

A micro-level claim count model with overdispersion and reporting delays

Benjamin Avanzi^{a,b}, Bernard Wong^a, Xinda Yang^{*,a}

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

^b*Département de Mathématiques et de Statistique, Université de Montréal
Montréal QC H3T 1J4, Canada*

Abstract

The accurate estimation of outstanding liabilities of an insurance company is an essential task. This is to meet regulatory requirements, but also to achieve efficient internal capital management. Over the recent years, there has been increasing interest in the utilisation of insurance data at a more granular level, and to model claims using stochastic processes. So far, this so-called ‘micro-level reserving’ approach has mainly focused on the Poisson process.

In this paper, we propose and apply a Cox process approach to model the arrival process and reporting pattern of insurance claims. This allows for over-dispersion and serial dependency in claim counts, which are typical features in real data. We explicitly consider risk exposure and reporting delays, and show how to use our model to predict the numbers of Incurred-But-Not-Reported (IBNR) claims. The model is calibrated and illustrated using real data from the AUSI data set.

Key words: Cox process, Shot noise, Insurance claims counts, Markov Chain Monte Carlo, Filtering
JEL codes: C51, C53, C55, G22

Accepted for publication in *Insurance: Mathematics and Economics*

1. Introduction

The financial liability of insurers due to outstanding claims typically represents more than half of the company’s total liabilities, and a factor of its economic capital. Its accurate estimation is thus of paramount importance. Some of the complexity of this *reserving* problem is due to reporting delays, leading to claims that have occurred, but have not been reported yet (“IBNR”). In this paper, we are interested in estimating the number of IBNR claims using micro-level data.

Nowadays, insurers record detailed information for each individual claim, which may include, for example, arrival and reporting dates of a claim, as well as the date and amount of each transaction. This is what we call a micro-level data set. If, on the other hand, information is aggregated over a (small) number of (long) discrete time periods, then data is qualified as ‘macro-level’. This is the case, for instance, of loss reserving triangles. A vast majority of the literature on modelling insurance claims is based on macro-level claims data, including the Mack’s chain-ladder model (Mack, 1993), where a typical choice is to use a random variable to model each data point of aggregated observations. For more examples, one can refer to Taylor (2000) and Wüthrich and Merz (2008).

A micro-level approach can arguably present advantages over a macro-level approach. Firstly, the aggregation of information may lead to the disappearance of useful, perhaps material information. For example, information of the arrival and reporting time of each individual claim (and their trends) may be critical for

*Corresponding author.

Email addresses: b.avanzi@unsw.edu.au (Benjamin Avanzi), bernard.wong@unsw.edu.au (Bernard Wong), xinda.yang@unsw.edu.au (Xinda Yang)

the quality of a model. Secondly, parameter uncertainty of a macro-level model can be high due to a small number of observations (England and Verrall, 2002), resulting in less predictive power. Some of the early theoretical work in modelling micro-level claims arrival and development can be traced back to Arjas (1989) and Norberg (1993, 1999, who adopted a marked Poisson process approach). In recent years, Antonio and Plat (2014) and Larsen (2007) have further implemented Norberg’s framework with real data sets. Moreover, Pigeon et al. (2013) and Pigeon et al. (2014) develop a discrete time framework, whereby numbers of claims follow Poisson distributions. Besides the papers that study the overall micro-level claims modelling, Jewell (1989), Zhao et al. (2009), Zhao and Zhou (2010) have focused on the issue of modelling the claims arrival with a Poisson process as well as the reporting delay distribution, while Taylor et al. (2008) model individual claims development using case estimates as additional information.

The natural choice of methodology to model micro-level data is to use a continuous stochastic process. The classical model for claims processes is the Poisson process (see, e.g., Mikosch, 2006), under which the average number of claims per time unit is a constant λ .

An alternative and more general approach is to adopt a deterministic function $\lambda(t)$ instead to model the claim intensity, which results in an inhomogeneous Poisson process. The increased flexibility permits an approach that more accurately represents the nature of claim frequencies in practice where the intensity is not stationary. However, it still does not capture overdispersion, which is frequently observed in claim counts data (for example, see Section 2.9 of de Jong and Heller, 2008) and such a deterministic intensity does not allow for serial dependency of claims counts (see Denuit et al., 2007).

The issues mentioned above can be solved by modelling the intensity as a non-negative stochastic process. This results in a *doubly stochastic Poisson process*, or *Cox process* (see, for example Cox, 1955; Lando, 1998). Doubly stochastic Poisson processes have been widely applied in varying research areas, such as finance (Frey and Runggaldier, 2001), credit risk modelling (Lando, 1998), risk theory (Björk and Grandell, 1988; Albrecher and Asmussen, 2006), mortality modelling (Biffis, 2005; Schrage, 2006), catastrophe modelling (Dassios and Jang, 2003; Jang and Fu, 2012), insurance claim modelling (Avanzi et al., 2016; Badescu et al., 2015, 2016), reinsurance pricing (Dassios and Jang, 2005) and operational risk modelling (Jang and Fu, 2008).

There are a number of possible choices for the intensity process under the Cox process approach, for example, a diffusion process (Frey and Runggaldier, 2001; Schrage, 2006), a continuous time Markov chain (Frey and Runggaldier, 2001), a discrete time process with state-dependent (Erlang) intensities (Badescu et al., 2015, 2016), a jump-diffusion process (Biffis, 2005) or a shot noise process (Dassios and Jang, 2003, 2005; Albrecher and Asmussen, 2006). In this paper, we focus our illustration using the shot noise process, which processes a number of attractive behaviours such as tractability and mean reverting intensity.

A complication arising from the use of a Cox process lies in its estimation. When a (homogeneous) Poisson process is assumed, standard likelihood techniques are available (see, for example Mikosch, 2006). However, the maximum likelihood estimation approach is in general not directly applicable to a Cox process. This is because the arrivals of insurance claims are not independent under the shot noise assumption. Secondly, although the complete likelihood of observing both the Cox process and the shot noise process is simple to derive, the likelihood of observing the Cox process unconditionally on the shot noise process involves a high dimensional integral, which is not computationally efficient to calculate. Furthermore, the prediction of the IBNR counts under a Cox process model also requires the estimation of the unobservable intensity. For all those reasons, the development of a filtering algorithm is necessary.

In the case of a shot noise intensity, two filtering methods have been proposed in the literature. One method is to use a Kalman filter, which involves Gaussian approximation and is suitable in the case with high frequency but low impact shots (Dassios and Jang, 2005). The other method is to use a Reversible Jump Markov Chain Monte Carlo (“RJMCMC”) filter (Centanni and Minozzo, 2006a,b), which is based on RJMCMC simulations of the shot noise trajectory. A comparison of these two methods can be found in Avanzi et al. (2016). However, the actual implementation of the model to insurance data is not straightforward, and is generally not discussed in the existing actuarial literature. In particular, the frequency of claims is subject to exposure and reporting delays. These require non trivial model extensions. The filtering algorithm must also be modified in order to allow for such features. In this paper, we address these issues in the model construction and estimation.

This paper is structured as follows. The model assumptions along with some of the main theoretical properties are introduced in Section 2. We consider estimation methods in detail in Section 3, and extend the existing methodology to allow for varying risk exposure, and for reporting delays. We illustrate the procedures and performance of the estimation and prediction algorithms with a simulated dataset in Section 4. Furthermore, we calibrate our model to the AUSI (real) insurance data set in Section 5, and provide prediction results.

For convenience, a table with notation used throughout the paper is provided in Appendix D.

2. A shot noise Cox process with exposure and reporting delays

In this section, we develop the shot noise Cox model that is considered in this paper. Section 2.1 reviews the stationary shot noise Cox model without reporting delays. Section 2.2 defines an appropriate non-stationary version in order to allow for exposure changes over time, and Section 2.3 explains how reporting delays can be incorporated.

2.1. A stationary shot noise Cox process model

Model assumption 2.1.1. (*Cox process*) (Grandell, 1976) Denote by $N(t)$ the number of claims up to time t . We assume that $\{N(t), t \geq 0\}$ is a Cox process, that is, there exists a non-negative stochastic process $\{\Lambda(t), t \geq 0\}$ and a homogeneous Poisson process with intensity rate 1, $\{\tilde{N}(t), t \geq 0\}$, such that $\{N(t), t \geq 0\}$ has the same distribution as $\tilde{N} \circ \left(\int_0^t \Lambda(s) ds\right)$.

A Cox process (Model assumption 2.1.1) can be interpreted as an extension of a Poisson process, where the intensity process $\{\Lambda(t), t \geq 0\}$ is stochastic. In particular, we assume that the stochastic intensity, $\Lambda(t)$, is a shot noise process (see, for example, Jang, 2004; Centanni and Minozzo, 2006a,b).

Model assumption 2.1.2. (*Homogeneous shot noise intensity process*) The stochastic intensity $\Lambda(t)$ is a stationary shot noise process if

$$\Lambda(t) = \Lambda(0)e^{-kt} + \sum_{j=1}^{J(t)} X_j e^{-k(t-\tau_j)}, \quad t \geq 0, \quad (2.1)$$

where τ_j represents the arrival time of the shots resulting from a homogeneous Poisson process $J(t)$ with a deterministic intensity ρ , and the shots X_j 's are independent and identically distributed random variables with a density function f_X (on the positive domain with finite mean). Furthermore, we assume that $\Lambda(0)$, the initial level of $\Lambda(t)$, follows the stationary distribution of the shot noise process. This is a convenient assumption and ensures that the shot noise process is stationary from $t = 0$.

Remark 2.1. In this paper, we follow the literature (see, for example, Dassios and Jang, 2003; Jang, 2004; Dassios and Jang, 2005; Centanni and Minozzo, 2006a,b; Avanzi et al., 2016) and consider an exponential decay function. However, it should be noted that alternative decay functions can be considered, at a slight loss of mathematical convenience (see, Schmidt, 2014, for details).

A possible interpretation of the shot is an external economic and environmental event which causes a sudden increase in claim intensity (e.g., rainy weather). The frequency of these shots reflects how often the external events occur while the sizes of these shot reflect the impact of the events on claim intensity (that is, the likelihood of a claim occurring). As time progresses, the event will have a diminished impact on claim intensity, as shown by the exponential decay of the claim intensity from the shot. In the following, we present an example of the shot noise process with exponentially distributed sizes.

