A marked Cox model for the number of IBNR claims: Theory

CrossMark

Andrei L. Badescu, X. Sheldon Lin *, Dameng Tang

Department of Statistical Sciences, University of Toronto, 100 St. George Street, Toronto, Ontario, Canada M5S 3G3

ABSTRACT

Incurred but not reported (IBNR) loss reserving is an important issue for Property & Casualty (P&C) insurers. To calculate IBNR reserve, one needs to model claim arrivals and then predict IBNR claims. However, factors such as temporal dependence among claim arrivals and environmental variation are often not incorporated in many of the current loss reserving models, which may greatly affect the accuracy of IBNR predictions.

In this paper, we propose to model the claim arrival process together with its reporting delays as a marked Cox process. Our model is versatile in modeling temporal dependence, allowing also for natural interpretations. This paper focuses mainly on the theoretical aspects of the proposed model. We show that the associated reported claim process and IBNR claim process are both marked Cox processes with easily convertible intensity functions and marking distributions. The proposed model can also account for fluctuations in the exposure. By an order statistics property, we show that the corresponding discretely observed process preserves all the information about the claim arrivals. Finally, we derive closed-form expressions for both the autocorrelation function (ACF) and the distributions of the numbers of reported claims and IBNR claims. Model estimation and its applications are considered in a subsequent paper, Badescu et al. (2015b).

1. Introduction

Loss reserving is fundamental for insurance companies because it affects various aspects of the business, such as ratemaking, solvency control and capital allocation. With the advent of Solvency II, insurers are required to not only provide a best estimate of their future liabilities, but also to have a better grasp of their uncertainty. For Property & Casualty (P&C) insurance companies, there are two important types of reserves, namely the incurred but not reported (IBNR) reserve and the reported but not settled (RBNS) reserve, which stem respectively from the potential time delays between the claim occurrence and its reporting time or between the reporting time and the settlement time. One potential benefit of separately estimating the RBNS reserve and the IBNR reserve is that the adequacy of the case reserves set by claim adjusters can be judged. This is especially important for actuaries to decide whether to include the case reserves for analysis or not (Friedland, 2010, page 14).

Although predicting the number of IBNR claims has only been investigated in a relatively few papers (see Jewell, 1989, 1990;

Zhao et al., 2009; Zhao and Zhou, 2010), the issue is of practical importance for several reasons. The number of IBNR claims can be used to calculate the claim frequency (the total number of reported and IBNR claims per exposure) and thus allows the incorporation of exposure information into loss reserving. In addition, estimating the claim frequency and the severity components separately makes it feasible to explicitly reflect inflation adjustments for the severity (Friedland, 2010, page 205) and to stabilize the uncertainty in projecting the ultimate claim amounts, especially for the most recent accident years (Friedland, 2010, page 212). In this paper we focus on modeling and predicting the number of IBNR claims. Nevertheless, our approach can be easily extended to incorporate claim amounts as well, as to be described in Section 7.

The existing practice to perform the loss reserving task is based on the so-called "run-off triangle" which sums up the claim data per combination of accident year and development year. Various deterministic algorithms, such as the chain-ladder (CL) method, the Bornhuetter–Ferguson (BF) method and the frequency-severity method, can be chosen to apply to this triangular data (see Friedland, 2010 from a practitioners' perspective). Meanwhile, there is a whole array of stochastic models, so called "macro-level" models, aiming to interpret these methods and to analyze the uncertainty of their results (Wuthrich and Merz, 2008, 2015). Due to the limited number of data points contained in the triangular structure, these

* Corresponding author.
E-mail address: sheldon@utstat.utoronto.ca (X.S. Lin).

macro-level models tend to be over-parameterized and thus produce unstable estimates (Verdonck et al., 2009). Furthermore, they cannot separately estimate the RBNS reserve and the IBNR reserve without some further granulation of the current data (Schnieper, 1991; Liu and Verrall, 2009), or the inclusion of new data such as the numbers of reported claims (Verrall et al., 2010). To resolve these two issues and other shortcomings, a class of "micro-level" models has emerged to use policy-level data to depict the development of individual claims. Norberg (1993a, 1999) proposed a marked nonhomogeneous Poisson model and a general mathematical framework for predicting IBNR claims and reserve calculation. The estimation of the model was considered in Norberg (1993b). A case study based on Norberg's model using a liability portfolio was presented in Antonio and Plat (2014). Several papers have demonstrated the advantages of this type of micro-level models over the macro-level models through case studies or simulation experiments (e.g., Jin and Frees, 2013; Huang et al., 2015). It is noted that the separate estimation of the RBNS reserve and the IBNR reserve is natural in micro-level models because reporting and settlement delays are explicitly modeled.

Naturally, predicting the IBNR reserve requires the modeling of the claim arrival process (e.g., Jewell, 1989; Norberg, 1993a). In actuarial science, a popular model for this purpose is the nonhomogeneous Poisson process (NHPP), which, to the best of our knowledge, has been used in all the micro-level loss reserving models. However, this aspect of modeling loss reserving may be improved based on several considerations. A NHPP might not be a reasonable approximation for the claim arrival process of a given portfolio when there exists dependence among the individual claim arrivals due to some environmental variations that affect the whole portfolio (Grandell, 1991). A Cox process is a more appropriate model in such situations (see the second paragraph of Section 2 for more details). In addition, modeling the claim arrival process as a NHPP implies independence among the numbers of claims from different accident years, which contradicts the calendar year effect exhibited in the run-off triangle (Holmberg, 1994; Shi et al., 2012; Wuthrich and Merz, 2015). Finally, Mikosch (2009, Section 2.1.7) held an empirical study of the NHPP by using the arrival times in the well-publicized Danish fire insurance data. The distribution of the claim inter-arrival times was found to be different from which was implied by a NHPP.

The afore-mentioned issues may be dealt with by incorporating a temporal dependence structure into a model for the claim arrival process. In the loss reserving context, this direction has only been taken up in a few papers recently. Shi et al. (2012) and Merz et al. (2013) respectively imposed additive and multiplicative structure upon a Bayesian Gaussian copula to model both calendar year and accident year dependence. Although lacking flexibility in modeling dependence, their approaches lead to an analytic formula for the cumulative claim amounts. Avanzi et al. (2015) used the shot noise Cox process to model the claim arrival process, whose estimation is dealt with by using a tailor-made reversible jump Markov Chain Monte Carlo (RJMCMC).

In this paper, we propose to model the claim arrival process together with reporting delays as a marked Cox process. The intensity function of the process will be a piecewise stochastic process generated by a hidden Markov model (HMM) with Erlang state-dependent distributions. The proposed model allows for the fluctuation of the exposure over time. As a Cox process, the proposed model shares a similar interpretation as that of Markov-modulated Poisson process (MMPP). On the other hand, our model is different from MMPP because its underlying Markov process is discrete-time based and its piecewise intensity function consists of random variables instead of constants. Under our model assumption, the associated reported claim process and the IBNR claim process are both marked Cox processes with

easily convertible intensity functions and marking distributions. We also derive an analytical formula for the number of reported claims and for the number of IBNR claims. Furthermore, the associated discretely observed process of the proposed model is a Pascal-HMM. Using an order statistics property for the proposed model, we show that this discrete-time model preserves all the information about the claim arrivals. The joint distribution of the discretely observed process is a multivariate Pascal mixture, which is known to be extremely flexible in modeling dependence (Badescu et al., 2015a). A closed-form expression for the ACF of the discretely observed process is also obtained in steady state.

In a subsequent paper (Badescu et al., 2015b), we will develop a fitting procedure to estimate all the model parameters including the number of states therein and the corresponding transition probabilities. The efficiency of the fitting procedure is illustrated through simulation studies. We also fit the model to a real insurance data set, which obtains very good results.

The paper is structured as follows. In Section 2 we describe the proposed model for the portfolio claim arrival process and discuss how it may be interpreted and justified. The associated reported claim process and IBNR claim process are presented and discussed in Section 3. The corresponding discretely observed processes along with their desirable properties are presented in Section 4. An order statistics property of the model is obtained in Section 5 and is used to explain how the information from the claim arrival process is preserved. The distributions for the number of reported claims and the number of IBNR claims are derived in Section 6. Finally, we provide some concluding remarks along with several directions for future research in Section 7.

