Bayesian Approach for Prediction Error in Chain-Ladder Claims Reserving

Ji Yao

Zurich Financial Services

126 Hagley Road

Birmingham, B16 9PF, UK

Tel: 0044-(0)121 697 4416

E-mail: ji.yao@uk.zurich.com

Abstract

Bayesian approach is applied to calculate the prediction error in chain ladder claim reserving. The philosophy of Bayesian approach to prediction error is first introduced and compared with Frequentist approach. All parameters in the model are then estimated by Bayesian approach, where three types of prior distributions are used. Posterior distribution could be standard student t-distribution and inverse Gamma distribution under non-informative and conjugate prior distribution. Finally, theory is illustrated by numerical examples. It concludes that prediction error by Bayesian approach is much higher than that by Frequentist approach.

Keywords. Bayesian approach; Prediction error; Chain ladder; Reserving; Student t distribution; Inverse Gamma distribution.

1. INTRODUCTION

The prediction error of chain-ladder claim reserving has been widely studied in the last twenty years. Based on three key assumptions, a close-form formula is derived in [1]. In [2] a recursive formula solution is provided, however it gives different results to [1] under the same three key assumptions. [3] and [4] present a nice picture of stochastic claim reserving but the formula used to calculate prediction error is same as [1]. Recently [5] derives a close-form formula based on time-series model, which gives same numerical results as [2]. The debate on which formula gives most accurate estimation of prediction error attracts lots of interests [6]-[8].

The approach taken so far to derive prediction error is classified as Frequentist approach in statistical point of view. Typically prediction error split into process error and parameter error. The maximum likelihood estimation (MLE) is used to estimate all parameters in the model and the process error is calculated based on these MLE parameters. Then by assuming all MLE parameters are random variables, the parameter error is considered by calculating the variance of MLE parameters around its true values.

Paralleling to Frequentist approach, Bayesian approach is another approach in statistics. The debate between Frequentist and Bayesian has been lasting for two century without clear outcome. However, in the context of prediction error for chain ladder claims reserving there is little study on Bayesian approach. So Bayesian approach is applied in this paper to prediction error.

There are different models and assumptions about stochastic reserving [3], [4]. However, Mack's model is one of most widely used model. So this paper focus on this model, but apparently the general theory could be applied to other models.

Bayesian approach is mentioned and studied in [3], [4], [8] and [9]. However, not all parameters are analyzed in a Bayesian approach: for example, although the parameter σ^2 [1] in Mack's model is defined as unknown, it was always assumed that this is known [8] or use the MLE variance as a "plug-in" parameter [3], [4]. These approaches are termed as a semi-Bayesian approach in this paper. In [9], although this parameter is included in Bayesian analysis, it largely uses simulation techniques by the nature of paper.

The remainder of the paper proceeds as follows. Section 2 introduces the basic claim reserving model and the Bayesian approach to prediction error. Section 3 illustrates assumptions of the model. Section 4 calculates the prediction error under assumption of Mack's model. Section 5 estimates parameters in the model by Bayesian approach. Numerical examples are presented in Section 6 and finally conclusion are made in Section 7.

2. BAYESIAN APRROACH TO PREDICTION ERROR

Let $X_{i,j}$ be the random variables of accumulated claim amounts of the accident year $i\ (1 \le i \le N)$ and development year $j\ (1 \le j \le N)$. By the end of N^{th} year, the left-up triangle of $X_{i,j}$ have been observed, that is

$$\begin{array}{ccccc} x_{1,1} & x_{1,2} & \cdots & \cdots & x_{1,N} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,N-1} & \\ \vdots & & \iddots & & \\ \vdots & \iddots & & & \\ x_{N,1} & & & & \end{array},$$

which are denoted in lower case as all are observed and therefore fixed. The whole observed triangle is denoted as $\mathbf{x}$. The task of claims reserving is to project the ultimate claim amounts based on this observation. In this paper it is assumed that the claim amount of 1st year has fully developed and therefore $X_{i,N}\ \ (2 \le i \le N)$ are considered the ultimate claim amounts to be estimated.

