

Filtering a Markov Modulated Random Measure

Robert J. Elliott, Tak Kuen Siu, and Hailiang Yang

Abstract—We develop a new exact filter when a hidden Markov chain influences both the sizes and times of a marked point process. An example would be an insurance claims process, where we assume that both the stochastic intensity of the claim arrivals and the distribution of the claim sizes depend on the states of an economy. We also develop the robust filter-based and smoother-based EM algorithms for the on-line recursive estimates of the unknown parameters in the Markov-modulated random measure. Our development is in the framework of modern theory of stochastic processes.

Index Terms—Insurance risk models, Markov-modulated random measures, martingales, model uncertainty, reference probability, robust EM algorithms.

I. INTRODUCTION

THE compound Poisson process is a standard, classical model in ruin theory which is used to describe an insurance claims process over a certain time horizon. The use of the compound Poisson process for modelling insurance claims has a natural interpretation. In particular, the number of claims over a time horizon can be modelled as a Poisson process while the amounts of individual claims can be modelled by a sequence of positive random variables, independent of the number of claims. One of the main reasons why the compound Poisson model is so popular is that analytically tractable results for ruin theory can be obtained. More specifically, some closed-form formulae for ruin probabilities or analytical formulae for the upper and/or lower bounds of ruin probabilities can be obtained. Besides the compound Poisson process, other point processes have been used to provide a more flexible and realistic way to describe insurance claims processes. For example, Møller [18] introduced a marked point process for an insurance claims process, in which the arrivals of claims follow an inhomogeneous Poisson process and claim amounts follow some time-dependent distributions.

Manuscript received November 30, 2007; revised August 07, 2008. First published November 24, 2009; current version published January 13, 2010. This paper was presented in part at the 35th Canadian Statistical Society Meeting, St John's, Newfoundland, Canada, 2007, the Daiwa Workshop, Tokyo, Japan, 2007, and at the American Control Conference, New York, NY. This work was supported by SSHRC, and by the Research Grants Council of the Hong Kong Special Administrative Region, China under Project HKU 754008H. Recommended by Associate Editor Z. Wang.

R. J. Elliott is with Haskayne School of Business, University of Calgary, Calgary, AB T2N 1N4, Canada. He is also with the School of Mathematical Sciences, University of Adelaide, Adelaide SA 5005, Australia (e-mail: relliott@ucalgary.ca).

T. K. Siu is with the Department of Actuarial Studies, Faculty of Business and Economics, Macquarie University, Sydney NSW 2109, Australia (e-mail: ken.siu@efs.mq.edu.au).

H. Yang is with Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong, China (e-mail: hlyang@hkusua.hku.hk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TAC.2009.2034227

For more details on the use of point processes for modeling insurance claims, see, for example, [22] and [2]. For an excellent account of point processes, interested readers may refer to the monograph [5].

In recent years there has been considerable interest in the applications of Markov-modulated point processes, such as Markov-modulated compound Poisson processes, in ruin theory. The history of Markov-modulated point processes can be tracked back to the original work of Neuts [19] and this term was formally coined in [20]. Reinhard [21] and Asmussen [1] explored the applications of Markov-modulated point processes to actuarial science, in particular, risk theory. In the context of risk theory, the states of the Markov chain underlying these models represent different economic environments. The switching of the states of the economy can be attributed to structural changes in (macro)-economic conditions, changes in political regimes and business cycles, etc. The states of the chain can also be interpreted as different states of risk characteristics of policyholders. In practice, insurance companies can observe individual claim sizes and claim numbers over a given time period. However, the underlying state of the economy, or the regime-switching process, is not directly observable. The unobservability of the state introduces model uncertainty, which is an important issue that needs to be addressed in modern actuarial modeling. Most of the literature about Markov-modulated point processes for risk theory considers the situation that the chain only modulates the intensity of claim arrivals. However, in practice, the stochastic behavior of claim sizes may also depend on the state of the economy. This situation is overlooked in the literature, but it is certainly of practical importance and relevance. Indeed, a risk model with both the stochastic intensity of claim arrivals and the distribution of claims sizes provides a more flexible and realistic way to describe the stochastic behavior of claims and risk processes than one only having a Markov-modulated intensity of claim arrivals. To build such a risk model, one needs to lift the Markov-modulated point processes to a more general process, namely, the Markov-modulated random measure. An early formalism of modulated random measures appeared in [8]. For a more recent account on modulated random measures, interested readers may refer to [9]. The classic monograph [15] is a key reference of random measures and provides an excellent and comprehensive account on the topic. Another issue that is overlooked in some of the literature about Markov-modulated point processes for risk theory is filtering and estimation of the hidden state of the underlying Markov chain. This issue is important for the practical implementation of risk models based on Markov-modulated point processes.

In this paper we develop filters and smoothers for the state of a hidden Markov chain, which influences both the jump times and jump sizes of a random measure. Robust filters and smoothers

for the Markov-modulated random measure are derived in the form of linear ordinary differential equations (O.D.E.s). An example would be an insurance claims process, where both the stochastic intensity of the claim arrivals and the distribution of the claim sizes depend on the hidden Markov chain whose states represent different states of an economy. This provides a convenient way to estimate the hidden state of the economy based on observed information about the number of claims and the claim sizes. The filters and smoothers of the hidden states provide an appropriate method to select, or estimate, an insurance risk model in the “mean-squared-error” sense. They also provide information about the underlying hidden economic states generating the claims data. We also discuss the estimation of the unknown parameters in the Markov-modulated random measure using the robust filter-based and smoother-based EM algorithms. The development here is in the framework of modern theory of stochastic processes and provides recursive estimates of the unknown parameters. A simulation example is presented to illustrate the implementation of the filters and the smoothers.

Bremaud [4] considered a filtering problem when a semi-martingale modulates a marked point process. However, the modified filtering and smoothing formulae for Markov chains derived in this paper appear to be new. The model considered here is also different from and certainly more general than those considered in Elliott and Malcolm [12], [13]. In the papers by Elliott and Malcolm [12], [13], the Markov chain only modulates the rate of the Poisson process. Our paper establishes new results for the case where the chain also influences the observed size of the jumps of a marked point process, or random measure. This result has not been obtained previously in the literature. Zeng [23] obtained some results for filters associated with an observed Poisson process. However, our results are more general than and different from those in [23]. In a recent paper by Elliott and Osakwe [14], a Markov-switching pure jump process was considered which is similar to the process considered here, but the chain was supposed fully observed. Bayraktar and Sezer [3] considered a Markov-modulated marked point process where the hidden Markov chain modulates the rate and the distribution of the jumps. However, the problem considered and approach used in our current paper are different from those in [3]. Their paper investigated the quickest detection problem of a Markov-modulated marked point process. The novelty of the current paper is that it is the first to develop the exact filter when the hidden Markov chain influences both the rate and size of the jumps of a random measure. We consider a situation where a Markov chain modulates a random measure, which is more general than the point processes considered in most of the existing literature. Our work indicates that similar techniques, such as Markovian regime-switching risk models, can be used in economic modeling and signal processing.

This paper is structured as follows. In Section II, we present the Markov-modulated random measure. Then we introduce a reference probability space by a measure change. Section III develops filters and smoothers for the hidden states of the economy. We then derive robust filters and smoothers for hidden states underlying the Markov-modulated random measure in the form of ordinary differential equations. Section IV presents robust filter-based and smoother-based EM algorithms.

A simulation example is given in Section V. The final section summarizes the paper.

II. MODEL DYNAMICS AND CHANGE OF MEASURES

In this section we shall present a Markov-modulated random measure for an insurance claims process. Møller [18] considered a marked point process for an insurance risk process, where the arrivals of claims follow an inhomogeneous Poisson process and claim amounts follow some time-dependent distributions. Here, we consider the situation when both the stochastic intensity of claim arrivals and the distribution of claim sizes switch over time according to a continuous-time, finite-state, hidden Markov chain. The hidden Markov chain is the state process in our model and its states represent different states of an economy. Both claim times and claim sizes are observable. We also suppose that the arrivals of claims and claim sizes are independent conditional on information about the hidden Markov chain. We shall first write the claims process in terms of a random measure with a Markov-switching compensator, as in [14]. To derive filters and smoothers for the hidden states, it is more convenient to work under a reference probability measure instead of the “real-world” probability measure under which the actual dynamics of the observable and unobservable processes are defined. We shall adopt the general Girsanov theorem for jump processes to change probability measures from a reference measure to a real-world one.

Consider a complete probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where \mathcal{P} is the “real-world” probability measure. We suppose that $(\Omega, \mathcal{F}, \mathcal{P})$ is rich enough to model the randomness of the observations process and the state process. We first describe the “real-world” dynamics under \mathcal{P} . However, \mathcal{P} will be constructed from a “reference probability” \mathcal{P}^\dagger under which the marked point process has simpler dynamics. We write \mathcal{T} for the time index set $[0, \infty)$ of the model. Let $X := \{X_t\}_{t \in \mathcal{T}}$ denote a continuous-time, finite-state, hidden Markov process, defined on $(\Omega, \mathcal{F}, \mathcal{P})$, with state space $\mathcal{S} := \{s_1, s_2, \dots, s_K\} \subseteq \mathbb{R}^K$. The states of the hidden Markov chain represent the hidden states of an economy. Without loss of generality, we can take the state space of X to be the set of unit vectors $\mathcal{L} := \{e_1, e_2, \dots, e_K\} \subseteq \mathbb{R}^K$, where $e_i := (0, \dots, 1, \dots, 0)^T \in \mathbb{R}^K$ with the “1” in the i^{th} position, and where y^T represents the transpose of a row vector y . This is called the canonical representation of the state space of the Markov chain X . Let A denote a constant rate matrix, or the Q -matrix, of the Markov chain. Then, using the canonical representation of the state space, the dynamics of the Markov chain X have the following semi-martingale representation, (see [11]):

$$X_t = X_0 + \int_0^t A X_s ds + M_t. \quad (1)$$

Here $\{M_t\}_{t \in \mathcal{T}}$ is an \mathbb{R}^K -valued martingale with respect to the natural filtration generated by $\{X_t\}_{t \in \mathcal{T}}$.

Consider a Poisson process $N := \{N_t\}_{t \in \mathcal{T}}$ on $(\Omega, \mathcal{F}, \mathcal{P})$, whose stochastic intensity is

$$\lambda_t := \langle \boldsymbol{\lambda}, X_t \rangle = \sum_{i=1}^K \langle \boldsymbol{\lambda}, e_i \rangle I_{\{X_t=e_i\}}. \quad (2)$$

Here $\boldsymbol{\lambda} := (\lambda_1, \lambda_2, \dots, \lambda_K) \in \mathbb{R}^K$ and $\lambda_k \geq 0$, for each $k = 1, 2, \dots, K$.

For each $t \in \mathcal{T}$, N_t will represent the number of claims over the time interval $[0, t]$.

Consider right-continuous, complete versions of the filtrations

$$\begin{aligned} \mathcal{F}^X &:= \{\mathcal{F}_t^X\}_{t \in \mathcal{T}}, \quad \mathcal{F}_t^X := \sigma\{X_u | u \in [0, t]\}, \\ \mathcal{F}^N &:= \{\mathcal{F}_t^N\}_{t \in \mathcal{T}}, \quad \mathcal{F}_t^N := \sigma\{N_u | u \in [0, t]\}, \\ \mathcal{G} &:= \{\mathcal{G}_t\}_{t \in \mathcal{T}}, \quad \mathcal{G}_t := \mathcal{F}_t^X \vee \mathcal{F}_t^N. \end{aligned}$$

Here \mathcal{F}^N is observable while \mathcal{F}^X is unobservable.

Then the Doob-Meyer decomposition for N is, (see, for example, [13])

$$N_t = \int_0^t \langle \boldsymbol{\lambda}, X_u \rangle du + V_t \quad (3)$$

where $V := \{V_t\}_{t \in \mathcal{T}}$ is a $(\mathcal{G}, \mathcal{P})$ -martingale.

