

Combined Analysis of Paid and Incurred Losses

B. Posthuma, E.A. Cator, W. Veerkamp, and E.W. van Zwet

Abstract

Motivation. The new solvency regimes now emerging, insist that capital requirements align with the underlying (insurance) risks. This paper explains how a stochastic model built on basic assumptions is used to monitor insurance risk in order to get a clear insight in the aligned economic capital including prudence margins for loss reserves.

Method. The incurred loss of an insurer consists of payments on claims and reserves for claims that have been reported. As all claims are settled eventually, the cumulative paid and incurred losses for a given loss period become equal. Therefore, a joint model for the paid and incurred loss arrays is constructed, following a multivariate normal distribution, conditioned on equality of the total paid and incurred losses for a given loss period. A new class of functions is designed specifically to model development curves.

Results. A simulation experiment proved that a joint model for both paid and incurred loss arrays as described under *Method*, leads to a more accurate prediction of loss reserves. While the standard way of estimating percentiles for the reserve is biased, the alternative method of bootstrapping will lead to more accurate outcomes.

Conclusions. Modeling paid and incurred losses jointly leads to a considerable improvement in loss reserving in terms of accuracy of predictions, as well as specification of percentiles.

Availability. This method is incorporated in software available from the authors.

Keywords: Solvency II, loss reserves, joint model for paid and incurred loss arrays.

1. INTRODUCTION

The new risk based solvency regimes now emerging, such as the Solvency II rules to be implemented in Europe in 2009, insist that capital requirements align with the underlying (insurance) risks. This makes a stochastic loss reserving model a necessity. Such a model needs straightforward assumptions that will allow that:

- risk for expired insurance contracts is integrated together with risk for future contracts, in order to get a complete insight into the risk of the insurance portfolio as a whole, and that
- incomplete data – such as imperfect loss triangles due to varying period lengths or even incidental missing values – is still constructive to the model.

Regression as a descriptive technique with basic probability assumptions often offers the possibility to efficiently create an appropriate stochastic framework.

In short, an insurer will have to examine previous payments to make predictions about all future financial obligations. However, the company needs to know more than just how much money it

should expect to pay. The model's stochastic ranges generate economic capital and prudence margins for reserves. Therefore, an adequate assessment of percentile ranges is crucial.

Typically, an insurer will arrange his payments by loss period and development period in a rectangular loss array, which is also sometimes called a run-off table. Since some of the payments lie in the future, this array is not fully observed. The observed part is often referred to as a run-off triangle. We regard the unobserved part of the loss array as a collection of random variables and the goal is to determine their probability distributions as well as possible on the basis of the available data.

Naturally, an extensive literature exists on this important problem. Perhaps the most widely used approach is the chain ladder. Renshaw and Verral (1998) identify the underlying assumptions and Mack (1993) and England and Verral (1999) present ways of estimating the standard error of the prediction. There are countless alternatives to the chain ladder and Schmidt (2007) has compiled a 35-page bibliography on the subject of loss reserving!

Much of the existing literature, however, concerns only a single array of payments—an exception is the Munich Chain Ladder introduced by Quarg and Mack (2004). Indeed, in most cases we have two arrays: an array of payments on settled claims and an array of reserves for claims that have been reported, but not yet settled. We refer to the sum of payments and reserves as “incurred loss.”

In this paper, we aim to analyze the paid and incurred loss arrays jointly. As all claims are settled eventually, the reserves vanish and the cumulative paid and incurred loss for a given loss period become equal. On the basis of this observation, we construct a joint model. In our description, each array follows a multivariate normal distribution, conditioned on equality of the total paid and incurred losses for a given loss period.

This paper is organized as follows. In the next section we present an overview of our multivariate normal model for the two arrays. We then proceed to give a more detailed description, defining a particular family of functions that is very useful for modeling development curves. In most cases, we observe only various aggregates of the arrays, but we show that this poses no difficulties. We discuss prediction and parameter estimation. To examine the advantage of our joint model we conducted a simulation experiment. We compared the results of the joint model to those obtained from using only a single array. We find that the joint model shows better results in terms of the mean squared prediction error. We report on these results in the final section.

2. MULTIVARIATE NORMAL MODEL

Let $Y_{lk}^{(1)}$ and $Y_{lk}^{(2)}$ denote the incremental paid and incurred losses for loss period $l = 1, 2, \dots, L$ in development period $k = 1, 2, \dots, K$. Suppose that they are all independent normally distributed with means

$$EY_{lk}^{(1)} = \mu_l \Pi_k^{(1)} \quad \text{and} \quad EY_{lk}^{(2)} = \mu_l \Pi_k^{(2)}$$

and variances

$$\text{var}(Y_{lk}^{(1)}) = \tilde{\Pi}_k^{(1)} \quad \text{and} \quad \text{var}(Y_{lk}^{(2)}) = \tilde{\Pi}_k^{(2)}.$$

We assume

$$\sum_k \Pi_k^{(1)} = \sum_k \Pi_k^{(2)} = 1.$$

It is of course sensible to assume a parametric form for the parameter vectors $\boldsymbol{\mu}$, $\boldsymbol{\Pi}^{(1)}$, $\boldsymbol{\Pi}^{(2)}$, $\tilde{\boldsymbol{\Pi}}^{(1)}$ and $\tilde{\boldsymbol{\Pi}}^{(2)}$. For ease of presentation, we defer this issue to the next section.

The assumed normal distribution of the entries of the loss arrays is often not appropriate. Occasional large claims result in distributions that are skewed to the right. To account for this skewness the entries are sometimes assumed to have the lognormal distribution. A disadvantage of such a model is the incompatibility of the log normal distribution with the negative values that do occur in practice in most arrays, and the incompatibility of the distribution when aggregating data (the sum of two lognormal random variables is not log normally distributed). Also, it will not be feasible to do what we are about to propose — that is, condition on the equality of the row sums of the loss arrays.

We should point out that as a result of the Central Limit Theorem, aggregates of the data are more normally distributed than the individual entries. We feel that the advantages of the multivariate normal model outweigh those of the multivariate lognormal model.