Example 1. (*exponentially distributed shots*) Suppose that

$$f_X(x) = \eta e^{-\eta x}, \quad x > 0, \quad (2.2)$$

Such a shot noise process presents attractive features, such as non-negativity, tractability, auto-correlation. In particular, its stationary distribution is a Gamma random variable (see Centanni and Minozzo, 2006a; Jang, 2004). Furthermore, the shot noise process is mean-reverting, with a mean-reverting level of $\rho/(\eta k)$ and a mean-reversion rate of k (see, for details, Avanzi et al., 2016). Such a process is not new in the actuarial literature; see Dassios and Jang (2003); Jang (2004); Dassios and Jang (2005); Albrecher and Asmussen (2006). One can refer to Avanzi et al. (2016) for the statistical properties of a shot noise process and a shot noise Cox process. In particular, the mean and variance of $N(t) - N(s)$ ($t > s > 0$) can be shown to be

$$E[N(t) - N(s)] = \frac{\rho(t-s)}{\eta k}; \quad (2.3)$$

$$\text{Var}(N(t) - N(s)) = \frac{2\rho}{\eta^2 k^2} \left(t - s - \frac{1 - e^{-k(t-s)}}{k} \right) + \frac{\rho(t-s)}{\eta k}; \quad (2.4)$$

which illustrates in particular that a shot noise Cox process allows for over-dispersion as opposed to a Poisson process. The autocovariance function $\gamma(h)$ of $N(t) - N(t - \Delta)$ for $h = 1, 2, 3, \dots$, $\Delta > 0$ is

$$\gamma(h) = \text{Cov}(N(t+h) - N(t-\Delta+h), N(t) - N(t-\Delta)) = \frac{\rho e^{-kh} \left(e^{\frac{k\Delta}{2}} - e^{-\frac{k\Delta}{2}} \right)^2}{\eta^2 k^3}. \quad (2.5)$$

One can generalise this example to allow for a mixture of exponential shots in (2.2). Such a model could be justified by the existence of multiple (different) sources of risk. The stationary distribution of such a shot noise is a mixed Gamma random variable (the proof of which is a straightforward extension of that for the special case (where $n = 2$) considered by Jang, 2004). In this case, moments in (2.3)-(2.5) are mixed in the same way as the exponentials for the shots density f_X .

In Example 1, the shot noise Cox process presents a number of advantages when compared to the Poisson process. Firstly, despite having stationary increments, the shot noise Cox process does not have independent increments. Secondly, the autocovariance function (and hence the autocorrelation function) of $N(t) - N(t - \Delta)$ (where $\Delta > 0$) is always positive for lags greater than zero and exponentially decays as h increases, where k is the measure of how quickly the autocovariance decays with h . Hence, a possible heuristic test to verify the validity of using a Cox process on claims data is to check whether the autocorrelations of the increments are positive and exponentially decreasing at increasing lags. Table 1 summarises the comparison between a Poisson process, a mixed Poisson process (see Section 2.3 Mikosch, 2006) and a shot noise Cox process.

	a Poisson process	a mixed Poisson process	a shot noise Cox process with exponential shots
over-dispersion	no	yes	yes
auto-correlation	no	yes, constant regardless of lags	yes, positive and exponentially decaying auto-correlation

Table 1: Comparison between the claim counts of a Poisson process, a mixed Poisson process process, and a shot noise Cox process with exponential shots

2.2. Risk exposure

In this section, we discuss alternative ways to extend the shot noise process to allow for exposure. This requires a non-stationary process. Werner and Modlin (2010, Chapter 4) define *exposure* as “the basic unit that measures a policy’s exposure to loss”. One primary criterion for choosing an exposure base is that it “should be directly proportional to expected loss”. In our case, we will assume that it directly affects the likelihood of a claim to occur - the intensity of the process. We further assume that it is an exogenous, deterministic quantity, denoted $\Omega(t)$, that can be determined for the timeframe of an existing data set.

Varying exposure means that we need to deal with an *inhomogeneous* intensity. In practice, the choice of $\Omega(t)$ depends on the nature of the line of business. Some traditional measures of exposure include the number of policyholders, total sum insured, or average weekly earnings. Model 2.1.2 above assumes a stationary arrival of claims, which is only suitable if the exposure of the business is homogeneous through time.

We extend Model 2.1.2 by using Model 2.2.1. Here we assume that the exposure scales the whole intensity process. For example, suppose that all the policyholders of a housing and contents insurance product live in the same area, and hence experience the same exposure conditions that drive the likelihood of insurance claims. An increasing number of policyholders of that particular kind will simply scale up the magnitude of the likelihood of claims.

Model assumption 2.2.1. (*Exposure scaling of the shot noise process*) *There exists a stationary shot noise process, $\Lambda'(t)$, such that*

$$\Lambda(t) = \Omega(t)\Lambda'(t). \quad (2.6)$$

Suppose that $\Omega(t)$ is a continuously differentiable function with $d\Omega(t) = \Omega_t(t) dt$. One should notice that $\Lambda'(t)$ is a semi-martingale (Schmidt, 2014) and by using the Itô formula for semi-martingales (Proposition 8.19, Cont and Tankov, 2004) we have

$$d\Lambda(t) = - \left(k - \frac{\Omega_t(t)}{\Omega(t)} \right) \Lambda(t)dt + \Omega(t)dK(t), \quad \text{where } K(t) = \sum_{i=1}^{J(t)} X_j \quad (2.7)$$

and therefore $\Lambda(t)$ is an inhomogeneous shot noise process with time-dependent parameters.

Example 2. *Suppose that $\Lambda'(t)$ follows the assumptions in Example 1 with parameters (ρ, η, k) . Then $\Lambda(t)$ is a non-stationary shot noise process of exponential shots with parameters $(\rho_{inh}(t), \eta_{inh}(t), k_{inh}(t))$ such that*

$$\begin{aligned} \rho_{inh}(t) &= \rho, \\ \eta_{inh}(t) &= \eta/\Omega(t), \\ k_{inh}(t) &= k - \Omega_t(t)/\Omega(t). \end{aligned} \quad (2.8)$$

One can interpret Equation (2.8) in the following way. The exposure will not affect the frequency of shots, however it will affect both the severity and decay effect of shots. Firstly, the higher the exposure, the larger the shots tend to be. Furthermore, an increasing exposure (that is, $\Omega_t(t) > 0$) will to some extent cancel the effect of the decay. This is because the exposure modifies the intensity process, not just the compound Poisson process of shots arrival. Therefore the decayed effect from previous shots will be amplified if the exposure increases, which is equivalent to weakening the decaying effect. If the exposure increases fast enough, the intensity may even increase without any shot arrival.

Besides its physical interpretation, Model 2.2.1 also has a nice property of one-to-one correspondence between $\Lambda(t)$ and $\Lambda'(t)$. This means one only needs to estimate the homogeneous shot noise process (that is, $\Lambda'(t)$), in order to estimate the stochastic intensity (that is, $\Lambda(t)$). In the following section, one can see that estimating $\Lambda'(t)$ involves a modification of the existing RJMCMC algorithm of Centanni and Minozzo (2006a,b). Furthermore, under Model 2.2.1, one can change the unit of $\Omega(t)$ and the resulting data can still be fitted to a shot noise Cox process (see Appendix A.1 for more details).

Remark 2.2. *There are alternative ways to allow for exposure in the intensity, and this is not a trivial choice. For example, suppose that the shot noise process follows Example 1. We could assume that exposure only affects the frequency parameter—write $\rho(t)$ instead of ρ —while both η and k remain constant, such that*

$$\rho(t) = \rho\Omega(t), \quad \eta(t) = \eta, \quad k(t) = k. \quad (2.9)$$

This assumption is valid if any two fractions of exposure are independent (for example, the claims from any two policyholders are independent), both of which contribute to the claims arrival likelihood in an identical and additive way. Therefore, an increasing level of exposure will lead to a proportionally increasing frequency of shots.

Remark 2.3. One can extend the definition of risk exposure to allow for seasonality. For instance, one possible extension is to define a risk exposure $\Omega(t)$ such that

$$\begin{aligned}\Omega(t) &= f(t) \times \text{a traditional measure of exposure,} \\ f(t) &= a_i \text{ if } t \text{ is in the } i^{\text{th}} \text{ season, } i = 1, \dots, s, \\ &\text{where } a_s = 1 \text{ and } s > 1.\end{aligned}\tag{2.10}$$

2.3. Reporting delays

Modelling the distribution of the reporting delay between the occurrence date (also called ‘accident date’) and (subsequent) reporting date of an insured event is essential in the insurance context (see, for instance, Jewell, 1989). This is because the historical data can only record claims that have been reported, therefore one needs to consider the fact that the numbers of observed claims are not necessarily the same as the actual number of claims arrivals (especially for recent periods). Reported claims may develop further into higher numbers. This creates complications, as explained in the next section.

The distribution of reporting delays may vary across different lines of business, where some lines may see a longer-tailed delay distribution while others have shorter-tailed delays. A typical observation is that a significant portion of claims are reported within one month or even one week, which motivates the separation of this range from longer delays. Denote by R the reporting delay, a random variable whose distribution is defined in Model assumption 2.3.1.

Model assumption 2.3.1. (*Reporting delay*) Denote by $F_R(x) = \Pr(R \leq x)$ the distribution function of reporting delays, where x is expressed in an appropriate unit of time. Denote by F_Y a continuous distribution function over the range $[0, \infty)$ and u a threshold value. Denote by β the vector of parameters. We assume that

$$F_R(x; u, \alpha, \beta) = \begin{cases} 0 & \forall x < 0 \\ \alpha & \forall 0 \leq x \leq u \\ \alpha + (1 - \alpha)F_Y(x - u; \beta) & \forall x > u. \end{cases}\tag{2.11}$$

Model assumption 2.3.1 assumes that there is a mass of probability α at 0 corresponding to delays of up to u units of time (e.g., days or weeks). The (conditional) distribution of delays (given they exceed threshold u) is modelled separately according to F_Y . In practice, the choice of u will depend on the characteristics of the data. Moreover, one should consider the horizon of forecasting in choosing the value of u since one does not model in details the distribution between $[0, u]$. For example, if the forecasting is over a one-year horizon, then a threshold u of 30 days may be adequate. On the contrary, if the forecasting is over a one-quarter horizon, a shorter term threshold may be required.

3. Parameter estimation and prediction

In this section, we extend the Monte Carlo Expectation Maximisation (“MCEM”) with RJMCMC algorithm developed by Centanni and Minozzo (2006a,b) (see also Avanzi et al., 2016) to allow for a general shot noise process model with non-constant exposure and reporting delays. Specifically, the RJMCMC filtering algorithm developed in Section 3.3 is valid for any distribution of shots within the framework of Definition 2.1.2. For completeness and convenience, we present the whole RJMCMC algorithm. In the next section, we introduce some notations, before developing the estimation and prediction algorithms in the later sections.

3.1. Notation

Suppose that the overall observation period is $[0, T]$, which is partitioned into a number of L sub-periods with equal lengths Δ (where $T = L \times \Delta$). In a real world scenario, one can only observe in which sub-period a claim arrives or is reported (e.g. day), rather than the exact arrival or reporting time (e.g. 2:51pm). Let N_i be the *ultimate* number of claims from accident period i (that is, the total number of claims that arrive

between time $(i-1)\Delta$ and $i\Delta$) and define $\mathbf{N} = \{N_i, i = 1, \dots, L\}$. Please note that \mathbf{N} is not observable due to the existence of the reporting delay distribution.