Among various reserving methods, chain-ladder method is the most widely used. Given $\mathbf{x}$, the development factor f_j is estimated by

$$\hat{f}_j \big| \mathbf{x} = \sum_{i=1}^{N-j} x_{i,j+1} \Big/ \sum_{i=1}^{N-j} x_{i,j} \tag{2.1}$$

and the ultimate claim amount for the i^{th} $(i \ge 2)$ year is estimated as

$$\hat{x}_{i,N} \big| \mathbf{x} = x_{i,N-i+1} \prod_{j=N-i+1}^{N-1} \hat{f}_j \,. \tag{2.2}$$

It is important to note that, at least from Bayesian point of view, given the observation $\mathbf{x}$ (2.1) and (2.2) are deterministic in nature.

Having the estimated ultimate claim amount, it is important to know how accurate this estimator is and what the prediction error is. One measure for this purpose is the mean square error (MSE), which is used in this paper. For each individual year, MSE is defined as

$$MSE_i = E\left[\left(\hat{x}_{i,N} - X_{i,N} \right)^2 \Big| \mathbf{x} \right], \tag{2.3}$$

and for aggregation of all years, MSE is defined as

$$MSE = E\left[\left(\sum_{i=2}^{N} \hat{x}_{i,N} - \sum_{i=2}^{N} X_{i,N} \right)^2 \Bigg| \mathbf{x} \right],$$

where the summation starts from 2nd year because the ultimate claim amount of 1st year has already been observed.

Because $\hat{x}_{i,N}$ is a fixed number given $\mathbf{x}$, (2.3) becomes

$$MSE_i = E\left[\left(\hat{x}_{i,N} - X_{i,N} \right)^2 \Big| \mathbf{x} \right]$$

$$= E\left[\left\{ \left(\hat{x}_{i,N} - E\left(X_{i,N}\right) \right)^2 - 2\left(\hat{x}_{i,N} - E\left(X_{i,N}\right) \right)\left(X_{i,N} - E\left(X_{i,N}\right) \right) + \left(X_{i,N} - E\left(X_{i,N}\right) \right)^2 \right\} \Big| \mathbf{x} \right]$$

$$= \left(\hat{x}_{i,N} \big| \mathbf{x} - E\left(X_{i,N} \big| \mathbf{x} \right) \right)^2 + E\left[\left(X_{i,N} - E\left(X_{i,N}\right) \right)^2 \Big| \mathbf{x} \right].$$

If $\hat{x}_{i,N}$ is an un-biased estimate of $X_{i,N}$, that is

$$\hat{x}_{i,N} \left| \mathbf{x} = E\left(X_{i,N} \middle| \mathbf{x} \right) \right.$$

which is the case for chain ladder reserving method [1], MSE is further simplified as

$$MSE_i = E\left[\left(X_{i,N} - E\left(X_{i,N} \right) \right)^2 \middle| \mathbf{x} \right] = \mathrm{var}\left(X_{i,N} \middle| \mathbf{x} \right), \tag{2.4}$$

If $\hat{x}_{i,N}$ is biased, (2.4) only gives a lower bound of MSE. An additional bias error is necessary to calculate the total MSE [10]. Similarly, the minimum MSE for aggregate ultimate claims amount is

$$MSE_i = \mathrm{var}\left(\sum_{i=2}^{N} X_{i,N} \middle| \mathbf{x} \right). \tag{2.5}$$

