Insurance Mathematics and Economics 116 (2024) 1–26

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

A Hawkes model with CARMA(p,q) intensity

Lorenzo Mercuri [a], Andrea Perchiazzo [b], Edit Rroji [c,*]

[a] *Department of Economics Management and Quantitative Methods, University of Milan, Milan, Italy*
[b] *Faculty of Economic and Social Sciences and Solvay Business School, Vrije Universiteit Brussel, Brussels, Belgium*
[c] *Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy*

ARTICLE INFO

JEL classification:
C00

Keywords:
Point processes
Autocorrelation
CARMA
Hawkes
Infinitesimal generator
Markov process

ABSTRACT

In this paper we introduce a new model, named CARMA(p,q)-Hawkes, as the Hawkes model with exponential kernel implies a strictly decreasing behavior of the autocorrelation function while empirical evidences reject its monotonicity. The proposed model is a Hawkes process where the intensity follows a Continuous Time Autoregressive Moving Average (CARMA) process. We also study the conditions for the stationarity and the positivity of the intensity and the strong mixing property for the increments. Furthermore, we present two estimation case studies based respectively on the likelihood and on the autocorrelation function.

1. Introduction

Point processes are useful mathematical models that describe the dynamics of observed event times and have been applied in several fields of study from queueing theory to forestry statistics. Among the family of point processes the Hawkes (1971a,b) process is widely the most established and widespread model in different areas, especially in quantitative finance, actuarial science and seismology (see Ogata 1988 and references therein for further details). Indeed the Hawkes process is particularly interesting since it is a self-exciting process, which means that each arrival excites the intensity such that the probability of the next arrival is increased for some period after the jump, and consequently it is well-suited to investigate, for instance, natural clustering effects and bank default in time. To show the versatility of the Hawkes process we mention a few other possible non-financial and non-insurance applications: a) social science area such as the modeling of urban crime (Mohler et al. 2011) and the population dynamics (Boumezoued 2016); b) social media sector as done in Rizoiu et al. (2017); and c) the modeling of disease spreading such as COVID-19 transmission as discussed in Chiang et al. (2022).

Recently the Hawkes process has gained a relevant role in financial modeling, in particular in the field of market microstructure. As a matter of fact it is used to model market activity, especially order arrivals in the limit order book (e.g., Bacry et al., 2013; Muni Toke and Yoshida, 2017; Clinet and Yoshida, 2017). For a complete overview of applications of the Hawkes process in finance, we suggest the works of Bacry et al. (2015) and Hawkes (2018). The Hawkes process has aroused its appeal among researchers and practitioners as well as in the insurance area. Indeed, as mentioned in Lesage et al. (2022), insurance companies are interested in point processes for the quantification of regulatory capital and in managing risks (e.g., computing ruin probabilities and measuring the effect of cyber-attacks as discussed respectively in Cheng and Seol 2020, Bessy-Roland et al. 2021 and, recently, Hillairet et al. (2023) for cyber-insurance derivatives). Swishchuk et al. (2021) show that the use of a Hawkes process with exponential kernel for modeling insurance claim occurrences provides an improvement over the fit of a classical Poisson model. However, they are not able to fit different empirical autocorrelation functions as exhibited in Swishchuk et al. (2021, Figures 3 and 5, p. 112). For recent results on Hawkes process we suggest Cattiaux et al. (2022) and references therein.

As stated in Errais et al. (2010), the Hawkes process with exponential kernel is Markov and shows a good level of tractability that makes it useful for real applications in the presence of large data sets (e.g., high-frequency market data). The specification of the kernel restricts the shape of the time dependence structure of the number of jumps observed in intervals with the same length. Indeed, as observed in Da Fonseca and Zaatour

* Corresponding author.
E-mail addresses: lorenzo.mercuri@unimi.it (L. Mercuri), andrea.perchiazzo@vub.be (A. Perchiazzo), edit.rroji@unimib.it (E. Rroji).

https://doi.org/10.1016/j.insmatheco.2024.01.007
Received June 2023; Received in revised form January 2024; Accepted 17 January 2024
Available online 2 February 2024
0167-6687/

(2014), the autocorrelation in a Hawkes model is a decaying function of lags which is not flexible enough to represent the dependence structure observed in many data sets (e.g., wind speed data in which the exponential autocorrelation overshoots the empirical one for small lags and vice versa for large lags as documented in Benth and Rohde 2019; and, as shown in Hitaj et al. 2019, mortality rates where the empirical autocorrelation function of the shock term appears to be non-monotonic).

To overcome the aforementioned drawback, in this paper we introduce a new model named CARMA(p,q)-Hawkes process. The proposed model is a Hawkes process where the intensity follows a Continuous-time Autoregressive Moving Average (CARMA) process and it is able to provide several shapes of the autocorrelation function as it removes the monotonicity constraint detected in the standard Hawkes process. The greater flexibility relies on the CARMA(p,q) component of our model, especially in the choice of the autoregressive and moving average parameters. The CARMA process, introduced in Doob (1944), is the continuous-time version of the ARMA model and has the advantage, other than to design different shapes of autocorrelation functions, to handle better irregular time series with respect to the ARMA process, especially for high-frequency market data, as discussed in Marquardt and Stelzer (2007) and Tómasson (2015). As a matter of fact, the CARMA model has found many applications in the literature. Here, we list a few of these applications: a) Andresen et al. (2014) use a CARMA(p,q) model for short and forward interest rates, while b) Hitaj et al. (2019) employ such a model in order to capture the dynamics of the shock term in mortality modeling; c) Benth et al. (2014) consider a non-Gaussian CARMA process for the dynamics of spot and derivative prices in electricity markets; and d) Mercuri et al. (2021) provide formulas for the futures term structure and options written on futures in the framework of a CARMA(p,q) model driven by a time-changed Brownian motion. As remarked in Iacus and Mercuri (2015), CARMA models have manifold interests: they can be used to describe directly the dynamics of time series and to construct the variance process in continuous time models (see Brockwell et al. 2006 and Iacus et al. 2017, 2018 for further details). Our paper presents a different type of application as we use CARMA(p,q) models for the intensity of a point process.

The paper is organized as follows. In Section 2, we review the Hawkes process with exponential kernel. In Section 3, we introduce the CARMA(p,q)-Hawkes process, study the conditions of stationarity and positivity for the intensity, and provide the likelihood function. In Section 4, we focus on the autocorrelation function of jumps in the proposed model and prove the strong mixing property of increments that leads to the asymptotic distribution of the empirical autocorrelation function. In Section 5, we present two estimation case studies in which we highlight methodological differences that may emerge from the level of data granularity and data storage since the estimation of parameters can be affected (see Shlomovich et al. 2022 and reference therein for details). In case of non-aggregate data, discussed in Section 5.1 with exact event times as for seismological data, we use the maximum likelihood estimation. Whereas in Section 5.2, that deals with aggregate data that refer to market orders with time span of one minute interval, we employ the autocorrelation function. Section 6 concludes the paper.

2. The Hawkes process

Point processes are useful to describe the dynamics of observed event times, i.e., a collection of realizations $\{t_i\}_{i=0}^{\infty}$, $t_i > 0$ for $i = 1, 2, \dots$ with $t_0 := 0$ of the non-decreasing non-negative process $\left\{T_i\right\}_{i\geq 1}$ called the time arrival process. The counting process N_t, representing the number of events up to time t, is obtained from the time arrival process as follows:

$$N_t := \sum_{i\geq 1} \mathbb{1}_{\{T_i \leq t\}} \tag{1}$$

for $t \geq 0$ with associated filtration $(\mathcal{F}_t)_{t\geq 0}$ that contains the information of the counting process N_t up to time t. An important quantity when dealing with a point process N_t is the conditional intensity λ_t defined as:

$$\lambda_t = \lim_{\Delta\to 0^+} \frac{\Pr[N_{t+\Delta} - N_t = 1|\mathcal{F}_t]}{\Delta}.$$

It then follows that the counting process satisfies the following properties

$$\Pr\left[N_{t+\Delta} - N_t = \eta\left|\mathcal{F}_t\right.\right] = \begin{cases} 1 - \lambda_t\Delta + o(\Delta) & \text{if } \eta = 0 \\ \lambda_t\Delta + o(\Delta) & \text{if } \eta = 1 \\ o(\Delta) & \text{if } \eta > 1. \end{cases}$$

The conditional intensity λ_t of a general self-exciting process has the following form:

$$\lambda_t = \mu + \int_{[0,t)} h(t-s)\,\mathrm{d}N_s \tag{2}$$

with baseline intensity parameter $\mu > 0$ and (excitation) kernel function $h(t) : [0, +\infty) \to [0, +\infty)$ that represents the contribution to the intensity at time t that is made by an event occurred at a previous time $T_i < t$. Following the general results about the Hawkes process in Brémaud and Massoulié (1996), the stationary condition reads:

$$\int_0^{+\infty} h(t)\,\mathrm{d}t < 1. \tag{3}$$

The most used kernel is the exponential function and specifically $h(t) = \alpha e^{-\beta t}$ with $\alpha, \beta \geq 0$. The stationary condition in (3) implies $\beta > \alpha$.

Exploiting the Markov property of the process $X_t := \left(\lambda_t, N_t\right)$, it is possible to get the infinitesimal generator (see Errais et al. 2010 and Da Fonseca and Zaatour 2014 for further details) associated to a function $f : \mathbb{R}_+ \times \mathbb{N} \to \mathbb{R}$ with continuous partial derivatives with respect to the first argument $\frac{\partial f}{\partial \lambda}(x)$. Starting from the definition of the infinitesimal operator for a Markov process X_t, that is,

$$\mathcal{A}f := \lim_{\Delta\to 0^+} \frac{\mathbb{E}\left[f\left(X_{t+\Delta}\right)|\mathcal{F}_t\right] - f\left(X_t\right)}{\Delta},$$

Errais et al. (2010) compute the infinitesimal generator for the Hawkes process with exponential kernel that writes as

$$\mathcal{A}f = \beta\left(\mu - \lambda_t\right)\frac{\partial f}{\partial \lambda}\left(\lambda_t, N_t\right) + \lambda_t\left[f\left(\lambda_t + \alpha, N_t + 1\right) - f\left(\lambda_t, N_t\right)\right]. \tag{4}$$

For every function f in the domain of the infinitesimal generator it is possible to build a martingale process M_t with respect to the natural filtration in the following way

$$M_t = f\left(\lambda_t, N_t\right) - f\left(\lambda_0, N_0\right) - \int_0^t \mathcal{A}f\left(\lambda_s, N_s\right)\mathrm{d}s,$$

which leads to the well-known Dynkin's formula

$$\mathbb{E}\left[f\left(\lambda_t, N_t\right)|\mathcal{F}_s\right] = f\left(\lambda_s, N_s\right) + \mathbb{E}\left[\int_s^t \mathcal{A}f\left(\lambda_u, N_u\right)\mathrm{d}u\middle|\mathcal{F}_s\right], \quad \forall t > s.$$

The above formula for $f \equiv N_t$ is used in Da Fonseca and Zaatour (2014) to compute the moments and the autocovariance function of jump increments observed in intervals of length τ with lag δ in which is shown that the Hawkes model with exponential kernel can only reproduce strictly decreasing autocorrelation functions for varying lag values δ. An interesting extension is given in Boswijk et al. (2018) where the self-excitation is identified through the modeling of common jumps between the log price process and its own jump intensity.

3. CARMA(p,q)-Hawkes model

In this section, we introduce the CARMA(p,q)-Hawkes model (Section 3.1), a point process where the intensity follows a CARMA(p,q) process, and its likelihood function (Section 3.2).

3.1. CARMA(p,q)-Hawkes: stationarity and positivity conditions for the intensity

Definition 1. A vector process $[X_{1,t}, \ldots, X_{p,t}, N_t]^\top$ of dimension $p+1$ is a CARMA(p,q)-Hawkes process if the conditional intensity λ_t is defined as

$$\lambda_t = \mu + \mathbf{b}^\top X_t, \tag{5}$$

where $\mu > 0$ is the baseline parameter and $\mathbf{b}$ is a p-dimensional column vector containing the $q+1$ moving average parameters $b_0, \ldots, b_q$ defined as

$$\mathbf{b} = \begin{cases} \left[b_0, \ldots, b_q\right]^\top & \text{if } p - q = 1 \\ \left[b_0, \ldots, b_q, b_{q+1} \ldots, b_{p-1}\right]^\top, \text{ with } b_{q+1} = \ldots = b_{p-1} = 0 & \text{if } p - q \geq 2. \end{cases} \tag{6}$$

The p-dimensional process $X_t = \left[X_{1,t}, \ldots, X_{1,p}\right]^\top$ is defined as:

$$X_t = \int_{[0,t)} e^{\mathbf{A}(t-s)}\mathbf{e}\mathrm{d}N_s, \tag{7}$$

where the exponential matrix $e^{\mathbf{A}} := \sum_{h=0}^{+\infty} \frac{1}{h!}\mathbf{A}^h$.

The $p \times p$ matrix $\mathbf{A}$, named companion matrix, has the following form

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ -a_p & -a_{p-1} & -a_{p-2} & \ldots & -a_1 \end{bmatrix}_{p\times p}, \tag{8}$$

where $a_1, \ldots a_p$ are the autoregressive parameters. The p-dimensional column vector $\mathbf{e}$ is defined as:

$$\mathbf{e} = [0, \ldots, 1]^\top. \tag{9}$$

Remark 1. The process X_t in (7) satisfies the following stochastic differential equation (sde)

$$\mathrm{d}X_t = \mathbf{A}X_t\mathrm{d}t + \mathbf{e}\mathrm{d}N_t \quad \text{ with } X_0 = \mathbf{0}. \tag{10}$$

The dynamics in (10) describes the state space process in a CARMA(p,q) model driven by N_t rather than a Lévy process as done in previous literature (see Brockwell, 2001; Brockwell et al., 2011; Tómasson, 2015, and references therein). As a result, the intensity λ_t in (5) is a CARMA(p,q) model. This means that the CARMA(p,q)-Hawkes process combines the self-exciting effect in a Hawkes process with the time-dependence structure of a CARMA(p,q) process (see Brockwell, 2004; Benth et al., 2014, for some examples). It is worth noting that N_t is right-continuous while X_t and λ_t are left-continuous.

Theorem 1. *The process X_t and the $(p+1)$-dimensional column vector process $\left[X_t, N_t\right]$ in Definition 1 are Markov.*[1]

[1] For the Markov property we refer to Cinlar (2011, Theorem 1.2, p. 444). Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space with filtration $\mathcal{F} = \left(\mathcal{F}_t\right)_{t\in\mathbb{T}}$ with $\mathbb{T} \subseteq \mathbb{R}$ and let $(E, \mathcal{E})$ be a measurable space. Consider a stochastic process $Z := \left\{Z_t : \Omega \to E\right\}_{t\in\mathbb{T}}$ adapted to the filtration $\mathcal{F}$. The process Z is Markov relative to $\mathcal{F}$ if and only if

Proof. See Appendix D.1 $\square$

To investigate the stationary regime of a CARMA(p,q)-Hawkes model, it is necessary to determine the conditions required for a non-negative kernel, i.e., $h(t) := \mathbf{b}^\top e^{\mathbf{A}t}\mathbf{e} \geq 0,\ \forall t \geq 0$. In case of a CARMA(p,q) driven by a non-negative Lévy process the conditions of a non-negative kernel are presented in Tsai and Chan (2005, Theorem 1, p. 592). In a similar fashion such conditions can be applied directly to our case due to the non-negative trajectories of the counting process N_t. Indeed, as done in Brockwell et al. (2006, Theorem 5.2) for COGARCH(p,q) models, in the next proposition we rephrase the main result that can be applied in a generic CARMA(p,q)-Hawkes process with $b_0 > 0$.

Proposition 1. *For a CARMA(p,q)-Hawkes process in which the real part of all eigenvalues of* $\mathbf{A}$ *is negative, the kernel function* $h(t) := \mathbf{b}^\top e^{\mathbf{A}t}\mathbf{e}\mathbb{1}_{\{t\geq 0\}}$ *is non-negative if and only if the ratio function* $\frac{b(z)}{a(z)}$ *is completely monotone*[2] *on* $(0, +\infty)$ *with the polynomials* $a(z)$ *and* $b(z)$ *defined respectively as*

$$a(z) := z^p + a_1 z^{p-1} + \ldots + a_p \quad \text{and} \quad b(z) := b_0 + b_1 z + \ldots + b_{p-1} z^{p-1}.$$

Remark 2. In the case of real negative eigenvalues the following results apply:

(a) Suppose all eigenvalues of $\mathbf{A}$ are negative real numbers sorted as follows $\tilde{\lambda}_p \leq, \ldots, \leq \tilde{\lambda}_1$ and that all the roots of $b(z) = 0$ are negative real numbers such that $\gamma_q \leq, \ldots, \leq \gamma_1 < 0$. If $\sum_{i=1}^k \gamma_i \leq \sum_{i=1}^k \tilde{\lambda}_i$ for $1 \leq k \leq q$, then the kernel of a CARMA(p,q)-Hawkes process is non-negative.
(b) A necessary and sufficient condition for a non-negative $h(t)$ in a CARMA(2,1)-Hawkes process is that $\tilde{\lambda}_2 \leq \tilde{\lambda}_1 < 0$ and $b_0 + \tilde{\lambda}_1 b_1 \geq 0$ with $b_1 \geq 0$.

We notice that the non-negativity requirement for the kernel implies a strictly positive intensity process λ_t as the baseline parameter μ is strictly positive.
Without loss of generality, we assume that matrix $\mathbf{A}$ is diagonalizable which corresponds to the assumption that the eigenvalues of $\mathbf{A}$ are distinct. The eigenvectors of $\mathbf{A}$ are

$$\left[1, \tilde{\lambda}_j, \ldots, \tilde{\lambda}_{p-1}\right]^\top,\ j = 1, \ldots, p$$

used to define a $p \times p$ matrix $\mathbf{S}$ as

$$\mathbf{S} := \begin{bmatrix} 1 & \ldots & 1 \\ \tilde{\lambda}_1 & \ldots & \tilde{\lambda}_p \\ \tilde{\lambda}_1^2 & \ldots & \tilde{\lambda}_p^2 \\ \vdots & & \vdots \\ \tilde{\lambda}_1^{p-1} & \ldots & \tilde{\lambda}_p^{p-1} \end{bmatrix}.$$

It follows that $\mathbf{S}$ satisfies $\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \mathsf{diag}\left(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_p\right)$, a result used to prove the next theorem on the stationarity conditions for a CARMA(p,q)-Hawkes process.