Denote by N_{ij} the number of claims that arrive during the i^{th} sub-period and are reported to the insurer in the $(i+j-1)^{\text{th}}$ sub-period. We say that N_{ij} is the number of claims that occur in the i^{th} accident period and are reported in the j^{th} reporting period. Since all the claims will be reported eventually

$$N_i = \sum_{j=1}^{\infty} N_{ij}. \quad (3.1)$$

At the end of the observation period T , we only observe $\{N_{ij} : i \geq 1, j \geq 1, i+j \leq L+1\}$, denoted by $\mathbf{N}_{\mathbf{D}}$, where the subscript \mathbf{D} indicates that this is a collection of claims counts with the information of reporting delays. It is, in a sense, a micro-level incremental triangle of claim counts. Its rows have sums

$$N_{iR} := \sum_{j=1}^{L-i+1} N_{ij}, \quad (3.2)$$

which represent the number of claims that occur during the i^{th} accident period that have been reported by time T . We will write $\mathbf{N}_{\mathbf{R}} = \{N_{iR}, i = 1, \dots, L\}$. Note that N_{iR}/N_i represents the proportion of claims from accident period i that have been reported by time T . In the reserving practice, this is often referred to as *operational time*, and is often plotted as a time series for diagnostic purposes.

Finally, we denote by n the number of shots, and we characterise the intensity process by $\boldsymbol{\theta} = \{\Lambda_0, \tau_1, \dots, \tau_n, X_1, \dots, X_n\}$. Furthermore, denote by

$$M_i = \int_{(i-1)\Delta}^{i\Delta} \Lambda(s) ds \quad (3.3)$$

the integrated intensity process in the i^{th} sub-period. Conditional on M_i , N_i is a Poisson random variable with intensity M_i ($1 \leq i \leq L$).

Table 2 summarises our notation.

	symbol	definition	mathematical relationship
unobservable	N_{ij}	the number of claims that occur in the i^{th} accident period and are reported in the j^{th} reporting period	
	N_i	the <i>ultimate</i> number of claims from accident period i	$N_i = \sum_{j=1}^{\infty} N_{ij} \quad i \geq 1$
	\mathbf{N}	the collection of N_i	$\mathbf{N} = \{N_i; i \geq 1\}$
	$\boldsymbol{\theta}$	the collection of the initial value of the shot noise process, arrival times and sizes of the shots	$\boldsymbol{\theta} = \{\Lambda_0, \tau_1, \dots, \tau_n, X_1, \dots, X_n\}$
observable	$\mathbf{N}_{\mathbf{D}}$	the subset of $\{N_{ij}, 1 \leq i \leq L, 1 \leq j \leq L\}$ that is observed by time T	$\{N_{ij} : i \geq 1, j \geq 1, i+j \leq L+1\}$
	N_{iR}	the number of claims that occur during the i^{th} accident period that have been reported by time T	$N_{iR} = \sum_{j=1}^{L-i+1} N_{ij} \quad 1 \leq i \leq L$
	$\mathbf{N}_{\mathbf{R}}$	the collection of N_{iR}	$\mathbf{N}_{\mathbf{R}} = \{N_{iR}; 1 \leq i \leq L\}$

Table 2: Notation related to the data structure

3.2. Initial parameters

In this section, we develop two methods to obtain initial estimates of the parameters of a homogeneous shot noise process defined in Example 1. The same idea applies to other assumptions. The first method is

based on moment matching. One can derive the theoretical mean, variance and auto-covariance at lag h (where $h \in \{1, 2, \dots\}$) for \mathbf{N} based on Equations (2.3), (2.4), and (2.5). By matching 3 empirical moments and the corresponding theoretical moments, one can obtain initial guesses of the parameters. Denote by m_1 , m_2 and m_3 the empirical mean, variance and auto-correlation at lag h and ρ_0 , η_0 and k_0 the initial guesses. We have

$$\frac{2m_2m_3}{m_2 - m_1} = \frac{\exp(-kh)(\exp(k\Delta/2) - \exp(-k\Delta/2))^2}{k\Delta - 1 + \exp(-k\Delta)}, \quad (3.4)$$

$$\eta_0 = \frac{2m_1 \left(\Delta - \frac{1 - \exp(-k_0\Delta)}{k_0} \right)}{(m_2 - m_1)\Delta k_0}, \quad (3.5)$$

$$\rho_0 = \frac{2m_1^2 \left(\Delta - \frac{1 - \exp(-k_0\Delta)}{k_0} \right)}{(m_2 - m_1)\Delta^2}, \quad (3.6)$$

where k_0 can be found first by solving (3.4) numerically, and then used to calculate ρ_0 and η_0 . One may also want to use other combinations of moments, for example, one could consider higher moments. Since we do not observe \mathbf{N} , we choose to use the observed claims counts, $\mathbf{N}_{\mathbf{R}}$, to approximate the ultimate claims counts, \mathbf{N} . There are two conditions to ensure a satisfactory level of approximation. Firstly, one should choose a subset of the data where the exposure is relatively constant. Therefore the homogeneity assumption implied by moment matching can be tolerated. Secondly, one should exclude the observations where the accident periods are close to the end of observations. This is to ensure that the proportion of unreported claims in the ultimate claims is not material. The length of observation period to be excluded depends on the reporting delay distribution. In practice, one can choose to exclude only a small portion of data if the observed reporting delays tend to be small.

The second method is based on a negative binomial approximation (see also Avanzi, Liu, and Wong, 2016). Based on the definition of a Cox process, we know that there exists a Poisson process \tilde{N} such that

$$N(t + \Delta) - N(t) \stackrel{d}{=} \tilde{N} \left(\int_t^{t+\Delta} \Lambda(t) dt \right) \approx \tilde{N}(\Lambda(t)\Delta). \quad (3.7)$$

Furthermore, at a given time t , the random variable $\Lambda(t)$ follows the stationary distribution of the shot noise process, which is known to be a gamma random variable (Centanni and Minozzo, 2006b). Then one can show (for example, by considering the moment generating function) that $\Lambda(t)\Delta$ is also a Gamma random variable but with parameters ρ/k and η/Δ . Now we consider the random variable $\tilde{N}(\Lambda(t)\Delta)$. It is a well known result that a mixed Poisson random variable with a gamma prior follows in fact a negative binomial distribution (for example, see section 6.3 of Klugman et al., 2012). In particular, the parameters of the negative binomial distribution in this case are $r = \rho/k$ and $p = \eta/(\eta + \Delta)$.

Therefore, one can fit \mathbf{N} to a negative binomial (r, p) distribution. Just as in the case of the first method, we use $\mathbf{N}_{\mathbf{R}}$ instead \mathbf{N} since the latter is not observable and the same two conditions apply. Using a method of moments yields

$$\frac{\rho_0}{k_0} = \hat{r} \quad \text{and} \quad \eta_0 = \frac{\hat{p}}{1 - \hat{p}} \Delta, \quad (3.8)$$

which provides us with two equations to solve for three parameters. One can use (3.4) as the third equation for an estimate of k .

3.3. A RJMCMC filtering algorithm

The unobservable intensity, $\Lambda(t)$, can be characterised by $\boldsymbol{\theta}$, representing the collection of the initial value of the shot noise process, as well as the arrival times and sizes of the shots. This is because there is a bijective relationship between $\Lambda(t)$ and $\Lambda'(t)$, where the latter is characterised by $\boldsymbol{\theta}$. Therefore the filtering of $\Lambda(t)$ is equivalent to estimation of $\boldsymbol{\theta}$ given the observed claim count $\mathbf{N}_{\mathbf{D}}$. This can be considered as

similar to a Bayesian problem and solved by MCMC. In particular, θ is a random vector whose length is unknown since the number of shots (n , a Poisson distributed quantity) is random. Therefore, one needs to use a RJMCMC filter, which extends the Metropolis-Hastings algorithm to allow for dimension changing. Centanni and Minozzo (2006a,b) have considered the situation where there is no reporting delay or non-constant risk exposure (that is, the estimation of θ given \mathbf{N} with $\Omega(t) = 1$). Here we extend the literature by allowing for such features in a real insurance dataset.

We start by presenting the general procedures of the RJMCMC algorithm, then proceed by deriving the required likelihood functions. For more details of the RJMCMC algorithm, one can refer to Green (1995) and Gelman et al. (2011). There are four steps of a RJMCMC filter:

- with a probability of $p(r|n)$, generate a move type r with a probability given the current estimate of (n, θ)
- given the move type r and (n, θ) , generate a random vector \mathbf{u} with a density function $q(\mathbf{u}|r, n, \theta)$
- propose a new state of the Markov Chain (n', θ') such that

$$(\theta', \mathbf{u}') = f_{r,n,n'}(\theta, \mathbf{u}) \quad (3.9)$$

where \mathbf{u}' is the random quantity in the reverse move (that is, to transfer θ' back to θ)

- the proposed Markov Chain is accepted with probability $\alpha_r(\theta, \theta')$, where

$$\alpha_r(\theta, \theta') = \min \left\{ 1, \underbrace{\frac{L(\mathbf{N}_D|n', \theta'; k)}{L(\mathbf{N}_D|n, \theta; k)}}_{\text{likelihood ratio}} \times \underbrace{\frac{p(n', \theta')}{p(n, \theta)}}_{\text{prior ratio}} \times \underbrace{\frac{p(r'|n') q(\mathbf{u}'|r', n', \theta')}{p(r|n) q(\mathbf{u}|r, n, \theta)}}_{\text{proposal ratio}} \times \underbrace{\left| \frac{\partial f_{r,n}(\theta, \mathbf{u})}{\partial(\theta, \mathbf{u})} \right|}_{\text{Jacobian}} \right\}. \quad (3.10)$$

One can decompose (3.10) into the multiplication of likelihood ratio, prior ratio, proposal ratio and Jacobian.