A comparison with Frequentist approach is necessary at this stage. In the Frequentist approach, MSE split into two parts, that is

$$MSE_i = \mathrm{var}\left(X_{i,N} \middle| \mathbf{x} \right) + \left(E\left(X_{i,N} \middle| \mathbf{x} \right) - \hat{x}_{i,N} \right)^2, \tag{2.6}$$

which is exactly copied from [1] but denoted in current definition of symbol. Comparing (2.4) with (2.6) plausibly indicates that Bayesian approach misses one term. However, this is not the case because of the different meaning of $\mathrm{var}\left(X_{i,N} \middle| \mathbf{x} \right)$. In Frequentist approach, $\mathrm{var}\left(X_{i,N} \middle| \mathbf{x} \right)$ is actually the variance of $X_{i,N}$ conditional on the MLE of all parameters. So stringently it is better to express (2.6) in this way

$$MSE_i = \mathrm{var}\left(X_{i,N} \middle| MLE\ parameters \right) + \left(E\left(X_{i,N} \middle| \mathbf{x} \right) - \hat{x}_{i,N} \right)^2.$$

By contrast, Bayesian approach includes all uncertainty in $\mathrm{var}\left(X_{i,N} \middle| \mathbf{x} \right)$. The key to Bayesian approach is to calculate the posterior distribution of all parameters, which contain uncertainty of parameters, and therefore posterior distribution of ultimate claims amount.

3. MODEL ASSUMPTIONS

To proceed with this analysis, a particular model has to be chosen. The Mack's model is used in this paper, but the methodology can be applied to other models. The key assumptions are

$$E\left(X_{i,j+1}\middle| X_{i,1},...,X_{i,j}\right)= f_j X_{i,j};$$

$$\left\{X_{i,1},...,X_{i,N}\right\},\ \left\{X_{k,1},...,X_{k,I}\right\} \text{ are independent;} \tag{3.1}$$

$$\text{and } \operatorname{var}\left(X_{i,j+1}\middle| X_{i,1},...,X_{i,j}\right)=\sigma_j^2 X_{i,j}.$$

Mack's model is claimed to be distribution free, that is, the results from Mack's model doesn't depend on the assumption of the conditional distribution of $X_{i,j}$. However to make this model comparable to other models and make simulation possible, it was often slightly changed to assume that $X_{i,j}$ is Normal distributed with mean $f_j X_{i,j}$ and variance $\sigma_j^2 X_{i,j}$ [3], [8], that is

$$X_{i,j+1}\left|\left(X_{i,1},...,X_{i,j}\right)\right.\sim N\left(f_j X_{i,j},\sigma_j^2 X_{i,j}\right). \tag{3.2}$$

Let $Y_{i,j}=X_{i,j+1}/X_{i,j}$, then this assumption is equivalent to

$$Y_{i,j}\left|\left(X_{i,1},...,X_{i,j}\right)\right.\sim N\left(f_j,\sigma_j^2/X_{i,j}\right).$$

Lower case $y_{i,j}$ is also defined as $x_{i,j+1}/x_{i,j}$ if both $x_{i,j}$ and $x_{i,j+1}$ are known.

Apparently Normal distribution is not the only assumption that can be made. Actually, it is also not a theoretically good one as Normal distribution could take negative value while accumulative claims amount never could. However, as one of the aims of this paper is to compare the analytical results with Frequentist approach, the same distribution assumption is made for consistency [3], [4], [8].

To simplify further denotation, these vectors are defined

$$\mathbf{f} = \left(f_1, f_2, ..., f_{N-1} \right)$$

and

$$\boldsymbol{\sigma}^2 = \left(\sigma_1^2, \sigma_2^2, ..., \sigma_{N-1}^2 \right).$$

4. CALCULATION OF PREDICTION ERROR

To calculate (2.4), the first step of Bayesian approach is to calculate the posterior distribution of all parameters by

$$p\left(\mathbf{f}, \boldsymbol{\sigma}^2 \middle| \mathbf{x}\right) \propto p\left(\mathbf{x} \middle| \mathbf{f}, \boldsymbol{\sigma}^2\right) \cdot p\left(\mathbf{f}, \boldsymbol{\sigma}^2\right) \tag{4.1}$$

