Asymptotic distribution of the score test for detecting marks in Hawkes processes

Simon Clinet · William T. M. Dunsmuir · Gareth W. Peters · Kylie-Anne Richards

Received: 1 July 2020 / Accepted: 13 May 2021 / Published online: 19 May 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract

The score test is a computationally efficient method for determining whether marks have a significant impact on the intensity of a Hawkes process. This paper provides theoretical justification for use of this test. It is shown that the score statistic has an asymptotic chi-squared distribution under the null hypothesis that marks do not impact the intensity process. For local power, the asymptotic distribution against local alternatives is proved to be non-central chi-squared. A stationary marked Hawkes process is constructed using a thinning method when the marks are observations on a continuous time stationary process and the joint likelihood of event times and marks is developed for this case, substantially extending existing results which only cover independent and identically distributed marks. These asymptotic chi-squared distributions required for the size and local power of the score test extend existing asymptotic results for likelihood estimates of the unmarked Hawkes process model under mild additional conditions on the moments and ergodicity of the marks process and an additional uniform boundedness assumption, shown to be true for the exponential decay Hawkes process.

Keywords Marked Hawkes point process · Ergodicity · Quasi likelihood · Score test · Inferential statistics · Local power
1 Introduction

Since their introduction over fifty years ago Hawkes self exciting process models (Hawkes 1971) have been used to model point processes in many fields of application including seismology (Ogata 1988), sociology (Crane and Sornette 2008), modelling neuronal systems and increasingly in recent years for modelling high frequency financial trading (for a general review, see Bacry et al. (2015), Hawkes (2018)). Extensions of the Hawkes process where parameters are time-varying and replicate the non-stationarity of intraday financial data have also been considered in Chen and Hall (2013), Clinet and Potiron (2018) for example. The theoretical properties of such models are quite well advanced, as is estimation methodology and its associated statistical theory. Increasingly, marked Hawkes processes in which marks attached to past event times influence future intensities, are being considered for a range of applications; in particular, in recent times, to modelling high frequency financial data. For example, Richards et al. (2019) develops marked Hawkes processes for modelling millisecond recordings of activity in the limit order book for a range of assets traded on international futures markets. In these applications there are numerous potential marks that are recorded at each event and it is helpful to efficiently screen out those that are not influential on future event arrival intensities before the joint models for the event times and associated marks are estimated.

Assessment of influential marks could be done, in theory, by simultaneously estimating the parameters of the marked Hawkes process using maximum likelihood methods and assessing the mark impact parameters for statistical significance using standard inferential techniques such as the Wald or likelihood ratio test. However, the relevant statistical inference methods and associated asymptotic distributional theory for marked Hawkes processes are not available at this time. In addition to these theoretical challenges for maximum likelihood methods, there are substantial computational challenges, increasingly so when many marks are jointly included in the model.

Because of these computational and theoretical challenges when fitting marked processes, Richards et al. (2019) argue that potential marks be first screened for their impact before fitting the full model. They propose using the familiar score test of statistics (Rao 2009), or equivalently, the Lagrange Multiplier test (Breusch and Pagan 1980). As is well known, the score test, is computed using the score of the likelihood evaluated under the null hypothesis which, in this application, is that marks do not impact the intensity and the event times are that of an unmarked Hawkes process. Because it can then be constructed easily based on a single fitted intensity for an unmarked process, the score test leads to substantial computational advantages particularly when relevant and significant marks need to be selected from a possibly large catalogue before requiring the effort of jointly fitting the marked process model. Richards et al. (2019) provides detailed mathematical derivation of the form of score statistic, practically useful variants of it, computational implementation, simulation to verify usability of asymptotic distributions, and, application to limit order book event series.

Apart from the obvious computational advantage afforded by using a single Hawkes model fit, the large sample distribution theory for the score test can be derived using existing asymptotic theory for unmarked Hawkes processes. The purpose of this paper is to provide the theoretical foundations for the methodology and applications detailed in Richards et al. (2019). This theory requires construction of suitable stationary marked Hawkes process and a formal definition of the associated quasi-likelihood. Next, existing results of Clinet and Yoshida (2017) are adapted to obtain the chi-squared distribution as the large sample distribution of the score test under the null hypothesis that the mark or marks under test do
not boost the intensity of events. Additionally we show that the power of the score test against local alternatives can be approximated using a non-central chi-squared distribution.

The log-likelihood and associated statistical properties for the unmarked Hawkes SEPP has a long history – see Ozaki (1979), Ogata (1978), Andersen et al. (1996) and, more recently, Clinet and Yoshida (2017) for example. In deriving the likelihood (Embrechts et al. 2011, Definition 3) assume that the marks are unpredictable as defined in (Daley and Vere-Jones 2002, Definition 6.4.II(b)) so that the distribution of $x_i$, the mark at time $t_i$, is independent of previous event times and marks, i.e. of $\{(t_j, x_j)\}$ for $t_j < t_i$. An example of unpredictable marks is where the marks are conditionally i.i.d. given the past of the process but the marks may impact on the future of the intensity. The simplest example of this is where the marks are actually i.i.d. unconditionally as considered in Embrechts et al. (2011). In our extensive empirical analysis (see Richards et al. (2019)) we have frequently observed that $(x_i)_{i \geq 1}$ is a time series of serially dependent marks. In this case the unpredictability property does not hold. As far as we can determine, in the literature on likelihood inference for marked Hawkes processes there is no existing treatment of the serially dependent marks case. Accordingly, we derive the asymptotic distribution of the score test for serially dependent multivariate valued marks. This requires us to show that stable marked Hawkes process exist when the marks are stationary serially dependent, something that is not currently available in the literature.

From now on we consider a univariate Hawkes self exciting marked point process (SEPP) $N_g \in \mathbb{N} \times \mathbb{X}$, observed over the interval $t \in [0, T]$ and which takes the value 0 at $t = 0$. The subscript $g$ refers to the so-called boost function which encodes the way the marks impact the intensity of the process as in equations (1) and (3) to follow. There are $N_T$ events observed in the interval $[0, T]$ at times $0 < t_1 < t_2 < \ldots < t_{N_T} \leq T$ and the marks associated with the $i$th event are observations on a $d$-dimensional random variable $x_i \in \mathbb{X} \subset \mathbb{R}^d$ giving pairs $\{(t_i, x_i), i = 1, \ldots, N_T\}$. Throughout, we assume that the $x_i$ have the same marginal distribution indexed by parameter $\phi$. Note, for later, that $\phi$ are the parameters in the full joint distribution of the marks process across time including for example parameters to describe serial dependence as well as marginal distributions. As discussed later only the parameters in $\phi$ required to specify the marginal distribution play a role in the definition of the score statistic since it depends only on estimation of marginal moments of functions of the marks. In Richards et al. (2019) relevant marks constitute a vector of correlated marks which are also serially dependent. In order to accommodate such examples we explain how to define a marked Hawkes process with serially dependent marks, give conditions for stationarity of the point process, and, define the relevant quasi-likelihood.

Two data generating mechanisms (DGM’s) will be discussed. The first is based on a non-stationary process starting from zero, i.e defined for $t \in \mathbb{R}_+$, while the second is a stationary version defined for $t \in \mathbb{R}$. The first DGM is that considered in Clinet and Yoshida (2017) and also used in applications of the score statistic to the limit order book in Richards et al. (2019). It is this DGM which is assumed in the definition of the quasi-likelihood (Sect. 3), the score statistic (Sect. 4) and the associated asymptotic theory (Sect. 5 and 6). The second DGM is introduced (see Sect. 2) primarily for construction of a stationary process to which the first DGM converges in distribution as $T \to \infty$. This is the version that is considered in Ogata (1988), Liniger (2009) and Embrechts et al. (2011) for example. However, construction of the exact likelihood for the second DGM requires approximation because it relies on the whole history of the process including the unobserved values for $t < 0$.

The remainder of the paper is organized as follows. Section 2 introduces the statistical model and proves (see Proposition 1), via a thinning construction, that a stationary marked Hawkes process can be constructed when the marks are picked from a continuous time stationary process and gives several examples of processes for which the conditions are met.
Section 3 extends the definition of the joint likelihood of event times and marks, beyond the i.i.d. case currently available in the literature, to marks which are serially dependent. Section 4 defines the score test in detail. Section 5 states the main result (see Theorem 1) that the score statistic is asymptotically chi-squared distributed under the null hypothesis that marks do not impact the intensity function. For this result, in addition to the conditions of Clinet and Yoshida (2017) for the consistency and asymptotic normality of the unmarked process, conditions are required on the existence of moments and ergodicity of the mark process itself together with an additional condition (Condition 3) which links the marks and the unmarked intensity process. Lemma 2 shows that the Condition 3 is satisfied for the case of exponential decay function $w$.

Section 6 proves that the score statistic is asymptotically non-central chi-squared distributed under local alternatives of the form $\psi_T^*/\sqrt{T}$ where $T \to \infty$ and $T$ is the length of the interval over which the point process is observed. Section 7 discusses possible extensions to the main results. The Appendices contain proofs.

2 Model definition and existence of a stable marked Hawkes process

2.1 Model

Using similar notation to that in Liniger (2009) and Embrechts et al. (2011), with modifications as used in Clinet and Yoshida (2017), the marked Hawkes SEPP under the first DGM has intensity process given by

$$\lambda_g(t; \theta, \phi, \psi) = \eta + \vartheta \int_{(0,t) \times \mathbb{X}} w(t - s; \alpha)g(x; \phi, \psi)N_g(ds \times dx)$$  \hspace{0.5cm} (1)

where $w$ is a non-negative decay function satisfying $\int_0^\infty w(s; \alpha)ds = 1$. Combined with a domination condition on the function $g$ stated in Proposition 1 below, the normalization of the kernel $w$ is the key ingredient to prove the existence of a non-explosive marked Hawkes process following (1). It is a natural extension of the classical stability condition for the unmarked Hawkes process as stated in, e.g. Brémaud and Massoulié (1996), Condition (3) p. 1567.

The immigration rate is $\eta$, the branching coefficient $\vartheta$ and the parameter $\alpha$, not necessarily scalar, specifies the decay function $w$. The marks impact the intensity through the scalar valued boost function $g(x; \phi, \psi)$, where $g(\cdot; \phi, \psi) : \mathbb{R}^d \to \mathbb{R}_+$ and $\psi$ is a vector parameter of length $r$ specifying the way in which marks enter the boost function. In addition, $g$ is assumed to be a continuous function of $x$ and depends on parameters $\phi$, of the marks distribution, because the normalization $E_\phi[g(x; \phi, \psi)] = 1$ is required along with $\vartheta < 1$, as in Embrechts et al. (2011), to obtain a solution to (1) which converges to stationarity as $T \to \infty$.

The second DGM that is considered has stochastic intensity

$$\lambda_g^\infty(t; \theta, \phi, \psi) = \eta + \vartheta \int_{(-\infty,t) \times \mathbb{X}} w(t - s; \alpha)g(x; \phi, \psi)N_g^\infty(ds \times dx).$$  \hspace{0.5cm} (2)

In Proposition 1 below, we construct a suitable point process on $\mathbb{R}$ with this stochastic intensity. For the case where the marks are independent and identically distributed (i.i.d.) as considered in Embrechts et al. (2011), a stationary solution exists when $g$ is normalized as above. When the marks are picked from an underlying stationary continuous-time process $(y_t)_{t \in \mathbb{R}}$, that is $x_i = y_{t_i}$ for any $i \in \{1, ..., N_T\}$, we will see that this is also true under a stronger condition on the conditional expectation of $g(y_i; \phi, \psi)$ given the past.
Note that for the first DGM even if the underlying continuous time process $y_t$ on $\mathbb{R}$ is conceptualized to be stationary, the $x_i$, for $i = 1, 2, \ldots$ do not form a discrete time stationary process on the positive integers simply because the observation times $t_i$ are not stationary under (1) except when $y_t$ is a process of independent random variables.

Henceforth, let $\theta = (\eta, \vartheta, \alpha) \in \Theta$, $\phi \in \Phi$ and $\psi \in \Psi$ for some parameter spaces $\Theta$, $\Phi$ and $\Psi$. Let $\nu = (\theta, \phi, \psi) \in \Theta \times \Phi \times \Psi$ be the collection of all parameters for the marked process with intensity function (1). We denote the true value of the parameters as $(\theta^*, \phi^*, \psi^*)$. Throughout this paper we will consider the null hypothesis $H_0 : \psi^* = 0$ under which the true parameter vector is denoted $\nu^* = (\theta^*, \phi^*, 0)$ and the boost function is chosen so that that $g(x; \phi, 0) \equiv 1$ and the intensity in (1) does not depend on the marks. Since the score test will be developed under the null hypothesis we only require details of $\Theta$ in the derivation and theory to follow. Specifically we let $\Theta$ be a finite dimensional relatively compact open subset of $\mathbb{R}^K$, $K > 1$. The parameter space $\Phi$ for the marks density will typically be the natural space of parameters for the specified density and the boost parameter space $\Psi$ is chosen as appropriate for the form of $g$.

Under $H_0$ the observed event times are those of an unmarked Hawkes SEPP, $N$ with intensity denoted by

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

(3)

The intensity process defined in Embrechts et al. (2011) is the stationary version with infinite, but unobserved, event history included. In Brémaud and Massoulié (1996) the authors show that a suitable probability space exists on which a stationary version on $\mathbb{R}$, $N_\infty$, can be defined and to which the non-stationary version in (1) converges. We assume that $N_\infty$ is ergodic and this is proven in Clinet and Yoshida (2017) for the exponential decay case. Ogata (1978) considers both the stationary version of the intensity process and the non-stationary version as in (3) along with the associated likelihoods.

2.2 Existence of a stable marked hawkes process

Assume the existence of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ bearing a continuous time process $(y_t)_{t \in \mathbb{R}}$ taking values in the Borel space $(\mathbb{X}, \mathcal{X})$. The goal of this section is to show that under certain assumptions on $y$ that we now detail, it is possible to construct a marked Hawkes process $N^\infty$ defined for any $t \in \mathbb{R}$, with stochastic intensity $\lambda^\infty$ given in (2), and with marks $(x_i)_i$ such that $x_i = y_{t_i}$, where $t_i$ is the $i$th jump time of the associated counting process. Moreover, we prove that $N^\infty$ can be made stationary as soon as $y$ is. In the present section, we drop the parameters involved in the expression of $\lambda^\infty$ since statistical inference is considered only from Sect. 3. All the processes can be thought of as taken at a given point $v$.

Let us define $(\mathcal{F}_t^y)_{t \in \mathbb{R}}$ the canonical filtration of $y$, which is hereafter assumed to be right-continuous, and such that $(\mathbb{E}[g(y_t)|\mathcal{F}_{t-}])_{t \in \mathbb{R}}$ admits a version that is progressively measurable. Note that the latter assumption is a mild one. For instance, it is satisfied when $y$ is an i.i.d. process, i.e. for any $t_1, \ldots, t_n \in \mathbb{R}$ different times, $(y_{t_1}, \ldots, y_{t_n})$ forms an i.i.d. finite sequence. It also holds when $y$ is $B(\mathbb{R}) \otimes \mathcal{X}$-measurable (Jacod and Shiryaev (2013), Theorem I.2.28) where $B(\mathbb{R})$ designates the Borel $\sigma$-field associated to $\mathbb{R}$. This proves, for instance, that any process $y$ that is right continuous with left limits (or left continuous with right limits) also satisfies the assumption, and, naturally, any mixture of the aforementioned processes as well. We now give a proof of the existence of the marked Hawkes process using a thinning method similar to (Liniger 2009, Chapter 6) and Brémaud and Massoulié (1996). To that end, we assume the existence on $(\Omega, \mathcal{F}, \mathbb{P})$ of a Poisson process $N$ with
intensity \( \eta \) on \( \mathbb{R}^2 \), with points denoted \((t_i, u_i)_{i \in \mathbb{Z}}\) and independent of \( y \). Then we consider the canonical process \((t_i, u_i, y_{t_i})_{i \in \mathbb{Z}}\) and see this process as a random measure \( \overline{N}_g \) on \( \mathbb{R}^2 \times \mathbb{X} \). We associate to \( \overline{N} \) the filtration generated by the \( \sigma \)-algebras \( \mathcal{F}^N_t = \sigma(\overline{N}(\langle -\infty, s \rangle \times A), s \in (-\infty, t], A \in \mathcal{B}(\mathbb{R})) \). Similarly, we associate to \( \overline{N}_g \) the filtration generated by the \( \sigma \)-algebras \( \mathcal{F}^N_t = \sigma(\overline{N}_g(\langle -\infty, s \rangle \times A \times B), s \in (-\infty, t], A \in \mathcal{B}(\mathbb{R}), B \in \mathcal{X}). \) Note that \( \overline{N}(dt \times du) = \overline{N}_g(dt \times du \times \mathbb{X}) \). Finally, consider the larger filtration generated by \( \mathcal{F}_t = \mathcal{F}^X_t \vee \mathcal{F}^N_t \). In Proposition 1 below, we show the existence of a marked point process \( N^\infty \) adapted to \( \mathcal{F}_t \), satisfying (2), and which is constructed as an integral over the canonical measure \( \overline{N}_g \). Let us first establish the expression of the predictable compensator of the canonical process \( \overline{N}_g \).

