Spectral analysis for noisy Hawkes processes inference

Anna Bonnet, Felix Cheysson, Miguel Martinez Herrera, Maxime Sangnier

Spectral analysis for noisy Hawkes processes inference

Anna Bonnet*, Félix Cheysson†, Miguel Martinez Herrera*, Maxime Sangnier*

May 20, 2024

Abstract

Classic estimation methods for Hawkes processes rely on the assumption that observed event times are indeed a realisation of a Hawkes process, without considering any potential perturbation of the model. However, in practice, observations are often altered by some noise, the form of which depends on the context. It is then required to model the alteration mechanism in order to infer accurately such a noisy Hawkes process. While several models exist, we consider, in this work, the observations to be the indistinguishable union of event times coming from a Hawkes process and from an independent Poisson process. Since standard inference methods (such as maximum likelihood or Expectation-Maximisation) are either unworkable or numerically prohibitive in this context, we propose an estimation procedure based on the spectral analysis of second order properties of the noisy Hawkes process. Novel results include sufficient conditions for identifiability of the ensuing statistical model with exponential interaction functions for both univariate and bivariate processes. Although we mainly focus on the exponential scenario, other types of kernels are investigated and discussed. A new estimator based on maximising the spectral log-likelihood is then described, and its behaviour is numerically illustrated on synthetic data. Besides being free from knowing the source of each observed time (Hawkes or Poisson process), the proposed estimator is shown to perform accurately in estimating both processes.

Keywords: Hawkes process, point process, spectral analysis, parametric estimation, superposition, identifiability.

1 Introduction

Hawkes processes, introduced in Hawkes (1971), are a class of point processes that have been originally used to model self-exciting phenomena and more recently other types of past-dependent behaviours. Their fields of applications are wide and include for instance seismology (Ogata, 1988, 1998), neuroscience (Chornoboy et al., 1988; Lambert et al., 2018), criminology (Olinde and Short, 2020), finance (Embrechts et al., 2011; Bacry et al., 2015) and biology (Gupta et al., 2018), to mention a few. Consequently, there has been a deep focus on estimation techniques for Hawkes processes. Among them, let us mention maximum likelihood approaches (Ogata, 1978; Ozaki, 1979; Guo et al., 2018), methods of moments (Da Fonseca and Zaatour, 2013), least-squares contrast minimisation (Reynaud-Bouret et al., 2014; Bacry et al., 2020), Expectation-Maximisation (EM) procedures (Lewis and Mohler, 2011), and methods using approximations through autoregressive models (Kirchner, 2017).

All of these methods are based on the assumption that the history of the process has been accurately observed, although this information is partial or noised in many contexts, in particular due to measurement or detection errors. Different models, described notably in Lund and Rudemo (2000), have been proposed to handle such errors for spatial point processes, with inference methods based on approximating the likelihood depending on each noise scenario. The first scenario, called *displacement*, is when the event times are observed with a shift. If deconvolution methods to recover the unnoised process are standard approaches for simpler point processes such as Poisson

*Sorbonne Université and Université Paris Cité, CNRS, Laboratoire de Probabilités, Statistique et Modélisation, F-75005 Paris, France

†Gustave Eiffel University, CNRS, Laboratoire d'Analyse et de Mathématiques Appliquées, 77420 Champs-sur-Marne, France

processes (see for instance Antoniadis and Bigot (2006); Bonnet et al. (2022) or the review by Hohage and Werner (2016)), the literature for Hawkes processes remains scarce and consists of the work of Trouleau et al. (2019) where the event times are observed with a delay and the work of Deutsch and Ross (2020) where the shift follows a Gaussian distribution. The latter also explores the framework where some event times are undetected, which is similar to that studied in Mei et al. (2019). This setting can either be referred to as *thinning* when the observations are randomly missing, or *censoring* when complete regions are unobserved. The last scenario, called *Superposition of ghost points* by Lund and Rudemo (2000), is the focus of this paper and describes situations where additional points are coming from an external point process, in our case a Poisson process. A real-world application that motivated this work comes from spike trains analysis in neuroscience: the membrane potential, which is a continuous signal, is recorded and when it exceeds a certain threshold, one considers that an event time, called a spike, occurred. However, since the original signal is noised, it is possible to detect spikes that do not correspond to real events. Regarding inference of such a noisy process, let us highlight that exact likelihood approaches are intractable due to the unknown origin of each occurrence (Hawkes process or noise process) while methods based on inferring this missing information, for instance Expectation-Maximisation algorithms, are computationally too demanding.

Inspired by the work of Cheysson and Lang (2022) when the event times of a Hawkes process are aggregated, we turn to spectral analysis (Section 2.1) to propose a novel estimation procedure that allows to infer the noisy Hawkes process. On the way, we present a general characterisation of the Bartlett spectrum of the superposition of two independent processes (Section 2.2) and identifiability results of the statistical model in the univariate (Section 3) and the bivariate (Section 4) settings. Inference in these two settings is numerically illustrated in Section 5. Before starting, we present the mathematical setting (Section 1.1) and review the references related to spectral approaches (Section 1.2).

1.1 Mathematical setting

Let $H = (H_1, \dots, H_d)$ be a stationary multivariate Hawkes process on $\mathbb{R}$ defined by its conditional intensity functions λ_i^H ($i \in \{1, \dots, d\}$): for all $t \in \mathbb{R}$,

$$\lambda_i^H(t) = \mu_i + \sum_{j=1}^{d} \int_{-\infty}^{t} h_{ij}(t-s)\, H_j(\mathrm{d}s) = \mu_i + \sum_{j=1}^{d} \sum_{T_k^{H_j} \le t} h_{ij}(t - T_k^{H_j})\,, \tag{1}$$

where $\mu_i > 0$ is the baseline intensity of process H_i, $h_{ij} : \mathbb{R} \to \mathbb{R}_{\ge 0}$ is the interaction or kernel function describing the effect of process H_j on process H_i, and $(T_k^{H_j})_{k \ge 1}$ denotes the event times of H_j.

By defining the matrix $S = (\|h_{ij}\|_1)_{1 \le i,j \le d}$, where

$$\|h_{ij}\|_1 = \int_{-\infty}^{+\infty} h_{ij}(t)\, \mathrm{d}t\,,$$

the stationarity condition of H reduces to controlling the spectral radius of S: $\rho(S) < 1$ (Brémaud and Massoulié, 1996).

The goal of this paper is to study a noisy version of the Hawkes process where the sequences of event times $(T_k^{H_1})_{k \ge 1}$, …, $(T_k^{H_d})_{k \ge 1}$ of H are contaminated by the event times from another process P. Since the latter process is aimed at modeling an external noise mechanism due to errors in the detection of event times, it is naturally assumed that each subprocess of H is perturbed uniformly with the same level of noise, such that P is chosen to be a multivariate Poisson process with same intensity across subprocesses. Formally, let $P = (P_1, \dots, P_d)$ be a multivariate homogeneous Poisson process on $\mathbb{R}$, supposed to be independent from H, where each univariate process P_i has the same intensity $\lambda_0 > 0$. We note the event times $(T_k^{P_i})_{k \ge 1}$ ($i \in \{1, \dots, d\}$). We consider then the point process $N = (N_1, \dots, N_d)$ defined as the superposition of H and P (Definition 2.1): the sequence of event times $(T_k^{N_i})_{k \ge 1}$ of N_i ($i \in \{1, \dots, d\}$) is the ordered union of $(T_k^{H_i})_{k \ge 1}$ and $(T_k^{P_i})_{k \ge 1}$.

Throughout this paper we will refer to N as the noisy Hawkes process, and it will be assumed that event times of N are observed without knowing their origin (Hawkes or Poisson process). Our goal is to estimate both processes (i.e. the baselines μ_i, the kernels h_{ij} and the shared Poisson intensity λ_0) from the sole observation of $(T_k^{N_i})_{k\geq 1}$, $i \in \{1, \dots, d\}$.

Inference procedures for point processes often leverage the intensity functions in order to devise maximum likelihood and method of moment estimators (Ogata, 1988; Ozaki, 1979; Da Fonseca and Zaatour, 2013). Here, the process of interest N being a superposition of two independent point processes, the intensity of each subprocess N_i reads (for any integer $i \in \{1, \dots, d\}$): for all $t \in \mathbb{R}$,

$$\lambda_i^N(t) = \lambda_0 + \mu_i + \sum_{j=1}^{d} \int_{-\infty}^{t} h_{ij}(t-s)\, H_j(\mathrm{d}s)\,. \tag{2}$$

However, it appears from Equation (2) that usual estimators cannot be designed from intensity functions $\lambda_1^N, \dots, \lambda_d^N$ since they are based on H, which is indistinguishable from P in our setting.

In order to estimate the Hawkes and the Poisson components of N, we propose to leverage the spectral analysis of point processes, recently advocated by Cheysson and Lang (2022) for inference of aggregated Hawkes processes. It consists in considering, for a multivariate point process, its matrix-valued spectral density function, denoted $\mathbf{f} : \mathbb{R} \to \mathbb{C}^{d\times d}$, which is related to second-order measures (Bartlett, 1963). Given some observed times $(T_k^{N_1})_{k\geq 1}$, …, $(T_k^{N_d})_{k\geq 1}$, (in a prescribed time window $[0, T]$), the spectral density is linked to cross-periodograms, defined for all pairs $(i, j) \in \{1, \dots, d\}^2$ and all $\omega \in \mathbb{R}$ by:

$$I_{ij}^T(\omega) = \frac{1}{T} \sum_{k=1}^{N_i(T)} \sum_{l=1}^{N_j(T)} \mathrm{e}^{-2\pi \mathrm{i}\omega(T_k^{N_i} - T_l^{N_j})}\,, \tag{3}$$

where $N_i(t) = N_i([0, t))$. Indeed, considering the matrix-valued function $\mathbf{I}^T : \omega \in \mathbb{R} \mapsto (I_{ij}^T(\omega))_{1\leq i,j\leq d}$, the aforementioned link is to be understood as $\mathbf{I}^T(\omega)$ (for all $\omega \in \mathbb{R}$) being asymptotically distributed according to a complex Wishart distribution with one degree of freedom and scale matrix $\mathbf{f}(\omega)$ (Tuan, 1981; Villani et al., 2022). In particular, this implies that $\mathbb{E}[\mathbf{I}^T(\omega)] = \mathbf{f}(\omega)$. Moreover, it is noteworthy that the periodogram $\mathbf{I}^T(\omega)$ can be computed regardless of knowing the source of the event times. This paves the way for estimation.

As it happens, in the scope of statistical inference, a parametric model for the matrix-valued spectral density function is considered:

$$\mathcal{P} = \left\{ \mathbf{f}_\theta^N : \mathbb{R} \to \mathbb{C}^{d\times d}, \theta = (\mu, \gamma, \lambda_0) \in \Theta \right\}\,,$$

where γ is a parameter that characterises the interaction functions. Then, for $\omega_k = k/T$ ($k \in \{1, 2, \dots\}$), it can be shown that $(\mathbf{I}^T(\omega_k))_{k\geq 1}$ are asymptotically independent, leading to the approximate spectral log-likelihood (Brillinger, 2012; Düker and Pipiras, 2019; Villani et al., 2022):

$$\ell_T(\theta) = -\frac{1}{T} \sum_{k=1}^{M} \left\{ \log\left(\det\left(\mathbf{f}_\theta^N(\omega_k)\right)\right) + \mathrm{tr}\left(\mathbf{f}_\theta^N(\omega_k)^{-1} \mathbf{I}^T(\omega_k)\right) \right\}, \tag{4}$$

where det and tr are respectively the determinant and trace of matrices. Then, the so-called Whittle (or spectral) estimator $\mathbf{f}_{\hat\theta}^N$ of $\mathbf{f}$ (Whittle, 1952) is obtained for $\hat\theta \in \Theta$ such that

$$\hat\theta \in \arg\max_{\theta\in\Theta}\ \ell_T(\theta)\,.$$

1.2 Related works

The spectral analysis of point processes was introduced in Bartlett (1963) and extended to 2-dimensional point processes in Bartlett (1964). Subsequent research works focusing on the theoretical properties of the Bartlett spectrum include Daley (1971); Daley and Vere-Jones (2003); Tuan (1981) for temporal settings and Mugglestone and Renshaw (1996, 2001); Rajala et al. (2023) for spatial contexts.

Despite this, practical applications remain scarcer in the literature. Adamopoulos (1976) studies earthquake arrivals through the analysis of Hawkes processes and Karavasilis and Rigas (2007) analyses the cross-correlation of bivariate point processes in the context of muscular stimulation.

In a recent contribution by Cheysson and Lang (2022), the authors employ the spectral analysis of point processes to infer an aggregated Hawkes process. More precisely, the observations are assumed to come from a standard Hawkes process but only the counts of events on fixed intervals are available. By leveraging the properties of the Bartlett spectrum, they propose an estimator obtained by means of maximisation of the spectral log-likelihood.

The main advantage of the spectral approach is its obvious ability to handle different kinds of partial observations. That is why we propose an inference procedure for noisy Hawkes processes based on Bartlett's spectral density.

Spectral results concerning the Hawkes model are built on the linearity of the intensity function along with its branching properties. The first spectral analysis of linear Hawkes processes appears in the original paper Hawkes (1971) and was then developed in Daley and Vere-Jones (2003) in both univariate and multivariate contexts. Explicit expressions of the spectral density of a Hawkes process are available as long as the Fourier transform of the kernel functions are known, which allows to work with a wide array of parametrisations.

In this paper, we mainly focus on studying identifiability of the statistical model with the classic exponential kernel (and also give insights about identifiability with other kernels). We subsequently derive a parametric inference procedure for estimating the parameters of a process composed of the superposition of a linear Hawkes process and a homogeneous Poisson process.

2 Spectral analysis

2.1 The Bartlett spectrum

In this section, we formally introduce the concept of matrix-valued spectral measure $\mathbf{\Gamma} : \mathbb{R} \to \mathbb{C}^{d\times d}$ for a multivariate stationary point process $N = (N_1, \dots, N_d)$. This is an extension of the Bartlett spectrum introduced by Bartlett (1963) for the analysis of univariate point processes. Let $\mathcal{S}$ be the space of real functions on $\mathbb{R}$ with rapid decay (Daley and Vere-Jones, 2003, Chapter 8.6.1):

$$\mathcal{S} = \left\{ f \in C^\infty, \forall k \in \{1, 2, \dots\}, \forall r \in \{1, 2, \dots\}, \sup_{x\in\mathbb{R}} \left| x^r f^{(k)}(x) \right| < \infty \right\},$$

where C^∞ is the set of smooth functions from $\mathbb{R}$ to $\mathbb{R}$ and $f^{(k)}$ is the k^{th} derivative of $f \in C^\infty$.

Then, the Bartlett spectrum of N is the matrix-valued function $\mathbf{\Gamma}^N : \omega \in \mathbb{R} \mapsto (\Gamma^N_{ij}(\omega))_{1\le i,j\le d} \in \mathbb{C}^{d\times d}$, such that for all $1 \le i, j \le d$, Γ^N_{ij} is a measure on $\mathbb{R}$ verifying (Daley and Vere-Jones, 2003, Equation 8.4.13):

$$\forall (\varphi, \psi) \in \mathcal{S} \times \mathcal{S} : \quad \operatorname{cov}\left(\int_{\mathbb{R}} \varphi(x)\, N_i(\mathrm{d}x), \int_{\mathbb{R}} \psi(x)\, N_j(\mathrm{d}x) \right) = \int_{\mathbb{R}} \tilde{\varphi}(\omega)\tilde{\psi}(-\omega)\, \Gamma^N_{ij}(\mathrm{d}\omega)\,,$$

where for all $f \in \mathcal{S}$, $\tilde{f} : \mathbb{R} \to \mathbb{C}$ denotes the Fourier transform of f:

$$\forall \omega \in \mathbb{R} : \quad \tilde{f}(\omega) = \int_{\mathbb{R}} f(x) \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \mathrm{d}x\,.$$

If, for all $1 \le i, j \le d$, the measure Γ^N_{ij} is absolutely continuous, we can define the matrix-valued spectral density function of N, denoted $\mathbf{f}^N : \mathbb{R} \to \mathbb{C}^{d\times d}$, such that for all $\omega \in \mathbb{R}$, $\mathbf{f}^N(\omega) = (f^N_{ij}(\omega))_{1\le i,j\le d}$ with $\Gamma^N_{ij}(\mathrm{d}\omega) = f^N_{ij}(\omega)\, \mathrm{d}\omega$.