where $p\left(\mathbf{f}, \boldsymbol{\sigma}^2\right)$ is the joint prior distribution of $\mathbf{f}$ and $\boldsymbol{\sigma}^2$, and $p\left(\mathbf{f}, \boldsymbol{\sigma}^2 \middle| \mathbf{x}\right)$ is the joint posterior distribution. $p\left(\mathbf{x} \middle| \mathbf{f}, \boldsymbol{\sigma}^2\right)$ is determined by assumptions of model. Using independency in (3.1) and (3.2), this probability is

$$p\left(\mathbf{x} \middle| \mathbf{f}, \boldsymbol{\sigma}^2\right) = \prod_{i=1}^{N} p\left(x_{i,1}, x_{i,2}, ..., x_{i,N-i+1} \middle| \mathbf{f}, \boldsymbol{\sigma}^2\right)$$

$$= \prod_{i=1}^{N} \left[p\left(x_{i,1} \middle| \mathbf{f}, \boldsymbol{\sigma}^2\right) \prod_{j=2}^{N-i+1} p\left(x_{i,j} \middle| x_{i,1}, x_{i,2}, ..., x_{i,j-1}, \mathbf{f}, \boldsymbol{\sigma}^2\right) \right]$$

$$= \left[\prod_{i=1}^{N} p\left(x_{i,1} \middle| \mathbf{f}, \boldsymbol{\sigma}^2\right) \right] \prod_{j=2}^{N} \left\{ \prod_{i=1}^{N-j+1} p\left(x_{i,j} \middle| x_{i,j-1}, x_{i,j-2} ..., x_{i,1}, \mathbf{f}, \boldsymbol{\sigma}^2\right) \right\}$$

$$\propto \prod_{j=2}^{N} \left\{ \prod_{i=1}^{N-j+1} \left\{ \frac{1}{\sqrt{2\pi\left(\sigma_{j-1}^2 x_{i,j-1}\right)}} \exp\left[-\frac{\left(x_{i,j} - f_{j-1} x_{i,j-1}\right)^2}{2\left(\sigma_{j-1}^2 x_{i,j-1}\right)} \right] \right\} \right\}$$

$$\propto \prod_{j=1}^{N-1}\left\{\prod_{i=1}^{N-j}\left\{\frac{1}{\sqrt{\sigma_j^2}}\exp\left[-\frac{\left(y_{i,j}-f_j\right)^2}{2\left(\sigma_j^2/x_{i,j}\right)}\right]\right\}\right\}. \tag{4.2}$$

There are several options for the prior distribution, which will be discussed in details in Section 5. However, to make analysis tractable one common assumption is made for all these prior distributions: pair $\left(f_j,\sigma_j^2\right)$ is independent to other pairs, i.e.

$$p\left(\mathbf{f},\boldsymbol{\sigma}^2\right)=\sum_{j=1}^{N-1}p\left(f_j,\sigma_j^2\right). \tag{4.3}$$

Note that non-informative prior used in [3], [4] and [8] satisfies this assumption. Substitute (4.2) and (4.3) into (4.1), there is

$$p\left(\mathbf{f},\boldsymbol{\sigma}^2\middle|\mathbf{x}\right)\propto \prod_{j=1}^{N-1}\left\{\prod_{i=1}^{N-j}\left\{\frac{1}{\sqrt{\sigma_j^2}}\exp\left[-\frac{\left(y_{i,j}-f_j\right)^2}{2\left(\sigma_j^2/x_{i,j}\right)}\right]\right\}\cdot p\left(f_j,\sigma_j^2\right)\right\} \tag{4.4}$$

which shows that the joint posterior distribution can be factorized. This gives an important conclusion that if the prior is believed to be independent, the joint posterior distribution of pair $\left(f_j,\sigma_j^2\right)\middle|\mathbf{x}$ is also independent to others. And each pair has a similar formation as

$$\begin{aligned}p\left(f_j,\sigma_j^2\middle|\mathbf{x}\right)&\propto \prod_{i=1}^{N-j}\left\{\frac{1}{\sqrt{\sigma_j^2}}\exp\left[-\frac{\left(y_{i,j}-f_j\right)^2}{2\left(\sigma_j^2/x_{i,j}\right)}\right]\right\}\cdot p\left(f_j,\sigma_j^2\right)\\ &\propto \left(\sigma_j^2\right)^{-(N-j)/2}\exp\left[-\frac{1}{2\sigma_j^2}\sum_{i=1}^{K}x_{i,j}\left(y_{i,j}-f_j\right)^2\right]\cdot p\left(f_j,\sigma_j^2\right).\end{aligned} \tag{4.5}$$

So the analysis on (4.4) can be done individually on each component.