**Lemma 1** The marked point process \( \overline{N}_g \) on \( \mathbb{R} \times (\mathbb{R} \times \mathbb{X}) \) admits the \( \mathcal{F}_t \)-compensator \( ds \times (d\mu \times F_s(dx)) \), where for \( s \in \mathbb{R} \), \( F_s(dx) \) is the conditional distribution of \( y_s \) with respect to \( \mathcal{F}^X_{s-} \).

The proof of Lemma 1 can be found in Appendix A. Recall now that a marked point process \( M \) seen as random measure on \( \mathbb{R} \times \mathbb{X} \) is said to be stationary if for any \( t \in \mathbb{R} \), \( r \in \mathbb{N} \) and any finite collections \( A_0, ..., A_r \in \mathcal{B}(\mathbb{R}) \) and \( B_0, ..., B_r \in \mathcal{X} \), we have that \((M((A_i + t) \times B_i))_{i \in \{0, ..., r\}}\) has the same distribution as \((M(A_i \times B_i))_{i \in \{0, ..., r\}}\). We now prove the existence of the marked Hawkes process.

**Proposition 1** Assume that for any \( t \in \mathbb{R} \), \( \mathbb{E}[g(y_t)|\mathcal{F}^X_{t-}] \leq C \) where \( C < 1/\theta^* \), along with \( \int_0^{+\infty} \psi(s)ds = 1 \). Then, there exists a marked point process of the form \((\tau_i, x_i)_{i \in \mathbb{Z}} := (t_i, y_{t_i})_{i \in \mathbb{Z}}\), also represented by the random measure \( N^\infty \) on \( \mathbb{R} \times \mathbb{X} \), such that:

1. The counting process associated to \((\tau_i)_{i \in \mathbb{Z}}\) is adapted to \((\mathcal{F}_t)_{t \in \mathbb{R}}\), and admits the stochastic intensity (with respect to \( \mathcal{F}_t \)) defined in (2), that is
   \[
   \lambda^\infty_g(t) = \eta^* + \theta^* \int_{(-\infty, t) \times \mathbb{X}} \psi(t - s)g(x)N^\infty_g(ds \times dx).
   \]

2. The random measure \( N^\infty_g \) admits \( \pi^\infty(ds \times dx) = \lambda^\infty_g(s)ds \times F_s(dx) \) as predictable compensator, where \( F_s(dx) \) is the conditional distribution of \( y_s \) given \( \mathcal{F}^X_{s-} \).

3. If the process \((y_t)_{t \in \mathbb{R}}\) is stationary, then so is \( N^\infty_g \).

The proof of Proposition 1 can also be found in “Appendix” A.

The construction in the proof yields the existence of a marked Hawkes process that is defined on the whole real line. For statistical purposes, it is perhaps more relevant to seek a version \( N_g \) that is defined on \( \mathbb{R}_+ \) only, and which yields the slightly different dynamics of (1) for its associated stochastic intensity \( \lambda_g \). It is immediate to check that the above construction can be adapted for a marked point process starting from 0 instead of \( -\infty \) by replacing \( -\infty \) by 0 in all the integrals. However, in that case, the resulting process \( N_g \) and its associated intensity \( \lambda_g \) are not stationary, but one can prove that \( N_g \) converges to the stationary version starting from \( -\infty \) by a straightforward adaptation of the proof of Theorem 1 in Brémaud and Massoulié (1996). Since it is not needed in the remainder of the paper details of the proof of this latter result are not provided.

In the unmarked case, a well-known alternative construction is the Poisson cluster representation, which consists in first drawing at random a Poisson process of intensity \( \eta^* \) (the immigration process), and then attaching to each immigrant point a cluster of inhomogeneous Poisson processes following the excitation kernel (the birth processes). The full Hawkes process is obtained by collapsing all these points on the real line (see e.g Hawkes...
Remark 1 If the marks \( y_s \) are i.i.d, the conditional expectation \( \mathbb{E}[g(y_s)|F^Y_{s^-}] \) reduces to the usual expectation \( \mathbb{E}[g(y_s)] \) which is equal to unity due to normalization of the boost function \( g \). Hence the condition that \( \mathbb{E}[g(y_s)|F^Y_{s^-}] \leq C < 1/\vartheta^* \) is obviously satisfied.

Remark 2 When \( y \) is a left-continuous process, or more generally a predictable process, then \( \mathbb{E}[g(y_s)|F^Y_{s^-}] = g(y_s) \) which leads to the condition \( g(y_s) \leq C < 1/\vartheta^* \). This may be very restrictive in practice. For example, for a parametric linear boost in a single mark this would require the mark process to be bounded from above by a constant which depends on \( \varphi^*, \psi^* \) and \( \vartheta^* \). It is also interesting to note that for mark processes with continuous sample paths, mixing conditions (which specify the rate at which dependence fades away with increasing time separation) will not lead to a weakening of the aforementioned stringent condition. The difficulty stems from the dependence of \( g(Y_t) \) and \( g(Y_s) \) when \( t \) and \( s \) are close together. If at some time \( t_0, g(Y_{t_0}) > 1/\vartheta \), by ‘continuity’, it will stay above that level for some time \([t_0, t_0 + \epsilon] \). On this interval the process becomes explosive, and regardless of the number of jumps, all the marks are highly correlated since \( g(y_s) \approx g(y_{t_0}) > 1/\vartheta \) for \( s \in [t_0, t_0 + \epsilon] \).

Remark 3 In view of the last remark, marks which arise from a stochastic process with continuous sample paths are probably not practical for use in marked Hawkes self exciting processes. On the other hand, marks based on a stochastic process which contains some degree of independence could more easily satisfy the condition of Proposition 1. For example, let \( y_t = U_t + V_t \) where \( U_t \) has continuous sample paths and \( V_t \) is a pure noise process independent of \( U_t \). More generally a conditionally independent specification in which \( U_t \) is as before and \( Y_t \sim_{i.i.d} f(\cdot|U_t) \) would also more easily satisfy the condition. For instance \( U_t \) may specify some of the parameters needed for the density \( f \).

Remark 4 The condition \( \mathbb{E}[g(Y_t)|F^Y_{t^-}] \leq C < 1/\vartheta \) can be slightly relaxed as we next explain. Let \( \rho_n(t) = \mathbb{E}[\lambda^\infty_{g,n}(t) - \lambda^\infty_{g,n-1}(t)] \). From the proof of Proposition 1 we know that a sufficient condition for non-explosion is \( \sum_{n=1}^{+\infty} \rho_n(t) < \infty \). A straightforward induction shows that

\[
\rho_n(t) = \vartheta^n \eta \mathbb{E} \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_{n-1}} w(t - t_1) \cdots w(t - t_{n-1} - t_{N_T}) \psi_{t_1} \cdots \psi_{t_{N_T}} dt_1 \cdots dt_{N_T},
\]

where \( \psi_t = \mathbb{E}[g(Y_t)|F^Y_{t^-}] \). Therefore, if we replace the above condition by \( \int_{-\infty}^{t} w(t - s) \mathbb{E}[g(y_s)|F^Y_{s^-}]ds \leq C < 1/\vartheta \) for any \( t \in \mathbb{R} \), then \( \sum_{n=1}^{+\infty} \rho_n(t) \leq \frac{\eta}{1 - C \vartheta} < +\infty \), and the process is stable.

3 Quasi-likelihood for marked hawkes processes

From now on, we will always assume that the observed marked Hawkes process \( N_g \) as defined in the introduction is generated as in Sect. 2, on the time interval \([0, +\infty) \). In particular, its
associated mark process \((x_i)_{i \geq 1}\) is sampled from a given stationary underlying continuous-time process \((y_t)_{t \in \mathbb{R}}\) through the relation \(x_i = y_{t_i}\) for any \(i \geq 1\), and the stability condition \(\mathbb{E}[L(y_t; \phi^*, \psi^*)] \leq C < \infty\) is satisfied for any \(s \in [0, +\infty)\).

In general, it is possible to represent the log-likelihood \(\mathcal{L}_g(v)\) when the marks are not i.i.d. and are sampled from \((y_t)_{t \in \mathbb{R}}\) as follows: recall that the integer-valued measure \(N_g(dt \times dx)\) admits a predictable compensator \(\pi(dt \times dx)\) by Proposition 1 (ii) of the form \(\pi(ds \times dx) = \lambda_g(s; v)ds \times F_t(dx, \phi)\). Assuming that for any \(s \in \mathbb{R}_+,\) the conditional distributions \(F_s(dx, \phi)\) are dominated by some measure \(c(dx)\) \((F_s(dx, \phi) = f_s(x; \phi)c(dx))\), using (Jacod and Shiryaev 2013, Theorem III.5.19)) we can generalize the log-likelihood for the pure point process with

\[
\mathcal{L}_g(v) = \bar{l}_g(v) - l_g(v^*)
\]

whose maximum can be obtained by maximizing

\[
\bar{l}_g(v) = \int_{[0,T] \times \mathbb{X}} \log[\lambda_g(t; v)f_t(x; \phi)]N_g(dt \times dx) - \int_{[0,T] \times \mathbb{X}} f_t(x; \phi)c(dx) \lambda_g(t; v)dt,
\]

and where we recall that \(v^* = (\theta^*, \phi^*, \psi^*)\) designates the actual parameter driving \(N_g\), and \(T > 0\) is the horizon time. Expanding the logarithm, we get

\[
l_g(v) = \int_{[0,T] \times \mathbb{X}} \log \lambda_g(t; v)N_g(dt \times dx) - \Lambda_g(T; v) + \int_{[0,T] \times \mathbb{X}} \log f_t(x; \phi)N_g(dt \times dx),
\]

where the compensator at \(T\) is

\[
\Lambda_g(T; v) = \int_{[0,T]} \lambda_g(t; v)dt.
\]

For instance, when the marks are i.i.d, the conditional density \(f_s\) is replaced by the unconditional marginal \(f\), as in Embrechts et al. (2011), and the last term in (4) becomes \(\int_{[0,T] \times \mathbb{X}} \log f(x; \phi)N_g(dt \times dx)\) which evaluates to \(\sum_{i=1}^{N_f} \log f(x_i; \phi)\) and the log-likelihood is (up to the term \(l_g(v^*)\), independent from \(v\))

\[
l_g(v) = \int_{[0,T] \times \mathbb{X}} \log \lambda_g(t; v)N_g(dt \times dx) - \Lambda_g(T; v) + \sum_{i=1}^{N_f} \log f(x_i; \phi).
\]

When the marks show more intricate dynamics, because of the third term, computing (4) requires that one observes the whole trajectory of the joint process \((N_t, y_t)_{t \in [0,T]}\) due to the fact that \(\pi\) is the compensator associated to \(F_t\), which is larger than the canonical filtration of the marked Hawkes process. However, as explained in the next section, the associated score statistic will depend on \(l_g(v)\) solely through its first two terms, and is therefore computable based on the discrete observations \((t_i, x_i)_{1 \leq i \leq N_f}\) only. It is nonetheless of interest to derive a feasible version of (4) when assuming that one only has discrete observations of the form \((t_i, x_i)_{1 \leq i \leq N_f} = (t_i, y_{t_i})_{1 \leq i \leq N_f}\). One option consists in changing the last term in (4) to \(\sum_{i=1}^{N_f} \ln f(y_{t_i}; \phi(t_j, y_{t_j}))_{1 \leq j < i}\), where \(f(\cdot; \phi(t_j, y_{t_j}))_{1 \leq j < i}\) corresponds to the conditional density of the \(i\)th mark given \((t_j, y_{t_j})_{1 \leq j < i}\), assuming that it depends on the parameter \(\phi\) only. This yields the objective function

\[
l_g(v) = \int_{[0,T] \times \mathbb{X}} \log \lambda_g(t; v)N_g(dt \times dx) - \Lambda_g(T; v) + \sum_{i=1}^{N_f} \ln f(y_{t_i}; \phi(t_j, y_{t_j}))_{1 \leq j < i}.
\]
Under the null hypothesis, $\psi = 0$ and the boost is the identity so that marks do not impact
the intensity. But the marks process and the event process may not be independent because
the conditional distribution of marks $x_i = y_{t_i}$ is not free of the event times. Nonetheless, the
objective function in (6) becomes a sum of two terms

$$l_g(\theta, \phi, 0) = l(\theta) = l(\theta) + \sum_{i=1}^{N_T} \ln f (y_{t_i}; \phi | (t_j; y_{t_j})_{1 \leq j < i})$$

where the first term is the log-likelihood for the unmarked process $N(t)$

$$l(\theta) = \int_{[0, T]} \log \lambda(t; \theta) N(dt) - \Lambda(T; \theta)$$

with corresponding compensator

$$\Lambda(T, \theta) = \int_0^T \lambda(t; \theta) dt.$$ Here $N(dt) = N_x(dt, X)$ and $\lambda(t; \theta) = \lambda_x(t; \theta, 0)$ which does not depend on $\phi \in \Phi$. The second term accounts for the distribution of the marks conditional on event times. Hence, under $H_0$, the parameters $\theta$ of the unmarked Hawkes process are decoupled from the parameters $\phi$ of the marks distribution so that these can be separately estimated.

### 4 The score test

We now derive the score statistic for testing the null hypothesis that marks do not impact the
intensity of the observed point process. Specifically, recall that $g(x; \phi, \psi)$ is the normalised
boost function in the general intensity process (1) and the null hypothesis is $H_0 : \psi = 0$ which corresponds to $g(x; \phi, 0) = 1$ for all $x$ and the alternative is that $\psi \neq 0$ in which case $g(x; \phi, \psi)$ is not identically equal to 1 so that the marks do impact the intensity. Note that in the log-likelihood (4), (and also in the alternative forms (5) and (6)), the third term involves only the parameter $\phi$ and, as a result, the score vector $\partial_\psi l_g$ with respect to the boost parameter $\psi$ involves the first two terms only. Hence the score for likelihoods in (4), (5) and (6) with respect to $\psi$ is the same under the null hypothesis.

Recall that $v^* = (\theta^*, \phi^*, 0)$ denotes the true value of the combined parameters under $H_0$. Let $\hat{\theta}_T = (\hat{\theta}_T, \hat{\phi}_T, 0)$ where $\hat{\theta}_T$ is the asymptotic quasi maximum likelihood estimate, as in (Clinet and Yoshida 2017, page 1804), based on the likelihood (7) under $H_0$ of the intensity process parameters and $\hat{\phi}_T$ is the MLE for the parameters of the joint marks density. Denote the derivatives of the log-likelihood with respect to $\nu$ as $\partial_\nu l_g(\nu)$ at the parameter value $\nu$ so that $\partial_\nu l_g(v^*)$ and $\partial_\nu l_g(\hat{\nu}_T)$ are evaluated at $v^*$ and $\hat{\nu}_T$ respectively. In general, the score (or Lagrange multiplier) test statistic (Breusch and Pagan 1980) is defined as

$$Q_T = \partial_\nu l_g(\hat{\nu}_T)^T \mathcal{I}(\hat{\nu}_T)^{-1} \partial_\nu l_g(\hat{\nu}_T)$$

where $\mathcal{I}(v^*) = \mathbb{E}_{v^*} [\partial_\nu l_g(v^*) \partial_\nu l_g(v^*)^T]$ is the Fisher information and $\mathcal{I}(\hat{\nu}_T)$ evaluates this at the parameters $\hat{\nu}_T$, estimated under $H_0$. Also (Breusch and Pagan 1980) the information matrix can be replaced by any matrix with the same limit in probability, for example the negative of the matrix of second derivatives of the log-likelihood, and the large sample properties of the score statistic will be the same.