Theorem 2. *Let us consider a non-negative kernel function and suppose* $\mu > 0$. *Then a CARMA(p,q)-Hawkes* $\left(X_{1,t}, \ldots, X_{p,t}, N_t\right)$ *is a stationary process if all eigenvalues of* $\mathbf{A}$ *are distinct with non-negative real part and* $-\mathbf{b}^\top\mathbf{A}^{-1}\mathbf{e} < 1$.

Proof. See Appendix D.2. $\square$

Assumption 1. We shall assume for the remainder of the paper that: i) the kernel is a non-negative function and $\mu > 0$; and ii) all eigenvalues of $\mathbf{A}$ are distinct with negative real part and $-\mathbf{b}^\top\mathbf{A}^{-1}\mathbf{e} < 1$.

For practical applications, instead of checking ex-post signs of eigenvalues of matrix $\mathbf{A}$, it is possible to enforce ex-ante the negativity of the real part for eigenvalues using some transformations on the parameters space as done, for example, in Tómasson (2015). As a CARMA(p,q)-Hawkes process is Markov, we are able to calculate the infinitesimal operator as described in the following theorem.

Theorem 3. *Let* $f(x_1, \ldots, x_p, n) : \mathbb{R}^p \times \mathbb{N} \to \mathbb{R}$ *be a function in which the first* p *derivatives* $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_p}$ *are required to be well defined and continuous. Under the same conditions in Assumption 1, the infinitesimal generator of function* f *for a CARMA(p,q)-Hawkes* $Y_t = \left[X_{1,t}, \ldots, X_{p,t}, N_t\right]$ *process is*[3]*:*

$$\begin{aligned} \mathcal{A}f_t = & \left(\mu + \sum_{j=1}^q b_j X_{j,t}\right)\left[f\left(X_{1,t}, \ldots, X_{p,t} + 1, N_t + 1\right) - f\left(X_{1,t}, \ldots, X_{p,t}, N_t\right)\right] \\ & + \sum_{i=1}^{p-1} \frac{\partial f}{\partial X_{i,t}} X_{i+1,t} + \frac{\partial f}{\partial X_{p,t}} \mathbf{A}_{[p,]} X_t \end{aligned} \tag{11}$$

where $\mathbf{A}_{[p,]}$ *is the p-th row of the companion matrix* $\mathbf{A}$ *and the intensity process* λ_t *is defined as in (5). Alternatively, denoting with* $f\left(X_t, N_t\right) := f\left(X_{1,t}, \ldots, X_{p,t}, N_t\right)$, *the infinitesimal generator can be written as*

for every time t and time $u > t$ and any positive $\mathcal{E}$−measurable function g we have $\mathbb{E}\left[g\left(Z_u\right)|\mathcal{F}_t\right] = \mathbb{E}\left[g\left(Z_u\right)|Z_t\right]$. For any indicator function $g = \mathbb{1}_A$ with $A \in \mathcal{E}$, the Markov property implies that $\mathbb{P}\left[Z_{t+u} \in A\,|\mathcal{F}_t\right] = \mathbb{P}\left[Z_{t+u} \in A\,|Z_t\right]$.

[2] A function $f(x)$ defined on $(0, +\infty)$ is said to be completely monotone if and only if it has derivatives of all orders and $(-1)^n \frac{\partial^n f(t)}{(\partial x)^n} \geq 0$ for $n = 0, 1, 3, \ldots$.

[3] The notation $\mathcal{A}f_t$ refers to the infinitesimal generator of f applied to $Y_t = \left[X_{1,t}, \ldots, X_{p,t}, N_t\right]$ i.e. $\mathcal{A}f_t := f\left(X_{1,t}, \ldots, X_{p,t}, N_t\right)$.

$$\mathcal{A} f_t = \left(\mu + \mathbf{b}^\top X_t\right)\left[f\left(X_t + \mathbf{e}, N_t + 1\right) - f\left(X_t, N_t\right)\right] + \nabla_p f^\top \mathbf{A} X_t \tag{12}$$

where $\nabla_p f := \left[\frac{\partial f}{\partial X_{1,t}}, \dots \frac{\partial f}{\partial X_{p,t}}\right]^\top$. *The quantities* X_t, $\mathbf{e}$ *and* $\mathbf{e}$ *have the same meaning as in Definition 1.*

Proof. See Appendix D.3. □

Under some mild conditions for the integrability of the transformation $f\left(X_T, N_T\right)$ at a generic finite final time $T < +\infty$ (see Errais et al., 2010; Cui et al., 2020, for instance), the conditional expected value for $f\left(X_T, N_T\right)$ can be computed applying the Dynkin's formula:

$$\mathbb{E}\left[f\left(X_T, N_T\right)\middle|\mathcal{F}_{t_0}\right] = f\left(X_{t_0}, N_{t_0}\right) + \mathbb{E}\left[\int_{t_0}^{T} \mathcal{A} f_t \mathrm{d}t \middle|\mathcal{F}_{t_0}\right] \tag{13}$$

that has a representation of the following form

$$\mathrm{d}\mathbb{E}\left[f\left(X_t, N_t\right)\middle|\mathcal{F}_{t_0}\right] = \mathbb{E}\left[\mathcal{A} f_t \middle|\mathcal{F}_{t_0}\right]\mathrm{d}t, \tag{14}$$

with initial condition $f\left(X_{t_0}, N_{t_0}\right)$. We use the infinitesimal generator (12) and the result in (13) to obtain the following theorem for the computation of the first moment of the counting process N_t. In the remainder of the paper, we use $\mathbb{E}_t[\cdot] := \mathbb{E}\left[\cdot\middle|\mathcal{F}_t\right]$.

Theorem 4. *Let* $\tilde{\mathbf{A}}$ *be a* $p \times p$ *companion matrix where the last row has the following structure*

$$\tilde{\mathbf{A}}_{[p,\cdot]} = \left[b_0 - a_p, b_1 - a_{p-1}, \dots, b_{p-1} - a_1\right]. \tag{15}$$

Under Assumption 1 and supposing that all eigenvalues of $\tilde{\mathbf{A}}$ *are distinct with negative real part, for any* $T > t_0 \geq 0$*, the conditional first moment of the counting process is*

$$\mathbb{E}_{t_0}\left[N_T\right] = N_{t_0} + \mu\left(1 - \mathbf{b}^\top \tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\left(T - t_0\right) + \mathbf{b}^\top \tilde{\mathbf{A}}^{-1}\left[e^{\tilde{\mathbf{A}}(T-t_0)} - I\right]\left[X_{t_0} + \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right], \tag{16}$$

while the conditional expected value of the process X_T *is*

$$\mathbb{E}_{t_0}\left[X_T\right] = e^{\tilde{\mathbf{A}}(T-t_0)}\left[X_{t_0} + \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right] - \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu. \tag{17}$$

The quantities in (16) and (17) satisfy respectively the following ordinary differential equations:

$$d\mathbb{E}_{t_0}\left[N_t\right] = \left[\mu\left(1 - \mathbf{b}^\top \tilde{\mathbf{A}}^{-1}\mathbf{e}\right) + \mathbf{b}^\top e^{\tilde{\mathbf{A}}(t-t_0)}\left[X_{t_0} + \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]\right] dt \tag{18}$$

and

$$d\mathbb{E}_{t_0}\left[X_t\right] = \left(\tilde{\mathbf{A}}\mathbb{E}_{t_0}\left[X_t\right] + \mu\mathbf{e}\right) dt \tag{19}$$

with initial conditions[4] X_{t_0} *and* N_{t_0}*. The long-run value for* $\mathbb{E}_{t_0}\left[X_T\right]$ *is obtained as follows*

$$\mathbb{E}\left[X_\infty\right] := \lim_{T \to +\infty} \mathbb{E}_{t_0}\left[X_T\right] = -\tilde{\mathbf{A}}\mathbf{e}\mu. \tag{20}$$

Moreover, the expected number of events that occurs in an interval with length τ*, i.e.,* $(T, T+\tau]$*, given the information at time* $t_0 < T$ *is*

$$\mathbb{E}_{t_0}\left[\left(N_{T+\tau} - N_T\right)\right] = \mu\left(1 - \mathbf{b}^\top \tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\tau + \mathbf{b}^\top \tilde{\mathbf{A}}^{-1} e^{\tilde{\mathbf{A}}(T-t_0)}\left(e^{\tilde{\mathbf{A}}\tau} - \mathbf{I}\right)\left(X_{t_0} + \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right) \tag{21}$$

and the stationary behavior of (21) is

$$\mathbb{E}\left[\Delta_\tau N_\infty\right] := \lim_{T \to +\infty} \mathbb{E}_{t_0}\left[N_{T+\tau} - N_T\right] = \mu\left(1 - \mathbf{b}^\top \tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\tau, \quad \forall \tau > 0. \tag{22}$$

Proof. See Appendix D.4. □

Using the same arguments in Brockwell et al. (2006, proof of Proposition 4.1, p. 815), all eigenvalues of matrix $\tilde{\mathbf{A}}$ have negative real parts if for some positive integer $r \geq 1$ the following inequality holds

$$\left\|\mathbf{S}^{-1}\mathbf{e}\mathbf{b}^\top \mathbf{S}\right\|_r < \mathsf{Re}\left(\tilde{\lambda}_1\right) \tag{23}$$

where, in this context, $\|\cdot\|_r$ denotes the natural matrix norm induced by the vector $\mathbb{L}^r$-norm. This result comes directly from an application of the Bauer-Fike Theorem (see Bauer and Fike 1960 for further details) since $\tilde{\mathbf{A}}$ is obtained by perturbing matrix $\mathbf{A}$ as $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{e}\mathbf{b}^\top$.
A sufficient condition for (23) is

[4] For $t_0 = 0$, then $\mathbb{E}_{t_0}\left[X_T\right] = \left(e^{\tilde{\mathbf{A}}(T-t_0)} - \mathbf{I}\right)\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu$ and
$\mathbb{E}_{t_0}\left[N_T\right] = \mu\left(1 - \mathbf{b}^\top \tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\left(T - t_0\right) + \mathbf{b}^\top \tilde{\mathbf{A}}^{-1}\left[e^{\tilde{\mathbf{A}}(T-t_0)} - I\right]\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu$.

$$\frac{\sigma_{\max}(\mathbf{S})}{\sigma_{\min}(\mathbf{S})}\|\mathbf{b}\|_2 < \mathrm{Re}\left(\tilde{\lambda}_1\right) \tag{24}$$

where $\|\mathbf{b}\|_2 := \sqrt{\sum_{i=1}^{p} b_{i-1}^2}$ is the Euclidean norm of $\mathbf{b}$, $\sigma_{\max}(\mathbf{S})$ and $\sigma_{\min}(\mathbf{S})$ are respectively maximal and minimal singular values of $\mathbf{S}$. In particular, we observe that

$$\left\|\mathbf{S}^{-1}\mathbf{e}\mathbf{b}^{\top}\mathbf{S}\right\|_2 \leq k_2(\mathbf{S})\left\|\mathbf{e}\mathbf{b}^{\top}\right\|_2 \tag{25}$$

and that $k_2(\mathbf{S}) := \|\mathbf{S}\|_2\left\|\mathbf{S}^{-1}\right\|_2$, the condition number in 2-norm, can be written as

$$k_2(\mathbf{S}) = \frac{\sigma_{\max}(\mathbf{S})}{\sigma_{\min}(\mathbf{S})}. \tag{26}$$

Moreover, denoting with $\|\mathbf{e}\mathbf{b}^{\top}\|_F$ the Frobenius norm of $\mathbf{e}\mathbf{b}^{\top}$, we obtain $\|\mathbf{e}\mathbf{b}^{\top}\|_2 \leq \|\mathbf{e}\mathbf{b}^{\top}\|_F$. Applying the definition of the Frobenius norm we have

$$\left\|\mathbf{e}\mathbf{b}^{\top}\right\|_2 \leq \|\mathbf{b}\|_2, \tag{27}$$

and combining (25), (26) and (27) we get

$$\left\|\mathbf{S}^{-1}\mathbf{e}\mathbf{b}^{\top}\mathbf{S}\right\|_2 \leq \frac{\sigma_{\max}(\mathbf{S})}{\sigma_{\min}(\mathbf{S})}\|\mathbf{b}\|_2. \tag{28}$$

Thus, the inequality in (24) implies (23).

3.2. Likelihood estimation of the CARMA(p,q)-Hawkes

As follows we present the likelihood of a CARMA(p,q)-Hawkes model. Consider that $\theta = \left(b_0, \ldots, b_q, a_1, \ldots, a_p\right)$, then the likelihood of a CARMA(p,q)-Hawkes model is given by

$$\mathcal{L}(\theta,\mu) = -\int_0^{T_k} \lambda_t \mathrm{d}t + \int_0^{T_k} \ln\left(\lambda_t\right)\mathrm{d}N_t. \tag{29}$$

Exploiting the fact that $\int_0^{T_k} \ln\left(\lambda_t\right)\mathrm{d}N_t = \sum_{i=1}^{k} \ln\left(\lambda_{T_i}\right)$, then (29) can be written as

$$\mathcal{L}(\theta,\mu) = -\int_0^{T_k} \left[\mu + \mathbf{b}^{\top}X_t\right]\mathrm{d}t + \sum_{i=1}^{k} \ln\left(\lambda_{T_i}\right) \tag{30}$$

and recalling once again that X_t can be expressed by (7) and rearranging the expression we have

$$\mathcal{L}(\theta,\mu) = -\mu T_k - \mathbf{b}^{\top}\int_0^{T_k}\int_0^{t} e^{\mathbf{A}(t-s)}\mathbf{e}\mathrm{d}N_s\mathrm{d}t + \sum_{i=1}^{k} \ln\left(\lambda_{T_i}\right). \tag{31}$$

Working on the inner integral, the likelihood becomes

$$\mathcal{L}(\theta,\mu) = -\mu\left(T_k\right) - \mathbf{b}^{\top}\int_0^{T_k}\left[\int_s^{T_k} e^{\mathbf{A}(t-s)}\mathrm{d}t\right]\mathrm{d}N_s\mathbf{e} + \sum_{i=1}^{k} \ln\left(\lambda_{T_i}\right), \tag{32}$$

while using the results in (A.1) we get

$$\mathcal{L}(\theta,\mu) = -\mu T_k - \mathbf{b}^{\top}\int_0^{T_k} \mathbf{A}^{-1}\left[e^{\mathbf{A}(T_k-s)} - \mathbf{I}\right]\mathrm{d}N_s\mathbf{e} + \sum_{i=1}^{k} \ln\left(\lambda_{T_i}\right). \tag{33}$$

Developing the integral in (33) and recalling that $S(k) := \sum_{i=1}^{k} e^{\mathbf{A}(T_k-T_i)}$, we finally obtain that the likelihood of a CARMA(p,q)-Hawkes model writes

$$\mathcal{L}(\theta,\mu) = -\mu T_k - \mathbf{b}^{\top}\mathbf{A}^{-1}S(k)\mathbf{e} + k\mathbf{b}^{\top}\mathbf{A}^{-1}\mathbf{e} + \sum_{i=1}^{k} \ln\left(\lambda_{T_i}\right). \tag{34}$$

4. Autocovariance and autocorrelation of a CARMA(p,q)-Hawkes process

In this section we compute the stationary autocorrelation and autocovariance functions for the number of jumps in non-overlapping time intervals of length τ. To this aim we introduce some quantities that are useful to compute the asymptotic covariance of a CARMA(p,q)-Hawkes process.

The first quantity we introduce is the $\frac{p(p+1)}{2} \times \frac{p(p+1)}{2}$ matrix $\tilde{\tilde{\mathbf{A}}}$ defined as follows

$$\tilde{\tilde{\mathbf{A}}} := \begin{bmatrix} D^1_{[p,p]} & U^{1,2}_{[p,p-1]} & 0_{[p,p-2]} & \dots & \dots & \dots \\ L^{2,1}_{[p-1,p]} & D^2_{[p-1,p-1]} & U^{2,3}_{[p-1,p-2]} & 0_{[p-1,p-3]} & \dots & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \dots \\ L^{j,1}_{[p-j+1,p]} & \dots & L^{j,j-1}_{[p-j+1,p-j+2]} & D^j_{[p-j+1,p-j+1]} & U^{j,j+1}_{[p-j+1,p-j]} & 0_{[p-j+1,p-j-1]} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \dots \\ L^{p,1}_{[1,p]} & \dots & \dots & \dots & \dots & D^p_{[1,1]} \end{bmatrix} \tag{35}$$

where the square matrices $D^j_{[p-j+1,p-j+1]}$, $j = 1, \dots, p-1$, have the following structure

$$D^j_{[p-j+1,p-j+1]} = \begin{bmatrix} 0 & 2 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 1 \\ b_{j-1} - a_{p-j+1} & b_j - a_{p-j} & \dots & \dots & b_{p-1} - a_1 \end{bmatrix},$$

with $D^p_{[1,1]} = 2(b_{p-1} - a_1)$. Matrices $L^{j,i}_{[p-j+1,p-i+1]}$ for $j = 2, \dots, p$ and $i = 1, \dots, j-1$ are characterized by the entries with the form

$$L^{j,i}(h,l) = \begin{cases} b_{j-2+i} - a_{p-j+1+(i-1)} & \text{if } h = p-j+1,\ l = j-i+1 \text{ and } j \neq p \\ 2\left(b_{j-2+i} - a_{p-j+1+(i-1)}\right) & \text{if } h = p-j+1,\ l = j-i+1 \text{ and } j = p \\ 0 & \text{otherwise} \end{cases}$$

while matrices $U^{i,i+1}_{[p-i+1,p-i]}$ for $i = 1, \dots, p-1$ have form

$$U^{i,i+1}_{[p-i+1,p-i]} = \begin{bmatrix} \mathbf{0}_{[1,p-i]} \\ \mathbf{I}_{[p-i,p-i]} \end{bmatrix}.$$

Here an example of the matrix $\tilde{\tilde{\mathbf{A}}}$ for a CARMA(3,2)-Hawkes model

$$\tilde{\tilde{\mathbf{A}}} = \begin{bmatrix} 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ b_0 - a_3 & b_1 - a_2 & b_2 - a_1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & b_0 - a_3 & 0 & b_1 - a_2 & b_2 - a_1 & 1 \\ 0 & 0 & 2(b_0 - a_3) & 0 & 2(b_1 - a_2) & 2(b_2 - a_1) \end{bmatrix}.$$

The second quantity introduced is the $p \times \frac{p(p+1)}{2}$ matrix $\mathbf{B}$ defined as:

$$\mathbf{B} := \begin{bmatrix} b_0 & b_1 & \dots & b_{p-1} & 0 & \dots & \dots & 0 & \dots & 0 \\ 0 & b_0 & \dots & 0 & b_1 & \dots & b_{p-1} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \dots & 0 \\ 0 & \dots & 0 & b_0 & 0 & \dots & b_1 & 0 & \dots & b_{p-1} \end{bmatrix} \tag{36}$$

where the generic i-th row is the result of a row concatenation of p vectors with dimensions p, $p-1$, …, $p-i$, … 1, respectively. The first $i-1$ vectors have zero entries except the element in position i that coincides with b_{i-1}, the vector with dimension $p-i$ contains the elements $b_i, \dots, b_{p-i}$ and the remaining vectors have zero entries.