From a practical point of view, the spectral density $\mathbf{f}^N$ can be derived from the reduced covariance densities. Let $\mathcal{B}^c_{\mathbb{R}}$ be the collection of all bounded Borel sets on $\mathbb{R}$ and $\ell_{\mathbb{R}} : \mathcal{B}^c_{\mathbb{R}} \to \mathbb{R}_{\ge 0}$ the Lebesgue measure on $\mathbb{R}$. For all $i \in \{1, \dots, d\}$, the first moment measure of process N_i is defined as $A \in \mathcal{B}^c_{\mathbb{R}} \mapsto \mathbb{E}[N_i(A)]$ and, by stationarity of the process, it comes:

$$\forall A \in \mathcal{B}^c_{\mathbb{R}} : \quad \mathbb{E}[N_i(A)] = m^N_i \ell_{\mathbb{R}}(A)\,,$$

where $m_i^N = \mathbb{E}[N_i([0,1))]$ is the mean intensity of process N_i. Then, for all $(i,j) \in \{1,\dots,d\}^2$ the second order moment measure $M_{ij}^N : \mathcal{B}_{\mathbb{R}}^c \times \mathcal{B}_{\mathbb{R}}^c \to \mathbb{R}_{\geq 0}$ is defined by (Daley and Vere-Jones, 2003, Section 5.4):

$$\forall (A,B) \in \mathcal{B}_{\mathbb{R}}^c \times \mathcal{B}_{\mathbb{R}}^c : \quad M_{ij}^N(A,B) = \mathbb{E}[N_i(A)N_j(B)] = \int_{A\times B} M_{ij}(\mathrm{d}x, \mathrm{d}y)\,.$$

Now, as the process N is stationary, M_{ij}^N can be decomposed in a product of $\ell_{\mathbb{R}}$ and a so-called reduced measure $\breve{M}_{ij}^N : \mathcal{B}_{\mathbb{R}}^c \to \mathbb{R}_{\geq 0}$, such that for any bounded measurable function $g : \mathbb{R}^2 \to \mathbb{R}$ with bounded support (Daley and Vere-Jones, 2003, Equation 8.1.1a):

$$\int_{\mathbb{R}^2} g(x,y)\, M_{ij}^N(\mathrm{d}x, \mathrm{d}y) = \int_{\mathbb{R}} \int_{\mathbb{R}} g(x, x+u)\, \ell_{\mathbb{R}}(\mathrm{d}x)\, \breve{M}_{ij}^N(\mathrm{d}u)\,, \tag{5}$$

which leads to the definition of the reduced covariance measure $\breve{C}_{ij}^N : \mathcal{B}_{\mathbb{R}}^c \to \mathbb{R}_{\geq 0}$:

$$\forall B \in \mathcal{B}_{\mathbb{R}}^c : \quad \breve{C}_{ij}^N(B) = \breve{M}_{ij}^N(B) - m_i^N m_j^N \ell_{\mathbb{R}}(B)\,. \tag{6}$$

Since the right-hand side is the difference of two positive, positive-definite measures (Daley and Vere-Jones, 2003, Section 8.6), we can define the Fourier transform of $\breve{C}_{ij}$ as the difference of their Fourier transforms (see for example (Pinsky, 2008, Equation 5.2.1) for the Fourier transform of a measure). The resulting quantity comes out to correspond exactly to the spectral density function f_{ij}^N:

$$\forall \omega \in \mathbb{R} : \quad f_{ij}^N(\omega) = \int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \breve{M}_{ij}^N(\mathrm{d}x) - m_i^N m_j^N \delta(\omega)\,, \tag{7}$$

where δ is the Dirac delta function.

2.2 Superposition of processes and noisy Hawkes process

The model we study considers the superposition of two point processes that we define as follows.

Definition 2.1 (Superposition of processes)**.** Let X and Y be two independent and stationary multivariate point processes with same dimension d. The superposition of X and Y, denoted $N = X + Y$, is the stationary multivariate point process defined for any integer $1 \leq i \leq d$ as:

$$\forall A \in \mathcal{B}_{\mathbb{R}}^c : \quad N_i(A) = X_i(A) + Y_i(A).$$

It comes from the definition that if X and Y have respectively event times $(T_k^{X_1})_{k\geq 1}$, …, $(T_k^{X_d})_{k\geq 1}$ and $(T_k^{Y_1})_{k\geq 1}$, …, $(T_k^{Y_d})_{k\geq 1}$, the sequences of event times of $N = X+Y$ are the ordered unions of $(T_k^{X_1})_{k\geq 1}$ and $(T_k^{Y_1})_{k\geq 1}$ up to $(T_k^{X_d})_{k\geq 1}$ and $(T_k^{Y_d})_{k\geq 1}$. In addition, Proposition 2.2 below states that the spectral density of N can be obtained easily from those of X and Y.

Proposition 2.2. *Let X and Y be two independent and stationary multivariate point processes with same dimension d, admitting respective spectral densities $\mathbf{f}^X$ and $\mathbf{f}^Y$.*

Then $N = X + Y$ admits a matrix-valued spectral density function $\mathbf{f}^N$ and

$$\mathbf{f}^N = \mathbf{f}^X + \mathbf{f}^Y\,. \tag{8}$$

Proof. The proof is given in Appendix A. □

We are now ready to define the noisy Hawkes model, which is the superposition of a Hawkes process and a homogeneous Poisson process.

Definition 2.3 (Noisy Hawkes process)**.** Let H be a multivariate Hawkes process and $P = (P_1, \dots P_d)$ be a multivariate homogenous Poisson process independent from H and with common intensity $\lambda_0 > 0$ (i.e. for any integer $1 \leq i \leq d$, P_i is a univariate Poisson process with constant intensity λ_0). The superposition $N = H + P$ is called a noisy Hawkes process.

Let us remark that, since the process P is aimed at modeling some kind of background noise, it is naturally assumed that subprocesses of P share the same intensity λ_0. However, if identifiability results from the forthcoming sections are specific to this assumption, the presentation done in this section can be trivially extended to a multivariate Poisson process with different intensities.

Now, let us recall that we aim at analysing the process $N = H + P$ through an observation of event times $(T_k^{N_1})_{k\geq 1}, \ldots, (T_k^{N_d})_{k\geq 1}$ with the inability of distinguishing between the times of H and P. To do so from a statistical estimation perspective, we leverage the spectral log-likelihood (Equation (4)), which makes use of the spectral density of N. The latter can be computed thanks to Proposition 2.2 and Example 8.3(c) from Daley and Vere-Jones (2003). Indeed, let $\tilde{h} : \mathbb{R} \to \mathbb{C}^{d\times d}$ be the matrix-valued Fourier transform of the interaction functions:

$$\forall w \in \mathbb{R} : \quad \tilde{h}(\omega) = \left(\tilde{h}_{ij}(\omega)\right)_{1\leq i,j\leq d} ,$$

under stationarity conditions, the spectral density $\mathbf{f}^H$ of the Hawkes process H (defined as in Equation (1)) is (Daley and Vere-Jones, 2003, Equation (8.3.11)):

$$\forall w \in \mathbb{R} : \quad \mathbf{f}^H(\omega) = \left(I_d - \tilde{h}(-\omega)\right)^{-1} \operatorname{diag}\left(m^H\right) \left(I_d - \tilde{h}(\omega)^T\right)^{-1} , \tag{9}$$

where I_d is the identity matrix of dimension d and $\operatorname{diag}\left(m^H\right)$ is the diagonal matrix formed by the vector of the mean intensities:

$$\begin{pmatrix} m_1^H \\ \vdots \\ m_d^H \end{pmatrix} = \left(I_d - \tilde{h}(0)\right)^{-1} \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix} .$$

Moreover, since the homogeneous Poisson process P (with common intensity λ_0) is a Hawkes process with null interactions, its spectral density $\mathbf{f}^P$ results easily from Equation (9):

$$\forall w \in \mathbb{R} : \quad \mathbf{f}^P(\omega) = \lambda_0 I_d ,$$

which leads to the spectral density of N:

$$\forall w \in \mathbb{R} : \quad \mathbf{f}^N(\omega) = \mathbf{f}^H(\omega) + \lambda_0 I_d , \tag{10}$$

with $\mathbf{f}^H$ expressed in Equation (9). Given this result, the inference procedure consists in maximising the spectral log-likelihood described in Equation (4).

In the forthcoming sections, we focus on this statistical estimation problem in low-dimensional settings ($d = 1$ and $d = 2$) and provide identifiability results when interactions are exponential.

3 The univariate noisy Hawkes process

3.1 General setting

Let us start with univariate processes ($d = 1$). In this case, Equation (10) simplifies, as stated in Corollary 3.1.

Corollary 3.1. *Let N be a noisy Hawkes process defined by the superposition of a stationary Hawkes process H (with baseline intensity $\mu > 0$ and kernel function $h : \mathbb{R} \to \mathbb{R}_{\geq 0}$) and an independent homogeneous Poisson process P (with constant intensity $\lambda_0 > 0$). Then the spectral density f^N of N reads:*

$$\forall \omega \in \mathbb{R}, \quad f^N(\omega) = \frac{\mu}{(1 - \|h\|_1)\left|1 - \tilde{h}(\omega)\right|^2} + \lambda_0 .$$

Proof. This is straightforward from Equations (9) and (10). See also (Daley and Vere-Jones, 2003, Example 8.2(e)) for the spectral density of a univariate exciting Hawkes process. □

The estimation procedure also simplifies since the periodogram of N and the spectral log-likelihood (see Equations (3) and (4)) respectively read:

$$\forall \omega \in \mathbb{R}, \quad I^T(\omega) = \frac{1}{T} \left| \sum_{k=1}^{N(T)} \mathrm{e}^{-2\pi \mathrm{i} \omega T_k^N} \right|^2 ,$$

where $(T_k^N)_{k \geq 1}$ is the sequence of event times of the noisy Hawkes process N, and:

$$\forall \theta \in \Theta, \quad \ell_T(\theta) = -\frac{1}{T} \sum_{k=1}^{M} \left(\log \left(f_\theta^N(\omega_k) \right) + \frac{I^T(\omega_k)}{f_\theta^N(\omega_k)} \right) . \tag{11}$$

As explained in Section 1.1, the Whittle estimator $\hat{\theta}$ is then obtained by maximising the function ℓ_T. In the next section, the exponential parametric model $\mathcal{Q}$ for f^N is described and analysed.

3.2 Exponential model

Let us consider the classic exponential kernel for the Hawkes process H:

$$\forall t \in \mathbb{R} : \quad h(t) = \alpha \beta \mathrm{e}^{-\beta t} \mathbb{1}_{t \geq 0} , \tag{12}$$

with $0 < \alpha < 1$ and $\beta > 0$. The kernel parameter is thus $\gamma = (\alpha, \beta)$ and the statistical model for a univariate noisy Hawkes process becomes:

$$\mathcal{Q} = \left\{ f_\theta^N : \mathbb{R} \to \mathbb{C}, \theta = (\mu, \alpha, \beta, \lambda_0) \in \mathbb{R}_{>0} \times (0,1) \times \mathbb{R}_{>0} \times \mathbb{R}_{>0} \right\} .$$

The exponential kernel parametrisation has been widely studied from an inference point of view (see for instance Ozaki (1979); Bacry and Muzy (2016)). In particular, its Fourier transform is:

$$\forall \omega \in \mathbb{R} : \quad \tilde{h}(\omega) = \frac{\alpha \beta}{\beta + 2\pi \mathrm{i} \omega} , \tag{13}$$

and the spectral density f^H of a univariate Hawkes process with baseline intensity $\mu > 0$ and exponential kernel is (Hawkes, 1971):

$$\forall \omega \in \mathbb{R} : \quad f^H(\omega) = m^H \left(1 + \frac{\beta^2 \alpha (2 - \alpha)}{\beta^2 (1-\alpha)^2 + 4\pi^2 \omega^2} \right) , \quad \text{where } m^H = \frac{\mu}{1-\alpha} .$$

It results that the spectral density $f_\theta^N \in \mathcal{Q}$ of a univariate noisy Hawkes process is:

$$\forall \omega \in \mathbb{R} : \quad f_\theta^N(\omega) = \left(\frac{\mu}{1-\alpha} \beta^2 \alpha (2-\alpha) \right) \frac{1}{\beta^2 (1-\alpha)^2 + 4\pi^2 \omega^2} + \left(\frac{\mu}{1-\alpha} + \lambda_0 \right) . \tag{14}$$

Now, we should be ready for implementing the inference procedure based on maximising the spectral log-likelihood as expressed in Equation (11). However, it appears that the model $\mathcal{Q}$, as currently defined, is not identifiable (see Proposition 3.2). Hopefully, this model becomes identifiable when restricted to only three parameters, thus allowing the practicability of the proposed estimation method, as numerically illustrated in Section 5.1.

Proposition 3.2 (Identifiability in the univariate setting)**.** *The model $\mathcal{Q}$ is not identifiable. In particular, for any admissible parameter $\theta = (\mu, \alpha, \beta, \lambda_0)$ there exists an infinite number of admissible parameters θ' such that $f_\theta^N = f_{\theta'}^N$.*

However, the four collections of models defined below are identifiable:

1. *for all $\mu^\circ > 0$, $\mathcal{Q}_\mu = \left\{ f^N_{(\mu,\alpha,\beta,\lambda_0)} \in \mathcal{Q}, \mu = \mu^\circ \right\}$;*
2. *for all $\alpha^\circ \in (0,1)$, $\mathcal{Q}_\alpha = \left\{ f^N_{(\mu,\alpha,\beta,\lambda_0)} \in \mathcal{Q}, \alpha = \alpha^\circ \right\}$;*
3. *for all $\beta^\circ > 0$, $\mathcal{Q}_\beta = \left\{ f^N_{(\mu,\alpha,\beta,\lambda_0)} \in \mathcal{Q}, \beta = \beta^\circ \right\}$;*

4. for all $\lambda_0^\circ > 0$, $\mathcal{Q}_{\lambda_0} = \left\{ f^N_{(\mu,\alpha,\beta,\lambda_0)} \in \mathcal{Q}, \lambda_0 = \lambda_0^\circ \right\}$.

Proof. The proof is given in Appendix B. □

The previous discussion raises the question of whether this non-identifiability also extends to the distribution of the noisy Hawkes process. It turns out that, in the exponential case, Markov properties of the intensity function λ^H of the underlying Hawkes process help ensuring identifiability. Indeed, from the definition of the Hawkes process H and the stationarity of the Poisson process P, $\left(\lambda^N(t)\right)_{t\geq 0}$ is a Markov process: it decreases with rate $\beta(\lambda^H(t)-\mu) = \beta(\lambda^N(t)-\mu-\lambda_0)$, and the jumps occurring from the Hawkes process, with rate $\lambda^H(t) = \lambda^N(t) - \lambda_0$, are of size $\alpha\beta$, while the jumps occurring from the Poisson component, with rate λ_0, have no impact on the intensity of the process. Then $\left(\lambda^N(t), N(t)\right)_{t\geq 0}$ is also a Markov process, more specifically a piecewise deterministic Markov process (Davis, 1984). This allows us to use results from Dassios and Zhao (2011) on the distribution of exponential Hawkes processes to show that the distribution of the exponential noisy Hawkes process is identifiable.

Proposition 3.3. *Let* $(\mu, \alpha, \beta, \lambda_0)$ *and* $(\mu', \alpha', \beta', \lambda_0')$ *be two admissible 4-tuples for the exponential noisy Hawkes model* $\mathcal{Q}$*, and* N *and* N' *respectively defined by these two tuples (Equation (2) with kernel (12)). Then, if* N *and* N' *have same distribution and* $\lambda^N(0)$ *(respectively* $\lambda^{N'}(0)$*) is distributed according to the stationary distribution of* $\left(\lambda^N(t)\right)_{t\geq 0}$ *(respectively* $\left(\lambda^{N'}(t)\right)_{t\geq 0}$*), then* $(\mu, \alpha, \beta, \lambda_0) = (\mu', \alpha', \beta', \lambda_0')$.

Proof. See Appendix C. □

A consequence of this result is that non-identifiability of our proposed method in the exponential case is a shortcoming of the spectral approach itself rather than an underlying property of the noisy Hawkes process, presumably stemming from the fact that the spectral density only encodes the second order moments of the process. In the forthcoming section, we briefly investigate whether this issue arises when considering other reproduction kernels.

3.3 Beyond the exponential model

Let N be a noisy Hawkes process defined by the superposition of a stationary Hawkes process with baseline intensity μ and kernel function αh, with $\alpha \in (0,1)$, $h : \mathbb{R} \to \mathbb{R}_{\geq 0}$ and $\|h\|_1 = 1$, and a homogeneous Poisson process with intensity λ_0. Per Corollary 3.1, its spectral density is given by

$$\forall \omega \in \mathbb{R} : \quad f^N(\omega) = \frac{\mu}{(1-\alpha)\left|1 - \alpha\tilde{h}(\omega)\right|^2} + \lambda_0 \,.$$

While it may be difficult to show that a model is identifiable from the spectral density expression, it may prove fruitful to look at its Taylor expansion around 0, and analyse the Taylor coefficients. For example, considering the uniform kernel and the corresponding Taylor expansion of the spectral density up to order 2, we get the following proposition.

Proposition 3.4. *Let us consider a rectangle interaction function*

$$h : t \in \mathbb{R} \mapsto \phi^{-1} \mathbb{1}_{0\leq t\leq \phi} \,,$$

for some kernel parameter $\phi > 0$*, and the corresponding statistical model for a univariate noisy Hawkes process:*

$$\mathcal{R} = \left\{ f^N_\theta : \mathbb{R} \to \mathbb{C}, \theta = (\mu, \alpha, \phi, \lambda_0) \in \mathbb{R}_{>0} \times (0,1) \times \mathbb{R}_{>0} \times \mathbb{R}_{>0} \right\} .$$

Then $\mathcal{R}$ *is identifiable.*

Proof. See Appendix D. □

This last proposition shows that non-identifiability of the spectral approach for the noisy exponential Hawkes process is more a consequence of the choice of the reproduction function h rather than a general shortcoming of the spectral approach. It is unexpected that the exponential reproduction function, usually chosen because the Markov properties for the resulting Hawkes intensity simplify calculations (Ozaki, 1979; Da Fonseca and Zaatour, 2013; Duarte et al., 2019), seems to be here the main culprit of non-identifiability for our proposed spectral approach. Still, we will show how, by imposing some constraints on the modelling of multivariate noisy Hawkes processes, we are able to ensure identifiability of the model even in this case.

4 The bivariate noisy Hawkes process

This section addresses bivariate noisy Hawkes processes ($d = 2$). More precisely, for such a process $N = H + P$, where H is a bivariate Hawkes process (see Equation (1)) and P a Poisson process with shared intensity λ_0, Corollary 4.1 gives the closed-form expression of the spectral density $\mathbf{f}^N$.