The score statistic (8) can be simplified using two facts. Firstly, while the first two terms in the
quasi-likelihoods do involve $\phi$ (as a result of the boost function being normalized using
moments of the marginal distribution of the marks), for the functional forms of boost function that we consider in (11) below it is shown in Richards et al. (2019) that the information matrix is block diagonal with respect to the groups of parameters \( \theta, \phi \) and \( \psi \). Secondly, \( \partial_{\psi} I_g(\hat{\psi}_T) = (0, 0, \partial_{\psi} I_g(\hat{\psi}_T))^T \). Hence, for all examples of likelihoods given above, the score test statistic (8) simplifies to

\[
Q_T = \partial_{\psi} I_g(\hat{\psi}_T)^T I(\hat{\psi})^{-1} \partial_{\psi} I_g(\hat{\psi}_T)
\]

(9)

where \( I(\hat{\psi}) \) is the \( r \times r \) diagonal block of \( I(\hat{\psi}_T) \) corresponding to \( \psi \). Because the third term in the log-likelihood (4) (and its variant (6)) do not depend on \( \psi \) it follows that

\[
\partial_{\psi} I_g(v) = \int_{[0,T]\times\mathbb{X}} \lambda_g(t; v)^{-1} \partial_{\psi} \lambda_g(t; v) N_g(dt \times dx) - \int_{[0,T]} \partial_{\psi} \lambda_g(t; v) dt
\]

(10)

where

\[
\partial_{\psi} \lambda_g(t; v) = \hat{\theta} \int_{[0,t]\times\mathbb{X}} w(t - s; \alpha) \partial_{\psi} g(x; \phi, \psi) N_g(ds \times dx).
\]

To proceed further the form of the boost function that is considered in this paper is now defined. Quite general normalized boost functions, \( g(x; \phi, \psi) \) can be constructed by starting with a function \( h(x; \psi) \) and defining the boost function

\[
g(x; \phi, \psi) = \frac{h(x; \psi)}{\mathbb{E}_\phi[h(x; \psi)]}.
\]

(11)

It is no loss of generality to require \( h(x; 0) \equiv 1 \). The vector of derivative of \( g \) with respect to \( \psi \) appearing in \( \partial_{\psi} \lambda_g(t; v) \) is then

\[
\partial_{\psi} g(x; \phi, \psi) = \frac{1}{\mathbb{E}_\phi[h(x; \psi)]} \left[ \partial_{\psi} h(x; \psi) - g(x; \phi, \psi) \mathbb{E}_\phi[\partial_{\psi} h(x; \psi)] \right].
\]

Recall that the null hypothesis being assessed with the score test is \( H_0 : \psi = 0 \), which is thus equivalent to \( g(x; \phi, 0) = 1 \) so that marks do not boost intensity. Denote \( H(x) = \partial_{\psi} h(x; 0) \) and \( \mu_H(\phi) = \mathbb{E}_\phi[H(x)] \) and put \( G(x; \phi) = \partial_{\psi} g(x; \phi, 0) \). Then \( G(x; \phi) = H(x) - \mu_H(\phi) \) is a vector of dimension \( r \) comprised of functions of the components of the vector mark centered at their expectations.

**Condition 1 Conditions on boost function specification:** Throughout we assume \( h \), used to define the boost function \( g \) in (11), and its first and second derivatives with respect to \( \psi \), denoted \( \partial_{\psi} h \) and \( \partial_{\psi}^2 h \), satisfy the following properties:

(i) \( h(x; 0) \equiv 1 \);

(ii) \( \mathbb{E}_\phi[h(x; \psi)] \) and \( \mathbb{E}_\phi[\partial_{\psi} h(x; \psi)] \) exist for all \( \psi \in \Psi, \phi \in \Phi \);

(iii) The density of \( x, f(x; \phi) \) with respect to the measure \( c(dx) \), satisfies

\[
\partial_{\phi} \int_{\mathbb{X}} h(x; \psi) f(x; \phi) c(dx) = \int_{\mathbb{X}} h(x; \psi) \partial_{\phi} f(x; \phi) c(dx)
\]

and

\[
\partial_{\psi} \mathbb{E}_\phi[h(x; \psi)] = \mathbb{E}_\phi[\partial_{\psi} h(x; \psi)].
\]

Obviously, based on these properties of \( h, \mathbb{E}_\phi[g(x; \phi, \psi)] = 1 \) for all \( \psi \in \Psi, g(x; \phi, 0) \equiv 1, \) and \( \mathbb{E}_\phi[G(x; \phi)] = 0 \). The requirements that \( \mathbb{E}_\phi[h(x; \psi)], \mathbb{E}_\phi[\partial_{\psi} h(x; \psi)] \) and \( \partial_{\psi} \mathbb{E}_\phi[\partial_{\psi} h(x; \psi)] \) exist impose obvious conditions on the marginal distribution of \( x \).
For example, when \( h(x; \psi) \) is a polynomial of degree \( p \) in \( x \) then \( \mathbb{E}_\phi[x^p] \) needs to exist. The first part of Condition (iii) is a consequence of the usual condition for definition of the Fisher information matrix (see Rao 2009, Sect. 5a.4). Condition (iii) ensures that the information matrix for all parameters has separate diagonal blocks for the parameters \( \theta, \phi \) and \( \psi \) which allows the score statistic to be simplified – see (Richards et al. 2019, Supplementary Material Appendix A). Many examples of boost functions including the polynomial, exponential and power function forms for \( h \), as presented in Liniger (2009), as well as additive and multiplicative combinations of individual elements of \( x \) used in Richards et al. (2019) satisfy these conditions.

We are now ready to derive an explicit form for the score vector and information matrix in terms of the function \( G(x; \phi) \) derived from the boost function. Under \( H_0 \), the derivative of (4) (or equivalently (6)) with respect to \( \psi \) at any values of \( \theta, \phi \) is

\[
\partial_{\psi} I_g(\theta, \phi, 0) = \int_{[0,T]} \lambda(t; \theta)^{-1} \partial_{\psi} \lambda_g(t; \theta, \phi, 0) N(dt) - \int_{[0,T]} \partial_{\psi} \lambda_g(t; \theta, \phi, 0) dt \tag{12}
\]

with

\[
\partial_{\psi} \lambda_g(t; \theta, \phi, 0) = \hat{\theta} \int_{[0,t) \times \mathbb{R}} w(t - s; \alpha) G(x; \phi) N_g(ds \times dx). \tag{13}
\]

When evaluated at the estimates \( (\hat{\theta}_T, \hat{\phi}_T, 0) \) under the null hypothesis the score (12) can be written

\[
\partial_{\psi} I_g(\hat{\nu}_T) = \int_{[0,T]} \lambda(t; \hat{\theta}_T)^{-1} \partial_{\psi} \lambda_g(t; \hat{\theta}_T, \hat{\phi}_T, 0) N(dt) - \int_{[0,T]} \partial_{\psi} \lambda_g(t; \hat{\theta}_T, \hat{\phi}_T, 0) dt \tag{14}
\]

and when evaluated at the true parameter vector, \( \nu^* = (\theta^*, \phi^*, 0) \) under \( H_0 \), it can be written as

\[
\partial_{\psi} I_g(\nu^*) = \int_{[0,T]} \lambda(t; \nu^*)^{-1} \partial_{\psi} \lambda_g(t; \nu^*) \tilde{N}(dt) \tag{15}
\]

where \( \tilde{N}(dt) = N(dt) - \lambda(t; \theta^*) dt \). Moreover, the Fisher information with respect to \( \psi \) admits the following representation

\[
\mathcal{I}_\psi(\nu^*) = \mathbb{E} \int_{[0,T]} \lambda(t; \nu^*)^{-2} (\partial_{\psi} \lambda_g(t; \nu^*)) \otimes^2 N(dt), \tag{16}
\]

where, for a vector \( x \in \mathbb{R}^r, x \otimes^2 = xx^T \in \mathbb{R}^{r \times r} \).

Noting that the expectation required to evaluate (16) is not computable in closed form, we suggest empirical evaluation replacing the expectation by the time average over events and using the estimate \( \hat{\nu}_T \) to get

\[
\hat{\mathcal{I}}_\psi = \int_{[0,T]} \lambda(t; \hat{\theta}_T)^{-2} (\partial_{\psi} \lambda_g(t; \hat{\nu}_T)) \otimes^2 N(dt). \tag{17}
\]

Under the conditions of Theorem 1 we show (Appendix B, Lemmas 5 and 7)

\[
T^{-1} \hat{\mathcal{I}}_\psi \rightarrow^P \Omega, \quad \text{as} \quad T \rightarrow \infty \tag{18}
\]

where \( \Omega \) is a finite, positive definite matrix.

Using these estimates in the definition (9), a feasible version of the score statistic can be implemented in practice as

\[
\hat{Q}_T = \partial_{\psi} I_g(\hat{\nu}_T) T^{\frac{1}{2}} \partial_{\psi} I_g(\hat{\nu}_T) \tag{19}
\]
where \( \partial_\phi I_g(\hat{\phi}_T) \) is defined above and \( \hat{I}_\psi \) is given by (17).

For practical implementation as well as the proof that the score statistic is asymptotically chi-squared (Theorem 1), the form of \( G(x, \phi_T) = H(x) - \mu_H(\phi_T) \) appearing in \( \partial_\phi \lambda_g(t; \theta_T, \phi_T, 0) \) of (13) allows substantial simplification while allowing many practically useful boost function specifications. Indeed, it is worth noting that the score statistic introduced in (19) depends on \( \hat{\phi}_T \) only through \( \mu_H(\hat{\phi}_T) \). For \( \mu_H(\phi) \) evaluated at \( \phi^* \) we put \( \mu_H = \mu_H(\phi^*) \) and, if evaluated at \( \hat{\phi}_T, \hat{\mu}_H \). We also use the same notation for any consistent estimate of \( \mu_H \) such that \( \hat{\mu}_H = \hat{H}(x) \), the vector of sample means of components. We let \( G(x) = H(x) - \mu_H \) at the true value and \( \hat{G}(x) = H(x) - \hat{\mu}_H \). The score statistic relies only on the marginal means of \( H(x) \) to be consistently estimated to get \( \hat{G}(x) \). One way is to use the MLE’s \( \hat{\phi}_T \) for \( \phi \) and use known formulae for \( \mu_H(\phi) \) evaluated at these estimates. Often these formulae for marginal means do not require all the parameters specifying the full joint distributions of the marks process. For example, serial dependence parameters may not be required. Alternatively, any consistent estimates \( \hat{\mu}_H \) can be used to define \( \hat{G}(x) \), for example sample moments for \( H(x) \), thereby avoiding the need for potentially complex computations to obtain the MLE’s for the full set of parameters \( \phi \). In practice this can be computationally very advantageous. Additionally, simulations in Richards et al. (2019), show that the size and power properties of the score statistic are very similar if sample moments or estimates using theoretical moments evaluated at the MLE’s for \( \phi \) are used. This is consistent with the results of Theorem 1 and Theorem 2 of this paper. In summary, the crucial quantity \( \hat{G} \) required to define the score statistic can be calculated using any consistent estimates \( \hat{\mu}_H \). Note that \( \hat{\mu}_H \rightarrow^p \mu_H \) holds for either the sample mean estimate (using ergodicity of the marks process), the parametric form, \( \hat{\mu}_H = \mu_H(\hat{\phi}_T) \) (using consistency of the maximum likelihood estimates \( \hat{\phi}_T \) under appropriate regularity conditions on \( f_1(x; \phi) \)) or any other consistent estimate of \( \phi \).

5 Asymptotic distribution of the score statistic

To prove that the score statistic \( \hat{Q}_T \) has a large sample chi-squared distribution under the null hypothesis conditions are required on the intensity process for the unboosted process. The extra conditions are those required for convergence of the quasi MLE for Hawkes processes under \( H_0 \), for which the intensity does not depend on marks. Because we adapt the proofs of (Clinet and Yoshida 2017, Theorems 3.9 and 3.11) to the score statistic we re-state their conditions [A1], [A2], [A3] and [A4] here. These generalize Conditions A, B and C of Ogata (1978) applied to the intensity process defined in (3) for the unmarked process. Ogata (1978) provided the first consistency and asymptotic normality results for the unmarked Hawkes process and verified that his conditions apply to the exponential decay function \( w(t; \alpha) \). Clinet and Yoshida (2017) give conditions for the convergence of moments of the quasi MLE and verify them for the exponential decay function case. As far as we are aware there has been no published verification of the conditions of Ogata (1978) or Clinet and Yoshida (2017) for the power law decay function.

Condition 2 Conditions on the intensity process under \( H_0 : \psi = 0 \). For clarity, these are restated from Clinet and Yoshida (2017) using notation of this paper and as relevant to the Hawkes process. These conditions refer to the intensity process defined in (3). Recall that \( \theta^* \) refers to the true parameter defining the intensity process under \( H_0 \).

A1 The mapping \( \lambda : \Omega \times \mathbb{R}_+ \times \Theta \rightarrow \mathbb{R}_+ \) is \( \mathcal{F} \otimes \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{B}(\Theta) \)-measurable. Moreover, almost surely:
(i) for any \( \theta \in \Theta \), \( s \to \lambda(s, \theta) \) is left continuous;

(ii) for any \( s \in \mathbb{R}_+ \), \( \theta \to \lambda(s, \theta) \) is in \( C^3(\Theta) \) and admits a continuous extension to \( \bar{\Theta} \).

A2 The intensity process \( \lambda \) and its derivatives satisfy, for any \( p > 1 \),
\[
\sup_{t \in \mathbb{R}_+} \sum_{i=0}^3 \| \sup_{\theta \in \Theta} |\partial_i^0 \lambda(t, \theta)| \|_p < \infty.
\]

A3 For a Borel space \((E, \mathcal{B}(E))\) let \( C_b(E, \mathbb{R}) \) be the set of continuous, bounded functions from \( E \) to \( \mathbb{R} \). For any \( \theta \in \Theta \) the triplet \((\lambda(\cdot, \theta^*), \lambda(\cdot, \theta), \partial_\theta \lambda(\cdot, \theta))\) is ergodic in the sense that there exists a mapping \( \pi : C_b(E, \mathbb{R}) \times \Theta \to \mathbb{R} \) such that for any \((\xi, \theta) \in C_b(E, \mathbb{R}) \times \Theta\),
\[
\frac{1}{T} \int_0^T \xi(\lambda(s, \theta^*), \lambda(s, \theta), \partial_\theta \lambda(s, \theta)) ds \to^P \pi(\xi, \theta).
\]

A4 Define
\[
\Upsilon_T(\theta) = \frac{1}{T} (l_T(\theta) - l_T(\theta^*)),
\]
which is shown in (Clinet and Yoshida 2017, Lemma 3.10) to satisfy
\[
\sup_{\theta \in \Theta} |\Upsilon_T(\theta) - \Upsilon(\theta)| \to^P 0
\]
and \( \Upsilon(\theta) \) is the ergodic limit of \( \Upsilon_T(\theta) \) as defined in (Clinet and Yoshida 2017, p. 1807).

Assume, for asymptotic identifiability, that for any \( \theta \in \bar{\Theta} - \{\theta^*\}, \Upsilon(\theta) \neq 0 \).

Under Condition 2: [A1] to [A4], Clinet and Yoshida (2017) show (Theorem 3.9) that any asymptotic QMLE \( \hat{\theta}_T \) is consistent, \( \hat{\theta}_T \to^P \theta^* \), and (Theorem 3.11) asymptotically normal \( \sqrt{T}(\hat{\theta}_T - \theta^*) \to^d \Gamma^{-\frac{1}{2}} \zeta \) where \( \zeta \) has a standard multivariate normal distribution and \( \Gamma \) is the asymptotic information matrix, assumed to be positive definite. Additionally they prove that \( \Gamma \) satisfies
\[
\sup_{\theta \in V_T} |T^{-1} \partial_\theta^2 l_T(\theta) + \Gamma| \to^P 0,
\]
where \( V_T \) is a ball shrinking to \( \theta^* \).

As noted above these conditions are met for the (multivariate) exponential decay Hawkes process without marks as shown in (Clinet and Yoshida 2017, Sect. 4) assuming each element of \( \theta = (\eta, \vartheta, \alpha) \) belongs to finite closed intervals of \( \mathbb{R} \). For example, for the exponential decay function \( w(s; \alpha) = \alpha \exp(-\alpha s), K = 3 \) and we assume that \( 0 < \eta \leq \eta \leq \tilde{\eta} < \infty, 0 < \vartheta \leq \vartheta \leq \tilde{\vartheta} < \infty, 0 < \alpha \leq \alpha \leq \tilde{\alpha} < \infty \) so that \( \Theta \) is a finite dimensional relatively compact open subset of \( \mathbb{R}^3 \).

In order to establish the asymptotic distribution of the score vector with respect to \( \psi \), Condition 2 A.2 needs to be extended to accommodate the contribution to the score vector from the marks as follows.

Condition 3 For \( p = (\dim(\Theta) + 1) \vee 4 \), where \( x \vee y = \max(x, y) \), under \( H_0 \) and with \( \phi \) fixed at \( \phi^* \), assume
\[
\sup_{t \in \mathbb{R}_+} \sum_{i=0}^2 \| \sup_{\theta \in \Theta} |\partial_i^0(\partial_\phi \lambda_x(t; \theta, \phi^*, 0))| \|_p < \infty.
\]
Lemma 2 Condition 3 is satisfied for the exponential decay function model (for which \( \text{dim}(\Theta) = 3 \)) and stationary ergodic marks for which \( \mathbb{E}_{\psi^*}[|G(y)|^4] < \infty \) where \( y \) has the marginal distribution of \((y_s)_{s \in \mathbb{R}}\).

