Modelling and understanding count processes through a Markov-modulated non-homogeneous Poisson process framework

Benjamin Avanzi[a], Greg Taylor[a], Bernard Wong[a], Alan Xian[a,*]

[a]*School of Risk and Actuarial Studies, UNSW Business School*
UNSW Sydney NSW 2052, Australia

Abstract

The Markov-modulated Poisson process is utilised for count modelling in a variety of areas such as queueing, reliability, network and insurance claims analysis. In this paper, we extend the Markov-modulated Poisson process framework through the introduction of a flexible frequency perturbation measure. This contribution enables known information of observed event arrivals to be naturally incorporated in a tractable manner, while the hidden Markov chain captures the effect of unobservable drivers of the data. In addition to increases in accuracy and interpretability, this method supplements analysis of the latent factors. Further, this procedure naturally incorporates data features such as over-dispersion and autocorrelation. Additional insights can be generated to assist analysis, including a procedure for iterative model improvement.

Implementation difficulties are also addressed with a focus on dealing with large data sets, where latent models are especially advantageous due the large number of observations facilitating identification of hidden factors. Namely, computational issues such as numerical underflow and high processing cost arise in this context and in this paper, we produce procedures to overcome these problems.

This modelling framework is demonstrated using a large insurance data set to illustrate theoretical, practical and computational contributions and an empirical comparison to other count models highlight the advantages of the proposed approach.

Keywords: Risk analysis, Markov processes, Count processes, Data analysis, EM algorithm

1. Introduction

1.1. Context

Count modelling and the analysis of the occurrence of events is common to a wide variety of fields. The Markov-modulated Poisson process (MMPP), which is a continuous time model within the general family of Hidden Markov models (HMMs), is used in a broad spectrum of count modelling applications. It provides a natural extension upon the widely-used Poisson process in situations where an observable event process responds to external environmental stimuli. The MMPP model is composed of two components: (i) a stochastic process representing observable events over time, and (ii) a latent component that modulates the observed process. One benefit of this approach compared to the classical Poisson model stems from the greater modelling flexibility provided by the modulation, in particular to achieve over-dispersion. For example, in the context of modelling economic demand such as in Nasr and Maddah (2015) and Arts (2017), the hidden Markov process proxies the external environmental drivers of demand observations which can consist of various financial, economic, social and political factors. Similarly, in insurance literature such as Guillou et al. (2013), claim arrivals are influenced by a variety of different environmental drivers depending

*Corresponding author.
Email addresses: b.avanzi@unsw.edu.au (Benjamin Avanzi), greg.taylor@unsw.edu.au (Greg Taylor), bernard.wong@unsw.edu.au (Bernard Wong), a.xian@unsw.edu.au (Alan Xian)

on the line of business under consideration. The MMPP model provides a more faithful representation of real-world circumstances in this case because such drivers may be unobservable or extremely difficult to model, resulting in "hidden" states.

The MMPP model is used in a number of other fields for comparable reasons. It is commonly found in the natural science literature to model phenomena from birth-related movements (Leroux and Puterman, 1992) and photon arrivals (Kou et al., 2005) to rainfall (Thayakaran and Ramesh, 2013b) and populations (Langrock et al., 2013). Aside from the various economic and financial areas previously mentioned, MMPPs are popular in the areas of network theory, telecommunications and data traffic modelling such as in Yoshihara et al. (2001), Salvador et al. (2003), Scott and Smyth (2003) and Casale et al. (2016). Finally, this approach is utilised in the literature on risk, inventory, reliability and queueing theory to address unrealistic Poisson process assumptions through modulation, for example, in Özekici and Parlar (1999), Ching (1997), Landon et al. (2013) and Arts et al. (2016).

In the above papers, it is sometimes possible to apply techniques (usually Bayesian in nature) to gain inference on the latent Markov process within the MMPP. This is generally an endeavour that produces significant value, as it enhances the tractability and interpretability of the model outputs. For example, in the case of software reliability analysis as presented in Landon et al. (2013), the hidden process is inferred to represent the state of the software, with labels assigned of "excellent", "good" or "bad". An intuitive extension upon this concept is the idea that there may exist seasonality in the modulating process which has led to the development of seasonal MMPPs in the literature, such as the model presented in Guillou et al. (2015) for insurance claim counts where the latent process was assumed to represent the seasonal influence of various geological and climatological phenomena.

However, in any modelling exercise, there usually exist known and modellable causes of event arrivals, and these do not necessarily need to be cyclical or seasonal effects. This situation has garnered little attention in the MMPP-related academic literature, although the idea itself is not new with some discussion in the insurance context in Norberg (1993). Consider the case of insurance claims where the number of insurance policies under cover is clearly related to the observed claim arrivals but is unable to be taken into account as this factor is not necessarily periodic. An argument may also be made that such an intuitive driver of claims should be modelled explicitly rather than delegated to approximation through hidden regimes. Alternative examples can be easily conjured for the other fields, such as promotions producing spikes in online traffic as well as the demand (and consequently, purchases) for goods and services. Unfortunately, it is difficult to structure this information into the standard MMPP model, as regime selection is generally an automatic procedure.

1.2. Contributions

In this paper, an extension to the standard MMPP model is introduced which includes a component to explicitly control for the frequency perturbation caused by *known* factors. This measure is chosen to be very flexible, so that *both* cyclical and non-recurring trends can be captured. The introduced heterogeneity here produces the Markov-modulated
emphnon-homogeneous Poisson process (MMNPP).

The advantages of the proposed framework are many. Firstly, the original benefits of MMPPs in analysing count processes that are outlined above are retained. Secondly, the distinct separation of the known and hidden drivers of event arrivals enhances the insights obtained. This greatly benefits the utility of the approach, as inappropriate or insufficient consideration of the known components can confound any inferences obtained about the unobservable regimes of the hidden Markov chain. By explicitly taking these factors into account, the modelling outputs may become more interpretable and this is demonstrated in an empirical case study in Section 4 where the MMNPP model is utilised as a diagnostic tool for generating inferences which further assist the modelling procedure. Data features such as over-dispersion, auto-correlation and persistence in regimes are also able to taken into account in an understandable manner. This included information, in addition to the domain knowledge structured into the model, can ultimately lead to gains in accuracy and precision.

While non-homogeneous extensions to the MMPP model have not been widely explored in the literature (see Thayakaran and Ramesh (2013a); Guillou et al. (2015) for the examples that the authors are aware of),

comparable models have been successfully utilised for the more general family of Hidden Markov models. In these cases, the latent model is implemented to capture various unobservable factors with the non-homogeneous component supplementing this analysis. For example, Montoya et al. (2010) apply a non-homogeneous HMM to assess the effects of pharmaceutical marketing activities on physicians, taking into account their heterogeneity and behaviour dynamics. Shi and Zhang (2014) apply a non-homogeneous HMM to investigate decision aids for online purchases with time-varying parameters across the latent Markov states. Further examples of applications include buyer-seller dynamics in Zhang et al. (2014) as well as the modelling of phenomena such as precipitation (Hughes et al., 1999; Greene et al., 2011) and pollution concentration (Lagona et al., 2011). This approach is also found in insurance analysis in Badescu et al. (2019), who use a Pascal-HMM to model claim arrivals. Analogously, our proposed non-homogeneous extension will provide a large amount of flexibility and practicability to the MMNPP model in order to improve analysis in the varied areas where it is applied. Further, in comparison to the HMM approach, our model does not require considerations of discretisation as both the arrival processes and the underlying Markov chain are continuous. It is noted in some papers such as Badescu et al. (2019) and Crevecoeur et al. (2019) that choice of granularity and appropriate assumptions can be a non-trivial task, and the time-scale invariance of the continuous approach is advantageous.

Another contribution in this paper is that we show that the MMNPP is a specific operational time transform of the standard MMPP. The form of the transform depends on the frequency perturbation measure which provides much intuition to the modelling process which aids the interpretability of the final model outputs.

Finally, implementations found in the literature generally utilise examples with sample sizes that are relatively small. While this may be appropriate in certain areas, consideration of larger data sets is of merit, particularly given the current popularity of analytic models that are applied to such data sets. Further, the large number of observations advantages latent models such as the MMNPP due to clearer identification of hidden regimes.

Under these circumstances, calibration problems such as long computation times and numerical instability are greatly exacerbated. Without addressing these problems, applicability of the proposed approach is more limited. However, it is exactly these situations that are ideal for analysis using hidden Markov models as the unobservable regimes are more clearly identified. In this paper, we produce an Expectation-Maximisation (EM) calibration algorithm that introduces several computational developments to reduce calibration time, allowing the model to be calibrated in a reasonable period of time when there is a large amount of data to be processed. Further, the issue of numerical underflow is dealt with through a scaling procedure applied to the forward-backwards recursions. These practical considerations and contributions are outlined in the next section with context provided to demonstrate the necessity of these developments, as otherwise, the model is unable to be implemented and thus provides little utility.

In the following section, we outline the insurance claim analysis context that serves as a motivating example for the developments within this paper. The MMNPP model specifications are then presented in Section 2. In Section 3, we develop a calibration procedure including various techniques to alleviate the computational problems explained previously. In Section 4, we implement the model on a real data set to highlight the inferential capabilities of our proposed approach. We also provide a comparison to other models to highlight the advantages provided by the MMNPP. Section 5 concludes.

1.3. Detailed motivating example: Insurance claim counts

The empirical case study in Section 4 examines daily claim counts for a motor insurance line of business for a large Australian non-life insurer. The analysis of insurance claim occurrences is a common procedure and serves to inform many insurance business operations. There is a large amount of actuarial science literature on this topic but until recently, analysis has exploited data aggregated over longer discrete time frames such as months or years. Weekly and daily granularity of claim count data has been investigated by only a few papers such as Verrall and Wüthrich (2016); Avanzi et al. (2016); Crevecoeur et al. (2019) which deal with the reporting delay aspect of insurance claims while other nuances such as seasonality are demonstrated in Guillou et al. (2015); Albrecher et al. (2019). The benefits of analysis at this greater level

of detail include the generation of insights that were previously masked by aggregation and discretisation, which can lead to greater accuracy and precision.

Insurance claim count modelling encounters several obstacles that are significantly alleviated by the innovations made in this paper. Particularly for well known lines of business such as Motor, insurers possess established domain knowledge of important covariates that influence claim observations and the inability to incorporate this information into analysis would hamper to utility of any such modelling approach. Secondly, analysis at greater granularity results in a very large number of individual data points as opposed to the original aggregated observations (significantly more than half a million in our real data case study). This aggravates calibration issues common to hidden Markov models such as numerical instability as well as considerations such as calibration time and processing power required. Our developments allow the model to be calibrated on standard computing software in a reasonable amount of time. Finally, the insights in Section 4 highlight the diagnostic abilities of the modelling framework, including the capacity for iterative model improvement. This capability is particularly relevant for insurance claim modelling as analysis at this level of detail is not commonly undertaken and we demonstrate that additional interpretable causes of claim observations are able to be extracted. These drivers can then be examined and incorporated as deemed fit by the modeller.

2. Model Specifications for the MMNPP

We assume that event arrivals, over the time interval $[0, T]$, follow a Markov-modulated non-homogeneous Poisson process $N_M = \{N_M(t); t \geq 0\}$. Here, $N_M(t)$ represents the total number of events arriving before time t, with $N_M(0) = 0$. There is also assumed to be an unobservable underlying environmental process that impacts the event arrival intensity. This process is modelled with a continuous time Markov chain $M(t)$ with finite discrete state space $1, 2, \ldots, r$. We denote this latent process $M = \{M(t); t \geq 0\}$, and it is assumed that the chain has constant stationary transition intensities between states i and j of $q_{i,j}$. Consequently, if the n-th state change time is denoted by u_n, then the duration of the n-th state is exponentially distributed and given by

$$\mathbb{P}[u_{n+1} - u_n > t | M(u_n) = i] = \exp\left(-q_i t\right), \text{ where } q_i = \sum_{j \neq i} q_{i,j}. \tag{2.1}$$

When the Markov process is in state i at time t, events arrive according to the process N_M with intensity rate

$$\lambda_M(t) = \lambda_{M(t)} \times \gamma(t). \tag{2.2}$$

Here, $\lambda_{M(t)}$ is the constant intensity conditional on the state of the underlying Markov chain at time t, and γ represents the known exogenous volume or exposure process. The latter is at the heart of the contributions of this paper. This γ-process is very flexible and can take the form of any desired function of time, under the restriction that is adequately approximated by a finite number of constant-valued intervals. This is a fairly weak assumption as the simple function approximation is very general, allowing for a wide range of functions to be adequately captured such as continuous and finitely dis-continuous functions. It would be difficult to envisage a volume process in any practical circumstance that would be suitably irregular to violate this assumption. The reasoning behind this constraint is that it greatly reduces the computational burden in evaluating several complex expressions involving matrix differential equations. This is detailed within the proof for Theorem 3.2 in Section 3.2. The multiplicative structure of the overall Poisson intensity is also theoretically justified in Section 2.1 but it is intuitively explainable. For example, if the number of consumers within the market for a certain good doubles, then it is reasonable to expect that the observed conditional Poisson intensity of purchases of the good also doubles, ceteris paribus.