Corollary 4.1. *Let $N = (N_1, N_2)$ be a bivariate noisy Hawkes process defined by the superposition of a stationary Hawkes process $H = (H_1, H_2)$ (with baseline intensities $\mu_1 > 0$ and $\mu_2 > 0$, and kernel functions $h_{ij} : \mathbb{R} \to \mathbb{R}_{\geq 0}$, $1 \leq i, j \leq 2$) and an independent homogeneous Poisson process $P = (P_1, P_2)$ (with same constant intensity $\lambda_0 > 0$). Then the spectral density $\mathbf{f}^N$ of N reads:*

$$\forall \omega \in \mathbb{R}, \quad \mathbf{f}^N(\omega) = \begin{pmatrix} f_{11}^H(\omega) + \lambda_0 & f_{12}^H(\omega) \\ f_{21}^H(\omega) & f_{22}^H(\omega) + \lambda_0 \end{pmatrix},$$

where

$$\begin{cases} f_{11}^H(\omega) = \frac{m_1^H \left|1-\tilde{h}_{22}(\omega)\right|^2 + m_2^H \left|\tilde{h}_{12}(\omega)\right|^2}{\left|\left(1-\tilde{h}_{11}(\omega)\right)\left(1-\tilde{h}_{22}(\omega)\right) - \tilde{h}_{12}(\omega)\tilde{h}_{21}(\omega)\right|^2} \\ f_{12}^H(\omega) = \frac{m_1^H \left(1-\tilde{h}_{22}(-\omega)\right)\tilde{h}_{21}(\omega) + m_2^H \left(1-\tilde{h}_{11}(\omega)\right)\tilde{h}_{12}(-\omega)}{\left|\left(1-\tilde{h}_{11}(\omega)\right)\left(1-\tilde{h}_{22}(\omega)\right) - \tilde{h}_{12}(\omega)\tilde{h}_{21}(\omega)\right|^2} \end{cases},$$

and

$$m_1^H = \frac{\mu_1 \left(1 - \|h_{22}\|_1\right) + \mu_2 \|h_{12}\|_1}{\left(1 - \|h_{11}\|_1\right)\left(1 - \|h_{22}\|_1\right) - \|h_{12}\|_1 \|h_{21}\|_1},$$

and f_{22}^H, f_{21}^H and m_2^H are obtained by symmetry of all indices.

Proof. This is straightforward from Equations (9) and (10). □

Then, the estimation procedure is exactly that described in Section 1.1, which is based on computing the cross periodogram $\mathbf{I}^T$ and on maximising the spectral log-likelihood ℓ_T (see Equations (3) and (4)). Now, similarly to the univariate case detailed in Section 3.2, we consider exponential interaction functions, i.e. for $1 \leq i, j \leq 2$:

$$\forall t \in \mathbb{R} : \quad h_{ij}(t) = \alpha_{ij} \beta_i \mathrm{e}^{-\beta_i t} \mathbb{1}_{t \geq 0},$$

with $\alpha_{ij} \geq 0$ and $\beta_i > 0$. The kernel parameter is thus $\gamma = (\alpha, \beta)$, where $\alpha \in \mathbb{R}_{\geq 0}^{2\times 2}$ and $\beta \in \mathbb{R}_{>0}^2$ and the statistical model for a bivariate noisy Hawkes process becomes:

$$\mathcal{Q}_\Lambda = \left\{ \mathbf{f}_\theta^N : \mathbb{R} \to \mathbb{C}^{2\times 2}, \theta = (\mu, \alpha, \beta, \lambda_0) \in \mathbb{R}_{>0}^2 \times \Lambda \times \mathbb{R}_{>0}^2 \times \mathbb{R}_{>0}, \beta \in \Omega_\alpha \right\},$$

where $\Lambda \subset \{\alpha \in \mathbb{R}_{\geq 0}^{2\times 2} : \rho(\alpha) < 1\}$ is subset of matrices α that will be specified later, and for all $\alpha \in \mathbb{R}_{\geq 0}^{2\times 2}$,

$$\Omega_\alpha = \left\{ \beta \in \mathbb{R}_{>0}^2, \beta_1 = 1 \text{ if } \alpha_{11} = \alpha_{12} = 0, \beta_2 = 1 \text{ if } \alpha_{21} = \alpha_{22} = 0 \right\},$$

is a subset of admissible values for β. The definition of Ω_α takes into account that when a row, say the first one, of the interaction matrix α is null, then the corresponding kernels verify $h_{11} = 0$ and $h_{12} = 0$ independently of the value of β_1. Thus, identifiability for the parameter β_1 is hopeless, which justifies that we get rid of it (by fixing it to an arbitrary value) from the model.

Remark 4.1. *Different versions of the multivariate exponential model exist. A first convention assumes that there is a unique $\beta \in \mathbb{R}_{>0}$ shared by all kernel functions (Chevallier et al., 2019; Bacry et al., 2020). A second and less restrictive option, which is that we opt for in this paper, assumes that the recovery rate $\beta_i \in \mathbb{R}_{>0}$ for each subprocess N_i $(1 \leq i \leq d)$ is shared among received interactions (Bonnet et al., 2023). These choices allow for simplified derivations of estimators in the time domain and in the frequency domain, as shown below.*

The aim of this section is to study identifiability of model $\mathcal{Q}_\Lambda$. A broad analysis (i.e. for $\Lambda = \{\alpha \in \mathbb{R}_{\geq 0}^{2\times 2} : \rho(\alpha) < 1\}$) being out of reach for complexity reasons, we exhibit some situations (i.e. subsets Λ) for which non-identifiability (Proposition 4.2) or identifiability (Proposition 4.3) can be proved.

Proposition 4.2 (Non-identifiability in the bivariate setting)**.** *The model $\mathcal{Q}_\Lambda$ is not identifiable in the three situations:*

1. $\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & 0 \\ 0 & \alpha_{22} \end{pmatrix}, 0 \leq \alpha_{11}, \alpha_{22} < 1 \right\}$, *that is for diagonal matrices α (with possibly null entries).*
2. $\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ 0 & 0 \end{pmatrix}, 0 < \alpha_{11} < 1, \alpha_{12} > 0 \right\}$, *that is for matrices with positive entries in the first row and null entries in the second.*
3. $\Lambda = \left\{ \begin{pmatrix} 0 & 0 \\ \alpha_{21} & \alpha_{22} \end{pmatrix}, \alpha_{21} > 0, 0 < \alpha_{22} < 1 \right\}$, *that is for matrices with null entries in the first row and positive entries in the second.*

Proof. The proof is given in Appendix E. □

Remark 4.2. *The proof of Proposition 4.2, Situation 1 reveals that non-identifiability stands actually for each subprocess (considered as a univariate process), such that all the submodels with null cross-interactions built by fixing α_{11} or α_{22} to zero, or by keeping them away from zero are also not identifiable.*

Proposition 4.3 (Identifiability in the bivariate setting)**.** *The model $\mathcal{Q}_\Lambda$ is identifiable in the six situations:*

1. $\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{21} & 0 \end{pmatrix}, 0 \leq \alpha_{11} < 1, \alpha_{21} > 0 \right\}$, *that is for matrices α with null entries in the second column and a positive entry on the antidiagonal.*
2. $\Lambda = \left\{ \begin{pmatrix} 0 & \alpha_{12} \\ 0 & \alpha_{22} \end{pmatrix}, \alpha_{12} > 0, 0 \leq \alpha_{22} < 1 \right\}$, *that is for matrices α with null entries in the first column and a positive entry on the antidiagonal.*
3. $\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{21} & \alpha_{22} \end{pmatrix}, 0 < \alpha_{11} < 1, \alpha_{21} > 0, 0 \leq \alpha_{22} < 1 \right\}$, *that is for matrices α with positive entries in the first column and null upper right entry.*
4. $\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ 0 & \alpha_{22} \end{pmatrix}, 0 \leq \alpha_{11} < 1, \alpha_{12} > 0, 0 < \alpha_{22} < 1 \right\}$, *that is for matrices α with a null lower left entry and positive entries in the second column.*

Proof. The proofs of Situations 1 and 3 are respectively in Appendices F and G. The other situations are obtained by symmetry of all indices. □

Several lessons can be learnt from Propositions 4.2 and 4.3 and Remark 4.2. First, the statistical model $\mathcal{Q}_\Lambda$ is not identifiable if H reduces to a bivariate homogenous Poisson process (Proposition 4.2, Situation 1 with $\alpha_{11} = \alpha_{22} = 0$) or to two independent univariate Hawkes processes (Proposition 4.2, Situation 1 with $\alpha_{11} > 0$ and $\alpha_{22} > 0$) even if the noise P shares the same intensity λ_0 for both subprocesses. This result actually still holds true for a dimension $d > 2$.

Second, Proposition 4.3 tells in a nutshell that there must exist cross-interactions in the Hawkes process H (i.e. $\alpha_{12} > 0$ or $\alpha_{21} > 0$) for the model $\mathcal{Q}_\Lambda$ to be identifiable. However, interactions must not come from a Poisson subprocess and reach a self-exciting Hawkes subprocess (Proposition 4.2, Situations 2 and 3), but rather they must reach a self-neutral (i.e. with null self-excitation) Hawkes subprocess (Proposition 4.3, Situations 1 and 2) or come from a self-exciting Hawkes subprocess (Proposition 4.3, Situations 3 and 4).

5 Numerical results

This section numerically illustrates the behaviour of the proposed estimator $\hat{\theta}$ (described in Section 1.1) for noisy exponential Hawkes processes. It investigates the effect of horizon T and hyperparameter M in the univariate setting (Section 5.1), and the impact of model sparsity and interaction strength in the bivariate setting (Section 5.2).

In the whole study, point processes are simulated thanks to Ogata's thinning method (Ogata, 1981) and numerical optimisation of the spectral log-likelihood is performed via the L-BFGS-B method (Byrd et al., 1995), implemented in the `scipy.optimize.minimize` Python function. Both simulation and estimation algorithms are freely available as a Python package on GitHub[1]

5.1 Univariate setting

5.1.1 Simulation and estimation scenarios

Data simulation We consider observations $(T_k^N)_{1\leq k\leq N(T)}$ coming from a univariate exponential noisy Hawkes process $N = H + P$, where P is a Poisson process with intensity $\lambda_0 > 0$ and H is a Hawkes process with baseline intensity $\mu = 1$ and kernel given by Equation (12) with parameters $\alpha = 0.5$ and $\beta = 1$.

In order to get close to stationarity while coping with inability to generate a process on the whole line $\mathbb{R}$, N is simulated on the window $[-100, T]$ with no points in $(-\infty, -100)$ but only observations falling in $[0, T]$ are considered.

The forthcoming section will illustrate the convergence of $\hat{\theta}$, thanks to its behaviour with respect to varying horizon $T \in \{250, 500, 1000, \dots, 8000\}$, and the impact of the intensity λ_0 of the noise process P on estimation accuracy, via varying noise-to-signal ratio $\lambda_0/m^H \in \{0.2, 0.4, \dots, 2.0\}$ (given the average intensity $m^H = \mu/(1-\alpha) = 2$ of H).

Statistical models According to Proposition 3.2, which states four collections of identifiable models for univariate exponential Hawkes processes, estimation is successively performed in models $\mathcal{Q}_\mu$, $\mathcal{Q}_\alpha$, $\mathcal{Q}_\beta$, $\mathcal{Q}_{\lambda_0}$, where the known parameter is fixed to the value of the generated process (see above).

In addition, the behaviour of $\hat{\theta}$ will be assessed thanks to its relative error $\|\hat{\theta} - \theta^\star\|_2 / \|\theta^\star\|_2$ (where $\theta^\star = (\mu, \alpha, \beta, \lambda_0)$ is the vector of the parameters of the generated process) averaged over 50 different trials.

5.1.2 Convergence, computation time and influence of parameter M

Up to now, the hyperparameter M, appearing in the spectral log-likelihood (Equations (4) and (11)) and determining the number of frequencies tested with the spectral density, has been let free. However, the theoretical literature suggests that its choice has a lot of influence on the convergence of the estimation procedure, and has to be guided by $M/T \xrightarrow[T\to\infty]{} \infty$ (Pham, 1996).

Since the rate of convergence is not specified but has an effect on the computational efficiency of the spectral estimator, we propose to study the compatible case $M = N(T) \log N(T)$ and the economy case $M = N(T)$.

Figure 1 displays the relative errors $\|\hat{\theta} - \theta^\star\|_2 / \|\theta^\star\|_2$ with respect to both the simulation horizon T (top panels) and the estimation time (bottom panels). As expected, the quality of the estimations improves as T increases, and this independently of which parameter is fixed. Estimations are

[1] https://github.com/migmtz/noisy-hawkes-process

slightly better when considering $M = N(T)\log N(T)$ (orange line) especially for smaller values of T but the trade-off is a ten times higher computation time. Therefore, the performance benefit of taking $M = N(T)\log N(T)$ rather than $M = N(T)$ seems minor when compared to the computational cost. For this reason, the forthcoming numerical experiments will be performed with $M = N(T)$.

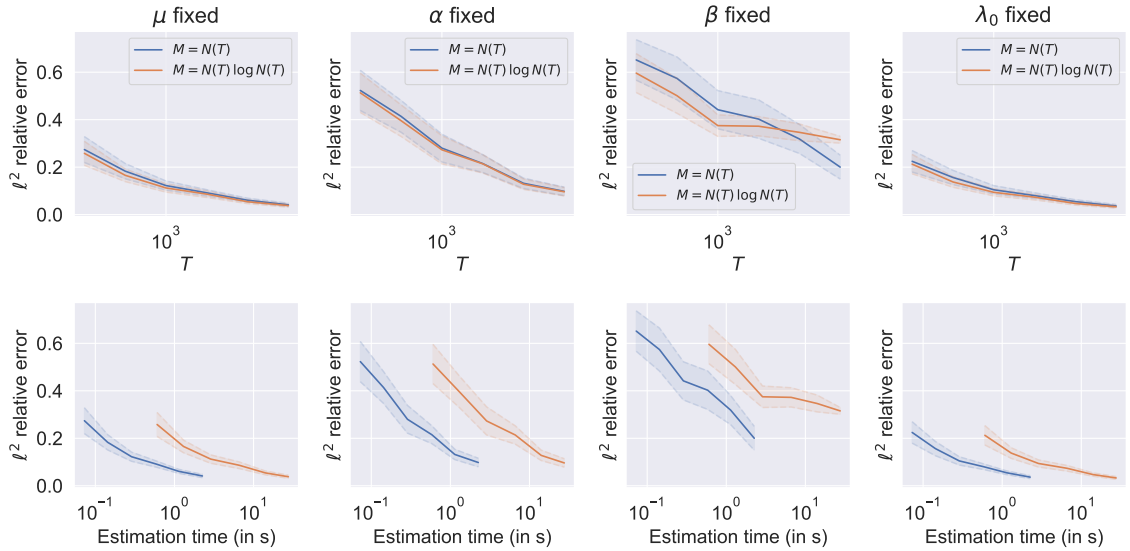

Figure 1: Relative estimation error with confidence bands (± 1.96 empirical standard deviation) respectively for μ, α, β and λ_0 fixed (columns from left to right) with respect to the time horizon T (top) and the computation time (bottom) in logarithmic scale. Level of noise $\lambda_0 = 1.6$.

5.1.3 Influence of the noise level

Figure 2 shows the relative error with respect to the ratio λ_0/m_H, obtained when increasing the value of λ_0 while keeping the other parameters fixed, for a given horizon $T = 8000$.

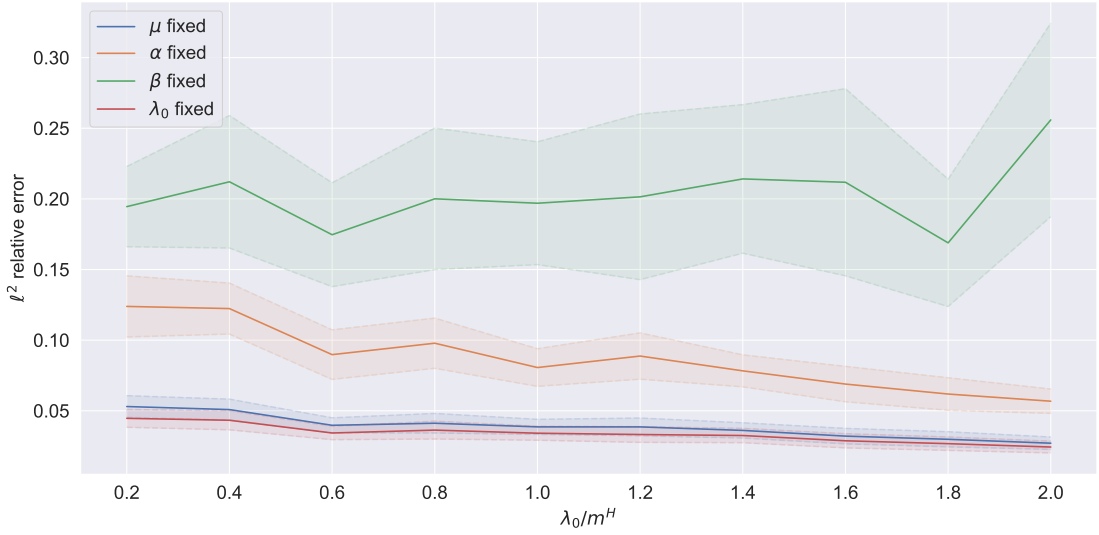

Figure 2: Relative estimation error with confidence bands (± 1.96 empirical standard deviation) respectively for μ, α, β and λ_0 fixed with respect to the noise-to-signal ratio for the maximal horizon $T = 8000$.