The proof of Lemma 2 is given in Appendix B.

The final condition concerns the marks.

Condition 4 The marks are from a strictly stationary ergodic process \((y_s)_{s \in \mathbb{R}}\) with \( \mathbb{E}_{\psi^*}[|G(y)|^4] < \infty \) such that for any \( c \in \mathbb{R}^r, c \neq 0 \), the process of linear combinations \( G_c(y_s) = c^T G(y_s) \) for \( s \in \mathbb{R} \) has a strictly positive definite autocovariance function \( \gamma_{G_c}(\cdot) \). We also assume that \( \hat{\mu}_H \to \mu_H \) as \( T \to \infty \).

We now state the main result.

Theorem 1 Assume Conditions 1, 2, 3 and 4. Under \( H_0 \), the score statistic defined in (19) using any consistent estimator \( \hat{\mu}_H \) of \( \mu_H \) in the definition of \( \hat{G}(x) \) and with information matrix estimated by \( \hat{I}_\psi \) defined in (17) satisfies

\[
\hat{Q}_T \xrightarrow{d} \chi(r) \quad \text{as} \ T \to \infty, \quad r = \text{dim}(\psi).
\]  

The proof, given in Appendix C, consists essentially of showing (18) for the limit of the information matrix and that the score vector scaled by \( 1/\sqrt{T} \) has a limiting multivariate normal distribution by using the additional Condition 3 to extend the functional central limit theorem in Clinet and Yoshida (2017). Note that the dimension of the chi-squared distribution in the limit (21) has the dimension of the mark parameter \( \psi \).

Richards et al. (2019) (and more extensively Richards (2019) ) investigate the accuracy of the large sample chi-square distribution in the context of the LOB. Details of how to implement the score statistic efficiently computationally are provided. Comprehensive simulations based on realistic scenarios concerning the variety of potential marks encountered in the LOB context and their statistical properties (distributions, serial and cross dependencies) are summarised and demonstrate that for realistic sample sizes the chi-squared distribution of Theorem 1 provides sufficiently accurate null distribution (Type I) significance levels. The score test is then applied to efficiently detect the significant marks from a substantial list of potential relevant marks and confirms that the use of the score statistic and its limit distribution are useful for complex practical applications.

6 Local power

We now investigate what happens to the distribution of the score statistic when \( H_0 \) fails, that is when the mark process impacts the distribution of the jump times of the point process. We adopt the local power approach, which consists in considering the sequence of local alternatives \( H_T^1 : \psi_T^* = \gamma^*/\sqrt{T} \) for some unknown \( \gamma^* \). We therefore assume that the marks weakly impact the distribution of the jump times (with a magnitude of order \( 1/\sqrt{T} \)), so that for a given \( T > 0 \), the associated counting process is nearly a pure Hawkes process. Accordingly, our goal is to derive the asymptotic distribution of the score statistic under the local alternatives \( H_T^1 \). The main result of this section, Theorem 2, states that the related limit distribution is that of a non-central chi-squared variable, whose non-centrality parameter directly depends on \( \gamma^* \) and on the partial Fisher information matrix \( \Omega \) introduced in (18).

Of course, taking \( \gamma^* = 0 \) yields the null hypothesis so that Theorem 2 is strictly stronger...
than Theorem 1, although for the sake of clarity the case $γ^* > 0$ has been postponed to the present section.

Following Proposition 1 and the discussion below, we assume that we observe a sequence of marked Hawkes processes $(N^T_g)_{T > 0}$ starting from 0, and all defined on (and adapted to) the same probability space $(Ω, F, P)$ endowed with the filtration $(F_t)_{t ∈ R_+}$. Note that we adopt the notation $N^T_g$ because, in contrast with the null hypothesis, the point process now depends on $T$ through the parameter $ψ^*_T$. Moreover, we assume that all the marked Hawkes processes indexed by $T$ are generated by the random measure $\bar{N}_g$ on $R^2 × X$, such that the normalized boost function of $N^T_g$ is $g(., φ^*_T)$, that is, for any $t ∈ R_+$, $N^T_g$ admits the following stochastic intensity:

$$
\lambda^T_g (t; θ^*, φ^*, ψ^*_T) = η^* + \theta^* \int_{(0,t)×X} w(t - s; α^*) g(x; φ^*, ψ^*_T) N^T_g (ds × dx),
$$

for some unknown parameter $ν^*_T = (θ^*, φ^*, ψ^*_T)$. Recall the expression of the score statistic

$$
\hat{Q}_{γ^*, T} = \frac{∂ψ}{I^T_g (\hat{ϕ}_T)} I^T_{\hat{ϕ}} \frac{∂ψ}{I^T_g (\hat{ϕ}_T)} − \frac{∂ψ}{I^T_g (\hat{ϕ}_T)} I^T_{\hat{ϕ}}\frac{∂ψ}{I^T_g (\hat{ϕ}_T)},
$$

where we now make appear $γ^*$ in the notation to emphasize the fact that the statistic depends on the parameter, and where $I^T_g$ admits the same expression as in (7), replacing the pure Hawkes process $N(dt)$ by the counting process $N^T (dt) = N^T_g (dt, X)$. Similarly, in (22), $\hat{ϕ}_T = (\hat{Θ}_T, \hat{ϕ}_T, 0)$, where $\hat{Θ}_T$ is one maximizer of $I^T_g$ in the interior of $Θ$, and $\hat{ϕ}_T$ is again a consistent estimator of $ϕ^*$. As stated in Theorem 2 below, it turns out that under $H^T_1$, $\hat{Q}_{γ^*, T}$ tends to a non central chi-squared distribution, whose non-centrality parameter depends on $γ^*$ and on the inverse of the Fisher information matrix (at point $ψ = 0$, $Ω$). In order to ensure the convergence of $\hat{Q}_{γ^*, T}$, we make the following assumptions.

**Condition 5** For $p = (dim(Θ) + 1) ∨ 4$, we assume the existence of $ε > 0$ such that, defining $U = Θ × {φ^*} × B(0, ε)$ where $B(0, ε)$ is the open ball of radius $ε$,

$$
\sup_{T ∈ R_+} \sup_{t ∈ [0,T]} \sum_{i=0}^3 \mathbb{E} \left[ \sup_{v ∈ U} |∂^i_ν^T φ^T (t; v)|^p \right] < +∞.
$$

Moreover,

$$
\sup_{T ∈ R_+} \sup_{t ∈ [0,T]} \sum_{i=0}^2 \mathbb{E} \left[ \sup_{v ∈ U} |∂^i_ν^T φ^T (t; v)|^p \right] < +∞.
$$

Moreover, assume that there exists $ε > 0$ such that

$$
\mathbb{E} \sup_{ψ ∈ B(0, ε)} |∂_ψ g(y; φ^*, ψ)|^p < +∞,
$$

where $y$ represents the marginal distribution of $(y_s)_{s ∈ R}$. Finally, for $q ∈ \{1, 2\}$, defining $A = [a|3(η, θ)| s.t. (η, θ, α) ∈ Θ$, we assume the existence of $w$ such that for any $α ∈ A$, for any $t ≥ 0$, $w(t; α) ≤ \bar{w}(t)$, and

$$
\int_0^{+∞} \bar{w}(t)^q dt < +∞.
$$

Condition (24) is satisfied for the exponential decay function under the conditions stated above for $α$. For suitable choice of a compact parameter space for the power law decay function, a two parameter family of decay functions, the condition is also satisfied without placing undue restrictions on the parameter space.
Lemma 3 Condition 5 is satisfied for the exponential kernel case ($\dim(\Theta) = 3$) and for stationary marks satisfying (23).

Proof The proof follows exactly the same path as that of Lemma 2, replacing the fourth order moment condition on $G(x)$ by the local uniform condition (23).

We can now state the following theorem.

Theorem 2 Assume Conditions 1, 2, 3, 4 and 5. Under $H_{1}^{T} : \psi_{T}^{*} = \gamma^{*}/\sqrt{T}$, we have

$$\hat{Q}_{\gamma^{*}, T} \rightarrow^{d} \chi^{2}(\Omega^{1/2}\gamma^{*}),$$

where $\Omega$ is defined in (18), $\chi^{2}(\Omega^{1/2}\gamma^{*}) \sim \|Z\|^{2}$ with $Z \sim \mathcal{N}(\Omega^{1/2}\gamma^{*}, I_{r\times r})$, and $I_{r\times r}$ is the $r \times r$ identity matrix.

The proof is in Appendix D. Note that, of course, when $\gamma^{*} = 0$, Theorem 2 coincides with 1. In all generality however, the limit distribution now follows a biased chi-squared distribution, whose bias parameter is directly proportional to $\gamma^{*}$, which controls the deviation of our model from the null hypothesis.

Richards (2019) and Richards et al. (2019) provide extensive simulations about the power of the score test against alternatives that increasingly diverge from the null under a variety of situations derived from empirical study of the limit order book. These simulations confirm that, as expected, power increases monotonically to unity for increasingly more distant alternative parameter values in the boost function and increase more rapidly with increasing sample size. The local power result of Theorem 2 could be used to compute the power for alternatives close to the null hypothesis as an alternative to using simulation.

7 Conclusions and future extensions

In this paper we have derived the asymptotic distribution of the score test proposed for determining if marks have no impact on the intensity of a single Hawkes process against the alternative that marks impact the intensity through a boost functions selected from a quite general class of such functions. We prove that the asymptotic distribution under the null hypothesis that there is no impact of the proposed marks on the intensity process is the usual chi-squared distribution with degrees of freedom equal to the number of parameters specified for the marks boost function. These asymptotic results rely heavily on the large sample results for quasi-likelihood estimation of multivariate unmarked Hawkes process considered in Clinet and Yoshida (2017). In addition to their assumptions on the null hypothesis model specification and parameters, because the score test involves functions of the marks, the additional Conditions 3 and 4 are required. Condition 3 is shown to hold in the exponential decay case (see Lemma 2). Condition 4 is concerned with properties of the marks process alone.

The marks process can be quite general and includes marks obtained from observations on a continuous time vector valued process in which there is serial dependence as well as dependence between components of the mark vector. In applications to the limit order book, Richards et al. (2019) modelled the marks $x_{i}$ at event times $t_{i}$ as observations on a discrete time stationary process indexed by event index $i$ instead of the continuous time process $(y_{t})_{t \in \mathbb{R}}$. It is not obvious that the thinning method, used to construct the marked Hawkes process in Sect. 2, can also be used to construct a marked Hawkes process using such a discrete time marks process. Moreover, while it is possible to construct a non-stationary marked Hawkes
process with discrete time marks by iterating the intensity function from some initial time $t_0$, it is not clear whether there exists a stationary version of this process on $\mathbb{R}$.

Notwithstanding the difficulties of constructing a stationary marked Hawkes process using such a discrete time process, it is possible to proceed with deriving the score statistic formally as follows. The third term in the likelihood (4) is replaced by $\log f(x_1, \ldots, x_{NT}; \phi)$ where $f$ now denotes the joint density for the discrete time stationary time series $\{x_i\}$ in which actual event times are ignored and only the indices, $i$, of event times are needed to model serial dependence structure. Note that this leads to the objective function

$$l_g(\nu) = \int_{[0,T] \times \mathbb{X}} \log \lambda_g(t; \nu) N_g(dt \times dx) - \Lambda_g(T; \nu) + \log f(x_1, \ldots, x_{NT}; \phi)$$

(25)

to be maximised over the parameters. It is not clear that this can be written as an integral with respect to $N_g(dt \times dx)$ corresponding to the third term in (4). However this is not a formal likelihood, nor does it seem possible to define a stationary Hawkes process, as we did in Sect. 2 for the case where marks are drawn from a stationary discrete time process. Of course in the absence of serial dependence both (4) and (25) lead to the i.i.d. version (5) considered in the literature to date. Using (25) to define the score statistic leads to the same form for the components $\partial_\psi l_g(\hat{\nu}_T)$ and $I^{-1}_{\psi}$ used to define the score statistic $\hat{Q}_T$ in (19). Because of this Theorems 1 and 2 also apply under the same conditions except for Condition 4 which is modified to require the discrete time process of marks is a stationary, ergodic process in discrete time with a strictly positive definite autocovariance function.

For local power computations, we have also derived the non-central chi-squared limiting distribution for the score test statistic under a sequence of local alternatives with the boost parameter converging to the null hypothesis value at rate $T^{-1/2}$. More generally, establishing consistency of the score test would require a proof that the power tends to unity for any value of $\psi \neq 0$. However, showing this rigorously requires proving the ergodicity of the point process along with substantial extensions to existing asymptotic theory for likelihood estimation in marked Hawkes processes. The main technical challenge for establishing this is showing that the asymptotic score w.r.t. $\phi$ is non-degenerate. Here a major difficulty arises because the existence of multiple stationary values in the limiting likelihood function of $(\theta, \phi)$ when $\psi \neq 0$ cannot be ruled out easily.

Crucial to establishing the conditions required for the results of Clinet and Yoshida (2017) as well as our additional Condition 3 is the Markovian nature of the Hawkes intensity process with an exponential decay function. More general decay functions, such as linear combinations of polynomial-exponential kernels retain the Markov property (see Durante et al. (2016) for the unmarked case, and Clinet (2020) for the case where the marks process yields a Markovian representation of the point process) and so extension of above results should be straightforward. For other kernels, such as the power law decay function, the Markov property does not hold and hence extension of our results would require substantial and fundamental theory to extend known results in the literature firstly in the unmarked Hawkes processes and secondly in the marked case.

We close with a brief discussion of extension to the multivariate Hawkes point process. The point process with intensity given by equation (1) is referred to as the ‘vector valued case’ in Embrechts et al. (2011) and defines a scalar valued marked point process with vector valued marks. Embrechts et al. (2011) also consider the so called ‘multivariate marked point process’ consisting of $d$ processes in which the $j$th has intensity given (for the non-stationary form) by
\[ \lambda_{g,j}(t; \theta, \phi, \psi) = \eta_j + \sum_{k=1}^{d} \vartheta_{j,k} \int_{(0,t] \times X_k} w_j(t-s; \alpha_j)g_k(x_k; \phi_k, \psi_k)N_{g,k}(ds \times dx_k) \]  

(26)

where \( w_k \) are non-negative decay functions satisfying \( \int_0^\infty w_k(s; \alpha)ds = 1 \), the branching matrix \( Q \) with \( (j,k) \)th element \( Q_{(j,k)} = [\vartheta_{j,k}] \) has spectral radius strictly less than unity, and the marks associated with the \( k \)th process are independent and identically distributed with density \( f_k(x_k; \phi_k) \) and are independent of the past of the process (and hence marks associated with the other processes).

In the formulation presented in Embrechts et al. (2011) the boost functions associated with each component process are distinct and parameterised individually with parameters \( \psi_k \) for the \( k \)th component. The likelihood for this process is given in Embrechts et al. (2011) for the stationary version of (26). The analogue for our set-up is a straightforward extension of (5)

\[
\bar{\ell}_g(\nu) = \sum_{j=1}^{d} \left\{ \int_{[0,T] \times X_j} \log \lambda_{g,j}(t; \nu)N_{g,j}(dt \times dx) - A_{g,j}(T; \nu) \right\}.
\]

In this particular case of independent identically distributed marks in each component, the score vector and information matrix can easily be derived in the same fashion and as a result the extension of our score test statistic to this case is straightforward.

Clinet and Yoshida (2017) show that for the multivariate unmarked Hawkes process the likelihood estimates are consistent and asymptotically normally distributed. Assuming that our assumptions concerning the form of the boost function and Condition 1 applies to each of the \( d \) boost functions \( g_j \) and that Condition 4 similarly holds for each component mark then the extension of Lemma 2, Theorem 1 and Theorem 2 to the multivariate case will follow from the general results of Clinet and Yoshida (2017) for the multivariate unmarked Hawkes process in the same way that we established these results for the vector valued marks in a scalar process case considered in this paper. In particular the score test of the null hypothesis that none of the component intensity processes is impacted by marks versus the alternative that at least one of these is so impacted, has a large sample \( \chi^2_{(d)} \)-distribution.

More complex formulations of the marks processes in the multivariate case and the way in which they enter the boost functions are possible. For example marks which are observations on a stationary process or marks which are vector valued for each component process could be considered. We anticipate that the distribution of the score statistic will extend in a straightforward way. However the details need to be worked out and would be the basis for future research and development in applications.

Acknowledgements K-A Richards gratefully acknowledges PhD scholarship support by Boronia Capital Pty. Ltd., Sydney, Australia. The research of S. Clinet is supported by a special grant from Keio University. W. T.M. Dunsmuir was supported by travel funds from the Faculty of Sciences, University of New South Wales. The authors thank the referees for comments and suggestions that have improved the clarity and scope of the paper.