There are five move types (see Table 3) and we generate all possible moves with equal probabilities. These are arbitrary, non-informative choices (see also Centanni and Minozzo, 2006b). Note that there are two possible sets of probabilities for generating move types. This is because that, given a current estimate of zero shot, one cannot generate a move of changing the height of a shot, changing the position of a shot or deleting a shot. In this case, one can only generate a new shot or modify the height of the initial value.

move type (' r ')	proposal of the next state	$p(r n)$
s	modifying the initial value of $\Lambda'(t)$ by drawing $\Lambda'^{new}(0)$ from the stationary distribution with density function f_S	0.5 ($n = 0$), 0.2 ($n > 0$)
b	generating a new shot by drawing a new position τ^* uniformly from $(0, t]$ and drawing a new jump height X^* from the shot size distribution f_X	0.5 ($n = 0$), 0.2 ($n > 0$)
h	changing the height of a shot by drawing j from the discrete uniform distribution over $\{1, \dots, n\}$ and drawing X_j' from the shot size distribution f_X to replace X_j	0 ($n = 0$), 0.2 ($n > 0$)
p	changing the position of a shot by drawing j from the discrete uniform distribution over $\{1, \dots, n\}$ and drawing a new position, τ_j' , uniformly over (τ_{j-1}, τ_{j+1}) (where $\tau_0 = 0$ and $\tau_{n+1} = t$)	0 ($n = 0$), 0.2 ($n > 0$)
d	deleting a shot by drawing j from the discrete uniform distribution over $\{1, \dots, n\}$ and deleting the j^{th} shot	0 ($n = 0$), 0.2 ($n > 0$)

Table 3: types of moves

The prior ratio, proposal ratio and Jacobian of each move type are summarised in Table 4. Observe that the Jacobian for a move that does not involve a dimension change (that is, moves s , h and p in our case) is always 1; the Jacobians for the dimensional changing moves (that is, moves b and d in our case) is 1 in our case, but will not be 1 in general

move type	prior ratio	proposal ratio	Jacobian
s	$f_S(\lambda^{new}(0))/f_S(\lambda'(0))$	$f_S(\lambda'(0))/f_S(\lambda^{new}(0))$	1
b	$\rho f_X(X'_j)$	$\frac{p(d n+1)}{p(b n)} \frac{(n+1)^{-1}}{t^{-1} f_X(X'_j)}$	1
h	$f_X(X'_j)/f_X(X_j)$	$f_X(X_j)/f_X(X'_j)$	1
p	1	1	1
d	$(\rho f_X(X_j))^{-1}$	$\frac{p(b n-1)}{p(d n)} \frac{t^{-1} f_X(X_j)}{n^{-1}}$	1

Table 4: the RJMCMC filtering algorithm

The conditional likelihood function of \mathbf{N}_D given $\boldsymbol{\theta}$ is

$$L(\mathbf{N}_D|\boldsymbol{\theta}; k, \alpha, \boldsymbol{\beta}) = \prod_{i=1}^L \sum_{m=N_{iR}}^{\infty} A(m, M_i; k) B(i, m) C(i, m; \alpha, \boldsymbol{\beta}) \quad (3.11)$$

where

$$A(m, M_i; k) = \frac{e^{-M_i} M_i^m}{m!}, \quad (3.12)$$

$$B(i, m) = \frac{m!}{\prod_{j=1}^{L-i+1} N_{ij}! \times (m - N_{iR})!}, \quad (3.13)$$

and

$$C(i, m; \alpha, \boldsymbol{\beta}) = \prod_{j=1}^{L-i+1} \Pr(\text{a claim is reported in reporting period } j)^{N_{ij}} \times (1 - F_R((L - i + 0.5)\Delta))^{m - N_{iR}}. \quad (3.14)$$

The index i refers to an accident period and m refers to the number of ultimate claims. The term $A(m, M_i; k)$ is the likelihood of having a number m of ultimate claims from accident period i . The term $B(i, m)C(i, m; \boldsymbol{\beta})$ is the likelihood of a multinomial distribution, which allocates the count of ultimate claims into each reporting delay period. Here $B(i, m)$ calculates the number of combinations of the order of the reporting delays in each reporting period; and $C(i, m; \boldsymbol{\beta})$ calculates the probability of a particular order of reporting delays (where there is a reporting bin for unreported claims). In calculating the probability of the reporting delay, we assume that the arrival time is the middle of the arrival period. For an unreported claim that arrives from the i^{th} accident period, the likelihood of its reporting delay is approximated by being at least $(L - i + 0.5)\Delta$, where the arrival time is approximated to be at the middle of the accident period. In particular, a number of $(m - N_{iR})$ IBNR claims from accident period i are reported in the (future) period that is between the end of T and ∞ .

For the allowance of changing exposure over time, we assume for simplicity that $\Omega(t)$ is a piece-wise constant function, where $\Omega(t) = \Omega^i$ if t is in the i^{th} accident period. The integrated intensity for subperiod i is then

$$M_i = \int_{(i-1)\Delta}^{i\Delta} \Omega(t) \Lambda'(t) dt = \Omega^i \int_{(i-1)\Delta}^{i\Delta} \Lambda'(t) dt \quad (3.15)$$

where $\Lambda'(t)$ is defined in Model assumption 2.1.2. The piece-wise constant assumption simplifies the issue of calculating M_i .

The log-likelihood simplifies to

$$\begin{aligned} \log L(\mathbf{N}_D | \boldsymbol{\theta}; k, \alpha, \beta) = & - \sum_{i=1}^L M_i + \sum_{i=1}^L N_{iR} \log M_i + \sum_{i=1}^L M_i (1 - F_R((L - i + 0.5)\Delta)) \\ & + \sum_{j=1}^L \left[\left(\sum_{i=1}^{L-j+1} N_{ij} \right) \log(\Pr(\text{a claim is reported in reporting period } j)) \right] + \text{constant}, \end{aligned} \quad (3.16)$$

where the constant term is a function of observations only.

Remark 3.1. *For consistency, in model calibration, one should take care in the preparation of the reporting delay data such that all the small delays (given the subjective choice of u) are merged into the first reporting delay period. Moreover, one should exclude the reporting delay periods during which the probability that a claim is reported is 0. Including these periods will result in calculating $\log(0)$ and lead to unsatisfactory results in numerical implementation. Furthermore, the mixture nature of Model 2.3.1 requires care in numerical calculation of the likelihood of reporting delays. Firstly, in the case that the reporting delay falls in the continuous range of the distribution, one can use the continuous density function evaluated at the middle point of the possible range. Secondly, in a case where a claim is reported in the arrival period, this probability should be the sum of α (that is, the initial mass at 0) and any mass from the continuous part of the distribution (approximated by the density at the middle point of the possible range).*

Remark 3.2. *There are alternative approaches to derive the conditional likelihood function to allow for the micro-level structure of the data. Firstly, conditional on its intensity, the likelihood of a Cox process is in fact the same as that of a Poisson process, where the latter has been developed in the literature. For example, Antonio and Plat (2014) utilised exact observations of the reporting time and arrival time of an individual claim. In practice, data might not be recorded in a such minute manner (e.g. to the second) for the exact inter-arrival times to be observed. Instead, typically it is recorded to the date of an event (for example, claim occurrence and notification). Even in a case where the exact time of a claim arrival is recorded, the small inter-arrival time may create computational difficulty (such as 0 likelihood as a result of rounding) and additional adjustment may be involved in such a nano structure (for example, incorporating the fact that claims arrive at different frequencies during the days and nights). Therefore we extend the literature by allowing for the discrete nature of the data.*

Alternatively, one can also derive the likelihood function based on the discrete observation of reported claims of each accident period, as opposed to taking a probability-weighted average of possible outcomes of IBNR counts. Although the reported claims of accident periods are simply independent Poisson random variables (conditionally, given the intensity), the intensity of each Poisson distribution involves a convolution of the intensity and the reporting delay distribution. Calculating such a likelihood involves more approximation than (3.11), and hence is not preferred.

Remark 3.3. *It is worth mentioning that one should not use multiplication of independent and identical distribution reporting delay likelihoods to replace the multinomial distribution component in (3.11). This results from the fact that the likelihood of reporting delays is derived conditionally on the ultimate number of claims; hence they are not independent.*

Remark 3.4. *In the insurance context, while the exposure data is typically continuously changing, it is recorded at discrete time points only. It is then natural to model $\Omega(t)$ as a piece-wise constant function. Alternative approaches (such as fitting splines through observations) are also possible if the piece-wise constant approximation is likely to lead to material distortions.*

3.4. The full MCEM-RJMCMC algorithm

The E-M algorithm is an iterative algorithm that updates the parameters in the presence of unobservable components— $\boldsymbol{\theta}$ in our case. Let us illustrate the steps of an E-M algorithm (refer to Rydén, 1996, for more

details). Let $\phi = (\rho, \eta, k)$ be the vector of parameters we want to estimate. The E-M algorithm involves two steps, the first of which is the *Expectation* (E) step, where one determines the conditional expectation (given current parameter estimates ϕ^k in the $(k + 1)^{th}$ iteration)

$$Q(\phi, \phi^k) = \mathbb{E}_{\phi^k}[\log \mathcal{L}_c(\mathbf{N}_D, \theta; \phi) | \mathbf{N}_D]. \quad (3.17)$$

The second step is the *Maximisation* (M) step, whereby the estimate is updated from ϕ^k to ϕ^{k+1} by maximising the condition expectation

$$\phi^{k+1} = \arg \max_{\phi} Q(\phi, \phi^k) \quad (3.18)$$

A complication arises in the computation of the expectation in the E-step, which requires finding the conditional density function of θ given \mathbf{N} , which cannot be computed directly. However, this can be approximated by using an (RJ)MCMC algorithm (hence resulting in an MCEM algorithm).

A full MCEM-RJMCMC algorithm that estimates all the parameters simultaneously includes the following steps:

- in the k^{th} iteration, generating a large number of RJMCMC iterations given ϕ^k (where ϕ^0 refer to the initial estimates);
- approximating the conditional expectation $Q(\phi, \phi^k)$ as an average of the conditional likelihoods given the RJMCMC simulations;
- updating the parameters to ϕ^{k+1} by maximising the approximated conditional expectation term.

3.5. A partial MCEM-RJMCMC algorithm

For computational ease, one can alternatively opt to perform a partial MCEM-RJMCMC algorithm to separate the calibration of the reporting delay model from that of the arrival process. This includes estimation of the reporting delay parameters by maximising the term

$$\sum_{j=1}^L \left[\left(\sum_{i=1}^{L-j+1} N_{ij} \right) \log(\Pr(\text{a claim is reported in reporting period } j)) \right], \quad (3.19)$$

which is equivalent to maximising the likelihood of independent observation, before proceeding to the MCEM algorithm with fixed reporting delay parameters. In this case, the conditional likelihood function can be further simplified into

$$\begin{aligned} & L(\mathbf{N}_D | \theta; k, \alpha, \beta) \\ &= - \sum_{i=1}^L M_i + \sum_{i=1}^L N_{iR} \log M_i + \sum_{i=1}^L M_i (1 - F_R((L - i + 0.5)\Delta)) + \text{constant}, \end{aligned} \quad (3.20)$$

which only depends on \mathbf{N}_D through \mathbf{N}_{iR} . With this partial algorithm, an implicit approximation is that $F_R((L - i + 0.5)\Delta) \approx 1$. This approximation is acceptable if the reporting delay distribution has a relatively short tail such that $F_R((L - i + 0.5)\Delta)$ is close to 1 for most of the accident periods.

The partial MCEM-RJMCMC algorithm is expected to result in less computational time. However, the independent estimation of the arrival process and reporting distribution may provide unsatisfactory results, which we illustrate numerically in the next section. In our work, we will use a partial MCEM-RJMCMC algorithm as an intermediate step in order to arrive at a better starting values for the full MCEM-RJMCMC fitting. Such an approach can potentially improve the computation speed and efficiency of the fitting procedure, and is akin to the Inference Functions of Margins (‘IFM’) approach used in copula fitting (see Chapter 10 of Joe, 1997).

