



HAL
open science

Estimating the parameters of a seasonal Markov-modulated Poisson process

Armelle Guillou, Stéphane Loisel, Gilles Stupfler

► **To cite this version:**

Armelle Guillou, Stéphane Loisel, Gilles Stupfler. Estimating the parameters of a seasonal Markov-modulated Poisson process. *Statistical Methodology*, 2015, 26, pp.103-123. 10.1016/j.stamet.2015.04.003 . hal-00965279v2

HAL Id: hal-00965279

<https://hal.science/hal-00965279v2>

Submitted on 24 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimating the parameters of a seasonal Markov-modulated Poisson process

Armelle Guillou⁽¹⁾, Stéphane Loisel⁽²⁾ & Gilles Stupfler⁽³⁾

⁽¹⁾ Université de Strasbourg & CNRS, IRMA, UMR 7501,
7 rue René Descartes, 67084 Strasbourg Cedex, France

⁽²⁾ Université Lyon 1, Institut de Science Financière et d'Assurances,
50 avenue Tony Garnier, 69007 Lyon, France

⁽³⁾ Aix Marseille Université, CNRS, EHESS, Centrale Marseille, GREQAM UMR 7316,
13002 Marseille, France

Abstract. Motivated by seasonality and regime-switching features of some insurance claim counting processes, we study the statistical analysis of a Markov-modulated Poisson process featuring seasonality. We prove the strong consistency and the asymptotic normality of a maximum split-time likelihood estimator of the parameters of this model, and present an algorithm to compute it in practice. The method is illustrated on a small simulation study and a real data analysis.

Keywords: Markov-modulated Poisson process, seasonality, split-time likelihood, strong consistency, asymptotic normality.

MSC 2010 Subject Classifications: Primary 62M05, 62F12; Secondary 60F05, 60F15.

1 Introduction

It is often the case that the insurance claim frequency is impacted by environment variables. For instance, flood risk is higher in a period of frequent heavy rains, and fire risk is more

intense when the weather is particularly dry. Such environment variables may be hidden to some extent to the practitioner: for instance, it is now accepted that the probabilities of severe floods in Australia, strong snowstorms in North America or hurricanes on the East Coast of the United States increase during La Niña episodes (see Neumann *et al.* [17], Cole and Pfaff [5], Parisi and Lund [20] and Landreneau [10]). This is now taken seriously by most reinsurers as well as Lloyd's and the UK Met Office [13]. However, observing and understanding the role of those variables is not easy, which makes it realistic to consider these variables as unobserved so far.

To take such a dependency into account, one may for instance assume that the underlying environment process is a Markov process J in continuous time and that in each state of J , the claim counting process N is a Poisson process. The resulting bivariate process (J, N) is then called a Markov-Modulated Poisson Process (MMPP). MMPPs have been used in different fields during the past forty years, in particular in data traffic systems (see *e.g.* Salvador *et al.* [27]), for ATM sources (see Kesidis *et al.* [9]), in manufacturing systems (see *e.g.* Ching *et al.* [4]) or even in ecology (see for example Skaug [29] for applications of MMPPs to clustered line transect data). The MMPP cookbook by Fischer and Meier-Hellstern [6] sums up the main results and ideas that were behind the rise of MMPP applications. The idea of considering Markov modulation in insurance was first introduced by Asmussen [2]; the obtained model can capture the fact that the insurance claim frequency may be modified if climatic, political or economic factors change. Such a model has gained considerable attention recently: see for instance Lu and Li [15], Ng and Yang [18], Zhu and Yang [33] and Wei *et al.* [32]. The parameters of an MMPP are often estimated using a Maximum Likelihood Estimator (MLE), whose consistency was proved in Rydén [22]. Various methods have been suggested to compute the MLE; a standard tool is the Expectation-Maximization (EM) algorithm, see Rydén [25] for the implementation of this procedure for the estimation of the parameters of an MMPP. We finally mention that in a recent paper, Guillou *et al.* [7] introduced a new MMPP-driven loss process in insurance with several lines of business, showed the strong consistency of the MLE and fitted their model to real sets of insurance data using an adaptation of the EM algorithm.

Of course, once the MMPP model is fitted, it is possible to use Bayesian techniques to determine probabilities to be in each state and consequently the average number of events during the

next period (see *e.g.* Scott [28]). If external information is present, then it is possible to enrich this Bayesian estimation. This is for example the case for some long-tailed non-life insurance businesses, where indices of sectorial inflation can provide useful information. For reinsurance cycles, large claims that may cause a cycle phase change as well as other aspects of competition or adverse development of reserves can sometimes be plugged into the Bayesian estimation process. For some other risks however, even if we feel that a phenomenon might have an impact on the claim frequency, it may be very hard to come up with a measure of such a phenomenon (for instance, the El Niño-La Niña phenomena). In that case our non-Bayesian framework is of interest for actuarial risk assessment.

Furthermore, many examples of practical applications in insurance display some sort of seasonal variation. For example, theft in garages are more frequent before Christmas as people tend to store Christmas gifts there, fire risk is more intense in the summer, and hurricanes occur mostly between June and November on the East Coast of the United States. These random, cyclic factors and their impact on insurance risk, which need to be taken into account to carry out a proper regime switching analysis, are yet to be understood and forecasted. In an inhomogeneous context with deterministic intensity function, Lu and Garrido [14] have fitted double-periodic Poisson intensity rates to hurricane data, for particular parametric forms (like double-beta and sine-beta intensities) to hurricane data. Helmers *et al.* [8] have provided an in-depth theoretical statistical analysis of such doubly periodic intensities. We aim at carrying out a theoretical statistical analysis in a stochastic intensity framework with seasonality.

An important aspect of pricing in non-life insurance concerns segmentation: thanks to generalized linear models or more sophisticated techniques, the insurer takes into account explanatory variables to adapt the price of the contract and avoid adverse selection (see Ohlsson and Johansson [19]). Besides, individual ratemaking is updated thanks to credibility adjustments in order to take into account the claim history of each policyholder or contract (see Bühlmann and Gisler [3]). Our approach is only operational and interesting at the aggregated risk management level, and it would be very challenging to try to combine it with regression techniques, from a theoretical point of view as well as the practical point of view since a very large number of data points would be needed to ensure that the estimation is reasonably accurate. The only simple way to combine the two approaches would be to assume that the seasonal and Marko-

vian intensities are multiplied for each class of contract by a number that corresponds to the risk of each class and does neither depend on the season nor on the state of the environment. In other terms, the ratio between claim intensities for classes A and B, say, should remain constant over time. This is of course a strong limitation; in wind-related risk for instance, some buildings are more subject to damage coming from wind effects than others, and consequently the rise of their risk should be sharper during windy episodes.

In this paper, we thus consider an MMPP featuring seasonality, and study estimation issues for this process when the environment process is unobserved. A further motivation for the introduction of this type of process appears in the Solvency II insurance regulation framework. In this framework, it is required to carry out an ORSA (Own Risk and Solvency Assessment). The ORSA has both qualitative and quantitative components, including the continuous compliance requirement: the insurer must be able to provide evidence that he/she satisfies solvency requirements at all times (see Vedani and Ramaharobandro [30]). This leads insurers to define confidence regions or to use a continuous time model. The MMPP with seasonality is one interesting potential tool for this new challenge for some non-life insurance companies.

A problem, however with this type of process is that contrary to when there is no seasonality, the random sequence of the inter-event times is not ergodic. Studying the asymptotic properties of the MLE, as Rydén [22] and Guillou *et al.* [7] do, is thus very difficult; furthermore, the particular structure of the likelihood makes it hard to compute the MLE in practice. To tackle this issue, we borrow an idea of Rydén [23, 24] and we introduce a Split-Time Likelihood (STL). The logarithm of this quantity is shown to be a sum of ergodic random variables; maximizing the STL then yields a Maximum Split-Time Likelihood Estimator (MSTLE) whose consistency and asymptotic normality can be proven using regenerative theory.

The outline of the paper is as follows. We give a precise definition of our model in Section 2 where we also introduce our estimator. The asymptotic properties are studied in Section 3. We explain how to implement our estimation technique in practice in Section 4 where we illustrate the performance of our estimator on a small simulation study and a real data analysis. Our paper ends by a conclusion and discussion of the results in Section 5. Technical lemmas are given and proven in the appendix.

2 Our model and estimation procedure

2.1 Our model

We consider an irreducible continuous-time Markov process J with generator L on the state space $\{1, \dots, r\}$, where $r \geq 2$. Consider further a counting process N and real numbers $\tau_0 = 0 < \tau_1 < \dots < \tau_{k-1} < 1 = \tau_k$ such that for every $q \in \mathbb{N}$, on the intervals $[q + \tau_{s-1}, q + \tau_s)$, if J is in state i , then N is a Poisson process with jump rate $\lambda_i^{(s)}$, where $\lambda_i^{(s)} \geq 0$. The time interval $[q + \tau_{s-1}, q + \tau_s)$ represents season s of the period $q + 1$. The case $k = 1$, in which there is actually no seasonality, is just the standard MMPP, which was considered from a statistical point of view by Rydén [22]; in this paper, we shall focus on the case $k \geq 2$. The jump intensity of the counting process N is thus modulated not only by the random switches of the Markov process J but also by switches from one season to another, which happen at nonrandom times.

The context of our work is the following: let us assume that the process N has been observed until time $n \in \mathbb{N} \setminus \{0\}$, so that the available data consists of

1. the number r of states of J and the times $\tau_1, \dots, \tau_{k-1}$,
2. the full knowledge of the process N between time 0 and time n .

The goal is to estimate the unknown parameters of the model, namely the elements ℓ_{ij} of the transition intensity matrix L of J and the jump intensities $\lambda_i^{(s)}$ of N . Since the modulating Markov process J is not observed, estimating the parameters of this model is not straightforward. For the sake of shortness, we let Φ be the global parameter of the model: Φ consists of the values ℓ_{ij} , for $1 \leq i, j \leq r$ and $i \neq j$ and of the $\lambda_i^{(s)}$, $1 \leq i \leq r$ and $1 \leq s \leq k$. The distribution of the process with parameter Φ is denoted by \mathbb{P}_Φ and we let \mathcal{E} be the parameter space

$$\mathcal{E} = \left\{ \Phi \mid L(\Phi) \text{ is irreducible and } \min_{1 \leq i \leq r} \min_{1 \leq s \leq k} \lambda_i^{(s)}(\Phi) > 0 \right\}.$$

The space \mathcal{E} can be seen as the set of those parameters for which in any state of J and in any season, an event, namely a jump of N , can occur with positive probability, while the irreducibility assumption makes sure that all states of J are visited infinitely often and thus that all parameters may be consistently estimated. Note that an element of \mathcal{E} is given by

$|\mathcal{E}| = r(r-1) + rs$ parameters. Finally, define, for $1 \leq s \leq k$ and $0 \leq y \leq \tau_s - \tau_{s-1}$ the matrix-valued functions

$$\begin{aligned} f(y, s, \Phi) &= \exp(y(L(\Phi) - \Lambda^{(s)}(\Phi)))\Lambda^{(s)}(\Phi) \\ \text{and } \bar{F}(y, s, \Phi) &= \exp(y(L(\Phi) - \Lambda^{(s)}(\Phi))) \end{aligned}$$

where $\Lambda^{(s)}(\Phi) = \text{diag}(\lambda_1^{(s)}(\Phi), \dots, \lambda_r^{(s)}(\Phi))$. The functions f and \bar{F} , which will be instrumental in defining our procedure, have a nice probabilistic interpretation (see Meier-Hellstern [16]): for any $i, j \in \{1, \dots, r\}$,

$$\begin{aligned} f_{ij}(y, s, \Phi) dy &:= \mathbb{P}(\mathcal{T}_1^{(s)} \in dy, J(y) = j | J(0) = i) \\ \text{and } \bar{F}_{ij}(y, s, \Phi) &:= \mathbb{P}(\mathcal{T}_1^{(s)} > y, J(y) = j | J(0) = i), \end{aligned}$$

where $\mathcal{T}_1^{(s)}$ is the time of the first event of an MMPP whose Markov-modulating process is J , with jump intensities given by $\Lambda^{(s)}$. The functions $f(\cdot, s, \Phi)$ and $\bar{F}(\cdot, s, \Phi)$ can thus be seen as the probability density and survival functions of the inter-event times for given states of J in season s , respectively.