Declarations

Conflict of Interest Statement On behalf of all authors, the corresponding author states that there is no conflict of interest.
A proof of lemma 1 and proposition 1

Proof (Lemma 1) By construction, we first show that $\overline{N}_g$, seen as a marked point process on $\mathbb{R} \times (\mathbb{R} \times \mathbb{X})$ is compensated by $ds \times (du \times F_s(d\mathbf{x}))$ where $F_s(d\mathbf{x})$ is the conditional distribution of $\mathbf{y}$, with respect to the filtration $\mathcal{F}_s^{x}$. We recall that by definition, all we have to show is that for any non-negative measurable predictable process $W$ (see Jacod and Shiryaev 2013, Theorem II.1.8) we have

$$
\mathbb{E}\left[\int_{\mathbb{R} \times \mathbb{X}} W(s, u, x)\overline{N}_g(ds \times du \times d\mathbf{x})\right] = \mathbb{E}\left[\int_{\mathbb{R} \times \mathbb{R} \times \mathbb{X}} W(s, u, x)dsduF_s(d\mathbf{x})\right].
$$

By a monotone class argument, it is sufficient to prove the martingale property for $W(s, u, x)(\omega) = 1_{\{(\eta', t) \in F\}}(s, \omega)1_{A \times B}(u, x)$ where $\eta' < t$, $F \in \mathcal{F}_{\eta'}$, $A \in B(\mathbb{R})$ and $B \in \mathcal{X}$. We have

$$
\begin{align*}
\mathbb{E}\left[\int_{\mathbb{R} \times \mathbb{X}} W(s, u, x)\overline{N}_g(ds \times du \times d\mathbf{x})\right] & = \mathbb{E}\left[1_{F} \sum_{i \in \mathbb{Z}, \eta' < t_i \leq t} 1_{A}(u_i)1_{B}(y_i)\right] \\
& = \mathbb{E}\left[1_{F} \mathbb{E}\left[\sum_{i \in \mathbb{Z}, \eta' < t_i \leq t} 1_{A}(u_i)1_{B}(y_i) \mid \mathcal{F}_{\eta}'\right]\right] \\
& = \mathbb{E}\left[1_{F} \mathbb{E}\left[\sum_{i \in \mathbb{Z}, \eta' < t_i \leq t} 1_{A}(u_i)1_{B}(y_i) \mid \mathcal{F}_{\eta}' \vee \mathcal{F}_{\eta}^{y}\right]\right] \\
& = \mathbb{E}\left[1_{F} \mathbb{E}\left[\sum_{i \in \mathbb{Z}, \eta' < t_i \leq t} 1_{A}(u_i)1_{B}(y_i) \mid \mathcal{F}_{\eta}' \right]\right] \\
& = \mathbb{E}\left[1_{F} \mathbb{E}\left[\int_{\{(\eta', t) \in F\}} 1_{A \times B}(u, x)F_\eta(d\mathbf{x})\overline{N}(ds \times du) \mid \mathcal{F}_{\eta}'\right]\right] \\
& = \mathbb{E}\left[1_{F} \mathbb{E}\left[\int_{\{(\eta', t) \in F\}} 1_{A \times B}(u, x)F_\eta(d\mathbf{x})d\mathbf{s}du \mid \mathcal{F}_{\eta}'\right]\right] \\
& = \mathbb{E}\left[\int_{\mathbb{R} \times \mathbb{R} \times \mathbb{X}} W(s, u, x)F_\eta(d\mathbf{x})d\mathbf{s}du\right],
\end{align*}
$$

where at the fourth line we have used the independence of $\mathbf{y}$ and $\overline{N}$, and at the seventh line we have used again the independence of $\overline{N}$ from $\mathbf{y}$ along with the fact that $\overline{N}$ is a Poisson process with intensity measure equal to $ds \times du$.

Proof (Proposition 1) We construct by thinning and a fixed point argument the marked point process $N_\infty^g$. Define $\lambda_0(\eta) = \eta^s$. By induction, we then define the sequences of processes $(N_{\eta}^g)_{\eta \in \mathbb{N}}$ and $(\lambda_\eta^g)_{\eta \in \mathbb{N}}$ as follows. For $n \in \mathbb{N}$, we let $N_{\eta}^g$ and $\lambda_\eta^g$ be such that for any $t', t \in \mathbb{R}$ with $t' < t$ and $A \in \mathcal{X}$
\begin{align}
N^\infty_{g,n}((t', t] \times A) &= \int_{(t', t] \times \mathbb{R} \times A} 1_{\{0 \leq u \leq \lambda^\infty_{g,n}(s)\}} N_g(ds \times du \times dx), \quad (27a) \\
\lambda^\infty_{g,n+1}(t) &= \eta^* + \vartheta^* \int_{(-\infty, t] \times \mathbb{X}} w(t-s)g(x)N^\infty_{g,n}(ds \times dx). \quad (27b)
\end{align}

Clearly (27a) uniquely defines a measure on \( \mathbb{R} \times \mathbb{X} \). By taking conditional expectations throughout (27a) it is immediate to see that \( t \rightarrow \lambda^\infty_{g,n}(t) \) is the stochastic intensity of the counting process \( t \rightarrow N^\infty_{g,n}((t', t] \times \mathbb{X}) \). Moreover, by positivity of \( w(t-s)g(x) \), we immediately deduce that for any \( t' < t \) and \( A \in \mathcal{X} \) the quantities \( N^\infty_{g,n}((t', t] \times A) \) and \( \lambda^\infty_{g,n}(t) \) are increasing with \( n \), so that we may define \( N^\infty_{g} \) and \( \lambda^\infty_{g} \) the (possibly infinite) point wise limits of \( N^\infty_{g,n} \) and \( \lambda^\infty_{g,n} \). Next, by the monotone convergence theorem, taking the limit \( n \rightarrow +\infty \) in the above equations yields the representation

\begin{align}
N^\infty_g((t', t] \times A) &= \int_{(t', t] \times \mathbb{R} \times A} 1_{\{0 \leq u \leq \lambda^\infty_g(s)\}} N_g(ds \times du \times dx), \quad (28a) \\
\lambda^\infty_g(t) &= \eta^* + \vartheta^* \int_{(-\infty, t] \times \mathbb{X}} w(t-s)g(x)N^\infty_g(ds \times dx). \quad (28b)
\end{align}

In particular, the first equation shows that \( t \rightarrow \lambda^\infty_g(t) \) is the stochastic intensity of \( t \rightarrow N^\infty_g((t', t] \times \mathbb{X}) \) for \( t' < t \) and the second equation proves that \( \lambda^\infty_g \) has the desired shape. All we have to check is the finiteness of the two limits in (28a)-(28b). Let \( \rho_n(t) = \mathbb{E}[\lambda^\infty_{g,n} - \lambda^\infty_{g,n-1}(t)] \). We have

\[
\rho_n(t) = \vartheta^* \mathbb{E} \left[ \int_{(-\infty, t] \times \mathbb{X}} w(t-s)g(x)\{N^\infty_{g,n-1} - N^\infty_{g,n-2}\}(ds \times dx) \right] \\
= \vartheta^* \mathbb{E} \left[ \int_{(-\infty, t] \times \mathbb{R} \times \mathbb{X}} w(t-s)g(x)1_{\{\lambda^\infty_{g,n-1}(s) < u \leq \lambda^\infty_{g,n-2}(s)\}} N_g(ds, du, dx) \right] \\
= \vartheta^* \mathbb{E} \left[ \int_{(-\infty, t] \times \mathbb{X}} w(t-s)g(x)\{\lambda^\infty_{g,n-1}(s) - \lambda^\infty_{g,n-2}(s)\}ds F_x(dx) \right] \\
= \vartheta^* \mathbb{E} \left[ \int_{(-\infty, t] \times \mathbb{X}} w(t-s)\mathbb{E}[g(y_s)|F^\mathcal{Y}_{s-}]\{\lambda^\infty_{g,n-1}(s) - \lambda^\infty_{g,n-2}(s)\}ds \right] \\
\leq C \vartheta^* \int_{(-\infty, t]} w(t-s)\rho_{n-1} ds,
\]

where we have used at the third step that \( \overline{N}_g \), seen as a marked point process on \( \mathbb{R} \times (\mathbb{R} \times \mathbb{X}) \) admits \( ds du F_x(dx) \) as \( \mathcal{F}_s \)-compensator where \( F_x(dx) \) is the conditional distribution of \( y_s \) with respect to \( \mathcal{F}^\mathcal{Y}_{s-} \) by Lemma 1, and at the fifth step have used the \( \mathcal{F}_s \)-measurability of the stochastic intensities and that \( \mathbb{E}[g(y_s)|F_{s-}] = \mathbb{E}[g(y_s)|\mathcal{F}^\mathcal{Y}_{s-}] \leq C \) (by independence of \( y \) and \( \overline{N} \)). From here, we deduce that \( \sup_{s \in (-\infty, t]} \rho_n(s) \leq C \vartheta^* \sup_{s \in (-\infty, t]} \rho_{n-1}(s) \) since \( \int_0^{+\infty} w(s)ds = 1 \). By a similar calculation, we also have \( \rho_1(t) \leq C \vartheta^* \eta^* \), so that by an immediate induction \( \sup_{s \in (-\infty, t]} \rho_n(s) \leq (C \vartheta^*)^n \eta^* \). Therefore, \( \mathbb{E}\lambda^\infty_g(t) = \eta^* + \sum_{k=1}^{+\infty} \rho_k(t) \leq \eta^*/(1 - C \vartheta^*) < +\infty \), which implies the almost sure finiteness of both \( N^\infty_g \) (on any set of the form \( (t', t] \times \mathbb{X}, -\infty < t' < t < +\infty \)) and \( \lambda^\infty_g(t) \). This proves the first claim. Now we prove the second point, and show first that for any \( n \in \mathbb{N} \), \( \pi_n^\infty(ds \times dx) = \lambda^\infty_{g,n}(s)ds \times F_x(dx) \) is the \( \mathcal{F}_s \)-compensator of \( N^\infty_{g,n} \). Let \( W \) be a non-negative measurable predictable process on \( \mathbb{R} \times \mathbb{X} \). By (27a), we have

\[
\mathbb{E}\left[ \int_{\mathbb{R} \times \mathbb{X}} W(s, x)N^\infty_{g,n}(ds \times dx) \right] \]
where we have used that $\tilde{N}_g$ admits $ds \times (du \times F_s(d\mathbf{x}))$ as $\mathcal{F}_s$-compensator by Lemma 1. Since $W$ is arbitrary, this proves that $\pi_n^n$ is the compensator of $N_{g,n}$. Moreover, taking the limit $n \to +\infty$ in the above expectations and using again the monotone convergence theorem yields that $\pi_\infty(ds \times d\mathbf{x})$ is the compensator of $N_g$. Finally, the third claim comes from the stationarity of $\lambda_{g,n}^\infty$ and $N_{g,n}^\infty$ which is in turn a consequence of the stationarity of $\tilde{N}_g$ and $y$.

\[ \square \]

**B proof of lemma 2**

For any $c \in \mathbb{R}^r$ we denote the linear combinations $G_c(\mathbf{x}) = c^T G(\mathbf{x})$ and similarly for $\hat{G}_c(\mathbf{x})$. Recall that under $H_0$, the marked point process $N_g$ has event intensity identical to that of $N$ defined in (3). Marks are observed at the event times of this process but do not impact the intensity of it. For the exponential decay specification, since $\dim(\Theta) = 3$, we need to show Condition 3 for $p = 4$, i.e that

$$\sup_{t \in \mathbb{R}^+} \mathbb{E} \left[ \sup_{\theta \in \Theta} \left| \frac{\partial^{i} \theta}{\partial \theta} \int_{[0,t] \times \mathbb{X}} w(t-s; \alpha) G_c(\mathbf{x}) N_g(ds \times d\mathbf{x}) \right|^p \right] < \infty$$

for $i = 0, 1, 2$. Notice that only the derivatives with respect to $\theta$ and $\alpha$ are required. These derivatives are linear combinations of terms of the form

$$\Phi^k \int_{[0,t] \times \mathbb{X}} \frac{\partial^{i} \theta}{\partial \theta} w(t-s; \alpha) G_c(\mathbf{x}) N_g(ds \times d\mathbf{x})$$

for $i = 0, 1, 2$ and $k = 0, 1$, and with $w(t-s; \alpha) = e^{-\alpha(t-s)}$. Since $\Theta$ is bounded, we consider the integrals which are finite combinations of terms of the form

$$\int_{[0,t] \times \mathbb{X}} (t-s)^i e^{-\alpha(t-s)} G_c(\mathbf{x}) N_g(ds \times d\mathbf{x}) \leq \int_{[0,t] \times \mathbb{X}} (t-s)^i e^{-\alpha(t-s)} G_c(\mathbf{x}) N_g(ds \times d\mathbf{x})$$

for $i = 0, 1, 2$. Therefore, we need to show to conclude the proof that

$$\sup_{t \in \mathbb{R}^+} \mathbb{E} \left[ \int_{[0,t] \times \mathbb{X}} (t-s)^i e^{-\alpha(t-s)} G_c(\mathbf{x}) N_g(ds \times d\mathbf{x}) \right]^p < \infty, \quad i = 0, 1, 2.$$
for some finite constant $C$ and where the compensator of $N_g(ds \times d\mathbf{x})$ is $\lambda(s; \theta^*) F_s(d\mathbf{x})ds$ where $F_s(d\mathbf{x})$ is the conditional distribution of $\mathbf{y}_s$ with respect to $\mathcal{F}_s^Y$.

First define the probability measure $\mu(ds) = (\int_{0}^{t} f_{i,t}(du))^{-1} f_{i,t} ds$ on $[0, t]$, and apply Jensen’s inequality to the second term to get

$$
\mathbb{E} \left| \int_{[0,t] \times \mathbb{X}} f_{i,t}(s)|G_c(\mathbf{x})|F_s(d\mathbf{x})\lambda(s; \theta^*) ds \right|^4
= \left( \int_{[0,t]} f_{i,t}(s)ds \right)^4 \mathbb{E} \left| \int_{[0,t] \times \mathbb{X}} |G_c(\mathbf{x})|F_s(d\mathbf{x})\lambda(s; \theta^*) \mu(ds) \right|^4
\leq \left( \int_{[0,t]} f_{i,t}(s)ds \right)^3 \mathbb{E} \int_{[0,t]} f_{i,t}(s) \left| \int_{\mathbb{X}} |G_c(\mathbf{x})|F_s(d\mathbf{x})\lambda(s; \theta^*) \right|^4ds
\leq \int_{[0,t]} f_{i,t}(s)ds \mathbb{E} \left| \int_{\mathbb{X}} |G_c(\mathbf{x})|F_s(d\mathbf{x})\lambda(s; \theta^*) \right|^4\mathbb{E} \left[ |G_c(\mathbf{y}_s)| |\mathcal{F}_{s-}^Y \right]^4 ds
\leq C,
$$

Where we have used the independence of $\mathbf{y}$ and $N_g$, the fact that $\mathbb{E}[|G_c(\mathbf{y}_s)| |\mathcal{F}_{s-}^Y] < \mathbb{E}[|G_c(\mathbf{y}_s)|^4] < K$ for some constant $K > 0$, and $\sup_{t \in \mathbb{R}^+} \mathbb{E}[\int_{(0,t]} f_{i,t}(s)\lambda(s; \theta^*) ds] < \infty$ by (Clenet and Yoshida 2017, Lemma A.5). Consider now the first expected value. Using Davis-Burkholder-Gundy inequality we have arguing similarly to (Clenet and Yoshida 2017, Lemma A.2), for some constant $C < \infty$ not necessarily the same as above,

$$
\mathbb{E} \left| \int_{[0,t] \times \mathbb{X}} f_{i,t}(s)|G_c(\mathbf{x})|\tilde{N}_g ds \times d\mathbf{x} \right|^4
\leq C \mathbb{E} \left| \int_{[0,t] \times \mathbb{X}} f_{i,t}(s)^2|G_c(\mathbf{x})|^2\tilde{N}_g ds \times d\mathbf{x} \right|^2
\leq 2C \mathbb{E} \left| \int_{[0,t] \times \mathbb{X}} f_{i,t}(s)^2|G_c(\mathbf{x})|^2\tilde{N}_g ds \times d\mathbf{x} \right|^2
+ 2C \mathbb{E} \left| \int_{[0,t] \times \mathbb{X}} f_{i,t}(s)^2|G_c(\mathbf{x})|^2\lambda(s; \theta^*) F_s(d\mathbf{x})ds \right|^2.
$$

Similarly to the previous argument the second term is uniformly bounded because

$$
\sup_{t \in \mathbb{R}^+} \mathbb{E} \left| \int_{(0,t]} f_{i,t}(s)^2\lambda(s; \theta^*) ds \right|^2 < \infty
$$

by (Clenet and Yoshida 2017, Lemma A.5) and $\mathbb{E}[|G(\mathbf{y}_s)^2| |\mathcal{F}_{s-}^Y] \leq \mathbb{E}[G_c(\mathbf{y}_s)^4] < K$ for some constant $K > 0$. For the first term we have

$$
\mathbb{E} \left| \int_{[0,t] \times \mathbb{X}} f_{i,t}(s)^2|G_c(\mathbf{x})|^2\tilde{N}_g ds \times d\mathbf{x} \right|^2
= \mathbb{E} \int_{[0,t] \times \mathbb{X}} f_{i,t}(s)^4|G_c(\mathbf{x})|^4\lambda(s; \theta^*) F_s(d\mathbf{x})ds
= \mathbb{E} \int_{0}^{t} f_{i,t}(s)^4\lambda(s; \theta^*) \mathbb{E}[G_c(\mathbf{y}_s)^4] ds,
$$
where we have used the independence of $y$ and $N_g$. Now, $E|G_c(y_s)|^4 < \infty$ and, once more by (Clinet and Yoshida 2017, Lemma A.5) we have
\[
\sup_{t \in \mathbb{R}^+} \mathbb{E} \int_{(0,t)} f_t, t(s)^4 \lambda(s; \theta^*) ds < \infty
\]
which completes the proof. \hfill \Box

### C proof of theorem 1

Define, for any fixed $c \in \mathbb{R}^r$,
\[
U(t; \theta, \phi) = c^T \partial_\psi \lambda_g(t; \theta, \phi, 0) = \partial \int_{[0,t] \times \mathbb{X}} w(t - s; \alpha) c^T G(x; \phi) N_g(ds \times dx). \tag{29}
\]

This notation is used repeatedly in the proof of the theorem as well as the lemmas used. The proof follows somewhat closely that of Clinet and Yoshida (2017). We first consider the normalized process corresponding to (15) and for any non zero vector of constants $c \in \mathbb{R}^r$ define the process in $u \in [0, 1]$
\[
S_u^T = \frac{1}{\sqrt{T}} \int_{[0,uT]} \lambda(t; \theta^*)^{-1} c^T \partial_\psi \lambda_g(t; \nu^*) \tilde{N}(dt)
= \frac{1}{\sqrt{T}} \int_{[0,uT]} \lambda(t; \theta^*)^{-1} U(t; \theta^*, \phi^*) \tilde{N}(dt). \tag{30}
\]

Note that $S_1^T = \frac{1}{\sqrt{T}} c^T \partial_\psi I_g(\nu^*)$. Similarly to Clinet and Yoshida (2017), we establish a functional CLT when $T \to \infty$.