The standard expressions for the conditional Poisson processes are easily obtained under this framework. The probability of n events occurring before time t is given by

$$\mathbb{P}[N_M(t) = n | M(s); 0 \leq s \leq t] = \exp\left(-\int_0^t \lambda_{M(s)} \gamma(s) ds\right) \frac{\left(\int_0^t \lambda_{M(s)} \gamma(s) ds\right)^n}{n!}. \tag{2.3}$$

From this expression, properties of the inter-arrival times of events may also be obtained. Let t_n be the arrival time of the n-th event. Then

$$\mathbb{P}[t_{n+1} > t_n + t | t_n, M(s); t_n \leq s \leq t_n + t] = \exp\left(-\int_{t_n}^{t_n+t} \lambda_{M(s)}\gamma(s)ds\right). \tag{2.4}$$

In the paper, the overall MMNPP process to describe the hidden regimes and event arrivals is denoted as $\{M, N_M\} = \{M(t), N_M(t); t \geq 0\}$.

2.1. Justification for the form of the exposure measure

We have defined the non-homogeneous Poisson intensity $\lambda_M(t) = \lambda_{M(t)} \times \gamma(t)$, where $\gamma(t)$ was a known exogenous process that reflected external risk exposures. The form of this intensity can be theoretically and intuitively justified in the following manner. If we have multiple (homogeneous) MMPP processes with the same modulating Markov process, then we can show that the standard additivity property of Poisson processes still holds. This result follows directly from Proposition 2.1 below.

Proposition 2.1. *Let there be two MMPP processes $\{M(t), N_M^1(t)\}$ and $\{M(t), N_M^2(t)\}$ with the same underlying Markov chain $\{M(t)\}$ and claim intensities $\lambda_M^1(t)$ and $\lambda_M^2(t)$ respectively. We also have that the conditional Poisson processes $N_M^i(t)$ are independent for $i = 1, 2$. Then the combined process*

$$\{M(t), N_M^{1,2}(t)\} = \{M(t), N_M^1(t) + N_M^2(t)\}$$

is also an MMPP with a conditional Poisson process $N_M^{1,2}(t)$ that has a claim intensity of $\lambda_M(t) = \lambda_M^1(t) + \lambda_M^2(t)$.

Proof of Proposition 2.1 This proposition is a very intuitive result but we provide a quick proof here for completeness. As the Markov process $M(t)$ is the same in both MMPP processes, it suffices to prove that the combined process $N_M^{1,2}(t)$ is a conditional Poisson process with intensity $\lambda_M(t) = \lambda_M^1(t) + \lambda_M^2(t)$ in order to obtain the result that $\{M(t), N_M^{1,2}(t)\}$ is the required MMPP. This follows directly from the definition of Poisson processes. We begin by noting that each $N_M^i(t)$ is a conditional Poisson process i.e. $N_M^i(t)$ for $i = 1, 2$ has the following four properties:

1. $N_M^i(0) = 0$;
2. $N_M^i(t)$ has independent increments;
3. $\mathbb{P}[N_M^i(t + \Delta t) - N_M^i(t) = 1] = \lambda_M^i(t)\Delta t + o(\Delta t)$; and
4. $\mathbb{P}[N_M^i(t + \Delta t) - N_M^i(t) \geq 2] = o(\Delta t)$,

where $o(\Delta t)$ is little-o notation for $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$. We wish to prove that these properties also hold for $N_M^{1,2}(t)$.

The first property is easily seen for $N_M^{1,2}(t)$ as $N_M^{1,2}(0) = N_M^1(0) + N_M^2(0) = 0 + 0 = 0$. The second property holds by using that the conditional Poisson processes $N_M^i(t)$ are independent for $i = 1, 2$ and that each process individually has independent increments. Thus, $N_M^i(t)$ must also have independent increments.

For the third property, we have that

$$\begin{aligned}
\mathbb{P}[N_M^{1,2}(t + \Delta t) - N_M^{1,2}(t) = 1] &= \mathbb{P}[N_M^1(t + \Delta t) - N_M^1(t) = 1, N_M^2(t + \Delta t) - N_M^2(t) = 0] \\
&+ \mathbb{P}[N_M^1(t + \Delta t) - N_M^1(t) = 0, N_M^2(t + \Delta t) - N_M^2(t) = 1] \\
&= (\lambda_{M(t)}^1(t)\Delta t + o(\Delta t))(1 - \lambda_{M(t)}^2(t)\Delta t + o(\Delta t)) \\
&+ (1 - \lambda_{M(t)}^1(t)\Delta t + o(\Delta t))(\lambda_{M(t)}^2(t)\Delta t + o(\Delta t)) \\
&\quad \text{due to the independence between the two processes} \\
&= [\lambda_{M(t)}^1(t) + \lambda_{M(t)}^2(t)]\Delta t + o(\Delta t).
\end{aligned}$$

The final property can be shown in a similar manner. Thus, the combined process $N_M^{1,2}(t)$ is a conditional Poisson process with a claim intensity given by $\lambda^1_{M(t)}(t) + \lambda^2_{M(t)}(t)$. □

Generalising Proposition 2.1 to n event processes of form $\{M(t), X^i(t)\}_{i=1,\ldots,n}$ with the same intensity $\lambda_{M(t)}$ (for example, risk events arriving from n similar insurance policies or arrivals to n similar queues at airport check-in counters), the combined MMPP process has a claim intensity of $\lambda_{M(t)} \times n$. If this exposure process is allowed to vary over time, then the event intensity function can be thought of as $\lambda_{M(t)} \times \gamma(t)$, where $\gamma(t)$ is the volume or risk exposure measure. This interpretation also applies when the known function γ is a combination of multiple factors, in which case the multiplicative assumption is convenient. In this case, the overall event intensity would also be of the form $\lambda_{M(t)} \times \gamma(t)$, where γ is now the multiplicative combination of the impact of the considered factors.

2.2. An operational time scaling interpretation

The MMNPP can be alternatively interpreted as a operational time transformation of the a homogeneous MMPP, which provides a convenient, intuitive interpretation of the model. This argument follows in a similar manner to the transformation applicable to non-homogeneous Poisson processes and is outlined below.

We begin with the definition of an operational time scale for Poisson processes.

Definition 2.2. *Define an operational time scaling ρ, defined as $\rho(t) = \tau$. We then define a new process ν_M relative to the time τ such that*

$$N_M(t) = \nu_M(\rho(t)) = \nu_M(\tau) \tag{2.5}$$

where $\rho(t)$ is a monotonically increasing function in t with $\rho(0) = 0$.

The inverse operational time function is also defined in the standard way as the following:

Definition 2.3. *Using Definition 2.2, we define the inverse operational time function $\rho^{-1}(\tau)$ as*

$$\rho^{-1}(\tau) = \inf_t \{t \geq 0; \rho(t) = \tau\} \tag{2.6}$$

In the context of the MMNPP model, the operational time function ρ takes the form

$$\rho(t) = \int_0^t \gamma(s)ds. \tag{2.7}$$

It is shown in Theorem 2.4 below that the original MMNPP process $\{M(t), N_M(t); t \geq 0\}$ under this transformation becomes $\{M(t), \nu_M(\rho(t)); t \geq 0\}$, where $\nu_M(\rho(t))$ is a homogeneous Poisson process with constant conditional intensity rate $\lambda_{M(t)}$ at time t. Observe that in the above, the time scaling has only been applied to the conditional Poisson process and not the underlying Markovian process.

Theorem 2.4. *Define the time transform from t to $\rho(t)$ as the following:*

$$\underset{\textit{old time}}{t} \rightarrow \underset{\textit{new time}}{\tau} = \rho(t) = \int_0^t \gamma(s)ds$$

where $\rho(t)$ is a monotonically increasing function in t with $\rho(0) = 0$. This transform is applied to the conditional Poisson component of a MMNPP process $\{M(t), N_M(t); 0 \leq t \leq T\}$ with transitions $q_{i,j}$ and intensities $\lambda_{M(t)} \times \gamma(t)$. The resultant process is the homogeneous MMPP process $\{M(t), \nu_M(\rho(t)); 0 \leq t \leq T\}$ with transitions $q_{i,j}$ and intensities $\lambda_{M(t)}$.

Proof of Theorem 2.4 Although the result is very intuitive, for completeness we provide a proof here. We require the following preliminary results.

Lemma 2.5. *The time transformed process $\{M(t), \nu_M(\rho(t))\}$ retains the Markov property*

Proof of Lemma 2.5 For some constants s, t, it is known that

$$\begin{aligned}\mathbb{P}\left[\nu_M(\rho(t))|\nu_M(\rho(i)); 0 \leq i < s\right] &= \mathbb{P}\left[N(t)|N(i); 0 \leq i < s\right] \\ &= \mathbb{P}\left[N(t)|N(s)\right] \\ &= \mathbb{P}\left[\nu_M(\rho(t))|\nu_M(\rho(s))\right]\end{aligned}$$

Thus, the new process $\{M(t), \nu_M(\rho(t))\}$ retains the Markov property. □

Lemma 2.6. *The MMPP process $\{M(t), N_M(t)\}$ is homogeneous over period $[t_1, t_2]$ if and only if*

$$\mathbb{P}\left[N_M(t_1) = n, N_M(t_2) = n\right] = \exp\left[-\sum_{k=1}^{m} \lambda_{s_k}(u_k - u_{k-1})\right]$$

where k is a counter for the number of regime periods that occur in $[t_1, t_2]$, $u_0 = t_1$ and $u_m = t_2$.

Proof of Proof of Lemma 2.6 ($\Longrightarrow$) Beginning from the definition of a homogeneous MMPP,

$$\begin{aligned}\mathbb{P}\left[N_M(t_1) = n, N_M(t_2) = n\right] &= \exp\left[-\int_{t_1}^{t_2} \lambda_{M(s)} ds\right] \\ &= \exp\left[-\sum_{k=1}^{m} \int_{u_{k-1}}^{u_k} \lambda_{s_k} ds\right] \\ &= \exp\left[-\sum_{k=1}^{m} \lambda_{s_k}(u_k - u_{k-1})\right]\end{aligned}$$

($\Longleftarrow$) Beginning from the expression,

$$\lim_{h \to 0} \frac{1 - \mathbb{P}\left[N_M(t) = n, N_M(t+h) = n\right]}{h} = \lim_{h \to 0} \frac{1 - \exp\left[-\sum_{k=1}^{m} \lambda_{s_k}(u_k - u_{k-1})\right]}{h}$$

for the regime intervals that fall between times t and $t + h$. Note that as h approaches zero, the right expression eventually becomes a time interval within a single regime period, so that

$$\lim_{h \to 0} \frac{1 - \exp\left[-\sum_{k=1}^{m} \lambda_{s_k}(u_k - u_{k-1})\right]}{h} = \lim_{h \to 0} \frac{1 - \exp\left[-\lambda_{M(t)}(t - (t+h))\right]}{h} = \lambda_{M(t)}.$$

Thus, homogeneity is achieved as $\lambda_{M(t)}$ is the constant claim intensity at time t for state of the process M at time t. □

In order to satisfy the non-decreasing property of ρ, γ is restricted to being strictly positive (which is an intuitively correct restriction for an exposure measure). Using k as the number of regime changes between $[t_1, t_2]$, it can be obtained that

$$\begin{aligned}\mathbb{P}\left[\nu_M(\rho(t_1)) = n, \nu_M(\rho(t_2)) = n\right] &= \mathbb{P}\left[N_M(t_1) = n, N_M(t_2) = n\right] \\ &= \exp\left[-\int_{t_1}^{t_2} \lambda(t) dt\right] \\ &= \exp\left[-\int_{t_1}^{t_2} \lambda_{M(t)} \times \gamma(t) dt\right] \\ &= \exp\left[-\sum_{i=1}^{k} \int_{u_{i-1}}^{u_i} \lambda_{s_i} \times \gamma(t) dt\right] \\ &= \exp\left[-\sum_{i=1}^{k} \lambda_{s_i} \int_{u_{i-1}}^{u_i} \gamma(t) dt\right] \\ &= \exp\left[-\sum_{i=1}^{k} \lambda_{s_i}[\rho(u_i) - \rho(u_{i-1})]\right]\end{aligned}$$

The proof is concluded with the observation that the times $\rho(u_i)$ correspond to the regime change times u_i in the underlying Markov chain M. Thus, using Lemmas 2.5 and 2.6, the operational time adjustment as defined above changes the original non-homogeneous MMPP into a homogeneous MMPP when the non-homogeneous intensities are in the form $\lambda(t) = \lambda_{M(t)} \times \gamma(t)$. □

3. The EM calibration algorithm

The calibration of the MMNPP model utilises an adapted form of the EM algorithm for MMPPs originally proposed in Rydén (1994). Alternative methods exist in the literature for the homogeneous MMPP models such as moment-matching (Rossiter, 1989; Gusella, 1991), other likelihood-based estimation approaches (Meier-Hellstern, 1984) and Markov chain Monte Carlo (MCMC) sampling, the last of which has been popular in the literature in recent times. Rydén et al. (2008) examine the EM vs MCMC estimation for the general family of Hidden Markov models (HMM). The paper takes a computational perspective, and concludes that in the case where only point estimates are required or model comparison only requires likelihoods, the EM algorithm is preferred. In other cases, either approach has advantages and disadvantages and the comparison is less clear.

When applying the model to large data sets, as is the case in Section 4, the relative computational simplicity of the EM algorithm is beneficial. For MMPPs, closed form expressions for the estimators of the algorithm can be obtained, alleviating the high computational requirements. We extend those results to the MMNPP case in this paper. Additional improvements to the algorithm have been made to further reduce computation times and to resolve the previously mentioned issue of numerical instability.

