

# CLAIMS RESERVING IN CONTINUOUS TIME; A NONPARAMETRIC BAYESIAN APPROACH

SVEND HAASTRUP AND ELJA ARJAS

*Laboratory of Actuarial Mathematics, Universitetsparken 5,  
2100 København Ø, Denmark*

*Department of Applied Mathematics and Statistics, University of Oulu,  
Linnanmaa 90570, Oulu, Finland*

## ABSTRACT

Occurrences and developments of claims are modelled as a marked point process. The individual claim consists of an occurrence time, two covariates, a reporting delay, and a process describing partial payments and settlement of the claim. Under certain likelihood assumptions the distribution of the process is described by 14 one-dimensional components. The modelling is nonparametric Bayesian. The posterior distribution of the components and the posterior distribution of the outstanding IBNR and RBNS liabilities are found simultaneously. The method is applied to a portfolio of accident insurances.

## KEYWORDS

IBNR and RBNS liabilities, marked point process, missing data, Markov chain Monte Carlo.

## 1. INTRODUCTION

A major issue in non-life insurance is prediction of outstanding liabilities. Outstanding liabilities are traditionally divided into occurred but not reported (IBNR) claims and reported but not settled (RBNS) claims. At each time the insurance company has to predict the outstanding liabilities and provide a reserve correspondingly.

A vast number of articles have been written on the subject. In most models the data are assumed to be discretized. Arjas (1989), Jewell (1989), Norberg (1993a,b,c) and Hesselager (1994) model in continuous time. Norberg (1993a) describes occurrence and development of the claims by a marked Poisson process. In Norberg (1993c) and Kirkegaard (1994) different parametric specifications of the model are considered, and real insurance data are analyzed. Furthermore, Norberg (1993a) considers an extended model where the occurrence intensity is assumed to be a stochastic process, and he finds the best linear predictor of the outstanding liabilities.

The present paper deals, by way of a case-study, with a portfolio of accident insurances. The model used is close to that of Norberg (1993a). The claims process generating occurrences, covariates and developments of the claims is modelled as a marked Poisson process. Our approach to estimation and prediction is nonpara-

metric Bayesian. Adopting the methods of Arjas and Gasbarra (1994), the distribution governing the process of occurrences, covariates and developments is modelled by piecewise constant conditional intensities. The intervals on which the intensities are constant, and the values (levels) of the intensities on the different intervals, are then viewed as model parameters. In principle, such a parameter space is of infinite dimension. A prior distribution (a distribution on the parameter space) is attached to the intensities of the claims process.

Both the model parameters and the outstanding liabilities (RBNS and IBNR claims) are unknown. The problem is to find the conditional distribution of such unknowns given the observations. This distribution will be called the posterior; it covers the conditional distribution of the unknown parameters which by standard usage is called posterior, and the conditional distribution of future observables which by standard usage is called predictive distribution. In complex models, it is often difficult to identify the posterior. The posterior can always be determined up to proportionality, but it can be difficult to normalize, which is necessary e.g. for the calculation of means. Recently, a technique called Markov chain Monte Carlo (MCMC) integration has been used to solve this problem numerically in connection with large statistical models. A general review of the topic can be found in Smith and Roberts (1993). The idea is to generate a Markov chain which has the posterior distribution as its equilibrium distribution. Using such a chain, all quantities of interest can be estimated/predicted. For example, at each step of the chain a new value of the RBNS claims is sampled; the empirical distribution of these sampled RBNS claims then converges towards the predictive (posterior) distribution of the RBNS claims.

Section 2 below describes the data. A claim is described by an occurrence time, two covariates, and a development. The development contains a reporting delay, a settlement delay, and a partial payment process containing the partial payments made from reporting to the settlement of the claim.

In Section 3 distributional assumptions are made. Claims are assumed to occur in accordance with a Poisson process, and covariates and developments are modelled as *marks* associated with the occurrences. The distribution of both occurrences and marks is specified by piecewise constant conditional intensities, and a prior distribution of these intensities is chosen.

Section 4 describes the MCMC algorithm (sampling algorithm). A Markov chain with the desired properties is generated. The algorithm is close to the one described by Arjas and Gasbarra (1994).

Section 5 describes the estimations and predictions. Using the Markov chain generated in the sampling algorithm we approximate both the distribution of the claims process and the distribution of the outstanding liabilities.

## 2. THE DATA

The data are a portfolio of accident insurances. In the following we first describe the general structure of the data, and then go on with a detailed description of the present data set.

### 2.1. Structure of the data

We use the set-up of Norberg (1993a). By a claim we understand a combination of a time of occurrence, a set of covariates, and a development. Let time 0 be the initial time, and let  $\tau$  be the time at which IBNR and RBNS liabilities are to be predicted. Each individual policy is described by the covariates  $(s, a)$  denoting sex and age of the policy holder. For each combination of calendar time  $t$  in  $(0, \tau]$ , sex  $s$  in  $\{\text{male, female}\}$ , and age  $a$  in  $\{1, 2, \dots\}$ , the number of policies covered is denoted

$$w(t, s, a).$$

The function  $w$  is called the exposure rate.

Using the notation of Norberg (1993a) the *development* of a claim can be described by

$$(U, (Y(v))_{v \in [0, V]}),$$

where  $U$  is the waiting time from occurrence until notification (the reporting delay),  $V$  is the waiting time from notification until final settlement, and  $Y(v)$  is the amount paid  $v$  time units after notification. The final claim amount is  $Y(V)$ , which is called  $Y$  for short. Moreover, when a claim is reported, the occurrence time  $T$  and the covariates sex and age  $(S, A)$  are known, age being the age at the time of occurrence. So the complete description of a claim is

$$(T, S, A, U, X) \quad \text{where, for short,} \quad X = (Y(v))_{v \in [0, V]}.$$

The *partial payments process*  $X = (Y(v))_{v \in [0, V]}$  consists of a series of lump payments. An illustration can be found in Norberg (1993a), Figure 2.

Not all claims which occurred before time  $\tau$  are actually observed. At time  $\tau$  we have only observed the reported claims, i.e. claims with  $T + U \leq \tau$ , and for each of these we only know the development up to time  $\tau$ . This means that for a reported claim we always know  $(T, S, A, U)$ . Furthermore we observe,

$$\text{for a settled claim,} \quad (V, (Y(v))_{v \in [0, V]}), \quad \text{and} \quad (2.1)$$

$$\text{for a reported but not settled claim,} \quad (1_{(V \leq \tau - T - U)} = 0, (Y(v))_{v \in [0, \tau - T - U]}), \quad (2.2)$$

where  $1_{(V \leq \tau - T - U)} = 0$  indicates that the claim is not settled. Note that the above is just the partial payment process,  $(Y(v))_{v \in [0, V]}$ , censored at calendar time  $\tau$ .

### 2.2. The present data set

The data are a portfolio of accident insurances, supplied by a Danish insurance company. There are four different kinds of claims; dental claims, spectacles claims, disability claims, and death claims. We have chosen to model the dental claims, and look at leisure time cover only. Claims which occurred between January 1, 1982 and December 31, 1990 and which were reported before March 3, 1992 are observed and contained in the data set. Here we have chosen to consider  $\tau = 6$  years. This means

that only data recorded by December 31, 1987 are considered. The rest of the data can be used to check the validity of our model.

Some modifications are needed to give the data the desired structure. The exposure is only known by years, not days. An estimate of the exposure rate  $w$  is obtained by interpolation. Information about claim developments is almost as we need them. Some of the claims records do not contain information about the covariates, and for simplicity such claims records (as well as exposure) were removed. Furthermore, times of occurrence, notification, partial payments and settlements are only known by days. However, we shall model occurrences and developments in continuous time, and view the reported times of occurrence, notification, etc. as approximations to the true values. Finally, we have chosen to modify the observed partial payment processes. Looking at a partial payment process, two payments made during the same day on the same claim are lumped together and viewed as a single payment. Furthermore, about 2% of the partial payments were negative. Some of these had a corresponding positive payment on the same day, and in that case both these payments were disregarded. Negative payments that did not have a corresponding positive payment never exceeded the accumulated claim amount paid, and they were set off against the previous positive payments. For example, a recorded partial payment process containing two payments, one of 1200 DKK at January 5, 1985 and one of -500 DKK at January 16, 1985, is transformed into a partial payment process containing only one partial payment of 700 DKK made at January 5, 1985. Lumping payments that are made during the same day is easy to justify, but to set off negative payments against previous positive payments requires some comments. We have chosen December 31, 1987 as the time of prediction. However, in our modification of the observed partial payments processes we included all partial payments made before March 3, 1992. This means that we use negative payments in the future to modify payments in the past. An insurance company can not do that as it does not know about payments in the future. Therefore, insurance companies with recorded negative payments will tend to slightly overestimate the outstanding liabilities if they use our model.

After these modifications, our data contain 434.000 exposure years. There are 2806 reported claims; of these 2191 are settled and 617 unsettled. There are 3718 observed partial payments, and they add up to 10.040.000 DKK.