First, we can see that the value of λ_0 has a low impact on the quality of estimations. However, let us notice that when β is fixed, the average error is substantially larger than when any of the other parameters is fixed. This could be explained by a compensation phenomenon inside the triplet (μ, α, λ_0) which occurs as our method implicitly adjusts the estimation to the mean intensity of the noisy Hawkes process:

$$m^N = \lambda_0 + \frac{\mu}{1-\alpha},$$

which is indeed a quantity independent of β.

This numerical compensation is illustrated in Figure 3, where we can see that overestimating μ is systematically balanced by underestimating α and λ_0 and vice versa, whereas the estimated mean intensities remain accurate. In this experiment, the level of noise has been arbitrarily fixed to $\lambda_0 = 1.2$ but the observed behaviour appears similarly for all possible values of λ_0 used in the previous section.

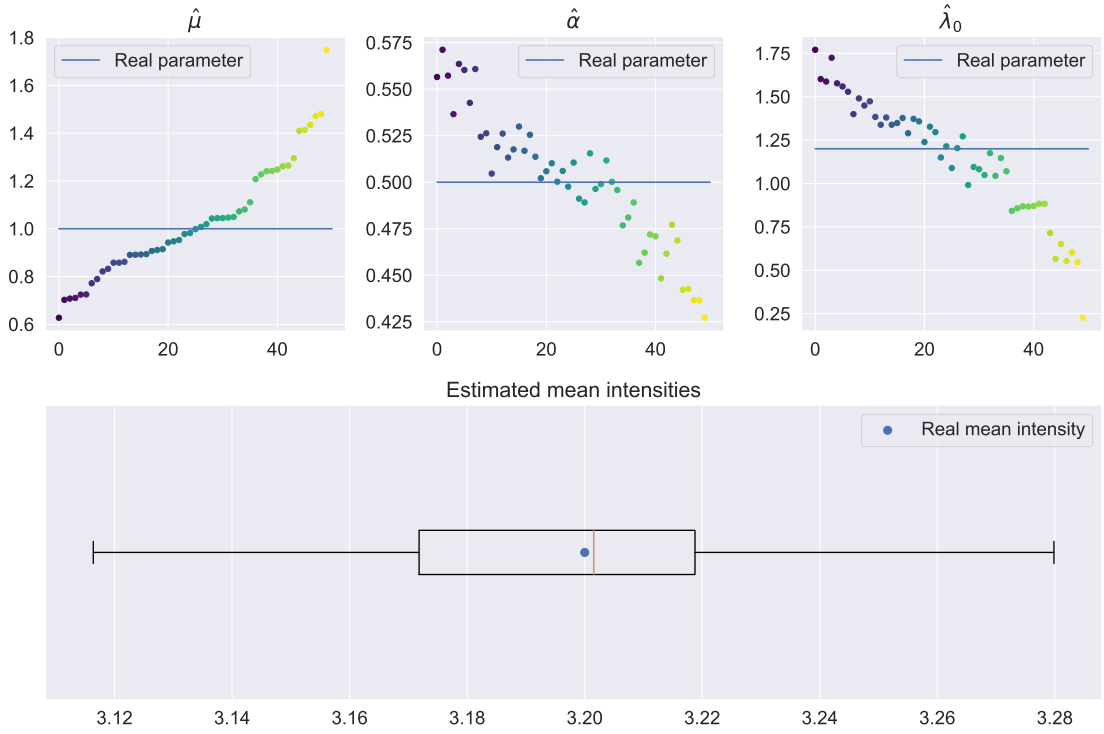

Figure 3: Estimations $\hat{\mu}, \hat{\alpha}$ and $\hat{\lambda}_0$ (of μ, α, and λ_0) for $\lambda_0 = 1.2$ when β is fixed, sorted by the values of $\hat{\mu}$ (top). In all plots, each color corresponds to one of the 50 repetitions. Boxplot of estimated mean intensities (bottom).

When performing estimation with α fixed (the case for which the average error is the second larger, as illustrated in Figure 2), the compensation appears only between $\hat{\mu}$ and $\hat{\lambda}_0$ whereas $\hat{\beta}$ does not seem impacted, as shown in Figure 4.

5.2 Bivariate setting

This section illustrates numerical estimation of bivariate exponential noisy Hawkes processes (see Section 4) when conditions of identifiability are met (Proposition 4.3). We carry out two different studies, exploring different scenarios: Section 5.2.1 studies the influence of the strength of the cross-interaction between the two subprocesses and Section 5.2.2 investigates the performance of the estimator with and without knowledge of the null components. Indeed, since identifiability conditions depend on knowing which components are non-null, an information that is unlikely to

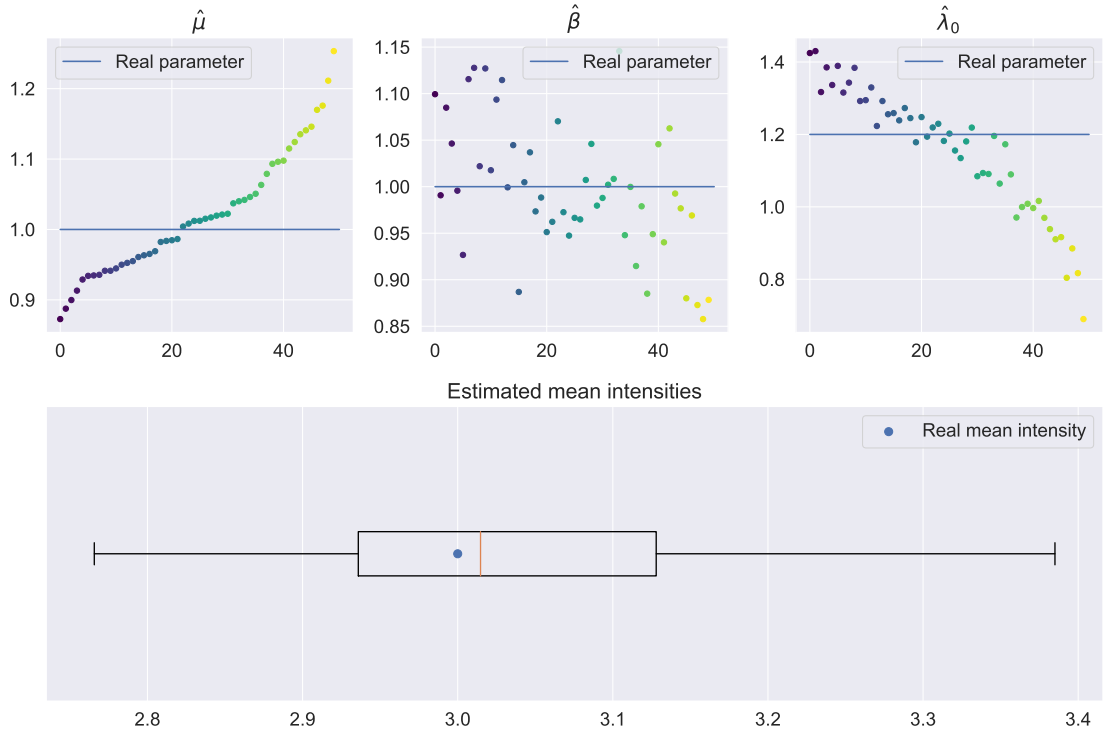

Figure 4: Estimations $\hat{\mu}, \hat{\beta}$ and $\hat{\lambda}_0$ (of μ, β, and λ_0) for $\lambda_0 = 1.2$ when α is fixed, sorted by the values of $\hat{\mu}$ (top). In all plots, each color corresponds to one of the 50 repetitions. Boxplot of estimated mean intensities (bottom).

be available in practical applications, we compare the performance of the estimator for both the reduced model $\mathcal{Q}_\Lambda$, where the null components are known, and the complete model,

$$\mathcal{Q} = \left\{ \mathbf{f}_\theta^N : \mathbb{R} \to \mathbb{C}^{2\times 2}, \theta = (\mu, \alpha, \beta, \lambda_0) \in \mathbb{R}_{>0}^2 \times \mathbb{R}_{\geq 0}^{2\times 2} \times \mathbb{R}_{>0}^2 \times \mathbb{R}_{>0}, \rho(\alpha) < 1 \right\},$$

with no prior information.

Throughout this section, we consider a Hawkes process with $\mu = \begin{pmatrix} 1.0 \\ 1.0 \end{pmatrix}$ and $\beta = \begin{pmatrix} 1.0 \\ 1.3 \end{pmatrix}$. In addition, fortified by the analysis of the univariate setting, the Poisson intensity is chosen to be $\lambda_0 = 0.5$ (the level of noise does not appear to have a significant impact on the quality of estimations, Figure 2) and it is considered $M = N(T)$ (which provides accurate estimations in a reasonable amount of time, Figure 1).

5.2.1 Influence of the cross-interaction

Let us consider one of the identifiable scenarios where the only non-null interaction in the Hawkes process is one of the two cross-interactions (see Proposition 4.3, Situation 1). More precisely, we consider the reduced model $\mathcal{Q}_\Lambda$, where

$$\Lambda = \left\{ \begin{pmatrix} 0 & 0 \\ \alpha_{21} & 0 \end{pmatrix} : \alpha_{21} > 0 \right\}.$$

The Hawkes process is then simulated with different levels of cross-interaction: $\alpha_{21} \in \{0.2, 0.4, 0.6, 0.8\}$, and estimations are obtained by optimising the spectral log-likelihood on the non-null parameters μ_1, μ_2, α_{21}, β_2, and λ_0.

Figure 5 illustrates the influence of the true parameter α_{21} on the quality of the estimations, through the relative error of the estimations for the different values of α_{21} and an increasing range

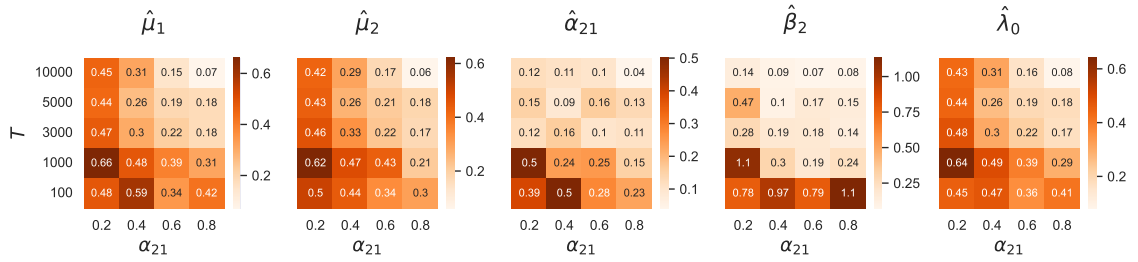

Figure 5: Heatmap of relative errors for each estimation $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\alpha}_{21}$, $\hat{\beta}_2$, and $\hat{\lambda}_0$ for different levels of interaction α_{21} (x-axis) and horizons T (y-axis).

of horizons T. As a complement to what has been observed in Figure 1, our estimator appears to behave particularly well for higher values of T, but also for higher values of α_{21}.

This is not surprising since, for smaller values of α_{21}, the Hawkes process behaves closely to a homogeneous Poisson process, and as proven in Proposition 4.2, the superposition of two Poisson processes leads to a non-identifiable model. Lower interactions necessitate then higher values of T to obtain satisfactory results. Inversely, for average and high interaction magnitudes, we start to obtain small errors for horizon values around $T = 3000$.

5.2.2 Influence of null interactions

In this section we simulate 50 repetitions with a fixed horizon $T = 3000$ for two identifiable scenarios regarding the Hawkes process $H = (H_1, H_2)$.

Scenario 1 The matrix of interactions is:

$$\alpha = \begin{pmatrix} 0.5 & 0 \\ 0.4 & 0 \end{pmatrix},$$

corresponding to Proposition 4.3, Situation 1. In other terms, H_1 excites both subprocesses whereas H_2 has no influence on the dynamics (See Figure 6, left).

Scenario 2 The matrix of interactions is:

$$\alpha = \begin{pmatrix} 0.5 & 0 \\ 0.4 & 0.4 \end{pmatrix},$$

corresponding to Proposition 4.3, Situation 3. In other terms, H_1 excites both subprocesses and H_2 excites itself (See Figure 6, right).

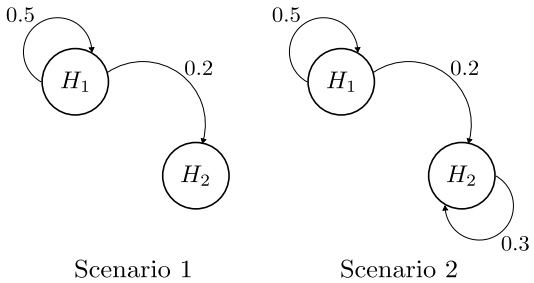

Figure 6: Interactions in the two numerical scenarios considered.

Graphics in the left column of Figure 7 present the boxplots of each parameter estimation when considering their respective reduced models $\mathcal{Q}_\Lambda$. These results show that our method provides unbiased estimates of all parameters and is particularly efficient at inferring the interaction matrix α (estimations have very low variance). The larger variances are observed for parameters μ_1, μ_2 and λ_0, which is probably due to compensation effects already mentioned in Section 5.1.3.

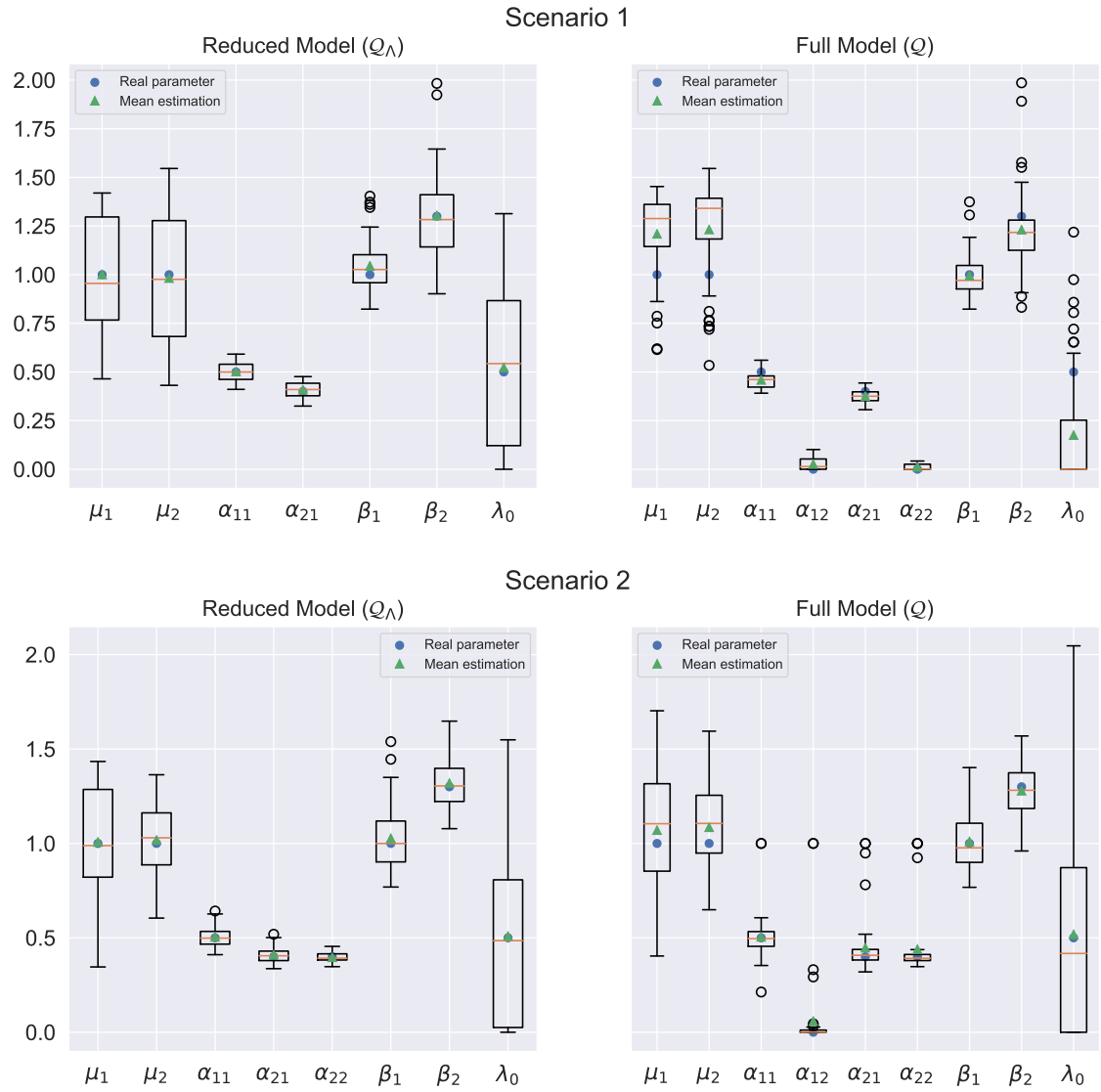

Figure 7: Boxplots of parameter estimations in the reduced model $\mathcal{Q}_\Lambda$ (left) and full model $\mathcal{Q}$ (right) in Scenario 1 (top) and Scenario 2 (bottom) (50 trials). Average estimation (green triangle) is to be compared to true parameter (blue point).