Let $Z := \{Z_t\}_{t \in \mathcal{T}}$ denote a real-valued, Markov-modulated, random jump process under the real-world probability \mathcal{P} . If $\Delta Z_u := Z_u - Z_{u-}$, we can write

$$Z_t = \sum_{0 < u \leq t} \Delta Z_u, \quad Z_0 = 0, \quad \mathcal{P} - \text{a.s.} \quad (4)$$

Note that the state space of random times of claim arrivals \mathcal{T} is $[0, \infty)$. The state space of the jump sizes is $\mathcal{Z} := (-\infty, \infty)$. Our results can be modified so they are true for jumps with values in a general measurable space. For modeling insurance claims, we have in mind that the jumps are always positive. Suppose \mathcal{X} is the product space $\mathcal{T} \times \mathcal{Z}$ and $\gamma(\cdot, \cdot)$ is a random measure on \mathcal{X} , which selects the random times of claim arrivals and random claim sizes $y := Z_u - Z_{u-}$. More rigorously, the random measure should be written as a family $\{\gamma(\cdot, \cdot, \omega) | \omega \in \Omega\}$ of measures on \mathcal{X} . To simplify the notation, we suppress the notation ω and write $\gamma(\cdot, \cdot)$ for $\gamma(\cdot, \cdot, \omega)$. Note that $y := Z_u - Z_{u-} = 0$ if no claim occurs at time u ; otherwise, it is positive. Then the process Z can be written as

$$Z_t = \int_0^t \int_{-\infty}^{\infty} y \gamma(dy, du). \quad (5)$$

Here γ is a sum of random delta functions

$$\gamma(dy, dt) = \sum_{k \geq 0} \delta_{Y_{T_k}}(dy) \delta_{T_k}(dt)$$

where $\delta_x(\cdot)$ is a Dirac delta function, or a point mass, at the point x .

Here T_k is the random time of the k^{th} claim and Y_{T_k} is the random jump size of the process Z at the random time T_k .

Then, for suitable integrands $g : (\Omega \times (-\infty, \infty) \times [0, \infty)) \rightarrow \mathbb{R}$

$$\int_0^t \int_{-\infty}^{\infty} g(\omega, y, u) \gamma(dy, du) = \sum_{T_k \leq t} g(\omega, Y_{T_k}, T_k). \quad (6)$$

In terms of the random measure $\gamma(\cdot, \cdot)$

$$N_t = \int_0^t \int_{-\infty}^{\infty} \gamma(dy, du), \quad t \in \mathcal{T}. \quad (7)$$

The statistical properties of Z under \mathcal{P} are determined by its compensator. In fact we wish to consider a Markov-modulated compensator for Z under the real-world probability \mathcal{P} .

First, for each $k = 1, 2, \dots, K$, write $f_k(y)$ for a probability density function of the random claim size $Y := Z_u - Z_{u-}$ when $X_{u-} = k$. The random times of claim arrivals and the random claim sizes are assumed to be independent conditional on the hidden states X . Now, consider right-continuous, complete versions of the filtrations

$$\begin{aligned} \mathcal{F}^Z &:= \{\mathcal{F}_t^Z\}_{t \in \mathcal{T}}, \quad \mathcal{F}_t^Z := \sigma\{Z_u | u \in [0, t]\}, \\ \bar{\mathcal{G}} &:= \{\bar{\mathcal{G}}_t\}_{t \in \mathcal{T}}, \quad \bar{\mathcal{G}}_t := \mathcal{F}_t^Z \vee \mathcal{F}_t^X \end{aligned}$$

where \mathcal{F}^Z is observable.

Then, the Markov-switching compensator of the random measure $\gamma(dy, du)$ under \mathcal{P} is

$$\nu(dy, du | X_{u-}) := \sum_{k=1}^K \langle X_{u-}, e_k \rangle \lambda_k f_k(y) dy du. \quad (8)$$

Consequently, under \mathcal{P}

$$\begin{aligned} \hat{M}_t &:= Z_t - \int_0^t \int_{-\infty}^{\infty} y \nu(dy, du | X_{u-}) \\ &= \int_0^t \int_{-\infty}^{\infty} y (\gamma(dy, du) - \nu(dy, du | X_{u-})) \\ &= \int_0^t \int_{-\infty}^{\infty} y \gamma(dy, du) \\ &\quad - \sum_{k=1}^K \int_0^t \lambda_k \langle X_{u-}, e_k \rangle \left(\int_{-\infty}^{\infty} y f_k(y) dy \right) du \end{aligned} \quad (9)$$

is a $(\bar{\mathcal{G}}, \mathcal{P})$ -local martingale.

Note that $\hat{M} := \{\hat{M}_t\}_{t \in \mathcal{T}}$ is the compensated Markov-modulated random measure under \mathcal{P} .

To construct the process Z with compensator given by (8) we, in fact, “start again” with a “reference probability” \mathcal{P}^\dagger under which the observation process Z does not depend on the chain. The main purpose of introducing the reference probability measure \mathcal{P}^\dagger is to facilitate the derivation of the Zakai forms of the filtering and smoothing equations governing the evolution of the filtered and smoothed estimates of the hidden state of the chain over time. This technique based on a reference probability measure plays an important role in the filtering literature

and was discussed in [11]. Here suppose we have a reference probability measure \mathcal{P}^\dagger under which the process Z is a marked point process with unit intensity for the jump times and a density function for the jump sizes $f(y)$, with $f(y) > 0$, which is independent of the hidden state X . That is, we assume that the compensator ν^\dagger of γ under \mathcal{P}^\dagger is:

$$\nu^\dagger(dy, du) := f(y)dydu. \quad (10)$$

Then, the compensated process $M^\dagger := \{M_t^\dagger\}_{t \in \mathcal{T}}$ of Z under \mathcal{P}^\dagger is

$$\begin{aligned} M_t^\dagger &:= Z_t - \int_0^t \int_{-\infty}^{\infty} y \nu^\dagger(dy, du) \\ &= \int_0^t \int_{-\infty}^{\infty} y \gamma(dy, du) - \int_0^t \int_{-\infty}^{\infty} y f(y) dy du \end{aligned} \quad (11)$$

where M^\dagger is a local martingale under \mathcal{P}^\dagger .

Definition 2.1: For each $k = 1, 2, \dots, K$, let $h_k(y) = (\lambda_k f_k(y)/f(y))$, where $f(y) > 0$.

We suppose that there is a probability measure \mathcal{P}^\dagger under which (10) holds. Taking the measure \mathcal{P}^\dagger as a starting point, we define a $\bar{\mathcal{G}}$ -adapted process $\Lambda := \{\Lambda_{0,t}\}_{t \in \mathcal{T}}$ by

$$\begin{aligned} \Lambda_{0,t} &:= 1 + \int_0^t \Lambda_{0,u-} \left(\sum_{k=1}^K \int_{-\infty}^{\infty} \langle X_{u-}, e_k \rangle \right. \\ &\quad \times (h_k(y) - 1) (\gamma(dy, du) - f(y) dy du) \Big). \end{aligned} \quad (12)$$

Then Λ is a $(\bar{\mathcal{G}}, \mathcal{P}^\dagger)$ -local martingale. We assume that Λ is a $(\bar{\mathcal{G}}, \mathcal{P}^\dagger)$ -martingale.

By Itô's differentiation rule

$$\begin{aligned} \Lambda_{0,t} &:= \exp \left(- \int_0^t \sum_{k=1}^K \langle X_{u-}, e_k \rangle \int_{-\infty}^{\infty} (h_k(y) - 1) f(y) dy du \right. \\ &\quad \left. + \int_0^t \sum_{k=1}^K \langle X_{u-}, e_k \rangle \int_{-\infty}^{\infty} \log h_k(y) \gamma(dy, du) \right). \end{aligned} \quad (13)$$

The “real world” probability measure \mathcal{P} is then constructed using Λ by setting

$$\left. \frac{d\mathcal{P}}{d\mathcal{P}^\dagger} \right|_{\bar{\mathcal{G}}(t)} := \Lambda_{0,t}.$$

Note, however, by the general Girsanov theorem for jump processes, it is under the real-world probability measure \mathcal{P} that

$$\begin{aligned} M_t &= - \sum_{k=1}^K \left(\int_0^t \int_{-\infty}^{\infty} y \langle X_{u-}, e_k \rangle \lambda_k f_k(y) dy du \right) \\ &\quad + \int_0^t \int_{-\infty}^{\infty} y \gamma(dy, du), \quad t \in \mathcal{T} \end{aligned} \quad (14)$$

is a $(\bar{\mathcal{G}}, \mathcal{P})$ -local martingale. So, it is under \mathcal{P} that the observation process Z is influenced by the chain.

Suppose $\gamma := \{\gamma_t\}_{t \in \mathcal{T}}$ is any $\bar{\mathcal{G}}$ -adapted, integrable process. Given \mathcal{F}_t^Z , we can estimate γ_t by its least-square estimate $E[\gamma_t | \mathcal{F}_t^Z]$. By a form of the Bayes' rule

$$E[\gamma_t | \mathcal{F}_t^Z] = \frac{E^\dagger[\Lambda_{0,t} \gamma_t | \mathcal{F}_t^Z]}{E^\dagger[\Lambda_{0,t} | \mathcal{F}_t^Z]} \quad (15)$$

where E^\dagger denotes expectation with respect to \mathcal{P}^\dagger .

Write, for each $t \in \mathcal{T}$, $\sigma_t(\gamma) := E^\dagger[\Lambda_{0,t} \gamma_t | \mathcal{F}_t^Z]$. Then (15) can be expressed as $\sigma_t(\gamma)/\sigma_t(1)$.

III. FILTERS AND SMOOTHERS

In this section, we shall derive filters and smoothers for the hidden Markov chain X . Theorems 3.2, 3.5 and 3.8 are the main results in this section. First, we shall derive the recursive Zakai equation for the filter $E(X_t | \mathcal{F}_t^Z)$. In particular, the Zakai stochastic differential equation governing the evolution of the unnormalized filter over time is given in Theorem 3.2. Then we shall derive robust filters for the hidden states using the gauge transformation technique introduced by Clark [6]. In particular, a linear ordinary differential equation (O.D.E) for the transformed process of the filtered estimates will be derived. The O.D.E. governing the robust filter is then summarized in Theorem 3.5. This provides a convenient way to evaluate the filtered estimates of the state process. By exploiting a duality as in [13], we also derive robust smoothers for the hidden process. In particular, a linear forward O.D.E. for the smoothed estimates and a linear backward O.D.E. for the process of extra information are derived. These results are presented in Theorem 3.8.

For each time $u \in (0, \infty)$, suppose $Y_u : \Omega \rightarrow (-\infty, \infty)$ is a random variable with a strictly positive density function f under \mathcal{P}^\dagger . Write

$$H_k(u, \omega) := \frac{\lambda_k f_k(Y_u(\omega))}{f(Y_u(\omega))} = h_k(Y_u(\omega)) \quad (16)$$

where $h_k(y)$ is given by Definition 2.1. To simplify the notation, we suppress the notation ω in the sequel unless otherwise stated.

Then

$$\begin{aligned} E^\dagger(H_k) &= \int_{-\infty}^{\infty} h_k(y) f(y) dy \\ &= \int_{-\infty}^{\infty} \frac{\lambda_k f_k(y)}{f(y)} f(y) dy = \lambda_k. \end{aligned} \quad (17)$$

We shall consider the diagonal matrices

$$\begin{aligned} \text{diag}(H(u)-1) &:= \text{diag}((H_1(u)-1), (H_2(u)-1), \dots, \\ &\quad (H_K(u)-1)), \\ \text{diag}(\lambda-1) &:= \text{diag}((\lambda_1-1), (\lambda_2-1), \dots, (\lambda_K-1)). \end{aligned}$$

Definition 3.1: For each $t \in \mathcal{T}$, we define $q_t = \sigma(X_t) = E^\dagger(\Lambda_{0,t} X_t | \mathcal{F}_t^Z)$, which is the unnormalized filtered estimate of X_t given \mathcal{F}_t^Z .