This section is organised as follows. In Section 3.1, the complete likelihood function that is maximised through the EM algorithm is derived to take into account the non-homogeneity component. In Section 3.2, the closed form expressions for the E-step estimators are derived. In order to overcome numerical instability issues, a scaling approach is applied to the recursion equations, as outlined in Section 3.3. This approach is adapted from the algorithm for MMPPs from Roberts et al. (2006). We also utilise a result from Van Loan (1978) to increase computational efficiency when evaluating integrals of matrix exponentials, which was previously the most computationally burdensome component of the algorithm. Finally, during the maximisation step shown in Section 3.1, the traditional time-based estimators presented in Rydén (1996) are now split into a time-based and operational time-based estimator, which follows intuitively from the operational time transform described in the previous section. Section 3.5 concludes with a brief summary of the iterative calibration procedure.

3.1. Complete likelihood function of an MMNPP

The complete form of likelihood function, which assumes knowledge of both event arrival times and regime change times, is used in the calibration algorithm. We introduce the following notation:

1. Jumps between states/regimes occurs at time points $0 = u_0 < u_1 < u_2 < \ldots < u_m < u_{m+1} = T$, where T is the end of the observation period.
2. The regime time interval is $I_k = [u_{k-1}, u_k)$ for $1 \leq k \leq m+1$.
3. The regime of the process $M(t)$ in the interval I_k is given by $s_k \in \{1, 2, \ldots, r\}$.
4. The number of events occurring in regime period I_k is given by $z_k \in \mathbb{N}$.
5. The total number of events is denoted n.
6. Event arrivals occur within I_k at time points $t_{k,1}, t_{k,2}, \ldots, t_{k,z_k}$ and $t_{k,0} = u_{k-1}$.

Further, the following result is required in order to derive the complete likelihood function:

Lemma 3.1. *Consider the regime time interval $I_k = [u_{k-1}, u_k)$ of a non-homogeneous MMPP with claim intensity $\lambda(t) = \lambda_{s_k} \times \gamma(t)$ where the underlying Markov chain $M(t)$ does not change states, and λ_{s_k} is the constant component of the claim intensity within I_k. The joint density of claim times T_i, given that z_k claims occur within this interval, is given by the following expression*

$$\mathbb{P}\left[T_{k,1} = t_{k,1}, T_{k,2} = t_{k,2}, \ldots, T_{k,z_k} = t_{k,z_k} | N_M(u_k) - N_M(u_{k-1}) = z_k\right] = \frac{z_k!}{\left(\int_{u_{k-1}}^{u_k} \gamma(s)ds\right)^{z_k}} \prod_{i=1}^{z_k} \gamma(t_{k,i})$$

Proof of Lemma 3.1 It can be seen that

$$\begin{aligned}
&\mathbb{P}\left[T_{k,1}=t_{k,1},T_{k,2}=t_{k,2},\ldots,T_{k,z_k}=t_{k,z_k}|N_M(u_k)-N_M(u_{k-1})=z_k\right]\\
&=\frac{\mathbb{P}\left[T_{k,1}=t_{k,1},T_{k,2}=t_{k,2},\ldots,T_{k,z_k}=t_{k,z_k},N_M(u_k)-N_M(u_{k-1})=z_k\right]}{\mathbb{P}[N_M(u_k)-N_M(u_{k-1})=z_k]}
\end{aligned}$$

Note that the numerator of the above can be rewritten in the following manner:

$$\begin{aligned}
&\mathbb{P}\left[T_{k,1}=t_{k,1},T_{k,2}=t_{k,2},\ldots,T_{k,z_k}=t_{k,z_k},N_M(u_k)-N_M(u_{k-1})=z_k\right]\\
&=\mathbb{P}\left[z_k \text{ claims between } [u_{k-1},u_k) \text{ with claim times } t_{k,1},t_{k,2},\ldots,t_{k,z_k}\right]\\
&=\prod_{i=1}^{z_k}\left[\mathbb{P}\left[\text{ no claims until time } t_{k,i}\right]\times\mathbb{P}\left[\text{ claim at time } t_{k,i}\right]\right]\times\mathbb{P}\left[\text{ no claims until end of interval } u_k\right]\\
&=\prod_{i=1}^{z_k}\left[\exp\left(-\int_{t_{k,i-1}}^{t_{k,i}}\lambda_{s_k}\gamma(y)dy\times\lambda_{s_k}\right)\gamma(t_{k,i})\right]\times\exp\left(-\int_{t_{k,z_k}}^{u_k}\lambda_{s_k}\gamma(y)dy\right)
\end{aligned}$$

Thus, we have that

$$\begin{aligned}
&\mathbb{P}\left[T_{k,1}=t_{k,1},T_{k,2}=t_{k,2},\ldots,T_{k,z_k}=t_{k,z_k}|N_M(u_k)-N_M(u_{k-1})=z_k\right]\\
&=\frac{\prod_{i=1}^{z_k}\left[\exp\left(-\int_{t_{k,i-1}}^{t_{k,i}}\lambda_{s_k}\gamma(y)dy\times\lambda_{s_k}\right)\gamma(t_{k,i})\right]\times\exp\left(-\int_{t_{k,z_k}}^{u_k}\lambda_{s_k}\gamma(y)dy\right)}{\exp\left(-\int_{u_{k-1}}^{u_k}\lambda_{s_k}\gamma(y)dy\right)\frac{(\int_{u_{k-1}}^{u_k}\lambda_{s_k}\gamma(y)dy)^{z_k}}{z_k!}}\\
&=\frac{z_k!}{\left(\int_{u_{k-1}}^{u_k}\gamma(s)ds\right)^{z_k}}\prod_{i=1}^{z_k}\gamma(t_{k,i}).
\end{aligned}$$

□

The complete likelihood function of the MMNPP is then given in Equation (3.2), using Lemma 3.1 to determine the last term:

$$\begin{aligned}
\mathcal{L}^c=&\pi_{s_0}\left\{\prod_{k=1}^{m}q_{s_k}\exp\left(-q_{s_k}\Delta u_k\right)\times\frac{q_{s_k,s_{k+1}}}{q_{s_k}}\right\}\exp\left(-q_{s_{m+1}}\Delta u_{m+1}\right)\times\\
&\left\{\prod_{k=1}^{m+1}\frac{\left(\int_{u_{k-1}}^{u_k}\lambda_{s_k}\gamma(t)dt\right)^{z_k}}{z_k!}\exp\left(-\int_{u_{k-1}}^{u_k}\lambda_{s_k}\gamma(t)dt\right)\times\frac{z_k!}{\left(\int_{u_{k-1}}^{u_k}\gamma(t)dt\right)^{z_k}}\times\prod_{i=1}^{z_k}\gamma(t_{k,i})\right\} \qquad (3.1)\\
=&\pi_{s_0}\left\{\prod_{k=1}^{m}\exp\left(-q_{s_k}\Delta u_k\right)\times q_{s_k,s_{k+1}}\right\}\exp\left(-q_{s_{m+1}}\Delta u_{m+1}\right)\times\\
&\left\{\prod_{k=1}^{m+1}\left[\lambda_{s_k}^{z_k}\exp\left(-\int_{u_{k-1}}^{u_k}\lambda_{s_k}\gamma(t)dt\right)\prod_{i=1}^{z_k}\gamma(t_{k,i})\right]\right\} \qquad (3.2)
\end{aligned}$$

In the above, π_{s_0} denotes the starting probabilities of the underlying Markov chain. If prior information is unavailable, a discrete uniform distribution is applicable. By taking logs, we obtain

$$\begin{aligned}
\log\mathcal{L}^c=&\log\pi_{s_1}+\sum_{k=1}^{m+1}q_{s_k}\Delta u_k+\sum_{k=1}^{m}\log q_{s_k,s_{k+1}}\\
&+\sum_{k=1}^{m+1}z_k\log\lambda_{s_k}-\sum_{k=1}^{m+1}\lambda_{s_k}\int_{u_{k-1}}^{u_k}\gamma(t)dt+\sum_{k=1}^{m+1}\sum_{i=1}^{z_k}\log(\gamma(t_{k,i})). \qquad (3.3)
\end{aligned}$$

These sums are then reformulated in terms of the order of the Markov chain r instead of the number of regime changes. Note that the last term in Equation (3.3) is the summation of all the logs of the exposures at the claim arrival times. As these values are known, this term is rewritten as the constant $C(\gamma)$.

We also require the following E-step estimators, which are generally analogous to the estimators from Rydén (1994). The first two, $m_{i,j}$ and n_i can be interpreted to the number of transitions from state i to state j and the number of claims arriving in state i respectively.

$$m_{i,j} = \text{card}\{t : 0 < t \leq T, N_M(t-) = i, N_M(t) = j\} \tag{3.4}$$

$$n_i = \sum_{k=1}^{n} \mathbb{I}_{\{M(t_k)=i\}}, \tag{3.5}$$

where # is a count function for the number of events. The total amount of time spent in state i is

$$T_i = \int_0^T \mathbb{I}_{\{M(t)=i\}} dt. \tag{3.6}$$

Compared to the original EM algorithm, there is an extra "time in state" estimator T_i^*, given by

$$T_i^* = \int_0^T \mathbb{I}_{\{M(t)=i\}} \gamma(t) dt, \tag{3.7}$$

which reduces to the definition of T_i in the case of a homogeneous MMPP where $\gamma(t) = 1 \, \forall t$. Finally, the complete loglikelihood expression may be written as

$$\log \mathcal{L}^c = \sum_{i=1}^{r} \mathcal{I}_{\{M(0)=i\}} \log \pi_i - \sum_{i=1}^{r} T_i q_i + \sum_{i=1}^{r} \sum_{j=1, i \neq j}^{r} m_{i,j} \log q_{i,j} + \sum_{i=1}^{r} n_i \log \lambda_i - \sum_{i=1}^{r} \lambda_i T_i^* + C(\gamma) \tag{3.8}$$

It is interesting to note that the forms of T_i and T_i^* point to the use of two different time scales to evaluate the MMNPP parameters q and λ, where the relationship between the different time scales is defined through the exposure measure γ. This is intuitive given our previous operational time interpretation of the MMNPP process.

3.2. E-step of the adapted EM algorithm for MMNPPs

In the next section, we will derive closed form expression for the expectation of the estimators defined above. We require the evaluation of

$$\mathbb{E}_{(Q^0, \Lambda^0)}[\log \mathcal{L}^c(Q, \Lambda) | \nu_M(\rho(t)), 0 \leq t \leq T]. \tag{3.9}$$

In the above, $Q = \{q_{i,j}\}$, $\Lambda = \text{diag}(\lambda_{M_1}, \lambda_{M_2}, \ldots, \lambda_{M_r})$ and the current estimates of the parameters of the MMNPP at each iteration of the EM algorithm are denoted by $q_{i,j}^0$ and λ_i^0. Their respective matrices are correspondingly Q^0 and Λ^0. For notational convenience, we will also rewrite the $t_{k,i}$ terms in chronological order as t_j, $j = 1, \ldots, n$. The following expressions for the expectations of the M-step estimators are obtained:

$$\hat{m}_{i,j} = \mathbb{E}[m_{i,j} | \nu_M(\rho(t)), 0 \leq t \leq T] = \int_0^T \mathbb{P}[M(s-) = i, M(s) = j | \nu_M(\rho(t)), 0 \leq t \leq T] ds \tag{3.10}$$

$$\hat{n}_i = \mathbb{E}[n_i | \nu_M(\rho(t)), 0 \leq t \leq T] = \sum_{k=1}^{n} \mathbb{P}[M(t_k) = i | \nu_M(\rho(t)), 0 \leq t \leq T] \tag{3.11}$$

$$\hat{T}_i = \mathbb{E}[T_i | \nu_M(\rho(t)), 0 \leq t \leq T] = \int_0^T \mathbb{P}[M(s) = i | \nu_M(\rho(t)), 0 \leq t \leq T] ds \tag{3.12}$$

$$\hat{T}_i^* = \mathbb{E}[T_i^* | \nu_M(\rho(t)), 0 \leq t \leq T] = \int_0^T \mathbb{P}[M(s) = i | \nu_M(\rho(t)), 0 \leq t \leq T] \gamma(s) ds \tag{3.13}$$

The above expressions are decomposed into event inter-arrival intervals and written in terms of the fundamental regime transition probabilities for MMNPPs within each interval. These probabilities are be derived from the Chapman-Kolmogorov equations, taking into account the varying exposure measure at event arrival times, and are given in Theorems 3.2 and 3.3 below.

Theorem 3.2. *The probability of regime transitions without any event arrivals by time $t + t^*$, given the regime at time t, is given (in matrix form) by*

$$\bar{\boldsymbol{F}}(t, t^*) = \exp\left[(Q - \Lambda\gamma_{t^*})\, t^*\right], \tag{3.14}$$

where γ_{t^*} is the constant exposure measure that applies to the interval $[t, t + t^*]$.