2. Model set-up and description

Following the notation of Norberg (1993a), suppose that the development of a claim until its reporting time is described as a pair of random variables (T, U), where T is the claim arrival time and U is its reporting delay. In chronology of their arrivals, $\{(T_i, U_i), i = 1, 2, \ldots, \}$ constitute the claim arrival process of the portfolio along with reporting delays. We denote the number of claims process by $\{N^a(t), t \geq 0\}$, where $N^a(t)$ is the number of claims up to time t. However, at a given valuation date τ, we are aware of a claim that has occurred only if it has been reported. Consequently, one cannot fully observe the claim arrival process up to time τ: $\{N^a(t), 0 \leq t \leq \tau\}$.

As stated in Grandell (1991) and also described in Norberg (1993a), a Cox process is a natural choice for modeling risk fluctuations exhibited in a portfolio claim arrival process. Depending on the line of business being considered, the stochastic intensity function can be interpreted as variations in an appropriate environment over time. For example, this environment may include weather conditions in automobile insurance. The environmental variation will affect every policy in the portfolio and all policies are assumed to be independent conditional on the environmental variation. Since the sum of a large number of independent and sparse point processes is approximately a Poisson process (Grigelionis, 1963; Grandell, 1991) argued that the portfolio claim arrival process, being unconditional on environmental variation, can be reasonably modeled as a Cox process.

We here propose to model $\{N^a(t), t \geq 0\}$ as a marked Cox process that is described through two components. First, the marks $\{U_1, U_2, \ldots\}$ are independent random variables with common density function $p_U(u)$ and cumulative distribution function $P_U(u)$. Second, the stochastic intensity function $\Lambda(t)$ is a piecewise stochastic process: $\Lambda(t) = \Lambda_l$, for $d_{l-1} \leq t < d_l$, $l = 1, 2, \ldots$ and $d_0 = 0$. Here $d_l, l = 1, 2, \ldots$, are pre-determined time points. In practice, data collection cannot be conducted in continuous time and these time points may thus be interpreted as data collecting

times. As a result, we have a continuous stochastic process with discrete observation times. We assume that $\{\Lambda_1, \Lambda_2, \ldots\}$ is generated by an Erlang hidden Markov model (Erlang-HMM) with the following structure:

- *The hidden parameter process* $\{C_1, C_2, \ldots\}$ *is a time-homogeneous Markov chain with a finite state space* $\{1, 2, \ldots, g\}$. Its initial distribution and transition probability matrix are respectively denoted by row vector $\boldsymbol{\pi}_1$ and matrix $\boldsymbol{\Gamma} = (\gamma_{ij})_{g \times g}$, where $\gamma_{ij} = P(C_l = j | C_{l-1} = i)$. We assume that the Markov chain is irreducible, aperiodic and all the states are positive recurrent. We denote the limiting distribution of the Markov chain by $\boldsymbol{\delta}$. The assumptions are very natural from a modeling perspective: each state will be revisited infinitely many times over time; the time between two consecutive visits to the same state is irregular; and the mean time is finite.

- *The state-dependent process* $\{\Lambda_1, \Lambda_2, \ldots\}$ *is defined such that each* Λ_l *depends only on the current state* C_l. Given that $C_l = i$, we assume that Λ_l follows an Erlang distribution with shape parameter m_i and scale parameter $\omega_l \theta$ whose density function is given by

$$f_{\Lambda_l | C_l = i}(\lambda) = \frac{\lambda^{m_i - 1} e^{-\frac{\lambda}{\omega_l \theta}}}{(\omega_l \theta)^{m_i}(m_i - 1)!} \triangleq f(\lambda; m_i, \omega_l \theta), \quad (2.1)$$

where ω_l represents the risk exposure of the considered portfolio for the lth period.

If a Poisson process is used to model the portfolio claim arrival process, then the risk exposure is usually incorporated as a multiplicative factor into the intensity function (see e.g., Grandell, 1991; Norberg, 1993a). Due to the scaling property of the class of Erlang distributions, (2.1) is indeed a reasonable way to reflect the fluctuation of the risk exposure over time for the claim arrival process.

The mixed Poisson process and the Ammeter process (Ammeter, 1948) are two commonly used classes of Cox processes, both of them being special cases of the proposed model. A mixed Poisson process is a Cox process with intensity process $\Lambda(t) \equiv \Lambda$, a single random variable. Our model reduces to a mixed Poisson process if $d_1 = \infty$. An Ammeter process is a Cox process with $\Lambda(t) = \Lambda_l$, $d_{l-1} \leq t < d_l$, where $\{\Lambda_1, \Lambda_2, \ldots\}$ is a sequence of i.i.d. random variables. When the initial distribution $\boldsymbol{\pi}_1$ is the limiting distribution $\boldsymbol{\delta}$, each row of $\boldsymbol{\Gamma}$ is $\boldsymbol{\delta}$ and there are no exposure fluctuation $\omega_l \equiv \omega$, our model degenerates to an Ammeter process.

With the above model specifications, one may identify the density function of each Λ_l as follows.

Proposition 2.1. *For the piecewise stochastic intensity function following the Erlang-HMM structure described above,* Λ_l *is an Erlang mixture with density function*

$$f_{\Lambda_l}(\lambda) = \sum_{i=1}^{g} \pi_{li} f(\lambda; m_i, \omega_l \theta), \quad l = 1, 2, \ldots, \quad (2.2)$$

where $\pi_{li} = P(C_l = i)$ *is the ith element of the row vector* $\boldsymbol{\pi}_l = \boldsymbol{\pi}_1 \boldsymbol{\Gamma}^{l-1}$.

Proof. It follows immediately from the law of total probability and the Chapman–Kolmogorov equation. □

Remark 2.2. Since the class of Erlang mixtures is dense in the space of positive continuous distributions (Tijms, 1994), one might think of extending the state-dependent density in (2.1) to that of mixed Erlang. However, the resulting Λ_l is still Erlang mixture distributed (Klugman et al., 2013), which implies that (2.1) is sufficient for constructing a flexible model of the piecewise stochastic intensity function.

3. Reported claim process and IBNR claim process

As stated in Section 2, we are aware of a claim that has occurred only if it has been reported. As a result, even though we specify the proposed model based on the claim arrival process, it can only be estimated through its associated reported claim process. The IBNR claim process plays a key role in predicting the number of IBNR claims. In this section, we show that both processes are still marked Cox processes and we identify their intensity functions. Although both the reported claim process and the IBNR claim process should be indexed with their corresponding valuation date τ, for notational simplicity we drop this index hereafter.

We denote the reported claim process with respect to valuation date τ as $\{N^r(t), 0 \leq t \leq \tau\}$. This process comprises those marked points from $\{(T_i, U_i), i = 1, 2, \ldots\}$ which satisfies the condition $T_i + U_i \leq \tau$. The total number of reported claims is then $N^r(\tau)$. When ordered in chronology of their arrival times, the selected marked points are denoted by $\{(T_i^r, U_i^r), i = 1, 2, \ldots, N^r(\tau)\}$. In a similar way, we can define the IBNR claim process with respect to valuation date τ, which is denoted by $\{N^{IBNR}(t), 0 \leq t \leq \tau\}$ and consists of marked points $\{(T_i^{IBNR}, U_i^{IBNR}), i = 1, 2, \ldots, N^{IBNR}(\tau)\}$, where $N^{IBNR}(\tau)$ is the total number of IBNR claims. It is noted that $\{N^r(t), 0 \leq t \leq \tau\}$ is observable while $\{N^{IBNR}(t), 0 \leq t \leq \tau\}$ is not.

In the following, we first prove that the marked Cox processes are closed under thinning. More discussions about the thinning operation of point processes can be found in Grandell (1997). For this, we will mainly use the tool of Laplace functional transform (LFT). For a point process $N = \{X_i, i = 1, 2, \ldots\}$, its LFT is given by

$$L_N(f(x)) = E(e^{-\sum_i f(X_i)}),$$

where $f(x)$ is a nonnegative function of the point process. Just like the fact that the Laplace transforms have a one-to-one correspondence with random variables, the mapping between LFTs and point processes is also one-to-one. For more properties about LFT, see Mikosch (2009).