For example, in the case of a CARMA(3,2)-Hawkes model, the structure of matrix $\mathbf{B}$ reads

$$\mathbf{B} = \begin{bmatrix} b_0 & b_1 & b_2 & 0 & 0 & 0 \\ 0 & b_0 & 0 & b_1 & b_2 & 0 \\ 0 & 0 & b_0 & 0 & b_1 & b_2 \end{bmatrix}.$$

The third quantity is the $\frac{p(p+1)}{2} \times p$ matrix $\tilde{\mathbf{C}}$ in which the entry in the $i-$ th row and in $j-$ th column has the following structure

$$c_{i,j} := \begin{cases} 0 & \text{if} \quad i \neq j\left(p - \frac{j-1}{2}\right) \text{ and } i \neq \frac{p(p+1)}{2} \\ \mu & \text{if} \quad i = j\left(p - \frac{j-1}{2}\right) \text{ and } i \neq \frac{p(p+1)}{2} \\ b_{j-1} & \text{if} \quad i = \frac{p(p+1)}{2} \text{ and } j \neq p \\ 2\mu + b_{p-1} & \text{if} \quad i = \frac{p(p+1)}{2} \text{ and } j = p \end{cases}. \tag{37}$$

Let H be a $p \times 1$ vector. Then we define the operator $vlt(\cdot)$ as a function that transforms the $p \times p$ matrix $HH^\top$ into a $\frac{p(p+1)}{2}$ vector containing the lower triangular part of the product $HH^\top$. Specifically:

$$vlt\left(HH^\top\right) := \Bigg[\underbrace{H_1H_1, \dots, H_pH_1}_{\text{p entries}}, \underbrace{H_2H_2, \dots, H_pH_2}_{\text{p-1 entries}}, \dots, \underbrace{H_iH_i, \dots, H_pH_i}_{\text{p-i+1 entries}}, \dots, H_pH_p\Bigg]^\top. \tag{38}$$

4.1. Conditions for existence of stationary autocovariance function

In the following section we present the variance and covariance of the number of jumps that occur in two non-overlapping time intervals of the same length for a CARMA(p,q)-Hawkes model. We rewrite the quantity $\mathbb{E}_{t_0}\left[X_T X_T^\top\right]\mathbf{b}$ using the $vlt(\cdot)$ operator defined in (38).

Lemma 1. *The following identity holds true*

$$\mathbb{E}_{t_0}\left[X_T X_T^\top\right]\mathbf{b} = \mathbf{B}vlt\left(\mathbb{E}_{t_0}\left(X_T X_T^\top\right)\right) \tag{39}$$

where the matrix $\mathbf{B}$ *is defined in* (36) *and the operator* $vlt(\cdot)$ *is defined as in* (38). *Moreover:*

$$\begin{aligned} vlt\left(\mathbb{E}_{t_0}\left(X_T X_T^\top\right)\right) &= e^{\tilde{\tilde{\mathbf{A}}}(T-t_0)} vlt\left(X_{t_0} X_{t_0}^\top\right) + \left[e^{\tilde{\tilde{\mathbf{A}}}(T-t_0)} - \mathbf{I}\right]\tilde{\tilde{\mathbf{A}}}^{-1}\mu\left(\tilde{\mathbf{e}} - \tilde{C}\tilde{\mathbf{A}}^{-1}\mathbf{e}\right) \\ &+ e^{\tilde{\tilde{\mathbf{A}}}T}\left[\int_{t_0}^{T} e^{-\tilde{\tilde{\mathbf{A}}}t}\tilde{\mathbf{C}}e^{\tilde{\mathbf{A}}t}dt\right] e^{-\tilde{\mathbf{A}}t_0}\left[X_{t_0} + \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]. \end{aligned} \tag{40}$$

Proof. See Appendix E.1. □

Theorem 5. *Under Assumption 1 and supposing that all eigenvalues of* $\tilde{\mathbf{A}}$ *and* $\tilde{\tilde{\mathbf{A}}}$ *have negative real parts, the long-run covariance*

$$Cov(\tau,\delta) := \lim_{t\to+\infty} \mathbb{E}[(N_{t+\tau} - N_t)(N_{t+2\tau+\delta} - N_{t+\tau+\delta})] - \mathbb{E}[N_{t+\tau} - N_t]\mathbb{E}[N_{t+2\tau+\delta} - N_{t+\tau+\delta}]$$

for a CARMA(p,q)-Hawkes process has the following form:

$$Cov(\tau,\delta) = \mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\left[e^{\tilde{\mathbf{A}}\tau} - \mathbf{I}\right]e^{\tilde{\mathbf{A}}\delta}g_\infty(\tau) \tag{41}$$

where $g_\infty(\tau)$ *is defined as*

$$g_\infty(\tau) := \left(\mathbf{I} - e^{\tilde{\mathbf{A}}\tau}\right)\tilde{\mathbf{A}}^{-1}\mu\left[\mathbf{e}\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e} - \mathbf{e} + \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\left(\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right) + \mathbf{B}\tilde{\tilde{\mathbf{A}}}^{-1}\left(\tilde{\mathbf{e}} - \tilde{\mathbf{C}}\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\right]. \tag{42}$$

Proof. See Appendix D.3. □

Theorem 6. *Under the same assumptions as in Theorem 5, the long-run variance*

$$Var(\tau) := \lim_{t\to+\infty} \mathbb{E}[(N_{t+\tau} - N_t)^2] - \mathbb{E}[N_{t+\tau} - N_t]^2$$

of the number of jumps in a interval with length τ *for a CARMA(p,q)-Hawkes process has the following form:*

$$\begin{aligned} Var(\tau) &= \left(1 - \mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\left(1 - 2\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\mu\tau + 2\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{A}}^{-1}\mathbf{e}\tau\mu^2\left(\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right) \\ &+ 2\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{B}\tilde{\tilde{\mathbf{A}}}^{-1}\mu\left(\tilde{\mathbf{e}} - \tilde{\mathbf{C}}\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\tau - 2\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\left[e^{\tilde{\mathbf{A}}\tau} - \mathbf{I}\right]h_\infty(0) \end{aligned} \tag{43}$$

where $h_\infty(0)$ *is defined as*

$$h_\infty(0) := -\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\left(1 - \mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right) + \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu^2\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e} + \tilde{\mathbf{A}}^{-1}\mathbf{B}\tilde{\tilde{\mathbf{A}}}^{-1}\mu\left(\tilde{\mathbf{e}} - \tilde{\mathbf{C}}\tilde{\mathbf{A}}^{-1}\mathbf{e}\right). \tag{44}$$

Proof. See Appendix E.2. □

Remark 3. Combining the results in Theorems 5 and 6, we determine the asymptotic autocorrelation function of the number of jumps in non-overlapping time intervals of length τ, i.e., $\rho_\tau(d)$, for a CARMA(p,q)-Hawkes in a closed-form formula:

$$\rho_\tau(d) = \frac{Cov(\tau, d-1)}{Var(\tau)}, \quad d = 1, 2, \ldots \tag{45}$$

where d denotes the lag order.

4.2. Strong mixing property for the increments of a CARMA(p,q)-Hawkes and asymptotic distribution of the autocorrelation function

The asymptotic distribution of the autocorrelation function of a CARMA(p,q)-Hawkes process can be easily obtained by showing that the increments of the process are strongly mixing.

Definition 2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{A}, \mathcal{B}$ two sub $\sigma-$algebras of $\mathcal{F}$. The strong-mixing coefficient is defined as:

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup\left\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \; A \in \mathcal{A}, B \in \mathcal{B}\right\}. \tag{46}$$

Following Poinas et al. (2019), the quantity in (46) can be reformulated for a point process N_t in the following way:

$$\alpha_N(r) := \sup_{t\in\mathbb{R}} \alpha\left(\xi_{-\infty}^{t}, \xi_{t+r}^{\infty}\right) \tag{47}$$

where ξ_a^b denotes the $\sigma-$algebra generated by the cylinder sets on the interval $(a,b]$.[5] Considering the sequence $\left(\Delta_1 N_k\right)_{k\in\mathbb{Z}}$ where $\Delta_1 N_k := N_{k+1} - N_k$ is the number of jumps in the interval of length 1 and extremes k, $k+1$, then the strong-mixing coefficient has the form

$$\alpha_{\Delta_1 N}(r) := \sup_{n\in\mathbb{Z}} \alpha\left(\mathcal{F}_{-\infty}^{n}, \mathcal{F}_{n+r}^{\infty}\right) \tag{48}$$

where $\mathcal{F}_a^b$ is the $\sigma-$algebra generated by the sequence $\left(\Delta_1 N_k\right)_{a\le k\le b}$. If $\alpha_N(r) \to 0$ (respectively $\alpha_{\Delta_1 N_k}(r) \to 0$) as $r \to +\infty$, the point process N_t (respectively $\Delta_1 N_k$) is said to be strongly-mixing.
Using Theorem 1 in Cheysson and Lang (2020), we obtain the following theorem.

Theorem 7. *A CARMA(p,q)-Hawkes process satisfying Assumption 1 is strongly mixing with exponential rate.*

Proof. See Appendix E.4. □

As shown in Cheysson and Lang (2020), we have that $\alpha_{\Delta_1 N}(r) \le \alpha_N(r)$ and the result in Theorem 7 implies that the sequence $\left(\Delta_1 N_k\right)_{k\in\mathbb{Z}}$ is strongly mixing. This result is useful to determine the asymptotic distribution of the sample autocovariance and autocorrelation functions associated to the sequence $\left(\Delta_1 N_k\right)_{k\in\mathbb{Z}}$. Following the result in Ibragimov and Linnik (1971), we obtain the following result for the asymptotic distribution of the sample mean, the sample variance and the sample autocovariance function.

Theorem 8. *Let $\left(N_t\right)_{t\ge 0}$ be a stationary CARMA(p,q)-Hawkes process that satisfies the assumptions in Theorem 7. We assume the existence of a positive constant ϕ such that $\mathbb{E}\left[\left(\Delta_1 N_1\right)^{4+\phi}\right] < +\infty$. Denoting with*

$$V_k := \begin{bmatrix} \Delta_1 N_k \\ \left(\Delta_1 N_k - \mathbb{E}\left(\Delta_1 N_\infty\right)\right)^2 \\ \left(\Delta_1 N_k - \mathbb{E}\left(\Delta_1 N_\infty\right)\right)\left(\Delta_1 N_{k+1} - \mathbb{E}\left(\Delta_1 N_\infty\right)\right) \\ \vdots \\ \left(\Delta_1 N_k - \mathbb{E}\left(\Delta_1 N_\infty\right)\right)\left(\Delta_1 N_{k+d} - \mathbb{E}\left(\Delta_1 N_\infty\right)\right) \end{bmatrix}, \text{ with } k = 1,\dots,n \text{ and } d < n$$

as $n \to +\infty$, we have:

$$\sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n} V_k - \begin{pmatrix} \mathbb{E}\left(\Delta_1 N_\infty\right) \\ Var\left(\Delta_1 N_\infty\right) \\ Acv(1) \\ \vdots \\ Acv(d) \end{pmatrix}\right) \to \mathcal{N}_{d+2}\left(\mathbf{0}, \Sigma\right) \tag{49}$$

where $Acv(d) := Cov(1, d-1)$ and

$$\Sigma := Var\left(V_1\right) + 2\sum_{k=2}^{+\infty} Cov\left(V_1 V_k^\top\right). \tag{50}$$

Proof. See Appendix E.5. □

Through the use of the Delta method, we study the asymptotic behavior of the random vector that contains the sample mean of the increments in the first position and the empirical autocorrelations in the remaining entries. As a first result, we report the asymptotic distribution of the empirical autocorrelation function. Denoting with $\hat{\rho}_{n,\tau} := \left[\hat{\rho}_{n,\tau}(1), \dots, \hat{\rho}_{n,\tau}(m)\right]^\top$ where $\hat{\rho}_\tau(d)$ is the sample estimator of $\rho_\tau(d)$ in (45). The asymptotic distribution of $\hat{\rho}_{n,\tau}$ is

$$\sqrt{n}\left(\hat{\rho}_{n,\tau} - \rho_\tau\right) \to \mathcal{N}_m\left(\mathbf{0}, \Sigma_\rho\right), \text{ as } n \to +\infty. \tag{51}$$

The variance - covariance matrix Σ_ρ has the following form:

$$\Sigma_\rho = J\rho_\tau \Sigma \left[J\rho_\tau\right]^\top,$$

where Σ is defined in (50); $J\rho_\tau$ is the Jacobian matrix of the autocorrelation that can be seen as a vector function of ϑ in (E.32). Therefore $J\rho_\tau$ is determined as:

$$J\rho_\tau = \left[\mathbf{0}_{m\times 1} \;\middle|\; -\frac{\rho_\tau}{Var\left(\Delta_1 N_\infty\right)} \;\middle|\; \frac{\mathbf{I}_{m\times m}}{Var\left(\Delta_1 N_\infty\right)}\right]. \tag{52}$$

[5] Let N be a counting process defined as a map from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to a measurable space $(\mathbb{M}, \mathcal{M})$ of locally finite counting measures on Ω. Then the $\sigma-$algebra ξ_a^b is defined as:

$$\xi_a^b := \sigma\left(\{N \in \mathbb{M} : N(A) = n\}; A \in \mathcal{B}((a,b]), n \in \mathbb{N}\right).$$

Following the same strategy, it is possible to determine the asymptotic distribution of the column vector $\hat{\chi}_{n,\tau} := \left[\widehat{\Delta_\tau N}_n, \hat{\rho}_{n,\tau}\right]^\top$ where $\widehat{\Delta_\tau N}_n$ is the sample estimator for $\mathbb{E}\left[\Delta_\tau N_\infty\right]$. Therefore, we have:

$$\sqrt{n}\left(\hat{\chi}_{n,\tau} - \chi_\tau\right) \to \mathcal{N}_{m+1}\left(\mathbf{0}, \Sigma_{\chi_\tau}\right), \text{ as } n \to +\infty \tag{53}$$

where χ_τ is a column vector containing the first moment $\mathbb{E}\left[\Delta_\tau N_\infty\right]$ and the first lag m autocorrelations. The variance-covariance asymptotic matrix Σ_{χ_τ} results to be

$$\Sigma_\chi = J\chi_\tau \Sigma \left[J\chi_\tau\right]^\top, \tag{54}$$

while the Jacobian matrix $J\chi_\tau$ can be written as

$$J\chi_\tau = \begin{bmatrix} e_1 \\ J\rho_\tau \end{bmatrix} \tag{55}$$

where the first element of row vector e_1 is equal to one and the others are all zeros.

4.3. Asymptotic distribution of the moment matching estimation based on the autocorrelation function

As discussed in Shlomovich et al. (2022), real event data can provide imprecision in the recording of event time-stamps (e.g., network traffic data). Furthermore, we have usually the following trade-off: expensive cost in recording event time with a high precision and poor accuracy of measurements. As a matter of fact, common practice is to work with binned data (that is, without loss of generality, the technique of aggregating data). To this aim, we propose a two-step estimation procedure, named Moment Matching Estimation (MME), for a CARMA(p,q)-Hawkes process and we provide the asymptotic distribution of the obtained estimators.

Consider a sequence of empirical observations for the increments of a counting process $\left(\Delta_\tau N_k\right)_{k=1,\dots,n}$. The first step is to compute the least squares estimator as

$$\hat{\theta}_n := \underset{\hat{\theta}_{n,\tau} \in \Theta \subseteq \mathbb{R}^{p+q+1}}{\operatorname{argmin}} M\left(\hat{\rho}_{n,\tau}, \theta\right) \tag{56}$$

where Θ is a compact subset of $\mathbb{R}^{p+q+1}$ such that the stationary condition is guaranteed, the kernel function is non-negative, higher order moments of a CARMA(p,q)-Hawkes process exist and the true vector parameter θ is an interior point of Θ. For a fixed $m \geq p+q+1$, $M : \mathbb{R}_+^m \times \Theta \to \mathbb{R}_0^+$ is defined as:

$$M\left(\hat{\rho}_{n,\tau}, \theta\right) := \sum_{d=1}^{m} \left(\hat{\rho}_{n,\tau}(d) - \rho_\tau(d)\right)^2 \tag{57}$$

in which d denotes the lag order, $\hat{\rho}_{n,\tau}(d)$ represents the empirical autocorrelation with lag d while $\rho_\tau(d)$ is its theoretical counterpart obtained in (45). The vector θ includes only the autoregressive $(a_1, \dots, a_p)$ and moving average $(b_0, \dots, b_q)$ parameters.

Let $\hat{\theta}_n$ be the solution of the minimization problem in (56). As the function in (57) is smooth (i.e., $M\left(\hat{\rho}_{n,\tau}, \theta\right) \in C^\infty$ with respect to both arguments), we compute its gradient vector $\bar{m}\left(\hat{\rho}_{n,\tau}, \theta\right)$ as follows

$$\bar{m}\left(\hat{\rho}_{n,\tau}, \theta\right) = \sum_{d=1}^{m} \left(\hat{\rho}_{n,\tau}(d) - \rho_{\tau,\theta}(d)\right) \nabla_\theta \rho_{\tau,\theta}(d).$$

$\hat{\theta}_n$ satisfies the first order condition, i.e.

$$\bar{m}\left(\hat{\rho}_{n,\tau}, \theta\right) = \mathbf{0}. \tag{58}$$

Applying the Implicit Function Theorem, we get a differentiable function that is the solution of the condition in (58). Specifically, that is

$$\hat{\theta}_n = f\left(\hat{\rho}_{n,\tau}\right). \tag{59}$$

Its Jacobian matrix $Jf(\cdot)$ reads

$$Jf(\hat{\rho}_{n,\tau}, \theta) = -\left[J_\theta \bar{m}\left(\hat{\rho}_{n,\tau}, \theta\right)\right]^{-1} J_{\hat{\rho}_{n,\tau}} \bar{m}\left(\hat{\rho}_{n,\tau}, \theta\right), \tag{60}$$

where

$$J_{\hat{\rho}_{n,\tau}} \bar{m}\left(\hat{\rho}_{n,\tau}, \theta\right) := \left[\nabla_\theta \rho_{\tau,\theta}(1) \middle| \dots \middle| \nabla_\theta \rho_{\tau,\theta}(m)\right], \tag{61}$$

$$J_\theta \bar{m}\left(\hat{\rho}_{n,\tau}, \theta\right) := -\sum_{d=1}^{m} \left(\nabla_\theta \rho_{\tau,\theta}(d) \left[\nabla_\theta \rho_{\tau,\theta}(d)\right]^\top\right) + \sum_{d=1}^{m} \left(\hat{\rho}_{n,\tau}(d) - \rho_{\tau,\theta}(d)\right) H_\theta \rho_{\tau,\theta}(d) \tag{62}$$

while $H_\theta \rho_{\tau,\theta}(d)$ is the Hessian matrix of the function $\rho_{\tau,\theta}(d)$ with respect to θ.