We give the recursive Zakai equation governing q_t in the following theorem:

Theorem 3.2:

$$q_t = q_0 + \int_0^t A q_u du + \int_0^t \text{diag}(H(u) - 1) q_u dN_u - \int_0^t \text{diag}(\lambda - 1) q_u du \quad (18)$$

where $q_0 := E[X_0]$ is the initial distribution of the chain X , which is assumed to be known.

Let $\mathbf{1} := (1, 1, \dots, 1)^T \in \mathbb{R}^K$; then, $\langle X_t, \mathbf{1} \rangle = 1$. Hence

$$\begin{aligned} \langle E^\dagger[\Lambda_{0,t} X_t | \mathcal{F}_t^Z], \mathbf{1} \rangle &= E^\dagger[\Lambda_{0,t} \langle X_t, \mathbf{1} \rangle | \mathcal{F}_t^Z] \\ &= E^\dagger[\Lambda_{0,t} | \mathcal{F}_t^Z]. \end{aligned} \quad (19)$$

Write $p_t := E[X_t | \mathcal{F}_t^Z]$. Then, by (15)

$$p_t = \frac{q_t}{\langle q_t, \mathbf{1} \rangle}. \quad (20)$$

Proof: See Appendix.

In the sequel, we shall derive a robust filter for the hidden Markov chain. First, we need to introduce a rotation, or transformation, matrix Γ_t for the gauge transformation of q_t , for each $t \in \mathcal{T}$. For each $k = 1, 2, \dots, N$, we consider a scalar-valued process $\gamma^k := \{\gamma_t^k\}_{t \in \mathcal{T}}$ defined by

$$\begin{aligned} \gamma_t^k &:= \exp \left[(1 - \lambda_k)t + \int_0^t \log H_k(u) dN_u \right] \\ &= \exp(L_t^k) \end{aligned} \quad (21)$$

where L_t^k is defined by

$$\begin{aligned} L_t^k &:= (1 - \lambda_k)t + \int_0^t \log H_k(u) dN_u \\ &= (1 - \lambda_k)t + \sum_{T_j \leq t} \log H_k(T_j) \\ &= (1 - \lambda_k)t + \log \left(\prod_{T_j \leq t} H_k(T_j) \right) \end{aligned} \quad (22)$$

where T_j is the random time of the j^{th} claim defined as above.

The “ dN ” integral in (22) is now given as a sum of the functions of the random times of claims. Indeed, the “ dN ” integral in the stochastic differential equation governing q in (18) is also really a sum. However, it involves the implicitly defined quantity q , which makes it difficult to implement (18) in practice. Equation (22) expresses the “ dN ” integral as an explicit sum. This eliminates the problem of evaluating the “ dN ” integral in (18) which implicitly involves the solution process q .

Then, we define the transformation matrix Γ_t as the diagonal matrix $\text{diag}(\gamma_t^1, \gamma_t^2, \dots, \gamma_t^K)$. Write, for each $t \in \mathcal{T}$, Γ_t^{-1} for the inverse of Γ_t . Then, the following lemma gives a stochastic differential equation governing the evolution of Γ^{-1} over time.

Lemma 3.3: For each $u \in \mathcal{T}$, define the diagonal matrix

$$\text{diag} \left(\frac{1 - H(u)}{H(u)} \right) := \text{diag} \left(\left(\frac{1 - H_1(u)}{H_1(u)} \right), \dots, \left(\frac{1 - H_K(u)}{H_K(u)} \right) \right).$$

Then

$$d\Gamma_t^{-1} = \text{diag}(\lambda - 1) \Gamma_t^{-1} dt + \text{diag} \left(\frac{1 - H(u)}{H(u)} \right) \Gamma_t^{-1} dN_t \quad (23)$$

where $\Gamma_0 = \Gamma_0^{-1} = I$ and I is the $(K \times K)$ -identity matrix.

Proof: See Appendix.

Definition 3.4: Define the transformed process $\bar{q} := \{\bar{q}_t\}_{t \in \mathcal{T}}$ by setting

$$\bar{q}_t := \Gamma_t^{-1} q_t, \quad t \in \mathcal{T}.$$

Then the following theorem gives the linear ordinary differential equation governing the evolution of the transformed process over time.

Theorem 3.5: \bar{q} satisfies the linear ordinary differential equation

$$\bar{q}_t = q_0 + \int_0^t \Gamma_u^{-1} A \Gamma_u \bar{q}_u du. \quad (24)$$

Proof: See Appendix.

Let $\pi(X_t)$ denote a version of the expectation $E[X_t | \mathcal{F}_t^Z]$. Since $q_t = \Gamma_t \bar{q}_t$

$$\pi(X_t) = \frac{\Gamma_t \bar{q}_t}{\langle \Gamma_t \bar{q}_t, \mathbf{1} \rangle}. \quad (25)$$

It has been shown in [17] that $\pi(X_t)$ is a version of the expectation $E[X_t | \mathcal{F}_t^Z]$, which is continuous in the observation process Z in the Skorokhod topology. The robust filter estimate of X_t given \mathcal{F}_t^Z is then given by $\pi(X_t)$.

In the sequel, we shall derive smoothed state estimates for X by exploiting a duality. That is, we calculate the conditional expectation $E[X_t | \mathcal{F}_T^Z]$, for $t \in [0, T]$. First, we define some useful notation.

Definition 3.6: Let $r_t := E^\dagger[\Lambda_{0,T} X_t | \mathcal{F}_T^Z]$, which represents the unnormalized smoothed estimate of X_t .

Then, we define the process of extra information and its transformed process as follows.

Definition 3.7: Consider the process ν^k ($k = 1, 2, \dots, K$)

$$\nu_t^k := E^\dagger[\Lambda_{t,T} | \mathcal{F}_T^Z, X_t = e_k]. \quad (26)$$

write ν_t for $(\nu_t^1, \nu_t^2, \dots, \nu_t^K)^T \in \mathbb{R}^K$ with $\nu_T = \mathbf{1}$, where ν_t , $t \in \mathcal{T}$, represents the process of extra information. Its transformed process is defined by

$$\bar{\nu}_t := \Gamma_t \nu_t, \quad t \in \mathcal{T}.$$

Let $\bar{p}_t := E[X_t | \mathcal{F}_T^Z]$, where \bar{p}_t is the smoothed estimate of X_t , for each $t \in \mathcal{T}$. Then, the smoothed estimate \bar{p}_t can be evaluated based on the formulae presented in the following theorem.

Theorem 3.8: For $t \in [0, T]$,

$$\begin{aligned}\bar{p}_t &= \frac{r_t}{\langle r_t, \mathbf{1} \rangle} \\ &= \frac{1}{\langle \bar{q}_t, \bar{\nu}_t \rangle} \sum_{k=1}^K \langle \bar{q}_t, e_k \rangle \langle \bar{\nu}_t, e_k \rangle e_k \\ &= \frac{1}{\langle q_t, \nu_t \rangle} \sum_{k=1}^K \langle q_t, e_k \rangle \langle \nu_t, e_k \rangle e_k\end{aligned}\quad (27)$$

where \bar{q}_t satisfies the following forward linear ordinary differential equation:

$$\frac{d\bar{q}_t}{dt} = \Gamma_t^{-1} A \Gamma_t \bar{q}_t, \quad \bar{q}_0 = q_0 \quad (28)$$

and $\bar{\nu}_t$ satisfies the following backward linear ordinary differential equation:

$$\frac{d\bar{\nu}_t}{dt} = -\Gamma_t A^T \Gamma_t^{-1} \bar{\nu}_t, \quad \bar{\nu}_T = \Gamma_T \mathbf{1}. \quad (29)$$

Proof: See Appendix.

IV. PARAMETER ESTIMATION BY THE EM ALGORITHM

In this section, we shall consider the estimation of the unknown parameters for the Markov-modulated random measure with Markov-switching stochastic intensity for claim arrivals and distribution for claim sizes using the EM algorithm. For the estimation of the parameters in a Markov-modulated Poisson process with Markov-switching stochastic intensity only, interested readers may refer to [13]. Here we shall compute the estimates of the parameters in the rate matrix $A := [a_{ij}]_{i,j=1,2,\dots,K}$ and the vector of intensity parameters $\lambda := (\lambda_1, \lambda_2, \dots, \lambda_K)$. We suppose that for each regime $k = 1, 2, \dots, K$, the distribution of claim sizes $F_k(y)$ is known. The estimation of the distribution of claim sizes is also an interesting and practically relevant research topic in actuarial science, (see, for example, [16]). Here, our main focus is to filter the hidden state of the chain and to estimate the rate matrix of the chain and the intensity parameter of the jumps. We leave the estimation of the distribution of claim sizes for future research.

We shall consider both a robust filter-based EM algorithm and a robust smoother-based EM algorithm for the estimation of the model parameters. We also provide practical forms of the robust dynamics involved in the estimation scheme by computing time domain discretization of these robust dynamics on a regular partition. Then the robust estimates of the unknown parameters can be evaluated recursively.

The main results obtained in this section are presented in Theorems 4.1, 4.3, 4.4, 4.6 and Corollary 4.5. Theorem 4.1 gives the filter-based estimates of a_{ij} and λ_i and the stochastic differential equations governing the measure-valued quantities to compute the estimates. Theorem 4.3 gives the dynamics governing the gauge transformed processes corresponding to the measure-valued quantities, which are then used to derive the robust filter-based EM algorithm. To implement the robust filter-based EM algorithm, we need to discretize the processes governing the evolution of the measure-valued quantities. The discretizations of these processes are presented in Theorem 4.4.

Corollary 4.5 gives the solutions to the discretized processes in Theorem 4.4. Then we consider the smoother-based estimates of the unknown parameters a_{ij} and λ_i by exploiting a duality and give these estimates in Theorem 4.6.

A. Robust Filter-Based EM Algorithm

First, we notice that since $F_k(y)$ is supposed given, the observation processes N and Z provide the same amount of information to estimate a_{ij} and λ_i . The computation of the estimates of the unknown parameters involves the filtered estimates of the state X , two quantities related to the state and one quantity involving the state X and the observation N . We shall list these quantities as follows:

- 1) X_t , the state of the Markov chain X . The unnormalized filter estimate q_t in Definition 2.1 is given by Theorem 3.2.
- 2) O_t^i , the occupation time of the process X in state e_i up to time t . That is

$$O_t^i = \int_0^t \langle X_u, e_i \rangle du \in \mathbb{R}. \quad (30)$$

- 3) N_t^{ij} , the number of transitions from e_i to e_j of X , where $i \neq j$, up to time t

$$N_t^{ij} = \int_0^t \langle X_u, e_i \rangle \langle dX_u, e_j \rangle \in \mathbb{R}. \quad (31)$$

- 4) G_t^i , the level integral for the state e_i

$$G_t^i = \int_0^t \langle X_u, e_i \rangle dN_u \in \mathbb{R}. \quad (32)$$

Our goal is to estimate the parameters $\lambda := (\lambda_1, \lambda_2, \dots, \lambda_K)$ and $A := [a_{ij}]$ by the EM algorithm. The estimates \hat{a}_{ij} and $\hat{\lambda}_i$ of a_{ij} and λ_i , respectively, based on the EM algorithm were given in [7] and are presented as follows:

$$\begin{aligned}\hat{a}_{ij} &= \frac{E[N_T^{ij} | \mathcal{F}_T^Z]}{E[O_T^{ij} | \mathcal{F}_T^Z]} = \frac{\sigma(N_T^{ij})}{\sigma(O_T^{ij})}, \quad \forall (i, j), i \neq j, \\ \hat{\lambda}_i &= \frac{E[G_T^i | \mathcal{F}_T^Z]}{E[O_T^i | \mathcal{F}_T^Z]} = \frac{\sigma(G_T^i)}{\sigma(O_T^i)}.\end{aligned}\quad (33)$$

Note that the estimate \hat{a}_{ii} of the diagonal element a_{ii} ($i = 1, 2, \dots, K$) of the rate matrix A can be computed by noticing that $\sum_{j=1}^K \hat{a}_{ji} = 0, \forall i = 1, 2, \dots, K$.