Theorem 3.1. *Assume that* $\bar{N}$ *is a marked Cox process on* $[0, \infty)$ *with a marking space of* R^d. *Its intensity function is* $\Lambda(t)$ *and its marks* $\boldsymbol{Z}_i$, $i = 1, 2, \ldots$, *are independent but position-dependent with density function* $p_{\boldsymbol{Z}|t}(\boldsymbol{z})$. *Now consider the following thinning probabilities:*

$$p(t, \boldsymbol{z}) = \begin{cases} 1 & \text{if } (t, \boldsymbol{z}) \in D, \\ 0 & \text{if } (t, \boldsymbol{z}) \notin D, \end{cases} \quad (3.1)$$

where D *is a subset of* $[0, \infty) \times R^d$. *Then the resulting thinned point process* $\bar{N}_p$ *is still a marked Cox process with intensity function* $\Lambda(t)P(Z \in D_t)I_{\{t \in T_D\}}$ *and independent yet position-dependent marks having density function* $\frac{p_{\boldsymbol{Z}|t}(\boldsymbol{z})}{P(\boldsymbol{Z} \in D_t)}I_{\{\boldsymbol{z} \in D_t\}}$, *where* $D_t = \{\boldsymbol{z} \in R^d | (t, \boldsymbol{z}) \in D\}$ *and* $T_D = \{t \in [0, \infty) | \exists \boldsymbol{z} \text{ s.t. } (t, \boldsymbol{z}) \in D\}$.

Proof. By Eq. (1.33) in Karr (1991), for any nonnegative function $f(t, \boldsymbol{z})$ on $[0, \infty) \times R^d$, the LFT of $\bar{N}_p$ equals $L_{\bar{N}}\left(-\log\left(1 - p(t, \boldsymbol{z}) + p(t, \boldsymbol{z})e^{-f(t, \boldsymbol{z})}\right)\right)$. Combining the results in Examples 1.16 and 1.28 in Karr (1991), this can be calculated as $E\left(e^{-\int_0^\infty \left(1 - \int_{R^d}(1 - p(t, \boldsymbol{z}) + p(t, \boldsymbol{z})e^{-f(t, \boldsymbol{z})})p_{\boldsymbol{Z}|t}(\boldsymbol{z})d\boldsymbol{z}\right)\Lambda(t)dt}\right)$.

Since the thinning probabilities are given in (3.1), the previous equation can be further simplified as given in Box I.

Again it follows from Examples 1.16 and 1.28 in Karr (1991) that $\bar{N}_p$ is a marked Cox process with intensity function $\Lambda(t)P(Z \in D_t)I_{\{t \in T_D\}}$ and position-dependent marks having density functions $\frac{p_{\boldsymbol{Z}|t}(\boldsymbol{z})}{P(\boldsymbol{Z} \in D_t)}I_{\{\boldsymbol{z} \in D_t\}}$. □

$$E\left(e^{-\int_{T_D}\left(1-\int_{D_t}\left(1-p(t,\mathbf{z})+p(t,\mathbf{z})e^{-f(t,\mathbf{z})}\right)p_{\mathbf{Z}|t}(\mathbf{z})d\mathbf{z}-\int_{D_t^c}\left(1-p(t,\mathbf{z})+p(t,\mathbf{z})e^{-f(t,\mathbf{z})}\right)p_{\mathbf{Z}|t}(\mathbf{z})d\mathbf{z}\right)\Lambda(t)dt}\right)$$

$$=E\left(e^{-\int_{T_D}\left(1-\int_{D_t}e^{-f(t,\mathbf{z})}p_{\mathbf{Z}|t}(\mathbf{z})d\mathbf{z}-\int_{D_t^c}p_{\mathbf{Z}|t}(\mathbf{z})d\mathbf{z}\right)\Lambda(t)dt}\right)$$

$$=E\left(e^{-\int_{T_D}\left(\int_{D_t}p_{\mathbf{Z}|t}(\mathbf{z})d\mathbf{z}-\int_{D_t}e^{-f(t,\mathbf{z})}p_{\mathbf{Z}|t}(\mathbf{z})d\mathbf{z}\right)\Lambda(t)dt}\right)$$

$$=E\left(e^{-\int_{T_D}\left(P(\mathbf{Z}\in D_t)-\int_{D_t}e^{-f(t,\mathbf{z})}\frac{p_{\mathbf{Z}|t}(\mathbf{z})}{P(\mathbf{Z}\in D_t)}d\mathbf{z}P(\mathbf{Z}\in D_t)\right)\Lambda(t)dt}\right)$$

$$=E\left(e^{-\int_{T_D}\left(1-\int_{D_t}e^{-f(t,\mathbf{z})}\frac{p_{\mathbf{Z}|t}(\mathbf{z})}{P(\mathbf{Z}\in D_t)}d\mathbf{z}\right)\Lambda(t)P(\mathbf{Z}\in D_t)dt}\right)$$

$$=E\left(e^{-\int_0^\infty\left(1-\int_{R^d}e^{-f(t,\mathbf{z})}\frac{p_{\mathbf{Z}|t}(\mathbf{z})}{P(\mathbf{Z}\in D_t)}I_{\{\mathbf{z}\in D_t\}}d\mathbf{z}\right)\Lambda(t)P(\mathbf{Z}\in D_t)I_{\{t\in T_D\}}dt}\right).$$

Box I.

We now show that if the claim arrival process $\{N^a(t), t \geq 0\}$ is a marked Cox process, then both the reported claim process $\{N^r(t), 0 \leq t \leq \tau\}$ and IBNR claim process $\{N^{IBNR}(t), 0 \leq t \leq \tau\}$ are still marked Cox processes with easily convertible stochastic intensity functions and mark densities.

Theorem 3.2. *Assume that the claim arrival process $\{N^a(t), t \geq 0\}$ is a marked Cox process with stochastic intensity function $\Lambda(t)$ and independent marks $\{U_i, i = 1, 2, \ldots\}$ following common density function $p_U(u)$. Then for a given valuation date τ, its associated reported claim process $\{N^r(t), 0 \leq t \leq \tau\}$ and IBNR claim process $\{N^{IBNR}(t), 0 \leq t \leq \tau\}$ are also marked Cox processes. Their adjusted stochastic intensity functions are $\Lambda^r(t) = \Lambda(t)P_U(\tau - t)I_{\{0 \leq t \leq \tau\}}$ and $\Lambda^{IBNR}(t) = \Lambda(t)(1 - P_U(\tau - t))I_{\{0 \leq t \leq \tau\}}$, respectively, and their independent marks follow adjusted position-dependent mark density functions $p^r_{U|t}(u) = \frac{p_U(u)}{P_U(\tau-t)}I_{\{0 \leq u \leq \tau-t\}}$ and $p^{IBNR}_{U|t}(u) = \frac{p_U(u)}{1-P_U(\tau-t)}I_{\{u \geq \tau-t\}}$, respectively.*

Proof. For a given valuation date τ, the reported claim process $\{N^r(t), 0 \leq t \leq \tau\}$ is a thinned point process of $\{N^a(t), t \geq 0\}$ by thinning probabilities

$$p(t, u) = \begin{cases} 1 & \text{if } t + u \leq \tau, \\ 0 & \text{if } t + u > \tau. \end{cases}$$

By Theorem 3.1, $\{N^r(t), 0 \leq t \leq \tau\}$ is again a marked Cox process, whose adjusted stochastic intensity and mark density are as given. For the IBNR claim process $\{N^{IBNR}(t), 0 \leq t \leq \tau\}$, the argument is similar except that one chooses $D = \{(t, u)|t + u > \tau\}$. $\square$

Theorem 3.2 shows that not only are both the reported claim process and the IBNR claim process marked Cox processes, but also that their intensity processes and mark densities have intuitive interpretations. For example, the intensity for the reported claim process $\Lambda^r(t)$ is the original intensity $\Lambda(t)$ times the probability that the claim has been reported by time τ. The mark density $p^r_{U|t}(u)$ is the original density $p_U(u)$ conditional on the fact that the claim has been reported by time τ.

4. The corresponding discretely observed processes

Data is usually aggregated in some discrete form before analysis in practice. As described in Section 2, d_l, $l = 0, 1, \ldots$, may be interpreted as data collecting times. Hence, the number of observable claims are $\{N_1, N_2, \ldots\}$, where N_l is the number of claims that arrived during $[d_{l-1}, d_l)$, no matter reported or not.