Proof of Theorem 3.2 The transition probabilities without claim arrivals at time $t^* + t$ given the state at time t is given by

$$\bar{F}_{i,j}(t, t^*) = \mathbb{P}[M(t^* + t) = j, \nu_M(\rho(t^* + t)) = n | M(t^*) = i, \nu_M(\rho(t^*)) = n]$$

for the process $\{M(t), \nu_M(\rho(t))\}$, given that n claims have occurred before time t^*. For any arbitrarily small increment Δt,

$$\bar{F}_{i,j}(t + \Delta t, t^*) = \bar{F}_{i,j}(t, t^*)(1 - \sum_{i \neq j}^{r} q_{j,i}\Delta t - \lambda_j \Delta\rho(t^* + t)) + \sum_{k \neq j}^{r} \bar{F}_{i,k}(t, t^*) q_{k,j}\Delta t + o(\Delta t)$$

where $\Delta\rho(t^* + t) = \rho(t^* + t + \Delta t) - \rho(t^* + t)$. Taking the limit as Δt goes to zero and using the previously specified constraint that $q_i = \sum_{j \neq i} q_{i,j}$, the following Chapman-Kolmogorov differential equations are obtained:

$$\bar{F}'_{i,j}(t, t^*) = \bar{F}_{i,j}(t, t^*)(-q_j - \lambda_j \lim_{\Delta t \to 0} \frac{\Delta\rho(t^* + t)}{\Delta t}) + \sum_{k \neq j}^{r} \bar{F}_{i,k}(t, t^*) q_{k,j} \tag{3.15}$$

Note that the limit in the above equation is simply $\rho'(t^* + t)$, which equals $\gamma(t^* + t)$ from the Fundamental Theorem of Calculus. Then Equation (3.15) may be rewritten as

$$\bar{F}'_{i,j}(t, t^*) = \bar{F}_{i,j}(t, t^*)(-q_j - \lambda_j \gamma(t^* + t)) + \sum_{k \neq j}^{r} \bar{F}_{i,k}(t, t^*) q_{k,j}$$

which in matrix form is $\bar{\boldsymbol{F}}'(t, t^*) = \bar{\boldsymbol{F}}(t, t^*)(Q - \Lambda\gamma(t^* + t))$.

The assumption that $\gamma(t)$ is constant between event arrivals and exposure component changes is applied here so that the above matrix differential equation may be simplified. In the proposed EM algorithm, each part of $\bar{F}_{i,j}$ is evaluated over disjoint intervals where there are no event arrivals or changes in the exposure component. In this circumstance, $\gamma(t^* + t) = \gamma(t^*) = \gamma_{t^*}$, where γ_{t^*} is the constant exposure applicable in the interval $[t^*, t^* + t)$. The matrix form of the differential equation is then

$$\bar{\boldsymbol{F}}'(t, t^*) = \bar{\boldsymbol{F}}(t, t^*)(Q - \Lambda\gamma_{t^*}),$$

with solution

$$\bar{\boldsymbol{F}}(t, t^*) = \exp\left[(Q - \Lambda\gamma_{t^*})\, t\right].$$

This is derived using the following power series definition for exponential matrices:

$$\exp(\boldsymbol{A}t) = \boldsymbol{I} + t\boldsymbol{A} + \frac{t^2}{2!}\boldsymbol{A}^2 + \ldots,$$

which holds when matrix $\boldsymbol{A}$ is commutative. In this case, this property is clearly satisifed as as $(Q - \Lambda\gamma_{t^*})$ is independent of t. It follows using the above definition that

$$\frac{d}{dt}\exp(\boldsymbol{A}t) = \exp(\boldsymbol{A}t)\boldsymbol{A}.$$

Substitution of $\boldsymbol{A} = (Q - \Lambda\gamma_{t^*})$ and $\bar{\boldsymbol{F}}(t, t^*) = \exp\left[(Q - \Lambda\gamma_{t^*})\, t\right]$ then gives the required result. □

Theorem 3.3. *The probability of regime transitions with a single event arrival at time $t^* + t$, given the regime at time t, is expressed in matrix form as*

$$\boldsymbol{f}(t, t^*) = \exp\left[(Q - \Lambda\gamma_{t^*})\, t^*\right] \Lambda\gamma_{t+t^*} \tag{3.16}$$

where γ_{t+t^*} is the constant exposure measure applicable at time $t + t^*$.

Proof of Theorem 3.3 The formal definition of $f(t, t^*)$ is given by

$$\begin{aligned} f_{i,j}(t, t^*) &= \mathbb{P}[M(t^* + t) = j, \nu_M(\rho(t^* + t^-)) = n | M(t^*) = i, \nu_M(\rho(t^*)) = n] \\ &\quad \times \mathbb{P}[M(t^* + t) = j, \nu_M(\rho(t^* + t)) = n + 1 | M(t^* + t) = j, \nu_M(\rho(t^* + t^-)) = n] \end{aligned}$$

where $\rho(t^-) = \lim_{\epsilon \to 0} \rho(t - \epsilon)$. By following a similar set of derivations as in the proof for Theorem 3.2, it can be seen that

$$\begin{aligned} \boldsymbol{f}(t, t^*) &= \exp\left[(Q - \Lambda\gamma_{t^*})\, t\right] \Lambda\gamma(t + t^*) \\ &= \exp\left[(Q - \Lambda\gamma_{t^*})\, t\right] \Lambda\gamma_{t+t^*} \end{aligned}$$

as required. □

These theorems can also be intuitively understood as the non-homogeneous extensions of the corresponding MMPP expressions provided in Rydén (1996). They provide the fundamental building blocks with which the EM algorithm for the MMNPP model is constructed. However, without further adjustments, use of these theorems would require the exposure measure to be unchanged within event inter-arrival time intervals. In situations where event occurrences are sparse, this may be a strong assumption. To remedy this issue, another event type is introduced for a change in the piecewise constant exposure function γ. For ease of notation, we rewrite the event times in chronological order. Denote all event times t_k (the event of interest and changes in the exposure function). For each event time, there is an accompanying indicator δ_k to denote the type of event:

$$\delta_k = \begin{cases} 0, & \text{if it is a change in the exposure function,} \\ 1, & \text{if it is an event of interest.} \end{cases}$$

Applying this to Theorem 3.2, the matrix function $\boldsymbol{f}$ becomes

$$\boldsymbol{f}^{\delta_k}(t_{k-1}, t_k - t_{k-1}) = \begin{cases} \boldsymbol{f}(t_{k-1}, t_k - t_{k-1}), & \delta_k = 0, \\ \bar{\boldsymbol{F}}(t_{k-1}, t_k - t_{k-1}), & \delta_k = 1. \end{cases} = \begin{cases} \exp[(Q - \Lambda\gamma^*_{t_{k-1}})(t_k - t_{k-1})]\Lambda\gamma^*_{t_k}, & \delta_k = 0, \\ \exp[(Q - \Lambda\gamma^*_{t_{k-1}})(t_k - t_{k-1})], & \delta_k = 1. \end{cases} \tag{3.17}$$

The analytical expressions for the expectation of the E-step estimators are now written in terms of the above. Each estimator will be dealt with individually in the following subsections.

3.2.1. The estimator for the number of regime changes: $\hat{\boldsymbol{m}}_{\boldsymbol{i},\boldsymbol{j}}$

Beginning with the estimator used for the number of regime/state changes, $\hat{m}_{i,j}$, the integrand in Equation (3.10) is

$$\begin{aligned} &\mathbb{P}[M(s^-) = i, M(s) = j | \nu_M(\rho(t)), 0 \le t \le T] \\ &= \frac{\mathbb{P}[M(s) = j | M(s^-) = i]\mathbb{P}[M(s^-) = i]\mathbb{P}[\nu_M(\rho(t)), 0 \le t \le T | M(s^-) = i, M(s) = j]}{\mathbb{P}[\nu_M(\rho(t)), 0 \le t \le T]}. \end{aligned}$$

The initial term of the numerator is simply $q_{i,j}$ and by using the conditional independence, the last term of the numerator is decomposed to

$$\begin{aligned} &\mathbb{P}[\nu_M(\rho(t)), 0 \le t \le T | M(s^-) = i, M(s) = j] \\ &= \frac{\mathbb{P}[\nu_M(\rho(t)), 0 \le t < s, M(s^-) = i]}{\mathbb{P}[M(s^-) = i]} \mathbb{P}[\nu_M(\rho(t)), s \le t \le T | M(s) = j]. \end{aligned}$$

Thus, we obtain

$$\hat{m}_{i,j} = \frac{q_{i,j}}{\mathbb{P}[\nu_M(\rho(t)), 0 \le t \le T]} \times \int_0^T \mathbb{P}[\nu_M(\rho(t)), 0 \le t < s, M(s^-) = i]\mathbb{P}[\nu_M(\rho(t)), s \le t \le T | M(s) = j] ds \tag{3.18}$$

Each of these probabilities can be rewritten in terms of the transition densities from the previous section. The probability $\mathbb{P}[\nu_M(\rho(t)), 0 \le t \le T]$ is simply the likelihood of the observed claim arrivals which can be expressed in matrix form as

$$\mathbb{P}[\nu_M(\rho(t)), 0 \le t \le T] = \boldsymbol{\pi}(Q^0, \Lambda^0)\left[\prod_{k=1}^{n} \boldsymbol{f}^{\delta_k}(t_{k-1}, t_k - t_{k-1})\right]\mathbf{1} \tag{3.19}$$

where $\boldsymbol{\pi}(Q^0, \Lambda^0)$ is the starting distribution of state probabilities given the initial parameter estimates Q^0 and Λ^0, $\mathbf{1}$ is a $r \times 1$ vector of ones and $t_0 = 0$. Also, k is now a counter for both event types and n is the total number of original events and exposure changes. In a similar manner, the probabilities within the integrand are

$$\mathbb{P}[\nu_M(\rho(t)), 0 \le t < s, M(s^-) = i] = \boldsymbol{\pi}(Q^0, \Lambda^0)\left[\prod_{k=1}^{\nu_M(\rho(s)^-)} \boldsymbol{f}^{\delta_k}(t_k - t_{k-1}, t_{k-1})\right]\bar{\boldsymbol{F}}(t_{\nu_M(\rho(s)^-)}, s - t_{\nu_M(\rho(s)^-)})\mathbf{1}_i, \tag{3.20}$$

$$\mathbb{P}[\nu_M(\rho(t)), s \le t \le T | M(s) = j] = \mathbf{1}_j^\intercal \boldsymbol{f}^{\delta_{\nu_M(\rho(s)^-)+1}}(s, t_{\nu_M(\rho(s)^-)+1} - s)\left[\prod_{k=\nu_M(\rho(s)^-)+2}^{n} \boldsymbol{f}^{\delta_k}(t_{k-1}, t_k - t_{k-1})\right]\mathbf{1}, \tag{3.21}$$

where $\mathbf{1}_i$ is a vector of zeroes except for the i-th entry which is one. Thus, the final expression for the estimator of $m_{i,j}$ is

$$\hat{m}_{i,j} = \frac{q^0_{i,j}}{\boldsymbol{\pi}(Q^0, \Lambda^0)\left[\prod_{k=1}^{n} \boldsymbol{f}^{\delta_k}(t_{k-1}, t_k - t_{k-1})\right]\mathbf{1}} \times \left[\int_0^T \boldsymbol{\pi}(Q^0, \Lambda^0)\left[\prod_{k=1}^{\nu_M(\rho(s)^-)} \boldsymbol{f}^{\delta_k}(t_{k-1}, t_k - t_{k-1})\right]\bar{\boldsymbol{F}}(t_{\nu_M(\rho(s)^-)}, s - t_{\nu_M(\rho(s)^-)})\mathbf{1}_i \times \mathbf{1}_j^\intercal \boldsymbol{f}^{\delta_{\nu_M(\rho(s)^-)+1}}(s, t_{\nu_M(\rho(s)^-)+1} - s)\left[\prod_{k=\nu_M(\rho(s)^-)+2}^{n} \boldsymbol{f}^{\delta_k}(t_{k-1}, t_k - t_{k-1})\right]\mathbf{1} ds\right]. \tag{3.22}$$

3.2.2. The estimator for the number of events of interest in each regime: $\hat{\boldsymbol{n}}_{\boldsymbol{i}}$

Continuing from Equation (3.11), we obtain

$$\begin{aligned}
\hat{n}_i &= \mathbb{E}[n_i|\nu_M(\rho(t)), 0 \le t \le T] \\
&= \sum_{k=1}^{n} \mathbb{P}[M(t_k) = i|\nu_M(\rho(t)), 0 \le t \le T] \\
&= \sum_{k=1}^{n} \frac{\mathbb{P}[M(t_k) = i, \nu_M(\rho(t)), 0 \le t \le T]}{\mathbb{P}[\nu_M(\rho(t)), 0 \le t \le T]} \\
&= \frac{1}{\boldsymbol{\pi}(Q^0, \Lambda^0)\left[\prod_{k=1}^{n} \boldsymbol{f}^{\delta_k}(t_{k-1}, t_k - t_{k-1})\right]\mathbf{1}} \\
&\times \sum_{k=1}^{n} \boldsymbol{\pi}(Q^0, \Lambda^0)\left[\prod_{l=1}^{\nu_M(\rho(t_k))} \boldsymbol{f}^{\delta_l}(t_{l-1}, t_l - t_{l-1})\right]\mathbf{1}_i\mathbf{1}_i^{\intercal}\left[\prod_{l=\nu_M(\rho(t_k)+1}^{n} \boldsymbol{f}^{\delta_l}(t_{l-1}, t_l - t_{l-1})\right]\mathbf{1}. \qquad (3.23)
\end{aligned}$$