Remark 3.5. *In both the full MCEM-RJMCMC and partial MCEM-RJMCMC algorithms, $\{N_{iR}, i = 1, \dots, L\}$ and $\{\sum_{i=1}^{L-j+1} N_{ij}, j = 1, \dots, L\}$ (that is, the row sums and column sums in the view of a micro-level count triangle) are sufficient for the estimation of θ , and subsequently one doesn’t need the whole dataset of \mathbf{N}_D in the estimation. This can be seen from the conditional log-likelihood functions 3.16 and 3.20.*

3.6. Prediction of IBNR counts

This prediction problem is equivalent to deriving the distribution of $(\mathbf{N} - \mathbf{N}_R | \mathbf{N}_D)$, that is, the distribution of the difference between the ultimate counts and observed claims counts given the observations. Here we have

$$\begin{aligned}
& \Pr(\mathbf{N} - \mathbf{N}_R = (n_1^{\text{IBNR}}, \dots, n_L^{\text{IBNR}}) | \mathbf{N}_D) \\
&= \mathbb{E}_{\Lambda(t) | \mathbf{N}_D} (\Pr(\mathbf{N} - \mathbf{N}_R = (n_1^{\text{IBNR}}, \dots, n_L^{\text{IBNR}}) | \mathbf{N}_D, \Lambda(t))) \\
&= \mathbb{E}_{\Lambda(t) | \mathbf{N}_D} (\Pr(\mathbf{N} - \mathbf{N}_R = (n_1^{\text{IBNR}}, \dots, n_L^{\text{IBNR}}) | \Lambda(t))) \\
&= \mathbb{E}_{\Lambda(t) | \mathbf{N}_D} \left(\prod_{i=1}^L \Pr(N_i - N_{iR} = n_i^{\text{IBNR}} | M_i) \right),
\end{aligned} \tag{3.21}$$

where $(N_i - N_{iR} | M_i)$ is a Poisson random variable with an intensity $M_i(1 - F_R((L - i + 0.5)\Delta))$. The last equality results from the conditional independence property of a Cox process given the intensity. The expectation operation with respect to the conditional distribution of $\Lambda(t)$ given \mathbf{N}_D can be numerically evaluated via simulation (using the RJMCMC algorithm. Suppose that there is a number of H ($H \geq 1$) RJMCMC simulations, and denote by $\hat{M}(i, h)$ the integrated intensity of the h^{th} filtering outcome over the i^{th} accident period, then

$$\Pr(N_i - N_{iR} = n_i^{\text{IBNR}} | \mathbf{N}_D) \approx \frac{1}{H} \sum_{i=1}^H \Pr(N_i - N_{iR} = n_i^{\text{IBNR}} | \hat{M}(i, h)). \tag{3.22}$$

Therefore, the conditional distribution of a predicted IBNR count (given \mathbf{N}_D) is in fact a mixture of H Poisson distributions (see, for example, Mikosch, 2006, for definition of mixture distribution). In particular, one can derive the first two moments and the probability mass function of the IBNR prediction,

$$\begin{aligned}
\mathbb{E}(N_i - N_{iR} | \mathbf{N}_D) &= \frac{1}{H} \sum_{i=1}^H Z(i, h), \\
\text{Var}(N_i - N_{iR} | \mathbf{N}_D) &= \frac{1}{H} \sum_{i=1}^H Z(i, h) + \frac{1}{H} \sum_{i=1}^H Z^2(i, h) - \frac{1}{H^2} \left(\sum_{i=1}^H Z(i, h) \right)^2, \\
\Pr(N_i - N_{iR} = x | \mathbf{N}_D) &= \frac{1}{H} \sum_{i=1}^H \frac{e^{-Z(i, h)} Z(i, h)^x}{x!}.
\end{aligned} \tag{3.23}$$

where

$$Z(i, h) = \hat{M}(i, h) F_R((L - i + 0.5)\Delta). \tag{3.24}$$

Furthermore, one can calculate the distribution of the IBNR count, which is an average of the distributions of the H individual Poisson distribution. This allows one to numerically evaluate the quantile function and hence construct a prediction interval of the IBNR count.

4. Fitting illustration

In this section, we illustrate the performance of the RJMCMC filter with simulated data. Here and in the following section, we adopt the parametric assumptions of Example 1, where the size of each shot is exponentially distributed. Because the intensity cannot be observed, and in absence of further evidence about its exact dynamics, we feel that a parsimonious approach must be followed.

Appendix B describes the simulation algorithm. The assumptions and parameters in the simulation are inspired by a domestic motor line of the AUSI data set; see Appendix C for a description of the dataset. Following are the details of our simulated dataset:

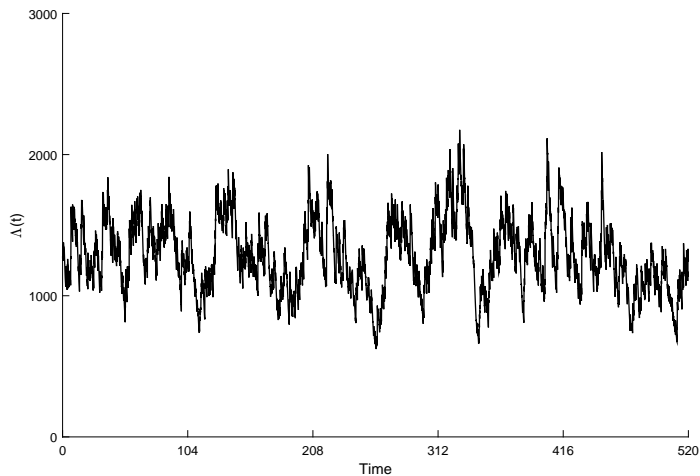


Figure 1: Simulated shot noise trajectory

- we simulate weekly data of ten years (that is, $T = 520$ weeks, $L = 520$ and $\Delta = 1$ week);
- the risk exposure, $\Omega(t)$, is a constant 1;
- the reporting delay distribution is modelled by Assumption 2.3.1 with a threshold of $u = 4$ weeks. There is a mass of $\alpha = 0.4164$ at 0 with a generalised Pareto tail of parameters $(\beta_1 = 0.5564, \beta_2 = 12.9427)$;
- the parameters of the stationary shot noise, $\Lambda'(t)$, are $\rho = 5.9995$, $\eta = 0.0216$ and $k = 0.2176$. Figure 1 shows the simulated shot noise trajectory.

One needs to estimate the parameters of two components: parameters from the shot noise process (that is, ρ , η and k) and parameters from the reporting delay component (that is, α and two Pareto parameters, namely β_1 and β_2). To start with, we obtain the initial estimates of the reporting delay parameters by maximum likelihood estimation. Furthermore, we obtain the initial guesses of the shot noise parameters by matching the theoretical and empirical values of mean, variance and autocorrelation at lag 1 of the weekly claims counts. In particular, we choose to remove the last few data points for initial estimation such that more than 95% of the claims are reported (given the estimated reporting delay parameters). This leads to the removal of 58 data points in obtaining the initial estimates (only).

We re-parametrise the shot noise process such that we estimate parameters $(\rho, \eta, \rho/k)$ directly. This is motivated by the observations that the estimates of ρ and k tend to move in the same direction, although the ratio seems to be relatively stable. This can be explained by the fact that if ρ is over-estimated, then there tends to be more jumps in the shot noise trajectory. Therefore a higher k is required to compensate that. By adopting such a re-parametrisation, one can improve the numerical efficiency in the M-step, where the parameter ρ/k is considered to move relatively independently from ρ .

Given the initial estimates, we use a two-step MCEM algorithm to update the parameters. In the first step, we fix the reporting delay parameters and only update the shot noise parameters. Since the reporting delay parameters are not updated, one can use the conditional likelihood function in (3.20); see Section 3.5. Then we update all parameters simultaneously with a full MCEM-RJMCMC algorithm with conditional likelihood in (3.16). We consider 70,000 RJMCMC simulations in each iteration. The first half are used as the “burn-in” period and 100 simulations of the second half are used in the M step. Figure 2 shows that the estimated integrated process (grey dots) follows closely to the simulated one (dark dots).

The final results of parameter estimates of the shot noise process are shown in Table 5. The initial estimation of the reporting delay parameters is not satisfactory, which is significantly improved in Step 2 where we perform a full MCEM algorithm. This shows the importance of a joint estimation of all parameters.

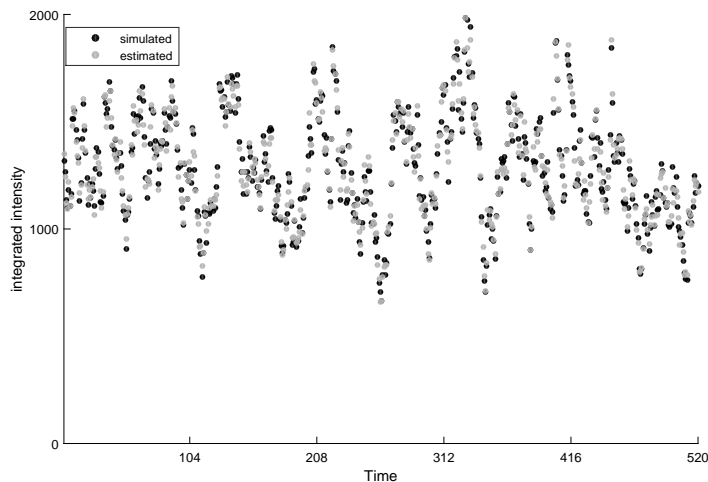


Figure 2: Filtering result

	shot noise parameters			shot noise moments			reporting delay parameters		
	ρ	η	ρ/k	mean	variance	autocor	α	β_1	β_2
true value	5.9995	0.0216	27.571	1276.4	56308	0.8476	0.4164	0.5564	12.9427
Initial	6.4005	0.0214	27.721	1297.0	57572	0.8405	0.4520	0.4068	13.0041
Step 1 (partial MCEM)	7.0881	0.0241	30.670	1272.7	50244	0.8380			
Step 2 (full MCEM)	6.4261	0.0219	28.075	1284.2	55793	0.8412	0.4165	0.5485	13.0675

Table 5: Results of parameter estimation

Furthermore, the estimation of the shot noise parameters also shows improvement in Step 2 compared to that from Step 1. We also notice that the moment matching algorithm is able to provide nice initial estimates. Moreover, the theoretical mean and auto-correlation at lag 1 implied by our final estimate are close to those implied by the true values, although the variance is slightly under-estimated.

We use the RJMCMC simulations in the last step of the MCEM algorithm as the estimate of the underlying shot noise process. This allows us to compute the estimate of the integrated intensity, conditional on which a claim count is a Poisson random variable. Furthermore, this also enables us to predict the counts of IBNR claims; see Section 3.6. In Table 6, we display the lower-left corner of the quarter count triangle. The upper triangle corresponds to observed data and the lower one corresponds to the expected values of the future IBNR counts. Please note that we aggregate the claim counts into quarterly intervals to reduce the size of the dataset and improve readability. The diagonal (that is, the most recent calendar period) of a full quarterly triangle involves future observations, hence we treat this whole diagonal as unobserved (note the last accident quarter is 39 instead of 40). Furthermore, the proportion of motor insurance claims that are reported beyond two years is quite small and therefore presenting only the first two reporting years already covers the majority of the data. The rest of the reporting periods are aggregated together. We have also included a column to display the total INBR counts for each accident quarter, and showed the relative error, that is, the ratio of the difference between the prediction and the true value (from the simulated sample path) over the latter. The small errors in Table 6 show that our algorithm is able to achieve good predictions. In fact, the aggregate prediction value (20466) is very close to the true IBNR count (20600).