The argument can be extended to the reported claims and the IBNR claims. In this section we characterize the three discretely observed processes, which not only are important in their own right, but also play a critical role in fitting the model to data.

We begin with the discretely observed claim process. The following proposition shows that under the model assumptions, this discretely observed process follows a Pascal-HMM.

Proposition 4.1. *With the proposed Cox process $\{N^a(t), t \geq 0\}$, its discretely observed process $\{N_1, N_2, \ldots\}$ follows a Pascal-HMM. Its hidden parameter process is given by $\{C_1, C_2, \ldots\}$. For $l = 1, 2, \ldots$; $i = 1, 2, \ldots, g$, its state-dependent distribution is Pascal with*

$$P(N_l = n|C_l = i) = p\left(n; m_i, (d_l - d_{l-1})\omega_l\theta\right), \tag{4.1}$$

where

$$p(n; m, \theta) = \binom{n+m-1}{m-1}\left(\frac{1}{1+\theta}\right)^m\left(\frac{\theta}{1+\theta}\right)^n. \tag{4.2}$$

Proof. Obviously the hidden parameter process is kept the same and we only need to check that the state-dependent distributions are Pascal. Using the law of total probability, we obtain

$$P(N_l = n|C_l = i) = \int_0^\infty P(N_l = n|\Lambda_l = \lambda, C_l = i)f_{\Lambda_l|C_l=i}(\lambda)d\lambda$$

$$= \int_0^\infty \frac{((d_l - d_{l-1})\lambda)^n e^{-(d_l-d_{l-1})\lambda}}{n!}f(\lambda; m_i, \omega_l\theta)d\lambda$$

$$= p(n; m_i, (d_l - d_{l-1})\omega_l\theta).$$

Thus $\{N_1, N_2, \ldots\}$ follows the given Pascal-HMM. $\square$

Without loss of generality, we assume that $\tau = d_k$. We denote the discretely observed reported claim process as $\{N_1^r, \ldots, N_k^r\}$, where N_l^r is the number of claims arrived during $[d_{l-1}, d_l)$ and reported by time τ. Similarly, $\{N_1^{IBNR}, \ldots, N_k^{IBNR}\}$ denotes the discretely observed IBNR claim process, where N_l^{IBNR} is the number of the claims arrived during $[d_{l-1}, d_l)$ but have not yet reported at time τ. Using similar arguments in the derivations of Proposition 4.1 and Theorem 3.2, it is straightforward to show that both discretely observed processes are also Pascal-HMM.

Corollary 4.2. *For the proposed claim arrival process, the discretely observed processes of the reported claim process and the IBNR claim*

process are both Pascal-HMMs. They share the same hidden parameter process $\{C_1, \ldots, C_k\}$ and their state-dependent distributions are

$$P(N_l^r = n | C_l = i) = p\left(n; m_i, \left(\int_{d_{l-1}}^{d_l} P_U(\tau - t) dt\right) \omega_l \theta\right)$$

and

$$P(N_l^{IBNR} = n | C_l = i)$$
$$= p\left(n; m_i, \left(\int_{d_{l-1}}^{d_l} (1 - P_U(\tau - t)) dt\right) \omega_l \theta\right),$$

where $l = 1, 2, \ldots, k$; $i = 1, 2, \ldots, g$.

As a direct consequence, the univariate and multivariate marginal distributions of $\{N_1, N_2, \ldots\}$ can be described below. We denote the k-step transition probability matrix by $\mathbf{\Gamma}^k = \left(\gamma_{ij}(k)\right)_{g \times g}$, that is, $\gamma_{ij}(k) = P(C_{l+k} = j | C_l = i)$.

Corollary 4.3. *For $l = 1, 2, \ldots, N_l$ follows a mixed Pascal distribution with probability function*

$$P(N_l = n) = \sum_{i=1}^{g} \pi_{li} p\left(n; m_i, (d_l - d_{l-1}) \omega_l \theta\right). \tag{4.3}$$

Proof. Can be proved in a similar way to the proof of Proposition 2.1. $\square$

Corollary 4.4. *The k-variate joint distribution of $(N_{l_1}, \ldots, N_{l_k})$ is*

$$P(N_{l_1} = n_{l_1}, \ldots, N_{l_k} = n_{l_k})$$
$$= \sum_{i_1=1}^{g} \cdots \sum_{i_k=1}^{g} \beta_{(i_1, \ldots, i_k)} \prod_{j=1}^{k} p(n_{l_j}; m_{i_j}, (d_{l_j} - d_{l_j-1}) \omega_{l_j} \theta), \tag{4.4}$$

where

$$\beta_{(i_1, \ldots, i_k)} = \pi_{l_1, i_1} \gamma_{i_1, i_2}(l_2 - l_1) \cdots \gamma_{i_{k-1}, i_k}(l_k - l_{k-1}). \tag{4.5}$$

Proof. We only prove the bivariate case and a similar argument can be applied to any higher dimension. For $l_1 < l_2$, we have

$$P(N_{l_1} = n_{l_1}, N_{l_2} = n_{l_2})$$
$$= \sum_{i_1=1}^{g} \sum_{i_2=1}^{g} P(N_{l_1} = n_{l_1}, N_{l_2} = n_{l_2}, C_{l_1} = i_1, C_{l_2} = i_2)$$
$$= \sum_{i_1=1}^{g} \sum_{i_2=1}^{g} P(C_{l_1} = i_1, C_{l_2} = i_2)$$
$$\quad \times P(N_{l_1} = n_{l_1}, N_{l_2} = n_{l_2} | C_{l_1} = i_1, C_{l_2} = i_2)$$
$$= \sum_{i_1=1}^{g} \sum_{i_2=1}^{g} P(C_{l_1} = i_1) P(C_{l_2} = i_2 | C_{l_1} = i_1)$$
$$\quad \times P(N_{l_1} = n_{l_1} | C_{l_1} = i_1) P(N_{l_2} = n_{l_2} | C_{l_2} = i_2)$$
$$= \sum_{i_1=1}^{g} \sum_{i_2=1}^{g} \pi_{l_1, i_1} \gamma_{i_1, i_2}(l_2 - l_1)$$
$$\quad \times \prod_{j=1}^{2} p(n_{l_j}; m_{i_j}, (d_{l_j} - d_{l_j-1}) \omega_{l_j} \theta)$$
$$= \sum_{i_1=1}^{g} \sum_{i_2=1}^{g} \beta_{(i_1, i_2)} \prod_{j=1}^{2} p(n_{l_j}; m_{i_j}, (d_{l_j} - d_{l_j-1}) \omega_{l_j} \theta). \quad \square$$

Corollary 4.4 shows that k-variate marginal of the discrete observations of the proposed model is a multivariate Pascal mixture, which is known to constitute a versatile class of discrete multivariate distributions (Badescu et al., 2015a). Consequently the proposed Pascal-HMM provides great flexibility in modeling temporal dependence. In The following, we provide a closed-form expression for the auto-correlation function (ACF) of $\{N_1, N_2, \ldots\}$.

Theorem 4.5. *If we further assume that:*

1. *$(d_l - d_{l-1}) \omega_l = 1$, $l = 1, 2, \ldots$,*
2. *the eigenvalues of $\mathbf{\Gamma}$ are all distinct and they are ordered as $1 = e_1 > e_2 > \cdots > e_g > -1$,*

then the ACF (in the limiting sense) for $\{N_1, N_2, \ldots\}$ is given by

$$\rho(k) = \frac{Cov(N_l, N_{l+k})}{Var(N_l)} = \sum_{i=2}^{g} c_i e_i^k, \quad k = 1, 2, \ldots, \tag{4.6}$$

where

$$c_i = \frac{\delta \mathbf{M} \mathbf{u}_i^T \mathbf{v}_i \mathbf{M} \mathbf{1}^T}{\sum_{i=1}^{g} \delta_i m_i \left(m_i + \frac{1+\theta}{\theta}\right) - \left(\sum_{i=1}^{g} \delta_i m_i\right)^2}.$$

Here $\mathbf{M} = diag\{m_1, \ldots, m_g\}$, $\mathbf{1} = (1, \ldots, 1)$, $\mathbf{v}_i$ and $\mathbf{u}_i^T$ are the left and right eigenvectors of $\mathbf{\Gamma}$ associated with e_i, and $\mathbf{v}_i \mathbf{u}_i^T = 1$.