The proof of this theorem proceeds via several lemmas. Convergence throughout is with $T \to \infty$. The first lemma is concerned with the ergodic properties of $U(t; \theta, \phi)$ defined in (29) when $\phi = \phi^*$, is fixed at the true value in which case we further abbreviate notation to $U(t; \theta) = U(t; \theta, \phi^*)$.

**Lemma 4** Under $H_0$, there exists a stationary marked Hawkes point process starting from $-\infty, N_g^\infty$, on the original probability space $(\Omega, \mathcal{F}, \mathbb{P})$, adapted to $\mathcal{F}_t$, such that: (i) $N^\infty = N_g^\infty(\cdot, \mathbb{X})$ and $y$ are independent. (ii) the stochastic intensity of $N^\infty$ admits the representation
\[
\lambda^\infty(t) = \eta^* + \theta^* \int_{(-\infty,t)} w(t - s, \alpha^*) N^\infty(ds).
\]

Moreover, let us define
\[
U^\infty(t; \theta^*) = \theta^* \int_{(-\infty,t) \times \mathbb{X}} w(t - s; \alpha^*) G_c(x) N^\infty_g(ds \times dx).
\]

Then, the joint process $(\lambda^\infty, U^\infty(\cdot; \theta^*))$ is stationary ergodic. Finally we have the convergence
\[
E|\lambda(t, \theta^*) - \lambda^\infty(t)| + E|U(t; \theta^*) - U^\infty(t; \theta^*)| \to 0, t \to +\infty. \tag{31}
\]

**Proof** The existence of $N_g^\infty$ along with property (ii) are direct consequences of Proposition 1. The fact that $N_g^\infty$ can be constructed on the same probability space as $N_g$ is ensured by building $N_g^\infty$ using the same canonical process $\tilde{N}_g$. The independence property (i) comes from the fact that under $H_0$ we have $\psi = 0$ so that $g(\cdot, \cdot, 0) = 1$ and the marks process $y$ does
not impact the marginal point process $N^\infty$. Next, since $y$ is ergodic by assumption, and the unmarked Hawkes process of jumps $N^\infty$ is stationary ergodic by assumption, and since both processes are independent from each other, the joint process $(N^\infty, y)$ is stationary ergodic as well. Since for any $t \in \mathbb{R}$, $(\lambda^\infty(t), U^\infty(t, \theta^*))$ admits a stationary representation and given the form of $(\lambda^\infty(t), U^\infty(t, \theta^*))_{t \in \mathbb{R}}$, we can deduce that they are also ergodic by Lemma 10.5 from Kallenberg (2006). Finally, we show (31). We first deal with the convergence of $\int_{-\infty}^{\infty} t w(t-s; \alpha^*) \lambda^\infty(s)ds$, and following the same reasoning as for the proof of Proposition 4.4 (iii) in Clinet and Yoshida (2017), some algebraic manipulations easily lead to the inequality

$$f(t) \leq r(t) + \vartheta^* w(.; \alpha^*) \ast f(t), \quad t \geq 0$$

where for two functions $a$ and $b$, and $t \in \mathbb{R}_+$, $a \ast b(t) = \int_0^t a(t-s)b(s)ds$ whenever the integral is well-defined. Iterating the above equation, we get for any $n \in \mathbb{N}$

$$f(t) \leq \sum_{k=0}^{n} \vartheta^* \ast^k w(.; \alpha^*) \ast^k r(t) + \vartheta^* \ast^{n+1} w(.; \alpha^*) \ast^{n+1} f(t).$$

Using the fact that $\int w(.; \alpha^*) = 1$, $\vartheta^* < 1$ and using Young’s convolution inequality we easily deduce that the second term tends to 0 as $n$ tends to infinity, so that $f$ is dominated by $R \ast r$ where $R := \sum_{k=0}^{+\infty} \vartheta^* \ast^k w(.; \alpha^*) \ast^k$. Note that $R$ is finite and integrable since $\int_{0}^{+\infty} R(s)ds \leq 1/(1 - \vartheta)$. We first prove that $r(t) \to 0$. To do so, note that $r(t) = \mathbb{E} [\lambda^\infty(0)] \int_{t}^{+\infty} w(u; \alpha^*)du \to 0$ since $w(., \alpha^*)$ is integrable. Now, since $R \ast r(t) = \int_{0}^{t} R(s)r(t-s)ds$, and $R(s)r(t-s)$ is dominated by $\sup_{u \in \mathbb{R}_+} r(u)R(s)$ which is integrable, we conclude by the dominated convergence theorem that $f(t) \leq R \ast r(t) \to 0$. Finally, we prove that $g(t) := \mathbb{E}[U(t; \theta^*) - U^\infty(t; \theta^*)] \to 0$. We have

$$g(t) \leq \mathbb{E} \left[ \int_{(0,t) \times \mathbb{X}} w(t-s; \alpha^*) G_c(x) (N_g - N^\infty_g)(ds \times dx) \right]$$

$$+ \mathbb{E} \left[ \int_{(-\infty,0) \times \mathbb{X}} w(t-s; \alpha^*) G_c(x) N^\infty_g(ds \times dx) \right]$$

$$\leq \mathbb{E} \int_{(0,t) \times \mathbb{X}} w(t-s; \alpha^*) |G_c(x)||N_g - N^\infty_g|(ds \times dx)$$

$$+ \mathbb{E} \int_{(-\infty,0) \times \mathbb{X}} w(t-s; \alpha^*) |G_c(x)| N^\infty_g(ds \times dx)$$

$$\leq \mathbb{E}[G_c(x)] \int_{(0,t)} w(t-s; \alpha^*) f(s) ds$$

$$+ \mathbb{E}[G_c(x)] \mathbb{E} \left[ \lambda^\infty(0) \right] \int_{t}^{+\infty} w(u; \alpha^*) du.$$
Proof Similarly to (Clinet and Yoshida 2017, proof of Lemma 3.13) we first show that
\[
\langle S^T, S^T \rangle_u = \frac{1}{T} \int_{[0,uT]} \lambda(t; \theta^*)^{-1} U(t; \theta^*)^2 dt
\]
converges in probability to \(uc^T \Omega c\). Introducing \(\lambda^\infty, U^\infty\) as in Lemma 4, we need to show that
\[
\frac{1}{T} \int_{[0,uT]} \{\lambda(t; \theta^*)^{-1} U(t; \theta^*)^2 - \lambda^\infty(t)^{-1} U^\infty(t; \theta^*)^2\} dt \to^P 0.
\]
Using the boundedness of \(\lambda(t; \theta^*)^{-1}\) and \(\lambda^\infty(t)^{-1}\), we have the domination
\[
A_t := \left| \frac{U(t; \theta^*)^2}{\lambda(t; \theta^*)} - \frac{U^\infty(t; \theta^*)^2}{\lambda^\infty(t)} \right| \leq K \left| U(t; \theta^*)^2 - U^\infty(t; \theta^*)^2 \right| + KU^\infty(t; \theta^*)^2 |\lambda(t; \theta^*) - \lambda^\infty(t)|,
\]
for some constant \(K > 0\). By Lemma 4, we thus have \(A_t \to^P 0\). Moreover, since by Condition 3, \(U(t; \theta^*, \phi^*)\) and \(U^\infty(t; \theta^*)\) are \(L^{2+\epsilon}\) bounded for some \(\epsilon > 0\), and \(\lambda(t; \theta^*)\) and \(\lambda^\infty(t)\) are \(L^p\) bounded for any \(p > 1\), we deduce that \(\mathbb{E}|A_t| \to 0\). This, in turn, easily implies that \(\mathbb{E}|T^{-1} \int_0^u A_t dt| \to 0\), and thus we get (32). By the ergodicity property of Lemma 4, we also have
\[
\frac{1}{T} \int_{[0,uT]} \lambda^\infty(t)^{-1} U^\infty(t; \theta^*)^2 dt \to^P u\mathbb{E} \left[ \lambda^\infty(0)^{-1} U^\infty(0; \theta^*)^2 \right] = uc^T \Omega c,
\]
where \(\Omega = \mathbb{E}[\lambda^\infty(0)^{-1} \partial_{\phi} \lambda^\infty(0) \partial_{\phi} \lambda^\infty(0)^3]\), which proves our claim. Note that the fact that \(\Omega\) corresponds to the limit (18) is proved in Lemma 7 below.

To prove that \(\Omega\) is positive definite note that \(\mathbb{E}\left[ \lambda^\infty(0)^{-1} U^\infty(0; \theta^*)^2 \right]\) can be computed by first calculating the conditional expectations of the \(G_c(y^\infty_i)G_c(y^\infty_j)\) terms (appearing in \(U^\infty(0; \theta^*)^2\)) given the event times to get for any non-zero \(c \in \mathbb{R}^r\)
\[
c^T \Omega c = \mathbb{E} \left[ \lambda^\infty(0)^{-1} \sum_{t_i^\infty < 0} \sum_{t_j^\infty < 0} w(-t_i^\infty; \alpha^*) w(-t_j^\infty; \alpha^*) \gamma_{G_c}(t_i^\infty - t_j^\infty) \right]
\]
where \(t_i^\infty\) are jump times of \(N^\infty\). Now, under Condition 4 the quadratic form is positive almost surely as is \(\lambda^\infty(0)^{-1}\) and hence the expectation is positive proving the claim.

Next, for Lindeberg’s condition, for any \(a > 0\), similarly to Clinet and Yoshida (2017)
\[
\mathbb{E} \sum_{s \leq u} (\Delta S^T_s)^2 1_{[\Delta S^T_s > a]} \leq \frac{1}{a^2} \mathbb{E} \sum_{s \leq u} (\Delta S^T_s)^4
\]
\[
= \frac{1}{a^2} \mathbb{E} \int_{[0,uT]} \left| \frac{1}{\sqrt{T}} \lambda(t; \theta^*)^{-1} U(t; \theta^*) \right|^4 N(dt)
\]
\[
= \frac{1}{a^2 T^2} \mathbb{E} \int_{[0,uT]} \lambda(t; \theta^*)^{-3} |U(t; \theta^*)|^4 dt
\]
\[
\leq \frac{uK}{T} \sup_{t \in \mathbb{R}^+} \mathbb{E} |U(t; \theta^*)|^4 \to 0,
\]
where we have used Condition 3 along with the boundedness of \(\lambda(t; \theta^*)^{-1}\). As in Clinet and Yoshida (2017), application of (Jacod and Shiryaev 2013, 3.24 chapter VIII) gives the required functional CLT. \(\square\)
Lemma 6 Under $H_0$, we have

$$\frac{1}{\sqrt{T}}(\partial_\mu l_g(\hat{\nu}_T) - \partial_\mu l_g(\nu^*)) \to^P 0.$$  

Proof Rewrite

$$\frac{1}{\sqrt{T}} c^T(\partial_\mu l_g(\hat{\nu}_T) - \partial_\mu l_g(\nu^*))$$

$$= \frac{1}{\sqrt{T}} c^T(\partial_\mu l_g(\hat{\theta}_T, \hat{\phi}_T, 0) - \partial_\mu l_g(\hat{\theta}_T, \phi^*, 0))$$

$$+ \frac{1}{\sqrt{T}} c^T(\partial_\mu l_g(\hat{\theta}_T, \phi^*, 0) - \partial_\mu l_g(\theta^*, \phi^*, 0)). \tag{33}$$

Consider the first term in (33). Recall that $G_c(x) = G_c(x) = \hat{\mu}_H - \mu_H$, we have

$$U(t; \hat{\theta}_T, \hat{\phi}_T) - U(t; \hat{\theta}_T, \phi^*)$$

$$= c^T \partial_\mu \lambda_g(t; \hat{\theta}_T, \hat{\phi}_T, 0) - c^T \partial_\mu \lambda_g(t; \hat{\theta}_T, \phi^*, 0)$$

$$= \hat{\theta}_T \int_{[0,t] \times \mathbb{X}} w(t-s; \hat{\alpha}_T)[\hat{G}_c(x) - G_c(x)] N_g(ds \times dx)$$

$$= \hat{\theta}_T \int_{[0,t]} w(t-s; \hat{\alpha}_T) N(ds) c^T(\hat{\mu}_H - \mu_H)$$

giving

$$\frac{1}{\sqrt{T}} c^T(\partial_\mu l_g(\hat{\theta}_T, \hat{\phi}_T) - \partial_\mu l_g(\hat{\theta}_T, \phi^*))$$

$$= \frac{1}{\sqrt{T}} \hat{\theta}_T \int_{[0,T]} \lambda(t; \hat{\theta}_T)^{-1} \int_{[0,t]} w(t-s; \hat{\alpha}_T) N(ds)[N(dt) - \lambda(t; \hat{\theta}_T)dt] c^T(\hat{\mu}_H - \mu_H).$$

Now by Condition 4, $\hat{\mu}_H - \mu_H \to^P 0$. Also, using the consistency of the quasi likelihood estimates for the unmarked process, $\hat{\phi} \to^P \phi^*$. Finally

$$\frac{1}{\sqrt{T}} \int_{[0,T]} \lambda(t; \hat{\theta}_T)^{-1} \int_{[0,t]} w(t-s; \hat{\alpha}_T) N(ds)[N(dt) - \lambda(t; \hat{\theta}_T)dt]$$

is precisely the same as the derivative of the nonboosted likelihood w.r.t. the branching ratio parameter $\phi$ and it converges in distribution to a normal random variable directly from (Clinet and Yoshida 2017, Proof of Theorem 3.11). Hence the first term in (33) converges to zero in probability.