The above estimates involve evaluating $\sigma(N_T^{ij})$, $\sigma(O_T^{ij})$ and $\sigma(G_T^i)$. In general, it is not possible to compute closed form dynamics for the processes $\sigma(N^{ij})$, $\sigma(O^i)$ and $\sigma(G^i)$ directly. However, it is possible to compute the associated measure-valued quantities $\sigma(N_t^{ij} X_t)$, $\sigma(O_t^i X_t)$ and $\sigma(G_t^i X_t)$, which are vectors in \mathbb{R}^K . Here we adopt these measure-valued quantities to compute the estimates \hat{a}_{ij} and $\hat{\lambda}_i$. The results are summarized in the following theorem.

Theorem 4.1: The estimates \hat{a}_{ij} and $\hat{\lambda}_i$ are given by

$$\hat{a}_{ij} = \frac{\langle \sigma(N_T^{ij} X_T), \mathbf{1} \rangle}{\langle \sigma(O_T^i X_T), \mathbf{1} \rangle} \quad (34)$$

and

$$\hat{\lambda}_i = \frac{\langle \sigma(G_T^i X_T), \mathbf{1} \rangle}{\langle \sigma(O_T^i X_T), \mathbf{1} \rangle} \quad (35)$$

where

$$\begin{aligned} \sigma(G_t^i X_t) &= \int_0^t A \sigma(G_u^i X_u) du \\ &+ \int_0^t \text{diag}(H(u) - 1) \sigma(G_{u-}^i X_{u-}) dN_u \\ &- \int_0^t \text{diag}(\lambda - 1) \sigma(G_u^i X_u) du \\ &+ \int_0^t \langle H(u), e_i \rangle \langle q_{u-}, e_i \rangle dN_u e_i \\ &- \int_0^t \langle \lambda - 1, e_i \rangle \langle q_u, e_i \rangle du e_i \\ \sigma(N_t^{ij} X_t) &= \int_0^t A \sigma(N_u^{ij} X_u) du \\ &+ \int_0^t \langle q_{u-}, e_i \rangle \langle A e_i, e_j \rangle du e_j \\ &+ \int_0^t \text{diag}(H(u) - 1) \sigma(N_{u-}^{ij} X_{u-}) dN_u \\ &- \int_0^t \text{diag}(\lambda - 1) \sigma(N_u^{ij} X_u) du \end{aligned} \quad (36)$$

and

$$\begin{aligned} \sigma(O_t^i X_t) &= \int_0^t A \sigma(O_u^i X_u) du + \int_0^t \langle q_u, e_i \rangle du e_i \\ &+ \int_0^t \text{diag}(H(u) - 1) \sigma(O_{u-}^i X_{u-}) dN_u \\ &- \int_0^t \text{diag}(\lambda - 1) \sigma(O_u^i X_u) du \end{aligned} \quad (37)$$

with $\sigma(G_0^i X_0) = \sigma(N_0^{ij} X_0) = \sigma(O_0^i X_0) = 0$, \mathcal{P} -a.s.

Proof: See Appendix.

To compute the above estimates, we can implement a filter bank consisting of recursive filters given by (18) and (36)–(38). The seminal work of Clark [6] utilized a version of a gauge transformation to simplify these recursive filters and developed

observation-parameterized forms of these filters. Here we shall adopt the gauge transformation of Clark [6] to compute new versions of the filters, which are robust.

First, the gauge transformed process \bar{q} in Definition 3.4 satisfies the linear ordinary differential equation in Theorem 3.5. We define the gauge transformed processes of the measure-valued quantities as follows.

Definition 4.2: Define the gauge transformed processes $\bar{\sigma}(G_t^i X_t) := \Gamma_t^{-1} \sigma(G_t^i X_t)$, $\bar{\sigma}(N_t^{ij} X_t) := \Gamma_t^{-1} \sigma(N_t^{ij} X_t)$ and $\bar{\sigma}(O_t^i X_t) := \Gamma_t^{-1} \sigma(O_t^i X_t)$.

Then the following theorem gives the dynamics of these gauge transformed processes.

Theorem 4.3: The gauge transformed processes in Definition 4.2 are governed by

$$\begin{aligned} \bar{\sigma}(G_t^i X_t) &= \int_0^t \Gamma_u^{-1} A \Gamma_u \bar{\sigma}(G_u^i X_u) du \\ &+ \text{diag}\left(\frac{1}{H(u)}\right) \langle H(u), e_i \rangle \langle \bar{q}_t, e_i \rangle N_u e_i \\ &- \int_0^t \langle \lambda - 1, e_i \rangle \langle \bar{q}_u, e_i \rangle du e_i \\ &- \int_0^t N_u \left\langle d \left[\text{diag}\left(\frac{1}{H(u)}\right) \langle H(u), e_i \rangle \bar{q}_u \right], e_i \right\rangle e_i, \end{aligned} \quad (39)$$

$$\begin{aligned} \bar{\sigma}(N_t^{ij} X_t) &= \int_0^t \Gamma_u^{-1} A \Gamma_u \bar{\sigma}(N_u^{ij} X_u) du \\ &+ \int_0^t \langle \bar{q}_u, e_i \rangle \langle A e_i, e_j \rangle du e_j \end{aligned} \quad (40)$$

and

$$\bar{\sigma}(O_t^i X_t) = \int_0^t \Gamma_u^{-1} A \Gamma_u \bar{\sigma}(O_u^i X_u) du + \int_0^t \langle \bar{q}_u, e_i \rangle du e_i. \quad (41)$$

Proof: See Appendix.

To compute the estimates \hat{a}_{ij} and $\hat{\lambda}_i$, we need to compute the filtered quantities $\sigma(N_T^{ij} X_T)$, $\sigma(O_T^i X_T)$ and $\sigma(G_T^i X_T)$, which also depend on q_t defined in Definition 3.1. In practice, we need to consider a version of the discrete-time approximation to these filtered quantities to implement these filters. Here we consider a regular partition on the time interval $[0, T]$ with equal length of each sub-interval $\Delta := T/M$, where M is a positive integer. We write $t_m := m\Delta$, where $m = 1, 2, \dots, M$.

Let

$$\Phi_{m,m-1} := \Gamma_{t_m} \Gamma_{t_{m-1}}^{-1}. \quad (42)$$

Then

$$q_{t_m} = \Gamma_{t_m} \bar{q}_{t_m} \approx \Gamma_{t_m} \Gamma_{t_{m-1}}^{-1} [\mathbf{I} + \Delta A] q_{t_{m-1}}. \quad (43)$$

Hence

$$q_{t_m} \approx \Phi_{m,m-1}[\mathbf{I} + \Delta A]q_{t_{m-1}}, \quad q_0 = E[X_0]. \quad (44)$$

By first calculating the time domain discretizations of $\bar{\sigma}(N_t^{ij} X_t)$, $\bar{\sigma}(O_t^i X_t)$ and $\bar{\sigma}(G_t^i X_t)$ in Definition 4.2 and noting that $\sigma(H_t) = \Gamma_t \bar{\sigma}(H_t)$, the time domain discretizations of $\sigma(N_t^{ij} X_t)$, $\sigma(O_t^i X_t)$ and $\sigma(G_t^i X_t)$ are obtained. The following theorem presents the time domain discretizations of $\sigma(N_t^{ij} X_t)$, $\sigma(O_t^i X_t)$ and $\sigma(G_t^i X_t)$.

Theorem 4.4: Let $Q_i(t) := \text{diag}(1/H(t))\langle H(t), e_i \rangle$, for each $t \in \mathcal{T}$ and $i = 1, 2, \dots, K$. Then

$$\begin{aligned} & \sigma(G_{t_m}^i X_{t_m}) \\ & \approx \Phi_{m,m-1}[\mathbf{I} + \Delta A]\sigma(G_{t_{m-1}}^i X_{t_{m-1}}) + \Gamma_{t_m} \\ & \quad \times [Q_i(t_m) \langle \bar{q}_{t_m}, e_i \rangle N_{t_m} e_i - Q_i(t_{m-1}) \langle \bar{q}_{t_{m-1}}, e_i \rangle \\ & \quad \times N_{t_{m-1}} e_i - \langle \lambda - 1, e_i \rangle \langle \bar{q}_{t_{m-1}}, e_i \rangle \Delta e_i \\ & \quad - N_{t_{m-1}} \langle Q_i(t_m) \bar{q}_{t_{m-1}}, e_i \rangle e_i - N_{t_{m-1}} \\ & \quad \times \langle Q_i(t_{m-1}) \Gamma_{t_{m-1}}^{-1} (\Delta A - \mathbf{I}) q_{t_{m-1}}, e_i \rangle e_i] \end{aligned} \quad (45)$$

$$\begin{aligned} & \sigma(N_{t_m}^{ij} X_{t_m}) \\ & \approx \Phi_{m,m-1}[\mathbf{I} + \Delta A]\sigma(N_{t_{m-1}}^{ij} X_{t_{m-1}}) \\ & \quad + \Phi_{m,m-1} \langle q_{t_{m-1}}, e_i \rangle \langle A e_i, e_j \rangle \Delta e_j \end{aligned} \quad (46)$$

and

$$\begin{aligned} \sigma(O_{t_m}^i X_{t_m}) & \approx \Phi_{m,m-1}[\mathbf{I} + \Delta A]\sigma(O_{t_{m-1}}^i X_{t_{m-1}}) \\ & \quad + \Phi_{m,m-1} \langle q_{t_{m-1}}, e_i \rangle \Delta e_i. \end{aligned} \quad (47)$$

Proof: See Appendix.

The following corollary gives explicit forms for \bar{q}_m , $\sigma(G_{t_m}^i X_{t_m})$, $\sigma(N_{t_m}^{ij} X_{t_m})$ and $\sigma(O_{t_m}^i X_{t_m})$.

Corollary 4.5: For each $k = 1, 2, \dots, K$ and each $m, l = 1, 2, \dots, M$ with $m > l$, let

$$L_{t_m, t_l}^k := (1 - \lambda_k)(t_m - t_l) + \log \left(\prod_{t_l < T_y \leq t_m} H_k(T_y) \right)$$

so that $L_{t_m, t_0}^k = L_{t_m}^k$.

Write, for each $m, l = 1, 2, \dots, M$ with $m > l$

$$U_{m,l} := (\exp(L_{t_m, t_l}^1), \exp(L_{t_m, t_l}^2), \dots, \exp(L_{t_m, t_l}^K)) \in \mathbb{R}^K.$$

Then for each $m = 1, 2, \dots, M$

$$q_{t_m} \approx q_0 \prod_{j=1}^m [\text{diag}(U_{j,j-1})(I + \Delta A)] \quad (48)$$

$$\begin{aligned} & \sigma(G_{t_m}^i X_{t_m}) \\ & \approx \sum_{l=1}^m \{ \text{diag}(U_{l,0}) \\ & \quad \times [Q_i(t_l) \langle \bar{q}_{t_l}, e_i \rangle N_{t_l} e_i - Q_i(t_{l-1}) \langle \bar{q}_{t_{l-1}}, e_i \rangle \\ & \quad \times N_{t_{l-1}} e_i - \langle \lambda - 1, e_i \rangle \langle \bar{q}_{t_{l-1}}, e_i \rangle \Delta e_i \\ & \quad - N_{t_{l-1}} \langle Q_i(t_l) \bar{q}_{t_{l-1}}, e_i \rangle e_i] \end{aligned}$$

$$\begin{aligned} & - N_{t_{l-1}} \left\langle Q_i(t_{l-1}) [\text{diag}(U_{l-1,0})]^{-1} \right. \\ & \quad \times (\Delta A - \mathbf{I}) q_{t_{l-1}}, e_i \rangle e_i \left. \right\} \\ & \times \left[\prod_{n=l+1}^m \text{diag}(U_{n,n-1})(I + \Delta A) \right] \end{aligned} \quad (49)$$

$$\begin{aligned} & \sigma(N_{t_m}^{ij} X_{t_m}) \\ & \approx \sum_{l=1}^m \{ \text{diag}(U_{l,l-1}) \langle q_{t_{l-1}}, e_i \rangle \langle A e_i, e_j \rangle \Delta e_j \\ & \quad \times \left[\prod_{n=l+1}^m \text{diag}(U_{n,n-1})(I + \Delta A) \right] \} \end{aligned} \quad (50)$$

and

$$\begin{aligned} \sigma(O_{t_m}^i X_{t_m}) & \approx \sum_{l=1}^m \{ \text{diag}(U_{l,l-1}) \langle q_{t_{l-1}}, e_i \rangle \Delta e_j \\ & \quad \times \left[\prod_{n=l+1}^m \text{diag}(U_{n,n-1})(I + \Delta A) \right] \}. \end{aligned} \quad (51)$$

Proof: See Appendix.