Remark 4.6. The first assumption ensures that, in the limiting sense, $\{N_1, N_2, \ldots\}$ is a stationary time series, which means that the ACF is sufficient for describing its temporal dependence in the long run. Since one can adjust the exposure by scaling, this assumption essentially means that $(d_l - d_{l-1}) \omega_l$ is irrelevant to l, which can be realized if we narrow/widen the lengths of periods when portfolio exposure is high/low. The second assumption is not restrictive in applications as matrices with distinct eigenvalues are dense in the matrix space.

Proof. According to the Perron–Frobenius Theorem (e.g., see Bremaud, 2013), we have the following decomposition

$$\mathbf{\Gamma}^k = \sum_{i=1}^{g} e_i^k \mathbf{u}_i^T \mathbf{v}_i,$$

where $1 = e_1 > \cdots > e_g > -1$ are the distinct eigenvalues of $\mathbf{\Gamma}$ and $\mathbf{v}_i(\mathbf{u}_i^T)$ are the left(right) eigenvectors of $\mathbf{\Gamma}$ associated with e_i, satisfying $\mathbf{v}_i \mathbf{u}_i^T = 1$. Since all the eigenvalues are distinct, all the left and right eigenvectors are determined up to multiplication by a non-zero scalar. In particular, we have $\mathbf{u}_1 = c_1(1, \ldots, 1)$ and $\mathbf{v}_1 = c_2(\delta_1, \ldots, \delta_g)$, where $c_1 c_2 = 1$ due to the constraint that $\mathbf{v}_i \mathbf{u}_i^T = 1$.

In a similar way to the proof of Corollary 4.4, one may check that

$$Cov(N_l, N_{l+k})$$
$$= E(N_l N_{l+k}) - E(N_l)^2$$
$$= \sum_{i=1}^{g} \sum_{j=1}^{g} \delta_i \gamma_{ij}(k) m_i \theta m_j \theta - \left(\sum_{i=1}^{g} \delta_i m_i \theta\right)^2$$
$$= \theta^2 \left(\delta \begin{pmatrix} m_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & m_g \end{pmatrix} \mathbf{\Gamma}^k \begin{pmatrix} m_1 \\ \vdots \\ m_g \end{pmatrix} - \left(\sum_{i=1}^{g} \delta_i m_i\right)^2\right)$$
$$= \theta^2 \left(\delta \begin{pmatrix} m_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & m_g \end{pmatrix} \left(\sum_{i=1}^{g} e_i^k \mathbf{u}_i^T \mathbf{v}_i\right) \begin{pmatrix} m_1 \\ \vdots \\ m_g \end{pmatrix}\right.$$

$$- \left(\sum_{i=1}^{g} \delta_i m_i \right)^2 \right)$$

$$= \theta^2 \delta \begin{pmatrix} m_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & m_g \end{pmatrix} \boldsymbol{u}_1^T \boldsymbol{v}_1 \begin{pmatrix} m_1 \\ \vdots \\ m_g \end{pmatrix}$$

$$+ \theta^2 \sum_{i=2}^{g} \delta \begin{pmatrix} m_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & m_g \end{pmatrix} \boldsymbol{u}_i^T \boldsymbol{v}_i \begin{pmatrix} m_1 \\ \vdots \\ m_g \end{pmatrix} e_i^k - \theta^2 \left(\sum_{i=1}^{g} \delta_i m_i \right)^2$$

$$= \theta^2 \left(\left(\sum_{i=1}^{g} \delta_i m_i \right)^2 + \sum_{i=2}^{g} \delta \begin{pmatrix} m_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & m_g \end{pmatrix} \right.$$

$$\times \boldsymbol{u}_i^T \boldsymbol{v}_i \begin{pmatrix} m_1 \\ \vdots \\ m_g \end{pmatrix} e_i^k - \left(\sum_{i=1}^{g} \delta_i m_i \right)^2 \right)$$

$$= \theta^2 \sum_{i=2}^{g} \delta \begin{pmatrix} m_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & m_g \end{pmatrix} \boldsymbol{u}_i^T \boldsymbol{v}_i \begin{pmatrix} m_1 \\ \vdots \\ m_g \end{pmatrix} e_i^k$$

$$= \theta^2 \sum_{i=2}^{g} \delta \boldsymbol{M} \boldsymbol{u}_i^T \boldsymbol{v}_i \boldsymbol{M} \boldsymbol{1}^T e_i^k.$$

The result follows by plugging

$$Var(N) = \theta^2 \left(\sum_{i=1}^{g} \delta_i m_i \left(m_i + \frac{1+\theta}{\theta} \right) - \left(\sum_{i=1}^{g} \delta_i m_i \right)^2 \right)$$

into the expression for $\rho(k)$. $\square$

Theorem 4.5 reveals that the ACF exhibits a power decaying pattern, which is similar to that of popular autoregressive integrated moving average (ARIMA) models. As a result, the proposed Pascal-HMM can achieve a wide range of temporal dependence structures by judiciously choosing its parameters. Below, we illustrate with a toy example that both $\rho(k) = 1$ and $\rho(k) = -1$ can be achieved. For this we choose $\boldsymbol{\Gamma} = \begin{pmatrix} \gamma & 1-\gamma \\ 1-\gamma & \gamma \end{pmatrix}$. It is easy to check that $\boldsymbol{\delta} = (0.5, 0.5)$, $e_2 = 2\gamma - 1$, $\boldsymbol{u}_2 = (1 - \gamma, \gamma - 1)$ and $\boldsymbol{v}_2 = \left(\frac{1}{2-2\gamma}, \frac{-1}{2-2\gamma} \right)$. One can calculate that

$$c_2 = \frac{\frac{1}{4}(m_1 - m_2)^2}{\frac{1}{4}(m_1 - m_2)^2 + \frac{1+\theta}{2\theta}(m_1 + m_2)}.$$

If we take $m_1 = 3m_2$ and $\theta = 1$, c_2 equals to $\frac{m_2}{m_2+4}$, which has a limit of 1 when $m_2 \to \infty$. We also notice that $e_2 = 2\gamma - 1 \to 1$ if $\gamma \to 1$ and $e_2 \to -1$ if $\gamma \to 0$.

5. An order statistics property

When we compress the original data to its discretely observed process $\{N_1, N_2, \ldots\}$, one may be concerned about any potential loss of information. We again note that here N_l includes all the claims occurred during $[d_{l-1}, d_l]$, no matter reported or not. In Theorem 5.2 we show that a well-known order statistics property of the Poisson process holds in this more general case as well. This shows that the discretely observed process preserves all the information about the claim arrival epochs. First we present an intermediate result in Lemma 5.1, which can be easily checked using the independent increment property of the Poisson process.

Lemma 5.1. *Assume that $\{N^a(t)\}$ is a marked Poisson process with intensity function $\lambda(t)$ and path-dependent mark U with density function $p_{U|t}(u)$. The likelihood for the observations up to a given time t is*

$$P(N^a(t) = n, (T_i, U_i) \in (dt_i, du_i), i = 1, \ldots, n)$$

$$= e^{-\int_0^t \lambda(s)ds} \prod_{i=1}^{n} \left(\lambda(t_i)dt_i \, p_{U_i|t_i}(du_i) \right).$$

Theorem 5.2. *For $l = 1, \ldots, k$, we assume that there are n_l claims that have occurred during the period $[d_{l-1}, d_l]$, together with their marks ordered chronologically according to the arrival epochs $\{(T_i^{(l)}, U_i^{(l)}), i = 1, 2, \ldots, n_l\}$. We then have*

$$P\big((T_i^{(l)}, U_i^{(l)}) \in (dt_i^{(l)}, du_i^{(l)}),$$

$$l = 1, \ldots, k, i = 1, \ldots, n_l | N_l = n_l, l = 1, \ldots, k\big)$$

$$= \prod_{l=1}^{k} \left(n_l! \prod_{i=1}^{n_l} \left(\frac{dt_i^{(l)}}{d_l - d_{l-1}} p_{U_i^{(l)}}(du_i^{(l)}) \right) \right). \tag{5.1}$$

In other words, given the discrete observations, the joint distribution of the claim arrival epochs with markings are mutually independent and the epochs are uniformly distributed.