Let $\mathbf{Y}^{(1)}\mathbf{1}$ and $\mathbf{Y}^{(2)}\mathbf{1}$ denote the row sums of the matrices $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$. Also, we can stretch out $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ as length KL vectors $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$, respectively.

Given the event $\{\mathbf{Y}^{(1)}\mathbf{1} = \mathbf{Y}^{(2)}\mathbf{1}\}$, the vectors $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$ have multivariate normal distributions. It is not difficult to determine the conditional mean and conditional covariance matrix. Refer to the Appendix for a general formulation.

Because $E\mathbf{Y}^{(1)}\mathbf{1} = E\mathbf{Y}^{(2)}\mathbf{1}$, the conditional mean of the vectors $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$ is the same as the unconditional mean. However they are of course no longer independent!

Let Σ_{11} denote the unconditional covariance matrix of the length $2KL$ vector $\mathbf{y} = (\mathbf{y}^{(1)}, -\mathbf{y}^{(2)})$.

$$\Sigma_{11} = \begin{pmatrix} Cov(\mathbf{y}^{(1)}) & 0 \\ 0 & Cov(\mathbf{y}^{(2)}) \end{pmatrix}, \quad (2.1)$$

where $Cov(\mathbf{y}^{(1)})$ and $Cov(\mathbf{y}^{(2)})$ are the diagonal covariance matrices of $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$. We use $-\mathbf{y}^{(2)}$ for convenience, since in that case the row sums add up to zero.

Let Σ_{22} denote the covariance matrix of $(\mathbf{Y}^{(1)} - \mathbf{Y}^{(2)})\mathbf{1}$. Then

$$\Sigma_{22} = \left(\sum_k \tilde{\Pi}_{(k)}^{(1)} + \sum_k \tilde{\Pi}_k^{(2)} \right) \mathbf{I},$$

where \mathbf{I} is the $L \times L$ identity matrix.

Let $\Sigma_{12} = \Sigma'_{21}$ denote the covariance between \mathbf{y} and $(\mathbf{Y}^{(1)} - \mathbf{Y}^{(2)})\mathbf{1}$.

The conditional covariance matrix of \mathbf{y} given the event $\{(\mathbf{Y}^{(1)} - \mathbf{Y}^{(2)})\mathbf{1} = \mathbf{0}\}$ is

$$\Sigma = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \quad (2.2)$$

This completes the global specification of our model. In the next section we give a more detailed description.

3. DETAILED SPECIFICATION OF THE MODEL

In the previous section, we introduced vectors $\boldsymbol{\mu} \in \mathfrak{R}^L$ and $\boldsymbol{\Pi}^{(1)}, \boldsymbol{\Pi}^{(2)} \in \mathfrak{R}^K$, to describe the expectations. Define

$$\boldsymbol{\Pi} = \begin{pmatrix} \boldsymbol{\Pi}^{(1)} \\ -\boldsymbol{\Pi}^{(2)} \end{pmatrix}.$$

Mostly, we have a vector of “exposures” $\mathbf{W} \in \mathfrak{R}^L$ representing a volume measure for each loss period, such as the total number of insurance policies. We choose an $L \times p$ matrix \mathbf{X} and a parameter vector $\boldsymbol{\beta} \in \mathfrak{R}^p$, and we model the expected total loss for loss period l

$$\mu_l = W_l e^{(\mathbf{X}\boldsymbol{\beta})_l} \quad (3.1)$$

We have

$$E(Y_{lk}^{(1)}) = W_l e^{(\mathbf{X}\boldsymbol{\beta})_l} \Pi_k^{(1)} \quad \text{and} \quad E(Y_{lk}^{(2)}) = W_l e^{(\mathbf{X}\boldsymbol{\beta})_l} \Pi_k^{(2)}.$$

This means that if we define for matrices \mathbf{A} and \mathbf{B} of equal size

$$\exp(\mathbf{A})_{ij} = e^{A_{ij}} \quad \text{and} \quad (\mathbf{A} \circ \mathbf{B})_{ij} = A_{ij} B_{ij},$$

then we can write

$$E(\mathbf{y}) = (\mathbf{W} \circ \exp(\mathbf{X}\boldsymbol{\beta})) \otimes \boldsymbol{\Pi},$$

where \otimes denotes the tensor product between two vectors.

Next, we recall the vectors $\tilde{\boldsymbol{\Pi}}^{(1)}, \tilde{\boldsymbol{\Pi}}^{(2)} \in \mathfrak{R}^K$, which represent the (unconditional) variances. Define their sums as σ_1^2 and σ_2^2

$$\sum_{k=1}^K \tilde{\Pi}_k^{(1)} = \sigma_1^2 \quad \text{and} \quad \sum_{k=1}^K \tilde{\Pi}_k^{(2)} = \sigma_2^2.$$

Also define

$$\tilde{\boldsymbol{\Pi}} = \begin{pmatrix} \tilde{\boldsymbol{\Pi}}^{(1)} \\ \tilde{\boldsymbol{\Pi}}^{(2)} \end{pmatrix}.$$

We model the unconditional variances for loss period l and development period k as

$$\tilde{V}_{lk}^{(1)} := \text{Var}(Y_{lk}^{(1)}) = W_l e^{(X\beta)_l} \tilde{\Pi}_k^{(1)} \quad \text{and} \quad \tilde{V}_{lk}^{(2)} := \text{Var}(Y_{lk}^{(2)}) = W_l e^{(X\beta)_l} \tilde{\Pi}_k^{(2)}.$$

Note that we use the same \mathbf{X} and $\boldsymbol{\beta}$ as we did for the expectations. In matrix notation this becomes

$$\text{Cov}(\mathbf{y}) = \left((\mathbf{W} \circ \exp(\mathbf{X}\boldsymbol{\beta})) \otimes \tilde{\boldsymbol{\Pi}} \right)_{\Delta}.$$

Here we denote the diagonal matrix with the vector \mathbf{v} as its diagonal by \mathbf{v}_{Δ} . This describes the unconditional distribution of the vector \mathbf{y} . We can now use (2.2) to find the conditional distribution of \mathbf{y} , which then completely specifies our model.