2.2 A Maximum Split-Time Likelihood Estimator

It is impossible to apply any known results on MMPPs here since the jump intensity in a given state of J changes as time goes by; especially, given that J is in state j , N is an inhomogeneous Poisson process. To overcome this issue, we introduce some notation: let $W_{q,s}$ be the number of jumps of N during season s of the period q , and let $T_1^{(q,s)}, \dots, T_{W_{q,s}}^{(q,s)}$ be the successive jump times of N during this season. Let $Y_l^{(q,s)} = T_l^{(q,s)} - T_{l-1}^{(q,s)}$ be the inter-event times in season s (with $T_0^{(q,s)} = (q-1) + \tau_{s-1}$). It is assumed that the starting distribution of J is its unique stationary distribution $a(\Phi)$ on $\{1, \dots, r\}$, that is, the only row vector $a(\Phi)$ such that $a(\Phi)L(\Phi) = 0$ and the sum of the entries of $a(\Phi)$ is equal to 1.

Let $Z_q = (W_{q,s}, Y_1^{(q,s)}, \dots, Y_{W_{q,s}}^{(q,s)})_{1 \leq s \leq k}$ be the random vector representing the information available for period number q . With this notation, the process $(Z_q)_{q \geq 1}$ is stationary because J is a stationary Markov process. Besides, given the states of the irreducible Markov chain $(J(q))_{q \in \mathbb{N}}$, the random variables $Z_q, q \geq 1$ are independent, so that arguing along the lines of

the proof of Lemma 1 in Leroux [12] shows that the process $(Z_q)_{q \geq 1}$ is ergodic. We denote by $Z = (W_s, Y_1^{(s)}, \dots, Y_{W_s}^{(s)})_{1 \leq s \leq k}$ a random vector which shares the distribution of the Z_q , $q \geq 1$.

Let then $\mathcal{L}_1(Z, \Phi)$ be the likelihood of the observations over one period, computed under the parameter Φ : if $z = (w_s, y_1^{(s)}, \dots, y_{w_s}^{(s)})_{1 \leq s \leq k}$,

$$\mathcal{L}_1(z, \Phi) = a(\Phi) \left[\prod_{s=1}^k \left\{ \left(\prod_{l=1}^{w_s} f(y_l^{(s)}, s, \Phi) \right) \bar{F} \left(\tau_s - \tau_{s-1} - \sum_{l=1}^{w_s} y_l^{(s)}, s, \Phi \right) \right\} \right] 1_r$$

where 1_r is the column vector of size r having all entries equal to 1. This quantity can be expanded as:

$$\begin{aligned} \mathcal{L}_1(z, \Phi) = \sum_{i_0, \dots, i_{k+1}} a_{i_0}(\Phi) \prod_{s=1}^k \left[e'_{i_{s-1}} \prod_{l=1}^{w_s} f(y_l^{(s)}, s, \Phi) \right. \\ \left. \times \bar{F} \left(\tau_s - \tau_{s-1} - \sum_{l=1}^{w_s} y_l^{(s)}, s, \Phi \right) e_{i_s} \right] \end{aligned} \quad (2.1)$$

where e_j is the column vector of size r having all entries equal to 0 except the j th which is equal to 1, and e'_j is the transpose of e_j . Following Rydén [24], we let

$$\mathcal{STL}_n(\Phi) = \prod_{q=1}^n \mathcal{L}_1(Z_q, \Phi)$$

be the split-time likelihood (STL) of the observations. If the random variables Z_q , $q \geq 1$ were independent, the STL would be the total likelihood of the observations. While these variables are actually not independent, the sequence $(Z_q)_{q \geq 1}$ possesses regenerative properties, which implies in particular that it possesses independence properties given a suitable sequence of increasing random times; this is the key idea to the proofs of our main results (see the proof of Theorem 2). An MSTL is then any parameter that maximizes the STL, or equivalently the log-STL

$$\log \mathcal{STL}_n(\Phi) = \sum_{q=1}^n \log \mathcal{L}_1(Z_q, \Phi)$$

and an MSTLE is any estimator of an MSTL.

3 Asymptotic results

We shall write $\Phi \sim \Phi'$ whenever the distributions of Z under \mathbb{P}_Φ and $\mathbb{P}_{\Phi'}$ agree. The identifiability of the parameter Φ , or equivalently the determination of the equivalence class of Φ

modulo \sim , is a difficult question; for the case of an MMPP, this was discussed in Rydén [26]. In the special case when there is a season for which the jump intensities of N are distinct, which shall be the case we consider in our simulation study and real data analysis, this problem does however possess a simple solution:

Proposition 1. *Let $\Phi \in \mathcal{E}$. Assume that for some $s \in \{1, \dots, k\}$, the matrix $\Lambda^{(s)}(\Phi)$ possesses distinct diagonal elements. Then the equivalence class of Φ modulo \sim reduces to those parameters Φ' obtained by permutation of the states of the underlying Markov process J .*

Proof of Proposition 1. Let Φ' be such that $\Phi \sim \Phi'$ and (J, N) (resp. (J', N')) have distribution \mathbb{P}_Φ (resp. $\mathbb{P}_{\Phi'}$). By stationarity, the restriction of the process (J, N) (resp. (J', N')) to the time interval $[\tau_{s-1}, \tau_s]$ actually has the same distribution as the restriction of an MMPP $(\mathcal{J}, \mathcal{N})$ (resp. $(\mathcal{J}', \mathcal{N}')$) with parameters $(L(\Phi), \Lambda^{(s)}(\Phi))$ (resp. $(L(\Phi'), \Lambda^{(s)}(\Phi'))$) to the time interval $[0, \tau_s - \tau_{s-1}]$. Using again the stationarity property of the underlying Markov process, it is then enough to show that these two processes are deduced from one another (in the distributional sense) by a permutation of the states. Applying Corollary 1 in Rydén [26] completes the proof. ■

Roughly speaking, in this case, the strong identifiability constraint on season s carries over, in some sense, to the whole process by the stationarity property of J . Back to the general case, the following consistency result holds:

Theorem 1. *Let $\Phi_0 \in \mathcal{E}$ be the true value of the parameter and $\overline{\Phi_0}$ be the equivalence class of Φ_0 modulo \sim . Let K be a compact subset of \mathcal{E} such that $\Phi_0 \in K$ and $\widehat{\Phi}_n$ be the MSTLE for Φ_0 on K , computed over n periods. Then if $O \subset K$ is an open set containing $K \cap \overline{\Phi_0}$, one has $\widehat{\Phi}_n \in O$ almost surely for n large enough.*

Proof of Theorem 1. The proof is similar to that of Theorem 1 in Rydén [24]: let $\Phi \in \mathcal{E}$ be such that $\Phi \not\sim \Phi_0$ and G_Φ be a neighborhood of Φ as in Lemma 2. If $B(\Phi, 1/q)$ denotes the Euclidean open ball in \mathcal{E} with center Φ and radius $1/q$, the continuity of the map $\varphi \mapsto \mathcal{L}_1(Z, \varphi)$ yields

$$\sup_{\varphi \in G_\Phi \cap B(\Phi, 1/q)} \log \mathcal{L}_1(Z, \varphi) \rightarrow \log \mathcal{L}_1(Z, \Phi) \quad \text{as } q \rightarrow \infty.$$

Noticing that

$$\left| \sup_{\varphi \in G_\Phi \cap B(\Phi, 1/q)} \log \mathcal{L}_1(Z, \varphi) \right| \leq \left| \sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \right| + |\log \mathcal{L}_1(Z, \Phi)|$$

the dominated convergence theorem implies

$$\mathbb{E}_{\Phi_0} \left[\sup_{\varphi \in G_\Phi \cap B(\Phi, 1/q)} \log \mathcal{L}_1(Z, \varphi) \right] \rightarrow \mathbb{E}_{\Phi_0} [\log \mathcal{L}_1(Z, \Phi)] \text{ as } q \rightarrow \infty. \quad (3.1)$$

Since $\Phi \not\sim \Phi_0$, the information inequality (see Rao [21]) gives

$$\mathbb{E}_{\Phi_0} [\log \mathcal{L}_1(Z, \Phi)] + 2\varepsilon < \mathbb{E}_{\Phi_0} [\log \mathcal{L}_1(Z, \Phi_0)] \quad (3.2)$$

for some $\varepsilon > 0$. It is thus a consequence of (3.1) and (3.2) that there exists a (possibly different) neighborhood G_Φ of Φ with

$$\mathbb{E}_{\Phi_0} \left[\sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \right] \leq \mathbb{E}_{\Phi_0} [\log \mathcal{L}_1(Z, \Phi_0)] - \varepsilon. \quad (3.3)$$

Besides, since $(Z_q)_{q \geq 1}$ is ergodic,

$$\begin{aligned} \frac{1}{n} \sup_{\varphi \in G_\Phi} \log \mathcal{STL}_n(\varphi) &\leq \frac{1}{n} \sum_{q=1}^n \sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z_q, \varphi) \rightarrow \mathbb{E}_{\Phi_0} \left[\sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \right] \\ \text{and } \frac{1}{n} \log \mathcal{STL}_n(\Phi_0) &= \frac{1}{n} \sum_{q=1}^n \log \mathcal{L}_1(Z_q, \Phi_0) \rightarrow \mathbb{E}_{\Phi_0} [\log \mathcal{L}_1(Z, \Phi_0)] \end{aligned}$$

almost surely as $n \rightarrow \infty$. Hence

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sup_{\varphi \in G_\Phi} \log \mathcal{STL}_n(\varphi) \leq \mathbb{E}_{\Phi_0} \left[\sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \right] \leq \mathbb{E}_{\Phi_0} [\log \mathcal{L}_1(Z, \Phi_0)] - \varepsilon$$

almost surely as $n \rightarrow \infty$, by (3.3). Finally, remark that the compact set $O^c \cap K$, where O^c is the complement of O , may be covered by a finite number of such neighborhoods G_{Φ_i} , $1 \leq i \leq d$; this yields

$$\sup_{\varphi \in O^c \cap K} \{\log \mathcal{STL}_n(\varphi) - \log \mathcal{STL}_n(\Phi_0)\} \leq \max_{1 \leq i \leq d} \left\{ \sup_{\varphi \in G_{\Phi_i}} \log \mathcal{STL}_n(\varphi) - \log \mathcal{STL}_n(\Phi_0) \right\}$$

which tends to $-\infty$ almost surely as $n \rightarrow \infty$. As a consequence, necessarily $\widehat{\Phi}_n \in O$ for n large enough, which completes the proof. \blacksquare

We now wish to obtain an asymptotic normality result for our estimator. In what follows, we pick $i_0 \in \{1, \dots, r\}$ and we let ω_k be the successive times when the Markov chain $(J(q))$ reaches i_0 :

$$\omega_1 = \min\{q \geq 1 \mid J(q) = i_0\} \quad \text{and} \quad \forall k \geq 1, \quad \omega_{k+1} = \min\{q > \omega_k \mid J(q) = i_0\}.$$

Let further $\mathbb{P}_\Phi^{i_0}(\cdot) = \mathbb{P}_\Phi(\cdot \mid J(0) = i_0)$ be the probability measure deduced from \mathbb{P}_Φ given that J starts at i_0 .