Proof. Since when given $\Lambda(t) = \lambda(t)$, the proposed model $\{N^a(t)\}$ is a marked Poisson process with intensity function $\lambda(t)$, according to Lemma 5.1 the likelihood for the observations up to time d_k is

$$P\left(N_l = n_l, (T_i^{(l)}, U_i^{(l)}) \in (dt_i^{(l)}, du_i^{(l)}), l = 1, \ldots, k, i = 1, \ldots, n_l \right)$$

$$= E_{\Lambda_1, \ldots, \Lambda_k} \left(\prod_{l=1}^{k} \left(e^{-(d_l - d_{l-1})\Lambda_l} \prod_{i=1}^{n_l} \left(\Lambda_l dt_i^{(l)} p_{U_i^{(l)}}(du_i^{(l)}) \right) \right) \right),$$

where $E_{\Lambda_1, \ldots, \Lambda_k}$ signifies taking expectation with respect to $\Lambda_1, \ldots, \Lambda_k$. This can be further calculated as

$$\sum_{i_1=1}^{g} \cdots \sum_{i_k=1}^{g} P(C_1 = i_1, \ldots, C_k = i_k)$$

$$\cdot \int_0^{\infty} \cdots \int_0^{\infty} \left(\prod_{l=1}^{k} \left(e^{-(d_l - d_{l-1})\lambda_l} \prod_{i=1}^{n_l} \left(\lambda_l dt_i^{(l)} p_{U_i^{(l)}}(du_i^{(l)}) \right) \right) \right)$$

$$\cdot f_{\Lambda_1, \ldots, \Lambda_k | C_1 = i_1, \ldots, C_k = i_k}(\lambda_1, \ldots, \lambda_k) d\lambda_1 \ldots d\lambda_k$$

$$= \sum_{i_1=1}^{g} \cdots \sum_{i_k=1}^{g} P(C_1 = i_1, \ldots, C_k = i_k)$$

$$\cdot \prod_{l=1}^{k} \left(\int_0^{\infty} \lambda_l^{n_l} e^{-(d_l - d_{l-1})\lambda_l} f_{\Lambda_l | C_l = i_l}(\lambda_l) d\lambda_l \right)$$

$$\times \prod_{l=1}^{k} \prod_{i=1}^{n_l} \left(dt_i^{(l)} p_{U_i^{(l)}}(du_i^{(l)}) \right).$$

At the same time,

$$P(N_1 = n_1, \ldots, N_k = n_k)$$

$$= \sum_{i_1=1}^{g} \cdots \sum_{i_k=1}^{g} P(C_1 = i_1, \ldots, C_k = i_k) \prod_{l=1}^{k} P(N_l = n_l | C_l = i_l)$$

$$= \sum_{i_1=1}^{g} \cdots \sum_{i_k=1}^{g} P(C_1 = i_1, \ldots, C_k = i_k)$$

$$\cdot \prod_{l=1}^{k} \left(\int_0^{\infty} \frac{((d_l - d_{l-1})\lambda_l)^{n_l} e^{-(d_l - d_{l-1})\lambda_l}}{n_l!} f_{\Lambda_l | C_l = i_l}(\lambda_l) d\lambda_l \right).$$

Combining these two equations yields the result. $\square$

According to (5.1), with the discrete observations given, the joint distribution of the claim arrival times with markings is completely specified. Furthermore, this order statistics is critical in the estimation of the proposed model.

6. The distributions of the numbers of reported and IBNR claims

In this section we derive explicit analytical expressions for the distribution of the number of reported claims and the distribution of the number of IBNR claims.

Proposition 6.1. *Recall that N_l^r is the number of reported claims in time interval $[d_{l-1}, d_l)$. Then the joint probability of N_l^r, $l = 1, \ldots, k$, where $\tau = d_k$ is the valuation date, can explicitly be expressed as*

$$P(N_1^r = n_1, \ldots, N_k^r = n_k)$$
$$= \sum_{i_1=1}^{g} \cdots \sum_{i_k=1}^{g} \beta_{(i_1,\ldots,i_k)} \prod_{j=1}^{k} p(n_j; m_{i_j}, \theta_j^r), \tag{6.1}$$

where

$$\beta_{(i_1,\ldots,i_k)} = \pi_{1,i_1} \gamma_{i_1,i_2} \cdots \gamma_{i_{k-1},i_k}$$

and

$$\theta_j^r = \left(\int_{d_{j-1}}^{d_j} P_U(\tau - t) dt \right) \omega_j \theta, \quad j = 1, 2, \ldots, k.$$

Similarly, the joint probability of N_l^{IBNR}, $l = 1, \ldots, k$, the numbers of IBNR claims in the same time intervals up to valuation date $\tau = d_k$ can explicitly be expressed as

$$P(N_1^{IBNR} = n_1, \ldots, N_k^{IBNR} = n_k)$$
$$= \sum_{i_1=1}^{g} \cdots \sum_{i_k=1}^{g} \beta_{(i_1,\ldots,i_k)} \prod_{j=1}^{k} p(n_j; m_{i_j}, \theta_j^{IBNR}), \tag{6.2}$$

where

$$\theta_j^{IBNR} = \left(\int_{d_{j-1}}^{d_j} (1 - P_U(\tau - t)) \, dt \right) \omega_j \theta, \quad j = 1, 2, \ldots, k.$$

Proof. It can be easily checked using Corollaries 4.2 and 4.4. □

In the following, we show that the total number of reported claims and the total number of IBNR claims up to the valuation date have a univariate Pascal mixture with a common scale parameter. For notational simplicity, we re-expressed the joint probabilities (6.1) and (6.2) as infinite series with a finite number of non-zero coefficients:

$$P(N_1^r = n_1, \ldots, N_k^r = n_k)$$
$$= \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \beta_{(m_1,\ldots,m_k)} \prod_{j=1}^{k} p(n_j; m_j, \theta_j^r), \tag{6.3}$$

and

$$P(N_1^{IBNR} = n_1, \ldots, N_k^{IBNR} = n_k)$$
$$= \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \beta_{(m_1,\ldots,m_k)} \prod_{j=1}^{k} p(n_j; m_j, \theta_j^{IBNR}), \tag{6.4}$$

where $\beta_{(m_1,\ldots,m_k)} = 0$ if one of the integers $m_1, \ldots, m_k$ is not a shape parameter in the Pascal mixture (6.1). Denote $N^r = \sum_{l=1}^{k} N_l^r$, the total number of reported claims up to the valuation date τ, and $N^{IBNR} = \sum_{l=1}^{k} N_l^{IBNR}$, the total number of IBNR claims. Note

that the both distributions are multivariate Pascal mixtures with scale parameters varying over different dimensions. To show that N^r and N^{IBNR} are univariate Pascal mixtures, we employ a two step procedure:

- First re-express (6.4) and (6.3) using a common scale parameter over different dimensions. This is an approach presented in Willmot and Woo (2015).
- Prove that the sum of the components of a multivariate Pascal mixture with a common scale parameter is a univariate Pascal mixture.

We begin with a lemma.