3.2.3. The estimator for the time spent in each regime: $\hat{\boldsymbol{T}}_{\boldsymbol{i}}$

Using the expression for $\hat{T}_i$ from Equation (3.12) and following a similar procedure to the $\hat{m}_{i,j}$ calculations, it can be seen that

$$\begin{aligned}
\hat{T}_i &= \int_0^T \mathbb{P}[M(s) = i|\nu_M(\rho(t)), 0 \le t \le T]ds \\
&= \frac{1}{\boldsymbol{\pi}(Q^0, \Lambda^0)\left[\prod_{k=1}^{n} \boldsymbol{f}^{\delta_k}(t_{k-1}, t_k - t_{k-1})\right]\mathbf{1}} \\
&\times \int_0^T \mathbb{P}[M(s) = i, \nu_M(\rho(t)); 0 \le t < s]\mathbb{P}[\nu_M(\rho(t)); s \le t \le T|M(s) = i]ds \\
&= \frac{1}{\boldsymbol{\pi}(Q^0, \Lambda^0)\left[\prod_{k=1}^{n} \boldsymbol{f}^{\delta_k}(t_{k-1}, t_k - t_{k-1})\right]\mathbf{1}} \\
&\times \left[\int_0^T \boldsymbol{\pi}(Q^0, \Lambda^0)\left[\prod_{l=1}^{\nu_M(\rho(s)^-)} \boldsymbol{f}^{\delta_l}(t_{l-1}, t_l - t_{l-1})\right]\bar{\boldsymbol{F}}(t_{\nu_M(\rho(s)^-)}, s - t_{\nu_M(\rho(s)^-)})\mathbf{1}_i\right. \\
&\times \left.\mathbf{1}_j^{\intercal}\boldsymbol{f}^{\delta_{\nu_M(\rho(s)^-)+1}}(s, t_{\nu_M(\rho(s)^-)+1} - s)\left[\prod_{l=\nu_M(\rho(s)^-)+2}^{n} \boldsymbol{f}^{\delta_l}(t_{l-1}, t_l - t_{l-1})\right]\mathbf{1}ds\right]. \qquad (3.24)
\end{aligned}$$

In the above, the event $\{M(s^-) = i\}$ has been replaced with $\{M(s) = i\}$ which does not change the density due to the fact that $\{M(s)\}$ is continuous in probability.

3.2.4. The estimator for the operational time spent in each regime: $\hat{\boldsymbol{T}}_{\boldsymbol{i}}^{*}$

The results from the previous section are easily adapted to obtain the following estimator for T_i^*:

$$\begin{aligned}
\hat{T}_i^* &= \int_0^T \mathbb{P}[M(s) = i | \nu_M(\rho(t)), 0 \le t \le T]\gamma(s)ds \\
&= \frac{1}{\boldsymbol{\pi}(Q^0, \Lambda^0)\left[\prod_{k=1}^n \boldsymbol{f}^{\delta_k}(t_{k-1}, t_k - t_{k-1})\right]\mathbf{1}} \\
&\times \left[\int_0^T \boldsymbol{\pi}(Q^0, \Lambda^0)\left[\prod_{l=1}^{\nu_M(\rho(s)^-)} \boldsymbol{f}^{\delta_l}(t_{l-1}, t_l - t_{l-1})\right]\bar{\boldsymbol{F}}(t_{\nu_M(\rho(s)^-)}, s - t_{\nu_M(\rho(s)^-)})\mathbf{1}_i\right. \\
&\times \left.\mathbf{1}_j^\intercal \boldsymbol{f}^{\delta_{\nu_M(\rho(s)^-)+1}}(s, t_{\nu_M(\rho(s)^-)+1} - s)\left[\prod_{l=\nu_M(\rho(s)^-)+2}^{n} \boldsymbol{f}^{\delta_l}(t_{l-1}, t_l - t_{l-1})\right]\mathbf{1}\gamma(s)ds\right] \qquad (3.25)
\end{aligned}$$

3.3. The scaled recursion algorithms in the adapted EM algorithm

The estimators derived in the previous section involve operations on the product of matrices, resulting in large computational cost. In Rydén (1996), the original versions of the above estimators were calculated using a recursive algorithm that took advantage of the decomposition into disjoint event inter-arrival times. Roberts et al. (2006) proposed a methodology to scale these forward-backward recursions to deal with the numerical instability of the EM algorithm. A further innovation outlined in this paper allowed the efficient computation of the matrix exponentials in the $\boldsymbol{f}$ and $\bar{\boldsymbol{F}}$ components of the estimators. In this next section, this approach is adapted to the MMNPP context and applied the derived estimators above.

Firstly, some further notation is established for convenience. Let

$$\begin{aligned}
c_k = \mathbb{P}[&\nu_M(\rho(t_k)) = k, \nu_M(\rho(t_k)^-) = k-1 | \nu_M(\rho(t_{k-1})) = k-1, \nu_M(\rho(t_{k-1})^-) = k-2, \\
&\nu_M(\rho(t_{k-2})) = k-2, \nu_M(\rho(t_{k-2})^-) = k-3, \ldots], \qquad (3.26)
\end{aligned}$$

be the conditional probability of observing the k-th event arrival at time t_k given all previous event arrival times. Also, define $c_1 = \mathbb{P}[\nu_M(\rho(t_1)) = 1, \nu_M(\rho(t_1)^-) = 0]$. An alternative formulation for the joint likelihood of observed event arrivals can be obtained using these c_k's:

$$\mathbb{P}[\nu_M(\rho(t)), 0 \le t \le T] = \prod_{k=1}^n c_k. \qquad (3.27)$$

In a similar manner to the approach in Roberts et al. (2006), the forwards/backward recursion equations are defined by

$$L(k) = \boldsymbol{\pi}(Q^0, \Lambda^0)\prod_{l=1}^k \frac{\boldsymbol{f}^{\delta_l}(t_{l-1}, t_l - t_{l-1})}{c_l}, \quad R(k) = \prod_{l=k}^n \frac{\boldsymbol{f}^{\delta_l}(t_{l-1}, t_l - t_{l-1})}{c_l}\mathbf{1}, \qquad k = 1, \ldots, n \qquad (3.28)$$

with $L(0) = \boldsymbol{\pi}(Q^0, \Lambda^0)$ and $R(n+1) = \mathbf{1}$. Note that for each k, $L(k)$ is a $1 \times r$ vector while $R(k)$ is a $r \times 1$ vector. Using the fact that

$$\begin{aligned}
c_k &= \boldsymbol{\pi}(Q^0, \Lambda^0)\prod_{l=1}^{k-1} \frac{\boldsymbol{f}^{\delta_l}(t_{l-1}, t_l - t_{l-1})}{c_l}\boldsymbol{f}^{\delta_l}(t_{l-1}, t_l - t_{l-1})\mathbf{1} \\
&= L(k-1)\boldsymbol{f}^{\delta_l}(t_{l-1}, t_l - t_{l-1})\mathbf{1}, \qquad (3.29)
\end{aligned}$$

the recursive functions can be iteratively computed as

$$L(k) = \frac{L(k-1)\boldsymbol{f}^{\delta_k}(t_{k-1}, t_k - t_{k-1})}{L(k-1)\boldsymbol{f}^{\delta_k}(t_{k-1}, t_k - t_{k-1})\mathbf{1}}, \quad R(k) = \frac{\boldsymbol{f}^{\delta_k}(t_{k-1}, t_k - t_{k-1})R(k+1)}{L(k-1)\boldsymbol{f}^{\delta_k}(t_{k-1}, t_k - t_{k-1})\mathbf{1}}. \qquad (3.30)$$

The L and R functions can also be expressed in an alternative manner that is more interpretable. The i-th element in each vector is

$$L(k)_i = \mathbb{P}[M(t_k) = i | \nu_M(\rho(t_{k-1})) = k-1, \nu_M(\rho(t_{k-1})^-) = k-2, \ldots, \nu_M(\rho(t_1)) = 1, \nu_M(\rho(t_1)^-) = 0], \tag{3.31}$$

$$R(k)_i = \frac{\mathbb{P}[\nu_M(\rho(t_n)) = n, \nu_M(\rho(t_{n-1})^-) = n-1, \ldots, \nu_M(\rho(t_k)) = k, \nu_M(\rho(t_k)^-) = k-1 | M(t_{k-1}) = i]}{B_k}, \tag{3.32}$$

$$\begin{aligned} \text{with } B_k = \mathbb{P}[&\nu_M(\rho(t_n)) = n, \nu_M(\rho(t_{n-1})^-) = n-1, \ldots, \nu_M(\rho(t_k)) = k, \nu_M(\rho(t_k)^-) = k-1 | \\ &\nu_M(\rho(t_{k-1})) = k-1, \nu_M(\rho(t_{k-1})^-) = k-2, \ldots, \nu_M(\rho(t_1)) = 1, \nu_M(\rho(t_1)^-) = 0]. \end{aligned}$$

The scaling procedure above overcomes the issue of numerical underflow when implementing this method on large data sets. In the following subsections, the scaled forwards/backwards recursions above will be applied to the E-step estimators derived in Section 3.2. It should also be noted that the derivations utilise the disjoint intervals of event/exposure change inter-arrivals where the exposure measure $\gamma(t)$ is constant so the results from Theorems 3.2, 3.3 and Equation (3.17) are applicable.

3.3.1. The rescaled recursive algorithm for $\hat{\boldsymbol{m}}_{\boldsymbol{i,j}}$

Beginning with the estimator for $\hat{m}_{i,j}$ obtained in Equation (3.22), factoring out the bracketed elements from the integrand and substituting the recursive functions L and R gives

$$\begin{aligned} \hat{m}_{i,j} = &\frac{q^0_{i,j}}{\boldsymbol{\pi}(Q^0, \Lambda^0)\left[\prod_{k=1}^n \boldsymbol{f}^{\delta_k}(t_{k-1}, t_k - t_{k-1})\right]\mathbf{1}} \\ &\times \left[\int_0^T \boldsymbol{\pi}(Q^0, \Lambda^0)\left[\prod_{l=1}^{\nu_M(\rho(s)^-)} \boldsymbol{f}^{\delta_l}(t_{l-1}, t_l - t_{l-1})\right]\bar{\boldsymbol{F}}(t_{\nu_M(\rho(s)^-)}, s - t_{\nu_M(\rho(s)^-)})\mathbf{1}_i \right. \\ &\left. \times \mathbf{1}_j^\intercal \boldsymbol{f}^{\delta_{\nu_M(\rho(s)^-)+1}}(s, t_{\nu_M(\rho(s)^-)+1} - s)\left[\prod_{l=\nu_M(\rho(s)^-)+2}^{n} \boldsymbol{f}^{\delta_l}(t_{l-1}, t_l - t_{l-1})\right]\mathbf{1} ds\right] \end{aligned} \tag{3.33}$$

$$\begin{aligned} &= \sum_{k=1}^n \left[\frac{q^0_{i,j}}{c_k} \times \left(\boldsymbol{\pi}(Q^0, \Lambda^0)\prod_{l=1}^{k-1}\frac{\boldsymbol{f}^{\delta_l}(t_{l-1}, t_l - t_{l-1})}{c_l}\right)\right. \\ &\times \int_{t_{k-1}}^{t_k^-} \bar{\boldsymbol{F}}(t_{k-1}, s - t_{k-1}))\mathbf{1}_i \mathbf{1}_j^\intercal \boldsymbol{f}^{\delta_k}(s, t_k - s) ds \\ &\left.\times \left(\prod_{l=k+1}^{n}\frac{\boldsymbol{f}^{\delta_l}(t_{l-1}, t_l - t_{l-1})}{c_l}\mathbf{1}\right)\right] \end{aligned} \tag{3.34}$$

$$= q^0_{i,j}\sum_{k=1}^n \frac{1}{c_k}\int_{t_{k-1}}^{t_k^-} \mathbf{1}_j^\intercal \boldsymbol{f}^{\delta_k}(s, t_k - s) R(k+1) L(k-1) \bar{\boldsymbol{F}}(t_{k-1}, s - t_{k-1}))\mathbf{1}_i ds \tag{3.35}$$

Using matrix notation, this becomes

$$\boldsymbol{m}^\intercal = \left(Q^0\right)^\intercal \odot \sum_{k=1}^n \frac{1}{c_k}\int_{t_{k-1}}^{t_k^-} \boldsymbol{f}^{\delta_k}(s, t_k - s) R(k+1) L(k-1) \bar{\boldsymbol{F}}(t_{k-1}, s - t_{k-1}) ds \tag{3.36}$$

$$= \left(Q \odot \sum_{k=1}^n \frac{\mathcal{I}_k^\intercal}{c_k}\right)^\intercal, \tag{3.37}$$

where $\odot$ is the Hadamard product, otherwise known as the element-wise product for matrices. Also, the integral component above is represented by $\mathcal{I}_k$ where

$$\mathcal{I}_k = \int_{t_{k-1}}^{t_k^-} \boldsymbol{f}^{\delta_k}(s, t_k - s) R(k+1) L(k-1) \bar{\boldsymbol{F}}(t_{k-1}, s - t_{k-1}) ds \tag{3.38}$$

$$= \begin{cases} \int_0^{t_k^- - t_{k-1}} \exp[(Q - \Lambda\gamma^*_{t_{k-1}})(t_k - t_{k-1} - s)] \Lambda\gamma^*_{t_k} R(k+1) L(k-1) \exp[(Q - \Lambda\gamma^*_{t_{k-1}})(t_k - t_{k-1} - s)] ds, & \delta_k = 0, \\ \int_0^{t_k^- - t_{k-1}} \exp[(Q - \Lambda\gamma^*_{t_{k-1}})(t_k - t_{k-1} - s)] R(k+1) L(k-1) \exp[(Q - \Lambda\gamma^*_{t_{k-1}})(t_k - t_{k-1} - s)] ds, & \delta_k = 1. \end{cases} \tag{3.39}$$

where again, the integration is performed over disjoint time intervals where the exposure measure $\gamma(t)$ does not change. Thus, within each inter-arrival interval $\mathcal{I}_k$, the constant exposure measure will be denoted γ_k. The integrals within the cases above can be evaluated using the algorithm presented in Van Loan (1978). Define the following $2r \times 2r$ block-triangular matrices:

$$C_k^{\delta_k} = \begin{cases} \begin{bmatrix} Q - \Lambda\gamma^*_{t_{k-1}} & \Lambda\gamma_{t_k} R(k+1) L(k-1) \\ 0 & Q - \Lambda\gamma^*_{t_{k-1}} \end{bmatrix}, & \delta_k = 0, \\ \begin{bmatrix} Q - \Lambda\gamma^*_{t_{k-1}} & R(k+1) L(k-1) \\ 0 & Q - \Lambda\gamma^*_{t_{k-1}} \end{bmatrix}, & \delta_k = 1. \end{cases} \tag{3.40}$$

$\mathcal{I}_k$ can then be evaluated as the upper-right block of the matrix $\exp(C_k^{\delta_k}(t_k - t_{k-1}))$. This exponential can be easily and efficiently obtained using a variety of methods such as by using the diagonal Padé's approximant with repeated squaring (Moler and Van Loan, 2003). Combining all of the above, the required estimator of $\boldsymbol{m}$ is obtained.