The second step of Bayesian approach is to calculate the marginal posterior distribution $p\left(f_j\middle|\mathbf{x}\right)$ and $p\left(\sigma_j^2\middle|\mathbf{x}\right)$. This could be calculated by integrating out the unwanted variables in the joint posterior distribution as

$$p\left(f_j \middle| \mathbf{x}\right) = \int_0^{+\infty} p\left(f_j \middle| \sigma_j^2, \mathbf{x}\right) p\left(\sigma_j^2 \middle| \mathbf{x}\right) d\sigma_j^2$$

$$= \int_0^{+\infty} p\left(f_j, \sigma_j^2 \middle| \mathbf{x}\right) d\sigma_j^2 \tag{4.6}$$

and similarly

$$p\left(\sigma_j^2 \middle| \mathbf{x}\right) = \int_0^{+\infty} p\left(\sigma_j^2 \middle| f_j, \mathbf{x}\right) p\left(f_j \middle| \mathbf{x}\right) df_j$$

$$= \int_0^{+\infty} p\left(f_j, \sigma_j^2 \middle| \mathbf{x}\right) df_j \,. \tag{4.7}$$

In cases where the integration in (4.6) and (4.7) cannot be analytically calculated, numerical simulation techniques have to be used to calculate the posterior marginal distribution. However, as will be shown in section 5, these two integrations are possible for Mack's model under some prior distributions, which leads to interesting standard statistics distribution.

Having derived the marginal posterior distribution, the final step is to calculate the variance in (2.4). In this paper, this is done in a recursive way. Because any pair $\left(f_j, \sigma_j^2\right)\big|\mathbf{x}$ is independent to another pair $\left(f_k, \sigma_k^2\right)\big|\mathbf{x}$ $(j \neq k)$, $\left(f_j, \sigma_j^2\right)\big|\mathbf{x}$ is also independent to $X_{i,k+1}\big|\mathbf{x}$ if $j \neq k$. Using this independency, the mean of $X_{i,j+1}\big|\mathbf{x}$ (for $j \geq i$) is

$$E\left(X_{i,j+1} \middle| \mathbf{x}\right) = E\left(f_j X_{i,j} \middle| \mathbf{x}\right) = E\left(f_j \middle| \mathbf{x}\right) E\left(X_{i,j} \middle| \mathbf{x}\right) \tag{4.8}$$

and the second central moment is

$$E\left(X_{i,j+1}^2 \middle| \mathbf{x}\right) = E\left\{\left[\left(f_j X_{i,j}\right)^2 + \sigma_j^2 X_{i,j}\right] \middle| \mathbf{x}\right\} = E\left(f_j^2 \middle| \mathbf{x}\right) E\left(X_{i,j}^2 \middle| \mathbf{x}\right) + E\left(\sigma_j^2 \middle| \mathbf{x}\right) E\left(X_{i,j} \middle| \mathbf{x}\right).$$