### 3. DISTRIBUTIONAL ASSUMPTIONS

We model the distribution of all claims which occur in our portfolio and their full development. Let

$$(T_i, S_i, A_i, U_i, X_i)_{i \geq 1} \quad (3.1)$$

denote these claims. It is important to note that many of the claims (3.1) are not observed completely, and some of them not at all. As mentioned in the previous section, we only observe the reported claims and their development is censored at calendar time  $\tau$ .

A Bayesian model is used. The modelling is done in two steps. First, the distribution of the claims process (3.1) is described by intensities, henceforth referred to as *components*. Our model will have 14 such components. Some describe the distribution of the occurrences and covariates, some describe the distribution of the reporting delays, and some the distribution of the partial payments processes. Then, a prior distribution is chosen. The intensities (components) are assumed to be piecewise constant. The intervals on which the intensities are constant, and the values (levels) of the intensities on the different intervals are the parameters. A prior distribution (a distribution on the parameter space) is attached to the intensities.

The two steps are described below. The first step contains ‘likelihood assumptions’, and here the major restrictions are made. The second step contains ‘prior assumptions’. Since we have only little knowledge at hand, we try to add only little structure in our choice of prior.

### 3.1. Likelihood assumptions

At calendar time  $t$  there are  $w(t, s, a)$  policies with sex  $s$  and age  $a$  in the portfolio. We assume that for an individual policy, claims occur according to a Poisson process with intensity

$$f(t, s, a), \quad t \in (0, \tau]$$

(age changing once a year). As a consequence we get that, amalgamating all policies in our portfolio, the occurrence times and covariates follow a marked Poisson process with intensity

$$w(t, s, a)f(t, s, a), \quad (t, s, a) \in (0, \tau] \times \{\text{male, female}\} \times \{1, 2, \dots\}.$$

It is assumed that the intensity  $f$  can be written as

$$f(t, s, a) = f_1(t)f_2(s, a). \quad (3.2)$$

In this way the distribution of occurrences and covariates is described by three components; a calendar time effect  $f_1$ , an age effect for males  $f_2(\text{male}, \cdot)$ , and an age effect for females  $f_2(\text{female}, \cdot)$ .

The development of a claim contains a reporting delay and a partial payment process. Following Norberg (1993a), we assume that the distribution of the development of a claim depends on the past history of the process only through the associated occurrence time and covariates. This kind of development distribution is called *position dependent*, see Karr (1991). In the following the distribution of the development of a claim, which occurred at time  $t$  and has covariates  $(s, a)$ , is described by the distribution of the reporting delay  $U$ ,  $P_{U|t,s,a}$ , and the distribution  $P_{X|u,t,s,a}$  of the partial payment process  $X$  given  $U = u$ .

The distribution  $P_{U|t,s,a}$  is assumed to be absolutely continuous (with respect to the Lebesgue measure). It is modelled by the corresponding hazard rate

$$g_{U|t,s,a}(u) = g(u), \quad (3.3)$$

which is assumed independent of both the time of occurrence and the covariates;

that a probability  $P$  on  $\mathcal{R}_+$  has hazard rate  $g$  means that  $g(u)du = P(du)/P((u, \infty))$ . There are many possible extensions. For example, we could assume that  $P_{U|t,s,a}$  has hazard rate

$$g_{U|t,s,a}(u) = g_1(s, u)g_2(s, a)g_3(t),$$

containing 5 components. Both sexes have then their individual reporting delay component and age component, and then there is a calendar time component. Such an extension would cause no mathematical problems, but the computational effort would increase.

The partial payment process  $X = (Y_\nu)_{\nu \in [0, \nu]}$  arising from a single claim is a jump process. Let

$$\mathcal{H}_\nu = \left\{ T, S, A, U, (Y(\nu'), 1_{(\nu > \nu')})_{\nu' \in [0, \nu]} \right\}$$

be the history of a reported but not settled claim; here  $\nu$  is the time since the claim was reported. Let  $dy$  denote a small interval of length  $dy$  around  $y$ . The distribution of  $X$  is described using intensities. Let  $h_{se}(\nu | \mathcal{H}_{\nu-})$  be the intensity of settling at time  $\nu$  without a partial payment at time  $\nu$ , let  $h_{sep}(\nu, dy | \mathcal{H}_{\nu-})$  be the intensity of settling at time  $\nu$  with partial payment of size  $dy$ , and finally let  $h_p(\nu, dy | \mathcal{H}_{\nu-})$  be the intensity at time  $\nu$  of having a partial payment of size  $dy$  without settlement. We have to decide how the intensities  $h_{se}$ ,  $h_{sep}$  and  $h_p$  depend of the history  $\mathcal{H}_{\nu-}$  just before time  $\nu$ .

The following information can be derived from  $\mathcal{H}_{\nu-}$ :

$N_\nu$  number of partial payments in  $[0, \nu)$ ,

$T_\nu$  time since the latest partial payment if any, else  $T_\nu = \nu$ .

We assume that the intensity of settling only is

$$h_{se}(\nu | \mathcal{H}_{\nu-}) = \begin{cases} h_{se}^0(\nu) & \text{if } N_\nu = 0 \\ h_{se}^1(T_\nu) & \text{if } N_\nu > 0, \end{cases} \quad (3.4)$$

the intensity of settling with a partial payment of size  $dy$  is

$$h_{sep}(\nu, dy | \mathcal{H}_{\nu-}) = \begin{cases} h_{sep}^0(\nu)p_{sep}^0(y)dy & \text{if } N_\nu = 0 \\ h_{sep}^1(T_\nu)p_{sep}^1(y)dy & \text{if } N_\nu > 0, \end{cases} \quad (3.5)$$

and the intensity of having a partial payment of size  $dy$  without settlement is

$$h_p(\nu, dy | \mathcal{H}_{\nu-}) = \begin{cases} h_p^0(\nu)p_p^0(y)dy & \text{if } N_\nu = 0 \\ h_p^1(T_\nu)p_p^1(y)dy & \text{if } N_\nu > 0. \end{cases} \quad (3.6)$$

As a consequence, partial payments are distributed according to the densities  $p_{sep}^0$ ,  $p_{sep}^1$ ,  $p_p^0$  and  $p_p^1$ . These will be described by their corresponding hazard rates denoted  $q_{sep}^0$ ,  $q_{sep}^1$ ,  $q_p^0$  and  $q_p^1$ .

The assumptions (3.4)-(3.6) need some comments. The distribution of the partial payment process  $P_{X|u,t,s,a}$  is assumed to be independent of both the time of occur-

rence  $t$ , the reporting delay  $u$  and the covariates  $(s, a)$ . The intensities of the partial payment process depend on the past history of the process only through the number of partial payments  $(N_v)$  and the time since latest partial payment if any, else the time since notification. These assumptions are not crucial, and at this stage the model has many possible extensions. For example, it could be reasonable to assume that the size of a partial payment depends on the sizes of the previous partial payments. And often it is reasonable to assume that the size of a payment also depends on the time since the latest partial payment (if any). Such extensions can be done without mathematical difficulties, but the computational effort would increase. For example, making the size of a partial payment dependent on the cumulated amount of the previous payments,  $Y(v-)$ , can be done assuming that the intensities  $q_{sep}^1$  and  $q_p^1$  have the structure

$$q_{sep}^1(y | \mathcal{H}_{v-}) = q_{sep}^{1a}(y)q_{sep}^{1b}(Y(v-))$$

and

$$q_p^1(y | \mathcal{H}_{v-}) = q_p^{1a}(y)q_p^{1b}(Y(v-)).$$

Even the product assumptions made here are not that crucial.

### 3.2. Prior assumptions

In the previous subsection the distribution of the claims process is described by 14 components (intensities). These are

$$f_1, f_2(\text{male}, \cdot), f_2(\text{female}, \cdot), g,$$

$$h_{sep}^0, q_{sep}^0, h_p^0, q_p^0, h_{se}^0, h_{sep}^1, q_{sep}^1, h_p^1, q_p^1, h_{se}^1.$$

Now a prior distribution is to be chosen. The intensities are assumed piecewise constant. The intervals on which the intensities are constant, and the values (levels) of the intensities on the different intervals are the parameters. A prior distribution (a distribution on the parameter space) is attached to the intensities. The main idea is taken from Arjas and Gasbarra (1994). Following their notation, we shall denote unknown parameters by Latin letters and parameters in the prior distributions, the so-called hyperparameters, by Greek letters.

To begin with we look at the calendar time effect  $f_1$ . It is assumed to have a piecewise constant structure

$$f_1(t) = \sum_{j \geq 0} 1_{(S_j < t \leq S_{j+1})} b_j, \tag{3.7}$$

where  $b_j$ , given  $(b_0, \dots, b_{j-1})$ , follows a lognormal distribution with parameters  $(\log(b_{j-1}), \sigma_{f_1}^2)$  denoting the mean and the variance in the associated normal distribution, and  $0 = S_0 < S_1 < S_2 < \dots$  follows a Poisson process with intensity  $\lambda_{f_1}$ . In short, we write

$$b_j | (b_0, \dots, b_{j-1}) \stackrel{D}{=} \text{LogN}(\log(b_{j-1}), \sigma_{f_1}^2), \quad (S_i)_{i \geq 0} \stackrel{D}{=} \text{Poisson}(\lambda_{f_1}). \tag{3.8}$$

For uniqueness it is assumed that the initial level  $b_0$  is 1.