Figure 3 shows the results of parameter estimation via MCEM iterations in Step 2. In each plot, the

accident quarter	reporting quarter									IBNR	relative error
	1	2	3	4	5	6	7	8	≥ 9		
32	13097	3166	1296	728	386	242	183	138	490.74	490.74	-0.10
33	11173	2622	1144	573	332	232	168	110.51	493.05	603.56	-0.01
34	11390	2594	1087	566	341	202	155.14	111.03	514.60	780.77	0.02
35	11778	2720	1104	639	374	236.73	161.92	115.88	461.86	976.40	-0.05
36	10464	2498	1082	567	329.73	212.47	145.32	104.00	364.32	1155.84	-0.06
37	8281	1960	840	438.17	260.09	167.60	114.63	82.04	431.97	1494.50	0.05
38	9910	2258	987.50	519.54	308.39	198.72	135.92	97.27	407.63	2654.97	-0.01
39	9352	2154.16	931.84	490.26	291.01	187.52	128.26	91.79	390.43	4665.27	0.01

Table 6: Simulated sample path (upper left triangle) and corresponding expected future claim counts (lower right triangle), aggregated into quarterly intervals. The relative error has been computed with respect to the simulated sample path of the lower right triangle (which we assumed was unobserved when calibrating the model).

value at 0 refers to the initial estimate. The results show nice convergence of all parameters. In particular, the initial estimates of the reporting delay parameters underestimate the distribution, which are corrected in Step 2. The under-estimation of the reporting delay distribution at the initial step results from the fact that the reporting delay model is calibrated independently from the arrival process. This assumes complete observations and ignores the existence of unreported claims, where the IBNR claims have longer reporting delays than those of the reported claims in the same accident periods.

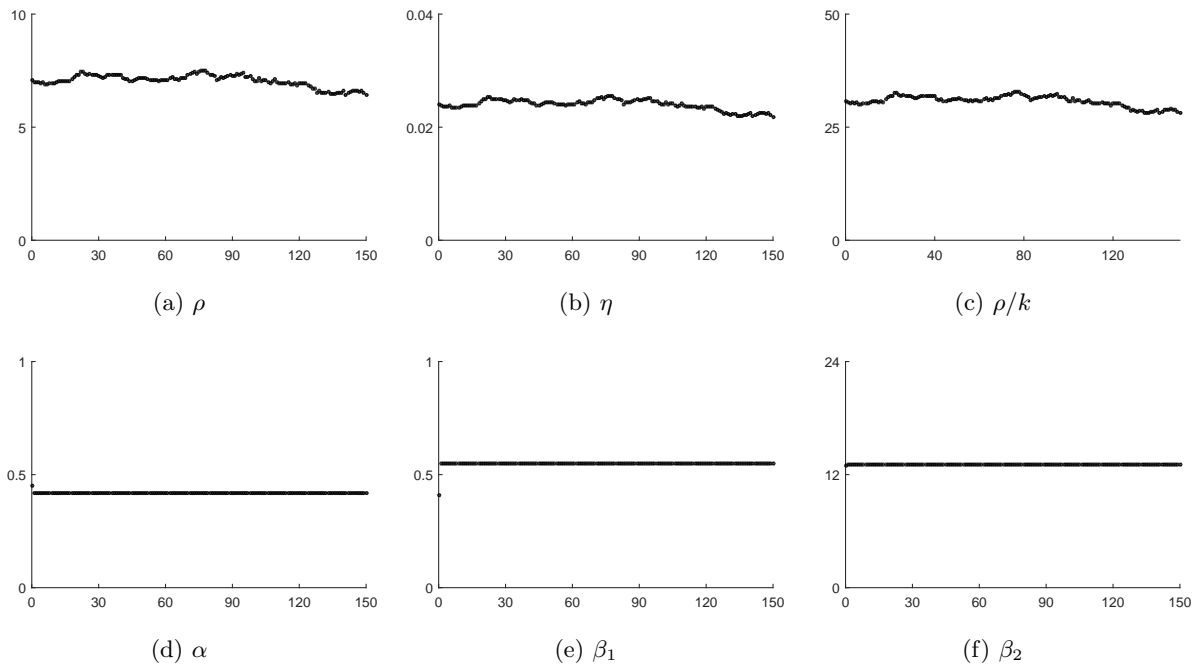


Figure 3: Parameter update via MCEM iterations

Overall, our estimation algorithm is able to filter out the underlying intensity process and estimate the parameters accurately. Furthermore, we are able to provide a close (point) prediction of the IBNR count and estimate the stochastic distribution of the IBNR count.

Remark 4.1. *The initial shot noise trajectory we use for a new MCEM iteration (after the first one) is the*

last RJMCMC simulation in the previous MCEM iteration. Therefore, once the parameter estimates have achieved reasonable convergence, the starting point of the RJMCMC simulation should belong to a suitable range of the stationary distribution of the shot noise trajectory. In such a case, a “burn-in” period may not be necessary. In this research, we do not investigate into the benefit a “burn-in” can bring considering the extra cost of computation. Here, we follow the same approach as Centanni and Minozzo (2006a,b) where a “burn-in” period is adopted.

Remark 4.2. *The number of RJMCMC simulation required in the “burn-in” period is chosen in the following way. Firstly, this number should be proportional to the number of shots in the shot noise trajectory, where the expected number of shots is ρT . Given a relatively good initial estimate of the trajectory, the expected number of simulations one needs to update each shot by changing its position and height once is 5.5 (given that we have 20% chance of choosing each particular move). Furthermore, one should consider that the estimate of ρ may not be accurate and other move types may be involved. Therefore we make a subjective choice of using at least $10\rho T$ (where the actual number can be larger for convenience) RJMCMC simulations in the “burn-in” period.*

5. Analysis of the AUSI data set

In this section, we illustrate our algorithm with a real insurance dataset. The dataset, which is part of the AUSI dataset (refer to Appendix C), corresponds to the Building and Contents business of a major Australian general insurer. Here we have observations of insurance claims and policies information from 01/July/2013 to 04/January/2015. We choose to use the number of policyholders as a measure of exposure for claim counts. Catastrophe claims are flagged in the dataset and have been excluded from our analysis. For confidentiality reasons, we only consider a random sample (a subset chosen randomly) of the claim observations and scale the risk exposure to an arbitrary number.

We start by examining the autocorrelation structure of the claim counts per unit exposure. The autocorrelation of the daily data, presented in Figure 4a, shows that there is a weekly cycle indicated by local peaks of autocorrelations lags at 7 and 14. A shot noise Cox process, whose autocovariance (as well as autocorrelation) is an exponential decay function of the lag (see Equation 2.5), may not provide nice fit in such a case. To overcome this, one can add covariates in the exposure term (see Remark 2.3) corresponding to weekly effects. However, this adds more parameters in the model and may lead to higher parameter uncertainty. Here, we adopt an alternative solution where we consider the weekly data. The autocorrelation of the claim counts per unit exposure (see Figure 4b) behaves more closely to that of a shot noise Cox process. In the rest of the analysis, we use weekly data instead of daily data. Figure 5 presents the observed weekly claim counts N_{iR} (dark dots) and risk exposure Ω^i (grey dots) for week i . Note that exposure is increasing over the period considered.

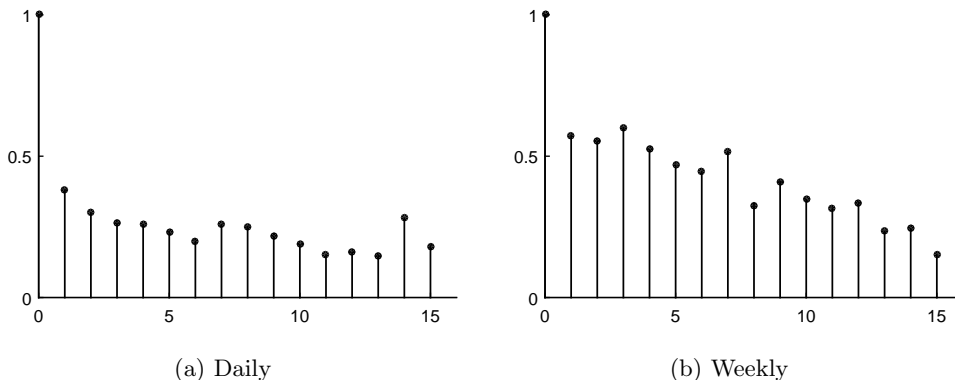


Figure 4: Sample autocorrelation in the AUSI data set

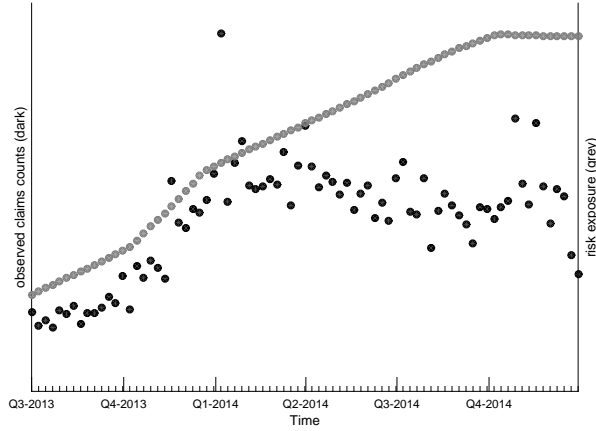


Figure 5: Observed claims counts N_{iR} (dark dots) and non-constant risk exposure Ω^i (grey dots) for week i

We obtain the initial guesses of the shot noise parameters by taking an average of the results of moment matching (where we consider the mean, variance and autocorrelation at lag 1) and negative binomial approximation. The reporting delay data is fitted to several candidate distributions, and the log-normal distribution leads to the highest likelihood. For illustration purpose, we assume that the reporting delay distribution is stationary. We follow the same two-step procedure as that for the simulated dataset. There are 50 and 1000 MCEM iterations in Steps 1 and 2 respectively, with 4600 RJMCMC in each iteration. Only 100 RJMCMC simulations from the second half of each iteration are used in the E step of the MCEM algorithm. The parameter estimation results are summarised in Table 7 and Figure 9 presents the parameter updates in each MCEM iterations in Step 2.

	shot noise parameters			shot noise moments			reporting delay parameters		
	ρ	η	ρ/k	mean	variance	autocor	α	β_1	β_2
Initial	2.9065	0.0211	7.3914	350.7465	15012	0.7579	0.8918	0.9283	1.5612
Step 1 (partial MCEM)	3.7661	0.0267	9.3098	348.9316	11828	0.7479			
Step 2 (full MCEM)	8.7231	0.0372	13.3547	358.8643	8205.1	0.6337	0.8578	1.5801	1.9376

Table 7: Results of parameter estimation

The convergence of the parameters is satisfactory (see Figure 9). Similarly to the observations from the simulated dataset, the reporting delay distribution is under-estimated in Step 1 and the parameter estimates quickly converge to stable levels in Step 2. Figure 6 presents the estimated integrated intensity (grey dots) compared to the observed claims counts (dark dots), together with the IBNR counts (dotted line). We observe that the estimated integrated intensity follows closely to the observed counts for most of the period. The difference between these two grows larger at the end of the observation period, which is caused by higher numbers of IBNR counts. We have also plotted the relative residuals of claim counts (see Figure 7), where we compare expected numbers of claim counts from our filter to the actual observations. 40 out of 79 of the residuals are positive. This further shows that our filtering algorithm is able to capture the movement of the claim frequencies.