Note that for every $\Phi \in \mathcal{E}$, since $L(\Phi)$ is the generator of an irreducible continuous-time Markov chain on a finite state space, then 0 is an eigenvalue of the transpose $L'(\Phi)$ of $L(\Phi)$ with multiplicity 1 and related normalized eigenvector $a'(\Phi)$. Since the map $\varphi \mapsto L'(\varphi)$ is infinitely continuously differentiable in a neighborhood of Φ , a straightforward extension of Theorem 8 in Chapter 9 of Lax [11] shows that the map $\varphi \mapsto a'(\varphi)$ is infinitely continuously differentiable in a neighborhood of Φ . The function $\varphi \mapsto \log \mathcal{L}_1(Z, \varphi)$ is thus infinitely continuously differentiable in a neighborhood of the true parameter Φ_0 ; if $\mathbb{E}_\Phi^{i_0}$ denotes the expectation under the measure $\mathbb{P}_\Phi^{i_0}$, we can set for Φ close enough to Φ_0

$$\begin{aligned} h_i(z, \Phi) &= \frac{\partial \log \mathcal{L}_1}{\partial \varphi_i}(z, \Phi), \\ A_{ij}(\Phi) &= \mathbb{E}_\Phi^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi) h_j(Z_q, \Phi) \right), \\ V_{ij}(\Phi) &= \mathbb{E}_\Phi^{i_0} \left(\sum_{p,q=1}^{\omega_1} h_i(Z_p, \Phi) h_j(Z_q, \Phi) \right). \end{aligned}$$

We assume that the matrix $A(\Phi) = (A_{ij}(\Phi))$ is invertible for $\Phi = \Phi_0$ and we let $V(\Phi) = (V_{ij}(\Phi))$, $C(\Phi) = [a_{i_0}(\Phi)]^{-1} A^{-1}(\Phi) V(\Phi) A^{-1}(\Phi)$ for Φ in a neighborhood of Φ_0 .

Our asymptotic normality result for $\widehat{\Phi}_n$ is as follows:

Theorem 2. *Let $\Phi_0 \in \mathcal{E}$ be the true value of the parameter and $\overline{\Phi_0}$ be the equivalence class of Φ_0 modulo \sim . Assume that Φ_0 lies in the interior of \mathcal{E} and let K be a compact subset of \mathcal{E} , whose interior contains Φ_0 , such that $K \cap \overline{\Phi_0} = \{\Phi_0\}$. Let $\widehat{\Phi}_n$ be the MSTLE computed on K over n periods. Then*

$$\sqrt{n}(\widehat{\Phi}_n - \Phi_0) \xrightarrow{d} \mathcal{N}(0, C(\Phi_0)) \quad \text{as } n \rightarrow \infty.$$

Proof of Theorem 2. We start as in the proof of Theorem 4 in Rydén [24]. Condition $K \cap \overline{\Phi_0} = \{\Phi_0\}$ ensures that $\widehat{\Phi}_n \rightarrow \Phi_0$ almost surely as $n \rightarrow \infty$. Especially, with probability 1, one has $\widehat{\Phi}_n \in K$ for n large enough by Theorem 1. For such n , a Taylor expansion of the i -th partial derivative of $\Phi \mapsto \log \mathcal{STL}_n(\Phi)$ at Φ_0 implies that

$$T_{1,n} := \frac{1}{\sqrt{n}} \sum_{q=1}^n h_i(Z_q, \Phi_0) = T_{2,n} + T_{3,n} \quad (3.4)$$

where

$$\begin{cases} T_{2,n} := - \sum_{j=1}^{|\mathcal{E}|} \sqrt{n}(\widehat{\Phi}_{n,j} - \Phi_{0,j}) \left[\frac{1}{n} \sum_{q=1}^n \frac{\partial^2 \log \mathcal{L}_1}{\partial \varphi_j \partial \varphi_i}(Z_q, \Phi_0) \right] \\ T_{3,n} := - \frac{1}{2} \sum_{j=1}^{|\mathcal{E}|} \sqrt{n}(\widehat{\Phi}_{n,j} - \Phi_{0,j}) \left[\sum_{k=1}^{|\mathcal{E}|} (\widehat{\Phi}_{n,k} - \Phi_{0,k}) \frac{1}{n} \sum_{q=1}^n \frac{\partial^3 \log \mathcal{L}_1}{\partial \varphi_k \partial \varphi_j \partial \varphi_i}(Z_q, \tilde{\Phi}_n) \right], \end{cases}$$

with $\tilde{\Phi}_n$ some point on the line connecting Φ_0 and $\widehat{\Phi}_n$. We start by dealing with the right-hand side of equality (3.4). To this end, use Lemma 4 and the fact that the process $(Z_q)_{q \geq 1}$ is ergodic to obtain

$$T_{2,n} = - \sum_{j=1}^{|\mathcal{E}|} \sqrt{n}(\widehat{\Phi}_{n,j} - \Phi_{0,j}) \mathbb{E}_{\Phi_0} \left(\frac{\partial^2 \log \mathcal{L}_1}{\partial \varphi_j \partial \varphi_i}(Z, \Phi_0) \right) (1 + o_{\mathbb{P}}(1)).$$

Moreover, since

$$\frac{\partial^2 \log \mathcal{L}_1}{\partial \varphi_j \partial \varphi_i} = \frac{1}{\mathcal{L}_1} \frac{\partial^2 \mathcal{L}_1}{\partial \varphi_j \partial \varphi_i} - \frac{1}{\mathcal{L}_1^2} \frac{\partial \mathcal{L}_1}{\partial \varphi_i} \frac{\partial \mathcal{L}_1}{\partial \varphi_j},$$

it is a consequence of Lemma 4 and of a differentiation under the expectation sign that

$$\begin{aligned} T_{2,n} &= \sum_{j=1}^{|\mathcal{E}|} \sqrt{n}(\widehat{\Phi}_{n,j} - \Phi_{0,j}) \mathbb{E}_{\Phi_0} (h_i(Z, \Phi_0) h_j(Z, \Phi_0)) (1 + o_{\mathbb{P}}(1)) \\ &= a_{i_0}(\Phi_0) \sum_{j=1}^{|\mathcal{E}|} \sqrt{n}(\widehat{\Phi}_{n,j} - \Phi_{0,j}) \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi_0) h_j(Z_q, \Phi_0) \right) (1 + o_{\mathbb{P}}(1)), \end{aligned} \quad (3.5)$$

by Lemma 6. Besides, Lemma 4, the ergodicity of $(Z_q)_{q \geq 1}$ and the consistency of $\widehat{\Phi}_n$ entail

$$T_{3,n} = o_{\mathbb{P}} \left(\sum_{j=1}^{|\mathcal{E}|} \sqrt{n}(\widehat{\Phi}_{n,j} - \Phi_{0,j}) \right) \quad (3.6)$$

as $n \rightarrow \infty$. Collecting (3.5) and (3.6), we obtain

$$T_{2,n} + T_{3,n} = a_{i_0}(\Phi_0) \sum_{j=1}^{|\mathcal{E}|} \sqrt{n} (\widehat{\Phi}_{n,j} - \Phi_{0,j}) \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi_0) h_j(Z_q, \Phi_0) \right) (1 + o_{\mathbb{P}}(1)) \quad (3.7)$$

as $n \rightarrow \infty$. We now examine the asymptotic properties of $T_{1,n}$: notice that $(Z_q)_{q \geq 1}$ is regenerative with associated cycle lengths $(C_q = \omega_q - \omega_{q-1})_{q \geq 1}$ (where we set $\omega_0 = 0$ for convenience), that is:

- for each $l \geq 2$, the random process $(C_{l+1+q}, Z_{\omega_{l+q}})_{q \geq 1}$ is independent of the ω_j , $1 \leq j \leq l - 1$ and its distribution does not depend on l ;
- given $(\omega_q)_{q \geq 1}$, $q \geq 1$, the random vectors Z_j , $\omega_{q-1} \leq j \leq \omega_q - 1$ are independent.

We may thus apply a law of large numbers similar to the one proposed in Theorem 2 in Rydén [24] (see also Asmussen [1]): this leads to

$$\frac{1}{\sqrt{n}} T_{1,n} = \frac{1}{n} \sum_{q=1}^n h_i(Z_q, \Phi_0) \xrightarrow{\mathbb{P}} a_{i_0}(\Phi_0) \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi_0) \right) \quad \text{as } n \rightarrow \infty$$

Besides, thanks to Lemma 4, we may once again differentiate under the integral sign to obtain $\mathbb{E}_{\Phi_0}(h_i(Z, \Phi_0)) = 0$. Lemma 4 and the ergodicity of $(Z_q)_{q \geq 1}$ thus yield

$$\frac{1}{\sqrt{n}} T_{1,n} = \frac{1}{n} \sum_{q=1}^n h_i(Z_q, \Phi_0) \rightarrow 0 \quad \text{almost surely as } n \rightarrow \infty.$$

Combining these last two convergences entails

$$\mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi_0) \right) = 0.$$

Finally $\mathbb{E}_{\Phi_0}^{i_0} \left| \sum_{q=1}^{\omega_1} h_i(Z_q, \Phi_0) \right|^2 < \infty$, see Lemmas 4 and 5; applying a central limit theorem similar to the one proposed in Theorem 3 in Rydén [24] then gives

$$\begin{aligned} T_{1,n} &= \frac{1}{\sqrt{n}} \sum_{q=1}^n \left[h_i(Z_q, \Phi_0) - a_{i_0}(\Phi_0) \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi_0) \right) \right] \\ &\xrightarrow{d} \mathcal{N}(0, a_{i_0}(\Phi_0) V(\Phi_0)) \quad \text{as } n \rightarrow \infty. \end{aligned} \quad (3.8)$$

Collecting (3.4), (3.7) and (3.8) concludes the proof. ■

The interesting point about MSTLEs is that confidence regions for Φ_0 may be computed fairly easily, see Lemma 7: it is enough to compute the MSTLE $\widehat{\Phi}_n$ with a numerical method, to generate random copies of the process under the law $\mathbb{P}_{\widehat{\Phi}_n}$ and to apply the strong law of large numbers to get an approximation of $C(\widehat{\Phi}_n)$ and hence of $C(\Phi_0)$.