Then the estimates \hat{a}_{ij} and $\hat{\lambda}_i$ can be computed by following the three steps of the filter-based EM algorithm described as below:

Step I: Select the initial values $\hat{a}_{ij}(0)$ and $\hat{\lambda}_i(0)$.

Step II: Compute the MLEs, $\hat{a}_{ij}(k+1)$ and $\hat{\lambda}_i(k+1)$ by (34), (35) and (48)–(51), respectively.

Step III: Stop when $|\hat{a}_{ij}(k+1) - \hat{a}_{ij}(k)| < \epsilon_1$ and $|\hat{\lambda}_i(k+1) - \hat{\lambda}_i(k)| < \epsilon_2$; otherwise, continue from Step II, where ϵ_1, ϵ_2 are the desirable levels of accuracy and $\epsilon_1, \epsilon_2 > 0$.

B. Robust Smoother-Based EM Algorithm

In this section, we shall derive the estimates of the unknown parameters based on a robust smoother-based EM algorithm. That is, the estimators are computed based on smoothing schemes rather than filtering schemes. The smoothing schemes are particularly useful when the expectation step is completed with smoothed estimates, rather than filtered estimates, in some implementations of the EM algorithm. One major difficulty of computing the smoothing schemes is that the backward dynamics are difficult to develop. The development of these backward dynamics involves the construction of stochastic integrals evolving backward in time. To avoid this difficulty, Elliott and Malcolm [12] used a duality between forward and backward robust dynamics to develop smoothing algorithms for Markov-modulated Poisson processes. The smoothers in [12] do not involve backward stochastic integration. In the sequel, we shall compute the estimates from the robust smoother-based EM algorithm for the Markov-modulated random measure using the approach in [12].

We need to derive the dynamics for the dual process $\bar{\nu}$ for ν in Definition 3.7. In particular, we need to find a process $\bar{\nu}$ such that the following duality holds:

$$\langle \bar{q}_t, \bar{\nu}_t \rangle = \langle \Gamma_t^{-1} q_t, \Gamma_t \nu_t \rangle = \langle q_t, \nu_t \rangle, \quad \forall t \in \mathcal{T} \quad (52)$$

where $\bar{\nu}_t$ satisfies (29) and Γ_t^{-1} satisfies (23).

By discretizing the O.D.E. for $\bar{\nu}$ satisfying (29) and recovering the process ν in Definition 3.7

$$\nu_{m-1} \approx (\mathbf{I} + \Delta A^T) \Phi_{m,m-1} \nu_m \quad (53)$$

where $\nu_T = \mathbf{1}$.

By adopting a duality, we obtain the smoother-based update equations for the estimates \hat{a}_{ij} and $\hat{\lambda}_i$ of a_{ij} and λ_i , respectively.

Theorem 4.6: The smoother-based update equations for \hat{a}_{ij} and $\hat{\lambda}_i$ are

$$\hat{a}_{ij}(k+1) = \hat{a}_{ij}(k) \frac{\int_0^T \langle q_t, e_i \rangle \langle \nu_t, e_j \rangle dt}{\int_0^T \langle q_t, e_i \rangle \langle \nu_t, e_i \rangle dt} \quad (54)$$

and

$$\begin{aligned} \hat{\lambda}_i(k+1) = & [Q_i(T) N_T \langle q_T, e_i \rangle \langle \nu_T, e_i \rangle \\ & - \int_0^T \langle \lambda - 1, e_i \rangle \langle \bar{q}_t, e_i \rangle \langle \bar{\nu}_t, e_i \rangle dt \\ & + \int_0^T N_t \langle \bar{q}_t, e_i \rangle \langle \Gamma_t^{-1} A \Gamma_t Q_i(t) e_i, \bar{\nu}_t \rangle dt \\ & - \int_0^T N_t \langle \bar{\nu}_t, e_i \rangle \langle d[Q_i(t) \bar{q}_t], e_i \rangle dt] \\ & \times \left(\int_0^T \langle q_t, e_i \rangle \langle \nu_t, e_i \rangle dt \right)^{-1}. \end{aligned} \quad (55)$$

Proof: See Appendix.

Note that the index $k \in \{1, 2, \dots\}$ represents each pass of the estimation scheme that determines the elements of the transition matrix A and the vector of intensity parameters λ . We need to update q_u and ν_u in each step in (54) and (55) using estimates obtained in the k^{th} pass.

To sum up, the smoother-based EM algorithm is:

Step I: Select $\hat{a}_{ij}(0)$ and $\hat{\lambda}_i(0)$.

Step II: Using the discrete time forward and backward equations, respectively, (48) and (53), evaluate (54) and (55) to compute the MLEs, $\hat{a}_{ij}(k+1)$ and $\hat{\lambda}_i(k+1)$.

Step III: Stop when $|\hat{a}_{ij}(k+1) - \hat{a}_{ij}(k)| < \epsilon_1$ and $|\hat{\lambda}_i(k+1) - \hat{\lambda}_i(k)| < \epsilon_2$; otherwise, continue from Step II, where ϵ_1, ϵ_2 are the desirable levels of accuracy and $\epsilon_1, \epsilon_2 > 0$.

V. A SIMULATION EXAMPLE

In this section we present a simulation example to illustrate the implementation of the filters and the smoothers derived here. We consider a two-state Markov chain with two-state distribution of jumps. We say that the chain X is in “State 1” (“State 2”) at time t if $X_t = e_1$ ($X_t = e_2$). We suppose that the distributions of jump sizes are two gamma distributions with different shape parameters, but the same scale parameter. That is

$$f_i(z) = \frac{1}{b^{a_i} \Gamma(a_i)} z^{a_i-1} e^{-z/b}, \quad z > 0, \quad i = 1, 2.$$

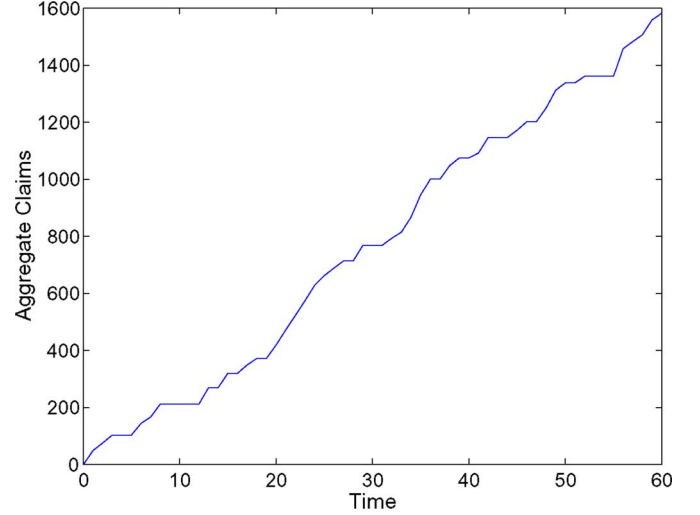


Fig. 1. Simulated aggregate claims process.

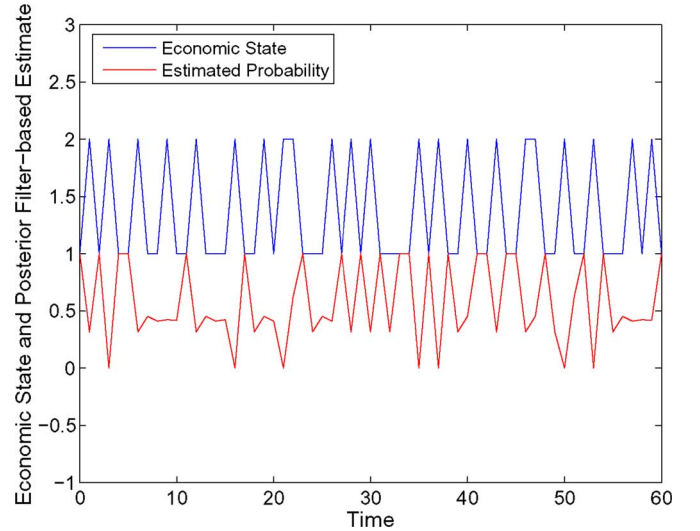


Fig. 2. $P(X = 1 | \text{Observation})$, filter.

In actuarial science, gamma distributions are commonly used distributions for claim sizes. We consider the following configuration of specimen values of the parameters in the simulation example

$$\begin{aligned} M &= 60, \Delta = 1/12, \\ \lambda_1 &= 10, \lambda_2 = 5, a_1 = 50, a_2 = 100, b = 0.5, \\ a_{11} &= -5, a_{12} = 5, a_{21} = 10, a_{22} = -10. \end{aligned}$$

We assume that the Markov chain is in “State 1” initially, (i.e. $X_0 = e_1$). Note that $Z_0 = 0$, which means that the initial claim amount is zero. All figures are presented in the Appendix after the proofs.

Fig. 1 depicts a simulation path of the aggregate claim process Z , which is our observation process.

Fig. 2 depicts a simulated hidden Markov chain and the robust, filter-based, estimated state probability $P(X = 1 | \text{Observation})$.