3.3.2. The rescaled recursive algorithm for $\hat{\boldsymbol{n}}_i$

$$\begin{aligned} \hat{n}_i &= \frac{1}{\boldsymbol{\pi}(Q^0, \Lambda^0) \left[\prod_{k=1}^n \boldsymbol{f}^{\delta_k}(t_{k-1}, t_k - t_{k-1}) \right] \mathbf{1}} \\ &\times \sum_{k=1}^n \boldsymbol{\pi}(Q^0, \Lambda^0) \left[\prod_{l=1}^{\nu_M(\rho(t_k))} \boldsymbol{f}^{\delta_l}(t_{l-1}, t_l - t_{l-1}) \right] \mathbf{1}_i \mathbf{1}_i^\intercal \left[\prod_{l=\nu_M(\rho(t_k))+1}^{n} \boldsymbol{f}^{\delta_l}(t_{l-1}, t_l - t_{l-1}) \right] \mathbf{1} \\ &= \sum_{k=1}^n \boldsymbol{\pi}(Q^0, \Lambda^0) \left[\prod_{l=1}^{k} \frac{\boldsymbol{f}^{\delta_l}(t_{l-1}, t_l - t_{l-1})}{c_l} \right] \mathbf{1}_i \mathbf{1}_i^\intercal \left[\prod_{l=k+1}^{n} \frac{\boldsymbol{f}^{\delta_l}(t_{l-1}, t_l - t_{l-1})}{c_l} \right] \mathbf{1} \\ &= \left[\sum_{k=1}^n L(k)^\intercal \odot R(k+1) \right]_i. \end{aligned} \tag{3.41}$$

Thus, in matrix form,

$$\hat{\boldsymbol{n}} = \sum_{k=1}^n L(k)^\intercal \odot R(k+1). \tag{3.42}$$

3.3.3. The rescaled recursive algorithm for $\hat{\boldsymbol{T}}_i$

The computation of $\hat{T}_i$ is very simple due to the following relationship between $\hat{m}_{i,i}$ and $\hat{T}_i$ (Equations (3.22) and (3.24) respectively):

$$\hat{T}_i = \frac{\hat{m}_{i,i}}{q^0_{i,i}}. \tag{3.43}$$

3.3.4. The rescaled recursive algorithm for $\hat{\boldsymbol{T}}_{\boldsymbol{i}}^{*}$

The calculations for $\hat{T}_i^*$ are slightly more complex than for $\hat{\boldsymbol{T}}_{\boldsymbol{i}}$. Continuing from Equation 3.25,

$$\begin{aligned}
\hat{T}_i^* &= \frac{1}{\boldsymbol{\pi}(Q^0,\Lambda^0)\left[\prod_{k=1}^n \boldsymbol{f}^{\delta_k}(t_{k-1},t_k-t_{k-1})\right]\mathbf{1}} \\
&\times \left[\int_0^T \boldsymbol{\pi}(Q^0,\Lambda^0)\left[\prod_{l=1}^{\nu_M(\rho(s)^-)} \boldsymbol{f}^{\delta_l}(t_{l-1},t_l-t_{l-1})\right]\bar{\boldsymbol{F}}(t_{\nu_M(\rho(s)^-)},s-t_{\nu_M(\rho(s)^-)})\mathbf{1}_i\right. \\
&\times \left.\mathbf{1}_j^\intercal \boldsymbol{f}^{\delta_{\nu_M(\rho(s)^-)+1}}(s,t_{\nu_M(\rho(s)^-)+1}-s)\left[\prod_{l=\nu_M(\rho(s)^-)+2}^{n} \boldsymbol{f}^{\delta_l}(t_{l-1},t_l-t_{l-1})\right]\mathbf{1}\gamma(s)ds\right] \qquad (3.44) \\
&= \sum_{k=1}^n \left[\frac{1}{c_k}L(k-1)\int_{t_{k-1}}^{t_k^-} \bar{\boldsymbol{F}}(t_{k-1},s-t_{k-1}))\mathbf{1}_i\mathbf{1}_i^\intercal \boldsymbol{f}^{\delta_k}(s,t_k-s)\gamma_k^* ds R(k+1)\right]. \qquad (3.45)
\end{aligned}$$

These expressions were similar to the ones obtained when evaluating $\hat{m}_{i,i}$ in Section 3.3.1, except that there is no $q_{i,j}$ constant in the summation and $i=j$ in this case. It is seen that

$$\boldsymbol{T}^* = \operatorname{diag}\left[\sum_{k=1}^n \frac{\gamma_k^* \mathcal{I}_k^\intercal}{c_k}\right] \qquad (3.46)$$

where $\mathcal{I}_k$ has already been evaluated in the calculation of $\hat{\boldsymbol{m}}$ and the function diag denotes taking the diagonal entries of the matrix (as $\hat{\boldsymbol{T}}^*$ is a $r \times 1$ vector).

3.4. The M-step of the EM algorithm

In comparison to the E-step, the M-step of the EM algorithm is very simple. Under the constraint that $q_i = \sum_{j\neq i} q_{i,j}$, the unique maximum likelihood estimators of the MMNPP parameters are given in terms of the EM algorithm parameters by the following expressions:

$$\hat{q}_{i,j} = \frac{m_{i,j}}{T_i}, \quad \hat{\lambda}_i = \frac{n_i}{T_i^*}. \qquad (3.47)$$

This is derived using the standard MLE procedure applied to the loglikelihood in Equation (3.8).

3.5. Summary of adapted EM algorithm for MMNPPs

In summary, one iteration of the EM algorithm is given by the following:

1. Using current estimates of Q and Λ, evaluate $\hat{\boldsymbol{m}}, \hat{\boldsymbol{n}}, \hat{\boldsymbol{T}}$ and $\hat{\boldsymbol{T}}^*$.
2. Using current estimates of $\hat{\boldsymbol{m}}, \hat{\boldsymbol{n}}, \hat{\boldsymbol{T}}$ and $\hat{\boldsymbol{T}}^*$, evaluate Q and Λ.

The last consideration is the determination of when the algorithm should be terminated. Two simple approaches may be applied. One method is to consider the change to the matrices $\boldsymbol{Q}$ and $\boldsymbol{\Lambda}$ and to stop the algorithm when the difference between successive EM iterations is smaller than some tolerance value. Another approach is to instead compare the value of the log-likelihood function, which is non-decreasing for successive iterations of the EM algorithm. As in the case from Roberts et al. (2006), it is clear from Equation (3.27) that the log-likelihood is given by

$$\begin{aligned}
&\log \mathbb{P}[\nu_M(\rho(t_k)) = k, \nu_M(\rho(t_k)^-) = k-1, \nu_M(\rho(t_{k-1})) = k-1, \nu_M(\rho(t_{k-1})^-) = k-2, \\
&\quad \nu_M(\rho(t_{k-2})) = k-2, \nu_M(\rho(t_{k-2})^-) = k-3, \ldots, \nu_M(\rho(t_1)) = 1, \nu_M(\rho(t_1)^-) = 0] \\
&= \sum_{k=1}^n \log \mathbb{P}[\nu_M(\rho(t_k)) = k, \nu_M(\rho(t_k)^-) = k-1 | \nu_M(\rho(t_{k-1})) = k-1, \nu_M(\rho(t_{k-1})^-) = k-2, \\
&\quad \nu_M(\rho(t_{k-2})) = k-2, \nu_M(\rho(t_{k-2})^-) = k-3, \ldots, \nu_M(\rho(t_1)) = 1, \nu_M(\rho(t_1)^-) = 0], \\
&= \sum_{k=1}^n \log c_k. \qquad (3.48)
\end{aligned}$$

This value is readily obtainable as it is explicitly calculated during the calibration procedure. In addition, the ease of obtaining the log-likelihood facilitates the use of likelihood-based model selection procedures. When choosing the optimal order of the MMNPP, the standard AIC and BIC criteria, as well as their extensions, are straightforward to evaluate. This is discussed in greater detail in Section 4.4.

Various simulation studies were carried out to ascertain the accuracy of the proposed algorithm. The results demonstrated a high level of accuracy and a number of other favorable attributes such as fast convergence of parameter estimates and robustness to initial inputs. For conciseness, the full details have been omitted from this paper but are available for interested readers in Avanzi et al. (2018).

4. Numerical illustration: Motor insurance claims

In this section, we apply the MMNPP model to a set of motor insurance real data to determine what insights can be gained in order to assist insurance operations in a real world setting; see also Section 1.3 for additional background. Insurance data sets have several characteristics that call for the additional features introduced in this paper. These data sets generally consists of a large number of event observations which, as seen in the previous section, facilities more accurate detection of the underlying regimes. Further, the modelling of claims in motor lines of business is standard, which assists in the analysis required to determine the appropriate exposure measure γ. How we decided which trends to include in γ is outlined in the Section 4.2. The residual components are modelled by a MMNPP. Additional insights available through the MMNPP model are discussed.

4.1. The AUSI data set

The Allianz, University of New South Wales, Suncorp and Insurance Australia Group (AUSI) dataset was developed as part of an Australian Research Council Linkage Project titled *"Modelling claim dependencies for the general insurance industry with economic capital in view: an innovative approach with stochastic processes"*, which is supported by the Australian Research Council and several major industry partners, namely the Allianz Group, the Suncorp Group and the Insurance Australian Group. The data set contains non-life insurance claims records in a standard format containing information on each claim transaction such as occurrence and notification dates, claim states and finalisation dates. It also includes a policy file which describes underwritten policies, containing information such as date of inception/expiry and sum insured.

This case study investigates the Private Motor line of business from the 1st of January, 2005 to the 31st of December, 2010. It is noted that the data set extends to the end of 2014 and due to the short-tailed nature of the reporting delay for motor claims, all accidents that occurred during this period are expected to have been reported in the data set. As a result, considerations of censoring and reporting delay, such as the ones investigated in Crevecoeur et al. (2019) are not necessary here. Further, all claims related to catastrophe events have been removed from this data set. In total, the utilized subset of the data consists of significantly more than half a million claim observations.

Because the level of granularity of the claims data is not finer than daily, an extra assumption is required to distribute the claim arrival times throughout the day, as the MMNPP is a continuous time model. In order to not introduce any artificial variability, the claims are assumed to arrive at equal intervals throughout the day. This approach was compared using simulation studies against simulating claim arrivals using a uniform distribution but the latter approach induced a small amount of additional variability, clouding results in terms of regime identification. Further, issues arose in terms of the appropriate method to aggregate these simulations. Finally, the use of simulations meant that the model had to be fit several times, increasing the computational burden of the approach. Thus, the more convenient approach of an equidistant distribution was taken instead.

4.2. The volume/exposure component

Within the context of motor claims analysis, several factors are typically incorporated into the frequency perturbation measure γ. Clearly, the number of policies in force are a relevant exposure component but other trends such as seasonality also exist. An over-dispersed Poisson (ODP) GLM was chosen as an appropriate

model to determine the exposure process γ. Such an approach is standard in actuarial modelling. Covariates of interest were identified using a combination of data analysis and domain knowledge. Statistical significance was then tested and the final model consisted of the features relating to the number of policies in force (introduced as an offset), various forms of seasonality such as weekday/weekend, public holiday, month and day of month fluctuations, and a linear residual component to capture the gradual decay of motor claim frequencies over time.

Further analysis was conducted of γ over time and no additional events involving sudden changes in the exposure component were deemed necessary. This is partly due to the large number of claim events within the data set.

4.3. Evidence of hidden regimes

A natural question that may arise is whether hidden regimes exist within the data after having accounted for the data features above. A number of statistical results indicate that this is indeed the case. Firstly, the dispersion parameter calibrated in the ODP GLM was 3.16, indicating some level of over-dispersion within the data relative to the standard Poisson process.