So the variance of $X_{i,k+1}\big|\mathbf{x}$ is

$$\operatorname{var}\left(X_{i,j+1}\middle|\mathbf{x}\right)=E\left(X_{i,j+1}^2\middle|\mathbf{x}\right)-E^2\left(X_{i,j+1}\middle|\mathbf{x}\right)$$

$$=E\left(f_j^2\middle|\mathbf{x}\right)E\left(X_{i,j}^2\middle|\mathbf{x}\right)+E\left(\sigma_j^2\middle|\mathbf{x}\right)E\left(X_{i,j}\middle|\mathbf{x}\right)-E^2\left(f_j\middle|\mathbf{x}\right)E^2\left(X_{i,j}\middle|\mathbf{x}\right)$$

$$=\operatorname{var}\left(f_j\middle|\mathbf{x}\right)E^2\left(X_{i,j}\middle|\mathbf{x}\right)+E\left(f_j^2\middle|\mathbf{x}\right)\operatorname{var}\left(X_{i,j}\middle|\mathbf{x}\right)+E\left(\sigma_j^2\middle|\mathbf{x}\right)E\left(X_{i,j}\middle|\mathbf{x}\right). \tag{4.9}$$

The value of $E\left(f_j\middle|\mathbf{x}\right)$, $\operatorname{var}\left(f_j\middle|\mathbf{x}\right)$ and $\operatorname{var}\left(\sigma_j^2\middle|\mathbf{x}\right)$ can be calculated from the posterior distribution in (4.6) and (4.7). As for the first term $X_{i,N-i+1}$ in the recursive formula, because there is

$$X_{i,N-i+1}\big|\mathbf{x}=x_{i,N-i+1},$$

the mean of it is

$$E\left(X_{i,N-i+1}\middle|\mathbf{x}\right)=x_{i,N-i+1} \tag{4.10}$$

and the variance is

$$\operatorname{var}\left(X_{i,N-i+1}\middle|\mathbf{x}\right)=0. \tag{4.11}$$

So by recursive formula (4.8), (4.9) and initial value (4.10), (4.11), MSE in (2.4) can be calculated for any i.

A comparison with the results from MLE approach is very interesting. One difference is the value of σ_j^2, which is due to the different philosophy between Frequentist and Bayesian approach. In Frequentist approach, the MLE $\hat{\sigma}_j^2$ is used while in Bayesian approach the mean of $\sigma_j^2\big|\mathbf{x}$ is used. As will be shown in Section 5, this difference between MLE and mean is very large when there are few data available, such as in the tail of reserving triangle.

Another is that MSE of Bayesian approach is larger than that of Frequentist approach. Because Frequentist approach always splits the total MSE into process error and parameter error, for comparison purpose (4.9) is artificially split into process component and parameter component, denoted as $\mathrm{var}_{pro}\left(X_{i,j}\middle|\mathbf{x}\right)$ and $\mathrm{var}_{par}\left(X_{i,j}\middle|\mathbf{x}\right)$, respectively, that is

$$\mathrm{var}\left(X_{i,j}\middle|\mathbf{x}\right)=\mathrm{var}_{pro}\left(X_{i,j}\middle|\mathbf{x}\right)+\mathrm{var}_{par}\left(X_{i,j}\middle|\mathbf{x}\right) \tag{4.12}$$

Substitute (4.12) into (4.9), (4.9) becomes

$$\begin{aligned}&\mathrm{var}_{pro}\left(X_{i,j+1}\middle|\mathbf{x}\right)+\mathrm{var}_{par}\left(X_{i,j+1}\middle|\mathbf{x}\right)\\&=\mathrm{var}\left(f_j\middle|\mathbf{x}\right)E^2\left(X_{i,j}\middle|\mathbf{x}\right)+E\left(\sigma_j^2\middle|\mathbf{x}\right)E\left(X_{i,j}\middle|\mathbf{x}\right)+E\left(f_j^2\middle|\mathbf{x}\right)\left[\mathrm{var}_{pro}\left(X_{i,j}\middle|\mathbf{x}\right)+\mathrm{var}_{par}\left(X_{i,j}\middle|\mathbf{x}\right)\right].\end{aligned} \tag{4.13}$$

If it is assumed that the process component follows the same recursive formula of process risk as in Frequentist approach, there is [2], [10]

$$\mathrm{var}_{pro}\left(X_{i,j+1}\middle|\mathbf{x}\right)=E^2\left(f_j\middle|\mathbf{x}\right)\mathrm{var}_{pro}\left(X_{i,j}\middle|\mathbf{x}\right)+E\left(\sigma_j^2\middle|\mathbf{x}\right)E\left(X_{i,j}\middle|\mathbf{x}\right). \tag{4.14}$$