3.1 Modeling the development curves

For a sensible approach to the estimation problem, it is necessary to limit the number of parameters by assuming a parametric model for the development vectors $\boldsymbol{\Pi}^{(1)}$, $\boldsymbol{\Pi}^{(2)}$, $\tilde{\boldsymbol{\Pi}}^{(1)}$, and $\tilde{\boldsymbol{\Pi}}^{(2)}$. To explain our method, let us concentrate on one of the arrays, for example $\mathbf{Y}^{(1)}$. We suppose that in loss period l , we expect a total loss of μ_l . Now suppose that the length of the loss period is T time units. The claims occurring in the small interval $[t, t + \Delta t]$, have an effect on the expected loss in the time interval $[s, s + \Delta s]$, $t \leq s$ equal to

$$\frac{\mu_l \Delta t}{T} f_{\boldsymbol{\theta}}(s - t) \Delta s,$$

where $f_{\boldsymbol{\theta}}$ is a (possibly negative) function such that

$$\int_0^{\infty} f_{\boldsymbol{\theta}}(x) dx = 1$$

for all possible choices of the parameter vector $\boldsymbol{\theta}$.

In the next section, we will describe a particular family of such functions, which possess some desirable properties. For now, let us note that the total loss over loss period l equals μ_l . Indeed, if $[t_l, t_l + T]$ denotes the loss period l ,

$$\frac{1}{T} \int_{t_l}^{t_l+T} \mu_l \int_t^{\infty} f_{\theta}(s-t) ds dt = \mu_l.$$

We are interested in the expected loss from loss period l in development period k . Denote with I_k the interval corresponding to this development period. We see that

$$\begin{aligned} \Pi_k &= \frac{1}{T} \int_{t_l}^{t_l+T} \int_{t_l \oplus I_k \cap [t, \infty)} f_{\theta}(s-t) ds dt \\ &= \frac{1}{T} \int_0^T \int_{I_k \cap [t, \infty)} f_{\theta}(s-t) ds dt. \end{aligned}$$

Usually, the loss and development periods have the same length T . We can choose $T = 1$ so that $I_k = [k-1, k]$. We get

$$\Pi_1 = \int_0^1 \int_t^1 f_{\theta}(s-t) ds dt \tag{3.2}$$

$$\Pi_k = \int_0^1 \int_{k-1}^k f_{\theta}(s-t) ds dt, \quad k \geq 2. \tag{3.3}$$

If we define the survival function

$$S_{\theta}(x) = \int_x^{\infty} f_{\theta}(y) dy$$

and the function

$$H_{\theta}(x) = \int_0^x S_{\theta}(y) dy$$

then we can rewrite (3.2) as

$$\Pi_1 = \int_0^1 (1 - S_{\theta}(1-t)) dt = 1 - H_{\theta}(1),$$

and (3.3) as

$$\Pi_k = \int_0^1 (S_\theta(k-1-t) - S_\theta(k-t)) dt = 2H_\theta(k-1) - H_\theta(k) - H_\theta(k-2).$$

We conclude that it is useful to choose the functions f_θ in such a way that we can calculate H_θ explicitly. In the next section we will do just that. We conclude this section by mentioning that the development of the variances is modeled in a similar way. In that case we do need to make sure that the functions f_θ are always positive.

3.2 A parametric family of functions

Now, we will introduce a parametric family of functions that meet the requirement of the previous section.

$$\{f(x; \beta, \gamma, \mu, \sigma) : \beta, \gamma, \sigma > 0, \mu \geq 0\},$$

where $x \in [0, \infty)$, These functions all satisfy

- $\int_0^\infty f(x; \beta, \gamma, \mu, \sigma) dx = 1 (\forall \beta, \gamma, \sigma > 0, \mu \geq 0)$.
- $f(x; \beta, \gamma, \mu, \sigma) = Cx^{\gamma-1} + o(x^{\gamma-1}) (x \downarrow 0)$ for some $C > 0$, depending on the parameters.
- $f(x; \beta, \gamma, \mu, \sigma) = Cx^{-2-\beta} + o(x^{-2-\beta}) (x \rightarrow \infty)$ for some $C \in \mathfrak{R}$, depending on the parameters.

Furthermore, there exist analytic expressions for both the first and the second primitive of the function $f(\cdot; \beta, \gamma, \mu, \sigma)$. Finally, for $\mu > 1$, $f(\cdot; \beta, \gamma, \mu, \sigma)$ will have a negative tail.

We will use an auxiliary variable y to define our parametric family, and at first ignore the dependence on the scaling parameter σ . Define

$$\mathbf{y}(x) = \int_0^x \left(1 + \frac{\left(t \mathbf{B} \left(\frac{1}{\gamma}, \frac{\beta}{\gamma} \right) \right)^\gamma}{\gamma} \right)^{\frac{1+\beta}{\gamma}} dt$$

where

$$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt$$

is the incomplete regularized beta-function. Now define

$$f(x; \lambda, \beta, \gamma, \mu, 1) \stackrel{\text{def}}{=} (1 + \beta)x^{\gamma-1} \left(B\left(\frac{1}{\gamma}, \frac{\beta}{\gamma}\right) / \gamma \right)^{\gamma} \left(1 + \frac{\left(xB\left(\frac{1}{\gamma}, \frac{\beta}{\gamma}\right) \right)^{\gamma}}{\gamma} \right)^{-1 - \frac{1+\beta}{\gamma}}$$

$$\times (1 - \mu y(x)^{\gamma}) + \gamma \mu y(x)^{\gamma-1} \left(1 + \frac{\left(xB\left(\frac{1}{\gamma}, \frac{\beta}{\gamma}\right) \right)^{\gamma}}{\gamma} \right)^{-\frac{2+2\beta}{\gamma}}.$$

Finally, we include the scale parameter σ so that

$$f(x; \lambda, \beta, \gamma, \mu, \sigma) \stackrel{\text{def}}{=} \frac{1}{\sigma} f\left(\frac{x}{\sigma}; \lambda, \beta, \gamma, \mu, 1\right). \tag{3.4}$$