Estimating in the reduced model $\mathcal{Q}_\Lambda$ requires prior knowledge on the null parameters α_{ij} $(1 \leq i, j \leq 2)$, which is unlikely in practical applications. Therefore, we compare the results obtained in the reduced model to those in the complete model $\mathcal{Q}$ (see the right column graphics in Figure 7). If the estimates of α and β seem still empirically unbiased, we observed several deteriorations compared to the previous results. First, we notice a bias in the estimates of μ_1, μ_2 and λ_0: more precisely, μ_1 and μ_2 are overestimated while λ_0 is underestimated in both scenarios. Moreover, we observe in Scenario 2 some outlier estimations for the α_{ij} $(1 \leq i, j \leq 2)$ coefficients, which did not appear when considering the reduced model.

Fortunately, Figure 7 also suggests that our estimator is able to detect the null interactions in the full model, which allows to re-estimate the parameters in the reduced model. To do so, we propose to look at the 5%-empirical quantile of each term of the estimated interaction matrix $\hat{\alpha}$, which are summarised in Table 1.

We can notice that the empirical 5%-quantiles are set to zero for each real null parameters α_{ij} $(1 \leq i, j \leq 2)$ in both scenarios. This suggests that when enough repetitions are available, it is possible to use these empirical quantiles to estimate the null interactions. An estimation

	α_{11}	α_{12}	α_{21}	α_{22}
Scenario 1	0.40	0.00	0.32	0.00
Scenario 2	0.41	0.00	0.34	0.35

Table 1: 5%-quantile of each parameter of the estimated interaction matrix $\hat{\alpha}$. The true parameter α_{12} is null in both scenarios and α_{22} is null only in Scenario 1.

procedure when no prior information is known about the support of the interaction graph would then consist in a three-step approach: first, estimating all parameters in the complete model $\mathcal{Q}$; second, computing the 5%-quantiles for all estimated interactions parameters in matrix $\hat{\alpha}$, and defining the support of α to be entries corresponding to a positive empirical quantile; finally, re-estimating all parameters in the reduced model defined by the support of α. Let us remark that the proposed support estimation step boils down to correspond to a multiple test that, when the noisy Hawkes process has significantly more that 2 dimensions, can be corrected thanks to usual procedures such as Bonferroni and Benjamini-Hochberg methods.

6 Discussion

In this paper, we propose a spectral approach to estimate the parameters of a noisy Hawkes process, the performance of which is illustrated on an extended numerical study. Although we highlight the great benefit of considering a spectral analysis when standard inference methods are not available, we also bring out identifiability issues that may arise, either from the model itself or from the spectral approach. While we exhibit several identifiable and non-identifiable scenarios in both univariate and bivariate contexts, a general result on identifiability is still to be established, in particular in higher dimensions. For this purpose, the number of non-null cross-interactions and the choice of the kernel functions appear to be key elements in order to obtain identifiability guarantees.

More generally, we believe that the spectral analysis can provide efficient estimators in many frameworks of inaccurately or partially observed data. A natural extension of this work is to consider alternative forms for the noise process, for instance an heterogeneous Poisson process. Another topic of interest is to investigate another mechanism for the noise when some points are randomly missing. This would be complementary to our work since it would allow to model both false positive and false negative occurrences. In practice, this could be of great interest for applications, especially for the tracking of epidemics. Finally, let us mention that this paper focuses on the linear Hawkes model, which excludes notably inhibition phenomenons. Though the nonlinear framework is particularly interesting for many applications, for instance in neuroscience, all the spectral theory for point processes only exists in a linear context so that we believe that developing a spectral inference procedure for nonlinear processes would be very challenging and remains a widely open topic.

References

Adamopoulos, L. (1976), "Cluster models for earthquakes: Regional comparisons," *Journal of the International Association for Mathematical Geology*, 8, 463–475.

Antoniadis, A. and Bigot, J. (2006), "Poisson inverse problems," *Annals of Statistics*, 34, 2132–2158.

Bacry, E., Bompaire, M., Gaïffas, S., and Muzy, J. (2020), "Sparse and low-rank multivariate Hawkes processes," *Journal of Machine Learning Research*, 21, 1–32.

Bacry, E., Mastromatteo, I., and Muzy, J.-F. (2015), "Hawkes Processes in Finance," *Market Microstructure and Liquidity*, 01, 1550005.

Bacry, E. and Muzy, J. (2016), "First- and Second-Order Statistics Characterization of Hawkes Processes and Non-Parametric Estimation," *IEEE Transactions on Information Theory*, 62, 2184–2202.

Bartlett, M. S. (1963), "The Spectral Analysis of Point Processes," *Journal of the Royal Statistical Society. Series B (Methodological)*, 25, 264–296.

— (1964), "The Spectral Analysis of Two-Dimensional Point Processes," *Biometrika*, 51, 299–311.

Bonnet, A., Lacour, C., Picard, F., and Rivoirard, V. (2022), "Uniform deconvolution for Poisson point processes," *Journal of Machine Learning Research*, 23.

Bonnet, A., Martinez Herrera, M., and Sangnier, M. (2023), "Inference of multivariate exponential Hawkes processes with inhibition and application to neuronal activity," *Statistics and Computing*, 33.

Brillinger, D. R. (2012), *Statistical Inference for Stationary Point Processes*, New York, NY: Springer New York, 499–543.

Brémaud, P. and Massoulié, L. (1996), "Stability of nonlinear Hawkes processes," *The Annals of Probability*, 24, 1563–1588.

Byrd, R., Lu, P., Nocedal, J., and Zhu, C. (1995), "A Limited Memory Algorithm for Bound Constrained Optimization," *SIAM Journal on Scientific and Statistical Computing*, 16, 1190–1208.

Chevallier, J., Duarte, A., Löcherbach, E., and Ost, G. (2019), "Mean field limits for nonlinear spatially extended Hawkes processes with exponential memory kernels," *Stochastic Processes and their Applications*, 129, 1–27.

Cheysson, F. and Lang, G. (2022), "Spectral estimation of Hawkes processes from count data," *The Annals of Statistics*, 50, 1722 – 1746.

Chornoboy, E. S., Schramm, L. P., and Karr, A. F. (1988), "Maximum likelihood identification of neural point process systems," *Biological Cybernetics*, 59, 265–275.

Da Fonseca, J. and Zaatour, R. (2013), "Hawkes Process: Fast Calibration, Application to Trade Clustering, and Diffusive Limit," *Journal of Futures Markets*, 34, 548–579.

Daley, D. and Vere-Jones, D. (2003), *An introduction to the theory of point processes. Vol. I*, New York: Springer-Verlag, second edition.

Daley, D. J. (1971), "Weakly Stationary Point Processes and Random Measures," *Journal of the Royal Statistical Society. Series B (Methodological)*, 33, 406–428.

Dassios, A. and Zhao, H. (2011), "A dynamic contagion process," *Advances in Applied Probability*, 43, 814–846.

Davis, M. H. A. (1984), "Piecewise-Deterministic Markov Processes : A General Class of Non-Diffusion Stochastic Models," *Journal of the Royal Statistical Society Series B (Methodological)*, 46, 353–388.

Deutsch, I. and Ross, G. (2020), "ABC Learning of Hawkes Processes with Missing or Noisy Event Times," Preprint at https://arxiv.org/abs/2006.09015.

Duarte, A., Löcherbach, E., and Ost, G. (2019), "Stability, convergence to equilibrium and simulation of non-linear Hawkes processes with memory kernels given by the sum of Erlang kernels," *ESAIM: Probability and Statistics*, 23, 770–796.

Düker, M.-C. and Pipiras, V. (2019), "Asymptotic results for multivariate local Whittle estimation with applications," in *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*.

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

Guo, X., Hu, A., Xu, R., and Zhang, J. (2018), "Consistency and Computation of Regularized MLEs for Multivariate Hawkes Processes," Preprint at https://arxiv.org/abs/1810.02955.

Gupta, A., Farajtabar, M., Dilkina, B., and Zha, H. (2018), "Discrete Interventions in Hawkes Processes with Applications in Invasive Species Management," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, International Joint Conferences on Artificial Intelligence Organization.

Hawkes, A. (1971), "Spectra of Some Self-Exciting and Mutually Exciting Point Processes," *Biometrika*, 58, 83–90.

Hohage, T. and Werner, F. (2016), "Inverse problems with Poisson data: statistical regularization theory, applications and algorithms," *Inverse Problems*, 32, 093001.

Karavasilis, G. J. and Rigas, A. G. (2007), "Spectral analysis techniques of stationary point processes used for the estimation of cross-correlation: Application to the study of a neurophysiological system," in *2007 15th European Signal Processing Conference*.

Kirchner, M. (2017), "An estimation procedure for the Hawkes process," *Quantitative Finance*, 17, 571–595.

Lambert, R., Tuleau-Malot, C., Bessaih, T., Rivoirard, V., Bouret, Y., Leresche, N., and Reynaud-Bouret, P. (2018), "Reconstructing the functional connectivity of multiple spike trains using Hawkes models," *Journal of Neuroscience Methods*, 297, 9–21.

Lewis, E. and Mohler, G. (2011), "A Nonparametric EM Algorithm for Multiscale Hawkes Processes," *Journal of Nonparametric Statistics*, 1, 1–20.

Lund, J. and Rudemo, M. (2000), "Models for Point Processes Observed with Noise," *Biometrika*, 87, 235–249.

Mei, H., Qin, G., and Eisner, J. (2019), "Imputing Missing Events in Continuous-Time Event Streams," in Chaudhuri, K. and Salakhutdinov, R. (editors), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, PMLR.

Mugglestone, M. and Renshaw, E. (1996), "A practical guide to the spectral analysis of spatial point processes," *Computational Statistics & Data Analysis*, 21, 43–65.

Mugglestone, M. A. and Renshaw, E. (2001), "Spectral tests of randomness for spatial point patterns," *Environmental and Ecological Statistics*, 8, 237–251.

Ogata, Y. (1978), "The asymptotic behaviour of maximum likelihood estimators for stationary point processes," *Annals of the Institute of Statistical Mathematics*, 30, 243–261.

— (1981), "On Lewis' simulation method for point processes," *IEEE Transactions on Information Theory*, 27, 23–30.

— (1988), "Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes," *Journal of the American Statistical Association*, 83, 9–27.

— (1998), "Space-Time Point-Process Models for Earthquake Occurrences," *Annals of the Institute of Statistical Mathematics*, 50, 379–402.

Olinde, J. and Short, M. (2020), "A Self-limiting Hawkes Process: Interpretation, Estimation, and Use in Crime Modeling," in *2020 IEEE International Conference on Big Data (Big Data)*.

Ozaki, T. (1979), "Maximum likelihood estimation of Hawkes' self-exciting point processes," *Annals of the Institute of Statistical Mathematics*, 31, 145–155.

Pham, D. (1996), "Blind separation of instantaneous mixture of sources via an independent component analysis," *IEEE Transactions on Signal Processing*, 44, 2768–2779.

Pinsky, M. (2008), *Introduction to Fourier Analysis and Wavelets*, Graduate studies in mathematics, American Mathematical Society.

Rajala, T., Olhede, S., Grainger, J., and Murrell, D. (2023), "What is the Fourier Transform of a Spatial Point Process?" *IEEE Transactions on Information Theory*.

Reynaud-Bouret, P., Rivoirard, V., Grammont, F., and Tuleau-Malot, C. (2014), "Goodness-of-Fit Tests and Nonparametric Adaptive Estimation for Spike Train Analysis," *The Journal of Mathematical Neuroscience*, 4, 3.

Trouleau, W., Etesami, J., Grossglauser, M., Kiyavash, N., and Thiran, P. (2019), "Learning Hawkes Processes Under Synchronization Noise," in Chaudhuri, K. and Salakhutdinov, R. (editors), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, PMLR.

Tuan, P. (1981), "Estimation of the Spectral Parameters of a Stationary Point Process," *The Annals of Statistics*, 9, 615 – 627.

Villani, M., Quiroz, M., Kohn, R., and Salomone, R. (2022), "Spectral Subsampling MCMC for Stationary Multivariate Time Series with Applications to Vector ARTFIMA Processes," *Econometrics and Statistics*.

Whittle, P. (1952), "Some results in time series analysis," *Scandinavian Actuarial Journal*, 1952, 48–60.

A Proof of Proposition 2.2

Lemma A.1. *Let X and Y be two independent and stationary multivariate point processes with same dimension d. If they admit second order moment measures, denoted respectively $(M^X_{ij})_{1\leq i,j\leq d}$ and $(M^Y_{ij})_{1\leq i,j\leq d}$, then the process $N = X + Y$ also admits a second order moment measure, noted $(M^N_{ij})_{1\leq i,j\leq d}$, and for any pair $(A, B) \in (\mathcal{B}^c_{\mathbb{R}})^2$ and all $1 \leq i, j \leq d$:*

$$M^N_{ij}(A,B) = M^X_{ij}(A,B) + M^Y_{ij}(A,B) + (m^X_i m^Y_j + m^X_j m^Y_i)\ell_{\mathbb{R}}(A)\ell_{\mathbb{R}}(B)\,, \tag{15}$$

where for all $i \in \{1,\dots,d\}$ and $m^X_i = \mathbb{E}[X_i([0,1))]$ and $m^Y_i = \mathbb{E}[Y_i([0,1))]$. Furthermore, the reduced measure of N reads:

$$\breve{M}^N_{ij}(B) = \breve{M}^X_{ij}(B) + \breve{M}^Y_{ij}(B) + (m^X_i m^Y_j + m^X_j m^Y_i)\ell_{\mathbb{R}}(B)\,. \tag{16}$$

Proof. Let $(A, B) \in (\mathcal{B}^c_{\mathbb{R}})^2$. Then for any $1 \leq i, j \leq d$:

$$\begin{aligned}
M^N_{ij}(A,B) &= \mathbb{E}[N_i(A)N_j(B)] \\
&= \mathbb{E}[(X_i(A) + Y_i(A))(X_j(B) + Y_j(B))] \\
&= \mathbb{E}[X_i(A)X_j(B)] + \mathbb{E}[X_i(A)Y_j(B)] + \mathbb{E}[X_j(A)Y_i(B)] + \mathbb{E}[Y_j(A)Y_j(B)] \\
&= M^X_{ij}(A,B) + \mathbb{E}[X_i(A)]\mathbb{E}[Y_j(B)] + \mathbb{E}[X_j(A)]\mathbb{E}[Y_i(B)] + M^Y_{ij}(A,B) \\
&= M^X_{ij}(A,B) + m^X_i m^Y_j \ell(A)\ell(B) + m^X_j m^Y_i \ell(A)\ell(B) + M^Y_{ij}(A,B)\,,
\end{aligned}$$

where the last line comes from the stationarity, which implies that $\mathbb{E}[X_i(A)] = m^X_i \ell(A)$.

By applying Equation (5) to the function $g(x,y) = \mathbb{1}_{x\in[0,1]}\mathbb{1}_{y-x\in B}$ for any $B \in \mathcal{B}^c_{\mathbb{R}}$, we can remark that:

$$\int_{\mathbb{R}^2} g(x,y)\, M^N_{ij}(\mathrm{d}x,\mathrm{d}y) = \int_{[0,1]} \ell_{\mathbb{R}}(\mathrm{d}x) \int_B \breve{M}^N_{ij}(\mathrm{d}u) = \breve{M}^N_{ij}(B)\,.$$

In particular this equality is also true if we replace N by X and Y. By leveraging Equation (15) on the left-side integral, we obtain:

$$\begin{aligned}\int_{\mathbb{R}^2} g(x,y)\, M_{ij}^N(\mathrm{d}x,\mathrm{d}y) &= \breve{M}_{ij}^X(B) + \breve{M}_{ij}^Y(B) + (m_i^X m_j^Y + m_j^X m_i^Y) \int_{x\in[0,1]} \int_{y\in B+x} \ell_{\mathbb{R}}(\mathrm{d}y)\, \ell_{\mathbb{R}}(\mathrm{d}x) \\ &= \breve{M}_{ij}^X(B) + \breve{M}_{ij}^Y(B) + (m_i^X m_j^Y + m_j^X m_i^Y)\ell_{\mathbb{R}}([0,1])\ell_{\mathbb{R}}(B) \\ &= \breve{M}_{ij}^X(B) + \breve{M}_{ij}^Y(B) + (m_i^X m_j^Y + m_j^X m_i^Y)\ell_{\mathbb{R}}(B)\,,\end{aligned}$$

which achieves the proof.

□

Proof of Proposition 2.2

By definition of the spectral density, for all $\omega \in \mathbb{R}$:

$$\begin{aligned}f_{ij}^N(\omega) &= \int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \breve{M}_{ij}^N(\mathrm{d}x) - m_i^N m_j^N \delta(\omega) \\ &= \int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \breve{M}_{ij}^X(\mathrm{d}x) + \int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \breve{M}_{ij}^Y(\mathrm{d}x) \\ &\quad + (m_i^X m_j^Y + m_j^X m_i^Y) \int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \ell_{\mathbb{R}}(\mathrm{d}x) - (m_i^X + m_i^Y)(m_j^X + m_j^Y)\delta(\omega) \\ &= \int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \breve{M}_{ij}^X(\mathrm{d}x) - m_i^X m_j^X \delta(\omega) + \int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \breve{M}_{ij}^Y(\mathrm{d}x) - m_i^Y m_j^Y \delta(\omega) \\ &\quad + (m_i^X m_j^Y + m_j^X m_i^Y) \int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \ell_{\mathbb{R}}(\mathrm{d}x) - (m_i^X m_j^Y + m_i^Y m_j^X)\delta(\omega) \\ &= f_{ij}^X(\omega) + f_{ij}^Y(\omega) + (m_i^X m_j^Y + m_j^X m_i^Y) \int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \mathrm{d}x - (m_i^X m_j^Y + m_i^Y m_j^X)\delta(\omega)\,. \qquad (17)\end{aligned}$$

By properties of the Dirac measure, for all $\omega \in \mathbb{R}$:

$$\int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega} \delta(x)\, \mathrm{d}x = 1\,,$$

and so by duality of the Fourier transform (Pinsky, 2008, Proposition 5.2.4.) it follows that:

$$\int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \mathrm{d}x = \delta(\omega)\,.$$

The last two terms of Equation (17) being equal, we obtain $f_{ij}^N = f_{ij}^X + f_{ij}^Y$.