The prior structure (3.7)-(3.8) is essentially used to model all 14 components. A prior should reflect the knowledge one has about the problem at hand, and it is therefore reasonable to discuss what kind of prior information the structure (3.7)-(3.8) will represent. To save notation we drop the subscript  $f_1$  on the hyperparameters.

The intensity  $f_1$  is by (3.7) assumed to be a positive simple function (a piecewise constant function). The prior distribution of  $f_1$  should therefore be a distribution on the space of positive simple functions. The prior (3.8) is a possible choice. The prior expected number of changes in level per year is  $\lambda$ , and the levels have a log-martingale structure,

$$\log(b_j) = \log(b_{j-1}) + \varepsilon_j,$$

where the  $\varepsilon$ 's are iid normally distributed with zero mean and variance  $\sigma^2$ . A small (large) value of  $\sigma^2$  corresponds to a high (low) correlation between the levels at different times. We consider this choice of prior very vague. It is our experience that, with a reasonable amount of data, the data will 'speak for themselves'. By adjusting the values of the hyperparameters we can, however, control the smoothness of the estimate, the estimate being the posterior mean, say. Now, our prior also has some weaknesses. The prior distribution of  $f_1$  is not stationary. The median in the prior distribution of  $b_j$  is  $b_0$ , but the mean is  $b_0 \exp(j\sigma^2/2)$ . This could be a problem if we wish to predict the occurrence intensity in the years to come. There we have no data, and if we choose to predict using the posterior mean, we will therefore get an increasing estimator. The reader might ask: why not control the mean by multiplying  $b_j$  with  $\exp(-j\sigma^2/2)$ ? If that is done, then it can be shown that  $b_j$  converges almost surely to 0, and new difficulties arise. As an alternative prior one could keep the distributional assumptions about the jump times, but assume that the levels are distributed as

$$b_j = \bar{b} \zeta_j, \text{ where } \bar{b} \stackrel{D}{=} \text{Gamma}(\gamma, \delta), \text{ and } \zeta_j \text{ are iid } \text{Gamma}(\kappa, 1/\kappa). \quad (3.9)$$

From a computational point of view, this prior gives no difficulties. Furthermore, it is stationary. The assumption is that the intensity varies around a level  $\bar{b}$ . It does not allow for permanent changes in the intensity. As a consequence, the prior (3.9) adds more structure to the intensity (smaller variance), and that can be useful when predicting the future occurrence intensity. We are going to use prior (3.7)-(3.8). The prior is chosen in an attempt to add only little prior information.

The age effects  $f_2(\text{male}, \cdot)$  and  $f_2(\text{female}, \cdot)$  are modelled as discrete versions of the construction (3.7)-(3.8). We assume that

$$f_2(\text{male}, a) = \sum_{j \geq 0} 1_{(a \in \{K_j + 1, \dots, K_{j+1}\})} c_j, \quad a = 1, 2, 3, \dots$$

where  $(c_j)_{j \geq 1}$  are modelled as  $(b_j)_{j \geq 1}$  above, but with new hyperparameters  $\sigma_{f_{2m}}^2$  and  $\lambda_{f_{2m}}$ , and where  $(K_{j+1} - K_j)_{j \geq 1}$  are assumed to be iid geometrically distributed with parameter  $\kappa_{f_{2m}}$ . Some readers might again find it reasonable to add the prior information that young males have a high intensity of making claims. Such

information is not added here, but our estimates will show that young males are a high risk group. The age effect for females  $f_2(\text{female}, \cdot)$  is assumed to have the same structure as the age effect for males, but with new hyperparameters. Furthermore, we need to specify the prior of the initial levels  $f_2(\text{male}, 0)$  and  $f_2(\text{female}, 0)$ . These are assumed to be lognormally distributed with parameters  $(\mu_{0,f_{2m}}, \sigma_{0,f_{2m}}^2)$  for males and  $(\mu_{0,f_{2f}}, \sigma_{0,f_{2f}}^2)$  for females; the  $\mu$ 's controlling the levels and the  $\sigma$ 's controlling the variability.

The remaining 11 components ( $g, h_{\text{sep}}^0, q_{\text{sep}}^0$ , etc.) are all assumed to have the same structure (3.7)-(3.8) as the calendar time effect  $f_1$ ; each component having its individual hyperparameters. The initial levels are assumed to be lognormally distributed, again with individual hyperparameters for each component.

Finally, unknown parameters associated with different components are assumed to be independent with respect to the prior. This independence assumption is not that crucial, and in some branches it is not reasonable. It might be that a large number of occurrences induces relatively small claims. In that case  $f_1$  should be positively correlated with the  $q$ 's. Also, it might not be reasonable to assume independence between the competing risks,  $h_{\text{se}}^0, h_{\text{sep}}^0$  and  $h_{\text{p}}^0$ , say. It could be that, if the intensity of settling only,  $h_{\text{se}}^0$ , is high, then the intensity of settling with a partial payment,  $h_{\text{sep}}^0$ , tends to be high too. There are ways of modelling such dependencies, but for simplicity we have chosen not to consider them here.

### 3.3. Additional remarks

The distribution of the claims process (3.1) is described by 14 components. These components are modelled nonparametrically. Some readers might want to add more structure by using parametric models. For example, the distribution of the partial payments could be described by mixtures of lognormal distributions. Such a choice can be motivated easily. We work with dental claims; there are examinations and there are operations. Operations are more expensive than examinations, and therefore it can be argued that the distribution of a partial payment should have two peaks. The estimates will show that the distributions of the partial payments actually have two peaks. Also, as discussed previously, it could be useful to add more structure into the calendar time effect  $f_1$  when predicting the occurrence intensity in the years to come. As time goes ( $\tau$  increases) more and more claims are observed, and the estimates of all components that are independent of calendar time will be consistent (the posterior distributions will converge towards unit mass distributions). However, the uncertainty about the occurrence intensity in the years to come will remain. Assuming that the occurrence intensity does not change with calendar time ( $f_1$  constant) is a possibility. This is, however, typically not realistic, and structures like (3.9) could be chosen.

Along the way we have pointed out possible extensions. Different branches of insurance (motor insurance, fire insurance, etc.) call for different specifications of the model. The insurance companies will have an idea about which dependencies the model should allow for, e.g. that a long settlement period typically induces large partial payments.

## 4. MARKOV CHAIN MONTE CARLO INTEGRATION

We wish to approximate the posterior distribution of the unknowns: i.e. parameters and outstanding liabilities. To do that, Markov chain Monte Carlo (MCMC) integration is used. A Markov chain, which has the posterior distribution as its equilibrium distribution, is generated. The algorithm contains an arbitrarily chosen number of steps where, at each step, new values of the unknown parameters, the remaining developments of the reported but not settled claims (RBNS) and the occurrence times, covariates and full developments of the occurred but not reported claims (IBNR), are sampled. As the number of steps increases, the empirical distribution of these sampled quantities converges towards the posterior distribution. Thereby the posterior can be approximated, and based on this approximation the estimations and predictions are made. By increasing the number of steps we can make the approximation as exact as we wish.

Figure 1 shows a model graph. Quantities surrounded by squares are known; these are either hyperparameters chosen by us, or they are data. Quantities surrounded by ellipses are unknowns; these are either unknown parameters, IBNR claims or RBNS claims. Arrows indicate dependencies. In each step of the algorithm new values of all quantities surrounded by circles are sampled.

Before we go on with a description of the algorithm we shall again refer to Smith and Roberts (1993) for an introduction to MCMC integration.

## 4.1. Likelihood

In order to solve the Bayesian inferential problem we need the likelihood of the data, i.e. the distribution of the (observed) data given the unknown parameters.