Observe that

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \to \mathcal{N}_{p+q}\left(\mathbf{0}, \Sigma_\theta\right), \text{ as } n \to +\infty \tag{63}$$

where

$$\Sigma_\theta = Jf\left(\rho_\tau, \theta\right) \Sigma_\rho \left[Jf\left(\rho_\tau, \theta\right)\right]^\top. \tag{64}$$

In the second step, we estimate the baseline intensity parameter μ using the analytical first moment. Indeed, by inverting the equation in (22), we get

$$\mu\left(\mathbb{E}\left[\Delta_\tau N_\infty\right],\theta\right)=\frac{\mathbb{E}\left[\Delta_\tau N_\infty\right]}{1-\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}}. \tag{65}$$

The quantity $\hat{\mu}_n$ is a plug-in estimator, where instead of $\mathbb{E}\left[\Delta_\tau N_\infty\right]$, we consider its sample version $\widehat{\Delta_\tau N}_n$; we also substitute $\theta=(a,b)$ with $\hat{\theta}_n=\left(\hat{a}_n,\hat{b}_n\right)$.

To get the asymptotic distribution of $\hat{\theta}_{0,n}:=\left[\hat{\mu}_n,\hat{\theta}_n\right]^\top$, we need to determine the asymptotic distribution of $\hat{\theta}_{1,n}:=\left[\widehat{\Delta_\tau N}_n,\hat{\theta}_n\right]^\top$ that results to be

$$\sqrt{n}\left[\begin{pmatrix}\widehat{\Delta_\tau N}_n\\ \hat{\theta}_n\end{pmatrix}-\begin{pmatrix}\mathbb{E}\left[\Delta_\tau N_\infty\right]\\ \theta\end{pmatrix}\right]\to\mathcal{N}_{p+q+1}\left(\mathbf{0},\Sigma_{\theta_1}\right)\text{ as }n\to+\infty \tag{66}$$

where

$$\Sigma_{\theta_1}=\begin{bmatrix}1 & \mathbf{0}_{1\times m}\\ \mathbf{0}_{1\times(p+q)} & Jf\left(\rho_\tau,\theta\right)\end{bmatrix}\Sigma_\chi\begin{bmatrix}1 & \mathbf{0}_{1\times m}\\ \mathbf{0}_{1\times(p+q)} & Jf\left(\rho_\tau,\theta\right)\end{bmatrix}^\top \tag{67}$$

with Σ_χ as in (54). The asymptotic distribution of $\hat{\theta}_{0,n}$ can be obtained straightforwardly using the results in (65) and in (67). Indeed

$$\sqrt{n}\left(\hat{\theta}_{0,n}-\theta\right)\to\mathcal{N}_{p+q+1}\left(\mathbf{0},\Sigma_{\theta_0}\right)\text{ as }n\to+\infty, \tag{68}$$

with

$$\Sigma_{\theta_0}=\begin{bmatrix}\nabla\mu\left(\mathbb{E}\left[\Delta_\tau N_\infty\right],\theta\right)^\top\\ \mathbf{I}\end{bmatrix}\Sigma_{\theta_1}\begin{bmatrix}\nabla\mu\left(\mathbb{E}\left[\Delta_\tau N_\infty\right],\theta\right)^\top\\ \mathbf{I}\end{bmatrix}^\top. \tag{69}$$

All partial derivatives used for the computation of the asymptotic behavior of the parameter estimators involve the parameters differentiation of a matrix exponential. This can be easily done using the procedure proposed in Tsai and Chan (2003) (see Das et al., 2022, for recent developments).

5. Empirical analysis

In this section we perform two estimation exercises using real data, showing how CARMA(p,q)-Hawkes models may find applications in various areas. In the first case study, we consider the occurrences of earthquake events with timestamp values accurate down to the second. Indeed an insurance company, could be interested in an accurate modeling of time arrivals of new events as a consequence of large-magnitude earthquakes in order to improve the forecasting of future losses. Given exact timestamps, we estimate model parameters based on the likelihood function (Section 3.2) in which we select optimal p and q orders for the intensity process.

The second case study considers intra-day orders of an Italian government bond indexed to the Italian inflation rate received during the first day of placement period (October 2, 2023) reserved to individual investors. The security in question is the “BTP Valore Sc Oct28 Eur” with ISIN IT0005565400, which has quarterly coupons with a “step-up” mechanism. Data are recorded in equidistant intervals of one minute, allowing for cumulative indistinguishable events, and thus the estimation of model parameters is based on the minimization of the squared distance between empirical and theoretical autocorrelation as discussed in Section 4.3.

5.1. Estimation procedure using the likelihood function

For the estimation procedure based on the likelihood function (34), we use a data set composed of earthquake events registered on the coast of Ancona (central-eastern Italian coast) in the period January 2, 1982 to January 2, 2023.[6] In Fig. 1 we report the events that define the counting process, observing that the coast of Ancona experienced two large-magnitude earthquakes in 2016 and 2022 followed respectively by subsequent events of smaller magnitude.

The estimation procedure with the selection of optimal p and q orders for the intensity process distinguishing for nested and non-nested models is performed as follows. The first candidate considered, which is also a natural choice of starting point, is a CARMA(1,0)-Hawkes that is compared with a CARMA(2,0)-Hawkes (i.e., the closest nested model in terms of the p and q orders) using the likelihood-ratio (LR) test. In the case that null hypothesis cannot be accepted at a desired significance level (in our case 5%), the procedure considers the next couple of models by increasing the order of p and/or q up to a fixed autoregressive order $\bar{p}$ until the null hypothesis fails to be rejected. Following this strategy for a fixed $\bar{p}=3$, the best fitting model in the subset of nested CARMA(p,q)-Hawkes processes is identified (see Table 1a); i.e., the CARMA(2,1)-Hawkes.

Then the selection is carried out for the case of non-nested models through the Akaike Information Criterion (AIC) and the Bayes Information Criterion (BIC); e.g., CARMA(2,1)-Hawkes and CARMA(3,0)-Hawkes (see Table 1b). From the combined results we observe that a CARMA(2,1)-Hawkes is the most appropriate model within the CARMA(p,q)-Hawkes family up to a fixed autoregressive order $\bar{p}=3$ for describing earthquake-time arrivals in the geographic area under investigation. Table 2 displays estimated parameters and standard errors of the best fitting CARMA(2,1)-Hawkes model.

To establish if the collected data are properly described by the estimated CARMA(2,1)-Hawkes process we implement the residual analysis discussed in Ogata (1988). In practice, the estimated residuals $\left\{\tau_i\right\}_{i=1,\dots,n}$ of a point process are defined as $\tilde{\tau}_i:=\int_0^{t_i}\tilde{\lambda}_t\mathrm{d}t$ where $\left\{t_i\right\}_{i=1,\dots,n}$ denote observed event times and $\tilde{\lambda}_t$ is the estimated intensity. A given model is appropriate for reproducing the time arrivals $\left\{t_i\right\}_{i=1,\dots,n}$ if the new counting

[6] For sake of clarity, an event is qualified as an earthquake if the seismograph records a movement of at least two magnitudes in the Richter scale. Data are downloaded from https://terremoti.ingv.it/.

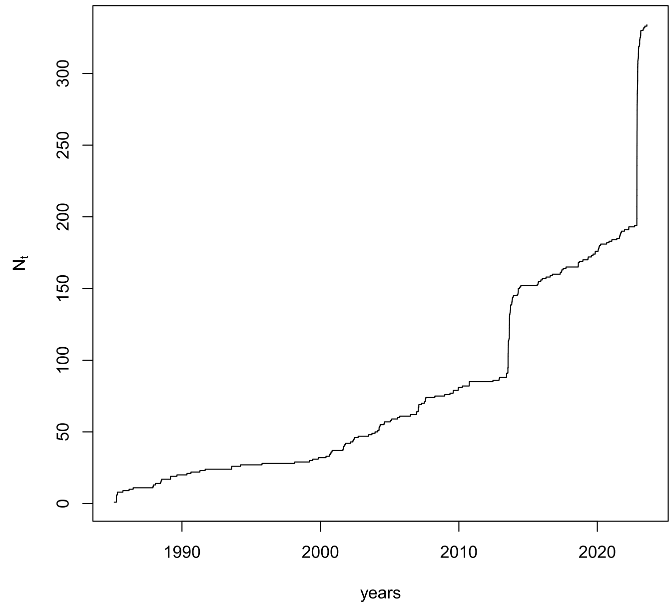

Fig. 1. Earthquake events on the coast of Ancona from January 2, 1985 to January 2, 2023.

Table 1
Estimation procedure of CARMA(p,q)-Hawkes models applied to the sequence of earthquake events with $p \leq 3$.

Model1	Model2	LR test	Best fitting
CH(1,0)	CH(2,0)	4.94 (0.0262)	CH(2,0)
CH(2,0)	CH(2,1)	112.46 (<0.0001)	CH(2,1)
CH(2,1)	CH(3,1)	1.63 (0.2087)	CH(2,1)

(a) LR test and corresponding p-value for nested models.

Model	AIC	BIC
CARMA(2,1)-Hawkes	1666	1685
CARMA(3,0)-Hawkes	1886	1691

(b) Fitting results for non-nested models.

Table 2
Estimated parameters (est. par.) and log-likelihood (loglik.) for the CARMA(2,1)-Hawkes model using MLE for earthquake events on the coast of Ancona.

	μ_0	a_1	a_2	b_0	b_1
est. par.	6.8532E-03	4.8540	0.2643	0.1874	1.7363
s.e.	7.9593E-04	1.6762	0.1411	0.0997	0.5009
loglik.	-828.03				

process $\tilde{N}_t := \sum_i \mathbb{1}_{\tilde{\tau}_i \leq t}$ results to be a homogeneous Poisson with intensity equal to one. Therefore, the estimated increments $\tilde{\tau}_i - \tilde{\tau}_{i-1}$ should be modeled with an exponential random variable with rate equal to one.

We apply two statistical tests to the increments of the fitted CARMA(2,1)-Hawkes model: the Kolmogorov-Smirnov (KS) test, as done in Ogata (1988), and the Anderson-Darling (AD) test as is more sensitive on the tails. In both cases we obtain a p-value greater than 5% (respectively 40% for the KS-test and 7.23% for the AD-test), confirming the appropriateness of the fitted CARMA(2,1)-Hawkes in modeling earthquake-time arrivals in the coast of Ancona.

5.2. Estimation procedure using the autocorrelation function

In the second estimation case study, data of orders collected during the first day of placement period regarding the Italian government bond are recorded in time intervals of length one minute, allowing for the presence of indistinguishable multiple events. In this setting, the only viable solution for the estimation of model parameters is minimizing the squared distance between empirical and theoretical autocorrelation functions of the number of jumps within intervals of the same length. Furthermore, in this context, it is essential to determine the optimal p and q orders and the number of autocorrelation lags to use in the minimization problem. Here, we use a graphical approach, reporting the empirical and theoretical autocorrelation function (acf) as shown in Fig. 2. The choice of the number of lags in the acf is done applying the same rule in the `acf` R function. Specifically, we consider lags that do not exceed the integer part of $10 \log_{10}\left(\bar{N}\right)$ with $\bar{N}$ being number of observations and, in this case, the maximum number of considered lags is equal to 26. Applying this idea, the best fitting model is the CARMA(2,1)-Hawkes (red line) which seems to fit better the curvature of the empirical acf with respect to the Hawkes with exponential kernel (blue line).

Using the estimation procedure discussed in Section 4.3 and the result in (68), we report in Table 3 respectively the estimated parameters and the asymptotic standard errors for the CARMA(2,1)-Hawkes model.

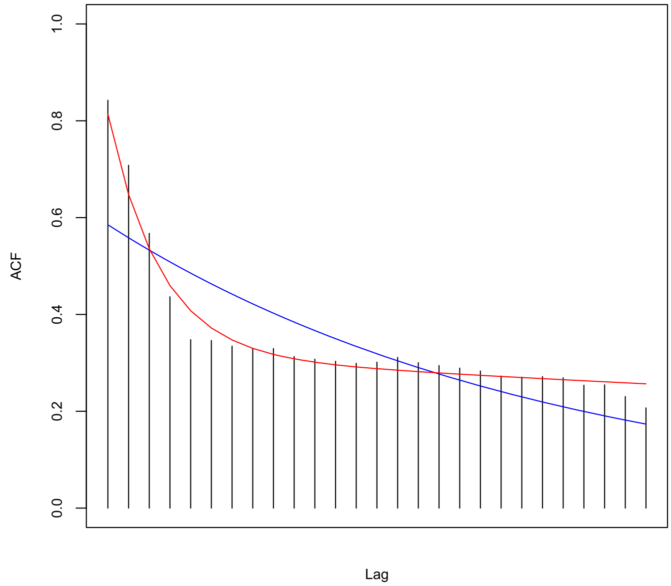

Fig. 2. Empirical and estimated autocorrelation functions. The red line is the estimated autocorrelation function of a CARMA(2,1)-Hawkes model and the blue line is the autocorrelation function of a standard Hawkes with exponential kernel. Vertical lines refer to empirical values of the autocorrelation function. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Table 3
Estimated parameters (est. par.) and asymptotic standard errors (s.e.) of a CARMA(2,1)-Hawkes fitted to the data set composed of orders arrivals recorded on October 2, 2023 for the Italian government bond “BTP Valore Sc Oct28 Eur”.

	μ_0	a_1	a_2	b_0	b_1
est. par.	6.2347	3.9375	0.1485	0.1453	3.5309
s.e.	(5.5106)	(1.9336)	(0.1248)	(0.1155)	(2.0412)

6. Conclusion

In this paper we introduce a Hawkes process where the intensity is a CARMA(p, q) model. We analyze the statistical properties of this process and obtain a closed-form expression for the autocorrelation function of the number of jumps observed in non-overlapping time intervals of the same length. The model is a generalization of the standard Hawkes with exponential kernel but it is able to reproduce more complex dependence structures observed in physical events or in finance.

CRediT authorship contribution statement

Lorenzo Mercuri: Conceptualization, Data curation, Funding acquisition, Methodology, Software, Supervision, Writing – original draft. **Andrea Perchiazzo:** Investigation, Methodology, Software, Writing – original draft, Writing – review & editing, Conceptualization. **Edit Rroji:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Writing – original draft, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare no competing interest.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by JST CREST Grant Number JPMJCR2115, Japan and by the PRIN2022 project “The effects of climate change in the evaluation of financial instruments” financed by the “Ministero dell’Università e della Ricerca” with grant number 20225PC98R, CUP Code: H53D23002200006.

Appendix A. Integration of matrix exponentials

Let $\mathbf{A}$ be a square matrix and $\mathbf{A}^{(i)} := \underbrace{\mathbf{A}\mathbf{A}\cdots\mathbf{A}}_{i \text{ times}}$. As the exponential of the matrix $\mathbf{A}$ can be computed as

$$\exp(\mathbf{A}t) = \mathbf{I} + \sum_{i=1}^{+\infty} \frac{A^{(i)}t^i}{i!},$$

it is straightforward to show that

$$\int_{t_0}^{T} e^{\mathbf{A}(T-t)}\mathrm{d}t = \mathbf{A}^{-1}\left(e^{\mathbf{A}(T-t_0)} - \mathbf{I}\right) = \left(e^{\mathbf{A}(T-t_0)} - \mathbf{I}\right)\mathbf{A}^{-1}. \tag{A.1}$$

Appendix B. Solution of a general linear ordinary differential equation

To solve $\mathrm{d}Y_t = \left(b_t + AY_t\right)\mathrm{d}t$, we consider the transformation $X_t = e^{-At}Y_t$ and observe that

$$\mathrm{d}X_t = -Ae^{-At}Y_t\mathrm{d}t + e^{-At}\mathrm{d}Y_t = e^{-At}b_t\mathrm{d}t,$$

from where we have $X_T = X_{t_0} + \int_{t_0}^{T} e^{-At}b_t\mathrm{d}t$ that in terms of Y_t reads

$$Y_T = e^{A(T-t_0)}Y_{t_0} + \int_{t_0}^{T} e^{A(T-t)}b_t\mathrm{d}t. \tag{B.1}$$

Appendix C. Computation of integrals with matrix exponentials

Some useful results for computing integrals that involve matrix exponentials are provided in Van Loan (1978) and Carbonell et al. (2008). In particular, we recall the result that deals with the computation of the following two integrals:

$$\int_{0}^{t} e^{\mathbf{H}_{11}(t-u)}\mathbf{H}_{12}e^{\mathbf{H}_{22}u}\mathrm{d}u \tag{C.1}$$

$$\int_{0}^{t}\int_{0}^{u} e^{\mathbf{H}_{11}(t-u)}\mathbf{H}_{12}e^{\mathbf{H}_{22}(u-r)}\mathbf{H}_{23}e^{\mathbf{H}_{33}r}\mathrm{d}r\mathrm{d}u \tag{C.2}$$

where $\mathbf{H}_{11}$, $\mathbf{H}_{12}$, $\mathbf{H}_{22}$, $\mathbf{H}_{23}$ and $\mathbf{H}_{33}$ have dimension $d_1 \times d_1$, $d_1 \times d_2$, $d_2 \times d_2$, $d_2 \times d_3$ and $d_3 \times d_3$, respectively. We need to define a block triangular matrix $\mathbf{H}$ as follows

$$\mathbf{H} := \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{22} & \mathbf{H}_{23} \\ \mathbf{0} & \mathbf{0} & \mathbf{H}_{33} \end{pmatrix}. \tag{C.3}$$

The integrals (C.1) and (C.2) coincide with the elements $\mathbf{B}_{12}(t)$ and $\mathbf{B}_{13}(t)$ in the matrix exponential:

$$e^{\mathbf{H}t} = \begin{pmatrix} \mathbf{B_{11}}(t) & \mathbf{B_{12}}(t) & \mathbf{B_{13}}(t) \\ \mathbf{0} & \mathbf{B_{22}}(t) & \mathbf{B_{23}}(t) \\ \mathbf{0} & \mathbf{0} & \mathbf{B_{33}}(t) \end{pmatrix} \tag{C.4}$$

while $\mathbf{B}_{11}(t) := e^{\mathbf{H}_{11}t}$, $\mathbf{B}_{22}(t) := e^{\mathbf{H}_{22}t}$ and $\mathbf{B}_{33}(t) := e^{\mathbf{H}_{33}t}$.