Lemma 6.2. *Assume that $(N_1, \ldots, N_k)$ is a multivariate Pascal mixture with varying scale parameters over different dimensions, i.e.,*

$$P(N_1 = n_1, \ldots, N_k = n_k)$$
$$= \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \beta_{(m_1,\ldots,m_k)} \prod_{j=1}^{k} p(n_j; m_j, \theta_j). \tag{6.5}$$

Then its probability generating function (PGF) is given by

$$P(z_1, \ldots, z_k) = Q \left(\frac{1}{1 - \theta_1(z_1 - 1)}, \ldots, \frac{1}{1 - \theta_k(z_k - 1)} \right),$$

where

$$Q(z_1, \ldots, z_k) = \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \beta_{(m_1,\ldots,m_k)} \prod_{j=1}^{k} z_j^{m_j}.$$

Proof. The proof is straightforward. □

Proposition 6.3. *Assume that $(N_1, \ldots, N_k)$ is a multivariate Pascal mixture with varying scale parameters over different dimensions and its joint probability is given by (6.5). For any $\theta \leq \min\{\theta_1, \ldots, \theta_k\}$, (6.5) can be re-expressed as*

$$P(N_1 = n_1, \ldots, N_k = n_k)$$
$$= \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \tilde{\beta}_{(m_1,\ldots,m_k)} p(n_j; m_j, \theta), \tag{6.6}$$

where

$$\tilde{\beta}_{(m_1,\ldots,m_k)} = \sum_{n_1=1}^{m_1} \cdots \sum_{n_k=1}^{m_k} \beta_{(n_1,\ldots,n_k)} \prod_{j=1}^{k} \binom{m_j - 1}{n_j - 1}$$
$$\times \left(\frac{\theta}{\theta_j} \right)^{n_j} \left(1 - \frac{\theta}{\theta_j} \right)^{m_j - n_j}.$$

Proof. The proof is similar to that in Willmot and Woo (2015). The PGF of $(N_1, \ldots, N_k)$ is

$$P(z_1, \ldots, z_k) = Q \left(\frac{1}{1 - \theta_1(z_1 - 1)}, \ldots, \frac{1}{1 - \theta_k(z_k - 1)} \right)$$

$$= Q \left(\frac{1}{1 - \theta(z_1 - 1)} \frac{\theta}{\theta_1} \frac{1}{1 - \frac{1 - \frac{\theta}{\theta_1}}{1 - \theta(z_1 - 1)}}, \ldots, \right.$$

$$\left. \frac{1}{1 - \theta(z_k - 1)} \frac{\theta}{\theta_k} \frac{1}{1 - \frac{1 - \frac{\theta}{\theta_k}}{1 - \theta(z_k - 1)}} \right)$$

$$= \tilde{Q} \left(\frac{1}{1 - \theta(z_1 - 1)}, \ldots, \frac{1}{1 - \theta(z_k - 1)} \right),$$

where

$\tilde{Q}(z_1, \ldots, z_k)$

$$= Q\left(\frac{\frac{\theta}{\theta_1} z_1}{1 - \left(1 - \frac{\theta}{\theta_1}\right) z_1}, \ldots, \frac{\frac{\theta}{\theta_k} z_k}{1 - \left(1 - \frac{\theta}{\theta_k}\right) z_k}\right)$$

$$= \sum_{m_1 \geq 1} \cdots \sum_{m_k \geq 1} \beta_{(m_1, \ldots, m_k)} \prod_{j=1}^{k} \left(\frac{\frac{\theta}{\theta_j} z_j}{1 - \left(1 - \frac{\theta}{\theta_j}\right) z_j}\right)^{m_j}$$

$$= \sum_{m_1 \geq 1} \cdots \sum_{m_k \geq 1} \beta_{(m_1, \ldots, m_k)}$$

$$\times \left(\sum_{n_j \geq m_j} \binom{n_j - 1}{m_j - 1} \left(\frac{\theta}{\theta_j}\right)^{m_j} \left(1 - \frac{\theta}{\theta_j}\right)^{n_j - m_j} z_j^{n_j}\right)$$

$$= \sum_{m_1 \geq 1} \cdots \sum_{m_k \geq 1} \beta_{(m_1, \ldots, m_k)} \sum_{n_1 \geq m_1} \cdots \sum_{n_k \geq m_k}$$

$$\times \left(\prod_{j=1}^{k} \binom{n_j - 1}{m_j - 1} \left(\frac{\theta}{\theta_j}\right)^{m_j} \left(1 - \frac{\theta}{\theta_j}\right)^{n_j - m_j} z_j^{n_j}\right)$$

$$= \sum_{n_1 \geq 1} \cdots \sum_{n_k \geq 1} \left(\sum_{m_1 = 1}^{n_1} \cdots \sum_{m_k = 1}^{n_k} \beta_{(m_1, \ldots, m_k)}\right)$$

$$\times \prod_{j=1}^{k} \binom{n_j - 1}{m_j - 1} \left(\frac{\theta}{\theta_j}\right)^{m_j} \left(1 - \frac{\theta}{\theta_j}\right)^{n_j - m_j}\right) \prod_{j=1}^{k} z_j^{n_j}$$

$$= \sum_{n_1 \geq 1} \cdots \sum_{n_k \geq 1} \tilde{\beta}_{(n_1, \ldots, n_k)} \prod_{j=1}^{k} z_j^{n_j}.$$

The conclusion follows from Lemma 6.2 and the uniqueness of the PGF. □

Lemma 6.4. *If* $(N_1, \ldots, N_k)$ *is a multivariate Pascal mixture with common scale parameter given by* (6.6), *then each marginal* N_i *has the following stochastic representation:*

$$N_i = \sum_{j=1}^{M_i} G_{ij}, \quad i = 1, 2, \ldots, k,$$

where G_{ij} *are i.i.d geometric random variables with mean* θ, *and the count variables* $(M_1, \ldots, M_k)$ *have a joint probability function*

$$P(M_1 = m_1, \ldots, M_k = m_k) = \tilde{\beta}_{(m_1, \ldots, m_k)}.$$

Proof. It is easy to check by calculating the PGFs of both representations. □

Proposition 6.5. *If* $(N_1, \ldots, N_k)$ *is a multivariate Pascal mixture with common scale parameter and its joint probability function given by* (6.6), *then* $N = N_1 + \cdots, N_k$ *is a univariate Pascal mixture with the same scale parameter and its mixing weights are*

$$\tilde{\beta}_m^N = \sum_{m_1 + \cdots + m_k = m} \tilde{\beta}_{(m_1, \ldots, m_k)}.$$

Proof. By Lemma 6.4, $N = N_1 + \cdots + N_k = \sum_{i=1}^{M_1 + \cdots + M_k} G_i$. As a result, N is a univariate Pascal mixture whose mixing weight for the ith component equals

$$\tilde{\beta}_m^N = P(M_1 + \cdots + M_k = m)$$

$$= \sum_{m_1 + \cdots + m_k = m} P(M_1 = m_1, \ldots, M_k = m_k)$$

$$= \sum_{m_1 + \cdots + m_k = m} \tilde{\beta}_{(m_1, \ldots, m_k)}. \quad \square$$

Theorem 6.6. *For any* $0 < \theta^r < \min\{\theta_1^r, \ldots, \theta_k^r\}$, *the total number of reported claims up to the valuation date* τ *is a univariate Pascal mixture with probability function*

$$P(N^r = n) = \sum_{m=1}^{\infty} \left(\sum_{m_1 + \cdots + m_k = m} \tilde{\beta}_{(m_1, \ldots, m_k)}\right) p(n; m, \theta^r),$$

where

$$\tilde{\beta}_{(m_1, \ldots, m_k)} = \sum_{n_1 = 1}^{m_1} \cdots \sum_{n_k = 1}^{m_k} \beta_{(n_1, \ldots, n_k)} \prod_{j=1}^{k} \binom{m_j - 1}{n_j - 1}$$

$$\times \left(\frac{\theta^r}{\theta_j^r}\right)^{n_j} \left(1 - \frac{\theta^r}{\theta_j^r}\right)^{m_j - n_j}.$$

Similarly, for any $0 < \theta^{IBNR} < \min\{\theta_1^{IBNR}, \ldots, \theta_k^{IBNR}\}$, *the total number of IBNR claims up to the valuation date* τ *is an univariate Pascal mixture, with probability function*

$$P(N^{IBNR} = n) = \sum_{m=1}^{\infty} \left(\sum_{m_1 + \cdots + m_k = m} \tilde{\beta}_{(m_1, \ldots, m_k)}\right) p(n; m, \theta^{IBNR}),$$

where

$$\tilde{\beta}_{(m_1, \ldots, m_k)} = \sum_{n_1 = 1}^{m_1} \cdots \sum_{n_k = 1}^{m_k} \beta_{(n_1, \ldots, n_k)} \prod_{j=1}^{k} \binom{m_j - 1}{n_j - 1}$$

$$\times \left(\frac{\theta^{IBNR}}{\theta_j^{IBNR}}\right)^{n_j} \left(1 - \frac{\theta^{IBNR}}{\theta_j^{IBNR}}\right)^{m_j - n_j}.$$

Proof. It is obvious by Propositions 6.3 and 6.5. □

We remark that although Theorem 6.6 shows that theoretically it has a closed-form expression under our model assumptions, there may be some computational issues. When one uses Proposition 6.3 to unify the different scale parameters over the time intervals, the resulting multivariate Pascal mixture will have infinite number of terms. There is no guarantee that a truncation of this infinite series can give an adequate approximation. More importantly, the truncated terms might play an important role in deciding the tail shape of the predictive distribution. Due to these considerations, one might resort to the simulation technique when calculating the distribution in practice.