The observations are the claims reported, including their development up to calendar time  $\tau$ . We denote these observations

$$(T_i^o, S_i^o, A_i^o, U_i^o, X_i^o)_{i \geq 1}. \quad (4.1)$$

Note that  $X_i^o$  is the partial payment process censored at calendar time  $\tau$ , i.e.  $\tau - T_i^o - U_i^o$  time units after notification; cf. (2,1)-(2,2). In the previous section the distribution of all claims which had *occurred*, including their *full* development were modelled. The results of Norberg (1993a) imply that, given the unknown parameters, the process generating the *reported* claims including their full development, even if these go beyond calendar time  $\tau$ , has occurrence intensity

$$w(t, s, a)f(t, s, a)P_U(\tau - t). \quad (4.2)$$

The conditional distribution of the reporting delay of a claim which occurred at time  $t$  is

$$\frac{P_U(du)1_{(u \leq \tau - t)}}{P_U(\tau - t)}, \quad (4.3)$$

and the distribution of the corresponding partial payment process is

$$P_x(dx).$$

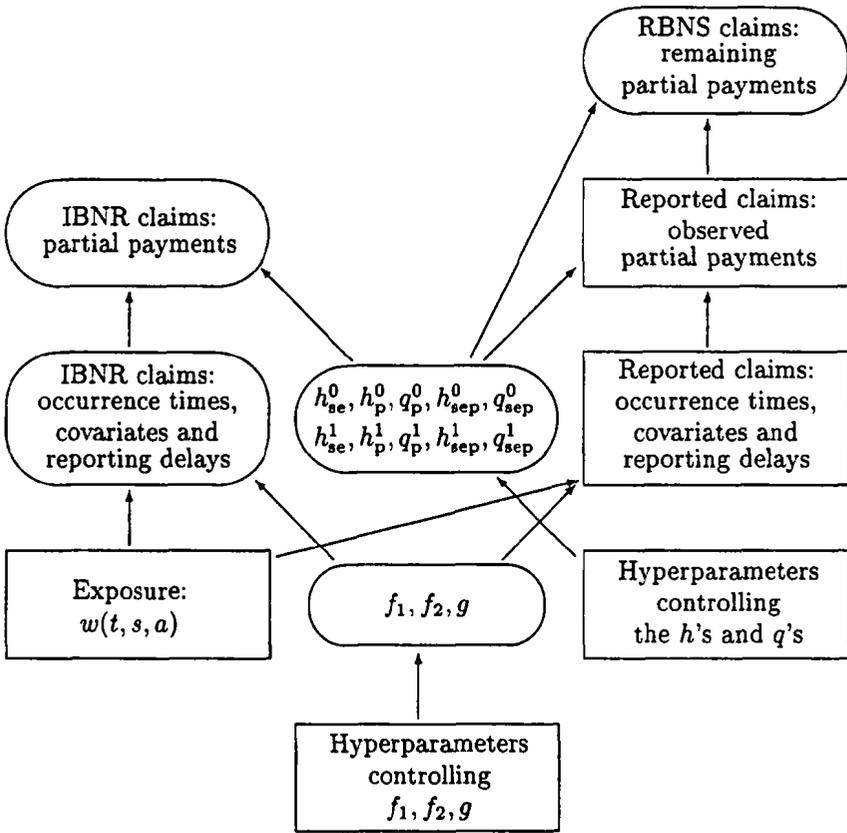


FIGURE 1 Model graph

The claims (4.1) occur with intensity (4.2). The distribution of a reporting delay of a claim which occurred at time  $t$  is given by (4.3). The distribution of the observed part of a partial payment process of a claim, which occurred at time  $t$  and was reported after  $u$  time units, is easily found using the ‘intensity construction’ of the distribution  $P_X$  of the partial payment process; see Subsection 3.1. We denote the distribution

$$P_X^{\tau-t-u},$$

where  $\tau - t - u$  refers to the censoring of the process  $\tau - t - u$  time units after notification. A likelihood is a density. In the following we use a somewhat sloppy notation which hopefully does not cause misunderstandings. The likelihood of the observations (4.1) is

$$\begin{aligned}
\Lambda(\text{obs}) &\propto \prod_{i \geq 1} f(T_i^o, S_i^o, A_i^o) P_U(\tau - T_i^o) \frac{P_U(dU_i^o)}{P_U(\tau - T_i^o)} P_X^{\tau - T_i^o - U_i^o}(dX_i^o) \\
&\quad \times \exp\left(-\sum_{s,a} \int_0^\tau w(t, s, a) f(t, s, a) P_U(\tau - t) dt\right) \\
&= \prod_{i \geq 1} f(T_i^o, S_i^o, A_i^o) P_U(dU_i^o) P_X^{\tau - T_i^o - U_i^o}(dX_i^o) \\
&\quad \times \exp\left(-\sum_{s,a} \int_0^\tau w(t, s, a) f(t, s, a) P_U(\tau - t) dt\right) \\
&\propto \prod_{i \geq 1} f_1(T_i^o) \prod_{i \geq 1} f_2(S_i^o, A_i^o) \\
&\quad \times \exp\left(-\sum_{s,a} \int_0^\tau w(t, s, a) f_1(t) f_2(s, a) \left(1 - e^{-\int_0^{\tau-t} g(u) du}\right) dt\right) \quad (4.4) \\
&\quad \times \prod_{i \geq 1} g(U_i^o) \times \exp\left(-\sum_i \int_0^{U_i^o} g(u) du\right) \\
&\quad \times \prod_{i \geq 1} P_X^{\tau - T_i^o - U_i^o}(dX_i^o). \quad (4.5)
\end{aligned}$$

Here the last part (4.5), dealing with the partial payment processes, has to be written in more detail. Recall that  $X_i^o$  is the observed part of the partial payment process of the  $i$ th claim. The superscript  $\tau - T_i^o - U_i^o$  refers to the censoring of the process at calendar time  $\tau$ . From the observed partial payment processes  $(X_i^o)_{i \geq 1}$  some useful quantities are derived. First we consider exposure:

$W^0(\nu)$ : The number of claims in which the waiting time from notification to first partial payment or settlement is at least  $\nu$ . Note that  $W^0$  is majorised by the number of reported claims and decreasing in  $\nu$ . We have  $W^0(0) = 2806$ , the number of reported claims.

$W^1(\nu)$ : The number of times any partial payment process is observed, with waiting time at least  $\nu$  since the latest partial payment. Also  $W^1$  is decreasing in  $\nu$ . We have  $W^1(0) = 1780$ . Here  $W^1$  is less than the number of reported claims, but in principle  $W^1$  can exceed the number of reported claims since a claim can have several partial payments.

Then 'events':

$(V_{se}^{0,j})_{j \geq 1}$ : Waiting times from notification to settlement without payment. There are only 9 of these in the data.

$(V_{sep}^{0,j})_{j \geq 1}$  and  $(Y_{sep}^{0,j})_{j \geq 1}$ : Waiting times from notification to settlement with one payment, and the corresponding amounts paid (in this case the payment equals the total claim amount). There are 1295 of these.

- $(V_p^{0,j})_{j \geq 1}$  and  $(Y_p^{0,j})_{j \geq 1}$ : Waiting times from notification to the first partial payment without settlement, and the corresponding amounts paid. There are 1376 of these.
- $(V_{se}^{1,j})_{j \geq 1}$ : Waiting times from a partial payment to settlement only (no partial payment at the time of settlement). There are 242 of these.
- $(V_{sep}^{1,j})_{j \geq 1}$  and  $(Y_{sep}^{1,j})_{j \geq 1}$ : Waiting times from a partial payment to settlement with payment, and corresponding amount paid at settlement. There are 643 of these.
- $(V_p^{1,j})_{j \geq 1}$  and  $(Y_p^{1,j})_{j \geq 1}$ : Waiting times from a partial payment to the next, when the latter does not settle the claim, and corresponding amounts. There are 404 of these.

As an illustration we have in Figure 2 shown two examples of partial payment processes. The first process settles within time  $\tau$ . There are two payments made, and no payment is made at the time of settlement. The second process arises from an RBNS claim. There is one payment made and the partial payment process is censored (calendar time  $\tau$  is reached). Events and amounts are shown at the figure; waiting time between jumps are events, and sizes of jumps are amounts. Note that  $V'$  is the waiting time from latest payment until censoring, and it only affects the exposure. The exposure functions are given by

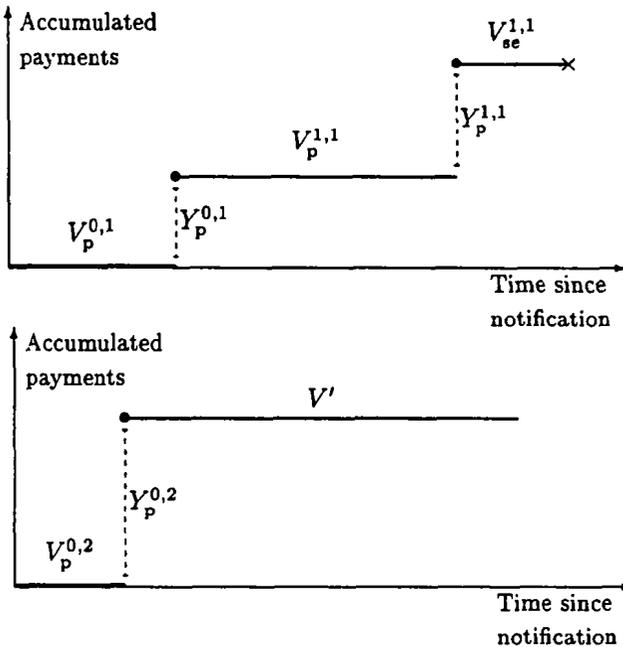


FIGURE 2 Examples of observed partial payment processes; x denotes a settlement

$$W^0(v) = \begin{cases} 2 & v \in (0, V_p^{0,2}] \\ 1 & v \in (V_p^{0,2}, V_p^{0,1}] \\ 0 & v \in (V_p^{0,1}, \infty) \end{cases} \quad \text{and} \quad W^1(v) = \begin{cases} 3 & v \in (0, V_{se}^{1,1}] \\ 2 & v \in (V_{se}^{1,1}, V_p^{1,1}] \\ 1 & v \in (V_p^{1,1}, V^1] \\ 0 & v \in (V^1, \infty) \end{cases}.$$