Substituting (4.14) into (4.13) gives the recursive formula for the parameter component

$$\begin{aligned}&\mathrm{var}_{par}\left(X_{i,j+1}\middle|\mathbf{x}\right)\\&=\mathrm{var}\left(f_j\middle|\mathbf{x}\right)E^2\left(X_{i,j}\middle|\mathbf{x}\right)+\mathrm{var}\left(f_j\middle|\mathbf{x}\right)\mathrm{var}_{pro}\left(X_{i,j}\middle|\mathbf{x}\right)+E\left(f_j^2\middle|\mathbf{x}\right)\mathrm{var}_{par}\left(X_{i,j}\middle|\mathbf{x}\right).\end{aligned} \tag{4.15}$$

The equivalent recursive formula for Mack's formula is [10]

$$\mathrm{var}_{par}\left(X_{i,j+1}\middle|\mathbf{x}\right)=\mathrm{var}\left(f_j\middle|\mathbf{x}\right)E^2\left(X_{i,j}\middle|\mathbf{x}\right)+E^2\left(f_j\middle|\mathbf{x}\right)\mathrm{var}_{par}\left(X_{i,j}\middle|\mathbf{x}\right), \tag{4.16}$$

which doesn't have the term $\mathrm{var}\left(f_j\middle|\mathbf{x}\right)\mathrm{var}\left(X_{i,j}\middle|\mathbf{x}\right)$ compared with (4.15). Murphy's formula [2], which is the recursive formula underlying BBMW's formula [5], is

$$\mathrm{var}_{par}\left(X_{i,j+1}\middle|\mathbf{x}\right)=\mathrm{var}\left(f_j\middle|\mathbf{x}\right)E^2\left(X_{i,j}\middle|\mathbf{x}\right)+E\left(f_j^2\middle|\mathbf{x}\right)\mathrm{var}_{par}\left(X_{i,j}\middle|\mathbf{x}\right), \tag{4.17}$$

which doesn't have $\mathrm{var}\left(f_j \middle| \mathbf{x}\right)\mathrm{var}_{pro}\left(X_{i,j} \middle| \mathbf{x}\right)$ compared with (4.15). So the parameter component of Bayesian approach is always larger than parameter error of Frequentist approach. However, because this separation of process component and parameter component is artificial for Bayesian approach, the only conclusion that can be made is that the total MSE of Bayesian approach is higher than that of Frequentist approach. However, in case of $\mathrm{var}\left(f_j \middle| \mathbf{x}\right) \ll E^2\left(f_j \middle| \mathbf{x}\right)$ or $\mathrm{var}\left(X_{i,j} \middle| \mathbf{x}\right) \ll E^2\left(X_{i,j} \middle| \mathbf{x}\right)$, the difference is very small.

To calculate the variance of aggregate claim amount in (2.5), a new sequence of random variables Z_j are introduced to express the aggregate ultimate claim amount in another way. Z_j is defined as

$$Z_j = x_{N-j+1,j} + \sum_{i=N-j+2}^{N} X_{i,j} \tag{4.18}$$

It is apparent that Z_N is the aggregate ultimate claim amount. Based on (3.2), it is shown in Appendix A that

$$Z_{j+1} \middle| \left(X_{N-j+2,j}, X_{N-j+3,j}, \ldots X_{N,j}\right) \sim N\left(f_j Z_j + x_{N-j,j+1}, \sigma_j^2 Z_j\right). \tag{4.19}$$