Remark 4. The eigenvalues of $\mathbf{H}$ coincide with the eigenvalues of $\mathbf{H}_{11}$, $\mathbf{H}_{22}$ and $\mathbf{H}_{33}$. If the real part of all eigenvalues of $\mathbf{H}_{11}$, $\mathbf{H}_{22}$ and $\mathbf{H}_{33}$ is negative, the following result holds

$$\lim_{t\to+\infty} e^{\mathbf{H}t} = \mathbf{0}$$

that implies

$$\lim_{t\to+\infty} \mathbf{B}_{12}(t) = \mathbf{0} \tag{C.5}$$

and

$$\lim_{t\to+\infty} \mathbf{B}_{13}(t) = \mathbf{0}. \tag{C.6}$$

Appendix D. Proofs of theorems in Section 3

D.1. Proof of Theorem 1

Proof. To show the Markov property for X_t in (7), we rewrite it as

$$X_{t+u} = \int_{[0,t)} e^{\mathbf{A}(u)+\mathbf{A}(t-s)}\mathbf{e}\mathrm{d}N_s + \phi(t,u)$$

where $\phi(t,u) := \int_{[t,t+u)} e^{\mathbf{A}(t+u-s)}\mathbf{e}\mathrm{d}N_s$. We recall that if H and K are two square matrices that commute, i.e. $HK = KH$, then the following result holds $e^{H+K} = e^H e^K$. We then consider the quantities $H = \mathbf{A}(u)$ and $K = \mathbf{A}(t-s)$ and notice that they commute. Thus:

$$X_{t+u} = e^{\mathbf{A}(u)} X_t + \phi(t,u). \tag{D.1}$$

Now, we show that the conditional distribution of $\phi(t,u)$ given the information at time t does not depend on history before t of the state process and counting process. Let us rewrite $\phi(t,u)$ as

$$\phi(t,u) = \int_{[0,u)} e^{\mathbf{A}(u-s)}\mathbf{e}\mathrm{d}\hat{N}_s$$

where for fixed t and $\forall u > 0$ the quantities

$$\hat{N}_u := N_{t+u} - N_t \tag{D.2}$$

coincide with the increments of the counting process over time intervals with extremes t and $t+u$, i.e. on a right-shifted time axis with the new origin corresponding to the current time t.
At this stage, we write the counting process N_t in terms of a Poisson random measure. Let M be a Poisson random measure on $\mathbb{R}_+ \times E$ where $\mathbb{R}_+$ refers to time and E to some physical space of events.[7] From Theorem 6.11 page 302 in Cinlar (2011) a counting process N_t with predictable intensity λ_t as the one defined as in (5), has the following (pathwise a.s.) form:

$$N_t(\omega) = \int_{[0,t]\times\mathbb{R}_+} \mathbb{1}_{\{(0,\lambda_s(\omega)]\}}(z)\, M(\mathrm{d}s,\mathrm{d}z).$$

Considering the increments $\hat{N}_u$ in (D.2) $\forall u \geq 0$, we get:

$$\hat{N}_u(\omega) = \int_{[t,t+u]\times\mathbb{R}_+} \mathbb{1}_{\{(0,\lambda_s(\omega)]\}}(z)\, M(\mathrm{d}s,\mathrm{d}z).$$

Now we define a shifted time axis such that $[t, t+u] \to [0,u]$, and notice that:

$$\hat{N}_u(\omega) = \int_{[0,u]\times\mathbb{R}_+} \mathbb{1}_{\{(0,\hat{\lambda}_s(\omega)]\}}(z)\, \hat{M}(\mathrm{d}s,\mathrm{d}z), \tag{D.3}$$

where $\hat{\lambda}_s(\omega)$ on the new axis coincides a.s. with $\lambda_{t+s}(\omega)$ on the original time axis while $\hat{M}$ is a (shifted) Poisson random measure.[8] The increments in (D.3) allow us to rewrite the process X_{t+u} in (D.1) as a process $\hat{X}_u$ with dynamics:

$$\hat{X}_u = e^{\mathbf{A}u}\hat{X}_0 + \int_{[0,u)} e^{\mathbf{A}(u-s)}\mathrm{d}\hat{N}_s, \quad \text{where} \quad \hat{X}_0 = X_t \text{ a.s.}, \tag{D.4}$$

and

$$\hat{\lambda}_u = \mu + \mathbf{b}^\top \hat{X}_u$$

which concludes the proof for the Markov property for X_t. Notice that, on the new axis, $\hat{X}_u$ has a similar form as (7) and the additional term $e^{\mathbf{A}u}\hat{X}_0$ is known at time t since $\hat{X}_0 = X_t$ almost surely. The integral in (D.4) is controlled only by $\{\hat{\lambda}_s\}_{0\leq s\leq u} := \{\lambda_{t+s}\}_{0\leq s\leq u}$, it does not depend to the past information up to t.
The Markov property for the vector process $[X_t, N_t]_{(p+1)\times 1}$ can be proved with similar steps as

$$\begin{bmatrix} X_{t+u} \\ N_{t+u} \end{bmatrix} = \begin{bmatrix} e^{\mathbf{A}u} & \mathbf{0}_{p\times 1} \\ \mathbf{0}_{1\times p} & 1 \end{bmatrix} \begin{bmatrix} X_t \\ N_t \end{bmatrix} + \int_t^{t+u} \begin{bmatrix} e^{\mathbf{A}(t+u-s)}\mathbf{e} \\ 1 \end{bmatrix} \mathrm{d}N_s. \tag{D.5}$$

Indeed, the result in (D.5) has the same structure as in (D.1) and, to compute its conditional distribution given the information at t, we need only the column vector $[X_t, N_t]$. This concludes the proof of the Markov property for $[X_t, N_t]$ and the whole proof. □

[7] See Definition 6.1, p. 299 in Cinlar (2011): the Poisson random measure M asserts the independence of the future of M from its past.

[8] If $\forall u \geq 0$ and $\forall z \in E$ we define $\hat{M}(u,z) := M(t+u,z)$, $\hat{M}$ is still a Poisson random measure independent of $\mathcal{F}_t$ and has the same law as M.

D.2. Proof of Theorem 2

Proof. For a non-negative kernel function, the stationary condition in (3) for a CARMA(p,q)-Hawkes process becomes

$$\int_0^{+\infty} \mathbf{b}^\top e^{\mathbf{A}t}\mathbf{e}\mathrm{d}t = \lim_{T\to+\infty}\int_0^{T} \mathbf{b}^\top e^{\mathbf{A}t}\mathbf{e}\mathrm{d}t = \lim_{T\to+\infty} \mathbf{b}^\top \mathbf{A}^{-1}\left(e^{\mathbf{A}T}-\mathbf{I}\right)\mathbf{e}, \tag{D.6}$$

where $\mathbf{I}$ is the identity matrix with dimension p. As $\mathbf{A}$ is diagonalizable,

$$e^{\mathbf{A}T} = \mathbf{S}e^{\Lambda T}\mathbf{S}^{-1}$$

where $\Lambda := \mathsf{diag}\left(\tilde{\lambda}_1,\dots,\tilde{\lambda}_p\right)$. Thus the limit in (D.6) is

$$\lim_{T\to+\infty} \mathbf{b}^\top \mathbf{A}^{-1}\left(e^{\mathbf{A}T}-\mathbf{I}\right)\mathbf{e} = \mathbf{b}^\top \mathbf{A}^{-1}\left[\mathbf{S}\left(\lim_{T\to+\infty} e^{\Lambda T}\right)\mathbf{S}^{-1}-\mathbf{I}\right]\mathbf{e}.$$

Recalling that all eigenvalues of $\mathbf{A}$ have negative real part, we notice that $e^{\Lambda T}$ tends to a $p\times p$ zero matrix. The integral in (D.6) becomes

$$\int_0^{+\infty} \mathbf{b}^\top e^{\mathbf{A}t}\mathbf{e}\mathrm{d}t = -\mathbf{b}^\top\mathbf{A}^{-1}\mathbf{e}. \tag{D.7}$$

The stationarity condition in (3) implies $-\mathbf{b}^\top\mathbf{A}^{-1}\mathbf{e} < 1$. □

D.3. Proof of Theorem 3

Proof. Let us consider two cases. If $N_{T+h} - N_T = 0$, the vector $X_T = \left[X_{1,t},\dots,X_{p,t}\right]^\top$ becomes $X_{T+h} = X^{\mathrm{NJ}}_{T+h}$ where X^{NJ}_{T+h} means no jump (NJ) occurred in the interval $(T, T+h]$ and can be written in the following way

$$X^{\mathrm{NJ}}_{T+h} = e^{\mathbf{A}(T+h-t_0)}X_{t_0} + \int_{[t_0,T)} e^{\mathbf{A}(T+h-t)}\mathbf{e}\mathrm{d}N_t$$

as the quantity $\int_{[T,T+h)} e^{\mathbf{A}(T+h-t)}\mathbf{e}\mathrm{d}N_t$ is zero due to the absence of jumps in the interval $(T, T+h]$. From

$$X^{\mathrm{NJ}}_{T+h} = e^{\mathbf{A}h}\left[e^{\mathbf{A}(T-t_0)}X_{t_0} + \int_{[t_0,T)} e^{\mathbf{A}(T-t)}\mathbf{e}\mathrm{d}N_t\right] = e^{\mathbf{A}h}X_T$$

we have that

$$\lim_{h\to0} X^{\mathrm{NJ}}_{T+h} = X_T. \tag{D.8}$$

If $N_{T+h} - N_T = 1$ then $X_{T+h} := X^{\mathrm{1J}}_{T+h}$ is computed as

$$X^{\mathrm{1J}}_{T+h} = e^{\mathbf{A}(T+h-t_0)}X_{t_0} + \int_{[t_0,T)} e^{\mathbf{A}(T+h-t)}\mathbf{e}\mathrm{d}N_t + \int_{[T,T+h)} e^{\mathbf{A}(T+h-t)}\mathbf{e}\mathrm{d}N_t.$$

Defining the jump time T_h in the time interval $(T, T+h]$ we get

$$\int_{[T,T+h)} e^{\mathbf{A}(T+h-t)}\mathbf{e}\mathrm{d}N_t = e^{\mathbf{A}(T+h-T_h)}\mathbf{e}.$$

As $\lim_{h\to0} T_h = T$, we observe that

$$\lim_{h\to0} X^{\mathrm{1J}}_{T+h} = \left[e^{\mathbf{A}(T-t_0)}X_{t_0} + \int_{[t_0,T)} e^{\mathbf{A}(T-t)}\mathbf{e}\mathrm{d}N_t\right] + \mathbf{e} = X_T + \mathbf{e}. \tag{D.9}$$

Note that $X_t + \mathbf{e} = \left[X_{t,1},\dots,X_{t,p}+1\right]^\top$ and consider the following quantity:

$$\begin{aligned}\mathbb{E}\left[f\left(X_{1,t+h},\dots,X_{p,t+h},N_{t+h}\right)|\mathcal{F}_t\right] &= f\left(X^{\mathrm{NJ}}_{1,t+h},\dots,X^{\mathrm{NJ}}_{p,t+h},N_t\right)\left(1-\lambda_t h\right)\\ &\quad + f\left(X^{\mathrm{1J}}_{1,t+h},\dots,X^{\mathrm{1J}}_{p,t+h},N_t+1\right)\lambda_t h + o(h).\end{aligned}$$

The infinitesimal generator is:

$$\mathcal{A}f_t := \lim_{h\to0}\frac{\mathbb{E}\left[f\left(X_{1,t+h},\dots,X_{p,t+h},N_{t+h}\right)|\mathcal{F}_t\right] - f\left(X_{1,t},\dots,X_{p,t},N_t\right)}{h}$$

$$= \lim_{h\to 0} \lambda_t \left[f\left(X^{1J}_{1,t+h},\dots,X^{1J}_{p,t+h},N_t+1\right) - f\left(X^{NJ}_{1,t+h},\dots,X^{NJ}_{p,t+h},N_t\right)\right] + \lim_{h\to 0} \frac{f\left(X^{NJ}_{1,t+h},\dots,X^{NJ}_{p,t+h},N_t\right) - f\left(X_{1,t},\dots,X_{p,t},N_t\right)}{h}.$$

From (D.8) and (D.9) we obtain

$$\mathcal{A} f_t := \lambda_t \left[f\left(X_{1,t},\dots,X_{p,t}+1,N_t+1\right) - f\left(X_{1,t},\dots,X_{p,t},N_t\right)\right] + \lim_{h\to 0} \frac{f\left(X^{NJ}_{1,t+h},\dots,X^{NJ}_{p,t+h},N_t\right) - f\left(X_{1,t},\dots,X_{p,t},N_t\right)}{h}. \tag{D.10}$$

To compute the limit (D.10) we use De l'Hôpital's rule

$$\begin{aligned}\lim_{h\to 0} \sum_{i=1}^{p} \frac{\partial f}{\partial X^{NJ}_{i,t+h}} \frac{\partial X^{NJ}_{i,t+h}}{\partial h} &= \lim_{h\to 0} \left[\frac{\partial f}{\partial X^{NJ}_{1,t+h}},\dots \frac{\partial f}{\partial X^{NJ}_{p,t+h}}\right] \mathbf{A} e^{\mathbf{A}h} X_t \\ &= \sum_{i=1}^{p-1} \frac{\partial f}{\partial X_{i,t}} X_{i+1,t} + \frac{\partial f}{\partial X_{p,t}} \mathbf{A}_{[p,]} X_t,\end{aligned} \tag{D.11}$$

and substituting (D.11) in (D.10), we finally obtain the result in (12). □

D.4. Proof of Theorem 4

Proof. Proof of Theorem 4. To determine the expected number of jumps in (16) we obtain first the infinitesimal generator of the function $f\left(X_{1,t},\dots,X_{p,t},N_t\right) = N_t$, that is $\mathcal{A} f_t = \lambda_t$ where the conditional intensity λ_t is defined in (7). Applying the Dynkin's formula in (14) we obtain the following ODE

$$\mathrm{d}\mathbb{E}_{t_0}\left[N_t\right] = \left[\mu + \mathbf{b}^\top \mathbb{E}_{t_0}\left(X_t\right)\right] \mathrm{d}t. \tag{D.12}$$

Then, we compute $\mathbb{E}_{t_0}\left[X_t\right]$ that requires a system of infinitesimal generators. In particular, for $i = 1,\dots,p-1$, we have

$$\mathcal{A} X_{t,i} = X_{t,i+1}$$

and

$$\mathcal{A} X_{t,p} = \left(\mu + \mathbf{b}^\top X_t\right) + \mathbf{A}_{[p,\cdot]} X_t = \mu + \sum_{i=1}^{p} \left(b_{i-1} - a_{p+1-i}\right) X_{t,i}.$$

Applying (14), we get

$$\mathrm{d}\mathbb{E}_{t_0}\left[X_t\right] = \left(\tilde{\mathbf{A}} \mathbb{E}_{t_0}\left[X_t\right] + \mu \mathbf{e}\right) \mathrm{d}t \tag{D.13}$$

where $\tilde{\mathbf{A}}$ is defined in (15). With the initial condition X_{t_0}, the solution of the system in (D.13) is (17). Substituting (17) in (D.12) we obtain the following ODE for the expected number of jumps

$$\mathrm{d}\mathbb{E}_{t_0}\left[N_t\right] = \left[\mu\left(1 - \mathbf{b}^\top \tilde{\mathbf{A}}^{-1} \mathbf{e}\right) + \mathbf{b}^\top e^{\tilde{\mathbf{A}}(t-t_0)} \left[X_{t_0} + \tilde{\mathbf{A}}^{-1} \mathbf{e}\mu\right]\right] \mathrm{d}t$$

whose solution is in (16) with initial condition N_{t_0}. Using the result in (16) we observe by straightforward calculations that the expected number of jumps in an interval of length τ reads as in (21). Due to the negativity assumption for the real part of the eigenvalues of matrix $\tilde{\mathbf{A}}$, we obtain the asymptotic behavior in (20) and (22) as $\lim_{T\to+\infty} e^{\tilde{\mathbf{A}}T} = \mathbf{0}$ where $\mathbf{0}$ is a $p \times p$ zero matrix (see (C.5)). □

Appendix E. Proofs of theorems in Section 4

E.1. Proof of Lemma 1

Proof. Using the definition of matrix $\mathbf{B}$ in (36), the identity in (39) is straightforward. To show the result in (40), we need first to compute the infinitesimal generator for each component of $vlt\left(X_t X_t^\top\right)$. From the definition in (38) we identify p blocks where the dimension of each block decreases by one unit. More precisely, the $j-$th block has $p-j+1$ elements. Considering the first block (i.e., $j=1$) we have p infinitesimal generators obtained applying the result in (11) of Theorem 3. For the first element in the first block, we have $\mathcal{A} X^2_{t,1} = 2X_{t,2}X_{t,1}$. While for the $i-$th element in the first block with $i = 2,\dots,p-1$ we get $\mathcal{A} X_{t,i}X_{t,1} = X_{t,i}X_{t,2} + X_{t,i+1}X_{t,1}$ and finally

$$\begin{aligned}\mathcal{A} X_{t,p}X_{t,1} &= \lambda_t \left[\left(X_{t,p}+1\right)X_{t,1} - X_{t,p}X_{t,1}\right] + X_{t,p}X_{t,2} + A_{[p,\cdot]} X_t X_{t,1} \\ &= \mu X_{t,1} + X_{t,p}X_{t,2} + \left(\mathbf{b}^\top + A_{[p,\cdot]}\right) X_t X_{t,1}.\end{aligned}$$

For a generic $j-$th block, we get $p-j+1$ infinitesimal generators. In particular for $i=j$ we have $\mathcal{A} X^2_{t,j} = 2X_{t,j}X_{t,j+1}$. For $i = j+1,\dots,p-1$ we have $\mathcal{A} X_{t,i}X_{t,j} = X_{t,i}X_{t,j+1} + X_{t,j}X_{t,i+1}$ and

$$\begin{aligned}\mathcal{A}X_{t,p}X_{t,j} &= \lambda_t\left[\left(X_{t,p}+1\right)X_{t,j}-X_{t,p}X_{t,j}\right]+X_{t,p}X_{t,j+1}+A_{[p,\cdot]}X_tX_{t,j}\\ &= \mu X_{t,j}+X_{t,p}X_{t,j+1}+\left(\mathbf{b}^\top+A_{[p,\cdot]}\right)X_tX_{t,j}.\end{aligned}$$

The last block contains only one infinitesimal generator of the form

$$\begin{aligned}\mathcal{A}X_{t,p}^2 &= \lambda_t\left[\left(X_{t,p}+1\right)^2-X_{t,p}^2\right]+2A_{[p,\cdot]}X_tX_{t,p}\\ &= \mu+\mathbf{b}^\top X_t+2\mu X_{t,p}+2\left(\mathbf{b}^\top+A_{[p,\cdot]}\right)X_tX_{t,p}.\end{aligned}$$

Using the Dynkin's formula in (14) we obtain the following system of linear ODE's:

$$\mathrm{d}vlt\left(\mathbb{E}_{t_0}\left(X_tX_t^\top\right)\right)=\left[\mu\tilde{\mathbf{e}}+\tilde{\mathbf{C}}\mathbb{E}_{t_0}\left(X_t\right)+\tilde{\tilde{\mathbf{A}}}vlt\left(\mathbb{E}_{t_0}\left(X_tX_t^\top\right)\right)\right]\mathrm{d}t \tag{E.1}$$

where the $\frac{p(p+1)}{2}$ vector $\tilde{\mathbf{e}}$ is composed of zero entries except the last position where the element is one; $\tilde{\tilde{\mathbf{A}}}$ and $\tilde{\mathbf{C}}$ are defined in (35) and (37) respectively.