As a function of exposure and events the last part (4.5) of the likelihood can be written as

$$\prod_{i \geq 1} P_x^{\tau - T_i^0 - U_i^0}(dX_i^0) \propto \prod_{j \geq 1} h_{se}^0(V_{se}^{0,j}) \prod_{j \geq 1} h_{sep}^0(V_{sep}^{0,j}) \prod_{j \geq 1} h_p^0(V_p^{0,j}) \tag{4.6}$$

$$\times \prod_{j \geq 1} h_{se}^1(V_{se}^{1,j}) \prod_{j \geq 1} h_{sep}^1(V_{sep}^{1,j}) \prod_{j \geq 1} h_p^1(V_p^{1,j}) \tag{4.7}$$

$$\times \exp\left(-\int_0^\infty W^0(v)(h_{se}^0(v) + h_{sep}^0(v) + h_p^0(v))dv\right) \tag{4.8}$$

$$-\int_0^\infty W^1(v)(h_{se}^1(v) + h_{sep}^1(v) + h_p^1(v))dv \tag{4.9}$$

$$\times \prod_{j \geq 1} p_{sep}^0(Y_{sep}^{0,j}) \prod_{j \geq 1} p_p^0(Y_p^{0,j}) \tag{4.10}$$

$$\times \prod_{j \geq 1} p_{sep}^1(Y_{sep}^{1,j}) \prod_{j \geq 1} p_p^1(Y_p^{1,j}), \tag{4.11}$$

where the  $V$ ,  $W$  and  $Y$  depend on the observed partial payment processes in an obvious way. The density  $p_{sep}^0$ , say, is given by its corresponding hazard rate  $q_{sep}^0$ . We have

$$\prod_{j \geq 1} p_{sep}^0(Y_{sep}^{0,j}) \propto \prod_{j \geq 1} q_{sep}^0(Y_{sep}^{0,j}) \times \exp\left(-\sum_{j \geq 1} \int_0^{Y_{sep}^{0,j}} q_{sep}^0(y)dy\right). \tag{4.12}$$

The last part (4.5) of the likelihood function can now be substituted with (4.6)-(4.12).

We have chosen priors along the lines of Arjas and Gasbarra (1994). To copy their sampling algorithm it is necessary that, in each of the 14 components, the likelihood is proportional to an expression of the form

$$\prod h(\cdot) \times \exp\left(-\int Z(\cdot)h(\cdot)\right),$$

where  $h$  is a component, and  $Z$  does not depend (functionally) on  $h$ . This is the case with all components but  $g$ , which unfortunately occurs in the part (4.4) of the likelihood. A way to deal with this problem is to use *data augmentation*.

#### 4.2. Missing data

The reason that the likelihood of the (observed) data does not have a tractable shape is that the occurred but not reported claims (IBNR) are not observed. Here data augmentation can be used. By sampling and adding to our data the occurrence times

and covariates of the IBNR claims, the likelihood obtains the desired shape. The occurrence times and covariates of the IBNR claims are called *missing data*. The missing data are denoted

$$(T_j^m, S_j^m, A_j^m)_{j \geq 1}. \quad (4.13)$$

Note that reporting delays and partial payment processes of the IBNR claims, and the remaining development of the RBNS claims, are not included as missing data. However, we are going to sample them at each step of the algorithm, but they are not used to sample unknown parameters. They are only used to predict outstanding liabilities, cf. Figure 1. The results of Norberg (1993a) imply that, given the unknown parameters, the missing data (4.13) are independent of the observations (4.1) and that the missing data are distributed as a marked Poisson process with intensity

$$w(t, s, a)f(t, s, a)(1 - P_U(\tau - t)), \quad (t, s, a) \in (0, \tau) \times \{\text{male, female}\} \times \{1, 2, \dots\}. \quad (4.14)$$

The likelihood of the missing data only is

$$\begin{aligned} \Lambda(\text{mis}) &\propto \prod_{j \geq 1} f(T_j^m, S_j^m, A_j^m)(1 - P_U(\tau - T_j^m)) \\ &\quad \times \exp\left(-\sum_{s,a} \int_0^\tau w(t, s, a)f(t, s, a)(1 - P_U(\tau - t))dt\right) \\ &\propto \prod_{j \geq 1} f_1(T_j^m)f_2(S_j^m, A_j^m) \exp\left(-\int_0^{\tau-T_j^m} g(u)du\right) \\ &\quad \times \exp\left(-\sum_{s,a} \int_0^\tau w(t, s, a)f_1(t)f_2(s, a)e^{-\int_0^{\tau-t} g(u)du}dt\right). \end{aligned}$$

Because of conditional independence of the observed and missing data, the likelihood of these combined data is the product of the corresponding two likelihood expressions, i.e.,

$$\Lambda(\text{total}) = \Lambda(\text{obs})\Lambda(\text{mis}) \quad (4.15)$$

$$\begin{aligned} &\propto \prod_{i \geq 1} f_1(T_i^o) \prod_{j \geq 1} f_1(T_j^m) \prod_{i \geq 1} f_2(S_i^o, A_i^o) \prod_{j \geq 1} f_2(S_j^m, A_j^m) \\ &\quad \times \exp\left(-\sum_{s,a} \int_0^\tau w(t, s, a)f_1(t)f_2(s, a)dt\right) \end{aligned} \quad (4.16)$$

$$\times \prod_{i \geq 1} g(U_i^o) \times \exp\left(-\sum_{i \geq 1} \int_0^{U_i^o} g(u)du - \sum_{j \geq 1} \int_0^{\tau-T_j^m} g(u)du\right) \quad (4.17)$$

$$\times \prod_{i \geq 1} \mathbf{P}_X^{\tau-T_i^o-U_i^o}(dX_i^o),$$

where the last part is written in detail in (4.6)-(4.12). Now the likelihood has the desired shape.

### 4.3. Sampling algorithm

The sampling algorithm goes as follows: To begin with, initial values of the unknown parameters are determined. These can either be sampled from the prior distribution or chosen arbitrarily; asymptotically it does not matter. A general step of the algorithm contains three substeps:

- Substep 1: Sample occurrence times and covariates of the IBNR claims (the missing data) given the unknown parameters sampled in the previous step of the algorithm;
- Substep 2: Sample the remaining development of the reported but not settled claims (RBNS) and the full developments of the IBNR claims, given the unknown parameters sampled in the previous step of the algorithm, and given both observed and missing data where the latter were sampled in Substep 1;
- Substep 3: Sample the unknown parameters given both the data (observations) and the missing data sampled in Substep 1.

These three substeps are discussed in the following.

#### 4.3.1. Sampling the missing data

The missing data are the occurrence times and covariates of the IBNR claims, see (4.13). Given the present value of the unknown parameters, sampled in the previous step of the algorithm, the missing data are distributed as a marked Poisson process with intensity (4.14). It follows from Norberg (1993a), Theorem 1, that the total number of IBNR claims is Poisson distributed with mean

$$W^{IBNR} = \int_0^\tau w(t, s, a) f_1(t) f_2(s, a) (1 - P_U(\tau - t)) dt,$$

and given this total number, the occurrence times and covariates of the IBNR claims are iid with density

$$w(t, s, a) f_1(t) f_2(s, a) (1 - P_U(\tau - t)) / W^{IBNR}$$

on  $(0, \tau] \times \{\text{male, female}\} \times \{1, 2, \dots\}$ . The missing data are sampled from this distribution.

#### 4.3.2. Sample IBNR and RBNS claims

In Substep 2 the unknown parameters sampled earlier by the algorithm, the data, and the occurrence times and covariates of the IBNR claims sampled in Substep 1 (the missing data), are all held fixed.

First we sample the reporting delays of the IBNR claims. We already know their corresponding occurrence times  $(T_j^m)_{j \geq 1}$ . So, the  $j$ th reporting delay must exceed  $\tau - T_j^m$ , and it is distributed according to the hazard rate  $g$ . Now the  $j$ th reporting

delay is sampled from a distribution on  $(\tau - T_j^m, \infty)$  with piecewise constant hazard rate  $g$ .

The partial payment process of an IBNR claim can be sampled as follows. First the waiting time to the first 'event' is sampled. This waiting time is distributed according to the piecewise constant intensity  $h_{se}^0 + h_{sep}^0 + h_p^0$ . Then the type of the event is sampled. With probability  $h_{se}^0 / (h_{se}^0 + h_{sep}^0 + h_p^0)$  it is a settlement only, with probability  $h_{sep}^0 / (h_{se}^0 + h_{sep}^0 + h_p^0)$  it is a settlement with partial payment, and with probability  $h_p^0 / (h_{se}^0 + h_{sep}^0 + h_p^0)$  it is a partial payment only. If the first event includes a partial payment then the size of this payment is sampled. If the first event also includes settlement, then the partial payment has hazard rate  $q_{sep}^0$ , else it has hazard rate  $q_p^0$ . If the first event was a partial payment only then the waiting time to the next event and the type of this event are sampled. Here the components needed are  $h_{se}^1, h_{sep}^1, h_p^1, q_{sep}^1$  and  $q_p^1$ . Waiting times and events are sampled until settlement is reached.