B Proof of Proposition 3.2

We first show that equality of two spectral densities with different parameters is equivalent to a system of equations (Equations (18)). This result will be used then to prove both parts of Proposition 3.2.

Let $(\mu, \alpha, \beta, \lambda_0)$ and $(\mu', \alpha', \beta', \lambda_0')$ be two admissible 4-tuples for the exponential noisy Hawkes model. Let us assume that for all $\omega \in \mathbb{R}$:

$$f^N_{(\mu,\alpha,\beta,\lambda_0)}(\omega) = f^N_{(\mu',\alpha',\beta',\lambda_0')}(\omega).$$

Thanks to Equation (14), this equality implies the following system of equations:

$$\begin{cases} \dfrac{\mu}{1-\alpha} + \lambda_0 = \dfrac{\mu'}{1-\alpha'} + \lambda_0' & (\omega \to +\infty) \\ \dfrac{\mu}{1-\alpha}\dfrac{\beta^2\alpha(2-\alpha)}{\beta^2(1-\alpha)^2} + \dfrac{\mu}{1-\alpha} + \lambda_0 = \dfrac{\mu'}{1-\alpha'}\dfrac{\beta'^2\alpha'(2-\alpha')}{\beta'^2(1-\alpha')^2} + \dfrac{\mu'}{1-\alpha'} + \lambda_0' & (\omega = 0) \\ \dfrac{\mu}{1-\alpha}\dfrac{\beta^2\alpha(2-\alpha)}{\beta^2(1-\alpha)^2+4\pi^2} + \dfrac{\mu}{1-\alpha} + \lambda_0 = \dfrac{\mu'}{1-\alpha'}\dfrac{\beta'^2\alpha'(2-\alpha')}{\beta'^2(1-\alpha')^2+4\pi^2} + \dfrac{\mu'}{1-\alpha'} + \lambda_0' & (\omega = 1). \end{cases}$$

The first equality can be used to simplify the second and third equalities, leading to:

$$\begin{cases} \dfrac{\mu}{1-\alpha} + \lambda_0 = \dfrac{\mu'}{1-\alpha'} + \lambda_0' \\ \dfrac{\mu}{1-\alpha}\dfrac{\alpha(2-\alpha)}{(1-\alpha)^2} = \dfrac{\mu'}{1-\alpha'}\dfrac{\alpha'(2-\alpha')}{(1-\alpha')^2} \\ \dfrac{\mu}{1-\alpha}\dfrac{\beta^2\alpha(2-\alpha)}{\beta^2(1-\alpha)^2+4\pi^2} = \dfrac{\mu'}{1-\alpha'}\dfrac{\beta'^2\alpha'(2-\alpha')}{\beta'^2(1-\alpha')^2+4\pi^2} \end{cases}.$$

Now, given that $\alpha \in (0,1)$ (same for α'), the last two equalities lead to:

$$\frac{\beta^2(1-\alpha)^2}{\beta^2(1-\alpha)^2+4\pi^2} = \frac{\beta'^2(1-\alpha')^2}{\beta'^2(1-\alpha')^2+4\pi^2},$$

which in turn implies $\beta(1-\alpha) = \beta'(1-\alpha')$.

Thus, it results the following system of equations:

$$\begin{cases} \dfrac{\mu}{1-\alpha} + \lambda_0 = \dfrac{\mu'}{1-\alpha'} + \lambda_0' & \text{(18a)} \\ \dfrac{\mu\alpha(2-\alpha)}{(1-\alpha)^3} = \dfrac{\mu'\alpha'(2-\alpha')}{(1-\alpha')^3} & \text{(18b)} \\ \beta(1-\alpha) = \beta'(1-\alpha'). & \text{(18c)} \end{cases}$$

Now, Equations (18) lead to

$$\begin{cases} \dfrac{\mu}{1-\alpha} + \lambda_0 = \dfrac{\mu'}{1-\alpha'} + \lambda_0' \\ \dfrac{\mu}{1-\alpha}\beta^2\alpha(2-\alpha) = \dfrac{\mu'}{1-\alpha'}\beta'^2\alpha'(2-\alpha') \\ \beta(1-\alpha) = \beta'(1-\alpha'), \end{cases}$$

which, by Equation (14), implies straightforwardly $f^N_{(\mu,\alpha,\beta,\lambda_0)} = f^N_{(\mu',\alpha',\beta',\lambda_0')}$. Consequently,

$$f^N_{(\mu,\alpha,\beta,\lambda_0)} = f^N_{(\mu',\alpha',\beta',\lambda_0')} \iff \text{Equations (18)}.$$

The model $\mathcal{Q}$ is not identifiable.

Let $\tau > -\lambda_0$ and $\lambda_0' = \lambda_0 + \tau > 0$.

Then, by denoting $\kappa = \frac{\mu}{1-\alpha}\beta^2\alpha(2-\alpha)$, Equations (18) are equivalent to:

$$\begin{cases} \mu' = \dfrac{\beta(1-\alpha)(\frac{\mu}{1-\alpha}-\tau)}{\sqrt{\beta^2(1-\alpha)^2+\frac{\kappa}{\frac{\mu}{1-\alpha}-\tau}}} \\ \alpha' = 1 - \dfrac{\beta(1-\alpha)}{\sqrt{\beta^2(1-\alpha)^2+\frac{\kappa}{\frac{\mu}{1-\alpha}-\tau}}} \\ \beta' = \sqrt{\beta^2(1-\alpha)^2 + \frac{\kappa}{\frac{\mu}{1-\alpha}-\tau}}. \end{cases} \tag{19}$$

From (19), it is clear that for all $\tau \in \left(-\lambda_0, \frac{\mu}{1-\alpha}\right) \setminus \{0\}$, $(\mu', \alpha', \beta', \lambda_0+\tau) \neq (\mu, \alpha, \beta, \lambda_0)$ is an admissible parameter for $\mathcal{Q}$ and $f^N_{(\mu',\alpha',\beta',\lambda_0+\tau)} = f^N_{(\mu,\alpha,\beta,\lambda_0)}$. Consequently, $\mathcal{Q}$ is not identifiable.

The reduced model defined by a triplet of admissible parameters is identifiable.

It will be shown that, for admissible parameters, $f^N_{(\mu,\alpha,\beta,\lambda_0)} = f^N_{(\mu',\alpha',\beta',\lambda_0')}$ implies

$$\mu = \mu' \iff \alpha = \alpha' \iff \beta = \beta' \iff \lambda_0 = \lambda_0',$$

from which we can deduce identifiability of the four collections of models mentioned in Proposition 3.2. Indeed, let, for instance, $\alpha^\circ \in (0,1)$ and consider

$$\mathcal{Q}_\alpha = \left\{ f^N_{(\mu,\alpha,\beta,\lambda_0)} \in \mathcal{Q}, \alpha = \alpha^\circ \right\} .$$

Then for all $f^N_{(\mu,\alpha,\beta,\lambda_0)} \in \mathcal{Q}_\alpha$ and $f^N_{(\mu',\alpha',\beta',\lambda_0')} \in \mathcal{Q}_\alpha$,

$$f^N_{(\mu,\alpha,\beta,\lambda_0)} = f^N_{(\mu',\alpha',\beta',\lambda_0')} \implies \begin{cases} \mu = \mu' \iff \alpha = \alpha' \iff \beta = \beta' \iff \lambda_0 = \lambda_0' \\ \alpha = \alpha^\circ = \alpha' \end{cases} \implies \begin{cases} \mu = \mu' \\ \beta = \beta' \\ \lambda_0 = \lambda_0' \\ \alpha = \alpha' . \end{cases}$$

So, let us assume that $f_{(\mu,\alpha,\beta,\lambda_0)} = f_{(\mu',\alpha',\beta',\lambda_0')}$, for some admissible parameters. As shown beforehand this implies Equations (18). From Equation (18c), we establish that

$$\alpha = \alpha' \iff \beta = \beta' ,$$

since, $\beta > 0$ and $\alpha < 1$.

From Equation (18b), since $\alpha > 0$, it is clear that $\alpha = \alpha' \implies \mu = \mu'$. Conversely, if $\mu = \mu' > 0$, Equation (18b) becomes:

$$g(\alpha) = \frac{\alpha(2-\alpha)}{(1-\alpha)^3} = \frac{\alpha'(2-\alpha')}{(1-\alpha')^3} = g(\alpha') .$$

In particular, g is a strictly increasing function on $(0,1)$ and so it can be deduced that $\alpha = \alpha'$. So

$$\mu = \mu' \iff \alpha = \alpha' .$$

Finally, as $\mu = \mu' \iff \alpha = \alpha'$, from Equation (18a) we conclude that $\alpha = \alpha' \implies \lambda_0 = \lambda_0'$. Let us now assume that $\lambda_0 = \lambda_0'$, Equation (18a) then reads

$$\frac{\mu}{1-\alpha} = \frac{\mu'}{1-\alpha'} .$$

As $\mu > 0$ and $\mu' > 0$, Equation (18b) can be then reduced to

$$s(\alpha) := \frac{\alpha(2-\alpha)}{(1-\alpha)^2} = \frac{\alpha(2-\alpha)}{(1-\alpha)^2} = s(\alpha') ,$$

where s is a strictly increasing function on $(0,1)$ and so $\alpha = \alpha'$. With this we have proved that $\alpha = \alpha' \iff \lambda_0 = \lambda_0'$ which achieves the proof.

C Proof of Proposition 3.3

Let $(\mathcal{H}^N_t)_{t\geq 0}$ (respectively $(\mathcal{H}^H_t)_{t\geq 0}$) be the natural filtration associated with $(\lambda^N(t))_{t\geq 0}$ (respectively $(\lambda^H(t))_{t\geq 0}$). Let us now consider the conditional survival function of the first non-negative jump τ_1 of N given the past (before 0): for all $t \geq 0$,

$$\begin{aligned} \mathbb{P}\left(\tau_1 > t \mid \mathcal{H}^N_0\right) &= \mathbb{P}\left(N(t) = 0 \mid \mathcal{H}^N_0\right) \\ &= \mathbb{P}\left(P(t) = 0, H(t) = 0 \mid \mathcal{H}^N_0\right) \\ &= \mathbb{P}\left(P(t) = 0 \mid \mathcal{H}^N_0\right) \mathbb{P}\left(H(t) = 0 \mid \mathcal{H}^N_0\right) && \text{(by independence)} \\ &= \mathbb{P}\left(P(t) = 0\right) \mathbb{P}\left(H(t) = 0 \mid \mathcal{H}^H_0\right) && \text{(by definition)} \\ &= \mathrm{e}^{-\lambda_0 t} e^{-\mu(t-u_t) - u_t \lambda^H(0)} , \end{aligned}$$

with $u_t = (1 - \mathrm{e}^{-\beta t})\beta^{-1}$ and where we have used that, given that $H(t) = 0$, for all $u \in [0,t]$,

$$\lambda^H(u) = \mu + \left(\lambda^H(0) - \mu\right) \mathrm{e}^{-\beta u}.$$

Let us remark that the last equality can also be deduced from (Dassios and Zhao, 2011, Corollary 3.3) with the same notation u_t:

$$\mathbb{P}\left(H(t)=0 \mid \mathcal{H}_0^H\right) = \exp\left(-\int_0^{u_t} \frac{\mu\beta v}{1-\beta v}\,\mathrm{d}v\right) \mathrm{e}^{-u_t\lambda^H(0)}\,.$$

It appears that $\mathbb{P}\left(\tau_1 > t \mid \mathcal{H}_0^N\right)$ depends on the past only through $\lambda^H(0)$. But $\lambda^H(0) = \lambda^N(0) - \lambda_0$ and $\lambda^N(0)$ is distributed according to the stationary distribution of $\left(\lambda^N(t)\right)_{t\geq 0}$. It results that $\lambda^H(0)$ is distributed according to the stationary distribution of $\left(\lambda^H(t)\right)_{t\geq 0}$, the Laplace transform of which is given by (Dassios and Zhao, 2011, Corollary 3.1). Plugging in this result, we have:

$$\begin{aligned}\mathbb{P}\left(\tau_1 > t\right) &= \mathrm{e}^{-\lambda_0 t-\mu(t-u_t)}\,\mathbb{E}\left[\mathrm{e}^{-u_t\lambda^H(0)}\right] \\ &= \mathrm{e}^{-\lambda_0 t-\mu(t-u_t)} \exp\left(-\int_0^{u_t} \frac{\mu\beta v}{\beta v + \mathrm{e}^{-\alpha\beta v}-1}\mathrm{d}v\right)\,.\end{aligned}$$

For ease of derivation, let $G : t \in \mathbb{R}_{\geq 0} \mapsto -\log \mathbb{P}\left(\tau_1 > t\right)$. The function G is differentiable and has derivative D given by:

$$D(t) = \lambda_0 + \mu\left(1-\mathrm{e}^{-\beta t}\right) + \frac{\mu\left(1-\mathrm{e}^{-\beta t}\right)\mathrm{e}^{-\beta t}}{\exp\left(-\alpha\left(1-\mathrm{e}^{-\beta t}\right)\right)-\mathrm{e}^{-\beta t}}\,.$$

Now define in the same manner τ_1', G', and D' from the process N' with parameter $(\mu', \alpha', \beta', \lambda_0')$, with $\lambda^{N'}(0)$ being distributed according to the stationary distribution of $\left(\lambda^{N'}(t)\right)_{t\geq 0}$, and assume that N and N' have the same distribution. Then it follows that both τ_1 and τ_1' have the same distribution, so that $G(t) = G'(t)$ for all $t \geq 0$. Since G and G' are everywhere differentiable and have the same initial value $G(0) = G'(0) = 0$, it results that $D(t) = D'(t)$ for all $t \geq 0$.

We want to establish a system of four equations satisfied by the parameters that leads to the equality of the 4-tuples. First, noting that $\lim_{t\to\infty} D(t) = \lambda_0 + \mu$, we get

$$\lambda_0 + \mu = \lambda_0' + \mu'\,. \tag{20}$$

Then, since $\lim_{t\to 0} D(t) = \lambda_0 + \mu/(1-\alpha)$, we get, using Equation (20),

$$\frac{\mu\alpha}{1-\alpha} = \frac{\mu'\alpha'}{1-\alpha'}\,. \tag{21}$$

To highlight two other equations on the parameters, we establish the Taylor expansion of $D(t)$ around $t = 0$ up to order 2. After some calculation, we find that

$$D(t) = \lambda_0 + \frac{\mu}{1-\alpha} + \frac{\mu\alpha}{1-\alpha}\left(\frac{1}{2}\beta t\frac{\alpha-2}{1-\alpha} + \frac{1}{12}\beta^2 t^2\frac{\alpha^3-\alpha^2-3\alpha+6}{(1-\alpha)^2} + o\left(t^2\right)\right)\,.$$

From the first-order term of the expansion, Equation (21) and the equality of D and D', we find that

$$\beta\frac{\alpha-2}{1-\alpha} = \beta'\frac{\alpha'-2}{1-\alpha'}\,. \tag{22}$$

Then the second-order term of the expansion can be rewritten

$$\frac{\mu\alpha}{1-\alpha}\frac{1}{12}\left(\frac{\beta(\alpha-2)}{1-\alpha}\right)^2\frac{\alpha^3-\alpha^2-3\alpha+6}{(\alpha-2)^2}\,,$$

so that from Equations (21) and (22), we find that

$$g(\alpha) = \frac{\alpha^3-\alpha^2-3\alpha+6}{(\alpha-2)^2} = \frac{\alpha'^3-\alpha'^2-3\alpha'+6}{(\alpha'-2)^2} = g(\alpha')\,. \tag{23}$$

$\alpha \mapsto g(\alpha)$ can be shown to be strictly increasing for $\alpha \in (0,1)$, so that Equation (23) yields that $\alpha = \alpha'$. With this remark, one can easily show from the system composed by Equations (20)–(23) that the tuples $(\mu, \alpha, \beta, \lambda_0)$ and $(\mu', \alpha', \beta', \lambda_0')$ are equal.

