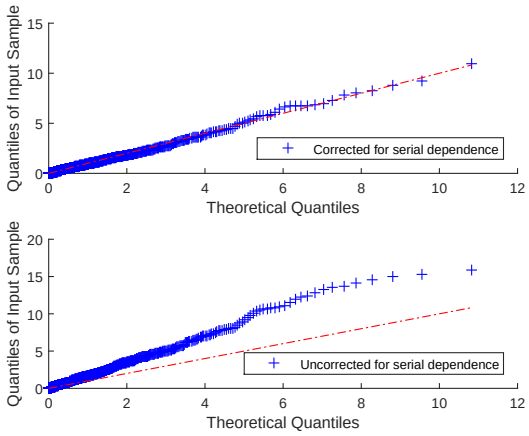

Fig. 13 Impact on the sampling distribution of the score test statistic, from adjusting and not adjusting for serial dependence, with a sample size $n = 4,000$. The marks are conditionally GPD, $X_i \sim \text{GPD}(\zeta = 0.10, \delta_i)$, with a linear boost and empirically estimated moments.

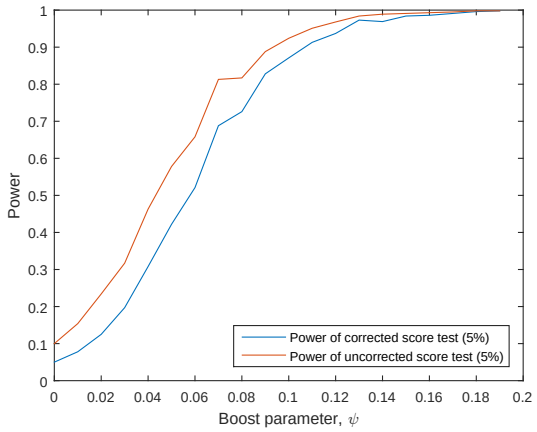

Fig. 14 Impact on the power curves of the score test statistic, from adjusting and not adjusting for serial dependence, with a sample size $n = 4,000$. The marks are conditionally GPD, $X_i \sim \text{GPD}(\zeta = 0.10, \delta_i)$, with a linear boost and empirically estimated moments.

Given the serial dependence is positive and reasonably persistent, the uncorrected score statistic is inflated because the variance obtained, ignoring the extra terms due to serial dependence, is too small. This is clear from Figure 13 in which the uncorrected statistic is substantially stochastically larger than the $\chi^2_{(1)}$ distribution, while the corrected score statistic conforms to the claimed asymptotic distribution very well. This results in the uncorrected score statistic realizing an actual size that is substantially larger than that from the $\chi^2_{(1)}$ distribution, and hence the power curve is substantially inflated (Figure 14).

H.5 Impact of Ignored Joint Dependence in Marks

To study the impact of ignoring joint dependence between marks on the score test, we consider a pair of linearly boosted, jointly dependent i.i.d. marks $X_1 \sim \text{GPD}(\zeta = 0.10, \delta = 1.00)$ and $X_2 \sim \text{GPD}(\zeta = 0.10, \delta = 1.00)$. The joint dependence is modeled using a Gaussian copula (Section F) with four different levels of dependence, $\rho_s \in \{0, 0.2, 0.4, 0.8\}$. For implementing the score statistic, both empirical and estimated theoretical moments are used. To assess the impact of ignoring joint dependence, that is, assuming $\rho_s = 0$ when in fact, $\rho_s > 0$, the estimated covariance matrix $\hat{\Sigma}_G$ for the bivariate marks is constrained to be diagonal.

Table 2 Upper tail probabilities for the nominal chi-squared upper 5% type I errors for spearman rank correlation, ρ_s.

Σ_G	$\rho_s = 0$	$\rho_s = 0.2$	$\rho_s = 0.4$	$\rho_s = 0.6$	$\rho_s = 0.8$
Empirical moments	0.043	0.052	0.049	0.041	0.050
Theoretical moments	0.045	0.054	0.059	0.053	0.065
Emp. block diagonal	0.044	0.051	0.050	0.067	0.084
Theo. block diagonal	0.044	0.051	0.049	0.065	0.084

Table 2 and graphically shown in Section G presents the impact of not accounting for the cross correlation terms by displaying the proportion of simulated scores statistics exceeding the nominal 5% upper tail probabilities from the $\chi^2_{(r)}$ distribution with $r = 2$ degrees of freedom. The score statistic when the covariance Σ_G is falsely assumed to be block diagonal is increasingly inflated relative to the $\chi^2_{(2)}$ distribution as the Spearman's rank correlation increases. This was further supported by a study of the power of the score test under three scenarios whereby the strength of dependence between the marks is chosen to be $\rho_s \in \{0, 0.4, 0.8\}$, presenting a shift up in the power curve for the block diagonal scenarios (Section G).