We close this section by noting that Lemma 6 makes it possible to give further insight on the non-singularity of $A(\Phi_0)$. If $A(\Phi_0)$ were singular then, by Lemma 6, we would have

$$\exists j \in \{1, \dots, |\mathcal{E}|\}, \forall i \in \{1, \dots, |\mathcal{E}|\}, \mathbb{E}_{\Phi_0} (h_i(Z, \Phi_0)h_j(Z, \Phi_0)) = \sum_{\substack{\ell=1 \\ \ell \neq j}}^{|\mathcal{E}|} \gamma_\ell \mathbb{E}_{\Phi_0} (h_i(Z, \Phi_0)h_\ell(Z, \Phi_0))$$

where the γ_ℓ are real constants. It is then straightforward that

$$\mathbb{E}_{\Phi_0} \left(\left[h_j(Z, \Phi_0) - \sum_{\substack{\ell=1 \\ \ell \neq j}}^{|\mathcal{E}|} \gamma_\ell h_\ell(Z, \Phi_0) \right]^2 \right) = 0$$

so that the partial derivatives of $\log \mathcal{L}_1(\cdot, \Phi_0)$ are linearly dependent. The structure of the log-likelihood (over one period) being highly non-linear, one may therefore expect that $A(\Phi_0)$ will be invertible in the vast majority of cases.

4 Simulation study and real data analysis

4.1 Practical implementation

The MSTLE is typically computed using a (quasi-)Newton algorithm. Such an algorithm being iterative, we give a method to start it in practice. Let $w_{q,s}$ be the number of jumps of N during season s of the period q , $t_1^{(q,s)}, \dots, t_{w_{q,s}}^{(q,s)}$ be the successive jump times of N during this season and let $u_l^{(q,s)} = t_l^{(q,s)} - \tau_{s-1} - (q-1)(1 - (\tau_s - \tau_{s-1}))$, $1 \leq l \leq w_{q,s}$.

For all $1 \leq s \leq k$, we treat the data $D_s = (u_l^{(q,s)})_{q,l}$ as if it were as a sample of a univariate MMPP with parameters $\Phi^{(s)} = (L, \lambda_1^{(s)}, \dots, \lambda_r^{(s)})$. An EM algorithm is then used to provide a first estimate of $\Phi^{(s)}$; the initial estimation procedure is a straightforward adaptation of the one in Guillou *et al.* [7].

Let $(\widetilde{L}^{(s)}, \widetilde{\lambda}_1^{(s)}, \dots, \widetilde{\lambda}_r^{(s)})$ be the EM estimations obtained with this procedure. The EM algorithms involved may have switched some of the parameters: we thus note, if $M = (m_{ij})$ is a

square matrix having size r and σ is a permutation of $\{1, \dots, r\}$, $M \circ \sigma = (m_{\sigma(i)\sigma(j)})$. Let, for such permutations $\sigma_1, \dots, \sigma_k$,

$$\tilde{L}^{(\sigma_1, \dots, \sigma_k)} = \frac{1}{k} \sum_{s=1}^k \tilde{L}^{(s)} \circ \sigma_s$$

and

$$\tilde{\Phi}^{(\sigma_1, \dots, \sigma_k)} = (\tilde{L}^{(\sigma_1, \dots, \sigma_k)}, (\tilde{\lambda}_{\sigma_s(1)}^{(s)}, \dots, \tilde{\lambda}_{\sigma_s(r)}^{(s)})_{1 \leq s \leq k}).$$

We then run the optimization algorithm starting from each of these values: in doing so, we obtain estimates $\hat{\Phi}^{(\sigma_1, \dots, \sigma_k)}$ of the parameters. Finally, we compute

$$(\sigma_1^0, \dots, \sigma_k^0) = \arg \max_{\sigma_1, \dots, \sigma_k} \mathcal{STL}_n(\hat{\Phi}^{(\sigma_1, \dots, \sigma_k)})$$

and our MSTLE is $\hat{\Phi} = \hat{\Phi}^{(\sigma_1^0, \dots, \sigma_k^0)}$.

4.2 Simulation study

In this section, we examine how our estimator behaves on a finite sample situation. We are motivated by climate-related statistical problems such as the impact of the El Niño-La Niña cycle on hurricane risk. We choose reasonable parameters for a simple illustration, but cannot claim that those parameters are estimated on a real data set, simply because it is very hard to understand the impact of this effect on hurricane frequency. We consider the situation $r = 2$, that is, the underlying Markov process J is a two-state continuous-time Markov process. We further choose $k = 2$ and $\tau_1 = 1/2$, so that any period is divided into two seasons of equal length. The following cases are considered:

- Case 1: $\ell_{12} = 1$, $\ell_{21} = 2$, $\lambda_1^{(1)} = 1$, $\lambda_2^{(1)} = 5$, $\lambda_1^{(2)} = 5$, $\lambda_2^{(2)} = 25$.
- Case 2: $\ell_{12} = 3$, $\ell_{21} = 10$, $\lambda_1^{(1)} = 1$, $\lambda_2^{(1)} = 5$, $\lambda_1^{(2)} = 5$, $\lambda_2^{(2)} = 25$.
- Case 3: $\ell_{12} = 1$, $\ell_{21} = 2$, $\lambda_1^{(1)} = 1/2$, $\lambda_2^{(1)} = 5$, $\lambda_1^{(2)} = 5/2$, $\lambda_2^{(2)} = 25$.

For each of these cases, we consider the situations $n = 50$ and $n = 100$. Our estimation procedure is carried out on $S = 100$ replications of the considered process. In each case, we compute the mean and median L^1 -error related to each parameter. We also computed asymptotic 95% confidence intervals (ACI) with the help of Theorem 2 and Lemma 7; we

display the mean and median half-length of these intervals as well as the estimated coverage probabilities in Tables 1–3.

Following the intuition, the jump intensities are estimated more accurately when the ratio of the claim frequencies in different states is further from 1. This is because identifying the hidden states of J is easier when both regimes of the process are further apart, and thus more information on the process can be recovered in such cases. Besides, estimating the elements of L is harder than estimating the jump intensities since J is unobserved, which is confirmed by our results. Coverage probabilities are globally satisfying; although they are fairly low in case 2 for $n = 50$, they improve to reasonable levels for $n = 100$.

From this simulation study, one can see that in practice, if an insurance company had around 50 or 100 years of data, which would be an ideal situation, for intensities like the ones we chose (as far as the El Niño-La Niña-related problem on hurricane risk is concerned, 5 times as many hurricanes during the hurricane season as in the other one, and 5 times as many hurricanes during a bad El Niño-La Niña phase as in a favorable one), it is feasible to estimate the parameters with an average relative error of around 10 to 20%. Of course, we have taken here a quite favorable case, as we have only two states of the environment and two seasons with fixed intensities. Considering 5 or more states and 4 seasons would require a period of observation too long for the approach to be reasonable. However, in most hidden Markov models in insurance and finance, the number of states is 2 or 3. In the near future, one could probably refine these estimations thanks to partial observations of the environment process, as climatologists understand better and better the El Niño-La Niña phenomenon and its impact on insurance perils.

4.3 Real data analysis

In this section, we focus on extreme climate events (large bushfires, cyclones, floods, hail, storms, tornados, tsunamis) experienced by Australia from March 1967 to March 2012. During this timeframe, 193 such extreme events were observed. Their frequency is largely influenced by a seasonality effect; the first “season” we consider is fall-winter, which in Australia roughly corresponds to the end of March until the end of September, and the second one is spring-summer, namely from the beginning of October to mid-March. Moreover, extreme climate

events tend to be influenced by external effects such as the El Niño-La Niña phenomenon or even global phenomena like climate change, which motivates the use of our technique on this data set. Our unit of time is the standard calendar year; our data is represented in Fig. 1 and our estimates are given in Table 4. A reconstruction of the underlying process J , carried out by a straightforward adaptation of the Viterbi algorithm (see Viterbi [31]) to our case, is given in Fig. 2.

The algorithm seems to identify an underlying process J with very small transition rates. A consequence of this is that the confidence intervals for the elements of L are poor. This is not the case however for the jump intensities, and the algorithm confirms that in a given state of J , the intensities in both seasons are significantly different, thus emphasizing the importance of taking into account the seasonality effect. Furthermore, the frequency of extreme climate events is significantly higher in the spring-summer season, which was expected at least as far as bushfires, cyclones and storms were concerned. We underline that, due to the fact that there are very few transitions of J , the Markov assumption on this process is questionable; nevertheless, as already observed in Guillou *et al.* [7] on real sets of non-life and life insurance data, the algorithm correctly matches a point where the slope of N sharply increases with a breakpoint in the data set where the hidden process switches from state 2 to state 1, thus leading to higher frequencies of jumps of the counting process. We thus believe that our approach is superior to a naive model using seasonal versions of several different Poisson processes since such an approach would not be helpful in identifying potential unknown breakpoints in the global structure of the model. It is also superior to the use of a classical MMPP approach as in Guillou *et al.* [7] which would not be able to handle the seasonal components of the data.

5 Conclusion and discussion

In this paper, we introduce an MMPP featuring seasonality for which we discuss estimation issues when the environment process is unobserved. After establishing the main asymptotic properties of the resulting estimator, we illustrate its performance on a small simulation study in which we observe that the number of states is crucial. We conclude our numerical section by a real data analysis on extreme climate events in Australia on which our developed methodology

seems to perform well. In the real data study, the Markov assumption is of course questionable, as well as the absence of impact of climate change on the parameters. Extending the number of states in order to get a Markov process is feasible in theory, but would lead to estimation problems that would not be tractable in practice. Another problem, when considering climate-related problems such as the motivating hurricane risk example, is that in different states of the environment, the seasons may be longer or shorter, as conditions are more or less met for events to happen. This is not considered here and would require further theoretical analysis; linked to this concern, if one had further details about the events and access to meteorological data, it would of course be relevant to distinguish between categories of extreme events and to carry out a more detailed analysis. Having said that, even if ENSO values were available, their measure would not be precise enough and their link with claim intensity and frequency would still be too loose to use credibility adjustments at this stage, although it is likely to be possible in the near future. Finally, note that we focus here on frequency of extreme events, but of course for the insurer the claim amounts are important. Nevertheless, our approach accounts for periods of time when there are significantly more events, which is very important for reinsurance and risk management purposes.

Acknowledgments

The authors thank the associate editor and two anonymous referees for their helpful comments which led to significant improvements of this article.

References

- [1] S. Asmussen, *Applied probability and queues*, Wiley, New York, 1987.
- [2] S. Asmussen, Risk theory in a Markovian environment, *Scand. Actuar. J.* 2 (1989) 69–100.
- [3] H. Bühlmann, A. Gisler, *A course in credibility theory and its applications*, Springer, 2005.