D Proof of Proposition 3.4

Lemma D.1. *Assume all moments of h are finite: $\forall n \geq 0, m_n = \int t^n h(t)\,\mathrm{d}t < \infty$. Then the spectral density f^N of the noisy Hawkes process N has a Taylor expansion around 0 given by:*

$$\forall t \in \mathbb{R}: \quad f^N(t) = \frac{\mu}{(1-\alpha)^3} + \frac{\mu}{(1-\alpha)^3} \sum_{q \geq 1} \left(\frac{\alpha}{1-\alpha} \sum_{n \geq 1} b_n t^{2n} \right)^q + \lambda_0 \,, \tag{24}$$

with

$$b_n = \frac{(-1)^n (2\pi)^{2n} a_n}{(2n)!} \,, \qquad \text{and} \qquad a_n = 2m_{2n} - \frac{\alpha}{1-\alpha} \sum_{k=1}^{2n-1} \binom{2n}{k} (-1)^k m_k m_{2n-k} \,. \tag{25}$$

Proof. Introduce the Taylor expansion of the Fourier transform of h:

$$\forall t \in \mathbb{R}: \quad \tilde{h}(t) = \sum_{n \geq 0} (\mathrm{i}\tau)^n \frac{m_n}{n!},$$

where $\tau = -2\pi t$. Then, given that $m_0 = 1$,

$$\begin{aligned}
\left| 1 - \alpha \tilde{h}(t) \right|^2 &= \left(1 - \alpha \sum_{n \geq 0} (\mathrm{i}\tau)^n \frac{m_n}{n!} \right) \left(1 - \alpha \sum_{n \geq 0} (-\mathrm{i}\tau)^n \frac{m_n}{n!} \right) \\
&= 1 - 2\alpha \sum_{n \geq 0} (-1)^n \tau^{2n} \frac{m_{2n}}{(2n)!} + \alpha^2 \sum_{n \geq 0} \sum_{k=0}^{n} (-1)^k (\mathrm{i}\tau)^n \frac{m_{n-k} m_k}{(n-k)!k!} \\
&= 1 - \alpha \sum_{n \geq 0} 2m_{2n} \frac{(-1)^n \tau^{2n}}{(2n)!} + \alpha^2 \sum_{n \geq 0} \sum_{k=0}^{2n} (-1)^k m_k m_{2n-k} \binom{2n}{k} \frac{(-1)^n \tau^{2n}}{(2n)!} \\
&= (1-\alpha)^2 \left(1 - \frac{\alpha}{1-\alpha} \sum_{n \geq 1} a_n \frac{(-1)^n \tau^{2n}}{(2n)!} \right).
\end{aligned}$$

Inverting this expression and taking the Taylor expansion of $x \mapsto (1-x)^{-1}$ around 0 yields the desired result. □

Remark D.1. *For $n = 1, 2$, a_n is given by*

$$\begin{cases} a_1 &= 2m_2 + \frac{\alpha}{1-\alpha} 2m_1^2 \\ a_2 &= 2m_4 + \frac{\alpha}{1-\alpha} \left(8m_1 m_3 - 6m_2^2 \right), \end{cases}$$

so that the Taylor expansion of f^N up to order 5 is:

$$f^N(t) = a + c_1 t^2 + c_2 t^4 + o(t^5) \,,$$

with

$$\begin{aligned}
a &= \frac{\mu}{(1-\alpha)^3} + \lambda_0 \\
c_1 &= 4 \frac{\mu \alpha \pi^2}{(1-\alpha)^4} \left[-m_2 - \frac{\alpha}{1-\alpha} m_1^2 \right] \\
c_2 &= 16 \frac{\mu \alpha \pi^4}{(1-\alpha)^4} \left[\frac{m_4}{12} + \frac{\alpha}{1-\alpha} \left(\frac{m_1 m_3}{3} + \frac{3m_2^2}{4} \right) + \left(\frac{\alpha}{1-\alpha} \right)^2 2m_2 m_1^2 + \left(\frac{\alpha}{1-\alpha} \right)^3 m_1^4 \right].
\end{aligned}$$

Proof of Proposition 3.4

Let $(\mu, \alpha, \phi, \lambda_0)$ and $(\mu', \alpha', \phi', \lambda_0')$ be two admissible 4-tuples for the uniform noisy Hawkes model $\mathcal{R}$, and assume that, for all $\omega \in \mathbb{R}$,

$$f^N_{(\mu,\alpha,\phi,\lambda_0)}(\omega) = f^N_{(\mu',\alpha',\phi',\lambda_0')}(\omega)\,.$$

First, noting that $\lim_{\omega\to\infty} f^N_{(\mu,\alpha,\phi,\lambda_0)}(\omega) = \lambda_0 + \mu/(1-\alpha)$, we get

$$\lambda_0 + \frac{\mu}{1-\alpha} = \lambda_0' + \frac{\mu'}{1-\alpha'}\,. \tag{26}$$

Then, since $f^N_{(\mu,\alpha,\phi,\lambda_0)}(0) = \lambda_0 + \mu/(1-\alpha)^3$ and using Equation (26), we get:

$$\frac{\mu\alpha(2-\alpha)}{(1-\alpha)^3} = \frac{\mu'\alpha'(2-\alpha')}{(1-\alpha')^3}\,. \tag{27}$$

Now, since the Taylor expansions, given by Equation (24), of $f^N_{(\mu,\alpha,\phi,\lambda_0)}$ and $f^N_{(\mu',\alpha',\phi',\lambda_0')}$ around 0 coincide, their respective Taylor coefficients $(c_1, c_2, \ldots)$ and $(c_1', c_2', \ldots)$ are equal. Plugging in the moments of the uniform distribution on $[0, \phi]$,

$$m_n = \frac{\phi^n}{n+1}\,, \qquad \text{for all } n \geq 0\,,$$

the first order coefficient c_1 can be written

$$c_1 = -\frac{\pi}{3}\frac{\mu\alpha(2-\alpha)}{(1-\alpha)^3}\phi^2\frac{4-\alpha}{(1-\alpha)^2(2-\alpha)}\,,$$

so that, with the use of Equation (27), we get

$$\phi^2\frac{4-\alpha}{(1-\alpha)^2(2-\alpha)} = \phi'^2\frac{4-\alpha'}{(1-\alpha')^2(2-\alpha')}\,. \tag{28}$$

Similarly, by plugging the moments in the second order coefficient c_2, we get

$$c_2 = \frac{\pi^4}{15}\frac{\mu\alpha(2-\alpha)}{(1-\alpha)^3}\left(\phi^2\frac{4-\alpha}{(1-\alpha)^2(2-\alpha)}\right)^2\frac{(2-\alpha)(\alpha^3-8\alpha^2+18\alpha+4)}{(4-\alpha)^2}\,,$$

so that, by Equations (27) and (28),

$$g(\alpha) = \frac{(2-\alpha)(\alpha^3-8\alpha^2+18\alpha+4)}{(4-\alpha)^2} = \frac{(2-\alpha')(\alpha'^3-8\alpha'^2+18\alpha'+4)}{(4-\alpha')^2} = g(\alpha')\,. \tag{29}$$

But $\alpha \mapsto g(\alpha)$ can be shown to be strictly increasing for $\alpha \in (0,1)$, so that Equation (29) yields that $\alpha = \alpha'$. Finally, it is easily proven from the system composed by Equations (26)–(29) that the tuples $(\mu, \alpha, \beta, \lambda_0)$ and $(\mu', \alpha', \beta', \lambda_0')$ are equal.

E Proof of Proposition 4.2

Let N be a bivariate noisy Hawkes process parametrised by the exponential model $\mathcal{Q}_\Lambda$. We will prove that if either of the conditions in Proposition 4.2 is fulfilled then $\mathcal{Q}_\Lambda$ is not identifiable.

- **Condition 1:** Let

$$\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & 0 \\ 0 & \alpha_{22} \end{pmatrix}, 0 \leq \alpha_{11}, \alpha_{22} < 1 \right\}\,.$$

  For any admissible $\theta = (\mu, \alpha, \beta, \lambda_0)$, the spectral matrix $\mathbf{f}^N_\theta$ is diagonal such that, for any integer $i \in \{1, 2\}$:

$$f_\theta^N{}_{ii}(\omega) = \frac{\mu_i}{1-\alpha_{ii}}\left(\frac{\beta_i^2\alpha_{ii}(2-\alpha_{ii})}{\beta_i^2(1-\alpha_{ii})^2+4\pi^2\omega^2}\right)+\left(\frac{\mu_i}{1-\alpha_{ii}}+\lambda_0\right).$$

We will show that there exists an admissible $\theta' = (\mu', \alpha', \beta', \lambda_0')$ such that $\theta' \neq \theta$ and $\mathbf{f}_\theta^N = \mathbf{f}_{\theta'}^N$. Let $\tau > -\lambda_0$ and $\lambda_0' = \lambda_0 + \tau$.

For every $i \in \{1,2\}$, on the one hand, if $\alpha_{ii} \neq 0$ then, as shown in the univariate case (Appendix B), letting $\kappa_i = \frac{\mu_i}{1-\alpha_{ii}}\beta_i^2\alpha_{ii}(2-\alpha_{ii})$ and defining the parameters:

$$\begin{cases} \mu_i' = \frac{\beta_i(1-\alpha_{ii})(\frac{\mu_i}{1-\alpha_{ii}}-\tau)}{\sqrt{\beta_i^2(1-\alpha_{ii})^2+\frac{\kappa_i}{\frac{\mu_i}{1-\alpha_{ii}}-\tau}}} \\ \alpha_{ii}' = 1-\frac{\beta_i(1-\alpha_{ii})}{\sqrt{\beta_i^2(1-\alpha_{ii})^2+\frac{\kappa_i}{\frac{\mu_i}{1-\alpha_{ii}}-\tau}}} \\ \beta_i' = \sqrt{\beta_i^2(1-\alpha_{ii})^2+\frac{\kappa_i}{\frac{\mu_i}{1-\alpha_{ii}}-\tau}}\,, \end{cases} \tag{30}$$

leads to $f_\theta^N{}_{ii} = f_{\theta'}^N{}_{ii}$.

On the other hand, if $\alpha_{ii} = 0$ then, $f_\theta^N{}_{ii} = \mu_i + \lambda_0$ and it is enough to consider:

$$\begin{cases} \mu_i' = \mu_i - \tau \\ \alpha_{ii}' = 0 \\ \beta_i' = \beta_i\,, \end{cases}$$

to get $f_\theta^N{}_{ii} = f_{\theta'}^N{}_{ii}$.

In both cases, for $\tau \in \left(-\lambda_0, \min_{1\leq i\leq 2}\left\{\frac{\mu_i}{1-\alpha_{ii}}\right\}\right)\setminus\{0\}$, we obtain an admissible parameter $\theta' \neq \theta$ and $\mathbf{f}_\theta^N = \mathbf{f}_{\theta'}^N$. Thus, the model is not identifiable.

- **Condition 2 and 3:** Without loss of generality, let

$$\Lambda = \left\{\begin{pmatrix}\alpha_{11} & \alpha_{12}\\ 0 & 0\end{pmatrix}, 0<\alpha_{11}<1, \alpha_{12}>0\right\}.$$

For any admissible parameter $\theta = (\mu, \alpha, \beta, \lambda_0)$ and for any $\omega \in \mathbb{R}$, the spectral matrix reads:

$$\mathbf{f}_\theta^N(\omega) = \begin{bmatrix} f_\theta^N{}_{11}(\omega) & f_\theta^N{}_{12}(\omega) \\ f_\theta^N{}_{21}(\omega) & f_\theta^N{}_{22}(\omega)\end{bmatrix} = \begin{bmatrix} f_\theta^N{}_{11}(\omega) & \frac{\mu_2\beta_1\alpha_{12}}{\beta_1(1-\alpha_{11})-2\pi\mathrm{i}\omega} \\ \frac{\mu_2\beta_1\alpha_{12}}{\beta_1(1-\alpha_{11})+2\pi\mathrm{i}\omega} & \mu_2+\lambda_0\end{bmatrix}$$

with

$$\begin{aligned} f_\theta^N{}_{11}(\omega) = &\left(\frac{\mu_1}{1-\alpha_{11}}+\frac{\mu_2\alpha_{12}}{1-\alpha_{11}}\right)\frac{\beta_1^2\alpha_{11}(2-\alpha_{11})}{\beta_1^2(1-\alpha_{11})^2+4\pi^2\omega^2}+\frac{\mu_2\beta_1^2\alpha_{12}^2}{\beta_1^2(1-\alpha_{11})^2+4\pi^2\omega^2} \\ &+\frac{\mu_1}{1-\alpha_{11}}+\frac{\mu_2\alpha_{12}}{1-\alpha_{11}}+\lambda_0\,. \end{aligned}$$

Let us introduce the following constants:

$$\begin{cases} A = \mu_2+\lambda_0 \\ B = \mu_2\beta_1\alpha_{12} \\ C = \beta_1(1-\alpha_{11}) \\ D = \frac{\mu_1+\mu_2\alpha_{12}}{1-\alpha_{11}}+\lambda_0 \\ E = (\frac{\mu_1+\mu_2\alpha_{12}}{1-\alpha_{11}})\beta_1^2\alpha_{11}(2-\alpha_{11})+\mu_2\beta_1^2\alpha_{12}^2\,, \end{cases}$$

which allow us to rewrite the spectral matrix as:

$$\mathbf{f}_\theta^N(\omega) = \begin{bmatrix} \dfrac{E}{C^2+4\pi^2\omega^2} + D & \dfrac{B}{C-2\pi\mathrm{i}\omega} \\ \dfrac{B}{C+2\pi\mathrm{i}\omega} & A \end{bmatrix} . \tag{31}$$

Let $\tau \in \mathbb{R} \setminus \{\mu_2, (\mu_1+\mu_2\alpha_{12})/(1-\alpha_{11})\}$ and:

$$\kappa_\tau = \frac{(\mu_1+\mu_2\alpha_{12})\alpha_{11}(2-\alpha_{11})(\mu_2-\tau) - \tau\mu_2^2\alpha_{12}^2(1-\alpha_{11})}{(\mu_2-\tau)(\mu_1+\mu_2\alpha_{12}-\tau(1-\alpha_{11}))} .$$

Now, consider the parameter $\theta' = (\mu', \alpha', \beta', \lambda_0')$, defined as follows:

$$\begin{cases} \mu_1' = \frac{\mu_1-\tau(1-\alpha_{11})}{\sqrt{(1-\alpha_{11})^2+\kappa_\tau}} \\ \mu_2' = \mu_2 - \tau \\ \alpha_{11}' = 1 - \frac{1-\alpha_{11}}{\sqrt{(1-\alpha_{11})^2+\kappa_\tau}} \\ \alpha_{12}' = \frac{\mu_2\alpha_{12}}{(\mu_2-\tau)\sqrt{(1-\alpha_{11})^2+\kappa_\tau}} \\ \beta_1' = \beta_1\sqrt{(1-\alpha_{11})^2+\kappa_\tau} \\ \lambda_0' = \lambda_0 + \tau . \end{cases} \tag{32}$$

The goal will be to show that there exists τ such that θ' is well defined, admissible for $\mathcal{Q}_\Lambda$, such that $\theta \neq \theta'$ and $\mathbf{f}_\theta^N = \mathbf{f}_{\theta'}^N$.

First, in order for θ' to be an admissible parameter, let us remark that $\rho(S) = \alpha_{11}$ and so we obtain the following constrains:

$$\begin{cases} \mu_1' > 0 \\ \mu_2' > 0 \\ 0 < \alpha_{11}' < 1 \\ \alpha_{12}' > 0 \\ \beta_1' > 0 \\ \lambda_0' > 0 \end{cases} \iff \begin{cases} \mu_1 - \tau(1-\alpha_{11}) > 0 \\ \mu_2 - \tau > 0 \\ \frac{1-\alpha_{11}}{\sqrt{(1-\alpha_{11})^2+\kappa_\tau}} < 1 \\ (1-\alpha_{11})^2 + \kappa_\tau > 0 \\ \lambda_0 + \tau > 0 \end{cases} \iff \begin{cases} \tau < \frac{\mu_1}{1-\alpha_{11}} \\ \tau < \mu_2 \\ \kappa_\tau > 0 \\ \tau > -\lambda_0 . \end{cases} \tag{33}$$

Since $\tau < \mu_1/(1-\alpha_{11}) \implies \mu_1 + \mu_2\alpha_{12} - \tau(1-\alpha_{11}) > 0$, the third inequality becomes

$$\tau < \frac{(\mu_1+\mu_2\alpha_{12})\alpha_{11}(2-\alpha_{11})\mu_2}{(\mu_1+\mu_2\alpha_{12})\alpha_{11}(2-\alpha_{11}) + \mu_2^2(1-\alpha_{11})\alpha_{12}^2} .$$

Then for

$$\tau \in \left(-\lambda_0, \min\left\{\frac{\mu_1}{1-\alpha_{11}}, \mu_2, \frac{(\mu_1+\mu_2\alpha_{12})\alpha_{11}(2-\alpha_{11})\mu_2}{(\mu_1+\mu_2\alpha_{12})\alpha_{11}(2-\alpha_{11}) + \mu_2^2(1-\alpha_{11})\alpha_{12}^2}\right\}\right) \setminus \{0\} ,$$

the right-hand side of Equations (33) is verified and so θ' defined by Equations (32) is well defined, admissible for $\mathcal{Q}_\Lambda$ and $\theta' \neq \theta$.

Then, we can show that:

$$\begin{cases} \mu_2' + \lambda_0' = \mu_2 + \lambda_0 = A \\ \mu_2'\beta_1'\alpha_{12}' = \mu_2\beta_1\alpha_{12} = B \\ \beta_1'(1-\alpha_{11}') = \beta_1(1-\alpha_{11}) = C \\ \frac{\mu_1'+\mu_2'\alpha_{12}'}{1-\alpha_{11}'} + \lambda_0' = \frac{\mu_1+\mu_2\alpha_{12}}{1-\alpha_{11}} + \lambda_0 = D \\ \left(\frac{\mu_1'+\mu_2'\alpha_{12}'}{1-\alpha_{11}'}\right)\beta_1'^2\alpha_{11}'(2-\alpha_{11}') + \mu_2'\beta_1'^2\alpha_{12}'^2 = \left(\frac{\mu_1+\mu_2\alpha_{12}}{1-\alpha_{11}}\right)\beta_1^2\alpha_{11}(2-\alpha_{11}) + \mu_2\beta_1^2\alpha_{12}^2 = E . \end{cases}$$

This assures that, for all $\omega \in \mathbb{R}$:

$$\mathbf{f}^N_{\theta'}(\omega) = \begin{bmatrix} \dfrac{E}{C^2+4\pi^2\omega^2} + D & \dfrac{B}{C-2\pi\mathrm{i}\omega} \\ \dfrac{B}{C+2\pi\mathrm{i}\omega} & A \end{bmatrix} = \mathbf{f}^N_{\theta}(\omega)\,,$$

Thus, the model is not identifiable.