The first step is to solve the ODE defined in (E.1) whose solution has the following form

$$\begin{aligned}vlt\left(\mathbb{E}_{t_0}\left(X_TX_T^\top\right)\right) &= e^{\tilde{\tilde{\mathbf{A}}}(T-t_0)}vlt\left(X_{t_0}X_{t_0}^\top\right)+e^{\tilde{\tilde{\mathbf{A}}}T}\int_{t_0}^{T}e^{-\tilde{\tilde{\mathbf{A}}}t}\left[\mu\tilde{\mathbf{e}}+\tilde{\mathbf{C}}\mathbb{E}_{t_0}\left(X_t\right)\right]\mathrm{d}t\\ &= e^{\tilde{\tilde{\mathbf{A}}}(T-t_0)}vlt\left(X_{t_0}X_{t_0}^\top\right)+\left[e^{\tilde{\tilde{\mathbf{A}}}(T-t_0)}-\mathbf{I}\right]\tilde{\tilde{\mathbf{A}}}^{-1}\mu\tilde{\mathbf{e}}\\ &\quad+e^{\tilde{\tilde{\mathbf{A}}}T}\int_{t_0}^{T}e^{-\tilde{\tilde{\mathbf{A}}}t}\tilde{\mathbf{C}}\mathbb{E}_{t_0}\left(X_t\right)\mathrm{d}t.\end{aligned} \tag{E.2}$$

We also observe that

$$\begin{aligned}e^{\tilde{\tilde{\mathbf{A}}}T}\int_{t_0}^{T}e^{-\tilde{\tilde{\mathbf{A}}}t}\tilde{\mathbf{C}}\mathbb{E}_{t_0}\left(X_t\right)\mathrm{d}t &= e^{\tilde{\tilde{\mathbf{A}}}T}\int_{t_0}^{T}e^{-\tilde{\tilde{\mathbf{A}}}t}\tilde{\mathbf{C}}\left[e^{\tilde{\mathbf{A}}(t-t_0)}\left[X_{t_0}+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]-\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]\mathrm{d}t\\ &= e^{\tilde{\tilde{\mathbf{A}}}T}\int_{t_0}^{T}e^{-\tilde{\tilde{\mathbf{A}}}t}\tilde{\mathbf{C}}e^{\tilde{\mathbf{A}}t}\mathrm{d}te^{-\tilde{\mathbf{A}}t_0}\left[X_{t_0}+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]\\ &\quad-\left[e^{\tilde{\tilde{\mathbf{A}}}(T-t_0)}-\mathbf{I}\right]\tilde{\tilde{\mathbf{A}}}^{-1}\tilde{C}\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu.\end{aligned} \tag{E.3}$$

Substituting (E.3) into (E.2) we obtain the result in (40). □

E.2. Proof of Theorem 5

We provide below the proof of Theorem 5 on the long-run covariance of the number of jumps in a CARMA(p,q)-Hawkes model.

Proof. We first determine the covariance of number of jumps in two non-overlapping time intervals given the information at time t_0. This quantity is formally defined as

$$\begin{aligned}Cov_{t_0}(\tau,\delta) &:= \mathbb{E}_{t_0}\left[\left(N_{t+\tau}-N_t\right)\left(N_{t+2\tau+\delta}-N_{t+\tau+\delta}\right)\right]\\ &\quad-\ \mathbb{E}_{t_0}\left[\left(N_{t+\tau}-N_t\right)\right]\mathbb{E}_{t_0}\left[\left(N_{t+2\tau+\delta}-N_{t+\tau+\delta}\right)\right].\end{aligned} \tag{E.4}$$

Using the iteration property of the conditional expected value, (E.4) becomes

$$\begin{aligned}Cov_{t_0}(\tau,\delta) &= \mathbb{E}_{t_0}\left[\left(N_{t+\tau}-N_t\right)\mathbb{E}_{t+\tau}\left[\left(N_{t+2\tau+\delta}-N_{t+\tau+\delta}\right)\right]\right]\\ &\quad-\mathbb{E}_{t_0}\left[\left(N_{t+\tau}-N_t\right)\right]\mathbb{E}_{t_0}\left[\left(N_{t+2\tau+\delta}-N_{t+\tau+\delta}\right)\right].\end{aligned}$$

Applying the result (21) in Theorem 4, we get

$$Cov_{t_0}(\tau,\delta)=\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\left[e^{\tilde{\mathbf{A}}(\tau+\delta)}-e^{\tilde{\mathbf{A}}\delta}\right]g_{t_0}(t,\tau) \tag{E.5}$$

where

$$\begin{aligned}g_{t_0}(t,\tau) &= \mathbb{E}_{t_0}\left[\left(N_{t+\tau}-N_t\right)X_{t+\tau}\right]-e^{\tilde{\mathbf{A}}(t+\tau-t_0)}\mathbb{E}_{t_0}\left[N_{t+\tau}-N_t\right]X_{t_0}\\ &\quad+\left(\mathbf{I}-e^{\tilde{\mathbf{A}}(t+\tau-t_0)}\right)\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\mathbb{E}_{t_0}\left[N_{t+\tau}-N_t\right]\\ &= \mathbb{E}_{t_0}\left[N_{t+\tau}X_{t+\tau}\right]+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\mathbb{E}_{t_0}\left[N_t\right]-e^{\tilde{\mathbf{A}}\tau}\left[\mathbb{E}_{t_0}\left(N_tX_t\right)+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\mathbb{E}_{t_0}\left[N_t\right]\right]\\ &\quad-e^{\tilde{\mathbf{A}}(t+\tau-t_0)}\mathbb{E}_{t_0}\left[N_{t+\tau}-N_t\right]X_{t_0}+\left(\mathbf{I}-e^{\tilde{\mathbf{A}}(t+\tau-t_0)}\right)\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\mathbb{E}_{t_0}\left[N_{t+\tau}-N_t\right].\end{aligned} \tag{E.6}$$

In the *rhs* of (E.6), the last two terms are stationary due to the result in (22) and to the negativity of the real part for the eigenvalues of $\tilde{\mathbf{A}}$; the third term converges to zero as $t \to +\infty$ while the fourth term has the following limit behavior

$$\left(\mathbf{I} - e^{\tilde{\mathbf{A}}(t+\tau-t_0)}\right)\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\mathbb{E}_{t_0}\left[N_{t+\tau} - N_t\right] \to \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu^2\left(1 - \mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\tau \text{ a.s. } t \to +\infty. \tag{E.7}$$

For the first two terms in the *rhs* (E.6) consider the quantity:

$$h_{t_0}(t,\tau) := \mathbb{E}_{t_0}\left[N_{t+\tau}X_{t+\tau}\right] + \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\mathbb{E}_{t_0}\left[N_t\right],\ \forall \tau \geq 0, t > t_0 \tag{E.8}$$

as $t \to +\infty$. In (E.8) the vector $\mathbb{E}_{t_0}\left[N_t X_t\right]$ requires the calculation of p infinitesimal generators. We then observe that for $i = 1, \ldots, p-1$, the infinitesimal generator of the function $N_t X_{t,i}$ is:

$$\begin{aligned}\mathcal{A}N_t X_{t,i} &= \left(\mu + \mathbf{b}^\top X_t\right)\left[\left(N_t + 1\right)X_{t,i} - N_t X_{t,i}\right] + N_t X_{t,i+1}\\ &= \left(\mu X_{t,i} + X_{t,i}X_t^\top\mathbf{b}\right) + N_t X_{t,i+1}\end{aligned}$$

while for $i = p$

$$\begin{aligned}\mathcal{A}N_t X_{t,p} &= \left(\mu + \mathbf{b}^\top X_t\right)\left[\left(N_t + 1\right)\left(X_{t,p} + 1\right) - N_t X_{t,p}\right] + N_t A_{[p,\cdot]}X_t\\ &= \left(\mu + \mathbf{b}^\top X_t + \mu N_t\right) + \left(\mu X_{t,p} + X_{t,p}X_t^\top\mathbf{b}\right) + \left(\mathbf{b}^\top + A_{[p,\cdot]}\right)N_t X_t,\end{aligned}$$

that implies

$$\mathrm{d}\mathbb{E}_{t_0}\left[X_t N_t\right] = \left[\left(\mu + \mathbf{b}^\top\mathbb{E}_{t_0}\left[X_t\right] + \mu\mathbb{E}_{t_0}\left[N_t\right]\right)\mathbf{e} + \mu\mathbb{E}_{t_0}\left[X_t\right] + \mathbb{E}_{t_0}\left[X_t X_t^\top\right]\mathbf{b} + \tilde{\mathbf{A}}\mathbb{E}_{t_0}\left[X_t N_t\right]\right]\mathrm{d}t \tag{E.9}$$

from where we get

$$\begin{aligned}\mathbb{E}_{t_0}\left[X_T N_T\right] &= e^{\tilde{\mathbf{A}}(T-t_0)}X_{t_0}N_{t_0} + \int_{t_0}^{T} e^{\tilde{\mathbf{A}}(T-t)}\left(\mu + \mathbf{b}^\top\mathbb{E}_{t_0}\left[X_t\right] + \mu\mathbb{E}_{t_0}\left[N_t\right]\right)\mathbf{e}\mathrm{d}t\\ &\quad + \int_{t_0}^{T} e^{\tilde{\mathbf{A}}(T-t)}\left[\mu\mathbb{E}_{t_0}\left[X_t\right] + \mathbb{E}_{t_0}\left[X_t X_t^\top\right]\mathbf{b}\right]\mathrm{d}t.\end{aligned} \tag{E.10}$$

The quantity $\mathbb{E}_{t_0}\left[X_T N_T\right]$ is not stationary but it is useful as it appears in the *rhs* of the function $h_{t_0}(t,\tau)$ introduced in (E.8) that can be rewritten as

$$\begin{aligned}h_{t_0}(t,\tau) &= e^{\tilde{\mathbf{A}}(t+\tau-t_0)}X_{t_0}N_{t_0} + \int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mu\mathbf{e}\mathrm{d}u + \int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbf{b}^\top\mathbb{E}_{t_0}\left[X_u\right]\mathbf{e}\mathrm{d}u\\ &\quad + \int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\left(\mu\mathbb{E}_{t_0}\left[N_u\right]\right)\mathbf{e}\mathrm{d}u + \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\mathbb{E}_{t_0}\left[N_t\right]\\ &\quad + \int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\left[\mu\mathbb{E}_{t_0}\left[X_u\right] + \mathbb{E}_{t_0}\left[X_u X_u^\top\right]\mathbf{b}\right]\mathrm{d}u.\end{aligned} \tag{E.11}$$

We analyze the long-run behavior of each term in the *rhs* of (E.11). We first observe that

$$\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathrm{d}u\mu\mathbf{e} = \left(e^{\tilde{\mathbf{A}}(t+\tau-t_0)} - \mathbf{I}\right)\tilde{\mathbf{A}}^{-1}\mu\mathbf{e}$$

with

$$\lim_{t\to+\infty}\left(e^{\tilde{\mathbf{A}}(t+\tau-t_0)} - \mathbf{I}\right)\tilde{\mathbf{A}}^{-1}\mu\mathbf{e} = -\tilde{\mathbf{A}}^{-1}\mu\mathbf{e}. \tag{E.12}$$

The formula for the conditional expected value of the process in (17) allows us to rewrite the third term in the *rhs* of (E.11) as follows

$$\begin{aligned}\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbf{e}\mathbf{b}^\top\mathbb{E}_{t_0}\left[X_u\right]\mathrm{d}u &= e^{\tilde{\mathbf{A}}(t+\tau)}\int_{t_0}^{t+\tau} e^{-\tilde{\mathbf{A}}u}\mathbf{e}\mathbf{b}^\top e^{\tilde{\mathbf{A}}u}\mathrm{d}u e^{-\tilde{\mathbf{A}}t_0}\left[X_{t_0} + \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]\\ &\quad - \int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathrm{d}u\mathbf{e}\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\\ &= e^{\tilde{\mathbf{A}}(t+\tau)}\int_{t_0}^{t+\tau} e^{-\tilde{\mathbf{A}}u}\mathbf{e}\mathbf{b}^\top e^{\tilde{\mathbf{A}}u}\mathrm{d}u e^{-\tilde{\mathbf{A}}t_0}\left[X_{t_0} + \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]\end{aligned}$$

$$- \left(e^{\tilde{\mathbf{A}}(t+\tau-t_0)} - \mathbf{I}\right) \tilde{\mathbf{A}}^{-1}\mathbf{e}\mathbf{b}^\top \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu. \tag{E.13}$$

To compute the integral $e^{\tilde{\mathbf{A}}(t+\tau)} \int_{t_0}^{t+\tau} e^{-\tilde{\mathbf{A}}u}\mathbf{e}\mathbf{b}^\top e^{\tilde{\mathbf{A}}u}\mathrm{d}u e^{-\tilde{\mathbf{A}}t_0}$ we use the result in (C.4) and exploiting its limit behavior (C.5), the long-run behavior of (E.13) becomes

$$\lim_{t\to+\infty} \int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbf{e}\mathbf{b}^\top \mathbb{E}_{t_0}\left[X_u\right] \mathrm{d}u = \tilde{\mathbf{A}}^{-1}\mathbf{e}\mathbf{b}^\top \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu. \tag{E.14}$$

The fourth term in the *rhs* of (E.11) can be written as

$$\begin{aligned}
&\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathrm{d}u\mathbf{e}\mu N_{t_0} + \int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)} \left(u-t_0\right) \mathrm{d}u\mathbf{e}\mu^2 \left(1-\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\\
&+ \int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbf{e}\mu\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\left[e^{\tilde{\mathbf{A}}(u-t_0)}-\mathbf{I}\right]\mathrm{d}u\left(X_{t_0}+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right)+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\mathbb{E}_{t_0}\left[N_t\right]\\
&= \left(e^{\tilde{\mathbf{A}}(t+\tau-t_0)}-\mathbf{I}\right)\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu N_{t_0} + \left[\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\left(u-t_0\right)\mathrm{d}u\right]\mathbf{e}\mu^2\left(1-\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\\
&+ e^{\tilde{\mathbf{A}}(t+\tau)}\int_{t_0}^{t+\tau} e^{-\tilde{\mathbf{A}}u}\mathbf{e}\mathbf{b}^\top e^{\tilde{\mathbf{A}}u}\mathrm{d}u e^{-\tilde{\mathbf{A}}t_0}\tilde{\mathbf{A}}^{-1}\left[X_{t_0}+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]\mu\\
&- \tilde{\mathbf{A}}^{-1}\left(e^{\tilde{\mathbf{A}}(t+\tau-t_0)}-\mathbf{I}\right)\mathbf{e}\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\left[X_{t_0}+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]\mu+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\mathbb{E}_{t_0}\left[N_t\right].
\end{aligned} \tag{E.15}$$

Integrating by parts we get

$$\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\left(u-t_0\right)\mathrm{d}u = \tilde{\mathbf{A}}^{-1}\left[\left(e^{\tilde{\mathbf{A}}(t+\tau-t_0)}-\mathbf{I}\right)\tilde{\mathbf{A}}^{-1}-\mathbf{I}\left(t+\tau-t_0\right)\right].$$

Thus (E.15) becomes

$$\begin{aligned}
&\left(e^{\tilde{\mathbf{A}}(t+\tau-t_0)}-\mathbf{I}\right)\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu N_{t_0}+\tilde{\mathbf{A}}^{-1}\left[\left(e^{\tilde{\mathbf{A}}(t+\tau-t_0)}-\mathbf{I}\right)\tilde{\mathbf{A}}^{-1}-\mathbf{I}\left(t+\tau-t_0\right)\right]\mathbf{e}\mu^2\left(1-\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\\
&+ e^{\tilde{\mathbf{A}}(t+\tau)}\int_{t_0}^{t+\tau} e^{-\tilde{\mathbf{A}}u}\mathbf{e}\mathbf{b}^\top e^{\tilde{\mathbf{A}}u}\mathrm{d}u e^{-\tilde{\mathbf{A}}t_0}\tilde{\mathbf{A}}^{-1}\left[X_{t_0}+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]\mu\\
&- \tilde{\mathbf{A}}^{-1}\left(e^{\tilde{\mathbf{A}}(t+\tau-t_0)}-\mathbf{I}\right)\mathbf{e}\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\left[X_{t_0}+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]\mu+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\mathbb{E}_{t_0}\left[N_t\right].
\end{aligned}$$