- [4] W.K. Ching, R.H. Chan, X.Y. Zhou, Circulant preconditioners for Markov-modulated Poisson processes and their applications to manufacturing systems, *SIAM J. Matrix Anal. Appl.* 18 (1997) 464–481.
- [5] J.D. Cole, S.R. Pfaff, A climatology of tropical cyclones affecting the Texas coast during El Niño/non-El Niño years: 1990-1996, Technical Attachment SR/SSD 97-37, National Weather Service Office, Corpus Christi, Texas, available at <http://www.srh.noaa.gov/topics/attach/html/ssd97-37.htm> (1997).
- [6] W. Fischer, K.S. Meier-Hellstern, The Markov-modulated Poisson process (MMPP) cookbook, *Perf. Eval.* 18 (1993) 149–171.
- [7] A. Guillou, S. Loisel, G. Stupfler, Estimation of the parameters of a Markov-modulated loss process in insurance, *Insurance Math. Econom.* 53 (2013) 388–404.
- [8] R. Helmers, I.W. Mangku, R. Zitikis, A non-parametric estimator for the doubly periodic Poisson intensity function, *Stat. Methodol.* 4 (2007) 481–492.
- [9] G. Kesidis, J. Walrand, C.-S. Chang, Effective bandwidths for multiclass Markov fluids and other ATM sources, *IEEE/ACM Trans. Netw.* 1 (1993) 424–428.
- [10] D. Landreneau, Atlantic tropical storms and hurricanes affecting the United States: 1899-2000, NOAA Technical Memorandum NWS SR-206 (updated through 2002), National Weather Service Office, Lake Charles, Louisiana, available at <http://www.srh.noaa.gov/lch/?n=tropical> (2001).
- [11] P.D. Lax, *Linear Algebra and its applications*, Wiley, New York, 2007.
- [12] B.G. Leroux, Maximum-likelihood estimation for hidden Markov models, *Stochastic Process. Appl.* 40 (1992) 127–143.
- [13] Lloyd’s Forecasting risk, The value of long-range forecasting for the insurance industry, joint report with the UK Met Office, 2010.
- [14] Y. Lu, J. Garrido, Doubly periodic non-homogeneous models for hurricane data, *Stat. Methodol.* 2 (2005) 17–35.

- [15] Y. Lu, S. Li, On the probability of ruin in a Markov-modulated risk model, *Insurance Math. Econom.* 37 (2005) 522–532.
- [16] K.S. Meier-Hellstern, A fitting algorithm for Markov-modulated Poisson processes having two arrival rates, *European J. Oper. Res.* 29 (1987) 970–977.
- [17] C.J. Neumann, B.R. Jarvinen, C.J. McAdie, J.D. Elms, *Tropical Cyclones of the North Atlantic Ocean, 1871-1992, Historical Climatology Series 6-2*, National Climatic Data Center, Asheville, North Carolina, 1993.
- [18] A. Ng, H. Yang, On the joint distribution of surplus prior and immediately after ruin under a Markovian regime switching model, *Stochastic Process. Appl.* 116 (2006) 244–266.
- [19] E. Ohlsson, B. Johansson, *Non-life insurance pricing with generalized linear models*, Springer, 2010.
- [20] F. Parisi, R. Lund, Seasonality and return periods of landfalling Atlantic basin hurricanes, *Aust. N. Z. J. Stat.* 42 (2000) 271–282.
- [21] C.R. Rao, *Linear statistical inference and its applications*, Wiley, New York, 1973.
- [22] T. Rydén, Parameter estimation for Markov modulated Poisson processes, *Comm. Statist. Stochastic Models* 10 (1994) 795–829.
- [23] T. Rydén, Consistent and asymptotically normal parameter estimates for hidden Markov models, *Ann. Statist.* 22 (1994) 1884–1895.
- [24] T. Rydén, Consistent and asymptotically normal parameter estimates for Markov modulated Poisson processes, *Scand. J. Stat.* 22 (1995) 295–303.
- [25] T. Rydén, An EM algorithm for estimation in Markov-modulated Poisson processes, *Comput. Statist. Data Anal.* 21 (1996) 431–447.
- [26] T. Rydén, On identifiability and order of continuous-time aggregated Markov chains, Markov-modulated Poisson processes, and phase-type distributions, *J. Appl. Probab.* 33 (1996) 640–653.

- [27] P. Salvador, R. Valadas, A. Pacheco, Multiscale fitting procedure using Markov modulated Poisson processes, *Telecommun. Syst.* 23 (2003) 123–148.
- [28] S.L. Scott, Bayesian analysis of a two-state Markov modulated Poisson process, *J. Comput. Graph. Statist.* 8 (1999) 662–670.
- [29] H.J. Skaug, Markov modulated Poisson processes for clustered line transect data, *Environ. Ecol. Stat.* 13 (2006) 199–211.
- [30] J. Vedani, F. Ramaharobandro, Continuous compliance: a proxy-based monitoring framework, preprint, available at <https://hal.archives-ouvertes.fr/hal-00866531> (2013).
- [31] A. Viterbi, Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, *IEEE Trans. Inform. Theory* IT-13 (1967) 260–269.
- [32] J. Wei, H. Yang, R. Wang, On the Markov-modulated insurance risk model with tax, *Blätter der DGVMF* 31 (2010) 65–78.
- [33] J. Zhu, H. Yang, Ruin theory for a Markov regime-switching model under a threshold dividend strategy, *Insurance Math. Econom.* 42 (2008) 311–318.

Appendix: technical lemmas and proofs

The first result is a necessary step to obtain the strong consistency of our estimator.

Lemma 1. *For every $\varphi \in \mathcal{E}$ and $z = (w_s, y_1^{(s)}, \dots, y_{w_s}^{(s)})_{1 \leq s \leq k}$, one has*

$$m(rm)^{k+2}(rm)^{w_1+\dots+w_k} \leq \mathcal{L}_1(z, \varphi) \leq M(rM)^{k+2}(rM)^{w_1+\dots+w_k}$$

where

$$M := \max \left\{ 1, \max_{j,s} \lambda_j^{(s)}(\varphi) \right\} < \infty$$

and if $K_s := [0, \tau_s - \tau_{s-1}]$,

$$m := \min \left\{ \min_i a_i(\varphi), \min_{i,j,s} \min_{y \in K_s} f_{ij}(y, s, \Phi), \min_{i,j,s} \min_{y \in K_s} \overline{F}_{ij}(y, s, \Phi) \right\} > 0.$$

Proof of Lemma 1. Start by remarking that

$$\overline{F}_{\alpha\beta}(y, s, \varphi) = [\exp(y(L(\varphi) - \Lambda^{(s)}(\varphi)))]_{\alpha\beta} = \mathbb{P}(J(y) = \beta, N^{(s)}(y) = 0 \mid J(0) = \alpha) \leq 1$$

when $(J, N^{(s)})$ is an MMPP with transition intensity matrix $L(\varphi)$ and jump intensity matrix $\Lambda^{(s)}(\varphi)$, see Meier-Hellstern [16]. Therefore, for every $y \geq 0$, one has $\overline{F}_{\alpha\beta}(y, s, \varphi) \leq M$ and $f_{\alpha\beta}(y, s, \varphi) \leq M$. Using (2.1), we immediately obtain

$$\mathcal{L}_1(z, \varphi) \leq M(rM)^{k+2}(rM)^{w_1+\dots+w_k}.$$

Besides, the compactness of K_s and the continuity of the maps involved entail

$$\min_{i,j,s} \min_{y \in K_s} f_{ij}(y, s, \Phi) > 0 \quad \text{and} \quad \min_{i,j,s} \min_{y \in K_s} \overline{F}_{ij}(y, s, \Phi) > 0$$

so that $m > 0$. Hence, using (2.1), the inequality

$$\mathcal{L}_1(z, \varphi) \geq m(rm)^{k+2}(rm)^{w_1+\dots+w_k}$$

which completes the proof. ■

A second pivotal tool is the following technical lemma:

Lemma 2. *For every $\Phi \in \mathcal{E}$, there exists a neighborhood G_Φ of Φ in \mathcal{E} such that*

$$\mathbb{E}_{\Phi_0} \left| \sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \right| < \infty.$$

Proof of Lemma 2. Let G_Φ be a neighborhood of Φ , $A = \left\{ \sup_{\varphi \in G_\Phi} \mathcal{L}_1(Z, \varphi) \leq 1 \right\}$ and write

$$\begin{aligned} \mathbb{E}_{\Phi_0} \left| \sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \right| &= \mathbb{E}_{\Phi_0} \left[\sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \mathbb{1}_{A^c} \right] - \mathbb{E}_{\Phi_0} \left[\sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \mathbb{1}_A \right] \\ &\leq \mathbb{E}_{\Phi_0} \left[\sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \mathbb{1}_{A^c} \right] - \mathbb{E}_{\Phi_0} [\log \mathcal{L}_1(Z, \Phi) \mathbb{1}_A]. \quad (\text{B.1}) \end{aligned}$$

The goal is to show that the quantity in the right-hand side of this inequality is finite for a suitable choice of the neighborhood G_Φ of Φ in \mathcal{E} . Let $z = (w_s, y_1^{(s)}, \dots, y_{w_s}^{(s)})_{1 \leq s \leq k}$ and G_Φ be a compact neighborhood of Φ such that $\varphi \mapsto a(\varphi)$ is continuous on G_Φ and $M := \sup_{\varphi \in G_\Phi} \max \left\{ 1, \max_{j,s} \lambda_j^{(s)}(\varphi) \right\} < \infty$. Since the logarithm function is increasing, using Lemma 1 we deduce that

$$\mathbb{E}_{\Phi_0} \left[\sup_{\varphi \in G_\Phi} \log \mathcal{L}_1(Z, \varphi) \mathbb{1}_{A^c} \right] \leq \log M + \left[k + 2 + \sum_{s=1}^k \mathbb{E}_{\Phi_0}(W_s) \right] \log rM < \infty \quad (\text{B.2})$$

since $\mathbb{E}_{\Phi_0}(W_s) < \infty$ for all s . Furthermore, using Lemma 1, we have

$$- \mathbb{E}_{\Phi_0} [\log \mathcal{L}_1(Z, \Phi) \mathbb{1}_A] \leq |\log m| + \left[k + 2 + \sum_{s=1}^k \mathbb{E}_{\Phi_0}(W_s) \right] |\log rm| < \infty. \quad (\text{B.3})$$

for some $m > 0$. Using together (B.1), (B.2) and (B.3) concludes the proof. \blacksquare

The next result shows that for every $i_0 \in \{1, \dots, r\}$, the survival function of the random time $\omega_1 = \min\{q \geq 1 \mid J(q) = i_0\}$ converges to 0 geometrically fast.

Lemma 3. *For every $i_0 \in \{1, \dots, r\}$, there exists a neighborhood G of Φ_0 and a constant $c \in (0, 1)$ such that*

$$\forall k \in \mathbb{N}, \sup_{\Phi \in G} \mathbb{P}_\Phi^{i_0}(\omega_1 > k) \leq c^k.$$

In particular, $\mathbb{E}_{\Phi_0}^{i_0}(\omega_1^k) < \infty$ for every $k \geq 1$.