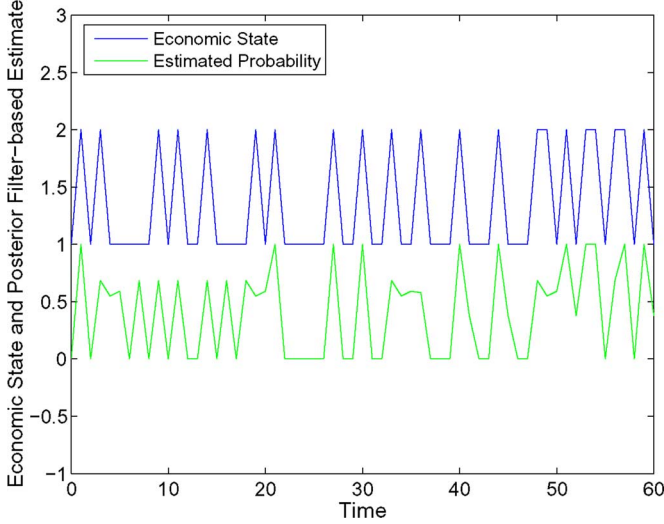
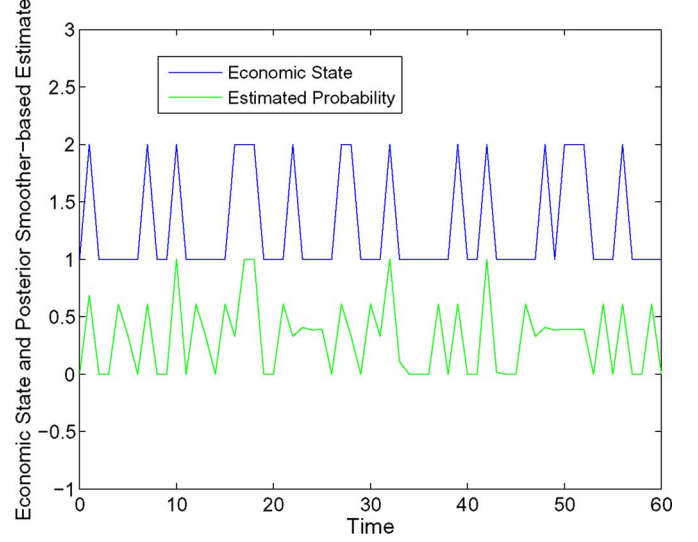
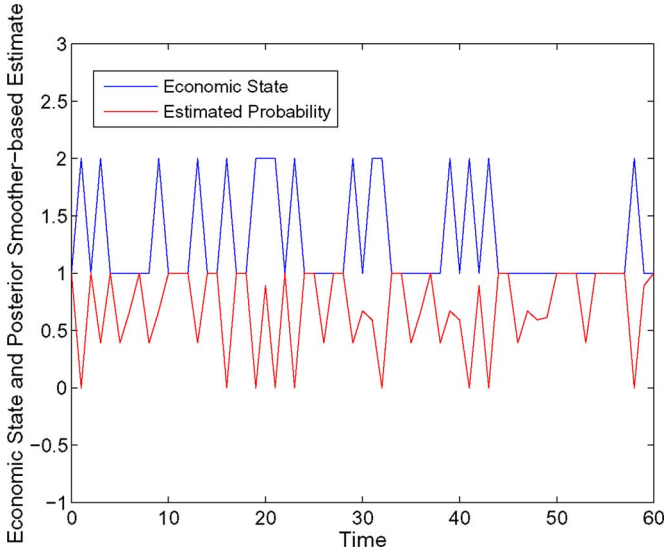
Fig. 3. $P(X = 2|\text{Observation})$, filter.Fig. 5. $P(X = 2|\text{Observation})$, smoother.Fig. 4. $P(X = 1|\text{Observation})$, smoother.

Fig. 3 depicts a simulated hidden Markov chain and the robust, filter-based, estimated state probability $P(X = 2|\text{Observation})$.

Fig. 4 depicts a simulated hidden Markov chain and the robust, smoother-based, estimated state probability $P(X = 1|\text{Observation})$.

Fig. 5 depicts a simulated hidden Markov chain and the robust, smoother-based, estimated state probability $P(X = 2|\text{Observation})$.

From Figs. 2 and 3, we see that the robust, filter-based, estimated state probabilities give reasonably good estimates for the hidden Markov chain. Comparing Figs. 2 and 3 with Figs. 4 and 5, the robust, filter-based, estimates seem outperforming the robust, smoother-based, estimates of the hidden Markov chain.

VI. CONCLUSION

We developed new recursive filter when the hidden Markov chain modulates both the jump sizes and times of

a Markov-modulated random measure. The model considered here is a natural generalization of marked point processes studied in [12], [13]. The Markov-modulated random measure is a flexible model for an insurance claims process, in which both the stochastic intensity of the claim arrivals and the distribution of the claim sizes switch over time according to a continuous-time, finite-state, hidden Markov chain. The filter developed here is new. The filter and the smoother, which can be obtained from solving linear ordinary differential equations, provide us with a convenient way to select a model that is the “best” estimate in the “mean-squared-error” sense. We provided the estimation method of the unknown parameters in the Markov-modulated random measure using the robust filter-based and smoother-based EM algorithms. The development provides recursive estimates in the framework of modern theory of stochastic processes. A simulation example was provided to illustrate the implementation of the filters and the smoothers. The simulation results revealed that the robust, filter-based, estimates seem outperforming the robust, smoother-based, estimates of the hidden Markov chain.

APPENDIX A

PROOF OF LEMMAS, THEOREMS AND COROLLARIES

Proof of Theorem 3.2: First, note that

$$X_t = X_0 + \int_0^t AX_u du + M_t. \quad (56)$$

Then, applying the product rule to the process ΛX

$$\begin{aligned} \Lambda_{0,t} X_t &= X_0 + \int_0^t \Lambda_{0,u-} X_u - \\ &\quad \times \left(\sum_{k=1}^K \int_{-\infty}^{\infty} \langle X_{u-}, e_k \rangle (h_k(y) - 1) \right. \\ &\quad \left. \times (\gamma(dy, du) - f(y) dy du) \right) \end{aligned}$$

$$+ \int_0^t \Lambda_{0,u} A X_u du + \int_0^t \Lambda_{0,u-} dM_u. \quad (57)$$

Conditioning on \mathcal{F}_t^Z under \mathcal{P}^\dagger (strictly, taking the \mathcal{F}^Z -optional projection under \mathcal{P}^\dagger), we have

$$q_t = q_0 + \int_0^t A q_u du + \int_0^t \left(\sum_{k=1}^K \int_{-\infty}^{\infty} \langle q_{u-}, e_k \rangle (h_k(y) - 1) \times (\gamma(dy, du) - f(y) dy du) \right) e_k. \quad (58)$$

Now

$$\int_{-\infty}^{\infty} (h_k(y) - 1) f(y) dy = \lambda_k - 1. \quad (59)$$

Thus

$$q_t = q_0 + \int_0^t A q_u du + \int_0^t \text{diag}(H(u) - 1) q_{u-} dN_u - \int_0^t \text{diag}(\lambda - 1) q_u du. \quad (60)$$

Proof of Lemma 3.3: First, by the differentiation rule (see, for example, Theorem 12.19 in [10]), for any real-valued function $F \in \mathcal{C}^1(\mathfrak{R})$ (i.e. $\mathcal{C}^1(\mathfrak{R})$ is the space of first-order differentiable functions on \mathfrak{R})

$$F(L_t) = F(L_0) + \int_0^t F'(L_{u-}) dL_u + \sum_{0 < u \leq t} [F(L_u) - F(L_{u-}) - F'(L_u) \Delta L_u]. \quad (61)$$

We shall apply this to $F(L) = e^{-L}$ so $F'(L) = -e^{-L}$, and with $L_t = L_t^k$

$$\Delta L_u = \log H_k(u) \Delta N_u. \quad (62)$$

Also

$$\begin{aligned} F(L_u) - F(L_{u-}) &= e^{-L_{u-}} [\exp(-\log H_k(u) \Delta N_u) - 1] \\ &= e^{-L_{u-}} \left(\frac{1}{H_k(u)} - 1 \right) \Delta N_u. \end{aligned} \quad (63)$$

Therefore

$$\begin{aligned} d(\gamma_t^k)^{-1} &= -(\gamma_{t-}^k)^{-1} [(1 - \lambda_k) dt + \log H_k(t) \Delta N_t] \\ &\quad + (\gamma_{t-}^k)^{-1} \left[\left(\frac{1}{H_k(t)} - 1 \right) \Delta N_t + \log H_k(t) \Delta N_t \right] \\ &= -(\gamma_t^k)^{-1} (1 - \lambda_k) dt \\ &\quad + (\gamma_{t-}^k)^{-1} \left(\frac{1 - H_k(t)}{H_k(t)} \right) \Delta N_t. \end{aligned} \quad (64)$$

Since

$$\Gamma_t^{-1} = \text{diag} \left((\gamma_t^1)^{-1}, (\gamma_t^2)^{-1}, \dots, (\gamma_t^K)^{-1} \right)$$

the result follows.

Proof of Theorem 3.5: First note that

$$\bar{q}_0 = q_0 = E^\dagger [X_0 | \mathcal{F}_0^Z] = E^\dagger [X_0]. \quad (65)$$

As diagonal matrices commute

$$\begin{aligned} \Gamma_t^{-1} q_t &= q_0 + \int_0^t \Gamma_u^{-1} dq_u + \int_0^t d\Gamma_u^{-1} q_{u-} + \sum_{0 < u \leq t} (\Delta \Gamma_u) (\Delta q_u) \\ &= q_0 + \int_0^t \Gamma_u^{-1} A q_u du + \int_0^t \text{diag}(H(u) - 1) \Gamma_u^{-1} q_{u-} dN_u \\ &\quad - \int_0^t \text{diag}(\lambda - 1) \Gamma_u^{-1} q_u du + \int_0^t \text{diag}(\lambda - 1) \Gamma_u^{-1} q_u du \\ &\quad + \int_0^t \text{diag} \left(\frac{1 - H(u)}{H(u)} \right) \Gamma_u^{-1} q_{u-} dN_u \\ &\quad + \int_0^t \text{diag}(H(u) - 1) \text{diag} \left(\frac{1 - H(u)}{H(u)} \right) \Gamma_u^{-1} q_{u-} dN_u \\ &= q_0 + \int_0^t \Gamma_u^{-1} A q_u du \\ &\quad + \int_0^t \left[\text{diag}(H(u) - 1) + \text{diag} \left(\frac{1 - H(u)}{H(u)} \right) \right. \\ &\quad \left. + \text{diag}(H(u) - 1) \text{diag} \left(\frac{1 - H(u)}{H(u)} \right) \right] \\ &\quad \times \Gamma_u^{-1} q_{u-} dN_u. \end{aligned} \quad (66)$$

Note that the sum of the k^{th} terms on the diagonal matrices in the dN_u integral is

$$H_k - 1 + \frac{1 - H_k}{H_k} + (H_k - 1) \left(\frac{1 - H_k}{H_k} \right) = 0 \quad (67)$$

and so the dN_u integral vanishes.

Therefore

$$\Gamma_t^{-1} q_t = \bar{q}_t = \int_0^t \Gamma_u^{-1} A q_u du. \quad (68)$$

The result follows by writing $q_u = \Gamma_u \bar{q}_u$.

Proof of Theorem 3.8: The proof is the same as the proof of Theorem 8 in [13]. So, we do not repeat the detail here. Instead, we only highlight some of the key steps of the proof.

First, by the semi-group property of Λ and the double expectation formula

$$r_t = E^\dagger [\Lambda_{0,t} X_t E^\dagger [\Lambda_{t,T} | \mathcal{F}_T^Z \vee \mathcal{F}_t^X] | \mathcal{F}_T^Z]. \quad (69)$$

Due to the Markov property of X under \mathcal{P}^\dagger

$$E^\dagger [\Lambda_{t,T} | \mathcal{F}_T^Z \vee \mathcal{F}_t^X] = E^\dagger [\Lambda_{t,T} | \mathcal{F}_T^Z \vee \sigma\{X_t\}]. \quad (70)$$

Since $\sum_{k=1}^K \langle X_t, e_k \rangle = 1$, ν_t^k is \mathcal{F}_T^Z -measurable and Z has independent increments under \mathcal{P}^\dagger

$$\begin{aligned} r_t &= E^\dagger \left[\Lambda_{0,t} \left(\sum_{k=1}^K \langle X_t, e_k \rangle \right) X_t E^\dagger [\Lambda_{t,T} | \mathcal{F}_T^Z \vee \mathcal{F}_t^X] | \mathcal{F}_T^Z \right] \\ &= \sum_{k=1}^K q_t^k \nu_t^k e_k \in \mathbb{R}^K. \end{aligned} \quad (71)$$

Then, the normalized smoothed estimate of X is

$$E[X_t | \mathcal{F}_T^Z] = \frac{r_t}{\langle r_t, \mathbf{1} \rangle}. \quad (72)$$

Note that the following duality holds, (see [13]):

$$\langle r_t, \mathbf{1} \rangle = \langle q_t, \nu_t \rangle. \quad (73)$$

Since the process $\langle r_t, \mathbf{1} \rangle$ is independent of time, so is $\langle q_t, \nu_t \rangle$, from the above equation. Then

$$\frac{d}{dt} \langle r_t, \mathbf{1} \rangle = \frac{d}{dt} \langle q_t, \nu_t \rangle = 0. \quad (74)$$

For the rest of the proof, interested readers may refer to [13].