We verify that

$$H(x; \lambda, \beta, \gamma, \mu, 1) = \int_0^x \int_t^{\infty} f(s; \lambda, \beta, \gamma, \mu, 1) ds dt = y(x) - \frac{\mu y(x)^{1+\gamma}}{1+\gamma}$$

and

$$H(x; \lambda, \beta, \gamma, \mu, \sigma) = \sigma H\left(\frac{x}{\sigma}; \lambda, \beta, \gamma, \mu, 1\right).$$

We will now describe the effect of the various parameters on the shape of the development function. The parameter μ is the most interesting parameter. If we choose $\mu \leq 1$, we get a positive density, whose left and right tail behavior is determined by γ and β respectively. As μ approaches 1, the bump around the mode becomes more pronounced. When $\mu > 1$, the density “falls through” the x -axis, only to approach it again as $x \rightarrow \infty$; the tail behavior is still determined by γ and β . See Figure 1. Note that from the previous section it follows that

$$\int_0^{\infty} xf(x; \beta, \gamma, \mu, \sigma) dx = \left(1 - \frac{\mu}{1 + \gamma}\right) \sigma$$

The effect of the parameters β and γ is similar to the behavior of these parameters in the parametric family of positive densities we get when we choose $\mu = 0$. The parameter β determines the right tail of the density, whereas γ determines the left-tail (near zero). In Figures 2 and 3 we chose $\mu = 5$.

Figure 1: Behavior of the density when varying μ

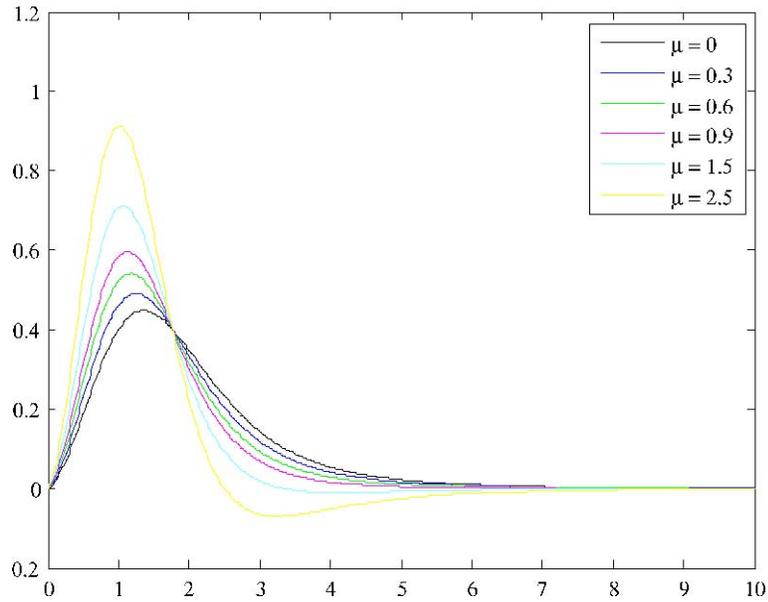
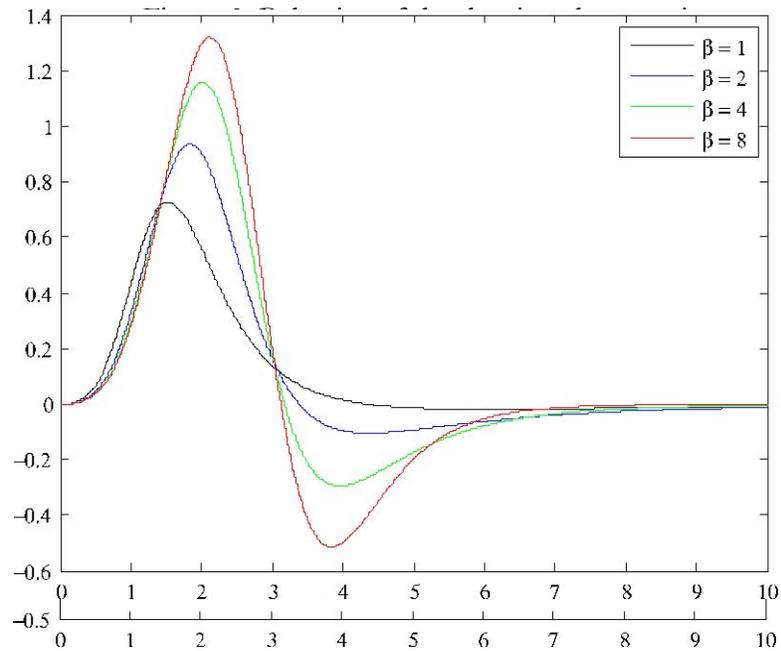


Figure 2: Behavior of the density when varying β



4. AGGREGATE OBSERVATIONS

Often we do not observe all the elements of the vector \mathbf{y} individually, but compounded in various aggregates. For instance, for certain years we may only have records of payments per quarter, while for other years payments per month are available.

Suppose we observe J aggregates. If we assume that different aggregates never involve the same payments, we can introduce a zero-one matrix \mathbf{S} with pair-wise orthogonal rows, of size $J \times 2KL$. Observing various independent sums of the elements of the vector \mathbf{y} then corresponds to $\mathbf{z} = \mathbf{S}\mathbf{y}$.

Conditionally on $\{(\mathbf{Y}^{(1)} - \mathbf{Y}^{(2)})\mathbf{1} = \mathbf{0}\}$, \mathbf{z} has a multivariate normal distribution with mean $\mathbf{S}\mathbf{E}\mathbf{y}$ and covariance matrix, $\mathbf{S}\mathbf{\Sigma}\mathbf{S}'$ where $\mathbf{\Sigma}$ is given in (2.2). The advantage of choosing a multivariate normal model is very prominent here, since in this case it is still feasible to determine the likelihood of the data \mathbf{z} .