The remaining partial payment process of an RBNS claim is sampled using the same method as above. For each individual RBNS claim we need to know whether any previous partial payments have been made, and the time since the latest partial payment if any, else the time from reporting.

#### 4.3.3. Sampling the unknown parameters

In this substep we condition on the observed data and the missing data sampled in Substep 1, and sample the unknown parameters (jump times and levels of the 14 components). The conditional distribution of these unknowns has a density that is proportional to the likelihood (4.15) of the combined observed and missing data, multiplied with the prior density of the unknown parameters. Direct sampling from this distribution is not possible. However, we can identify the conditional distributions of each of the unknown parameters given the remaining part of the unknown parameters and given the observed and missing data. From these one-dimensional distributions the parameters can be sampled one by one. A complete description of the sampling of the unknown parameters would take quite a few pages. We shall here give a short overview.

The unknown parameters are the jump times and levels of all the 14 components. To save notation we shall denote the unknown parameters by  $X_1, X_2, \dots$  (how they are ordered will be discussed later). Assume that the present step is the  $k$ th, and denote by  $X_1^{k-1}, X_2^{k-1}, \dots$  the values of the unknown parameters sampled in the previous step of the algorithm. They are now sampled one by one. As mentioned, the conditional distribution of each individual unknown parameter, given the remaining parts of the unknowns and given observed and missing data, is known (at least up to proportionality). First a new value  $X_1^k$  of  $X_1$  is sampled from the distribution of

$$X_1 \text{ given } X_2^{k-1}, X_3^{k-1}, \dots, \quad (4.18)$$

then a new value  $X_2^k$  of  $X_2$  is sampled from the distribution of

$$X_2 \text{ given } X_1^k, X_3^{k-1}, \dots, \quad (4.19)$$

and so the algorithm continues. The algorithm is called Gibbs sampling, cf. Smith and Roberts (1993). Recall that the unknown parameters are the levels and jump times of the 14 components. When the unknown parameter (an  $X$ ) is a level, it is not tractable to sample from the above one-dimensional distributions and then Metropolis-Hastings is used. With this modification the algorithm is called a 'variable-at-a-time Metropolis-Hastings', cf. Chan and Geyer's discussion of the paper by Tierney (1994).

The method used for doing the sampling (4.18)-(4.19) is taken from Arjas and Gasbarra (1994). The unknown parameters can be grouped into 14 groups according to the component they determine ( $f_1, f_2(\text{male}, \cdot), f_2(\text{female}, \cdot), g, \dots$ ). In Arjas and Gasbarra (1994) it is shown how the unknown parameters associated with one component can be sampled. Denote by ( $b_j$ ) and ( $S_j$ ) the levels and jump times of the component  $f_j$ , say. Arjas and Gasbarra order them as

$$b_0, S_1, b_1, S_2, \dots, \quad (4.20)$$

and sample new values as indicated in (4.18)-(4.19) above. Their algorithm carries over with minor modifications; their levels are correlated gamma distributed, while our levels are correlated lognormally distributed, and when sampling the levels they use rejection sampling which corresponds to repeating Metropolis-Hastings until acceptance is reached. When sampling the unknown parameters associated with an individual component, we adopt the methods of Arjas and Gasbarra, and thereby the components are sampled one by one, following the order (4.20). But in which order should the different components be sampled? According to the prior the 14 components are independent, and by inspection of the likelihood (4.15) of the combined observed and missing data, we get that according to the conditional distribution of the unknown parameters given observed and missing data, all components but  $f_1, f_2(\text{male}, \cdot)$  and  $f_2(\text{female}, \cdot)$  are independent. Therefore, except for the mutual order of  $f_1, f_2(\text{male}, \cdot)$  and  $f_2(\text{female}, \cdot)$ , the order in which the different components are sampled does not play a role. We have chosen to sample first  $f_1$ , then  $f_2(\text{male}, \cdot)$ , then  $f_2(\text{female}, \cdot)$ , and then the remaining 11 components. Asymptotically, the order does not play any role.

## 5. RESULTS OF THE ANALYSIS

Suppose we stop the sampling algorithm after  $n$  steps. What we then have is the first  $n$  steps of a Markov chain. That the chain is Markov follows from the construction. Each step of the chain contains sampled values of the unknown

parameters,

occurrence times, covariates and full development of the IBNR claims, and

remaining development of the RBNS claims.

By construction, the posterior distribution, i.e. the conditional distribution of the above unknowns given the observations, is invariant for that chain. It is obvious that

the chain is both irreducible and aperiodic. From Tierney (1994) it now follows that the chain is (positive) recurrent. We need the chain to be ergodic, and for that a stronger type of recurrence, called Harris recurrence, is needed. In their discussion of the paper by Tierney (1994), Chan and Geyer show that a certain type of algorithm, which they call a ‘variable-at-a-time Metropolis-Hastings’, is often Harris recurrent. Their proof carries over to our situation with only minor modifications. It then follows that our chain is ergodic. A survey of the theory can be found in Tierney (1994).

Let  $(M_j)_{j=1,2,\dots}$  denote the chain, let  $h$  be a function on the space of possible values of the unknowns above, and assume that  $h$  has finite mean with respect to the posterior distribution. Then

$$\frac{1}{n} \sum_{j=1}^n h(M_j) \xrightarrow{\text{a.s.}} E_{\text{posterior}} h. \quad (5.1)$$

As an example, suppose we wish to find (approximate) the posterior distribution of the size of the outstanding liabilities. The outstanding liabilities are a function of the claims process (3.1). The liabilities are divided into liabilities arising from the IBNR claims,

$$X^{\text{IBNR}} = \sum_{i \geq 1} \mathbf{1}_{(T_i \leq \tau < T_i + U_i)} Y_i,$$

and liabilities arising from the RBNS claims,

$$X^{\text{RBNS}} = \sum_{i \geq 1} \mathbf{1}_{(T_i + U_i \leq \tau < T_i + U_i + V_i)} (Y_i - Y_i(\tau - T_i - U_i)).$$

Let  $X_j^{\text{IBNR}}$  and  $X_j^{\text{RBNS}}$  denote the values sampled in the  $j$ th step of the algorithm (both are functions of  $M_j$ ). By (5.1) we have

$$\frac{1}{n} \sum_{j=1}^n \mathbf{1}_{(X_j^{\text{IBNR}} + X_j^{\text{RBNS}} \leq x)} \xrightarrow{\text{a.s.}} P_{\text{posterior}}(X^{\text{IBNR}} + X^{\text{RBNS}} \leq x), \quad (5.2)$$

and thereby the posterior (predictive) distribution of the outstanding liabilities can be approximated by the left hand side of (5.2). The theory does not give an exact answer as to how large  $n$  should be chosen. However, what one usually does is to run several independent simulations and use these to evaluate the stability of the algorithm.

To begin with we have to choose values of the hyperparameters, and the initial values of the unknown parameters. Recall that each of the 14 components (intensities) can be written as

$$\sum_{j \geq 0} \mathbf{1}_{(S_j < t \leq S_{j+1})} b_j. \quad (5.3)$$

The hyperparameters  $\sigma_0^2$  and  $\mu_0$  determine the prior variance and mean of the initial level ( $b_0$ ) while  $\lambda$  (or  $\kappa$ ) and  $\sigma^2$  determine the prior variability of the intensity; cf. the discussion in Subsection 3.2. Table 1 shows the chosen values of the hyper-

TABLE 1  
HYPERPARAMETERS

	$\sigma_0^2$	$\mu_0$	$\lambda^{-1}$	$\sigma^2$
$f_1$	.	.	70	0.01
$g$	0.2	-9	6	0.01
$h_{scp}^0$	0.2	-3	10	0.02
$h_p^0$	0.2	-6	10	0.02
$h_{se}^0$	0.2	-6	10	0.02
$h_{scp}^1$	0.2	-3	10	0.02
$h_p^1$	0.2	-6	10	0.02
$h_{se}^1$	0.2	-6	10	0.02
$q_{scp}^0$	0.02	-10	250	0.04
$q_p^0$	0.02	-10	250	0.04
$q_{scp}^1$	0.02	-10	250	0.04
$q_p^1$	0.02	-10	250	0.04
	$\sigma_0^2$	$\mu_0$	$-\log(1 - \kappa)$	$\sigma^2$
$f_2(\text{male}, \cdot)$	0.2	-11	5	0.01
$f_2(\text{female}, \cdot)$	0.2	-11	5	0.01

parameters. We shall not try to motivate the choice of all the values. However, a few comments are needed. The individual values of  $\lambda$  and  $\sigma^2$  were chosen as follows. We started out with some arbitrary values. If the estimate (posterior mean of the intensity) turned out either too ragged or too smooth, then new values of  $\lambda$  and  $\sigma^2$  were chosen. This way of choosing hyperparameters corresponds to how window size (bandwidth) is chosen in kernel density estimation, and it is not orthodox Bayesian in spirit. On the other hand, our choice of the hyperparameters  $\sigma_0^2$  and  $\mu_0$  represents prior knowledge. For example, we believe that small payments (from 0 to 25 DKK) are very unlikely, and therefore the values of the  $q$ 's should start out small. By choosing  $\mu_0$  small we express the prior knowledge that the intensity (hazard rate) starts out small, and by choosing  $\sigma_0^2$  small we express a high degree of belief in this knowledge. Other values of the hyperparameters were tried out. But, within what we believe were reasonable values, the hyperparameters had only minor influence on our estimations/predictions.