Proof of Theorem 4.1: Using the product rule, the decomposition of the process $\Lambda G^i X$ is computed as

$$\begin{aligned} \Lambda_{0,t} G_t^i X_t &= \int_0^t \Lambda_{0,u-} X_u \langle X_u, e_i \rangle dN_u + \int_0^t \Lambda_u G_u^i A X_u du \\ &\quad + \int_0^t \Lambda_{0,u} G_u^i dM_u + \int_0^t G_u^i X_u \Lambda_{0,u-} \\ &\quad \times \left(\sum_{k=1}^K \int_{-\infty}^{\infty} \langle X_u, e_k \rangle (h_k(y) - 1) \right. \\ &\quad \times (\gamma(dy, du) - f(y)dydu)) \\ &\quad + \int_0^t \int_{-\infty}^{\infty} X_u \Lambda_{0,u-} \langle X_u, e_i \rangle (h_i(y) - 1) \\ &\quad \times (\gamma(dy, du) - f(y)dydu). \end{aligned} \quad (75)$$

Taking the \mathcal{F}_t^Z -optional projection under \mathcal{P}^\dagger

$$\begin{aligned} \sigma(G_t^i X_t) &= \int_0^t \langle q_{u-}, e_i \rangle dN_u e_i + \int_0^t A \sigma(G_u^i X_u) du \\ &\quad + \int_0^t \left(\sum_{k=1}^K \int_{-\infty}^{\infty} \langle \sigma(G_{u-}^i X_{u-}), e_k \rangle (h_k(y) - 1) \right. \\ &\quad \times (\gamma(dy, du) - f(y)dydu)) e_k \\ &\quad + \int_0^t \int_{-\infty}^{\infty} \langle q_{u-}, e_i \rangle (h_i(y) - 1) \\ &\quad \times (\gamma(dy, du) - f(y)dydu) e_i \\ &= \int_0^t \langle q_{u-}, e_i \rangle dN_u e_i + \int_0^t A \sigma(G_u^i X_u) du \end{aligned}$$

$$\begin{aligned} &+ \int_0^t \text{diag}(H(u) - 1) \sigma(G_{u-}^i X_{u-}) dN_u \\ &- \int_0^t \text{diag}(\lambda - 1) \sigma(G_u^i X_u) du \\ &+ \int_0^t (H_i(Y_u(\omega)) - 1) \langle q_{u-}, e_i \rangle dN_u e_i \\ &- \int_0^t (\lambda_i - 1) \langle q_u, e_i \rangle du e_i \\ &= \int_0^t \text{diag}(H(u) - 1) \sigma(G_{u-}^i X_{u-}) dN_u \\ &+ \int_0^t A \sigma(G_u^i X_u) du \\ &- \int_0^t \text{diag}(\lambda - 1) \sigma(G_u^i X_u) du \\ &+ \int_0^t \langle H(u), e_i \rangle \langle q_{u-}, e_i \rangle dN_u e_i \\ &- \int_0^t \langle \lambda - 1, e_i \rangle \langle q_u, e_i \rangle du e_i \end{aligned} \quad (76)$$

where $\sigma(G_0^i X_0) = 0$, \mathcal{P} -a.s.

Similarly, it can be shown that

$$\begin{aligned} \sigma(N_t^{ij} X_t) &= \int_0^t A \sigma(N_u^{ij} X_u) du + \int_0^t \langle q_{u-}, e_i \rangle \langle A e_i, e_j \rangle du e_j \\ &\quad + \int_0^t \text{diag}(H(u) - 1) \sigma(N_{u-}^{ij} X_{u-}) dN_u \\ &\quad - \int_0^t \text{diag}(\lambda - 1) \sigma(N_u^{ij} X_u) du \end{aligned} \quad (77)$$

and

$$\begin{aligned} \sigma(O_t^i X_t) &= \int_0^t A \sigma(O_u^i X_u) du + \int_0^t \langle q_u, e_i \rangle du e_i \\ &\quad + \int_0^t \text{diag}(H(u) - 1) \sigma(O_{u-}^i X_{u-}) dN_u \\ &\quad - \int_0^t \text{diag}(\lambda - 1) \sigma(O_u^i X_u) du \end{aligned} \quad (78)$$

where $\sigma(N_0^{ij} X_0) = \sigma(O_0^i X_0) = 0$, \mathcal{P} -a.s.

Note that

$$\begin{aligned} \langle \sigma(N_t^{ij} X_t), \mathbf{1} \rangle &= \langle E^\dagger [\Lambda_{0,t} N_t^{ij} X_t | \mathcal{F}_t^Z], \mathbf{1} \rangle \\ &= \sigma(N_t^{ij}). \end{aligned} \quad (79)$$

Similarly

$$\langle \sigma(O_t^i X_t), \mathbf{1} \rangle = \sigma(O_t^i) \quad (80)$$

and

$$\langle \sigma(G_t^i X_t), \mathbf{1} \rangle = \sigma(G_t^i). \quad (81)$$

Proof of Theorem 4.3: Write $Q_i(t) := \text{diag}(1/H(t)) \langle H(t), e_i \rangle$, for each $t \in \mathcal{T}$ and $i = 1, 2, \dots, N$. Since diagonal matrices commute, from (23) and (36)

$$\begin{aligned} \Gamma_t^{-1} \sigma(G_t^i X_t) &= \int_0^t \Gamma_{u-}^{-1} d\sigma(G_u^i X_u) + \int_0^t d\Gamma_u^{-1} \sigma(G_{u-}^i X_u) \\ &\quad + \sum_{0 < u \leq t} (\Delta \Gamma_u^{-1}) (\Delta \sigma(G_u^i X_u)) \end{aligned} \quad (82)$$

where $\Delta H_t := H_t - H_{t-}$.

Then

$$\begin{aligned} \bar{\sigma}(G_t^i X_t) &= \int_0^t \Gamma_u^{-1} A \Gamma_u \bar{\sigma}(G_u^i X_u) du \\ &\quad + \int_0^t \langle H(u), e_i \rangle \langle \bar{q}_{u-}, e_i \rangle dN_u e_i \\ &\quad - \int_0^t \langle \lambda - 1, e_i \rangle \langle \bar{q}_u, e_i \rangle du e_i \\ &\quad - \int_0^t \text{diag}(\lambda - 1) \bar{\sigma}(G_u^i X_u) du \\ &\quad + \int_0^t \text{diag}(H(u) - 1) \bar{\sigma}(G_{u-}^i X_{u-}) dN_u \\ &\quad + \int_0^t \text{diag}(\lambda - 1) \bar{\sigma}(G_{u-}^i X_{u-}) dN_u \\ &\quad + \int_0^t \text{diag}(\lambda - 1) \bar{\sigma}(G_u^i X_u) du \\ &\quad + \int_0^t \text{diag}\left(\frac{1 - H(u)}{H(u)}\right) \bar{\sigma}(G_{u-}^i X_{u-}) dN_u \\ &\quad + \int_0^t \text{diag}\left(\frac{1 - H(u)}{H(u)}\right) \text{diag}(H(u) - 1) \\ &\quad \times \bar{\sigma}(G_{u-}^i X_{u-}) dN_u \end{aligned}$$

$$\begin{aligned} &+ \int_0^t \text{diag}\left(\frac{1 - H(u)}{H(u)}\right) \langle H(u), e_i \rangle \\ &\quad \times \langle \bar{q}_{u-}, e_i \rangle dN_u e_i. \end{aligned} \quad (83)$$

Note that

$$\begin{aligned} &\text{diag}(H(u) - 1) + \text{diag}\left(\frac{1 - H(u)}{H(u)}\right) + \\ &\quad \text{diag}(H(u) - 1) \text{diag}\left(\frac{1 - H(u)}{H(u)}\right) = \mathbf{0} \in \mathbb{R}^{K \times K}. \end{aligned} \quad (84)$$

This implies that

$$\begin{aligned} \bar{\sigma}(G_t^i X_t) &= \int_0^t \Gamma_u^{-1} A \Gamma_u \bar{\sigma}(G_u^i X_u) du \\ &\quad + \int_0^t Q_i(u-) \langle \bar{q}_{u-}, e_i \rangle dN_u e_i - \int_0^t \langle \lambda - 1, e_i \rangle \langle \bar{q}_u, e_i \rangle du e_i. \end{aligned} \quad (85)$$

By the stochastic integration by parts

$$\begin{aligned} \int_0^t Q_i(u-) \langle \bar{q}_{u-}, e_i \rangle dN_u e_i &= Q_i(t) \langle \bar{q}_t, e_i \rangle N_t e_i \\ &\quad - \int_0^t N_u \langle d[Q_i(u) \bar{q}_u], e_i \rangle e_i. \end{aligned} \quad (86)$$

Hence

$$\begin{aligned} \bar{\sigma}(G_t^i X_t) &= \int_0^t \Gamma_u^{-1} A \Gamma_u \bar{\sigma}(G_u^i X_u) du + Q_i(t) \langle \bar{q}_t, e_i \rangle N_t e_i \\ &\quad - \int_0^t \langle \lambda - 1, e_i \rangle \langle \bar{q}_u, e_i \rangle du e_i - \int_0^t N_u \langle d[Q_i(u) \bar{q}_u], e_i \rangle e_i. \end{aligned} \quad (87)$$

Similarly, we can derive the dynamics of the gauge transformed processes $\bar{\sigma}(N_t^{ij} X_t)$ and $\bar{\sigma}(O_t^i X_t)$.

Proof of Theorem 4.4: Note that

$$\begin{aligned} \bar{\sigma}(G_{t_m}^i X_{t_m}) &= \bar{\sigma}(G_{t_{m-1}}^i X_{t_{m-1}}) \\ &\quad + \int_{t_{m-1}}^{t_m} \Gamma_u^{-1} A \Gamma_u \bar{\sigma}(G_u^i X_u) du \\ &\quad + Q_i(t_m) \langle \bar{q}_{t_m}, e_i \rangle N_{t_m} e_i \\ &\quad - Q_i(t_{m-1}) \langle \bar{q}_{t_{m-1}}, e_i \rangle N_{t_{m-1}} e_i \\ &\quad - \int_{t_{m-1}}^{t_m} \langle \lambda - 1, e_i \rangle \langle \bar{q}_u, e_i \rangle du e_i \\ &\quad - \int_{t_{m-1}}^{t_m} N_u \langle d[Q_i(u) \bar{q}_u], e_i \rangle e_i. \end{aligned} \quad (88)$$

Now

$$\begin{aligned}
& \int_{t_{m-1}}^{t_m} N_u \langle d[Q_i(u)\bar{q}_u], e_i \rangle e_i \\
&= \int_{t_{m-1}}^{t_m} N_u \langle \bar{q}_u dQ_i(u), e_i \rangle e_i \\
&\quad + \int_{t_{m-1}}^{t_m} N_u \langle Q_i(u)\Gamma_u^{-1} A \Gamma_u \bar{q}_u du, e_i \rangle e_i \\
&\approx N_{t_{m-1}} \langle \bar{q}_{t_{m-1}} Q_i(t_m), e_i \rangle e_i \\
&\quad - N_{t_{m-1}} \langle \bar{q}_{t_{m-1}} Q_i(t_{m-1}), e_i \rangle e_i \\
&\quad + N_{t_{m-1}} \langle Q_i(t_{m-1})\Gamma_{t_{m-1}}^{-1} A q_{t_{m-1}}, e_i \rangle \Delta e_i \\
&= N_{t_{m-1}} \langle \bar{q}_{t_{m-1}} Q_i(t_m), e_i \rangle e_i \\
&\quad + N_{t_{m-1}} \langle Q_i(t_{m-1})\Gamma_{t_{m-1}}^{-1} (\Delta A - \mathbf{I}) q_{t_{m-1}}, e_i \rangle e_i. \quad (89)
\end{aligned}$$