Using the formula for the conditional expected value of the counting process in (16) we get

$$\begin{aligned}
&\left(e^{\tilde{\mathbf{A}}(t+\tau-t_0)}-\mathbf{I}\right)\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu N_{t_0}+\tilde{\mathbf{A}}^{-1}\left[\left(e^{\tilde{\mathbf{A}}(t+\tau-t_0)}-\mathbf{I}\right)\tilde{\mathbf{A}}^{-1}-\mathbf{I}\left(t+\tau-t_0\right)\right]\mathbf{e}\mu^2\left(1-\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\\
&+ e^{\tilde{\mathbf{A}}(t+\tau)}\int_{t_0}^{t+\tau} e^{-\tilde{\mathbf{A}}u}\mathbf{e}\mathbf{b}^\top e^{\tilde{\mathbf{A}}u}\mathrm{d}u e^{-\tilde{\mathbf{A}}t_0}\tilde{\mathbf{A}}^{-1}\left[X_{t_0}+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]\mu\\
&- \tilde{\mathbf{A}}^{-1}\left(e^{\tilde{\mathbf{A}}(t+\tau-t_0)}-I\right)\mathbf{e}\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\left[X_{t_0}+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]\mu\\
&+ \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\left[N_{t_0}+\mu\left(1-\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\left(t-t_0\right)+\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\left(e^{\tilde{\mathbf{A}}(t_1-t_0)}-\mathbf{I}\right)\left[X_{t_0}+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]\right]\\
&= e^{\tilde{\mathbf{A}}(t+\tau-t_0)}\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu N_{t_0}+\tilde{\mathbf{A}}^{-1}\left[\left(e^{\tilde{\mathbf{A}}(t+\tau-t_0)}-\mathbf{I}\right)\tilde{\mathbf{A}}^{-1}-\mathbf{I}\tau\right]\mathbf{e}\mu^2\left(1-\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\\
&+ e^{\tilde{\mathbf{A}}(t+\tau)}\int_{t_0}^{t+\tau} e^{-\tilde{\mathbf{A}}u}\mathbf{e}\mathbf{b}^\top e^{\tilde{\mathbf{A}}u}\mathrm{d}u e^{-\tilde{\mathbf{A}}t_0}\tilde{\mathbf{A}}^{-1}\left[X_{t_0}+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]\mu\\
&- \tilde{\mathbf{A}}^{-1}e^{\tilde{\mathbf{A}}(t+\tau-t_0)}\mathbf{e}\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\left[X_{t_0}+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]\mu+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}e^{\tilde{\mathbf{A}}(t-t_0)}\left[X_{t_0}+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]
\end{aligned}$$

and its long-run behavior is established considering $t \to +\infty$, that is

$$-\tilde{\mathbf{A}}^{-1}\left[\tilde{\mathbf{A}}^{-1}+\mathbf{I}\tau\right]\mathbf{e}\mu^2\left(1-\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right). \tag{E.16}$$

The fifth term in the right-hand side of (E.11) can be rewritten as

$$\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbb{E}_{t_0}\left[X_u\right]\mathrm{d}u\mu = \int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}e^{\tilde{\mathbf{A}}(u-t_0)}\mathrm{d}u\left[X_{t_0}+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]\mu$$
$$-\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathrm{d}u\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu^2$$
$$= e^{\tilde{\mathbf{A}}(t+\tau-t_0)}\left(t+\tau-t_0\right)\left[X_{t_0}+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]\mu$$
$$-\tilde{\mathbf{A}}^{-1}\left(e^{\tilde{\mathbf{A}}(t+\tau-t_0)}-\mathbf{I}\right)\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu^2,$$

that has the following long-run behavior

$$\lim_{t\to+\infty}\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbb{E}_{t_0}\left[X_u\right]\mathrm{d}u\mu = \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu^2. \tag{E.17}$$

Lemma 1 suggests that the last term in the *rhs* of (E.11) can be written as

$$\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbb{E}_{t_0}\left[X_uX_u^\top\right]\mathbf{b}\mathrm{d}u = \int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbf{B}e^{\tilde{\tilde{\mathbf{A}}}(u-t_0)}\mathrm{d}u vlt\left(X_{t_0}X_{t_0}^\top\right)$$
$$+\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbf{B}e^{\tilde{\tilde{\mathbf{A}}}(u-t_0)}\mathrm{d}u\tilde{\tilde{\mathbf{A}}}^{-1}\mu\left(\tilde{\mathbf{e}}-\tilde{\mathbf{C}}\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)$$
$$-\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathrm{d}u\mathbf{B}\tilde{\tilde{\mathbf{A}}}^{-1}\mu\left(\tilde{\mathbf{e}}-\tilde{\mathbf{C}}\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)$$
$$+\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbf{B}e^{\tilde{\tilde{\mathbf{A}}}u}\left[\int_{t_0}^{u} e^{-\tilde{\tilde{\mathbf{A}}}h}\tilde{\mathbf{C}}e^{\tilde{\mathbf{A}}h}\mathrm{d}h\right]e^{-\tilde{\mathbf{A}}t_0}\left[X_{t_0}+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]\mathrm{d}u.$$

The result in (A.1) implies that

$$\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbb{E}_{t_0}\left[X_uX_u^\top\right]\mathbf{b}\mathrm{d}u = \left[\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbf{B}e^{\tilde{\tilde{\mathbf{A}}}(u-t_0)}\mathrm{d}u\right]vlt\left(X_{t_0}X_{t_0}^\top\right)$$
$$+\left[\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbf{B}e^{\tilde{\tilde{\mathbf{A}}}(u-t_0)}\mathrm{d}u\right]\tilde{\tilde{\mathbf{A}}}^{-1}\mu\left(\tilde{\mathbf{e}}-\tilde{\mathbf{C}}\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)$$
$$-\left(e^{\tilde{\mathbf{A}}(t+\tau-t_0)}-\mathbf{I}\right)\tilde{\mathbf{A}}^{-1}\mathbf{B}\tilde{\tilde{\mathbf{A}}}^{-1}\mu\left(\tilde{\mathbf{e}}-\tilde{\mathbf{C}}\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)$$
$$+\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbf{B}e^{\tilde{\tilde{\mathbf{A}}}u}\left[\int_{t_0}^{u} e^{-\tilde{\tilde{\mathbf{A}}}h}\tilde{\mathbf{C}}e^{\tilde{\mathbf{A}}h}\mathrm{d}h\right]e^{-\tilde{\mathbf{A}}t_0}\mathrm{d}u\left[X_{t_0}+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right].$$

To determine the asymptotic behavior of this term, we analyze the long-run behavior of the integral $\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbf{B}e^{\tilde{\tilde{\mathbf{A}}}(u-t_0)}\mathrm{d}u$. Exploiting the result in Appendix C, we have

$$\lim_{t\to+\infty}\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbf{B}e^{\tilde{\tilde{\mathbf{A}}}(u-t_0)}\mathrm{d}u = \mathbf{0}$$

as all eigenvalues of $\tilde{\mathbf{A}}$ and $\tilde{\tilde{\mathbf{A}}}$ have negative real part. Using the Fubini-Tonelli's Theorem the last integral becomes

$$\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbf{B}e^{\tilde{\tilde{\mathbf{A}}}u}\left[\int_{t_0}^{u} e^{-\tilde{\tilde{\mathbf{A}}}h}\tilde{\mathbf{C}}e^{\tilde{\mathbf{A}}h}\mathrm{d}h\right]e^{-\tilde{\mathbf{A}}t_0}\mathrm{d}u = \int_{t_0}^{t+\tau}\int_{t_0}^{u} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbf{B}e^{\tilde{\tilde{\mathbf{A}}}(u-h)}\tilde{\mathbf{C}}e^{\tilde{\mathbf{A}}(h-t_0)}\mathrm{d}h\mathrm{d}u.$$

Its long-run behavior is obtained using the result in (C.6), that is

$$\lim_{t\to+\infty}\int_{t_0}^{t+\tau}\int_{t_0}^{u} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbf{B}e^{\tilde{\tilde{\mathbf{A}}}(u-h)}\tilde{\mathbf{C}}e^{\tilde{\mathbf{A}}(h-t_0)}\mathrm{d}h\mathrm{d}u = \mathbf{0}. \tag{E.18}$$

Finally, we have

$$\lim_{t\to+\infty}\int_{t_0}^{t+\tau} e^{\tilde{\mathbf{A}}(t+\tau-u)}\mathbb{E}_{t_0}\left[X_u X_u^\top\right]\mathbf{b}\mathrm{d}u = \tilde{\mathbf{A}}^{-1}\mathbf{B}\tilde{\tilde{\mathbf{A}}}^{-1}\mu\left(\tilde{\mathbf{e}}-\tilde{\mathbf{C}}\tilde{\mathbf{A}}^{-1}\mathbf{e}\right). \tag{E.19}$$

From (E.12), (E.14), (E.16), (E.17) and (E.19) we obtain the limit behavior for the quantity in (E.11)

$$\begin{aligned}
h_\infty(\tau) &:= \lim_{t\to+\infty} h_{t_0}(t,\tau)\\
&= -\tilde{\mathbf{A}}^{-1}\mu\mathbf{e}+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu-\tilde{\mathbf{A}}^{-1}\left[\tilde{\mathbf{A}}^{-1}+\mathbf{I}\tau\right]\mathbf{e}\mu^2\left(1-\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\\
&+\ \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu^2\tilde{\mathbf{A}}^{-1}\mathbf{B}\tilde{\tilde{\mathbf{A}}}^{-1}\mu\left(\tilde{\mathbf{e}}-\tilde{\mathbf{C}}\tilde{\mathbf{A}}^{-1}\mathbf{e}\right).
\end{aligned} \tag{E.20}$$

Using (E.20) we can determine the asymptotic behavior of (E.6) and we get

$$\begin{aligned}
g_\infty(\tau) &:= \lim_{t\to+\infty} g_{t_0}(t,\tau)\\
&= \lim_{t\to+\infty} h_{t_0}(t,\tau)-e^{\tilde{\mathbf{A}}\tau}\left[\lim_{t\to+\infty} h_{t_0}(t,0)\right]+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu^2\left(1-\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\tau\\
&= h_\infty(\tau)-e^{\tilde{\mathbf{A}}\tau}h_\infty(0)+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu^2\left(1-\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\tau.
\end{aligned} \tag{E.21}$$

By straightforward calculations (E.21) becomes (42) and the covariance reads as in (41). □

E.3. Proof of Theorem 6

Here we provide the proof of Theorem 6.

Proof. For the asymptotic variance we need to compute the conditional variance of the number of jumps in an interval with length τ. First we observe that

$$\sigma_{t_0}^2(t,\tau) := \mathsf{Var}_{t_0}\left(N_{t+\tau}-N_t\right)=\mathbb{E}_{t_0}\left[\left(N_{t+\tau}-N_t\right)^2\right]-\mathbb{E}_{t_0}^2\left[N_{t+\tau}-N_t\right].$$

We then compute the second moment of the increments

$$\begin{aligned}
\mathbb{E}_{t_0}\left[\left(N_{t+\tau}-N_t\right)^2\right] &= \mathbb{E}_{t_0}\left[N_{t+\tau}^2\right]+\mathbb{E}_{t_0}\left[N_t^2\right]-2\mathbb{E}_{t_0}\left[N_t\mathbb{E}_t\left[N_{t+\tau}\right]\right]\\
&= \mathbb{E}_{t_0}\left[N_{t+\tau}^2\right]-\mathbb{E}_{t_0}\left[N_t^2\right]-2\mathbb{E}_{t_0}\left[N_t\right]\mu\left(1-\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\tau\\
&-2\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\left[e^{\tilde{\mathbf{A}}\tau}-\mathbf{I}\right]\left[\mathbb{E}_{t_0}\left[N_tX_t\right]+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\mathbb{E}_{t_0}\left[N_t\right]\right].
\end{aligned}$$

For $\mathbb{E}_{t_0}\left[N_t^2\right]$ it is useful to compute the infinitesimal operator for the function $f\left(X_{1,t},\dots,X_{p,t},N_t,\right)=N_t^2$, that reads

$$\mathcal{A}f_t=\mu\left(2N_t+1\right)+2\mathbf{b}^\top N_tX_t+\mathbf{b}^\top X_t.$$

Applying the Dynkin's formula, we have

$$\mathbb{E}_{t_0}\left[N_t^2\right]=N_{t_0}^2+2\mu\int_{t_0}^{t}\mathbb{E}_{t_0}\left[N_u\right]\mathrm{d}u+\mu\left(t-t_0\right)+2\mathbf{b}^\top\int_{t_0}^{t}\mathbb{E}_{t_0}\left[N_uX_u\right]\mathrm{d}u+\mathbf{b}^\top\int_{t_0}^{t}\mathbb{E}_{t_0}\left[X_u\right]\mathrm{d}t.$$

Therefore

$$\begin{aligned}
\mathbb{E}_{t_0}\left[\left(N_{t+\tau}-N_t\right)^2\right] &= 2\mu\int_t^{t+\tau}\mathbb{E}_{t_0}\left[N_u\right]\mathrm{d}u+\mu\tau+2\mathbf{b}^\top\int_t^{t+\tau}\mathbb{E}_{t_0}\left[N_uX_u\right]\mathrm{d}u+\mathbf{b}^\top\int_t^{t+\tau}\mathbb{E}_{t_0}\left[X_u\right]\mathrm{d}u\\
&-2\mathbb{E}_{t_0}\left[N_t\right]\mu\left(1-\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\tau-2\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\left[e^{\tilde{\mathbf{A}}\tau}-\mathbf{I}\right]\left[\mathbb{E}_{t_0}\left[N_tX_t\right]+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\mathbb{E}_{t_0}\left[N_t\right]\right]\\
&= 2\mu\int_t^{t+\tau}\mathbb{E}_{t_0}\left[N_u-N_t\right]\mathrm{d}u+\mu\tau+\mathbf{b}^\top\int_t^{t+\tau}\mathbb{E}_{t_0}\left[X_u\right]\mathrm{d}u\\
&+2\mathbf{b}^\top\int_t^{t+\tau}\left[\mathbb{E}_{t_0}\left[N_uX_u\right]+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\mathbb{E}_{t_0}\left[N_t\right]\right]\mathrm{d}u\\
&-2\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\left[e^{\tilde{\mathbf{A}}\tau}-\mathbf{I}\right]\left[\mathbb{E}_{t_0}\left[N_tX_t\right]+\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\mathbb{E}_{t_0}\left[N_t\right]\right].
\end{aligned} \tag{E.22}$$

We study the asymptotic behavior of the terms in (E.22). Denoting with $a_{t_0}(t,\tau):=\int_t^{t+\tau}\mathbb{E}_{t_0}\left[N_u-N_t\right]\mathrm{d}u$, we obtain

$$a_{t_0}(t,\tau)=\int_t^{t+\tau}\mu\left(1-\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)(u-t)\,\mathrm{d}u+\int_t^{t+\tau}\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\left[e^{\tilde{\mathbf{A}}(u-t_0)}-e^{\tilde{\mathbf{A}}(t-t_0)}\right]\mathrm{d}u\left[X_{t_0}+\tilde{\mathbf{A}}^{-1}\mathbf{e}\right]$$

$$= \mu \left(1 - \mathbf{b}^\top \tilde{\mathbf{A}}^{-1}\mathbf{e}\right) \frac{\tau^2}{2} + \int_t^{t+\tau} \mathbf{b}^\top \tilde{\mathbf{A}}^{-1} \left[e^{\tilde{\mathbf{A}}(u-t)} - \mathbf{I}\right] \mathrm{d}u e^{\tilde{\mathbf{A}}\left(t-t_0\right)} \left[X_{t_0} + \tilde{\mathbf{A}}^{-1}\mathbf{e}\right].$$

We observe that the following integral is finite

$$\int_t^{t+\tau} \mathbf{b}^\top \tilde{\mathbf{A}}^{-1} \left[e^{\tilde{\mathbf{A}}(u-t)} - \mathbf{I}\right] \mathrm{d}u < +\infty$$

from where we deduce that

$$a_\infty(\tau) := \lim_{t\to+\infty} a_{t_0}(t,\tau) = \mu \left(1 - \mathbf{b}^\top \tilde{\mathbf{A}}^{-1}\mathbf{e}\right) \frac{\tau^2}{2}. \tag{E.23}$$

We then focus on the quantity $b_{t_0}(t,\tau) := \mu\tau + \mathbf{b}^\top \int_t^{t+\tau} \mathbb{E}_{t_0}\left[X_u\right] \mathrm{d}u$ that through straightforward computations can be written as

$$b_{t_0}(t,\tau) = \mu\tau + \mathbf{b}^\top \int_t^{t+\tau} \left[e^{\tilde{\mathbf{A}}\left(u-t_0\right)} \left(X_{t_0} + \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right) - \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right] \mathrm{d}u$$

$$= \left(1 - \mathbf{b}^\top \tilde{\mathbf{A}}^{-1}\mathbf{e}\right) \mu\tau + \mathbf{b}^\top e^{\tilde{\mathbf{A}}\left(t-t_0\right)} \int_t^{t+\tau} e^{\tilde{\mathbf{A}}(u-t)} \left(X_{t_0} + \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right) \mathrm{d}u.$$

Since we have a continuous integrand in a compact support

$$\int_t^{t+\tau} e^{\tilde{\mathbf{A}}(u-t)} \mathrm{d}u < +\infty,$$

we have

$$b_\infty(\tau) := \lim_{t\to+\infty} b_{t_0}(t,\tau) = \left(1 - \mathbf{b}^\top \tilde{\mathbf{A}}^{-1}\mathbf{e}\right) \mu\tau. \tag{E.24}$$

Denoting with $c_{t_0}(t,\tau) := \int_t^{t+\tau} \left[\mathbb{E}_{t_0}\left[N_u X_u\right] + \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu \mathbb{E}_{t_0}\left[N_t\right]\right] \mathrm{d}u$, we obtain

$$c_{t_0}(t,\tau) = I_{0,t_0}(t,\tau) + I_{1,t_0}(t,\tau) + I_{2,t_0}(t,\tau) + I_{3,t_0}(t,\tau) + I_{4,t_0}(t,\tau) + I_{5,t_0}(t,\tau)$$

where $I_{0,t_0}(t,\tau) := \int_t^{t+\tau} e^{\tilde{\mathbf{A}}\left(u-t_0\right)} X_{t_0} N_{t_0} \mathrm{d}u$ is rewritten as

$$I_{0,t_0}(t,\tau) = e^{\mathbf{A}(t\tilde{-}t_0)\tau} \int_t^{t+\tau} e^{\tilde{\mathbf{A}}(u-t)} X_{t_0} N_{t_0} \mathrm{d}u$$

and using the same arguments as above, we get

$$I_{0,\infty}(t,\tau) := \lim_{t\to+\infty} I_{0,t_0}(t,\tau) = \mathbf{0}.$$

The quantity $I_{1,t_0}(t,\tau) := \int_t^{t+\tau} \left(e^{\tilde{\mathbf{A}}\left(u-t_0\right)} - \mathbf{I}\right) \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu \mathrm{d}u$ can be rewritten as

$$I_{1,t_0}(t,\tau) = \int_t^{t+\tau} e^{\tilde{\mathbf{A}}\left(u-t_0\right)} \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu \mathrm{d}u - \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\tau$$

while taking the limit as $t \to +\infty$, we have

$$I_{1,\infty}(t,\tau) := \lim_{t\to+\infty} I_{1,t_0}(t,\tau) = -\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\tau. \tag{E.25}$$