The unknown parameters are the jump times and levels (the  $S$ 's and  $b$ 's, cf. (5.3)) of each of the 14 components. Their initial values can, for example, be sampled from the prior distribution or be chosen arbitrarily. The theory says that asymptotically it does not matter what you do, which was also our experience. However, the algorithm turned out to be time consuming, and to save iterations we, therefore, wanted to start the algorithm at a place with high posterior probability. For each component, the initial values of the jump times were chosen such that areas with many observations (occurrences, delays or payments) had many jumps. Given the jump times, the maximum likelihood estimators of the levels were used as their initial values.

The algorithm was written in S-Plus and run on a DEC workstation. As mentioned, the algorithm turned out to be rather time consuming; about 1 hour for each 100 steps, but much could have been gained by writing parts of the algorithm in C, say. We did two runs; one of length 1000 and one of length 500. After a few iterations (40-60) the algorithm seemed stable (we plotted the approximations to the posterior mean of the individual 14 components, and found that these did not change much). As approximation to the posterior distribution of all unknowns we use the empirical distribution of all 1500 realizations.

Each of the 14 components was estimated by its (approximated) posterior mean (an average of the 1500 iterations). In Figure 3 the estimates of the components  $f_1$ ,  $f_2$  and  $g$  are shown. From a mathematical point of view, it is easy to calculate (posterior) pointwise 95% credible intervals for each component, but the computational effort would increase. Even though we have no credible intervals, we allow ourselves to comment on the estimates. The occurrence intensity is the product of the calendar time effect  $f_1$  and the age effect  $f_2$ . Looking at  $f_1$ , there seems to be no obvious seasonal effect. In observation year no. 3 (1984) the occurrence intensity is high (we have no explanation why), otherwise there are only minor variations. Looking at  $f_2$  it is obvious that there is an age effect. Young males are a high risk group, while young females are a low risk group. From age 30 and on there are only minor differences between the sexes. The last graph shows the reporting delay hazard rate  $g$ . Based on the estimate from Figure 3, the 95% quantile of the

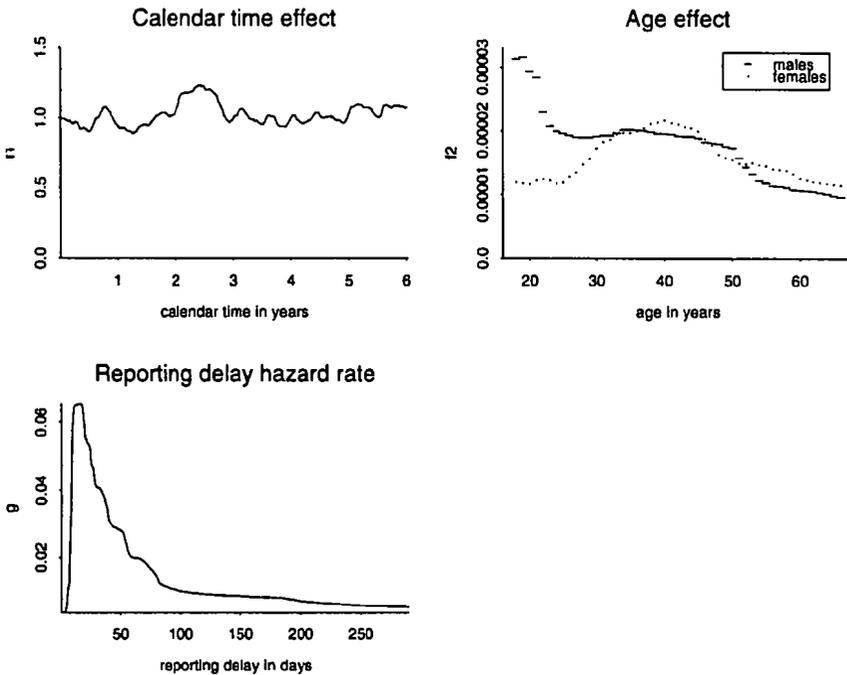


FIGURE 3 Posterior mean of  $f_1$ ,  $f_2$  and  $g$

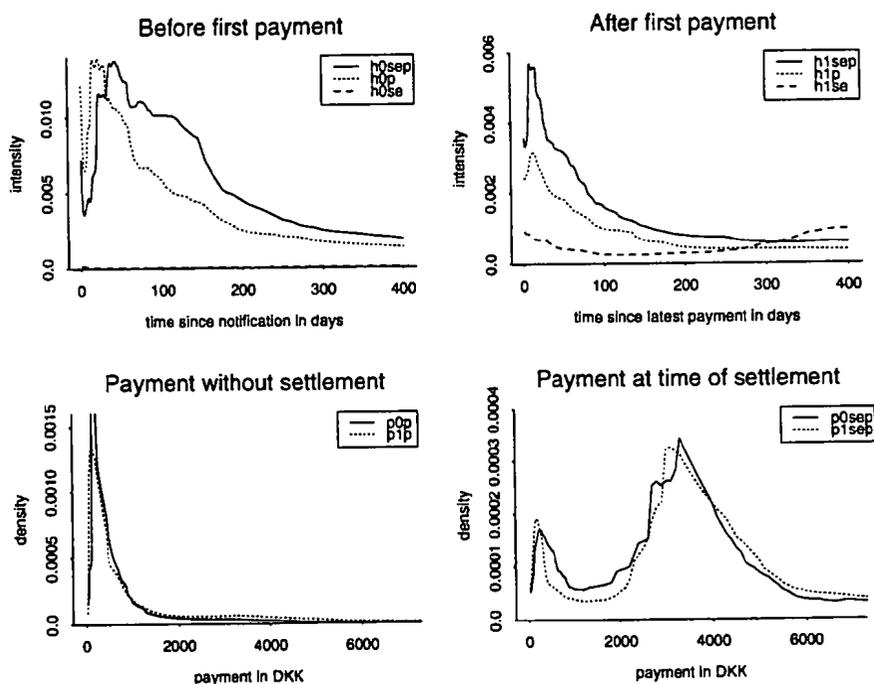


FIGURE 4 Posterior mean of  $h$ 's and  $q$ 's; the  $q$  hazard rates are shown by their corresponding densities, the  $p$ 's

distribution of the reporting delays is 4 months, and almost 75% of the claims are reported within a month.

Figure 4 shows the estimates (posterior means) of the  $h$ 's and  $q$ 's. The posterior means of the  $h$ 's and  $q$ 's are smooth functions. However, for computational simplicity the estimates are calculated at a relatively small number of points, and the estimates may therefore appear a bit ragged. Before the first payment, the different types of events, settlement with payment, payment only, and settlement only, occur with intensities  $h_{sep}^0$ ,  $h_p^0$  and  $h_{se}^0$ , respectively. The intensity of settling only is almost 0. The intensity of settling with a payment starts out high within the first couple of days from notification. It then decreases, and it has its maximum a month after notification. The intensity of having a partial payment only has essentially the same structure. There is 5% probability that no events have occurred within 6 months from notification. After the first payment, the different types of events occur with intensities  $h_{sep}^1$ ,  $h_p^1$  and  $h_{se}^1$ , respectively. Here the probability of settling only ( $h_{se}^1$ ) seems to be considerable, especially when the time since the latest payment is long. There is almost 50% probability that no events have occurred within 6 months from notification, and there is 5% probability that no events have occurred within 5 years. This means that for a claim with observed previous payments the expected settling time is long. Payments that are not made at the time of settlement are distributed with hazard rates  $q_p^0$  and  $q_p^1$ . The corresponding densities are called  $p_p^0$  and  $p_p^1$ , respectively. Now, 19% of all observed first-time

payments not made at the time of settlement were of the size of 178, 185, 186 or 189 DKK. These amounts are probably fees for some standard dental examination. Our estimate of  $p_p^0$  therefore has a high peak in this area (the peak has been truncated on the graph), and it is questionable whether it is reasonable to model the distribution of these payments as a continuous distribution. None of the other types of payment (first-time payment at time of settlement, or subsequent payments) had such very frequently occurring sizes. The average first-time payment not made at time of settlement is approximately 1000 DKK. The distribution of the subsequent payments not made at time of settlement,  $p_p^1$ , has the same shape as  $p_p^0$ , except for the high peak. Based on the estimate from Figure 4, the mean of these payments is approximately 2100 DKK. The distributions of the payments that are made at the time of settlement,  $p_{sep}^0$  and  $p_{sep}^1$ , seem to be independent of whether any previous payments have been made. The distributions have 2 peaks; one at approximately 200 DKK and one at approximately 3500 DKK. The average amounts are approximately 4200 and 4600 DKK.