Proof of Lemma 3. The result is obviously true for $k = 0$. Pick $k \geq 1$ and note that

$$\begin{aligned} \mathbb{P}_\Phi^{i_0}(\omega_1 > k) &= \sum_{j \neq i_0} \mathbb{P}_\Phi^{i_0}(J(1) \neq i_0, \dots, J(k-2) \neq i_0, J(k-1) = j, J(k) \neq i_0) \\ &= \sum_{j \neq i_0} \mathbb{P}_\Phi^{i_0}(J(1) \neq i_0, \dots, J(k-2) \neq i_0, J(k-1) = j) \mathbb{P}_\Phi^j(J(1) \neq i_0). \quad (\text{B.4}) \end{aligned}$$

Set, for $i, j \in \{1, \dots, r\}$, $P_{ij}(\Phi) := \mathbb{P}_{\Phi}^i(J(1) = j) = [\exp(L(\Phi))]_{ij}$. In particular, the maps $\Phi \mapsto P_{ij}(\Phi)$ are continuous. Moreover, since the Markov process J is irreducible, it holds that $P_{ij}(\Phi) > 0$ for all i and j . Consequently

$$0 < c = \max_{j \neq i_0} \sup_{\Phi \in G} \mathbb{P}_{\Phi}^j(J(1) \neq i_0) < 1.$$

Using (B.4) entails

$$\sup_{\Phi \in G} \mathbb{P}_{\Phi}^{i_0}(\omega_1 > k) \leq c \sup_{\Phi \in G} \mathbb{P}_{\Phi}^{i_0}(\omega_1 > k - 1)$$

which gives the desired result by induction on k . ■

A key element in the proof of Theorem 2 is to show the following technical result:

Lemma 4. *There exists a neighborhood G of Φ_0 in \mathcal{E} and positive constants C, C' such that for any i, j, ℓ :*

$$\begin{aligned} & \sup_{\varphi \in G} \max \left\{ \frac{1}{\mathcal{L}_1(Z, \varphi)}, \mathcal{L}_1(Z, \varphi), \left| \frac{\partial \mathcal{L}_1}{\partial \varphi_i}(Z, \varphi) \right|, \left| \frac{\partial^2 \mathcal{L}_1}{\partial \varphi_i \partial \varphi_j}(Z, \varphi) \right|, \left| \frac{\partial^3 \mathcal{L}_1}{\partial \varphi_i \partial \varphi_j \partial \varphi_\ell}(Z, \varphi) \right| \right\} \\ & \leq C \exp \left(C' \sum_{s=1}^k W_s \right). \end{aligned}$$

Epecially, there exist (possibly different) positive constants C, C' such that for any i, j, ℓ :

$$\begin{aligned} & \sup_{\varphi \in G} \max \left\{ \left| \log \mathcal{L}_1(Z, \varphi) \right|, \left| \frac{\partial \log \mathcal{L}_1}{\partial \varphi_i}(Z, \varphi) \right|, \left| \frac{\partial^2 \log \mathcal{L}_1}{\partial \varphi_i \partial \varphi_j}(Z, \varphi) \right|, \left| \frac{\partial^3 \log \mathcal{L}_1}{\partial \varphi_i \partial \varphi_j \partial \varphi_\ell}(Z, \varphi) \right| \right\} \\ & \leq C \exp \left(C' \sum_{s=1}^k W_s \right), \end{aligned}$$

the right-hand side of the above inequality defines a random variable with finite \mathbb{P}_{Φ_0} -moments and for all i_1, i and j ,

$$\mathbb{E}_{\Phi_0}^{i_1} \left[\sup_{\varphi \in G} |h_i(Z, \varphi) h_j(Z, \varphi)| \right] < \infty.$$

Proof of Lemma 4. Assume that G is a neighborhood of Φ_0 such that $\varphi \mapsto a(\varphi)$ is continuous on G and $\inf_{\varphi \in G} \min_{j,s} \lambda_j^{(s)}(\varphi) > 0$. If $K_s = [0, \tau_s - \tau_{s-1}]$ then for any $\varphi \in G$, the ideas of the proof of Lemma 1 yield

$$m := \inf_{\varphi \in G} \left\{ \min_i a_i(\varphi), \min_{y \in K_s} \min_{i,j,s} f_{ij}(y, s, \Phi), \min_{y \in K_s} \min_{i,j,s} \bar{F}_{ij}(y, s, \Phi) \right\} > 0.$$

For all $\varphi \in G$ and every $z = (w_s, y_1^{(s)}, \dots, y_{w_s}^{(s)})_{1 \leq s \leq k}$, Lemma 1 entails

$$\mathcal{L}_1(z, \varphi) \geq m(rm)^{k+2}(rm)^{w_1+\dots+w_k}. \quad (\text{B.5})$$

If moreover G is a neighborhood of Φ_0 with $\sup_{\varphi \in G} \max_{i,j} \ell_{ij}(\varphi) < \infty$ and $\sup_{\varphi \in G} \max_{j,s} \lambda_j^{(s)}(\varphi) < \infty$ then similarly

$$M_1 := \sup_{\varphi \in G} \left\{ \max_i a_i(\varphi), \max_{i,j,s} \max_{y \in K_s} f_{ij}(y, s, \Phi), \max_{i,j,s} \max_{y \in K_s} \bar{F}_{ij}(y, s, \Phi) \right\} < \infty.$$

Let further

$$\begin{aligned} M_2 &= \sup_{\varphi \in G} \max_{n,p,s} \max_{\ell} \max_{y \in K_s} \left\{ \left| \frac{\partial f_{np}}{\partial \varphi_\ell}(y, s, \varphi) \right|, \left| \frac{\partial \bar{F}_{np}}{\partial \varphi_\ell}(y, s, \varphi) \right| \right\}, \\ M_3 &= \sup_{\varphi \in G} \max_{n,p,s} \max_{\ell,q} \max_{y \in K_s} \left\{ \left| \frac{\partial^2 f_{np}}{\partial \varphi_\ell \partial \varphi_q}(y, s, \varphi) \right|, \left| \frac{\partial^2 \bar{F}_{np}}{\partial \varphi_\ell \partial \varphi_q}(y, s, \varphi) \right| \right\} \\ \text{and } M_4 &= \sup_{\varphi \in G} \max_{n,p,s} \max_{\ell,q,r} \max_{y \in K_s} \left\{ \left| \frac{\partial^3 f_{np}}{\partial \varphi_\ell \partial \varphi_q \partial \varphi_r}(y, s, \varphi) \right|, \left| \frac{\partial^3 \bar{F}_{np}}{\partial \varphi_\ell \partial \varphi_q \partial \varphi_r}(y, s, \varphi) \right| \right\} \end{aligned}$$

and note that $M = \max\{M_1, M_2, M_3, M_4\} < \infty$. Equation (2.1), inequality (B.5) and tedious computations may then be used to show that one may find positive constants C, C' such that

$$\begin{aligned} & \sup_{\varphi \in G} \max \left\{ \frac{1}{\mathcal{L}_1(Z, \varphi)}, \mathcal{L}_1(Z, \varphi), \left| \frac{\partial \mathcal{L}_1}{\partial \varphi_i}(Z, \varphi) \right|, \left| \frac{\partial^2 \mathcal{L}_1}{\partial \varphi_i \partial \varphi_j}(Z, \varphi) \right|, \left| \frac{\partial^3 \mathcal{L}_1}{\partial \varphi_i \partial \varphi_j \partial \varphi_\ell}(Z, \varphi) \right| \right\} \\ & \leq C \exp \left(C' \sum_{s=1}^k W_s \right). \end{aligned}$$

Consequently

$$\begin{aligned} & \sup_{\varphi \in G} \max \left\{ \left| \log \mathcal{L}_1(Z, \varphi) \right|, \left| \frac{\partial \log \mathcal{L}_1}{\partial \varphi_i}(Z, \varphi) \right|, \left| \frac{\partial^2 \log \mathcal{L}_1}{\partial \varphi_i \partial \varphi_j}(Z, \varphi) \right|, \left| \frac{\partial^3 \log \mathcal{L}_1}{\partial \varphi_i \partial \varphi_j \partial \varphi_\ell}(Z, \varphi) \right| \right\} \\ & \leq C \exp \left(C' \sum_{s=1}^k W_s \right), \end{aligned}$$

for (possibly different) positive constants C, C' . Finally, since for every $s \in \{1, \dots, k\}$, W_s is a Poisson distributed random variable, one has $\mathbb{E}_{\Phi_0}(x^{W_s}) < \infty$ for all s and $x > 0$, which proves that the right-hand side of the inequality above defines a random variable having finite \mathbb{P}_{Φ_0} -moments. The second part of the lemma is thus a straightforward consequence of the above inequality. \blacksquare

The next result shall be used to check a couple of integrability conditions used in the proof of Theorem 2.

Lemma 5. *Assume that ψ is a Borel measurable nonnegative function. We consider the random variable $U = \sum_{q=1}^{\omega_1} \psi(Z_q)$.*

- If $\mathbb{E}_{\Phi_0}(\psi(Z)) < \infty$ then for every $i_0 \in \{1, \dots, r\}$, $\mathbb{E}_{\Phi_0}^{i_0}(U) < \infty$.
- If $\mathbb{E}_{\Phi_0}(\psi^2(Z)) < \infty$ then for every $i_0 \in \{1, \dots, r\}$, $\mathbb{E}_{\Phi_0}^{i_0}(U^2) < \infty$.

Proof of Lemma 5. We start by remarking that for any $l \geq 1$, if $\mathbb{E}_{\Phi_0}(\psi^l(Z)) < \infty$ then for any i_1 :

$$\mathbb{E}_{\Phi_0}^{i_1}|\psi^l(Z)| \leq \frac{\mathbb{E}_{\Phi_0}|\psi^l(Z)|}{a_{i_1}(\Phi_0)} < \infty$$

since $a_{i_1}(\Phi_0) > 0$. To prove the first statement, we then write

$$\mathbb{E}_{\Phi_0}^{i_0}(U) = \sum_{N=1}^{\infty} \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^N \psi(Z_q) \mathbb{1}_{\{\omega_1=N\}} \right) = \sum_{q=1}^{\infty} \mathbb{E}_{\Phi_0}^{i_0}(\psi(Z_q) \mathbb{1}_{\{\omega_1 \geq q\}}).$$

Since $\{\omega_1 \geq q\} = \bigcap_{l=0}^{q-1} \{J(l) \neq i_0\}$, the hidden Markov structure of (Z_q) yields

$$\mathbb{E}_{\Phi_0}^{i_0}(U) = \sum_{q=1}^{\infty} \sum_{i_1 \neq i_0} \mathbb{E}_{\Phi_0}^{i_1}(\psi(Z)) \mathbb{P}_{\Phi_0}^{i_0}(\omega_1 \geq q, J(q) = i_1) \leq \mathbb{E}_{\Phi_0}^{i_0}(\omega_1) \max_{i_1} \mathbb{E}_{\Phi_0}^{i_1}(\psi(Z)).$$

Since the Markov chain $(J(q))$ is irreducible on $\{1, \dots, r\}$ and has stationary distribution $a(\Phi_0)$, one has $\mathbb{E}_{\Phi_0}^{i_0}(\omega_1) = 1/a_{i_0}(\Phi_0) < \infty$, from which we deduce that the right-hand side is finite.