Hence

$$\begin{aligned}
& \sigma(G_{t_m}^i X_{t_m}) \\
&\approx \Phi_{m,m-1} [\mathbf{I} + \Delta A] \sigma(G_{t_{m-1}}^i X_{t_{m-1}}) \\
&\quad + \Gamma_{t_m} [Q_i(t_m) \langle \bar{q}_{t_m}, e_i \rangle N_{t_m} e_i \\
&\quad \quad - Q_i(t_{m-1}) \langle \bar{q}_{t_{m-1}}, e_i \rangle N_{t_{m-1}} e_i \\
&\quad \quad - \langle \lambda - 1, e_i \rangle \langle \bar{q}_{t_{m-1}}, e_i \rangle \Delta e_i \\
&\quad \quad - N_{t_{m-1}} \langle Q_i(t_m) \bar{q}_{t_{m-1}}, e_i \rangle e_i - N_{t_{m-1}} \\
&\quad \quad \times \langle Q_i(t_{m-1}) \Gamma_{t_{m-1}}^{-1} (\Delta A - \mathbf{I}) q_{t_{m-1}}, e_i \rangle e_i]. \quad (90)
\end{aligned}$$

Similarly, it can be shown that

$$\begin{aligned}
& \sigma(N_{t_m}^{ij} X_{t_m}) \approx \Phi_{m,m-1} [\mathbf{I} + \Delta A] \sigma(N_{t_{m-1}}^{ij} X_{t_{m-1}}) \\
&\quad + \Phi_{m,m-1} \langle q_{t_{m-1}}, e_i \rangle \langle A e_i, e_j \rangle \Delta e_j \quad (91)
\end{aligned}$$

and

$$\begin{aligned}
& \sigma(O_{t_m}^i X_{t_m}) \approx \Phi_{m,m-1} [\mathbf{I} + \Delta A] \sigma(O_{t_{m-1}}^i X_{t_{m-1}}) \\
&\quad + \Phi_{m,m-1} \langle q_{t_{m-1}}, e_i \rangle \Delta e_i. \quad (92)
\end{aligned}$$

Proof of Corollary 4.5: Consider the following stochastic difference equation:

$$W_m = C_m W_{m-1} + D_m, \quad m = 1, 2, \dots, M. \quad (93)$$

Then it is easy to show that the solution of the stochastic difference equation is given by

$$\begin{aligned}
W_m &= W_0 \prod_{j=1}^m C_j + \sum_{l=1}^m D_l \left(\prod_{n=l+1}^m C_n \right), \\
m &= 1, 2, \dots, M. \quad (94)
\end{aligned}$$

Set $W_m = q_m$, $C_m = \Phi_{m,m-1}(I + \Delta A)$ and $D_m = 0$. Then

$$q_m = q_0 \left[\prod_{j=1}^m \Phi_{j,j-1}(I + \Delta A) \right]. \quad (95)$$

By noticing that $\Phi_{j,j-1} = \text{diag}(U_{j,j-1})$, the formula (48) follows. Note that the matrices in the product $\prod_{j=1}^m \Phi_{j,j-1}(I + \Delta A)$ do not commute. So we leave the solution q as the product form.

Similarly, the formulas (49)–(51) can be proved by applying (A.39) and by noting that $\sigma(G_{t_0}^i X_{t_0}) = \sigma(N_{t_0}^{ij} X_{t_0}) = \sigma(O_{t_0}^i X_{t_0}) = 0$.

Proof of Theorem 4.6: By exploiting the duality

$$\begin{aligned}
\langle \sigma(G_T^i X_T), \nu_T \rangle &= \langle \bar{\sigma}(G_T^i X_T), \bar{\nu}_T \rangle \\
&= Q_i(T) N_T \langle q_T, e_i \rangle \langle \nu_T, e_i \rangle \\
&\quad - \int_0^T \langle \lambda - 1, e_i \rangle \langle \bar{q}_t, e_i \rangle \langle \bar{\nu}_t, e_i \rangle dt \\
&\quad + \int_0^T N_t \langle \bar{q}_t, e_i \rangle \langle \Gamma_t^{-1} A \Gamma_t Q_i(t) e_i, \bar{\nu}_t \rangle dt \\
&\quad - \int_0^T N_t \langle \bar{\nu}_t, e_i \rangle \langle d[Q_i(t) \bar{q}_t], e_i \rangle. \quad (96)
\end{aligned}$$

Similarly

$$\begin{aligned}
\langle \sigma(N_T^{ij} X_T), \nu_T \rangle &= \langle \bar{\sigma}(N_T^{ij} X_T), \bar{\nu}_T \rangle \\
&= \int_0^T \langle A e_i, e_j \rangle \langle q_t, e_i \rangle \langle \nu_t, e_j \rangle dt \quad (97)
\end{aligned}$$

and

$$\begin{aligned}
\langle \sigma(O_T^i X_T), \nu_T \rangle &= \langle \bar{\sigma}(O_T^i X_T), \bar{\nu}_T \rangle \\
&= \int_0^T \langle q_t, e_i \rangle \langle \nu_t, e_i \rangle dt. \quad (98)
\end{aligned}$$

ACKNOWLEDGMENT

The authors would like to thank the referees for their many helpful and valuable comments and suggestions to the paper, and the participants of the conferences, workshops, seminars and colloquium talks (where preliminary versions of this paper were presented) for their valuable comments and suggestions.

REFERENCES

- [1] S. Asmussen, "Risk theory in a Markovian environment," *Scandinavian Actuarial J.*, vol. 2, pp. 69–100, 1989.
- [2] S. Asmussen, *Ruin Probability*. Singapore: World Scientific, 2000.
- [3] E. Bayraktar and S. Sezer, "Quickest detection for a Poisson process with a Phase-type change-time distribution," *Sequential Analysis*, to be published.
- [4] P. Bremaud, "La methode des semi-martingales en filtrage quand L'observation est un processus ponctuel marque," in *Seminaire De Probabilites, Lecture Notes in Mathematics 511*. Berlin, Germany: Springer-Verlag, 1976, pp. 1–18.

- [5] P. Bremaud, *Point Processes and Queues Martingale Dynamics*. Berlin, Germany: Springer-Verlag, 1981.
- [6] J. M. Clark, "The design of robust approximations to the stochastic differential equations for nonlinear filtering," in *Communications Systems and Random Process Theory*, J. K. Skwirzynski, Ed. Amsterdam, The Netherlands: Sijthoff and Noorhoff, 1978, pp. 721–734.
- [7] A. Dembo and O. Zeitouni, "Parameter estimation of partially observed continuous time stochastic process via the EM algorithm," *Stoch. Processes Appl.*, vol. 23, pp. 91–113, 1989.
- [8] J. H. Dshalalow, "On modulated random measures," *J. Appl. Math. Stoch. Anal.*, vol. 4, pp. 305–312, 1991.
- [9] J. H. Dshalalow, "Characterization of modulated Cox measures on topological spaces," *Int. J. Appl. Math. Stat.*, vol. 11, pp. 21–37, 2007.
- [10] R. J. Elliott, *Stochastic Calculus and Applications*. Berlin, Germany: Springer-Verlag, 1982.
- [11] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models: Estimation and Control*. Berlin, Germany: Springer-Verlag, 1994.
- [12] R. J. Elliott and W. P. Malcolm, "Robust EM algorithms for Markov-modulated Poisson processes," in *Proc. 39th IEEE Conf. Decision Control*, Sydney, Australia, 2000, pp. 4678–4685.
- [13] R. J. Elliott and W. P. Malcolm, "General smoothing formulas for Markov-modulated Poisson observations," *IEEE Trans. Autom. Control*, vol. 50, no. 8, pp. 1123–1134, Aug. 2005.
- [14] R. J. Elliott and C.-J. U. Osakwe, "Option pricing for pure jump processes with Markov switching compensators," *Finance Stoch.*, vol. 10, pp. 250–275, 2006.
- [15] O. Kallenberg, *Random Measures*, 2nd ed. New York: Academic Press, 1986.
- [16] S. A. Klugman, H. H. Panjer, and G. E. Willmot, *Loss Models: Data, Decisions, and Risks*. New York: Wiley, 2004.
- [17] W. P. Malcolm, R. J. Elliott, and J. van der Hoeck, "On the numerical stability of time-discretized state estimation via Clark transformations," in *Proc. IEEE Conf. Decision Control*, Maui, HI, Dec. 2003, pp. 1406–1412.
- [18] T. Møller, "Stochastic orders in dynamic reinsurance markets," *Finance Stoch.*, vol. 8, pp. 479–499, 2004.
- [19] M. F. Neuts, "A versatile Markovian point process," *J. Appl. Prob.*, vol. 16, pp. 764–779, 1979.
- [20] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. Baltimore, MD: J. Hopkins Univ. Press, 1981.
- [21] J. M. Reinhard, "On a class of semi-Markov risk models obtained as classical risk models in a Markovian environment," *ASTIN Bulletin*, vol. 31, pp. 255–273, 1984.
- [22] T. Rolski, H. Schmidli, V. Schmidt, and J. Teugels, *Stochastic Processes for Finance and Insurance*. New York: Wiley, 1999.
- [23] Y. Zeng, "A partially observed model for micromovement of asset prices with Bayes estimation via filtering," *Math. Finance*, vol. 13, pp. 411–444, 2003.



Robert J. Elliott received the B.S. and M.S. degrees from Oxford University, Oxford, U.K. and the Ph.D. and D.Sc. degrees from Cambridge University, Cambridge, U.K.

He has held positions at Newcastle, Yale, Oxford, Warwick, Hull, Alberta, and visiting positions in Toronto, Northwestern, Kentucky, Brown, Paris, Denmark, Hong Kong and Australia. From 2001 to 2009, he was the RBC Financial Group Professor of Finance at the University of Calgary, Calgary, AB, Canada, where he is also an Adjunct Professor in

both the Department of Mathematics and the Department of Electrical Engineering. Currently, he is an Australian Professorial Fellow at the University of Adelaide, Adelaide, Australia. He has authored nine books and over 300 papers. He wrote *Stochastic Calculus and Applications* (New York: Springer, 1982). He co-wrote *Mathematics of Financial Markets* (New York: Springer, 1999), *Binomial Methods in Finance* (New York: Springer, 2006), *Hidden Markov Models: Estimation and Control* (Berlin, Germany: Springer Verlag, 1995), and *Measure Theory and Filtering* (Cambridge, U.K.: Cambridge Univ. Press, 2004).



Tak Kuen Siu received the B.S. degree from the Hong Kong University of Science and Technology and the Ph.D. degree from the University of Hong Kong.

He is an Associate Professor in actuarial studies, Faculty of Business and Economics, Macquarie University, Sydney, Australia. He serves as a member in the editorial boards of several journals, including *Stochastics* and the *IMA Journal of Management Mathematics*. He is a Research Affiliate of the Centre for Research into Industry, Enterprise, Finance and

the Firm (CRIEFF), School of Economics and Finance, University of St. Andrews, U.K., and a Member of the Advanced Modeling and Applied Computing Laboratory, Department of Mathematics, University of Hong Kong. He has authored over 70 papers. His research interests include mathematical finance, actuarial science, and risk management.

Dr. Siu was an Honorary Fellow of the College of Science and Engineering, The University of Edinburgh, from 2004 to 2008. He has been an Affiliate Member of the Financial Integrity Research Network (FIRN), an ARC Research Network, since 2009.



Hailiang Yang was born in Inner Mongolia, China. He received the Ph.D. degree from the University of Alberta, Edmonton, Canada.

He is a Professor with the Department of Statistics and Actuarial Science, The University of Hong Kong. He has published over 100 papers in these fields. He serves as a member in the editorial boards of several journals including *Insurance: Mathematics and Economics*, *Stochastics*, and *Acta Mathematicae Applicatae Sinica*. His research interests are on mathematical finance and actuarial science.

Dr. Yang is an Associate of the Society of Actuaries and an elected member of the International Statistical Institute.