Consider the second term in (33) which is written as

$$\frac{1}{\sqrt{T}} c^T(\partial_\mu l_g(\hat{\theta}_T, \phi^*, 0) - \partial_\mu l_g(\hat{\theta}_T, \phi^*, 0))$$

$$= \frac{1}{\sqrt{T}} \int_{[0,T]} \lambda(t; \hat{\theta}_T)^{-1} U(t; \hat{\theta}_T) N(dt) - \int_{[0,T]} U(t; \hat{\theta}_T)dt$$

$$- \frac{1}{\sqrt{T}} \int_{[0,T]} \lambda(t; \theta^*)^{-1} U(t; \theta^*) N(dt) - \int_{[0,T]} U(t; \theta^*)dt$$

$$= \frac{1}{T} \int_{[0,T]} \partial_\theta \lambda(t; \hat{\theta}_T) \int_{[0,T]} \partial_\theta U(t; \hat{\theta}_T) dt \sqrt{\hat{\theta}_T - \theta^*}$$
using a first order Taylor series expansion where \( \hat{\theta}_T \in [\theta^*, \hat{\theta}_T] \). By the central limit theorem in Clinet and Yoshida (2017) \( \sqrt{T}(\hat{\theta}_T - \theta^*) \) is asymptotically normal. We show that the term multiplying this converges to zero in probability using a similar argument as to that in (Clinet and Yoshida 2017, Proof of Lemma 3.12). Now, at any \( \theta \) we have

\[
\frac{1}{T} \left\{ \int_{[0,T]} \partial_\theta \{ \lambda(t; \theta)^{-1} U(t; \theta) \} \right\} N(dt) - \frac{1}{T} \int_{[0,T]} \partial_\theta U(t; \theta) dt
\]

\[
= \frac{1}{T} \int_{[0,T]} \partial_\theta \{ \lambda(t; \theta)^{-1} U(t; \theta) \} \tilde{N}(dt) - \frac{1}{T} \int_{[0,T]} \lambda(t, \theta)^{-2} \partial_\theta \lambda(t; \theta) U(t; \theta) \lambda(t; \theta^*) dt
\]

\[
- \frac{1}{T} \int_{[0,T]} \partial_\theta U(t; \theta) \lambda(t; \theta)^{-1} [\lambda(t; \theta) - \lambda(t; \theta^*)] dt
\]

These three terms are analogous to the three terms appearing in the expression for \( \partial^2_T l_T(\theta) \) in (Clinet and Yoshida 2017, middle p. 1809) and are listed in the same order.

The third term converges in probability to zero uniformly on a ball, \( V_T \) centered on \( \theta^* \) shrinking to \( \{\theta^*\} \) using similar arguments to those in (Clinet and Yoshida 2017, p. 1810) for their third term and Lemma 4.

The second term also converges to a limit uniformly on a ball, \( V_T \) centered on \( \theta^* \) shrinking to \( \{\theta^*\} \) and uses ergodicity from Lemma 4 and similar arguments to Clinet and Yoshida (2017) but note that the limit is a matrix of zeros because its expectation is zero corresponding to the block diagonal structure of the full information matrix.

Finally consider the first, martingale term,

\[
M_T(\theta) = \frac{1}{T} \int_{[0,T]} \partial_\theta \{ \lambda(t; \theta)^{-1} U(t; \theta) \} \tilde{N}(dt)
\]

which we will show converges to zero in probability uniformly in \( \theta \in \Theta \) (uniformity allow us to deal with the evaluation at \( \hat{\theta}_T \) and use \( \mathbb{E}[|M_{a,T}(\hat{\theta}_T)|^p] \leq \mathbb{E}[\sup_{\theta \in \Theta} |M_{a,T}(\theta)|^p] \) where \( M_{a,T} \) is the \( a \)’th component. For \( p = dim(\Theta) + 1 \)

\[
\mathbb{E}[\sup_{\theta \in \Theta} |M_{a,T}(\theta)|^p] \leq K(\Theta, p) \left\{ \int_{\Theta} d\theta \mathbb{E}[|M_T(\theta)|^p] + \int_{\Theta} d\theta \mathbb{E}[|\partial_\theta M_T(\theta)|^p] \right\}
\]

where \( K(\Theta, p) < \infty \) using Sobolev’s inequality as in (Clinet and Yoshida 2017, Proof of Lemma 3.10). We next apply the Davis-Burkholder-Gundy inequality followed by Jensen’s inequality to each of \( \mathbb{E}[|M_T(\theta)|^p] \) and \( \mathbb{E}[|\partial_\theta M_T(\theta)|^p] \).

First

\[
\mathbb{E}[|M_T(\theta)|^p] \leq C T^{-p} \mathbb{E} \left[ \int_{[0,T]} (\partial_\theta \{ \lambda(t; \theta)^{-1} U(t; \theta) \})^2 \lambda(t; \theta^*) dt \right]^{\frac{p}{2}}
\]

\[
\leq C T^{-p+\frac{p}{2}-1} \int_{[0,T]} \mathbb{E} \left[ |\partial_\theta \{ \lambda(t; \theta)^{-1} U(t; \theta) \}|^p \lambda(t; \theta^*)^{\frac{p}{2}} \right] dt
\]

\[
\leq C T^{-\frac{p}{2}} \sup_{t \in \mathbb{R}_+} \mathbb{E} \left[ \sup_{\theta \in \Theta} |\partial_\theta \{ \lambda(t; \theta)^{-1} U(t; \theta) \}|^p \lambda(t; \theta^*)^{\frac{p}{2}} \right]
\]

Similarly

\[
\mathbb{E}[|\partial_\theta M_T(\theta)|^p] \leq C T^{-p} \mathbb{E} \left[ \int_{[0,T]} (\partial^2_\theta \{ \lambda(t; \theta)^{-1} U(t; \theta) \})^2 \lambda(t; \theta^*) dt \right]^{\frac{p}{2}}
\]
\[
\leq CT^{-\frac{p}{2}} \sup_{t \in \mathbb{R}_+} \mathbb{E} \left[ \sup_{\theta \in \Theta} |\partial_\theta^2 \{\lambda(t; \theta)^{-1} U(t; \theta)\}|^p \lambda(t; \theta^*)^\frac{p}{2} \right]
\]

Now, as in Clinet and Yoshida (2017) proof of Lemma 3.12, the processes \(|\partial_\theta \{\lambda(t; \theta)^{-1} U(t; \theta)\}|^p \lambda(t; \theta^*)^\frac{p}{2}\) and \(|\partial_\theta^2 \{\lambda(t; \theta)^{-1} U(t; \theta)\}|^p \lambda(t; \theta^*)^\frac{p}{2}\) are dominated by polynomials in \(\lambda(t; \theta)^{-1}, \partial_\theta \lambda(t; \theta)\) and \(\partial_\theta^2 U(t; \theta)\) for \(i \leq 0, 1, 2\). The first two terms are covered by Clinet and Yoshida (2017) condition A2. The terms \(\partial_\theta^2 U(t; \theta)\) are covered by Condition 3 (and are shown to be true for the exponential decay model in Lemma 2).

\[\square\]

**Lemma 7** Under \(H_0\), the estimated information matrix, \(\hat{I}_\psi\) defined in (17) satisfies

\[
\frac{1}{T} \hat{I}_\psi \rightarrow_P \Omega.
\]

**Proof** Recall from (17)

\[
\hat{I}_\psi = \int_{[0,T]} \lambda(t; \hat{\theta})^{-2}(\partial_\theta \lambda(t; \hat{\theta})) \otimes^2 N(dt)
\]

and let

\[
\hat{I}_\psi(v^*) = \int_{[0,T]} \lambda(t; v^*)^{-2}(\partial_\theta \lambda(t; v^*)) \otimes^2 N(dt).
\]

Note that, by similar arguments to that of the proof of Lemma 3.12 in Clinet and Yoshida (2017), we have

\[
T^{-1} \hat{I}_\psi(v^*) = T^{-1} \int_{[0,T]} \lambda(t; \theta^*)^{-1}(\partial_\theta \lambda(t; \theta^*)) \otimes^2 N(dt) + M_T,
\]

where \(M_T\) is a martingale of order \(O_P(T^{-1/2})\). By ergodicity, we thus have that \(T^{-1} \hat{I}_\psi(v^*) \rightarrow \Omega\) where \(\Omega\) is the same positive definite matrix as in Lemma 5. Hence to prove Lemma 7 it is sufficient to show that \(\frac{1}{T} c^T (\hat{I}_\psi - \hat{I}_\psi(v^*)) c \rightarrow 0\) for any \(c \in \mathbb{R}^T\). Let \(R(t; \theta, \phi) = \lambda(t; \theta)^{-1} U(t; \theta, \phi)\). Then

\[
\frac{1}{T} c^T (\hat{I}_\psi - \hat{I}_\psi(v^*)) c = \frac{1}{T} \int_{[0,T]} \{R(t; \hat{\theta}_T, \hat{\phi}_T)^2 - R(t; \theta^*, \phi^*)^2\} N(dt)
\]

\[= \frac{1}{T} \int_{[0,T]} \{R(t; \hat{\theta}_T, \hat{\phi}_T)^2 - R(t; \hat{\theta}_T, \phi^*)^2\} N(dt)
\]

\[+ \frac{1}{T} \int_{[0,T]} \{R(t; \hat{\theta}_T, \phi^*)^2 - R(t; \theta^*, \phi^*)^2\} N(dt)
\]

Now, using a Taylor series expansion

\[
\frac{1}{T} \int_{[0,T]} \{R(t; \hat{\theta}_T, \hat{\phi}_T)^2 - R(t; \theta^*, \phi^*)^2\} N(dt)
\]

\[= 2c^T (\hat{\mu}_H - \mu_H) \frac{1}{T} \int_{[0,T]} \{\lambda(t; \hat{\theta}_T)^{-1} \int_{[0,t]} \hat{\phi}_T w(t-s; \hat{\alpha}_T) N(ds)\} R(t; \hat{\theta}_T, \phi^*) N(dt)
\]

\[+ \{c^T (\hat{\mu}_H - \mu_H) \frac{2}{T} \int_{[0,T]} \{\lambda(t; \hat{\theta}_T)^{-1} \int_{[0,t]} \hat{\phi}_T w(t-s; \hat{\alpha}_T) N(ds)\}^2 N(dt).
\]

Now \(c^T (\hat{\mu}_H - \mu_H) \rightarrow P 0\) and, similarly to Clinet and Yoshida (2017), both integrals are uniformly bounded in probability for all \(T\) hence \(\frac{1}{T} c^T (\hat{I}_\psi - \hat{I}_\psi(v^*)) c\) converges to zero in probability, completing the proof.

\[\square\]
D proof of theorem 2

We have divided the proof of Theorem 2 into a series of Lemmas. Before we derive the asymptotic distribution of the score statistic we need some definitions. For the sake of simplicity, we will use the notation \( n \). By Proposition 1, we may assume the existence of an unboosted marked Hawkes process \( \tilde{N}_g(0) \) generated by the same measure \( \tilde{N}_g \) on \( \mathbb{R}^2 \times X \) as the sequence of processes \( N^T \). \( \tilde{N}_g(0) \) is thus a marked Hawkes process with boost function \( g(\cdot, \phi^*, 0) = 1 \), and corresponds to the process \( N_g \) studied in the asymptotic theory under the null hypothesis. Hereafter, we use the notation \( \theta \) to emphasize this fact. We call \( \lambda(0) \) its associated stochastic intensity, that is for any \( \theta \in \Theta \)

\[
\lambda^{(0)}(t; \theta) = \eta + \theta \int_{( - \infty, t) \times X} w(t - s; \alpha)g(x; \phi^*, 0)N_g^{(0)}(ds \times dx)
\]

\[
= \eta + \theta \int_{( - \infty, t) \times X} w(t - s; \alpha)N^{(0)}(ds)
\]

\( N^{(0)} = N^{(0)}_g(\cdot \times X) \) where \( \lambda^{(0)}(t; \theta^*) \) is the actual stochastic intensity of \( N^{(0)} \), that is \( \int_0^T \lambda^{(0)}(s; \theta^*)ds \) is the predictable compensator of \( N_t^{(0)} \). Finally, we define for \( i \in \{0, 1\} \), \( \theta \in \Theta, \phi \in \Phi \),

\[
\lambda^{(0), i}(t; \theta, \phi) = \theta \int_{(0, t) \times X} w(t - s; \alpha)\partial^i_{\phi}g(x; \phi, 0)N^{(0)}_g(ds \times dx).
\]

We first show that in the sense of (34) and (35) below, \( N^T \) is asymptotically close to \( N^{(0)}_g \) when \( T \to + \infty \).

Lemma 8 Let \( f \) be a predictable process depending on \( \theta \in \Theta \) such that

\[
\sup_{t \in [0, T]} \mathbb{E} \sup_{\theta \in \Theta} |f(t, \theta)|^p < +\infty
\]

for some \( p \geq 2 \). Then we have

\[
\mathbb{E} \sup_{\theta \in \Theta} \left| \int_{[0, T) \times X} f(t, \theta)(N^T_g - N^{(0)}_g)(dt \times dx) \right| = O(T^{1/2})
\]

(34)

and for any \( i \in \{0, 1\} \)

\[
\sup_{t \in \mathbb{R}^+} \mathbb{E} \sup_{\theta \in \Theta} \left| \partial^i_{\phi} \lambda^{(0)}_g(t; \theta, \phi^*, 0) - \lambda^{(0), i}(t; \theta, \phi^*) \right|^2 = O(T^{-1/2}).
\]

(35)

Proof We prove our claim in three steps.

Step 1. Letting \( \delta^T(t) = \mathbb{E}[\lambda^T_g(t; \theta^*, \phi^*, \psi^+_T) - \lambda^{(0)}(t; \theta^*)] \), we prove \( \sup_{t \in [0, T]} \delta^T(t) = O(T^{-1/2}) \). We have

\[
\delta^T(t) \leq \lambda^* \mathbb{E} \int_{[0, T) \times X} w(t - s; \alpha^*)g(x; \phi^*, \psi^+_T)|N^T_g - N^{(0)}_g|(ds \times dx)
\]

\[
+ \eta^* \mathbb{E} \int_{[0, T) \times X} w(t - s; \alpha^*)g(x; \phi^*, \psi^+_T) - 1|N^{(0)}_g|(ds \times dx)
\]

\[
\leq \lambda^* \mathbb{E} \int_{[0, T)} w(t - s; \alpha^*)[g(x; \phi^*, \psi^+_T)]\mathcal{F}^X_s\lambda^T_g(s; \theta^*, \phi^*, \psi^+_T) - \lambda^{(0)}(s; \theta^*)|ds
\]

\[
\leq \lambda^* \mathbb{E} \int_{[0, T)} w(t - s; \alpha^*)[g(x; \phi^*, \psi^+_T)]\mathcal{F}^X_s\lambda^T_g(s; \theta^*, \phi^*, \psi^+_T) - \lambda^{(0)}(s; \theta^*)|ds
\]
\[
\int_{[0, t]} w(t-s; \alpha^*) \sup_{\psi \in [0, \psi_T^+]} \partial_\psi g(x; \phi^*, \psi)|N_g^0(ds \times dx)
\]
\[
\leq \partial_\psi \mathbb{E} \int_{[0, t]} w(t-s; \alpha^*) g(y_s; \phi^*, \psi_T^+)|\mathcal{F}_{s-}^Y| \lambda^T_g(s; \theta^*, \phi^*, \psi_T^+)-\lambda^{(0)}(s; \theta^*)|ds
\]
\[
+ T^{-1/2} K
\]
\[
\leq C \partial_\psi \int_{[0, t]} w(t-s; \alpha^*) \delta^T(s)ds + T^{-1/2} K
\]
\[
\leq C \partial_\psi \sup_{s \in [0, T]} \delta^T(s) + T^{-1/2} K,
\]
for some constant \( K > 0 \), where we have used that \( \mathbb{E}[g(y_s; \phi^*, \psi_T^+)|\mathcal{F}_{s-}^Y] \leq C < 1/\partial_\psi \), that \( \int^\infty_0 w(\cdot; \alpha^*) = 1 \), and Condition 5. Moreover, for a vector \( x \), we have used the notation \( |x| = \sum_i |x_i| \). Taking the supremum over \([0, T]\) on the left hand side, we deduce \( \sup_{s \in [0, T]} \delta^T(s) \leq KT^{-1/2}/(1 - C \partial_\psi) \) and we are done.