Further, a Wald-Wolfowitz runs test was also applied in order to test for persistence of the frequency process within regimes. This procedure tested for randomness in the runs above and below the Poisson estimate of the intensity, having adjusted for the factors in the volume component outlined in the previous subsection. The obtained p-value was 0, suggesting that the frequencies tended to remain at certain levels over time.

Finally, a white-noise residual testing procedure was utilised for order selection, and an order 4 MMNPP was identified as optimal; see Section 4.4 for details.

4.4. Model order selection

An important aspect of the model is the selection of the order of the MMNPP. Some papers such as Landon et al. (2013) have proposed procedures for doing so although discussion of this aspect is relatively uncommon and generally, simple models such as order 2 or 3 are chosen for the purposes of demonstration.

The proposed methodology for order selection in the absence of (or to complement) expert opinion is as follows. For a chosen order x (starting at order 2, provided there is evidence of regimes as discussed in Section 4.3), daily residuals are calculated as the difference between the observed number of daily claims and the predicted number of claims. A white-noise test then applied, and if the white-noise property is rejected, then the MMNPP model of order $x+1$ is fitted instead and the process iterates until a suitable p-value for the hypothesis of white noise residuals is obtained. In the empirical case study, the Bartlett B test for white noise is used (see Bartlett, 1967).

4.5. Modelling results

The final calibrated MMNPP parameters for the order 4 model were

$$Q = \begin{bmatrix} -0.39 & 0.08 & 0.28 & 0.02 \\ 0.00 & -0.06 & 0.05 & 0.00 \\ 0.38 & 0.05 & -0.43 & 0.00 \\ 1.00 & 0.00 & 0.00 & -1.00 \end{bmatrix}, \quad \lambda = (135, 177, 204, 518)\,. \tag{4.1}$$

The distance between Poisson claim intensities suggest that the identified regimes are distinct. The generator matrix provides information about to the transition probabilities between regimes averaged over the entire period and modellers can also obtain other inferences such as how many claims occurred within each regime and how much time was spent in each regime through the estimators used in the EM algorithm. This is displayed in Table 1. For confidentiality reasons, the number of claims in each state statistic has been provided as a proportion of total claims instead.

Statistic of interest	**Estimator**
Number of changes to each state	$\begin{bmatrix} -20.2 & 5.9 & 12.4 & 1.0 \\ 4.4 & -106.1 & 101.3 & 0.0 \\ 98.9 & 14.8 & -113.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & -1.0 \end{bmatrix}$
Proportion of claims in each state	(2.3%, 88.9%, 8.7%, 0.0%)
Time spent in each state (days)	(52.4, 1870.5, 267.1, 1.0)

Table 1: M-estimator quantities of interest

Furthermore, Equation (3.48) allows for the generation of Figure 1, which depicts the most likely regime at each claim arrival time. Several insights are able to be drawn here. Firstly, there is a single day (a public holiday coinciding with a weekend) with an extremely high number of claims compared to expectations, sufficiently so to warrant a separate regime (regime 4). As insurance claims related to catastrophic events have been removed, such an explanation is unlikely. This result demonstrates some robustness of the MMNPP model to outliers, as the observation has been objectively identified and its impact on the other observations and regimes has been mitigated. This mitigation effect is due to the fact that the high intensity associated with the fourth regime will not skew the parameter estimates for the other regimes as significantly. For example, if the order of the MMNPP model was instead three, then this outlying observation would have increased the frequency parameter associated with the highest regime.

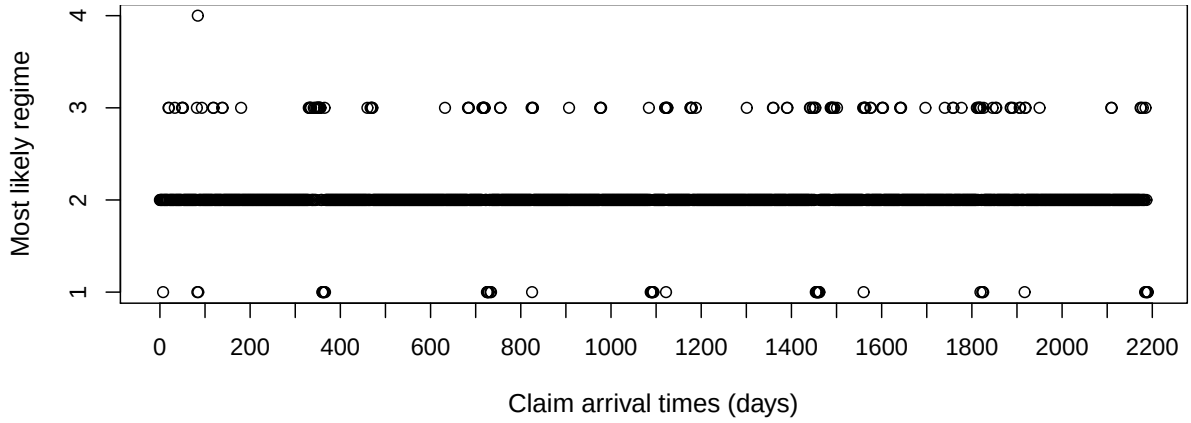

Figure 1: Most likely regimes over time

The remaining sequence of most likely regimes seems readily interpretable. There is a "normal" state (State 2) within which the process spends the majority of the investigation period. There are occasional jumps to the higher frequency regime (State 3) and somewhat periodic jumps to the lower frequency regime (State 1). This latter result produces some interesting insights which we discuss in the next section.

4.6. Utilising the MMNPP model for model development

A secondary observation demonstrates the diagnostic property of the MMNPP model. Figure 1 suggests that the adjusted frequency process remains in State 2 with occasional jumps to a higher or lower state. A closer look leads to the identification of periodic transitions to State 1. This appears to be due to the effect of workers taking leave from the 27th of December until the new year. This effect was not originally captured in the known component γ but is fairly intuitive in hindsight. The reduction in claims on the 25th and 26th was already captured (due to these dates being public holidays), but to further include the following dates until New Year's Eve was not obvious, and could easily be challenged in absence of evidence such as provided by our methodology.

Modellers may encounter difficulties in properly identifying all relevant factors of interest for analysis in practice. In the example above, upon discovery of the "end-of-year" effect on claim frequencies, this covariate can be included within the γ component of the model. This iterative exercise produces model improvement as the more interpretable volume/exposure component is expanded. Further, the impact of any additionally identified factors are removed from the hidden Markov chain, allowing this component to more realistically represent data features that are truly unobservable to the modeller. Repeated application of this diagnostic feature of the MMNPP model is facilitated by the computational speed increases produced in the previous sections. The final result is ideally a model where all known features are accounted for through the frequency perturbation measure γ while the hidden Markov regimes capture the aggregated effects of factors that are unable to be explicitly modelled.

4.7. Model comparisons

In order to demonstrate the benefits provided by the MMNPP approach as opposed to more elementary ones, we now provide a comparison between the MMNPP model and four special (simpler) cases, namely: (1) a homogeneous Poisson process (which serves as the null model), (2) a non-homogeneous Poisson process (with the non-homogeneous component incorporating the effect of covariates through the ODP GLM), (3) a homogeneous Markov-modulated Poisson process with 3 states, and (4) another homogeneous MMPP model with 10 states.

Model (2) extends Model (1) by allowing a non-homogeneous (but deterministic) intensity, whereas Model (3) extends Model (1) by making the constant intensity regime-dependent. However, it can be seen below that the predictions produced by Model (3) are suboptimal and thus Model (4) is introduced to extend Model (3) through the use of a greater number of regimes. The MMNPP considered in this paper combines the extensions outlined above, and the comparisons below suggest that both are necessary for the data set under consideration.

Remark 4.1. *We note here that after identification of the outlying regime and end of year effects discussed in the previous sections, these covariates were included in the ODP GLM associated with the γ frequency perturbation component. The order selection methodology described in Section 4.4 was again applied, resulting in a 3 regime MMNPP being chosen as optimal. Analysis of model outputs shows that essentially, the outlying regime (regime 4) has been removed while the lower regime (regime 1) has been retained as there are still a sufficient number of events to warrant a separate regime. Thus, we emphasise that in these comparisons, the chosen ODP GLM incorporates the additional features described in Section 4.6. This is generous, because without use of the MMNPP framework, some of these data features may not have been identified.*

Remark 4.2. *The chosen number of states/regimes in the MMPP models is also based on the residual-testing methodology outlined previously. The order 3 MMPP model was chosen to match the MMNPP model. Model (4) attempted to utilise a similar residual-testing method to obtain the optimal number of states for an MMPP model. However, even upon reaching 10 states, the residuals still did not resemble white noise. A higher number of states would continue to improve in-sample fit but this comes at the cost of reduced model interpretability. Thus, our chosen number of states is 10 in this case.*

For each of the models, we produce daily in-sample predictions and then examine the residuals. The daily granularity was chosen to match the granularity of the original data. Figure 2 below plots the predictions in the left column and the residuals for each model in the right column. Note that for confidentiality reasons, the y-axis has been removed.

Table 2 provides summary statistics of the residuals for each model, as well as the p-value for the Ljung-Box test for auto-correlation with a lag of 120 days and the Wolf-Waldowitz runs test for randomness/persistence.

It is clear from the plots above that the homogeneous Poisson prediction (HPP) produces a poor fit. Further, all residual summary statistics are significantly greater than the ones for any other considered model.

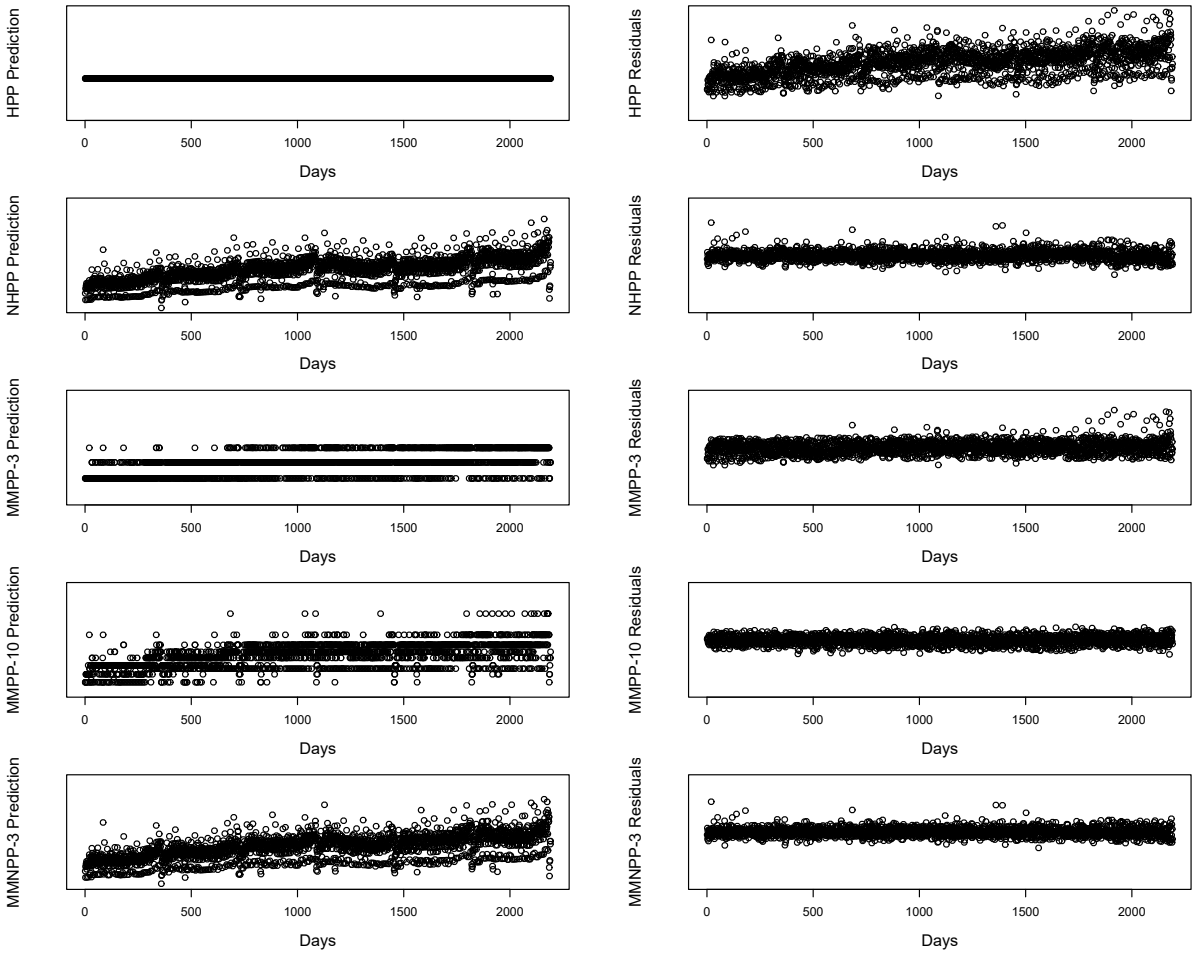

Figure 2: Predictions and residuals for each model

Model NHPP (2) uses the same over-dispersed Poisson GLM for the non-homogeneous intensity function as is used for the MMNPP-3 model. The aim of this approach is to incorporate domain knowledge within the model-building process. In addition, this method is standard in the industry both due to ease of implementation as well as model understandability and tractability. From Table 2, this approach provides a marked improvement upon Model (1). The residual plot also seem more reasonable in comparison. However, both statistical tests applied to Model NHPP (2) indicate that the residuals do not resemble white noise and thus, there remain some data features that have still not been adequately accounted for.