We now turn to the second part of the lemma. Note that

$$\begin{aligned} \mathbb{E}_{\Phi_0}^{i_0}(U^2) &= \sum_{N=1}^{\infty} \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{p,q=1}^N \psi(Z_p) \psi(Z_q) \mathbb{1}_{\{\omega_1=N\}} \right) \\ &= \sum_{N=1}^{\infty} \sum_{q=1}^N \mathbb{E}_{\Phi_0}^{i_0}(\psi^2(Z_q) \mathbb{1}_{\{\omega_1=N\}}) + 2 \sum_{N=1}^{\infty} \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{\substack{p,q=1 \\ p < q}}^N \psi(Z_p) \psi(Z_q) \mathbb{1}_{\{\omega_1=N\}} \right). \end{aligned}$$

We already know from the first statement of the lemma that

$$\sum_{N=1}^{\infty} \sum_{q=1}^N \mathbb{E}_{\Phi_0}^{i_0}(\psi^2(Z_q) \mathbb{1}_{\{\omega_1=N\}}) = \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} \psi^2(Z_q) \right) < \infty.$$

Further, for all Borel nonnegative functions f_1, \dots, f_N and all i_1, \dots, i_N , the hidden Markov structure of (Z_q) entails

$$\begin{aligned} & \mathbb{E}_{\Phi_0}^{i_0} (f_1(Z_1) \cdots f_N(Z_N) \mathbb{1}_{\{J(1)=i_1, \dots, J(N)=i_N\}}) \\ &= \left[\prod_{q=1}^N \mathbb{E}_{\Phi_0}^{i_0} (f_q(Z_q) | J(q-1) = i_{q-1}, J(q) = i_q) \right] \mathbb{P}_{\Phi_0}^{i_0} \left(\bigcap_{q=1}^N \{J(q) = i_q\} \right) \\ &= \left[\prod_{q=1}^N \mathbb{E}_{\Phi_0}^{i_{q-1}} (f_q(Z_1) | J(1) = i_q) \right] \mathbb{P}_{\Phi_0}^{i_0} \left(\bigcap_{q=1}^N \{J(q) = i_q\} \right). \end{aligned}$$

Since we may write

$$\{\omega_1 = N\} = \left(\bigcup_{j_1, \dots, j_{N-1} \neq i_0} \bigcap_{q=1}^{N-1} \{J(q) = j_q\} \right) \cap \{J(N) = i_0\}$$

it is straightforward that

$$\sum_{N=1}^{\infty} \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{\substack{p, q=1 \\ p < q}}^N \psi(Z_p) \psi(Z_q) \mathbb{1}_{\{\omega_1 = N\}} \right) \leq \left[\max_{i_1, i_2} \mathbb{E}_{\Phi_0}^{i_1} (\psi(Z) | J(1) = i_2) \right]^2 \mathbb{E}_{\Phi_0}^{i_0} (\omega_1^2)$$

which is finite since $\mathbb{E}_{\Phi_0}^{i_0} (\omega_1^2) < \infty$, see Lemma 3. ■

Lemma 6 is the key to the conclusion of the proof of Theorem 2.

Lemma 6. *Assume that ψ is a Borel measurable function such that $\mathbb{E}_{\Phi_0} |\psi(Z)| < \infty$. Then*

$$\mathbb{E}_{\Phi_0} (\psi(Z)) = a_{i_0}(\Phi_0) \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} \psi(Z_q) \right).$$

Proof of Lemma 6. The ergodicity of the process $(Z_q)_{q \geq 1}$ entails

$$\frac{1}{n} \sum_{q=1}^n \psi(Z_q) \rightarrow \mathbb{E}_{\Phi_0} (\psi(Z)) \quad \text{almost surely as } n \rightarrow \infty.$$

Besides, Lemma 5 yields $\mathbb{E} |\sum_{q=1}^{\omega_1} \psi(Z_q)| < \infty$, so that the regenerative properties of (Z_q) and a law of large numbers similar to the one proposed in Theorem 2 in Rydén [24] give

$$\frac{1}{n} \sum_{q=1}^n \psi(Z_q) \xrightarrow{\mathbb{P}} a_{i_0}(\Phi_0) \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} \psi(Z_q) \right) \quad \text{as } n \rightarrow \infty$$

from which the result follows. ■

Lemma 7 below is the essential result needed to construct confidence regions for the estimator.

Lemma 7. *If $\Phi_n \rightarrow \Phi_0$ as $n \rightarrow \infty$, one has $C(\Phi_n) \rightarrow C(\Phi_0)$ as $n \rightarrow \infty$.*

Proof of Lemma 7. We shall show that if $\Phi_n \rightarrow \Phi_0$, then $A(\Phi_n) \rightarrow A(\Phi_0)$ as $n \rightarrow \infty$. Let n be so large that Φ_n belongs to a (compact) neighborhood G of Φ_0 as in Lemma 4. Using the definition of A , one has:

$$\begin{aligned} |A_{ij}(\Phi_n) - A_{ij}(\Phi_0)| &\leq \left| \mathbb{E}_{\Phi_n}^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi_n) h_j(Z_q, \Phi_n) \right) - \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi_n) h_j(Z_q, \Phi_n) \right) \right| \\ &\quad + \left| \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi_n) h_j(Z_q, \Phi_n) \right) - \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi_0) h_j(Z_q, \Phi_0) \right) \right|. \end{aligned}$$

Applying Lemma 6 and differentiating under the expectation sign yields for any $\Phi \in G$

$$\mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi) h_j(Z_q, \Phi) \right) = \frac{1}{a_{i_0}(\Phi_0)} \mathbb{E}_{\Phi_0} \left[-\frac{\partial^2 \log \mathcal{L}_1}{\partial \varphi_i \partial \varphi_j}(Z, \Phi) \right].$$

Lemma 4 and the dominated convergence theorem thus entail

$$\left| \mathbb{E}_{\Phi_n}^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi_n) h_j(Z_q, \Phi_n) \right) - \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi_0) h_j(Z_q, \Phi_0) \right) \right| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Afterwards, notice that for every $\Phi \in \mathcal{E}$

$$\mathbb{E}_{\Phi}^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi) h_j(Z_q, \Phi) \right) = \sum_{i_1 \neq i_0} \sum_{q=1}^{\infty} \mathbb{E}_{\Phi}^{i_1} (h_i(Z, \Phi) h_j(Z, \Phi)) \mathbb{P}_{\Phi}^{i_0}(\omega_1 \geq q, J(q-1) = i_1)$$

so that

$$\left| \mathbb{E}_{\Phi_n}^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi_n) h_j(Z_q, \Phi_n) \right) - \mathbb{E}_{\Phi_0}^{i_0} \left(\sum_{q=1}^{\omega_1} h_i(Z_q, \Phi_n) h_j(Z_q, \Phi_n) \right) \right| \leq r_{1,n} + r_{2,n}$$

with

$$r_{1,n} = C \sum_{\substack{i_1 \neq i_0 \\ q \geq 1}} \left| \mathbb{P}_{\Phi_n}^{i_0}(\omega_1 \geq q, J(q-1) = i_1) - \mathbb{P}_{\Phi_0}^{i_0}(\omega_1 \geq q, J(q-1) = i_1) \right|$$

$$\text{and } r_{2,n} = \max_{i_1} \left| \mathbb{E}_{\Phi_n}^{i_1} (h_i(Z, \Phi_n) h_j(Z, \Phi_n)) - \mathbb{E}_{\Phi_0}^{i_1} (h_i(Z, \Phi_n) h_j(Z, \Phi_n)) \right| \sum_{q \geq 1} \mathbb{P}_{\Phi_n}^{i_0}(\omega_1 \geq q)$$

where we set

$$C = \max_{i_1} \mathbb{E}_{\Phi_0}^{i_1} \left[\sup_{\varphi \in G} |h_i(Z, \varphi) h_j(Z, \varphi)| \right]$$

which is finite by Lemma 4. We start by controlling $r_{1,n}$. Set $P_{ij}(\Phi) = \mathbb{P}_\Phi^i(J(1) = j)$ as in the proof of Lemma 3. For all $q \geq 1$ and $j \neq i_0$,

$$\begin{aligned} \mathbb{P}_{\Phi_n}^{i_0}(\omega_1 \geq q, J(q-1) = j) &= \sum_{i_1, \dots, i_{q-2} \neq i_0} P_{i_0, i_1}(\Phi_n) \cdots P_{i_{q-2}, j}(\Phi_n) \\ &\rightarrow \sum_{i_1, \dots, i_{q-2} \neq i_0} P_{i_0, i_1}(\Phi_0) \cdots P_{i_{q-2}, j}(\Phi_0) \text{ as } n \rightarrow \infty \\ &= \mathbb{P}_{\Phi_0}^{i_0}(\omega_1 \geq q, J(q-1) = j). \end{aligned}$$

Besides, Lemma 3 entails that there exists a constant $c \in (0, 1)$ such that for all $j \neq i_0$ and $q \geq 1$:

$$\sup_{n \geq 1} |\mathbb{P}_{\Phi_n}^{i_0}(\omega_1 \geq q, J(q-1) = j) - \mathbb{P}_{\Phi_0}^{i_0}(\omega_1 \geq q, J(q-1) = j)| \leq 2c^{q-1}$$

so that the dominated convergence theorem gives $r_{1,n} \rightarrow 0$ as $n \rightarrow \infty$. To control $r_{2,n}$, use Lemma 3 to get for all n

$$r_{2,n} \leq \frac{1}{1-c} \max_{i_1} |\mathbb{E}_{\Phi_n}^{i_1}(h_i(Z, \Phi_n)h_j(Z, \Phi_n)) - \mathbb{E}_{\Phi_0}^{i_1}(h_i(Z, \Phi_n)h_j(Z, \Phi_n))|.$$

To prove that the right-hand side of this inequality converges to 0 as $n \rightarrow \infty$, pick $p \geq 1$ and let $\|W\| = \sum_{s=1}^k W_s$ be the total number of events over one period. For any i_1 , it holds that

$$\begin{aligned} &|\mathbb{E}_{\Phi_n}^{i_1}(h_i(Z, \Phi_n)h_j(Z, \Phi_n)) - \mathbb{E}_{\Phi_0}^{i_1}(h_i(Z, \Phi_n)h_j(Z, \Phi_n))| \\ &\leq |\mathbb{E}_{\Phi_n}^{i_1}(h_i(Z, \Phi_n)h_j(Z, \Phi_n)\mathbb{1}_{\{\|W\| \leq p\}}) - \mathbb{E}_{\Phi_0}^{i_1}(h_i(Z, \Phi_n)h_j(Z, \Phi_n)\mathbb{1}_{\{\|W\| \leq p\}})| \\ &+ \mathbb{E}_{\Phi_n}^{i_1} \left[\sup_{\varphi \in G} |h_i(Z, \varphi)h_j(Z, \varphi)| \mathbb{1}_{\{\|W\| > p\}} \right] + \mathbb{E}_{\Phi_0}^{i_1} \left[\sup_{\varphi \in G} |h_i(Z, \varphi)h_j(Z, \varphi)| \mathbb{1}_{\{\|W\| > p\}} \right]. \end{aligned}$$

Since the state space of J is finite, it is enough to prove that the right-hand side of this inequality converges to 0 as $n \rightarrow \infty$. Using Lemma 4, we may find positive constants C_1, C_2 such that

$$\begin{aligned} &\mathbb{E}_{\Phi_n}^{i_1} \left[\sup_{\varphi \in G} |h_i(Z, \varphi)h_j(Z, \varphi)| \mathbb{1}_{\{\|W\| > p\}} \right] + \mathbb{E}_{\Phi_0}^{i_1} \left[\sup_{\varphi \in G} |h_i(Z, \varphi)h_j(Z, \varphi)| \mathbb{1}_{\{\|W\| > p\}} \right] \\ &\leq C_1 \left[\mathbb{E}_{\Phi_n}^{i_1}(\exp(C_2\|W\|)\mathbb{1}_{\{\|W\| > p\}}) + \mathbb{E}_{\Phi_0}^{i_1}(\exp(C_2\|W\|)\mathbb{1}_{\{\|W\| > p\}}) \right]. \end{aligned}$$