We present the counts triangle in Table 8. The upper triangle corresponds to observed counts and the lower triangle consists of the estimated IBNR counts. To improve readability, we aggregate the results into quarterly intervals. This requires the removal of the observations in the last quarter (which involves

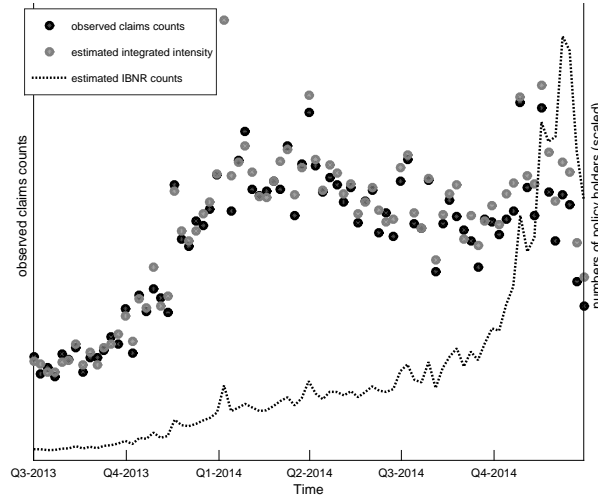


Figure 6: Filtering result

aggregating future observations). Furthermore, we have ignored the first accident week so we have exactly 5 accident quarters. We also notice that the proportion of unreported claims after the maximum reporting delay (as a result of the triangle shape) is material. This is because of a small sample size where the count triangle is not fully developed yet. In this case, our model is able to allow for the IBNR counts beyond the restriction of a fixed size triangle. Figure 8 shows the prediction of the total IBNR count. The expected value of the total IBNR count is 1109.1.

accident quarter	reporting quarter						IBNR
	1	2	3	4	5	≥ 6	
2013-Q3	2039	70	37	12	5	29.86	29.86
2013-Q4	3884	153	50	24	13.85	57.83	71.68
2013-Q1	5931	147	47	32.03	20.46	85.41	137.91
2013-Q2	5196	121	51.15	28.25	18.05	75.34	172.79
2013-Q3	4808	118.95	47.21	26.08	16.66	69.54	278.43

Table 8: The count triangle (aggregated into quarterly intervals)

In conclusion, our filtering results suggest that our shot noise Cox model is adequate for the modelling of the data set we consider. The model allows for an easy prediction of the IBNR counts.

Acknowledgments

The authors acknowledge insightful comments from a reviewer, whose suggestions led to significant improvements of the paper.

This paper (in some cases with other or earlier versions of the data set) was presented at the 6th *Australasian Actuarial Education and Research Symposium* in Perth (Australia), the Actuarial and Financial Mathematics Seminar of the Quantact Lab in Montreal (Canada), the 19th *International Congress on Insurance: Mathematics and Economics* in Liverpool (UK), and the 7th *Australasian Actuarial Education and Research Symposium* on the Gold Coast (Australia). The authors are grateful for constructive comments received from colleagues who attended those events.

This research was partially supported by an Australian Actuarial Research Grant, as well as under Australian Research Council's Linkage Projects funding scheme (project number LP130100723). Furthermore,

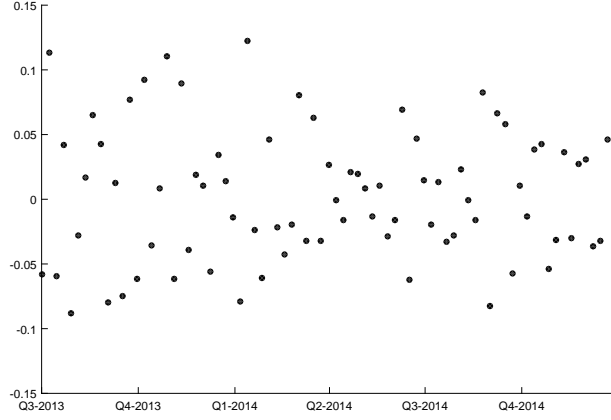


Figure 7: Relative residuals of reported claim counts

Xinda Yang acknowledges financial support from an Australian Postgraduate Award and supplementary scholarships provided by the UNSW Business School, as well as travel funds from a grant of the Natural Science and Engineering Research Council of Canada. The views expressed herein are those of the authors and are not necessarily those of the supporting organisations.

A. Proofs

A.1. Scaling the exposure under Model 2.2.1

Recall Model 2.2.1 and suppose that there exists a homogeneous shot noise process $\Lambda'(t)$ with parameters (ρ, η, k) such that

$$\Lambda'(t) = \Lambda'(0)e^{-kt} + \sum_{j=1}^{J(t)} X_j e^{-k(t-\tau_j)}, \quad t \geq 0, \quad (\text{A.1})$$

where all the notations have the same meaning as in Model 2.1.2 and the intensity process is

$$\Lambda(t) = \Omega(t)\Lambda'(t). \quad (\text{A.2})$$

Define a new exposure process $\Omega'(t) = a\Omega(t)$, which scales $\Omega(t)$ by a constant, then

$$\Lambda(t) = \Omega(t)\Lambda'(t) = \Omega'(t)\frac{\Lambda'(t)}{a} \quad (\text{A.3})$$

with

$$\frac{\Lambda'(t)}{a} = \frac{\Lambda'(0)}{a}e^{-kt} + \sum_{j=1}^{J(t)} \frac{X_j}{a}e^{-k(t-\tau_j)}, \quad t \geq 0. \quad (\text{A.4})$$

Now the characteristic function of X_j/a is

$$\mathbb{E}(e^{\frac{X_j}{a}t}) = \mathbb{E}(e^{X_j \frac{t}{a}}) = \frac{\eta}{\eta - t/a} = \frac{a\eta}{a\eta - t}, \quad (\text{A.5})$$

which suggests an exponential random variable with parameter η . Furthermore, the characteristic function of $\frac{\Lambda'(0)}{a}$ is

$$\mathbb{E}(e^{\frac{\Lambda'(0)}{a}t}) = \mathbb{E}(e^{\Lambda'(0)\frac{t}{a}}) = \left(\frac{\eta}{\eta - t/a}\right)^{\rho/k} = \left(\frac{a\eta}{a\eta - t}\right)^{\rho/k}, \quad (\text{A.6})$$

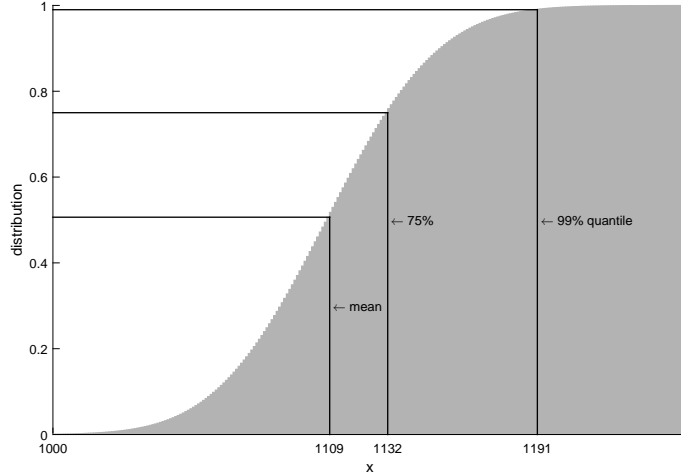


Figure 8: Probability density function the total IBNR count

which suggests a Gamma random variable with parameter $(\rho/k, a\eta)$. Therefore $\Lambda'(t)/a$ is still a shot noise process but with parameters $(\rho, a\eta, k)$.

B. Simulation algorithm

We extend the algorithm developed in Avanzi et al. (2016) to simulate the trajectory of a shot noise process. In this section, we will present the algorithm to generate discrete observations of claims count in the presence of reporting delays. We use the same notations as in Section 3. For the i^{th} ($i = 1, 2, \dots, L$) accident period,

- calculate the integrated intensity, M_i with Equation (3.15);
- simulate the ultimate claim count as a Poisson random variable with intensity M_i ;
- allocate the ultimate claims into reporting bins (where we treat the unreported claims as falling into a special reporting bin) with a multinomial distribution. The probability of allocating a claim to the j^{th} reporting period is

$$p_i(j) = \begin{cases} F_R(0.5\Delta) & j = 1 \\ F_R((j - 0.5)\Delta) - F_R((j - 1.5)\Delta) & j > 1, \end{cases} \quad (\text{B.1})$$

and the probability of a claim reporting after T (that is, it is an IBNR claim) is

$$p_i^{\text{IBNR}} = 1 - F_R((L - i + 0.5)\Delta). \quad (\text{B.2})$$

One can show that $\sum_{j=1}^{L-i+1} p_i(j) + p_i^{\text{IBNR}} = 1$, that is, a claim is either reported in one of the reporting periods or being an IBNR claim.