Model MMPP-3 (3) attempts to perform the same task as Model NHPP (2) through the use of the more-objective Markov-modulated Poisson process. However, it can be seen from the prediction plots of Models NHPP (2), MMPP-10 (4) and the MMNPP-3 that there is a increasing trend in the raw claims frequency over time. The prediction plot for Model MMPP-3 (3) illustrates that with only 3 regimes, the MMPP model provides a poor fit to the data. Evidence of inferior fit compared to other models under consideration is also seen in Table 2.

The 10-state MMPP approach (4) is a more data-driven approach compared to Model NHPP (2), and the high number of regimes means that the fit is able to better accommodate the trends/patterns within the data, including both the increasing frequency over time as well as the periodic jumps to lower frequency regimes. The residual plot depicts lower magnitudes as well when compared to Model MMPP-3 (3). However, Table 2 provides strong evidence that there still exists auto-correlation within the residuals, indicating the presence of data features not taken into account by the Markov modulation. Further, the high number of regimes,

	HPP	NHPP	MMPP-3	MMPP-10	MMNPP-3
Σ residuals	0.0	0.0	748.8	-1130	1418
Σ absolute residuals	130,644	45,211	59,501	44,487	40,203
Σ residuals squared	12,051,219	1,670,274	2,725,133	1,385,661	1,295,155
p-value for Ljung-Box test	0	1e-06	0	0	0.85
p-value for runs test	0	9e-08	1e-05	0.96	0.20

Table 2: Summary statistics for model comparison

while attempting to compensate for the increasing trend, unfortunately detract from understanding of the model. Ultimately, this is an issue with the structure of the MMPP model. A more natural approach may be to account for this trend elsewhere instead of using discretised regimes. This removes the dominating effect of the trend and ideally, reduces the number of regimes. It may be difficult to extract explainable information from a model with a very high number of regimes, as evidenced by the MMPP prediction plot above (which we note resembles the regime filtering plot for the most likely regime per day). Aside from the obvious upwards trend, it requires significant effort to develop additional insight into the causes of the hidden regimes.

Finally, the 3-state MMNPP residuals resemble white-noise. This model serves as a useful composite model that combines the 'non-parametric' fit of the MMPP with the model interpretability provided by the non-homogeneous Poisson model. The latter component allows for the tractable incorporation of domain knowledge whilst maintaining superior goodness-of-fit. The number of regimes in the MMNPP is only 3, which suitably caters for further diagnosis and model improvement. Also, by adding structural features through the ODP GLM, we have actually improved marginally on the fit of the MMPP, as can be seen in the residual statistics shown in Table 2. This has also removed the auto-correlation present in the NHPP and MMPP residuals. In summary, this case study demonstrates that the MMNPP produces a significant number of advantages through its extensions to the NHPP and MMPP models by allowing for expert experience and features extracted through data analysis to be integrated into the modelling process, while still producing excellent fits to data and accounting for characteristics such as auto-correlation.

Remark 4.3. *We remark that the MMNPP provides clear advantages for out-of-sample comparisons. The MMNPP model overcomes a key weakness of the original MMPP model in terms of out-of-sample predictions, in that MMPP predictions are (without adjustment) unable to maintain the time-structure of regimes that are extracted from the data set. In other words, the MMNPP allows for known effects which do not follow Markov arrivals to be modelled by the non-homogeneous component. Further, we have already commented that the structure of the MMPP seems unsuitable to capture the upwards trend displayed in the data and it is expected that out-of-sample comparisons will heavily favour the MMNPP model. For example, if this trend were continue in the future, the MMPP model would be unable to account for the higher frequency. Thus, we do not provide numerical results of the out-of-sample comparison here.*

5. Conclusion

In this paper, a Markov-modulated non-homogeneous Poisson process framework was introduced to model counting processes. We produced an extension of the standard MMPP models through a flexible frequency perturbation measure, which allows for the inclusion of both non-recurring and periodic data features to be readily incorporated into analysis. The proposed extension is beneficial as it not only allows known factors of interest to be included in a tractable manner but also provides some control over the underlying drivers of the observed events that are modelled through the hidden Markov chain. This produces a more comprehensive, accurate and realistic count model compared to the original MMPP approach. This method is also intuitive as the MMNPP is shown to be an operational time transform of the original MMPP. An efficient EM calibration procedure is derived, allowing the model to be applied in cases where there are large numbers of observations. In these situations, issues of computational cost and numerical stability are

of great concern. These barriers to practical implementation are addressed through various computational and theoretical innovations such as a scaling procedure for the backwards-forwards recursions utilised in the calibration algorithm and methods to improve calibration speeds. Our theoretical results and proposed implementation procedure are illustrated through a case study using real insurance data and insights that are provided through the MMNPP model are discussed. The procedure is shown to generate interpretable regimes in the empirical study and the MMNPP analysis framework is shown to generate additional results that can allow modellers to delve more deeply into the data if it is deemed necessary. Comparisons to alternative models demonstrate the advantages of the MMNPP model as an attractive composite model that provides both superior goodness-of-fit as well as significant model interpretability and tractability.

Acknowledgements

Earlier versions of this paper were presented at the 8th and 9th *Australasian Actuarial Education and Research Symposium* in Sydney (Australia), the *22nd International Congress on Insurance: Mathematics and Economics* in Sydney (Australia), the *Actuarial Risk Modelling and Extreme Values Workshop* in Canberra (Australia), the 4th *European Actuarial Journal Conference* in Leuven (Belgium), the ASTIN Colloquium in Cape Town (South Africa) as well as at seminars held at the universities of Lyon and Copenhagen in 2019. Some results in this paper were also presented at the Australian Actuaries Institute General Insurance Seminar in November 2018 (and awarded the Taylor Fry Silver Award); see Avanzi et al. (2018). The authors are grateful for constructive comments received from colleagues who attended these events.

This research was supported under the Australian Research Council's Linkage Projects funding scheme (project number LP130100723, with funding partners Allianz Australia Insurance Ltd, Insurance Australia Group Ltd, and Suncorp Metway Ltd). Furthermore, Alan Xian acknowledges financial support by the Australian Government Research Training program, as well the UNSW Business School through supplementary scholarships. The views expressed herein are those of the authors and are not necessarily those of the supporting organisations.

Data and Code

We are unable provide the dataset that was used in the empirical case study due to confidentiality. However, simulated data sets with similar features, as well as all relevant codes, can be found at `https://github.com/agi-lab/reserving-MMNPP`.

References

Albrecher, H., Araujo-Acuna, J., Beirlant, J., 2019. Fitting non-stationary Cox processes - an application to fire insurance data. Preprint, University of Lausanne.

Arts, J., 2017. A multi-item approach to repairable stocking and expediting in a fluctuating demand environment. European Journal of Operational Research 256 (1), 102–115.

Arts, J., Basten, R., Van Houtum, G.-J., 2016. Repairable stocking and expediting in a fluctuating demand environment: Optimal policy and heuristics. Operations Research 64 (6), 1285–1301.

Avanzi, B., Taylor, G., Wong, B., Xian, A., 2018. How to proxy the unmodellable: Analysing granular insurance claims in the presence of unobservable or complex drivers. 2018 General Insurance Seminar.

Avanzi, B., Wong, B., Yang, X., 2016. A micro-level claim count model with overdispersion and reporting delays. Insurance: Mathematics and Economics 71, 1–14.

Badescu, A. L., Chen, T., Lin, X. S., Tang, D., 2019. A markoed Cox model for the number of IBNR claims: Estimation and application. ASTIN Bulletin, 131.

Bartlett, M., 1967. Some remarks on the analysis of time-series. Biometrika 54 (1-2), 25–38.

Casale, G., Sansottera, A., Cremonesi, P., 2016. Compact Markov-modulated models for multiclass trace fitting. European Journal of Operational Research 255 (3), 822–833.

Ching, W. K., 1997. Markov-modulated Poisson processes for multi-location inventory problems. International Journal of Production Economics 53 (2), 217–223.

Crevecoeur, J., Antonio, K., Verbelen, R., 2019. Modeling the number of hidden events subject to observation delay. European Journal of Operational Research.

Greene, A. M., Robertson, A. W., Smyth, P., Triglia, S., 2011. Downscaling projections of Indian monsoon rainfall using a non-homogeneous hidden Markov model. Quarterly Journal of the Royal Meteorological Society 137 (655), 347–359.

Guillou, A., Loisel, S., Stupfler, G., 2013. Estimation of the parameters of a Markov-modulated loss process in insurance. Insurance: Mathematics and Economics 53 (2), 388–404.
Guillou, A., Loisel, S., Stupfler, G., 2015. Estimating the parameters of a seasonal Markov-modulated Poisson process. Statistical Methodology 26 (0), 103 – 123.
Gusella, R., 1991. Characterizing the variability of arrival processes with indexes of dispersion. Selected Areas in Communications, IEEE Journal on 9 (2), 203–211.
Hughes, J. P., Guttorp, P., Charles, S. P., 1999. A non-homogeneous hidden Markov model for precipitation occurrence. Journal of the Royal Statistical Society: Series C (Applied Statistics) 48 (1), 15–30.
Kou, S., Sunney Xie, X., Liu, J. S., 2005. Bayesian analysis of single-molecule experimental data. Journal of the Royal Statistical Society: Series C (Applied Statistics) 54 (3), 469–506.
Lagona, F., Maruotti, A., Picone, M., 2011. A non-homogeneous hidden Markov model for the analysis of multi-pollutant exceedances data. In: Hidden Markov Models, Theory and Applications. IntechOpen.
Landon, J., Özekici, S., Soyer, R., 2013. A Markov modulated Poisson model for software reliability. European Journal of Operational Research 229 (2), 404–410.
Langrock, R., Borchers, D. L., Skaug, H. J., 2013. Markov-modulated nonhomogeneous Poisson processes for modeling detections in surveys of marine mammal abundance. Journal of the American Statistical Association 108 (503), 840–851.
Leroux, B. G., Puterman, M. L., 1992. Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. Biometrics, 545–558.
Meier-Hellstern, K. S., 1984. A statistical procedure for fitting Markov-modulated Poisson processes. Technical report, Univ. of Delaware.
Moler, C., Van Loan, C., 2003. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. SIAM review 45 (1), 3–49.
Montoya, R., Netzer, O., Jedidi, K., 2010. Dynamic allocation of pharmaceutical detailing and sampling for long-term profitability. Marketing Science 29 (5), 909–924.
Nasr, W. W., Maddah, B., 2015. Continuous (s, s) policy with MMPP correlated demand. European Journal of Operational Research 246 (3), 874–885.
Norberg, R., 1993. Prediction of outstanding liabilities in non-life insurance. ASTIN Bulletin 23, 95–115.
Özekici, S., Parlar, M., 1999. Inventory models with unreliable suppliers in a random environment. Annals of Operations Research 91, 123–136.
Roberts, W. J., Ephraim, Y., Dieguez, E., 2006. On Rydén's EM algorithm for estimating MMPPs. Signal Processing Letters, IEEE 13 (6), 373–376.
Rossiter, M. H., 1989. Characterizing a random point process by a switched Poisson process. Phd Thesis, Monash University, Melbourne.
Rydén, T., 1994. Parameter estimation for Markov modulated Poisson processes. Communication in Statistics. Stochastic Models 10 (4), 795–829.
Rydén, T., 1996. An EM algorithm for estimation in Markov-modulated Poisson processes. Computional Statistics & Data Analysis 21 (4), 431–447.
Rydén, T., et al., 2008. EM versus Markov chain Monte Carlo for estimation of hidden Markov models: A computational perspective. Bayesian Analysis 3 (4), 659–688.
Salvador, P., Valadas, R., Pacheco, A., 2003. Multiscale fitting procedure using Markov modulated Poisson processes. Telecommunication systems 23 (1-2), 123–148.
Scott, S., Smyth, P., 2003. The Markov modulated Poisson process and Markov Poisson Cascade with applications to web traffic modelling. Bayesian Statistics 7.
Shi, S. W., Zhang, J., 2014. Usage experience with decision aids and evolution of online purchase behavior. Marketing Science 33 (6), 871–882.
Thayakaran, R., Ramesh, N., 2013a. Markov modulated Poisson process models incorporating covariates for rainfall intensity. Water Science and Technology 67 (8), 1786–1792.
Thayakaran, R., Ramesh, N. I., 2013b. Multivariate models for rainfall based on Markov modulated Poisson processes. Hydrology Research 44 (4), 631–643.
URL `http://hr.iwaponline.com/content/44/4/631`
Van Loan, C., 1978. Computing integrals involving the matrix exponential. IEEE transactions on automatic control 23 (3), 395–404.
Verrall, R., Wüthrich, M., 2016. Understanding reporting delay in general insurance. Risks 4 (3), 25.
Yoshihara, T., Kasahara, S., Takahashi, Y., 2001. Practical time-scale fitting of self-similar traffic with Markov-modulated Poisson process. Telecommunication Systems 17 (1-2), 185–211.
Zhang, J. Z., Netzer, O., Ansari, A., 2014. Dynamic targeted pricing in B2B relationships. Marketing Science 33 (3), 317–337.