Figure 5 contains our main results. Here the (approximated) predictive distribution of the outstanding liabilities are shown. Recall that the original data set contained information about occurrence times, covariates and reporting delays, on all claims which occurred between January 1, 1982 and December 31, 1990 and which were reported before March 3, 1992. Furthermore, all payments made upon

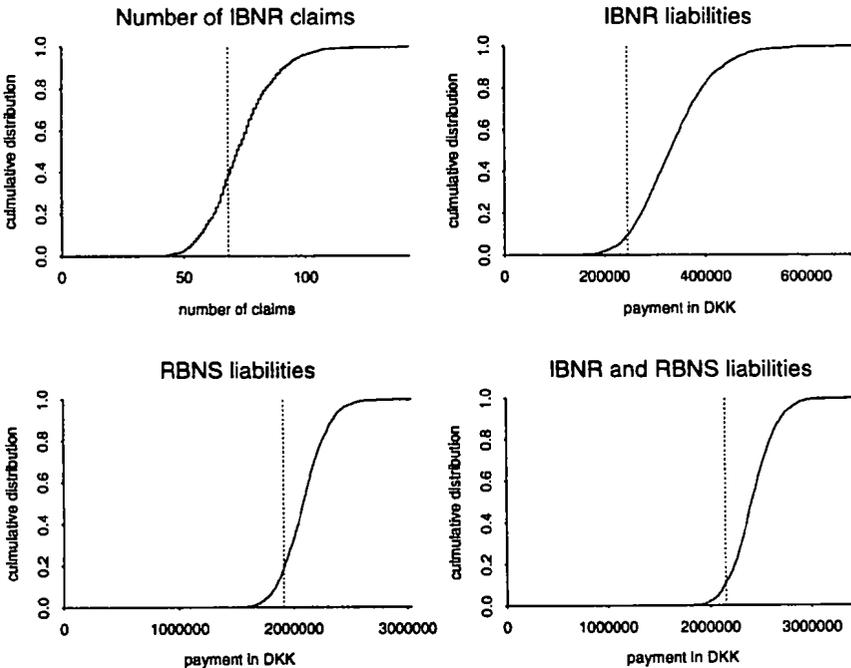


FIGURE 5 Predictive distribution of outstanding liabilities; observed accumulated run-off March 3, 1992 (4.2 years after time  $\tau$ ) is shown with dotted lines

these claims before March 3, 1992, were observed. In our estimation/prediction we only use data that were available at December 31, 1987 ( $\tau = 6$  years after January 1, 1982). The predictive distributions of the outstanding liabilities could therefore have been calculated at that time. The distributions are calculated using the method indicated in (5.2). We have the observed run-off for a period of 4.2 years. On the graphs the accumulated run-offs are shown with dotted vertical lines. There are 68 observed IBNR claims. This corresponds to the 39% quantile of the predictive distribution of the IBNR claims. The posterior distribution of the reporting delays has almost no mass for delays exceeding 3 years, and we therefore expect that all IBNR claims are reported within the 4.2 years. Table 2 shows mean, coefficient of variation (c.v.) and quantiles, from the predictive distributions of the outstanding liabilities, cf. Figure 5. The coefficient of variation is 0.22 in the predictive distribution of the outstanding IBNR liabilities, while it is only 0.09 for the RBNS liabilities. Intuitively, this is obvious; only the sizes of the RBNS claims are unknown, while for the IBNR claims also the number of claims is unknown. After 4.2 years the observed run-off for both the IBNR and the RBNS liabilities are below the 0.2 quantile of the predictive distribution of the outstanding liabilities. That was to be expected: 10 of the 68 observed IBNR claims were not yet settled after 4.2 years, and therefore there still remained some payments on the IBNR claims. Recall that RBNS claims with previous payments tend to have a long settlement delay. Of the observed 617 RBNS claims, 182 were not settled after 4.2 years. Some of these will settle without further payments, but still, the outstanding liabilities will exceed the observed 1.910.000 DKK. By combining the sampled values of the IBNR and the RBNS liabilities and using (5.2), we get an approximation to the predictive distribution of the outstanding IBNR and RBNS liabilities, shown at the last graph in Figure 5. The mean and quantiles are found in Table 2.

With the predictive distribution of the outstanding liabilities in hand, the insurance company can decide the size of the reserve to be set aside. Now, there is a possible extension at this stage. For deciding on an investment policy, it is useful for the company to know when the payments are due. This calls for the posterior (predictive) distribution of the run-off over calendar time. As previously, it is no mathematical problem to approximate this distribution; as a function of the un-

TABLE 2  
MEAN, C.V. AND QUANTILES OF THE PREDICTIVE DISTRIBUTIONS OF THE OUTSTANDING LIABILITIES

	IBNR	RBNS	IBNR and RBNS
mean	337.000 DKK	2.090.000 DKK	2.430.000 DKK
5% quantile	229.000 DKK	1.780.000 DKK	2.080.000 DKK
95% quantile	468.000 DKK	2.410.000 DKK	2.800.000 DKK
c.v.	0.22	0.09	0.09
4.2 years run-off	246.000 DKK	1.910.000 DKK	2.160.000 DKK
quantile for run-off	9%	18%	11%

knowns, sampled at each step of the algorithm, we get a value of the run-off over calendar time, and using these the desired distribution can be approximated.

#### CLOSING REMARKS

The distribution of the claims process was described by 14 one-dimensional components which were modelled in a nonparametric Bayesian way. We find the Bayesian approach very apt. A standard parametric approach to the prediction problem could be as follows. First, the distribution of the outstanding liabilities is found as a function of the parameters. Then, estimators (maximum likelihood, say) of the parameters are found. And finally, a reserve is calculated as a function of the estimators. Furthermore, the uncertainty about the parameter estimates can be incorporated into the reserve estimate. With the Bayesian approach the procedure takes place in a single step. The posterior distribution of unknown parameters and the predictive distribution of the outstanding liabilities are found simultaneously. Only the latter distribution is needed to predict the outstanding liabilities, since the uncertainty about the parameters is a part of the variation (variance) in that distribution.

We were a little less enthusiastic about the nonparametric modelling; the computations turned out very time consuming, and sometimes additional structure is needed (cf. the discussion in Subsection 3.3). In the future, we might want to model some components nonparametrically and some parametrically.

If we were to expand the model, then we would look at the partial payment processes. Claim handlers sometimes have additional information about the claims reported. Often, a claim handler forms an idea of the size of the total claim amount, and such information can be useful when predicting the outstanding payments on the RBNS claims.

#### ACKNOWLEDGMENT

Helpful comments and suggestions from Dario Gasbarra, Martin Jacobsen and Ragnar Norberg are acknowledged. Data were supplied by the Danish insurance company Topdanmark. This research was supported by The Academy of Finland.

#### REFERENCES

- ARJAS, E. 1989. The claim reserving problem in non-life insurance: some structural ideas. *ASTIN Bull. Vol. 19, no. 2*, 139-152.
- ARJAS, E., GASBARRA, D. 1994. Nonparametric Bayesian inference from right censored survival data using the Gibbs sampler. *Statistica Sinica*, 4, 505-524.
- HESSELAGER, O. 1994. A Markov model for loss reserving. *ASTIN Bull. Vol. 24, no. 2*, 183-193.
- JEWELL, W.S. 1989. Predicting IBNYR events and delays I. Continuous time. *ASTIN Bull. Vol. 19, no. 1*, 25-56.
- KARR, A. 1991. Point processes and their statistical inference. *Marcel Dekker Inc.*

- KIRKEGAARD, T. 1994. Praktisk anvendelse af en reservemodel på ulykkesforsikringer. *Master's thesis (in Danish), University of Copenhagen.*
- NORBERG, R. 1993a. Prediction of outstanding liabilities in non-life insurance. *ASTIN Bull. Vol. 23, no. 1, 95-115.*
- NORBERG, R. 1993b. Prediction of outstanding liabilities: model variations and extensions. *Unpublished.*
- NORBERG, R. 1993c. Prediction of outstanding liabilities: parameter estimation. *Proceedings of the XXIV ASTIN Coll., 255-266.*
- SMITH, A.F.M., ROBERTS, G.O. 1993. Bayesian computations via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. R. Statist. Soc. B. Vol. 55, no. 1, 3-23.*
- TIERNEY, L. 1994. Markov chains for exploring posterior distributions (with discussions). *Annals of Statistics. Vol. 22, no. 4. 1701-1762.*