**Step 2.** Letting \( \epsilon^T(t) = \mathbb{E}[\lambda^T_g(t; \theta^*, \phi^*, \psi_T^+)-\lambda^{(0)}(t; \theta^*)^2] \), we prove \( \sup_{t \in [0, T]} \epsilon^T(t) = O(T^{-1/2}) \). We have for some \( c > 0 \) arbitrary small,

\[
\epsilon^T(t) \leq (1 + c) \partial_\psi \mathbb{E} \left| \int_{[0, t]} w(t-s; \alpha^*) g(x; \phi^*, \psi_T^+)(N^T_g-N^0_g)(ds \times dx) \right|^2
\]
\[
+(1 + c^{-1}) \partial_\psi \mathbb{E} \left| \int_{[0, t]} w(t-s; \alpha^*) g(x; \phi^*, \psi_T^+)|N^0_g(ds \times dx) \right|^2
\]
\[
= I + II,
\]
where we have used the inequality \((x+y)^2 \leq (1+c)x^2 + (1+c^{-1})y^2 \) for any \( c > 0 \). First, we have

\[
I \leq (1 + c)(1 + c^{-1}) \partial_\psi \mathbb{E} \left| \int_{[0, t]} w(t-s; \alpha^*) g(x; \phi^*, \psi_T^+)(\tilde{N}^T_g - \tilde{N}^0_g)(ds \times dx) \right|^2
\]
\[
+(1 + c^2) \partial_\psi \mathbb{E} \left| \int_{[0, t]} w(t-s; \alpha^*) g(x; \phi^*, \psi_T^+)|\mathcal{F}_{s-}^Y|\lambda^T_g(s; \theta^*, \phi^*, \psi_T^+)-\lambda^{(0)}(s; \theta^*)|ds \right|^2
\]
\[
= I_A + I_B.
\]

Now, applying Jensen’s inequality with respect to the probability measure \( w(s; \alpha^*)ds / \int^T_0 w(s; \alpha^*)ds \), and then using \( \int^\infty_0 w(\cdot; \alpha^*) = 1 \), and \( \mathbb{E}[g(y_s; \phi^*, \psi_T^+)|\mathcal{F}_{s-}^Y] \leq C < 1/\partial_\psi \) yields

\[
I_B \leq (1 + c)^2 \partial_\psi \mathbb{E} \left| \int_{[0, t]} w(t-s; \alpha^*) \epsilon^T(s)ds \right|^2
\]
\[
\leq (1 + c^2) \partial_\psi \mathbb{E} \right| \sup_{s \in [0, T]} \epsilon^T(s).
\]

Now, for \( I_A \), we have

\[
I_A \leq K \mathbb{E} \left| \int_{[0, t]} w(t-s; \alpha^*) (g(x; \phi^*, \psi_T^+)-1)(\tilde{N}^T_g - \tilde{N}^0_g)(ds \times dx) \right|^2
\]
\[
+ K \mathbb{E} \left| \int_{[0, t]} w(t-s; \alpha^*) (\tilde{N}^T_g - \tilde{N}^0_g)(ds \times dx) \right|^2
\]
\[
\leq K \mathbb{E} \left| \int_{[0, t]} w(t-s; \alpha^*)^2 (g(y_s; \phi^*, \psi_T^+)-1)^2 |\mathcal{F}_{s-}^Y| \right|
\]
\[
|\tilde{\lambda}^T_g(s; \theta^*, \phi^*, \psi^*_T) - \lambda(0)(s; \theta^*)|ds \\
+ K\mathbb{E} \int_{[0,t]} w(t-s; \alpha^*)^2 |\tilde{\lambda}^T_g(s; \theta^*, \phi^*, \psi^*_T) - \lambda(0)(s; \theta^*)|ds \\
\leq K \int_{[0,t]} w(t-s; \alpha^*)^2 |\tilde{\lambda}^T_g(s; \theta^*, \phi^*, \psi^*_T) - \lambda(0)(s; \theta^*)|ds \\
+ K \sup_{s \in [0,T]} \delta^T(s) \\
\leq K \sup_{s \in [0,T]} \left( \epsilon^T(s) \lor 1 \right) + K \sup_{s \in [0,T]} \delta^T(s),
\]

where we have used Cauchy-Schwarz inequality along with (23) and (24). Moreover, following a similar path as for Step 1, we also have that \( II \leq KT^{-1/2} \) by (23). Thus, overall, using that \( \sup_{s \in [0,T]} \delta^T(s) \leq KT^{-1/2} \) by Step 1, we obtain for some constant \( K > 0 \)
\[
\epsilon^T(t) \leq K(T^{-1} + T^{-1/2}) + ((1+c)^2\beta^2\sigma^2 + KT^{-1}) \sup_{s \in [0,T]} \epsilon^T(s),
\]
and taking the supremum over \([0,T]\) on the left hand side, taking \( c > 0 \) close enough to 0 and \( T \) large enough so that \(((1+c)^2\beta^2\sigma^2 + KT^{-1}) < A\) for some constant \( A < 1 \), we get
\[
\sup_{s \in [0,T]} \epsilon^T(s) \leq K(T^{-1} + T^{-1/2})/(1-A) \leq \tilde{K}T^{-1/2}
\]
for some \( \tilde{K} > 0 \).

**Step 3.** We prove (34) and (35). For (34), this is a direct consequence of the fact that the compensator of \( N^T_g - N^0_g \) is \( \int_0^T |\tilde{\lambda}^T_g(t; \theta^*, \phi^*, \psi^*_T) - \lambda(0)(t; \theta^*)|dt \), Cauchy-Schwarz inequality and the uniform condition on \( f \). For (35), let \( i \in \{0, 1\} \). We have
\[
\mathbb{E} \sup_{\theta \in \Theta} \left| \frac{\partial^i \tilde{\lambda}^T_g}{\partial \phi^i} (t; \theta, \phi^*, 0) - \lambda(0).i(t; \theta, \phi^*) \right|^2 \\
\leq K \mathbb{E} \sup_{\alpha \in \mathcal{A}} \left| \int_{[0,t] \times \mathbb{X}} w(t-s; \alpha) \frac{\partial^i \tilde{\lambda}^T_g}{\partial \phi^i} (\mathbf{x}; \phi^*, 0)(N^T_g - N^0_g)(ds \times d\mathbf{x}) \right|^2 \\
\leq K \mathbb{E} \left| \int_{[0,t] \times \mathbb{X}} \tilde{w}(t-s) |\frac{\partial^i \tilde{\lambda}^T_g}{\partial \phi^i} (\mathbf{x}; \phi^*, 0)||N^T_g - N^0_g|(ds \times d\mathbf{x}) \right|^2
\]
And from here, using Burkholder-Davis-Gundy inequality, Step 2 of this proof along with conditions (23) and (24) we deduce that the above term is dominated by \( KT^{-1/2} \) for some \( K > 0 \) uniformly in \( t \in \mathbb{R}_+ \).

**Lemma 9** (Consistency of \( \hat{\nu}_T \) under the local alternatives) Under \( H^T_1 \), we have
\[
\hat{\nu}_T \to^P \nu^* := (\theta^*, \phi^*, 0).
\]

**Proof** The convergence of the third component is obvious, and the convergence of the second one is assumed. All we have to show is that \( \hat{\theta}_T \to^P \theta^* \). Let \( I^{(0)}_T(\theta) = \int_{[0,T] \times \mathbb{X}} \log \tilde{\lambda}^0(t; \theta)N^0_g(dt \times d\mathbf{x}) - \int_0^T \lambda(0)(t; \theta)dt \), where we recall that \( \lambda(0) \) is the stochastic intensity of \( N^0_g \). It suffices to show that uniformly in \( \theta \in \Theta \), we have the convergence \( T^{-1}(I_T(\theta) - I^{(0)}_T(\theta)) \to^P 0 \). But note that \( T^{-1}(I_T(\theta) - I^{(0)}_T(\theta)) = I + II \) with
\[
I = -T^{-1} \int_{[0,T] \times \mathbb{X}} \left\{ \log \tilde{\lambda}(0)(t; \theta)N^0_g(dt \times d\mathbf{x}) - \log \lambda^T(t; \theta)N^T_g(dt \times d\mathbf{x}) \right\}
\]
and
\[
II = -T^{-1} \int_0^T \{\lambda^{(0)}(t; \theta) - \lambda^T(t; \theta)\} dt.
\]

By (35), we immediately have that \(\mathbb{E} \sup_{\theta \in \Theta} |\{\lambda^{(0)}(t; \theta) - \lambda^T(t; \theta)\}| = O(T^{-1/2})\) uniformly in \(t \in [0, T]\), so that \(II \rightarrow P_0\) uniformly in \(\theta \in \Theta\). Writing \(I\) as the sum
\[
T^{-1} \int_{[0,T] \times \mathbb{X}} \left\{ \log \lambda^{(0)}(t; \theta) - \log \lambda^T(t; \theta) \right\} N_\theta(0) (dt \times dx) \\
+ T^{-1} \int_{[0,T] \times \mathbb{X}} \log \lambda^T(t; \theta) \left\{ N^T_g(dt \times dx) - N^T_g(dt \times dx) \right\}
= A + B,
\]
we need to show that both terms tend to 0. Since \(|\log \lambda^{(0)}(t; \theta) - \log \lambda^T(t; \theta)| \leq \eta^{-1}|\lambda^{(0)}(t; \theta) - \lambda^T(t; \theta)|\), we easily get by Cauchy-Schwarz inequality and (35) that \(\mathbb{E} \sup_{\theta \in \Theta} |A| \rightarrow 0\). Moreover, using \(\log \lambda^T(t; \theta) \leq \lambda^T(t; \theta) - 1\), by (34) and Condition 5 we have that \(\mathbb{E} \sup_{\theta \in \Theta} |B| \rightarrow 0\) and we are done. \(\square\)

**Lemma 10** Under \(H_1^T\), we have
\[
T^{-1/2} \partial \psi l_\theta(\hat{\nu}_T) \rightarrow^d N(\Omega \gamma^*, \Omega)
\]

**Proof** First, note that by application of Lemma 8, Lemma 9, and following the same path as for the proof of Lemma 6, we deduce
\[
T^{-1/2} \partial \psi l_\theta(\hat{\nu}_T) - T^{-1/2} \partial \psi l_\theta(\nu^*) \rightarrow^P 0.
\]

Next, we have
\[
T^{-1/2} \partial \psi l_\theta(\theta^*, \phi^*, 0) = T^{-1/2} \int_{(0,T)} \partial \psi \log \lambda^T_g(t; \theta^*, \phi^*, 0) \tilde{N}^T_g(dt) \\
+ T^{-1/2} \int_{(0,T)} \partial \psi \lambda^T_g(t; \theta^*, \phi^*, 0) \left( \frac{\lambda^T_g(t; \theta^*, \phi^*, \psi^*_T)}{\lambda^T(t; \theta^*)} - 1 \right) dt
= I + II,
\]

where we have used the notation \(\tilde{N}^T_g(dt) = N^T_g(dt, \mathbb{X}) - \lambda^T_g(t; \theta^*, \phi^*, \psi^*_T)dt\). We derive the limit of the first term following the same path as for the proof of Lemma 5. Letting \(S^T_u = T^{-1/2} \int_{(0,uT)} \partial \psi \log \lambda^T_g(t; \theta^*, \phi^*, 0) \tilde{N}^T_g(dt)\), we directly have that
\[
\langle S^T, S^T \rangle_u = T^{-1} \int_0^u \frac{\partial \psi \lambda^T_g(t; \theta^*, \phi^*, 0) \partial \psi \lambda^T_g(t; \theta^*, \phi^*, 0)^T}{\lambda^T(t; \theta^*)^2} \lambda^T_g(t; \theta^*, \phi^*, \psi^*_T) dt.
\]

By (35), the boundedness of moments of \(\lambda^T_g\) and its derivatives and Hölder’s inequality we easily deduce that
\[
\langle S^T, S^T \rangle_u = T^{-1} \int_0^u \frac{\lambda^{(0),1}(t; \theta^*, \phi^*) \partial \psi \lambda^{(0),1}(t; \theta^*, \phi^*)^T}{\lambda^{(0)}(t; \theta^*)} dt + o_P(1),
\]
which converges in probability to \(u \Omega\) by Lemma 5. Similarly, Lindeberg’s condition
\[
\mathbb{E} \sum_{t \leq u} (\Delta S^T_s)^2 1_{|\Delta S^T_s| > a} \rightarrow 0
\]
for any $a > 0$ is satisfied, so that by 3.24, Chapter VIII in Jacod and Shiryaev (2013), we get that $I = S_T^1 \rightarrow^d N(0, \Omega)$. Now we derive the limit for $II$. We have for some $\tilde{\gamma}_T \in [0, \tilde{\gamma}_T]$

$$II = T^{-1} \int_{(0,T)} \lambda^T(t; \theta^*, \phi^*)^{-1} \partial_\psi \lambda^T_g(t; \theta^*, \phi^*, 0) \partial_\psi \lambda^T_g(t; \theta^*, \phi^*, \tilde{\gamma}_T)^T \gamma^* dt.$$ 

Now, using Hölder’s inequality, the uniform boundedness of moments of $\lambda^T_g$ in $\nu$, and (35), we deduce as previously that

$$II = T^{-1} \int_{(0,T)} \lambda^{(0)}(t; \theta^*)^{-1} \lambda^{(0),1}(t; \theta^*, \phi^*) \lambda^{(0),1}(t; \theta^*, \phi^*)^T \gamma^* dt + o_P(1),$$

which, by the proof of Lemma 7, tends in probability to the limit $\Omega \gamma^*$. By Slutsky’s Lemma, we get the desired convergence in distribution for $T^{-1/2} \partial_\psi I^*_g(\tilde{\nu}_T)$. 

Lemma 11 Under $H^T_1$, we have

$$T^{-1} \hat{I}_\psi \rightarrow^P \Omega.$$ 

Proof First, as for Lemma 10, note that by application of Lemma 8, Lemma 9, and following the same path as for the proof of Lemma 7, we have

$$T^{-1} (\hat{I}_\psi - \hat{I}_\psi(\nu^*)) \rightarrow^P 0.$$ 

Now recall that

$$T^{-1} \hat{I}_\psi(\nu^*) = T^{-1} \int_{[0,T] \times \mathbb{X}} \lambda^T(t; \theta^*)^{-2} (\partial_\psi \lambda^T_g(t; \nu^*)) \otimes^2 N^T_g (dt \times d\mathbf{x}).$$

By (34), (35), the boundedness of moments of $\lambda^T_g$ and its derivatives and Hölder’s inequality we get

$$T^{-1} \hat{I}_\psi(\nu^*) = T^{-1} \int_{[0,T] \times \mathbb{X}} \lambda^{(0)}(t; \theta^*)^{-2} (\lambda^{(0),1}(t; \nu^*)) \otimes^2 N^{(0)}_g (dt \times d\mathbf{x}) + o_P(1),$$

and by Lemma 7, the right-hand side converges in probability to $\Omega$. 

References

Andersen P, Borgan O, Gill R, Keiding N (1996) Statistical Models Based on Counting Processes. Springer Series in Statistics, Springer, New York

Bacry E, Mastromatteo I, Muzy J-F (2015) Hawkes processes in finance. Market Microstruct Liq 1(01):1550005

Brémaud P, Massoulié L (1996) Stability of nonlinear hawkes processes. Ann Probab 24(3):1563–1588

Breusch TS, Pagan A (1980) The lagrange multiplier test and its applications to model specification in econometrics. Rev Econ Stud 47(1):239–253

Chen F, Hall P (2013) Inference for a nonstationary self-exciting point process with an application in ultra-high frequency financial data modeling. J Appl Probab 50(4):1006–1024

Clinet S (2020) Quasi-likelihood analysis for marked point processes and application to marked hawkes processes. arXiv preprint arXiv:2001.11624

Clinet S, Potiron Y (2018) Statistical inference for the doubly stochastic self-exciting process. Bernoulli 24(4B):3469–3493

Clinet S, Yoshida N (2017) Statistical inference for ergodic point processes and application to limit order book. Stoch Process Appl 127(6):1800–1839

Crane R, Sornette D (2008) Robust dynamic classes revealed by measuring the response function of a social system. Proc Nat Academy Sci United States Am 105(41):15649–15653

 Springer
Daley D, Vere-Jones D (2002) An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods. Probability and Its Applications. Springer
Duarte A, Löcherbach E, Ost G (2016) Stability, convergence to equilibrium and simulation of non-linear hawkes processes with memory kernels given by the sum of erlang kernels. arXiv preprint arXiv:1610.03300
Embrechts P, Liniger T, Lin L (2011) Multivariate hawkes processes: an application to financial data. J Appl Probab 48A:367–378
Hawkes A (1971) Point spectra of some mutually exciting point processes. J Royal Statist Soc Series B 33(3):438–443
Hawkes AG (2018) Hawkes processes and their applications to finance: a review. Quantitat Finance 18(2):193–198
Hawkes AG, Oakes D (1974) A cluster process representation of a self-exciting process. J Appl Probab 1:493–503
Jacod J, Shiryaev A (2013) Limit theorems for stochastic processes, Volume 288. Springer Science & Business Media
Kallenberg O (2006) Foundations of Modern Probability. Probability and Its Applications, Springer, New York
Liniger TJ (2009) Multivariate hawkes processes. Ph. D. thesis, ETH Zurich
Ogata Y (1978) The asymptotic behaviour of maximum likelihood estimators for stationary point processes. Annals Institut Statist Math 30(1):243–261
Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. J Am Statist Assoc 83(401):9–27
Ozaki T (1979) Maximum likelihood estimation of hawkes’ self-exciting point processes. Annals Inst Statist Math 31(1):145–155
Rao C (2009) Linear Statistical Inference and its Applications. Wiley Series in Probability and Statistics. Wiley
Richards K-A, Dunsmuir W, Peters G (2019) Score test for marks in hawkes processes. Available SSRN. https://doi.org/10.2139/ssrn.3381976
Richards, Kylie-Anne, M. . S. F. o. S. U. (2019) Modelling the dynamics of the limit order book in financial markets

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.