F Proof of Proposition 4.3, Situations 1 and 2

Without loss of generality, let

$$\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{21} & 0 \end{pmatrix}, 0 \leq \alpha_{11} < 1, \alpha_{21} > 0 \right\}.$$

Then, for any admissible parameter $\theta = (\mu, \alpha, \beta, \lambda_0)$ and all $\omega \in \mathbb{R}$:

$$\begin{cases} f^N_{\theta\,11}(\omega) = \dfrac{\mu_1}{1-\alpha_{11}} \dfrac{\beta_1^2 + 4\pi^2\omega^2}{\beta_1^2(1-\alpha_{11})^2 + 4\pi^2\omega^2} + \lambda_0 \\ f^N_{\theta\,12}(\omega) = \dfrac{\mu_1}{1-\alpha_{11}} \dfrac{\beta_1^2 + 4\pi^2\omega^2}{\beta_1^2(1-\alpha_{11})^2 + 4\pi^2\omega^2} \dfrac{\alpha_{21}\beta_2}{\beta_2 + 2\pi\mathrm{i}\omega} \\ f^N_{\theta\,22}(\omega) = \dfrac{\mu_1}{1-\alpha_{11}} \dfrac{\beta_1^2 + 4\pi^2\omega^2}{\beta_1^2(1-\alpha_{11})^2 + 4\pi^2\omega^2} \dfrac{\alpha_{21}^2\beta_2^2}{\beta_2^2 + 4\pi^2\omega^2} + \mu_2 + \dfrac{\mu_1\alpha_{21}}{1-\alpha_{11}} + \lambda_0\,, \end{cases}$$

and $f^N_{\theta\,21}(\omega) = \overline{f^N_{\theta\,12}(\omega)}$.

These expressions can be reformulated as follows:

$$\begin{cases} f^N_{\theta\,11}(\omega) = m_1^H \dfrac{1}{|1-\tilde{h}_{\theta\,11}(\omega)|^2} + \lambda_0 \\ f^N_{\theta\,12}(\omega) = (f^N_{\theta\,11}(\omega) - \lambda_0)\tilde{h}_{\theta\,21}(\omega) \\ f^N_{\theta\,22}(\omega) = \dfrac{|f^N_{\theta\,12}(\omega)|^2}{f^N_{\theta\,11}(\omega) - \lambda_0} + \lim\limits_{\omega' \to +\infty} f^N_{\theta\,22}(\omega')\,. \end{cases}$$

where $m_1^H = \mu_1/(1-\alpha_{11})$, $m_2^H = \mu_2 + \mu_1\alpha_{21}/(1-\alpha_{11})$ and $\lim_{\omega'\to+\infty} f^N_{\theta\,22}(\omega') = m_2^H + \lambda_0$.

Let $\theta' = (\mu', \alpha', \beta', \lambda_0')$ be an admissible parameter such that $\mathbf{f}^N_\theta = \mathbf{f}^N_{\theta'}$. For $\omega = 0$, it comes $f^N_{\theta\,22}(0) = f^N_{\theta'\,22}(0)$, which implies that

$$\frac{|f^N_{\theta\,12}(0)|^2}{f^N_{\theta\,11}(0) - \lambda_0} = \frac{|f^N_{\theta'\,12}(0)|^2}{f^N_{\theta'\,11}(0) - \lambda_0'}\,.$$

Since $|f^N_{\theta\,12}(0)|^2 = \mu_1\alpha_{21}(1-\alpha_{11})^{-3} \neq 0$ (as $\alpha_{11} < 1$ and $\alpha_{21} > 0$), $f^N_{\theta\,12}(0) = f^N_{\theta'\,12}(0)$ and $f^N_{\theta\,11}(0) = f^N_{\theta'\,11}(0)$, it results that:

$$\lambda_0 = \lambda_0'\,.$$

Now, for all $\omega \in \mathbb{R}$, since $f^N_{\theta\,12}(\omega) = f^N_{\theta'\,12}(\omega)$, it comes $(f^N_{\theta\,11}(\omega) - \lambda_0)\tilde{h}_{\theta\,21}(\omega) = (f^N_{\theta'\,11}(\omega) - \lambda_0')\tilde{h}_{\theta',21}(\omega)$, then $\tilde{h}_{\theta\,21}(\omega) = \tilde{h}_{\theta',21}(\omega)$. By Equation (13), it results that:

$$\frac{\alpha_{21}\beta_2}{\beta_2 + 2\pi\mathrm{i}\omega} = \frac{\alpha_{21}'\beta_2'}{\beta_2' + 2\pi\mathrm{i}\omega}\,.$$

For $\omega = 0$, this simplifies to

$$\alpha_{21} = \alpha_{21}'\,.$$

For $\omega = 1$, as $\alpha_{21} \neq 0$ it comes $\beta_2/(\beta_2 + 2\pi\mathrm{i}) = \beta_2'/(\beta_2' + 2\pi\mathrm{i})$, then

$$\beta_2 = \beta_2'\,.$$

Now, by considering the limits as $\omega \to \infty$, we obtain the two following equations:

$$\begin{cases} \dfrac{\mu_1}{1-\alpha_{11}} + \lambda_0 = \dfrac{\mu_1'}{1-\alpha_{11}'} + \lambda_0 \\ \mu_2 + \dfrac{\mu_1\alpha_{21}}{1-\alpha_{11}} + \lambda_0 = \mu_2' + \dfrac{\mu_1'\alpha_{21}}{1-\alpha_{11}'} + \lambda_0 \,, \end{cases}$$

which, using that $\mu_1 > 0$, can be simplified to:

$$\begin{cases} \dfrac{\mu_1}{1-\alpha_{11}} = \dfrac{\mu_1'}{1-\alpha_{11}'} \\ \mu_2 = \mu_2' \,. \end{cases}$$

Now, as ${f_\theta^N}_{11}(0) = \mu_1/(1-\alpha_{11})^3$, we have the two following equations:

$$\begin{cases} \dfrac{\mu_1}{1-\alpha_{11}} = \dfrac{\mu_1'}{1-\alpha_{11}'} \\ \dfrac{\mu_1}{(1-\alpha_{11})^3} = \dfrac{\mu_1'}{(1-\alpha_{11}')^3} \,, \end{cases}$$

which imply

$$\begin{cases} \mu_1 = \mu_1' \\ \alpha_{11} = \alpha_{11}' \,. \end{cases}$$

At last, since ${f_\theta^N}_{11}(1) = {f_{\theta'}^N}_{11}(1)$ and $\mu_1 > 0$,

$$\frac{\beta_1^2 + 4\pi^2}{\beta_1^2(1-\alpha_{11})^2 + 4\pi^2} = \frac{{\beta_1'}^2 + 4\pi^2}{{\beta_1'}^2(1-\alpha_{11})^2 + 4\pi^2} \,,$$

which implies

$$\alpha_{11}(2-\alpha_{11})\beta_1^2 = \alpha_{11}(2-\alpha_{11}){\beta_1'}^2 \,.$$

Either $\alpha_{11} > 0$, so the previous equation leads to

$$\beta_1 = \beta_1' \,,$$

or $\alpha_{11} = 0$, so $\alpha_{11}' = \alpha_{11} = 0$ and $\beta_1 = \beta_1' = 1$ (since θ and θ' are admissible for the model) .

This proves that $\mathbf{f}_\theta^N = \mathbf{f}_{\theta'}^N \implies \theta = \theta'$, which achieves the proof.

G Proof of Proposition 4.3, Situations 3 and 4

Without loss of generality, let

$$\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{21} & \alpha_{22} \end{pmatrix}, 0 < \alpha_{11} < 1, \alpha_{21} > 0, 0 \le \alpha_{22} < 1 \right\} .$$

Then, for any admissible parameter $\theta = (\mu, \alpha, \beta, \lambda_0)$ and all $\omega \in \mathbb{R}$:

$$\begin{cases} {f_\theta^N}_{11}(\omega) = \dfrac{\mu_1}{1-\alpha_{11}} \dfrac{\beta_1^2\alpha_{11}(2-\alpha_{11})}{\beta_1^2(1-\alpha_{11})^2 + 4\pi^2\omega^2} + \dfrac{\mu_1}{1-\alpha_{11}} + \lambda_0 \\ {f_\theta^N}_{12}(\omega) = \left({f_\theta^N}_{11}(\omega) - \lambda_0\right) \dfrac{\alpha_{21}\beta_2}{\beta_2^2(1-\alpha_{22})^2 + 4\pi^2\omega^2} \left(\beta_2(1-\alpha_{22}) - 2\pi \mathrm{i}\omega\right) \\ {f_\theta^N}_{22}(\omega) = \left(\dfrac{\mu_1\alpha_{21}}{(1-\alpha_{22})(1-\alpha_{11})} + \dfrac{\mu_2}{1-\alpha_{22}} \right) \dfrac{\beta_2^2 + 4\pi^2\omega^2}{\beta_2^2(1-\alpha_{22})^2 + 4\pi^2\omega^2} + \\ \qquad \dfrac{\mu_1\alpha_{21}^2\beta_2^2}{1-\alpha_{11}} \dfrac{\beta_1^2 + 4\pi^2\omega^2}{(\beta_2^2(1-\alpha_{22})^2 + 4\pi^2\omega^2)(\beta_1^2(1-\alpha_{11})^2 + 4\pi^2\omega^2)} + \lambda_0 \,, \end{cases}$$

and ${f_\theta^N}_{21}(\omega) = \overline{{f_\theta^N}_{12}(\omega)}$.

Let $\theta' = (\mu', \alpha', \beta', \lambda_0')$ be an admissible parameter such that $\mathbf{f}_\theta^N = \mathbf{f}_{\theta'}^N$. We start by showing that $\lambda_0 = \lambda_0'$. From $\mathfrak{Re}\left(f_\theta^{N}{}_{12}(1)\right) = \mathfrak{Re}\left(f_{\theta'}^{N}{}_{12}(1)\right)$ and $\mathfrak{Im}\left(f_\theta^{N}{}_{12}(1)\right) = \mathfrak{Im}\left(f_{\theta'}^{N}{}_{12}(1)\right)$, the following system of equations is established:

$$\begin{cases} \dfrac{(f_\theta^{N}{}_{11}(1) - \lambda_0)\alpha_{21}\beta_2}{\beta_2^2(1-\alpha_{22})^2 + 4\pi^2}\beta_2(1-\alpha_{22}) = \dfrac{(f_{\theta'}^{N}{}_{11}(1) - \lambda_0')\alpha_{21}'\beta_2'}{\beta_2'^{\,2}(1-\alpha_{22}')^2 + 4\pi^2}\beta_2'(1-\alpha_{22}') \\ \dfrac{(f_\theta^{N}{}_{11}(1) - \lambda_0)\alpha_{21}\beta_2}{\beta_2^2(1-\alpha_{22})^2 + 4\pi^2} = \dfrac{(f_{\theta'}^{N}{}_{11}(1) - \lambda_0')\alpha_{21}'\beta_2'}{\beta_2'^{\,2}(1-\alpha_{22}')^2 + 4\pi^2}\,. \end{cases} \tag{34}$$

Since

$$f_\theta^{N}{}_{11}(1) - \lambda_0 = \frac{\mu_1}{1-\alpha_{11}}\frac{\beta_1^2 + 4\pi^2}{\beta_1^2(1-\alpha_{11})^2 + 4\pi^2} \neq 0\,,$$

and $\alpha_{21}\beta_2 > 0$, Equations (34) imply that

$$\beta_2(1-\alpha_{22}) = \beta_2'(1-\alpha_{22}')\,. \tag{35}$$

Now, from $f_\theta^{N}{}_{12}(0) = \left(f_\theta^{N}{}_{11}(0) - \lambda_0\right)\alpha_{21}/(1-\alpha_{22})$ and $f_\theta^{N}{}_{11}(0) - \lambda_0 = \mu_1/(1-\alpha_{11})^3 \neq 0$, it comes $\alpha_{21} = f_\theta^{N}{}_{12}(0)(1-\alpha_{22})/\left(f_\theta^{N}{}_{11}(0) - \lambda_0\right)$ and the second equality of Equations (34) implies that

$$\frac{f_\theta^{N}{}_{11}(1) - \lambda_0}{f_\theta^{N}{}_{11}(0) - \lambda_0}\frac{f_\theta^{N}{}_{12}(0)\beta_2(1-\alpha_{22})}{\beta_2^2(1-\alpha_{22})^2 + 4\pi^2} = \frac{f_{\theta'}^{N}{}_{11}(1) - \lambda_0'}{f_{\theta'}^{N}{}_{11}(0) - \lambda_0'}\frac{f_{\theta'}^{N}{}_{12}(0)\beta_2'(1-\alpha_{22}')}{\beta_2'^{\,2}(1-\alpha_{22}')^2 + 4\pi^2}\,. \tag{36}$$

By Equation (35)

$$\frac{f_\theta^{N}{}_{12}(0)\beta_2(1-\alpha_{22})}{\beta_2^2(1-\alpha_{22})^2 + 4\pi^2} = \frac{f_{\theta'}^{N}{}_{12}(0)\beta_2'(1-\alpha_{22}')}{\beta_2'^{\,2}(1-\alpha_{22}')^2 + 4\pi^2}\,,$$

and since $\alpha_{21} > 0$, $f_\theta^{N}{}_{12}(0) \neq 0$ and Equation (36) leads to

$$\frac{f_\theta^{N}{}_{11}(1) - \lambda_0}{f_\theta^{N}{}_{11}(0) - \lambda_0} = \frac{f_{\theta'}^{N}{}_{11}(1) - \lambda_0'}{f_{\theta'}^{N}{}_{11}(0) - \lambda_0'}\,.$$

Since $f_\theta^{N}{}_{11}$ is strictly decreasing, $f_\theta^{N}{}_{11}(1) - f_\theta^{N}{}_{11}(0) = f_{\theta'}^{N}{}_{11}(1) - f_{\theta'}^{N}{}_{11}(0) \neq 0$, and it comes $\lambda_0 = \lambda_0'$.

Now, the expression of $f_\theta^{N}{}_{11}$ is similar to that of the univariate spectral density f_θ^N in Proposition 3.2. Thus, following the proof in Appendix B, from $f_\theta^{N}{}_{11} = f_{\theta'}^{N}{}_{11}$ and $\lambda_0 = \lambda_0'$ (since $\alpha_{11} > 0$ and $\alpha_{11}' > 0$) it comes:

$$\begin{cases} \mu_1 = \mu_1' \\ \alpha_{11} = \alpha_{11}' \\ \beta_1 = \beta_1'\,. \end{cases}$$

Next, from $f_\theta^{N}{}_{12}(0) = f_{\theta'}^{N}{}_{12}(0)$, it comes:

$$\left(f_\theta^{N}{}_{11}(0) - \lambda_0\right)\frac{\alpha_{21}}{1-\alpha_{22}} = \left(f_{\theta'}^{N}{}_{11}(0) - \lambda_0'\right)\frac{\alpha_{21}'}{1-\alpha_{22}'}\,,$$

which implies, with $f_\theta^{N}{}_{11}(0) - \lambda_0 = f_{\theta'}^{N}{}_{11}(0) - \lambda_0' \neq 0$:

$$\frac{\alpha_{21}}{1-\alpha_{22}} = \frac{\alpha_{21}'}{1-\alpha_{22}'}\,. \tag{37}$$

Now, from $\lim_{\omega'\to\infty} f_\theta^{N}{}_{22}(\omega') = \lim_{\omega'\to\infty} f_{\theta'}^{N}{}_{22}(\omega')$,

$$\frac{\mu_1\alpha_{21}}{(1-\alpha_{11})(1-\alpha_{22})} + \frac{\mu_2}{1-\alpha_{22}} = \frac{\mu_1\alpha_{21}'}{(1-\alpha_{11})(1-\alpha_{22}')} + \frac{\mu_2'}{1-\alpha_{22}'}\,,$$

and leveraging Equation (37), it comes:

$$\frac{\mu_2}{1-\alpha_{22}} = \frac{\mu_2'}{1-\alpha_{22}'}\,. \tag{38}$$

Moreover,

$$f_\theta^N{}_{22}(0) = \left(\frac{\mu_1 \alpha_{21}}{(1-\alpha_{11})(1-\alpha_{22})} + \frac{\mu_2}{1-\alpha_{22}} + \frac{\mu_1 \alpha_{21}^2}{(1-\alpha_{11})^3} \right) \frac{1}{(1-\alpha_{22})^2} + \lambda_0$$
$$= \left(\frac{\mu_1}{1-\alpha_{11}} \frac{\alpha_{21}}{1-\alpha_{22}} + \frac{\mu_2}{1-\alpha_{22}} \right) \frac{1}{(1-\alpha_{22})^2} + \frac{\mu_1}{(1-\alpha_{11})^3} \frac{\alpha_{21}^2}{(1-\alpha_{22})^2} + \lambda_0 \,.$$

Thus, by Equations (37) and (38), we obtain from $f_\theta^N{}_{22}(0) = f_{\theta'}^N{}_{22}(0)$:

$$\frac{1}{(1-\alpha_{22})^2} = \frac{1}{(1-\alpha'_{22})^2} \,,$$

which implies

$$\alpha_{22} = \alpha'_{22} \,.$$

To conclude, from Equations (35), (38) and (37), $\beta_2 = \beta'_2$, $\mu_2 = \mu'_2$ and $\alpha_{21} = \alpha'_{21}$. This proves that $\mathbf{f}_\theta^N = \mathbf{f}_{\theta'}^N \implies \theta = \theta'$, which achieves the proof.