5. ESTIMATION AND PREDICTION

We can estimate the parameters of our model by maximizing the likelihood of the data. If we call the vector of parameters $\boldsymbol{\theta}$, then we maximize

$$\text{lik}(\boldsymbol{\theta}) = P_{\boldsymbol{\theta}}(\mathbf{z} = \mathbf{z} \mid \mathbf{Y}^{(1)} - \mathbf{Y}^{(2)})\mathbf{1} = \mathbf{0}). \quad (5.1)$$

The parameter vector $\boldsymbol{\theta}$ is very high dimensional. Indeed, there are at least 16 parameters describing the (unconditional) means and variances of the $Y_{lk}^{(1)}$ and $Y_{lk}^{(2)}$. Maximizing (5.1) is a delicate affair and must involve some iterative procedure. The speed and success will depend on the algorithm which is used and, perhaps even more importantly, on the starting point. The starting point should be some ad hoc estimator, which is relatively easy to compute but still reasonably close to the true maximum likelihood estimator.

To evaluate the accuracy of our estimates we use standard theory for maximum likelihood estimation. That is, we use the Hessian of the log likelihood at the maximum likelihood estimate to approximate the Fisher information.

Typically, we are not so much interested in the parameters, as we are in a prediction of the reserve. Conditionally on the data and the equality of the row sums, the reserve has a multivariate normal distribution and we can use the conditional expectation as a prediction. The uncertainty in this prediction is a combination of the stochastic uncertainty of the model and the uncertainty in the parameter estimates.

6. SIMULATION

To evaluate the effect of conditioning on the equality of the row-sums, we conduct a simulation experiment. We estimate the reserve with conditioning on equal row-sums, using the run-off tables simultaneously, as described in this paper. We refer to this approach as the *joint* method. For comparison, we also estimated the reserve without conditioning on the row-sums, essentially only using the paid table. We call this approach the *marginal* method. Of course, the marginal method is much easier, as it involves no conditioning. However, results in this section show that the more complicated joint method does produce better results.

We carry out the following simulation experiment. We consider a set of actual insurance data that, for reasons of privacy, we have made anonymous by multiplying with some undisclosed factor. We fit a model using the parametric family of densities described in section 3 for the development curves of the expectation and variance of both the paid and the incurred table. This results in a 19 dimensional parameter θ_0 , which contained

- $2 \times 4 = 8$ parameters for the two expectation development curves.
- $2 \times 4 = 8$ parameters for the two variance development curves.
- Three parameters for the exposure (β) to account for two regime changes.

We define: R , the reserve as the sum of future payments and \hat{R} , the estimator for R .

For this data set we estimate the reserve \hat{R}_0 for the paid table and its variance \hat{V}_0 conditioned on the aggregated data and taking into account both the stochastic uncertainty and the parameter uncertainties. We find $\hat{R}_0 = 5.24$ and $\hat{V}_0 = 2.04$.

Next, we simulate two entire tables (paid and incurred) from the multivariate normal model determined by the estimated parameter vector θ_0 , and repeat this about 6000 times. By using the estimated parameter vector, we make sure that our simulated data resembles realistic data. Of course, for each simulated data set, we know the “true” reserve R , as the sum of total simulated future payments. Hereafter, the estimated reserve \hat{R} is based on the simulated data set of historical payments. The error in the estimated reserve, as the difference between R and \hat{R} , is compared for the two methods.

6.1 Reserve Estimation

Denote R as the true reserve in a given simulated data set, \hat{R}_1 and \hat{R}_2 as the estimates for the reserve for the joint and marginal methods, respectively. One of the most important measures for the quality of a prediction is the Mean Squared Error (MSE). Our simulation showed that

$$\begin{aligned} E(R - \hat{R}_1)^2 &= 2.18 \\ E(R - \hat{R}_2)^2 &= 6.90. \end{aligned}$$

Clearly, by using both tables simultaneously we achieve superior performance. We remark here, that using more simulations would not have changed this conclusion. In Figure 4 we show the convergence of the average of the squared error for the joint method as the number of simulations increases. We see that the average has sufficiently stabilized towards the end. The convergence for the marginal method is very similar.

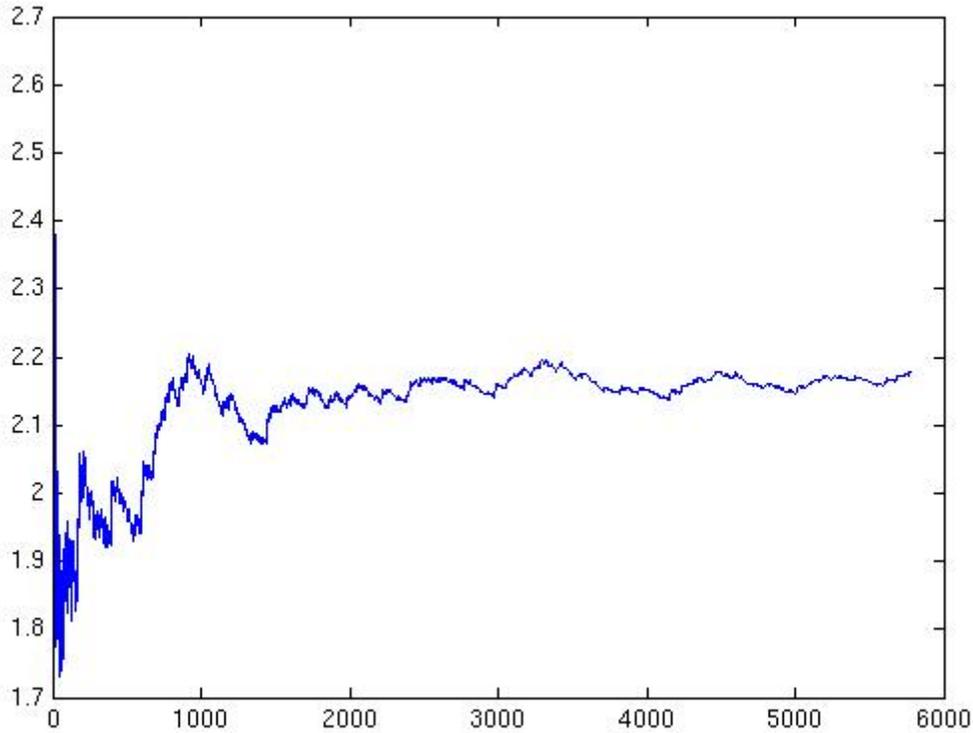


Figure 4: Convergence of the average mean squared error for method 1.