7. Concluding remarks

In this paper, we proposed a marked Cox model for a portfolio claim arrival process along with its reporting delays. The model takes into consideration the exposure fluctuations and permits an intuitive insurance interpretation. We proved that the associated reported claim process and the IBNR claim process are also marked Cox processes with easily convertible intensity functions and marking distributions. Furthermore, the proposed model produces an equivalent discretely observed process by an order statistics property. All these are critically important from a data fitting and prediction perspective. The flexibility of the proposed Cox process in modeling temporal dependence is guaranteed as the joint distribution of the discretely observed claim arrival process is a multivariate Pascal mixture (Badescu et al., 2015a). In addition, we derived an analytical formula for the ACF of the discretely observed process, whose power-decaying pattern is attractive for temporal dependence interpretation. Finally, we showed that both the distributions of the numbers of reported claims and IBNR claims come from the class of Pascal mixtures with closed-form expressions for their mixing weights.

In Badescu et al. (2015b), we will present an algorithm to fit the proposed model to data and to estimate all the model parameters including the number of states and the transition probabilities of the Markov chain. The efficiency of the fitting algorithm and the versatility of the proposed model are illustrated through detailed simulation studies. The usefulness of the proposed model is also tested by applying it to a real insurance data set. We compare the predictive distribution of the number of IBNR claims by our model to that of the over-dispersed Poisson model (ODP) (Renshaw and Verrall, 1998), one of the several stochastic models that underpin the widely used chain-ladder method. The results show that our model can yield more accurate best estimates and more realistic predictive distributions.

Our current work opens several potential research directions. One could introduce the time trend and the seasonal effect into the claim arrival process by incorporating time covariates either in the state-dependent distributions or in the transition probability matrix. While we only include the reporting delay as the single marker in the marked Cox model, the model can easily be extended to the situation that has multiple markers, e.g., the multiple payments of a reported claim. They can be potentially modeled as recurrent events or one can generalize the model to be a marked Cox cluster process. The temporal dependence structure in the proposed model can also be enhanced in at least two directions. One could replace the underlying Markov chain structure in the current model with a Markov process to make it a full generalization of the Markov-Modulated Poisson Process (MMPP) (Asmussen and Albrecher, 2010). Another path is to impose a more complex temporal dependence structure, e.g., at both the latent process level and the observation level. This would be related to self-excited processes such as the Hawkes process (Stabile and Torrisi, 2010).

Acknowledgments

References

Ammeter, H., 1948. A generalization of the collective theory of risk in regard to fluctuating basic-probabilities. Scand. Actuar. J. 1948 (1–2), 171–198.

Antonio, K., Plat, R., 2014. Micro-level stochastic loss reserving for general insurance. Scand. Actuar. J. 2014 (7), 649–669.

Asmussen, S., Albrecher, H., 2010. Ruin Probabilities, Vol. 14. World Scientific.

Avanzi, B., Wong, B., Yang, X., 2015. A micro-level claim count model with overdispersion and reporting delays. Available at http://ssrn.com/abstract=2705241.

Badescu, A.L., Gong, L., Lin, X.S., Tang, D., 2015a. Modeling correlated frequencies with application in operational risk management. J. Oper. Risk 10 (1), 1–43.

Badescu, A.L., Lin, X.S., Tang, D., 2015b. A marked Cox model for the number of IBNR claims: estimation and application. Available at http://ssrn.com/abstract=2747223.

Bremaud, P., 2013. Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues, Vol. 31. Springer Science & Business Media.

Friedland, J., 2010. Estimating Unpaid Claims Using Basic Techniques, Vol. 201, No. 0. Casualty Actuarial Society.

Grandell, J., 1991. Aspects of Risk Theory. Springer-Verlag, New York.

Grandell, J., 1997. Mixed Poisson Processes, Vol. 77. CRC Press.

Grigelionis, B., 1963. On the convergence of sums of random step processes to a Poisson process. Theory Probab. Appl. 8 (2), 177–182.

Holmberg, R.D., 1994. Correlation and the measurement of loss reserve variability. In: Casualty Actuarial Society Forum, Vol. 1. pp. 247–278.

Huang, J., Qiu, C., Wu, X., Zhou, X., 2015. An individual loss reserving model with independent reporting and settlement. Insurance Math. Econom. 64, 232–245.

Jewell, W.S., 1989. Predicting IBNYR events and delays I: continuous time. Astin Bull. 19 (01), 25–55.

Jewell, W.S., 1990. Predicting IBNYR events and delays II: discrete time. Astin Bull. 20 (01), 93–111.

Jin, X., Frees, E.W.J., 2013. Comparing micro-and macro-Level loss reserving models. Presentation at ARIA.

Karr, A., 1991. Point Processes and Their Statistical Inference, Vol. 7. CRC Press.

Klugman, S.A., Panjer, H.H., Willmot, G.E., 2013. Loss Models: Further Topics. John Wiley & Sons.

Liu, H., Verrall, R., 2009. Predictive distributions for reserves which separate true IBNR and IBNER claims. Astin Bull. 39 (01), 35–60.

Merz, M., Wuthrich, M.V., Hashorva, E., 2013. Dependence modelling in multivariate claims run-off triangles. Ann. Actuar. Sci. 7 (01), 3–25.

Mikosch, T., 2009. Non-life Insurance Mathematics: an Introduction with the Poisson Process. Springer Science & Business Media.

Norberg, R., 1993a. Prediction of outstanding liabilities in non-life insurance. Astin Bull. 23 (01), 95–115.

Norberg, R., 1993b. Prediction of outstanding liabilities: parameter estimation. In: Proceedings of the XXIV ASTIN Coll., pp. 255–266.

Norberg, R., 1999. Prediction of outstanding liabilities II: model variations and extensions. Astin Bull. 23 (2), 213–225.

Renshaw, A.E., Verrall, R.J., 1998. A stochastic model underlying the chain-ladder technique. Br. Actuar. J. 4 (04), 903–923.

Schnieper, R., 1991. Separating true IBNR and IBNER claims. Astin Bull. 21 (01), 111–127.

Shi, P., Basu, S., Meyers, G.G., 2012. A Bayesian log-normal model for multivariate loss reserving. N. Am. Actuar. J. 16 (1), 29–51.

Stabile, G., Torrisi, G.L., 2010. Risk processes with non-stationary Hawkes claims arrivals. Methodol. Comput. Appl. Probab. 12 (3), 415–429.

Tijms, H.C., 1994. Stochastic Models: An Algorithmic Approach. John Wiley & Sons Inc..

Verdonck, T., Van Wouwe, M., Dhaene, J., 2009. A robustification of the chain-ladder method. N. Am. Actuar. J. 13 (2), 280–298.

Verrall, R., Nielsen, J.P., Jessen, A.H., 2010. Prediction of RBNS and IBNR claims using claim amounts and claim counts. Astin Bull. 40 (02), 871–887.

Willmot, G.E., Woo, J.K., 2015. On some properties of a class of multivariate Erlang mixtures with insurance applications. Astin Bull. 45 (01), 151–173.

Wuthrich, M.V., Merz, M., 2008. Stochastic Claims Reserving Methods in Insurance, Vol. 435. John Wiley & Sons.

Wuthrich, M.V., Merz, M., 2015. Stochastic Claims Reserving Manual: Advances in Dynamic Modeling. Available at SSRN 2649057.

Zhao, X., Zhou, X., 2010. Applying copula models to individual claim loss reserving methods. Insurance Math. Econom. 46 (2), 290–299.

Zhao, X.B., Zhou, X., Wang, J.L., 2009. Semiparametric model for prediction of individual claim loss reserving. Insurance Math. Econom. 45 (1), 1–8.