Pick now $\varepsilon > 0$. Since $\sup_{\varphi \in G} \max_{i,j} \lambda_j^{(i)}(\varphi) < \infty$, one has

$$\sup_{\Phi \in G} \mathbb{E}_\Phi^{i_1}(C_3^{\|W\|} \exp(C_4\|W\|)) < \infty.$$

Consequently, one can choose $p \in \mathbb{N} \setminus \{0\}$ such that for every n

$$\mathbb{E}_{\Phi_n}^{i_1} \left[\sup_{\varphi \in G} |h_i(Z, \varphi)h_j(Z, \varphi)| \mathbb{1}_{\{\|W\| > p\}} \right] + \mathbb{E}_{\Phi_0}^{i_1} \left[\sup_{\varphi \in G} |h_i(Z, \varphi)h_j(Z, \varphi)| \mathbb{1}_{\{\|W\| > p\}} \right] \leq 2\varepsilon/3.$$

It is then enough to prove that for all fixed $q \geq 0$,

$$\left| \mathbb{E}_{\Phi_n}^{i_1} (h_i(Z, \Phi_n)h_j(Z, \Phi_n) \mathbb{1}_{\{\|W\|=q\}}) - \mathbb{E}_{\Phi_0}^{i_1} (h_i(Z, \Phi_n)h_j(Z, \Phi_n) \mathbb{1}_{\{\|W\|=q\}}) \right| \rightarrow 0$$

as $n \rightarrow \infty$. Note that Z may be written $Z = (W, Y)$ where $W = (W_1, \dots, W_k)$ and Y is the list of inter-event times in year one. Therefore, if $\mathcal{L}_1(w, y, \Phi_0 | i_1)$ is the likelihood of (w, y) given that $J(0) = i_1$, we may write

$$\begin{aligned} & \left| \mathbb{E}_{\Phi_n}^{i_1} (h_i(Z, \Phi_n)h_j(Z, \Phi_n) \mathbb{1}_{\{\|W\|=q\}}) - \mathbb{E}_{\Phi_0}^{i_1} (h_i(Z, \Phi_n)h_j(Z, \Phi_n) \mathbb{1}_{\{\|W\|=q\}}) \right| \\ & \leq \sum_{\|w\|=q} \int \sup_{\varphi \in G} |h_i(w, y, \varphi)h_j(w, y, \varphi)| \mathbb{1}_{\{\|w\|=q\}} |\mathcal{L}_1(w, y, \Phi_n | i_1) - \mathcal{L}_1(w, y, \Phi_0 | i_1)| dy \end{aligned}$$

where, for a given $w = (w_1, \dots, w_k)$, the integral in the right-hand side is on the $y = (y_1^{(s)}, \dots, y_{w_s}^{(s)})_{1 \leq s \leq k}$ whose components are nonnegative and are such that for every $s \in \{1, \dots, k\}$, $\sum_{l=1}^{w_s} y_l^{(s)} \leq \tau_s - \tau_{s-1}$. Since \mathcal{L}_1 is a continuous function of the parameters, the integrand goes to 0 everywhere as $n \rightarrow \infty$. Furthermore, this function is bounded from above by the function

$$(w, y) \mapsto 2 \sup_{\varphi \in G} |h_i(w, y, \varphi)h_j(w, y, \varphi)| \mathbb{1}_{\{\|w\|=q\}} \sup_{\varphi \in G} \mathcal{L}_1(w, y, \varphi | i_1).$$

By Lemma 4 and the irreducibility of J , there exists a positive constant C_q such that

$$2 \sup_{\varphi \in G} |h_i(w, y, \varphi)h_j(w, y, \varphi)| \mathbb{1}_{\{\|w\|=q\}} \sup_{\varphi \in G} \mathcal{L}_1(w, y, \varphi | i_1) \leq C_q.$$

Finally, for every $t > 0$, the simplex

$$\left\{ (y_1, \dots, y_{w_s}) \mid \sum_{i=1}^{w_s} y_i \leq t \text{ and } \forall i, y_i \geq 0 \right\}$$

is contained in the hypercube $\prod_{i=1}^{w_s} \{y_i | 0 \leq y_i \leq t\}$, whose volume is $t^{w_s} < \infty$. The dominated convergence theorem consequently gives $A(\Phi_n) \rightarrow A(\Phi_0)$ as $n \rightarrow \infty$. The convergence $V(\Phi_n) \rightarrow V(\Phi_0)$ as $n \rightarrow \infty$ is obtained in the same way. \blacksquare

Table 1: Mean and median L^1 -errors associated to the estimators in case 1.

	ℓ_{12}	ℓ_{21}	$\lambda_1^{(1)}$	$\lambda_2^{(1)}$	$\lambda_1^{(2)}$	$\lambda_2^{(2)}$
Case 1						
$n = 50$						
Mean L^1 -error	0.303	0.683	0.381	0.866	0.774	2.506
Median L^1 -error	0.247	0.544	0.239	0.586	0.708	1.893
Mean half-length of the 95% ACI	0.818	1.740	0.791	2.177	1.611	5.380
Median half-length of the 95% ACI	0.766	0.837	0.886	2.082	1.570	5.184
Coverage probabilities of the 95% ACI	0.93	0.93	0.92	0.95	0.92	0.88
$n = 100$						
Mean L^1 -error	0.307	0.547	0.248	0.657	0.487	1.629
Median L^1 -error	0.234	0.423	0.176	0.510	0.384	1.259
Mean half-length of the 95% ACI	0.632	1.208	0.561	1.529	1.194	3.704
Median half-length of the 95% ACI	0.613	1.097	0.551	1.460	1.155	3.582
Coverage probabilities of the 95% ACI	0.88	0.92	0.94	0.91	0.93	0.94

Table 2: Mean and median L^1 -errors associated to the estimators in case 2.

	ℓ_{12}	ℓ_{21}	$\lambda_1^{(1)}$	$\lambda_2^{(1)}$	$\lambda_1^{(2)}$	$\lambda_2^{(2)}$
Case 2						
$n = 50$						
Mean L^1 -error	1.593	4.228	0.692	2.448	1.207	5.341
Median L^1 -error	1.372	3.504	0.472	1.773	1.030	4.182
Mean half-length of the 95% ACI	3.528	9.221	1.219	4.549	2.510	11.258
Median half-length of the 95% ACI	3.036	7.936	1.034	3.682	2.362	10.473
Coverage probabilities of the 95% ACI	0.79	0.83	0.84	0.79	0.86	0.91
$n = 100$						
Mean L^1 -error	1.304	3.145	0.483	1.715	0.847	3.599
Median L^1 -error	0.816	2.184	0.262	1.272	0.654	2.980
Mean half-length of the 95% ACI	2.893	7.323	0.882	3.115	1.875	8.269
Median half-length of the 95% ACI	2.429	6.393	0.764	2.788	1.749	7.860
Coverage probabilities of the 95% ACI	0.9	0.95	0.86	0.87	0.92	0.93

Table 3: Mean and median L^1 -errors associated to the estimators in case 3.

	ℓ_{12}	ℓ_{21}	$\lambda_1^{(1)}$	$\lambda_2^{(1)}$	$\lambda_1^{(2)}$	$\lambda_2^{(2)}$
Case 3						
$n = 50$						
Mean L^1 -error	0.272	0.551	0.255	0.812	0.370	1.671
Median L^1 -error	0.213	0.490	0.197	0.611	0.289	1.435
Mean half-length of the 95% ACI	0.715	1.487	0.593	1.972	1.073	4.707
Median half-length of the 95% ACI	0.720	1.452	0.594	1.963	1.059	4.709
Coverage probabilities of the 95% ACI	0.94	0.96	0.91	0.92	0.98	0.96
$n = 100$						
Mean L^1 -error	0.197	0.410	0.194	0.568	0.298	1.392
Median L^1 -error	0.169	0.332	0.173	0.505	0.238	1.200
Mean half-length of the 95% ACI	0.479	0.978	0.421	1.386	0.757	3.240
Median half-length of the 95% ACI	0.476	0.951	0.416	1.371	0.753	3.221
Coverage probabilities of the 95% ACI	0.91	0.92	0.9	0.9	0.96	0.96

Table 4: Estimates of the parameters on our real data set.

Parameters	Estimates	95% ACI
ℓ_{12}	0.0771	$[-1.162, 1.316]$
ℓ_{21}	0.0230	$[-0.345, 0.391]$
$\lambda_1^{(1)}$	5.434	$[2.329, 8.540]$
$\lambda_2^{(1)}$	0.762	$[0.142, 1.381]$
$\lambda_1^{(2)}$	10.659	$[6.567, 14.750]$
$\lambda_2^{(2)}$	5.361	$[3.945, 6.776]$

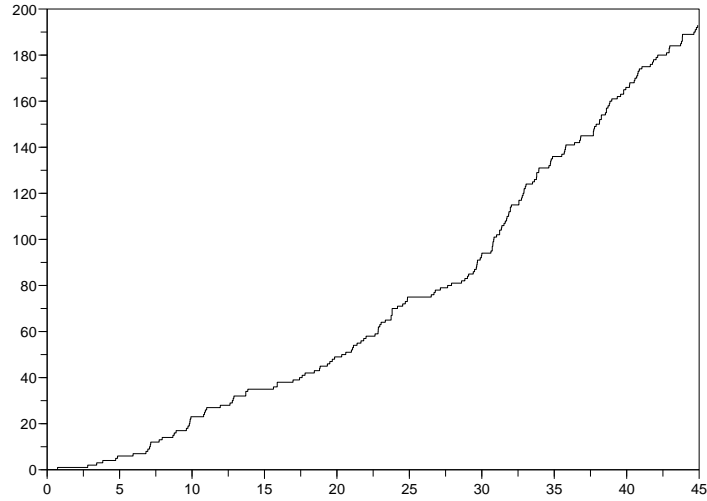


Figure 1: Counting process N of our real data set.

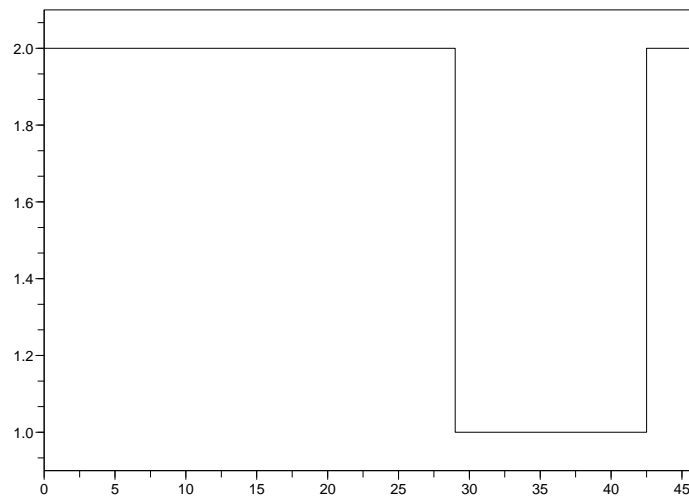


Figure 2: Reconstruction of the states of the underlying process J for our real data set.