The bias of the estimators is also important

$$E(\hat{R}_1 - R) = -0.18$$

$$E(\hat{R}_2 - R) = -0.32$$

We note that the bias of both methods is very small compared to the MSE. It is not surprising that we find a similar bias for both methods, since the expectation structure in both models is the same. Recalling that the MSE consists of the estimator's variance and its squared bias, we conclude that large MSE of the marginal method is an immediate consequence of its inability to correctly estimate this covariance structure.

It is also interesting to see how well both methods do at determining the accuracy of the estimate. We have calculated a conditional variance of the estimated reserve, given the data, taking into account the uncertainty in the parameter estimates. This leads to

$$\text{median}(\hat{V}_1) = 1.63$$

$$\text{median}(\hat{V}_2) = 4.28 .$$

Since the reserve estimates are almost unbiased, these values should be close to the mean-squared errors. This is not the case; both methods underestimate the variance. This is also clearly visible in Figure 5 where we plot the histogram of the estimated variances. The skewness indicates that we frequently underestimate the variance. This is a problem, when we want to estimate percentiles. We address this issue in the next sub-section.

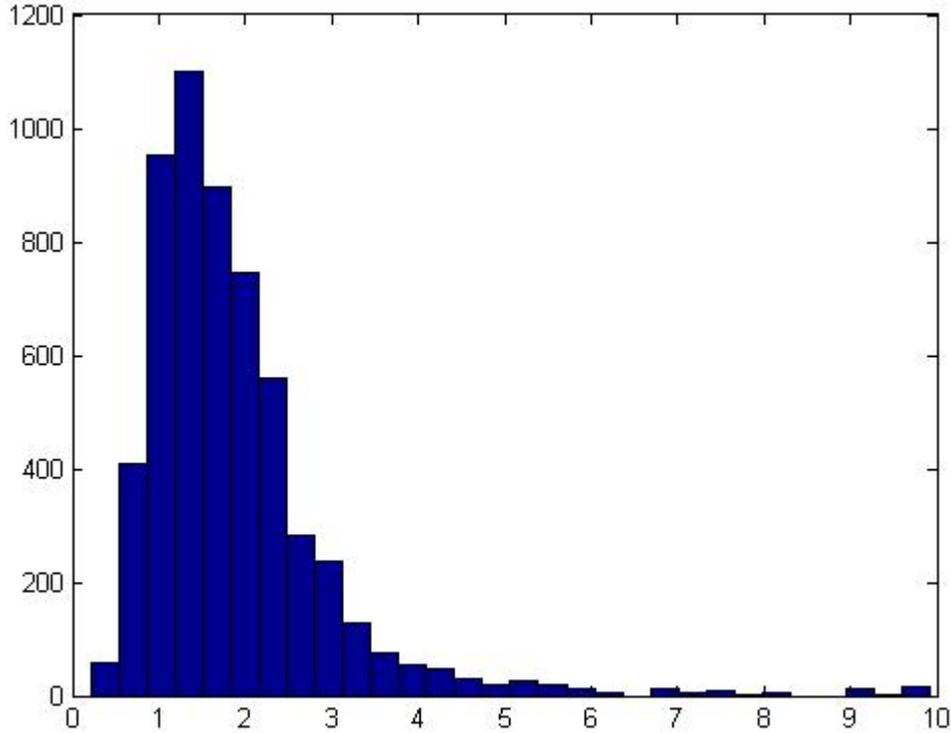


Figure 5: Histogram of the estimated variance for the joint method.

6.2 Estimating Percentiles

For loss reserving it is typically not sufficient to only have a point estimate of the reserve; percentiles are also needed. In this sub-section we discuss why the standard approach to estimating the percentiles does not work well in our case. We also provide an alternative.

The standard way of estimating percentiles is based on the following idea. When we estimate R by \hat{R} , and we estimate the variance by \hat{V} , we assume that the standardized residuals are approximately standard normal distributed. That is,

$$\frac{R - \hat{R}}{\sqrt{\hat{V}}} \sim N(0,1).$$

Then we can use the percentiles of the standard normal to find approximate percentiles for the reserve. We used this method to estimate the 75% and the 95% percentiles. To verify the results, we look at the percentage of times the true (simulated) reserve was larger than the estimated percentile. This gives

- $P(R > \hat{q}_{75}) = 0.32$ and $P(R > \hat{q}_{95}) = 0.12$ for the joint method.
- $P(R > \hat{q}_{75}) = 0.36$ and $P(R > \hat{q}_{95}) = 0.19$ for the marginal method.

Both methods seem to underestimate the percentiles, and we checked that this effect does not disappear as we increase the number of simulations. This is very troubling as it will lead to overoptimistic loss reserving.

In Figure 6 we plot the histogram of the standardized residuals for the joint method, and note that the distribution is not standard normal at all! Not only is it skewed, but also its mean is 0.24 instead of 0 and its variance is 1.49 instead of 1. This explains why the percentiles are not estimated accurately. The problem originates with the underestimation of the variance we discussed in the previous section, since having a small variance leads to a small percentile.

We conclude that the standard approach to estimating the percentiles does not work well. Therefore, we would like to suggest an alternative approach. The idea is simple: use the distribution of Figure 6 instead of the standard normal to calculate percentiles. This is essentially an application of the bootstrap. This would lead to a 75th percentile of 0.95, (instead of 0.67 for the standard normal) and a 95th percentile of 2.37 (instead of the familiar 1.65). For our original data set with $\hat{R}_0 = 5.24$ and $\hat{V}_0 = 2.04$ this means that the percentiles for the reserve are given by

- $\hat{q}_{75} = 6.60$ and $\hat{q}_{95} = 8.62$ using the bootstrap method.
- $\hat{q}_{75} = 6.20$ and $\hat{q}_{95} = 7.59$ using the normal method.