H.6 Increasing Joint Dependency of Bivariate, Serially Dependent Marks

To complete the analysis of the power properties of the score test, we investigate the effect on the power of the score test of serially dependent, bivariate, heavy-tailed marks with joint dependence, all of which are features considered in the subsequent application utilizing the LOB data set (see Richards [8] for properties of the marks). The parameter specification is chosen to match closely what is observed in the LOB data, setting the intensity parameters to: $\eta = 0.0010$; $\vartheta = 0.7000$; and $\alpha = 0.0100$. The marks are combined and multiplicative, each with a linear boost. They are distributed with a generalized Pareto distribution

$X_i \sim \text{GPD}(\zeta = 0.10, \delta_i)$. The conditional generalized Pareto distribution scale parameter is defined as $\delta_i = 0.9\delta_{i-1} + \epsilon_i$. The joint dependence is modeled via a Gaussian copula model with varying Spearman's rank correlation $\rho_s \in \{0, 0.1, \ldots, 1\}$. We utilize empirical moments, and the simulation sample size is $n = 1,000$.

Figure 15 exhibits strong size and power characteristics for the score test, and as the correlation of the copula increases, so does the power of the score test. The impact to the boost function of combining serial and joint dependent marks in a multiplicative form leads to increased realizations in the tails of the histogram of the boost function as the joint correlation increases (Figure 16). This provides assurance that the score test is reliable for using marks constructed from LOB data used in financial applications and, in addition, robust to such features that may exist in other applications.

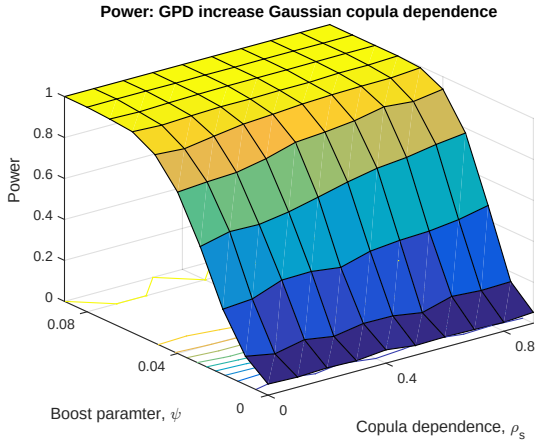

Fig. 15 Power of the score test, with a linear boost, bivariate marks with Gaussian copula dependence, assuming serially dependent marks with a conditional GPD $X \sim \text{GPD}(\zeta = 0.10, \delta_i)$, where $\delta_i = 0.9\delta_{i-1} + \epsilon_i$. The sample size is $n = 1,000$.

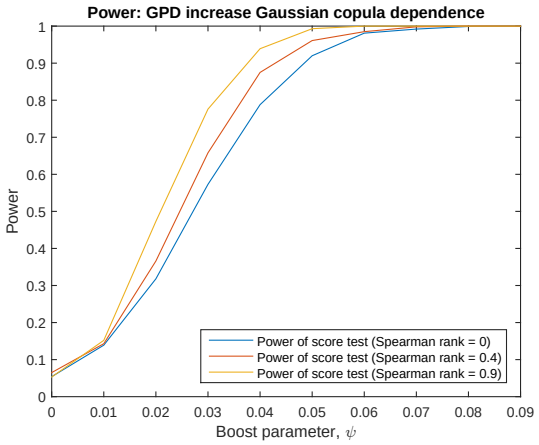

Fig. 16 Power of the score test, with a linear boost, bivariate marks with Gaussian copula dependence, assuming serially dependent marks with a conditional GPD $X \sim \text{GPD}(\zeta = 0.10, \delta_i)$, where $\delta_i = 0.9\delta_{i-1} + \epsilon_i$. The sample size is $n = 1,000$.

References

[1] Cruz, Marcelo G., Gareth W. Peters, and Pavel V. Shevchenko. 2015. *Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk.* John Wiley & Sons, Inc.

[2] Diggle, Peter, P. J. Heagerty, K. Y. Liang, and S. L. Zeger. 2002. *Analysis of Longitudinal Data.* Oxford Statistical Science Series. OUP Oxford.

[3] Gould, Martin D. and Julius Bonart. 2016. Queue imbalance as a one-tick-ahead price predictor in a limit order book. *Market Microstructure and Liquidity 2*(2) .

[4] Gould, Martin D., Mason A. Porter, Stacy Williams, Mark McDonald, Daniel J. Fenn, and Sam D. Howison. 2013. Limit order books. *Quantitative Finance 13*(11): 1709–1742 .

[5] Laub, Patrick J., Thomas Taimre, and Philip K. Pollett. 2015. Hawkes processes. *arXiv preprint arXiv:1507.02822* .

[6] Liniger, Thomas 2009. *Multivariate Hawkes Processes.* Ph. D. thesis, ETH Zürich (ethz-a-006037599).

[7] Ogata, Yosihiko. 1981. On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory 27*(1): 23–31 .

[8] Richards, Kylie-Anne L. 2019. *Modelling the Dynamics of the Limit Order Book in Financial Markets*. Ph. D. thesis, University of New South Wales (1959.4/61579).

[9] Richards, Kylie-Anne L., William T. M. Dunsmuir, and Gareth W. Peters. 2022. Score Test for Marks in Hawkes Processes. *Working Paper SSRN: https://ssrn.com/abstract=3381976* .