The quantity

$$\begin{aligned} I_{2,t_0}(t,\tau) := &\int_t^{t+\tau} \int_{t_0}^{u} e^{\tilde{\mathbf{A}}(u-s)} \mathbf{e}\mathbf{b}^\top e^{\tilde{\mathbf{A}}\left(s-t_0\right)} \mathrm{d}s\mathrm{d}u \left[X_{t_0} + \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right] \\ &- \int_t^{t+\tau} \left(e^{\tilde{\mathbf{A}}\left(u-t_0\right)} - \mathbf{I}\right) \mathrm{d}u \tilde{\mathbf{A}}^{-1}\mathbf{e}\mathbf{b}^\top \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu \end{aligned} \tag{E.26}$$

depends on the integral $\int_t^{t+\tau} \int_{t_0}^{u} e^{\tilde{\mathbf{A}}(u-s)} \mathbf{e}\mathbf{b}^\top e^{\tilde{\mathbf{A}}\left(s-t_0\right)} \mathrm{d}s\mathrm{d}u$ where from the substitutions $s - t_0 = h$ and $r = u - t$ we get

$$\int_0^{\tau} \int_0^{t+r-t_0} e^{\tilde{\mathbf{A}}\left(t-t_0+r-h\right)} \mathbf{e}\mathbf{b}^\top e^{\tilde{\mathbf{A}}h} \mathrm{d}h\mathrm{d}r. \tag{E.27}$$

Defining

$$\ddot{\mathbf{A}} := \begin{bmatrix} \tilde{\mathbf{A}} & \mathbf{e}\mathbf{b}^\top \\ \mathbf{0}_{p,p} & \tilde{\mathbf{A}} \end{bmatrix}$$

and applying the result in Appendix C, the inner integral in (E.27) becomes

$$\left[\mathbf{I}_{p,p}; \mathbf{0}_{p,p}\right] e^{\ddot{\mathbf{A}}(t-t_0+r)} \begin{bmatrix} \mathbf{0}_{p,p} \\ \mathbf{I}_{p,p} \end{bmatrix}. \tag{E.28}$$

Thus the integral in (E.27) can be computed as follows

$$\left[\mathbf{I}_{p,p}; \mathbf{0}_{p,p}\right] e^{\ddot{\mathbf{A}}(t-t_0)} \int_0^\tau e^{\ddot{\mathbf{A}}r} \mathrm{d}r \begin{bmatrix} \mathbf{0}_{p,p} \\ \mathbf{I}_{p,p} \end{bmatrix}. \tag{E.29}$$

We notice that as $\int_0^\tau e^{\ddot{\mathbf{A}}r}\mathrm{d}r < +\infty$ and all eigenvalues of $\ddot{\mathbf{A}}$ have negative real part, then

$$I_{2,\infty}(\tau) := \lim_{t\to+\infty} I_{2,t_0}(t,\tau) = \tilde{\mathbf{A}}^{-1}\mathbf{e}\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\tau.$$

Similarly, we get the limit for the term $I_{3,t_0}(t,\tau) := \int_t^{t+\tau}\left[\int_{t_0}^u e^{\tilde{\mathbf{A}}(u-s)}\mu\mathbb{E}_{t_0}\left(N_s\right)\mathbf{e}\mathrm{d}s + \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\mathbb{E}_{t_0}\left(N_u\right)\right]\mathrm{d}u$ as $t\to+\infty$:

$$I_{3,\infty}(t,\tau) := \lim_{t\to+\infty} I_{3,t_0}(t,\tau) = -\tilde{\mathbf{A}}^{-1}\left[\mathbf{I}\frac{\tau^2}{2} + \tilde{\mathbf{A}}^{-1}\tau\right]\mathbf{e}\mu^2\left(1-\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right).$$

We define the following quantity

$$\begin{aligned} I_{4,t_0}(t,\tau) := &\left[\int_t^{t+\tau} e^{\tilde{\mathbf{A}}(u-t_0)}\left(u-t_0\right)\mathrm{d}u\right]\left[X_{t_0} + \tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\right]\mu + \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu^2\tau \\ &- \tilde{\mathbf{A}}^{-1}\int_t^{t+\tau} e^{\tilde{\mathbf{A}}(u-t_0)}\mathrm{d}u\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu^2 \end{aligned}$$

and observe that the first integral can be rewritten as

$$\int_t^{t+\tau} e^{\tilde{\mathbf{A}}(u-t_0)}\left(u-t_0\right)\mathrm{d}u = e^{\tilde{\mathbf{A}}(t-t_0)}\int_t^{t+\tau} e^{\tilde{\mathbf{A}}(u-t)}(u-t)\,\mathrm{d}u + e^{\tilde{\mathbf{A}}(t-t_0)}\left(t-t_0\right)\int_t^{t+\tau} e^{\tilde{\mathbf{A}}(u-t)}\mathrm{d}u$$

where both terms in the *rhs* tend to be zero as $t\to+\infty$ thus

$$I_{4,\infty}(\tau) = \lim_{t\to+\infty} I_{4,t_0}(t,\tau) = \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu^2\tau.$$

Similar arguments are used to determine the limit as $t\to+\infty$ for the quantity $I_{5,t_0}(\tau) := \int_t^{t+\tau}\int_{t_0}^u e^{\tilde{\mathbf{A}}(u-s)}\mathbb{E}_{t_0}\left[X_s, X_s^\top\right]\mathbf{b}\mathrm{d}s\mathrm{d}u$ as follows

$$I_{5,\infty}(\tau) = \lim_{t\to+\infty} I_{5,t_0}(t,\tau) = \tilde{\mathbf{A}}^{-1}\mathbf{B}\tilde{\tilde{\mathbf{A}}}^{-1}\mu\left(\tilde{\mathbf{e}} - \tilde{\mathbf{C}}\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\tau.$$

Combining all results, we get the stationary behavior for the quantity $c_\infty(\tau) := \lim_{t\to+\infty} c_{t_0}(t,\tau)$ that reads

$$\begin{aligned} c_\infty(\tau) = &-\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu\tau\left(1-\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right) - \tilde{\mathbf{A}}^{-1}\left[\mathbf{I}\frac{\tau^2}{2} + \tilde{\mathbf{A}}^{-1}\tau\right]\mathbf{e}\mu^2\left(1-\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right) + \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{A}}^{-1}\mathbf{e}\mu^2\tau \\ &+ \tilde{\mathbf{A}}^{-1}\mathbf{B}\tilde{\tilde{\mathbf{A}}}^{-1}\mu\left(\tilde{\mathbf{e}} - \tilde{\mathbf{C}}\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\tau. \end{aligned} \tag{E.30}$$

Furthermore,

$$\begin{aligned} \lim_{t\to+\infty}\mathbb{E}_{t_0}\left[\left(N_{t+\tau}-N_t\right)^2\right] &= 2\mu a_\infty(\tau) + b_\infty(\tau) + 2\mathbf{b}^\top c_\infty(\tau) - 2\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\left[e^{\tilde{\mathbf{A}}\tau} - \mathbf{I}\right]h_\infty(0) \\ &= \mu^2\left(1-\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)^2\tau^2 + \left(1-\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\left(1-2\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\mu\tau \\ &+ 2\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{A}}^{-1}\mathbf{e}\tau\mu^2\left(\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{e}\right) + 2\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\mathbf{B}\tilde{\tilde{\mathbf{A}}}^{-1}\mu\left(\tilde{\mathbf{e}} - \tilde{\mathbf{C}}\tilde{\mathbf{A}}^{-1}\mathbf{e}\right)\tau \\ &- 2\mathbf{b}^\top\tilde{\mathbf{A}}^{-1}\left[e^{\tilde{\mathbf{A}}\tau} - \mathbf{I}\right]h_\infty(0). \end{aligned}$$

By straightforward calculations, we obtain the result in (43) for the asymptotic variance. □

E.4. Proof of Theorem 7

Proof. Proof of Theorem 7. We first prove the existence of a positive constant $a_0 > 0$ such that the kernel function satisfies the condition

$$\int_{\mathbb{R}} e^{a_0|t|}h(t)\,\mathrm{d}t < +\infty. \tag{E.31}$$

We notice that Assumption 1 implies that

$$\int_{\mathbb{R}} e^{a_0|t|} h(t)\,\mathrm{d}t = \mathbf{b}^\top \int_0^{+\infty} e^{a_0 t} e^{\mathbf{A}t} \mathrm{d}t\mathbf{e} = \mathbf{b}^\top \mathbf{S} \int_0^{+\infty} e^{a_0 t} e^{\Lambda t} \mathrm{d}t \mathbf{S}^{-1}\mathbf{e}.$$

Choosing $a_0 \in \left(0, \left|\mathrm{Re}\left(\lambda_1\right)\right|\right)$ the condition in (E.31) is ensured and thus we can apply the result in Theorem 1 proved by Cheysson and Lang (2020), and the strong-mixing coefficient results to be $\alpha_N(r) = O\left(e^{-ar}\right)$ where $a \in \left(0, a_0\right)$. □

E.5. Proof of Theorem 8

Proof. Proof of Theorem 8. The proof is quite standard and is an application of Theorem 18.5.3 in Ibragimov and Linnik (1971) and Cramér-Wold device. Denoting with

$$\vartheta := \left[\mathbb{E}\left(\Delta_1 N_\infty\right), Var\left(\Delta_1 N_\infty\right), Acv(1), \ldots, Acv(d)\right]^\top \tag{E.32}$$

we apply Theorem 18.5.3 in Ibragimov and Linnik (1971) to the linear combination $\left(c^\top V_k\right)_{k=1,2,\ldots n}$ where c is a generic $d+2$ real vector such that $c\top\Sigma c > 0$. Since the strong mixing property is preserved under linear transformations as well as the rate we have

$$\sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n} c^\top V_k - c^\top \vartheta\right) \to \mathcal{N}\left(0, c^\top Var\left(V_1\right)c + 2\sum_{k=1}^{+\infty} c^\top Cov\left(V_1 V_k^\top\right)c\right), \text{ as } n \to +\infty$$

that is

$$\sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n} c^\top V_k - c^\top \vartheta\right) \to \mathcal{N}\left(0, c\top\Sigma c\right), \text{ as } n \to +\infty.$$

Applying Cramér-Wold device we obtain the asymptotic behavior in (49). □

Appendix F. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.insmatheco.2024.01.007.

References

Andresen, A., Benth, F.E., Koekebakker, S., Zakamulin, V., 2014. The CARMA interest rate model. International Journal of Theoretical and Applied Finance 17 (02), 1450008.
Bacry, E., Delattre, S., Hoffmann, M., Muzy, J.F., 2013. Modelling microstructure noise with mutually exciting point processes. Quantitative Finance 13 (1), 65–77.
Bacry, E., Mastromatteo, I., Muzy, J.F., 2015. Hawkes processes in finance. Market Microstructure and Liquidity 1 (01), 1550005.
Bauer, F.L., Fike, C.T., 1960. Norms and exclusion theorems. Numerische Mathematik 2 (1), 137–141.
Benth, F.E., Klüppelberg, C., Müller, G., Vos, L., 2014. Futures pricing in electricity markets based on stable CARMA spot models. Energy Economics 44, 392–406.
Benth, F.E., Rohde, V., 2019. On non-negative modeling with CARMA processes. Journal of Mathematical Analysis and Applications 476 (1), 196–214.
Bessy-Roland, Y., Boumezoued, A., Hillairet, C., 2021. Multivariate Hawkes process for cyber insurance. Annals of Actuarial Science 15 (1), 14–39.
Boswijk, H.P., Laeven, R.J., Yang, X., 2018. Testing for self-excitation in jumps. Journal of Econometrics 203 (2), 256–266.
Boumezoued, A., 2016. Population viewpoint on Hawkes processes. Advances in Applied Probability 48 (2), 463–480.
Brémaud, P., Massoulié, L., 1996. Stability of nonlinear Hawkes processes. Annals of Probability, 1563–1588.
Brockwell, P., Chadraa, E., Lindner, A., 2006. Continuous-time GARCH processes. The Annals of Applied Probability 16 (2), 790–826.
Brockwell, P.J., 2001. Lévy-driven carma processes. Annals of the Institute of Statistical Mathematics 53, 113–124.
Brockwell, P.J., 2004. Representations of continuous-time arma processes. Journal of Applied Probability 41, 375–382.
Brockwell, P.J., Davis, R.A., Yang, Y., 2011. Estimation for non-negative Lévy-driven CARMA processes. Journal of Business & Economic Statistics 29 (2), 250–259.
Carbonell, F., Jimenez, J., Pedroso, L., 2008. Computing multiple integrals involving matrix exponentials. Journal of Computational and Applied Mathematics 213 (1), 300–305.
Cattiaux, P., Colombani, L., Costa, M., 2022. Limit theorems for Hawkes processes including inhibition. Stochastic Processes and Their Applications 149, 404–426.
Cheng, Z., Seol, Y., 2020. Diffusion approximation of a risk model with non-stationary Hawkes arrivals of claims. Methodology and Computing in Applied Probability 22 (2), 555–571.
Cheysson, F., Lang, G., 2020. Strong mixing condition for Hawkes processes and application to Whittle estimation from count data. arXiv preprint arXiv:2003.04314.
Chiang, W.H., Liu, X., Mohler, G., 2022. Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates. International Journal of Forecasting 38 (2), 505–520.
Cinlar, E., 2011. Probability and Stochastics, vol. 261. Springer.
Clinet, S., Yoshida, N., 2017. Statistical inference for ergodic point processes and application to limit order book. Stochastic Processes and Their Applications 127 (6), 1800–1839.
Cui, L., Hawkes, A., Yi, H., 2020. An elementary derivation of moments of Hawkes processes. Advances in Applied Probability 52 (1), 102–137.
Da Fonseca, J., Zaatour, R., 2014. Hawkes process: fast calibration, application to trade clustering, and diffusive limit. The Journal of Futures Markets 34 (6), 548–579.
Das, M.K., Tsai, H., Kyriakou, I., Fusai, G., 2022. Technical note—on matrix exponential differentiation with application to weighted sum distributions. Operations Research 70 (4), 1984–1995.
Doob, J.L., 1944. The elementary Gaussian processes. The Annals of Mathematical Statistics 15 (3), 229–282.
Errais, E., Giesecke, K., Goldberg, L.R., 2010. Affine point processes and portfolio credit risk. SIAM Journal on Financial Mathematics 1 (1), 642–665.
Hawkes, A.G., 1971a. Point spectra of some mutually exciting point processes. Journal of the Royal Statistical Society, Series B, Methodological 33 (3), 438–443.
Hawkes, A.G., 1971b. Spectra of some self-exciting and mutually exciting point processes. Biometrika 58 (1), 83–90.
Hawkes, A.G., 2018. Hawkes processes and their applications to finance: a review. Quantitative Finance 18 (2), 193–198.
Hillairet, C., Réveillac, A., Rosenbaum, M., 2023. An expansion formula for Hawkes processes and application to cyber-insurance derivatives. Stochastic Processes and Their Applications.
Hitaj, A., Mercuri, L., Lévy, Rroji E., 2019. CARMA models for shocks in mortality. Decisions in Economics and Finance 42 (1), 205–227.
Iacus, S.M., Mercuri, L., 2015. Implementation of Lévy CARMA model in YUIMA package. Computational Statistics 30 (4), 1111–1141.
Iacus, S.M., Mercuri, L., Rroji, E., 2017. COGARCH(p,q): simulation and inference with the YUIMA package. Journal of Statistical Software 80, 1–49.
Iacus, S.M., Mercuri, L., Rroji, E., 2018. Discrete-time approximation of a COGARCH(p,q) model and its estimation. Journal of Time Series Analysis 39 (5), 787–809.
Ibragimov, I., Linnik, Y.V., 1971. Independent and Stationary Sequences of Random Variables. Wolters, Noordhoff Pub.
Lesage, L., Deaconu, M., Lejay, A., Meira, J.A., Nichil, G., et al., 2022. Hawkes processes framework with a Gamma density as excitation function: application to natural disasters for insurance. Methodology and Computing in Applied Probability, 1–29.
Marquardt, T., Stelzer, R., 2007. Multivariate CARMA processes. Stochastic Processes and Their Applications 117 (1), 96–120.

Mercuri, L., Perchiazzo, A., Rroji, E., 2021. Finite mixture approximation of CARMA(p,q) models. SIAM Journal on Financial Mathematics 12 (4), 1416–1458.
Mohler, G.O., Short, M.B., Brantingham, P.J., Schoenberg, F.P., Tita, G.E., 2011. Self-exciting point process modeling of crime. Journal of the American Statistical Association 106 (493), 100–108.
Muni Toke, I., Yoshida, N., 2017. Modelling intensities of order flows in a limit order book. Quantitative Finance 17 (5), 683–701.
Ogata, Y., 1988. Statistical models for earthquake occurrences and residual analysis for point processes. Journal of the American Statistical Association 83 (401), 9–27.
Poinas, A., Delyon, B., Lavancier, F., 2019. Mixing properties and central limit theorem for associated point processes. Bernoulli 25 (3), 1724–1754.
Rizoiu, M.A., Lee, Y., Mishra, S., Xie, L., 2017. Hawkes processes for events in social media. In: Frontiers of Multimedia Research, pp. 191–218.
Shlomovich, L., Cohen, E.A., Adams, N., Patel, L., 2022. Parameter estimation of binned Hawkes processes. Journal of Computational and Graphical Statistics 31 (4), 990–1000.
Swishchuk, A., Zagst, R., Zeller, G., 2021. Hawkes processes in insurance: risk model, application to empirical data and optimal investment. Insurance: Mathematics and Economics 101, 107–124.
Tómasson, H., 2015. Some computational aspects of Gaussian CARMA modelling. Statistics and Computing 25 (2), 375–387.
Tsai, H., Chan, K., 2003. A note on parameter differentiation of matrix exponentials, with applications to continuous-time modelling. Bernoulli 9 (5), 895–919.
Tsai, H., Chan, K., 2005. A note on non-negative continuous time processes. Journal of the Royal Statistical Society, Series B, Statistical Methodology 67 (4), 589–597.
Van Loan, C., 1978. Computing integrals involving the matrix exponential. IEEE Transactions on Automatic Control 23 (3), 395–404.