The relative difference between the two methods becomes more pronounced for higher percentiles, mainly because the relative contribution of the estimated reserve \hat{R}_0 diminishes.

Performing the many simulations needed to determine the distribution of the standardized residuals is a substantial computational burden. It took us four days to create Figure 6. In certain applications this is prohibitive.

Although the distribution of Figure 6 is specific to our particular data set, it is certainly conceivable that similar distributions would result from other data sets. Indeed, for data sets concerning similar insurance products this seems plausible at least. This suggests the following approach. We perform the simulations for a number of different data sets with varying characteristics. Then, when confronted with a new data set, we choose the histogram that is most appropriate, and use it instead of the standard normal to calculate percentiles.

Another suggestion to deal with this problem is judging the standardized residuals of the original loss triangle data set, given the parameter estimates. While the kurtosis of these residuals differs from the normality 3-value, the percentiles for loss reserves should be adjusted by taking these percentiles from a t -distribution, whereby the degree of freedom depends on the magnitude of the difference for the calculated kurtosis and the value 3.

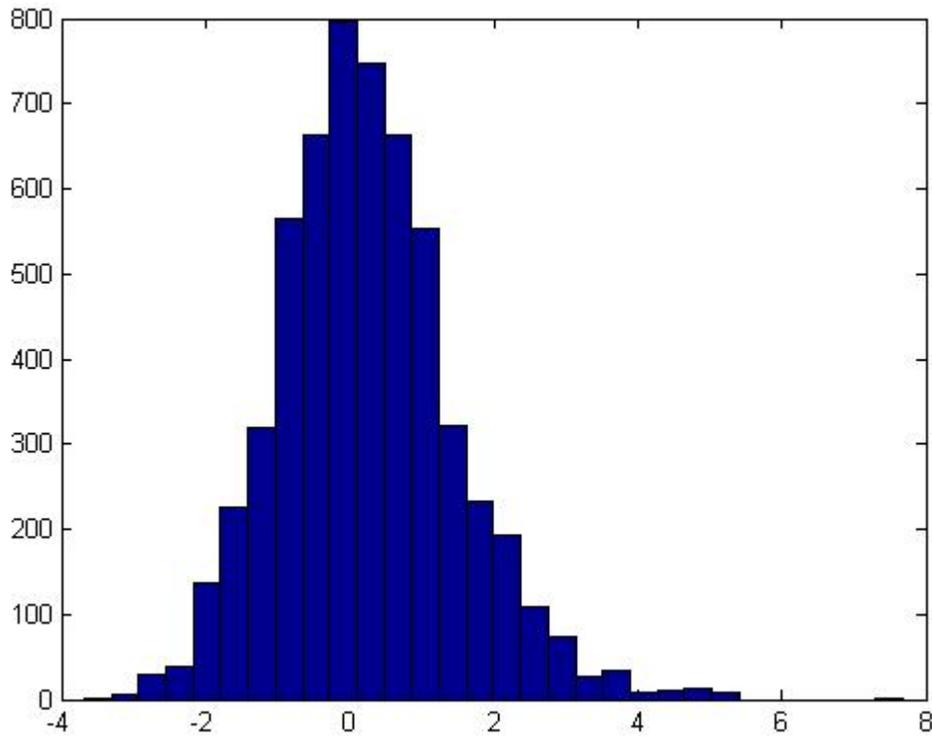
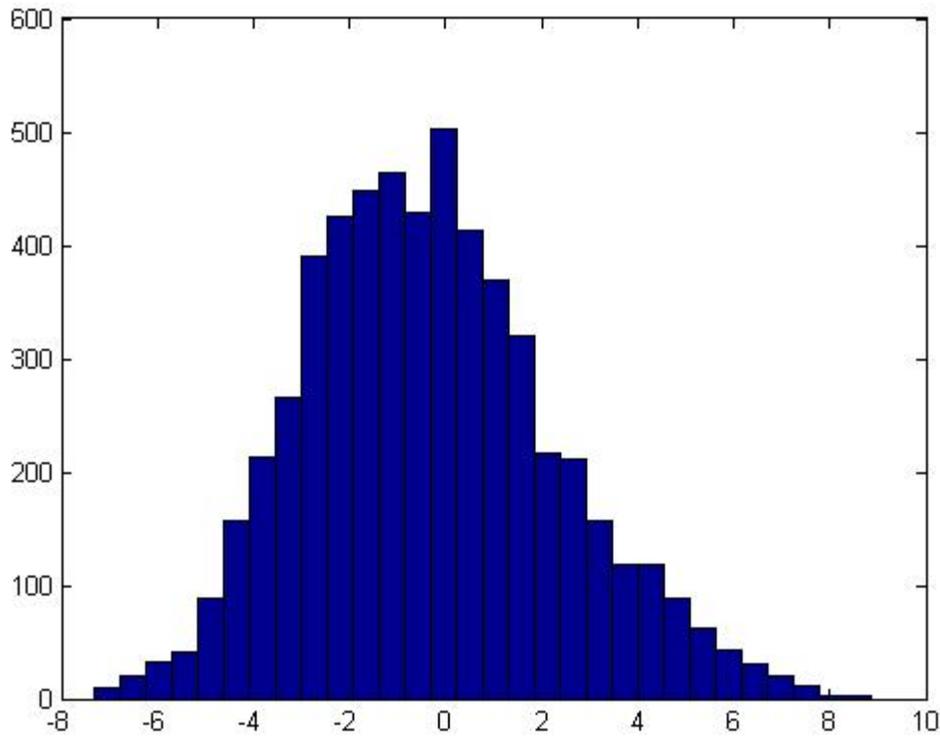
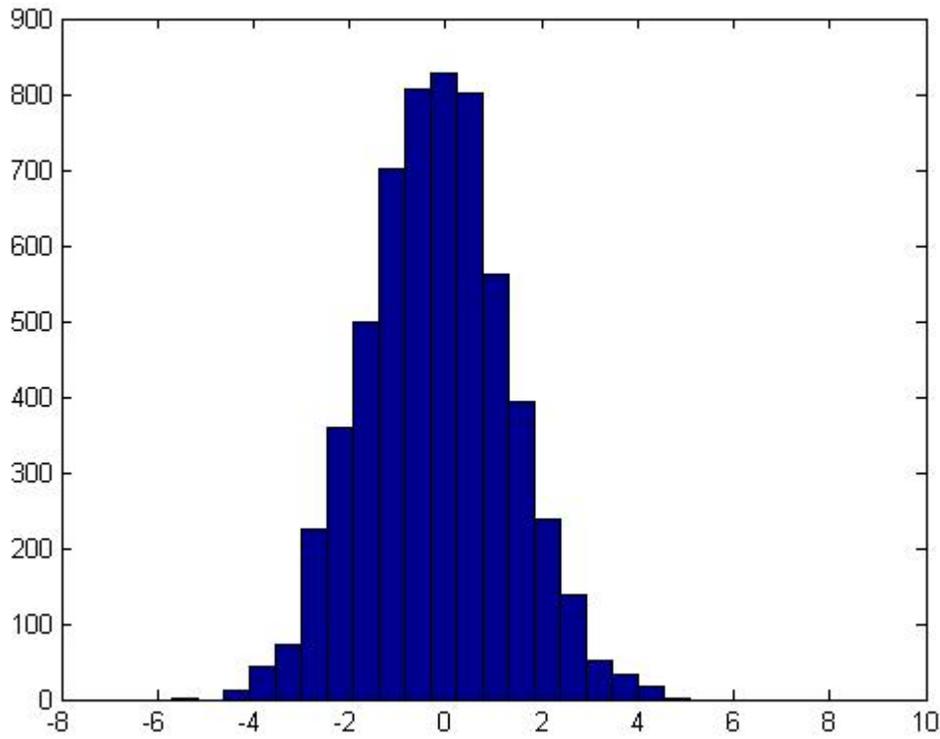