Remark B.1. *Our simulation algorithm is consistent with the likelihood function we use in our estimation algorithm (see Section 3). As discussed in Remark 3.2, one may adopt a different estimation algorithm and this may require a corresponding simulation method. For example, one can refer to Centanni and Minozzo (2006b) and Avanzi et al. (2016) for algorithms of simulating continuous observations of a homogeneous shot noise Cox process.*

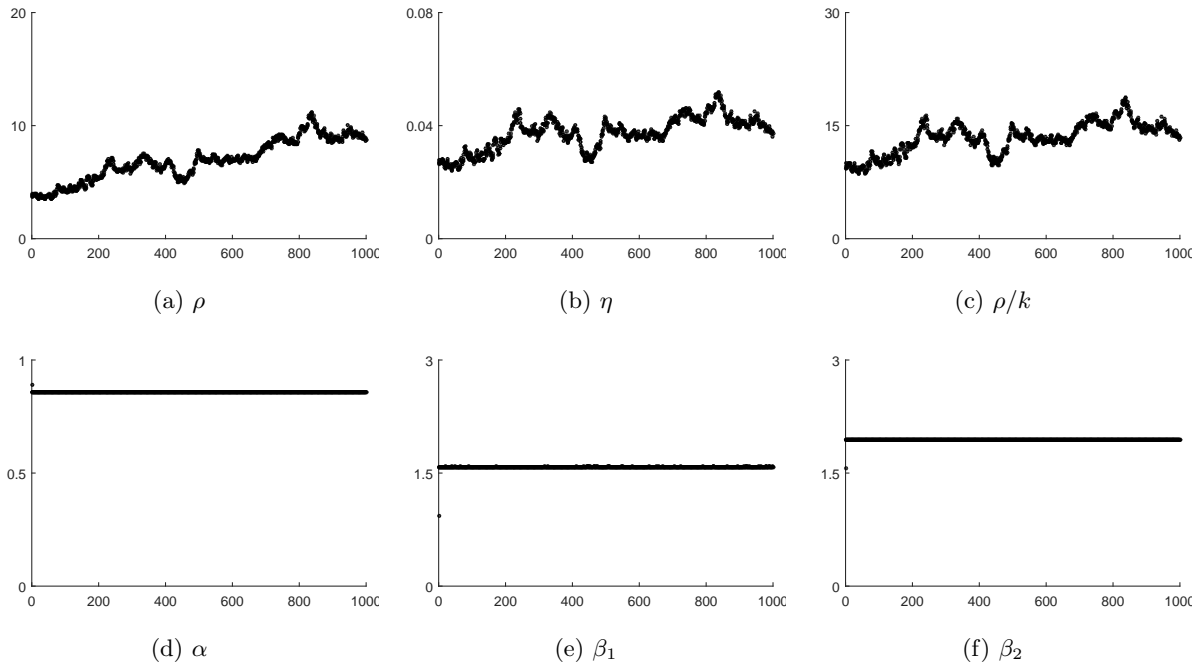


Figure 9: Parameter estimates via EM iterations

C. The AUSI dataset

The dataset was developed as part of a Linkage Project grant awarded by the Australian Research Council (ARC) until 2016 for a project titled *Modelling claim dependencies for the general insurance industry with economic capital in view: an innovative approach with stochastic processes*. It is referred to as the **AUSI data set**, an acronym of the names of the project partners (Allianz Australia Insurance Ltd, UNSW Australia, Suncorp Metway Ltd, and Insurance Australia Group Ltd).

The data set currently includes data from two insurers. LoBs covered comprise two property and two casualty lines, as follows, referred to initially in Australian nomenclature, and then US:

- Home (Homeowners), including both Buildings and Contents coverages;
- Private Motor (Private Auto Property Damage);
- Compulsory Third Party (“CTP”) (Auto Bodily Injury);
- Public Liability.

Data are provided in respect of a defined period (“investigation period”) in essentially the standard format, consisting of:

- A **policy file**, containing detail of each policy underwritten during the investigation period, e.g. dates of inception and expiry, sum insured, etc.;
- A **claim header file**, containing static information on each claim notified during the investigation period, e.g. dates of claim occurrence and notification, finalisation date (if finalised), claim state, etc.;
- A **claim transaction file**, containing detail of each claim transaction occurring during the investigation period, e.g. transaction date, type of transaction (claim payment or case estimate adjustment), amount of transaction, payment type (e.g. peril or head of damage under which payment is made), claim status (open/closed) after the transaction, etc.

D. Table of notations

symbol	definition
$N(t)$	insurance claims count process
$\Omega(t)$	risk exposure at time t
$\Lambda(t)$	stochastic intensity of claim arrival
$M_i(t)$	integral of the stochastic intensity over the i^{th} sub-period
$\Lambda'(t)$	stationary shot noise process ($\Lambda(t) = \Omega(t)\Lambda'(t)$)
(ρ, η, k)	frequency, severity and decay parameters of the shot noise process where the expected shot size is $1/\eta$
(ρ_0, η_0, k_0)	initial estimates of (ρ, η, k)
$J(t)$	arrival process of shots in the shot noise intensity
τ_j	arrival time of the j^{th} shot
X_j	severity of the j^{th} shot
R	reporting delay of a claim
u	threshold parameter of the reporting delay distribution
α	mass of reporting delays at 0
β	vector of parameters of the continuous reporting delay distribution above u
n	number of shots in a shot noise trajectory
ϕ	parameter vector to be estimated (e.g. $\phi = (\rho, \eta, k)$ in the case of a shot noise process with exponentially distributed shots)
$f_X(\cdot)$	probability density function of a random variable X
$F_X(\cdot)$	cumulative distribution function of a random variable X
Δ	length of each sub-period
L	number of sub-periods
T	length of the data observation period ($T = L \times \Delta$)

Table A: Notation used in the paper

References

- Albrecher, H., Asmussen, S., 2006. Ruin probabilities and aggregate claims distributions for shot noise Cox processes. *Scandinavian Actuarial Journal* 2006 (2), 86–110.
- Antonio, K., Plat, R., 2014. Micro-level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal* 2014 (7), 649–669.
- Arjas, E., 1989. The claims reserving problem in non-life insurance: Some structural ideas. *ASTIN Bulletin* 19 (2), 139–152.
- Avanzi, B., Liu, C.-Y., Wong, B., 2016. On the modelling of claims arrival processes with doubly stochastic (Cox) processes.
- Badescu, A. L., Lin, X. S., Tang, D., 2015. A marked Cox model for IBNR claims: Model and theory. <http://arxiv.org/abs/1512.06273>.
- Badescu, A. L., Lin, X. S., Tang, D., 2016. A marked Cox model for the number of IBNR claims: Estimation and application. <http://ssrn.com/abstract=2747223>.
- Biffis, E., 2005. Affine processes for dynamic mortality and actuarial valuations. *Insurance: Mathematics and Economics* 37 (3), 443–468.
- Björk, T., Grandell, J., 1988. Exponential inequalities for ruin probabilities in the Cox case. *Scandinavian Actuarial Journal* 1988 (1-3), 77–111.
- Centanni, S., Minozzo, M., 2006a. Estimation and filtering by reversible jump MCMC for a doubly stochastic Poisson model for ultra-high-frequency financial data. *Statistical Modelling* 6 (2), 97–118.
- Centanni, S., Minozzo, M., 2006b. A Monte Carlo approach to filtering for a class of marked doubly stochastic Poisson processes. *Journal of the American Statistical Association* 101 (476), 1582–1597.
- Cont, R., Tankov, P., 2004. *Financial modelling with jump processes*. Chapman & Hall/CRC, London.
- Cox, D., 1955. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B (Methodological)*, 129–164.
- Dassios, A., Jang, J., 2003. Pricing of catastrophe reinsurance and derivatives using the Cox process with shot noise intensity. *Finance and Stochastics* 7 (1), 73–95.
- Dassios, A., Jang, J., 2005. Kalman-bucy filtering for linear systems driven by the Cox process with shot noise intensity and its application to the pricing of reinsurance contracts. *Finance and Stochastics* 42 (1), 93–107.

- de Jong, P., Heller, G. Z., 2008. Generalized linear models for insurance data. Cambridge, UK ; New York : Cambridge University Press.
- Denuit, M., Maréchal, X., Pitrebois, S., Walhin, J., 2007. Actuarial modelling of claim counts. Wiley Online Library.
- England, P., Verrall, R., 2002. Stochastic claims reserving in general insurance. *British Actuarial Journal* 8 (3), 443–518.
- Frey, R., Runggaldier, W., 2001. A nonlinear filtering approach to volatility estimation with a view towards high frequency data. *International Journal of Theoretical and Applied Finance* 4 (2), 199–210.
- Gelman, A., Jones, G., Brooks, S., Meng, X.-L. (Eds.), 2011. Handbook of Markov chain Monte Carlo. Boca Raton : CRC Press.
- Grandell, J., 1976. Doubly stochastic Poisson processes. Springer-Verlag Berlin, Heidelberg, New York.
- Green, P., 1995. Reversible jump Markov chain Monte Carlo computation and bayesian model determination. *Biometrika* 82 (4), 711–732.
- Jang, J., 2004. Martingale approach for moments of discounted aggregate claims. *Journal of Risk and Insurance* 71, 201–211.
- Jang, J., Fu, G., 2008. Transform approach for operational risk modelling: VaR and TCE. *Journal of Operational Risk* 3, 45–61.
- Jang, J., Fu, G., 2012. Measuring tail dependence for aggregate collateral losses using bivariate compound shot-noise Cox process. *Applied Mathematics* 12A, 2191–2204.
- Jewell, W. S., 1989. Predicting IBNR events and delays. *ASTIN Bulletin* 19 (1), 25–55.
- Joe, H., 1997. Multivariate models and dependence concepts. Chapman & Hall, London.
- Klugman, S. A., Panjer, H. H., Willmot, G. E., 2012. Loss models : from data to decisions. Hoboken, N.J.:Wiley.
- Lando, D., 1998. On Cox processes and credit risky securities. *Review of Derivatives research* 2 (2), 99–120.
- Larsen, C. R., 2007. An individual claim reserving model. *ASTIN Bulletin* 37 (1), 113–132.
- Mack, T., 1993. Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin* 23 (2), 213–225.
- Mikosch, T., 2006. Non-life insurance mathematics: an introduction with stochastic processes. Springer.
- Norberg, R., 1993. Prediction of outstanding liabilities in non-life insurance. *ASTIN Bulletin* 23 (1), 95–115.
- Norberg, R., 1999. Prediction of outstanding liabilities - ii model variations and extensions. *ASTIN Bulletin* 29 (1), 5–27.
- Pigeon, M., Antonio, K., Denuit, M., 2013. Individual loss reserving with the multivariate skew normal framework. *ASTIN Bulletin* 43 (3), 399–428.
- Pigeon, M., Antonio, K., Denuit, M., 2014. Individual loss reserving using paid-incurred data. *Insurance: Mathematics and Economics* 58, 121–131.
- Rydén, T., 1996. An EM algorithm for estimation in Markov-modulated Poisson processes. *Computational Statistics & Data Analysis* 21 (4), 431–447.
- Schmidt, T., 2014. Catastrophe insurance modeled by shot-noise processes. *Risks* 2 (1), 3–24.
- Schrager, D., 2006. Affine stochastic mortality. *Insurance: Mathematics and Economics* 38 (1), 81–97.
- Taylor, G., 2000. Loss reserving: an actuarial perspective. Huebner International Series on Risk, Insurance and Economic Security. Kluwer Academic Publishers.
- Taylor, G., McGuire, G., Sullivan, J., 2008. Individual claim loss reserving conditioned by case estimates. *Annals of Actuarial Science* 3 (1-2), 215–256.
- Werner, G., Modlin, C., 2010. Basic ratemaking. Casualty Actuarial Society.
- Wüthrich, M., Merz, M., 2008. Stochastic claims reserving methods in insurance. John Wiley & Sons.
- Zhao, X., Zhou, X., 2010. Applying copula models to individual claim loss reserving methods. *Insurance: Mathematics and Economics* 46 (2), 290–299.
- Zhao, X. B., Zhou, X., Wong, J. L., 2009. Semiparametric model for prediction of individual claim loss reserving. *Insurance: Mathematics and Economics* 45 (1), 1–8.