Figure 6: Histogram of the standardized residuals for the joint method.

7. CONCLUSION

The incurred loss of an insurer consists of payments on claims and reserves for claims that have been reported. As all claims are settled eventually, the cumulative paid and incurred losses for a given loss period become equal. On the basis of this observation, we construct a joint model for the paid and incurred loss arrays. In our description, each follows a multivariate normal distribution, conditioned on equality of the total paid and incurred losses for a given loss period. On the basis of this model, we make predictions for future payments.



Figures 7 and 8: Histogram of the difference between the true and estimated reserve based on the joint model (top) and the marginal model (bottom)

A rather technical, but important feature of our model is the use of a new parametric family of functions that are ideally suited for modeling development curves.

We have compared the performance of the joint model of the paid and incurred losses to an approach where we analyze only the paid table. In Figure 7 and 8 we present the results of a simulation study. These figures show histograms of the difference of the true and predicted reserves for both methods. While both methods are approximately unbiased, the one based on the joint model has much smaller variance. A more detailed discussion of this result is found in the previous section, but here we conclude that joint modeling is to be preferred over utilizing only the paid table.

Since the practice of loss reserving also takes the distribution of the reserve into account, we have studied the estimation of percentiles as well. We noted that inference from the normal assumptions does not produce good results. In fact, the results would lead to over-optimistic assessment of economic capital and prudence margins, which is, of course, to be avoided. We have proposed an alternative approach based on the bootstrap. It entails performing many simulations, to replace the assumed standard normal distribution of the standardized residuals with a more accurate description. Carrying out this method requires substantial computational effort, which in practice is only feasible on a highly aggregated level.

APPENDIX A

For ease of reference, we recall a well-known fact about the multivariate normal distribution.

Consider a random vector \mathbf{X} , which is distributed according to the multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Suppose we partition \mathbf{X} into two sub-vectors

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix}.$$

Correspondingly, we write

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Now, if $\det(\boldsymbol{\Sigma}_{22}) > 0$, then the conditional distribution of $\mathbf{X}^{(1)}$ given $\mathbf{X}^{(2)}$ is multivariate normal with mean

$$\boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})$$

and covariance matrix

$$\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}.$$

8. REFERENCES

- [1] England, P. and R.J. Verral, "Analytic and bootstrap estimates of prediction errors in claims reserving," **1999**, *Insurance: Mathematics and Economics* 25, 281-291.
- [2] Mack, T., "Distribution-free calculation of the standard error of chain ladder reserve estimates," **1993**, *ASTIN Bulletin* 23, 213-225.
- [3] Quarg, G. and T. Mack,, "Munich chain ladder: A reserving method that reduces the gap between IBNR projections based on paid losses and IBNR projections based on incurred losses," **2004**, *Blätter DGVFM*, Heft 4, 597-630.
- [4] Renshaw, A.E. and R.J. Verral, "A stochastic model underlying the chain ladder technique," **1998**, *British Actuarial Journal* 4, 903-923.
- [5] Schmidt, K.D., "A bibliography on loss reserving," 2007, www.math.tu-dresden.de/sto/schmidt/dsvm/reserve.pdf

Biographies of Authors

Bouke Posthuma is Founder and President and Consulting Actuary of Posthuma Partners, a Dutch actuarial firm. He was chairman of the ASTIN board in the Netherlands and a member of the International ASTIN Committee. This year, he received the Honorary Member Award of Recognition and Merit from the Actuarial Association of The Netherlands (Actuariel Genootschap) www.posthuma-partners.nl.

Eric Cator is an Associated Professor at the Faculty of Information Technology and Systems of the Delft Technical University and holds a PhD in mathematics. E.A.Cator@TUDelft.nl.

Wim Veerkamp is a Consulting Actuary with Posthuma Partners and involved in loss reserving. He holds a master in econometrics, a PhD in psychometrics and a minor in actuarial science. He is an associated member of the Actuarial Association of The Netherlands (Actuariel Genootschap). wim.veerkamp@posthuma-partners.nl.

Erik van Zwet is an Associated Professor at the Mathematical Institute of Leiden University and holds a PhD in mathematics. evanzwet@math.leidenuniv.nl