Then the total risk can be calculated in the same way as individual year claims amount. Because $Z_1 = x_{N,1}$, which is fixed, the mean and variance of Z_1 is

$$E\left(Z_1 \middle| \mathbf{x}\right) = x_{N,1}$$

and

$$\mathrm{var}\left(Z_1 \middle| \mathbf{x}\right) = 0\ .$$

The recursive formula about mean of Z_{j+1} is

$$E\left(Z_{j+1} \middle| \mathbf{x}\right) = E\left(f_j Z_j + x_{N-j,j+1} \middle| \mathbf{x}\right) = E\left(f_j \middle| \mathbf{x}\right)E\left(Z_j \middle| \mathbf{x}\right) + x_{N-j,j+1} \tag{4.20}$$

and about variance is

$$\operatorname{var}\left(Z_{j+1}\middle|\mathbf{x}\right)=E\left(Z_{j+1}^2\middle|\mathbf{x}\right)-E^2\left(Z_{j+1}\middle|\mathbf{x}\right)$$

$$=E\left(f_j^2\middle|\mathbf{x}\right)E\left(Z_j^2\middle|\mathbf{x}\right)+E\left(\sigma_j^2\middle|\mathbf{x}\right)E\left(Z_j\middle|\mathbf{x}\right)-E^2\left(f_j\middle|\mathbf{x}\right)E^2\left(Z_j\middle|\mathbf{x}\right)$$

$$=\operatorname{var}\left(f_j\middle|\mathbf{x}\right)E^2\left(Z_j\middle|\mathbf{x}\right)+E\left(f_j^2\middle|\mathbf{x}\right)\operatorname{var}\left(Z_j\middle|\mathbf{x}\right)+E\left(\sigma_j^2\middle|\mathbf{x}\right)E\left(Z_j\middle|\mathbf{x}\right), \tag{4.21}$$

which is exactly same as the recursive formula for individual year.

5. PARAMETER ESTIMATION

As shown in last section, each pair of parameter $\left(f_j,\sigma_j^2\right)$ can be calculated individually and the posterior distribution in (4.5) takes a similar formation for different j. To make denotation in further analysis more concise, the analysis in this section focus on this term

$$p\left(f,\sigma^2\middle|\mathbf{x}\right)\propto\left(\sigma^2\right)^{-K/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{K}x_i\left(y_i-f\right)^2\right]\cdot p\left(f,\sigma^2\right). \tag{5.1}$$

In this paper, f is always assumed unknown, while σ^2 could be known or unknown.

5.1 Known σ^2

This was already done in [8]. But for completeness of the whole picture and make comparison, this is briefly analyzed in this section. In the case that σ^2 is known, (5.1) is simplified to

$$p\left(f\middle|\mathbf{x}\right)\propto\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{K}x_i\left(y_i-f\right)^2\right]\cdot p\left(f\right) \tag{5.2}$$

One typical non-informative prior distribution is

$$p\left(f\right)=1. \tag{5.3}$$

Note that actually (5.3) is not a distribution as the integration over f is infinite, which gains its name 'improper' prior distribution. By substituting (5.3) into (5.2), there is

$$p\left(f\,|\,\mathbf{x}\right) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{K} x_i\left(y_i - f\right)^2\right]$$

$$\propto \exp\left[-\frac{1}{2\sigma^2}\left(f - \hat{f}\right)^2\sum_{i=1}^{K} x_i\right],$$

where

$$\hat{f} = \sum_{i=1}^{K} x_i y_i \Big/ \sum_{i=1}^{K} x_i \,. \tag{5.4}$$

So the posterior distribution of f is Normal distribution

$$f\,|\,\mathbf{x} \sim N\left(\hat{f}, \sigma^2 \Big/ \sum_{i=1}^{K} x_i\right), \tag{5.5}$$

and the mean is

$$E\left(f\,|\,\mathbf{x}\right) = \hat{f} \tag{5.6}$$

and the variance is

$$\operatorname{var}\left(f\,|\,\mathbf{x}\right) = \sigma^2 \Big/ \sum_{i=1}^{K} x_i \tag{5.7}$$

If there is prior knowledge of f, it is useful to use informative prior distribution. One common option is the Normal distribution, i.e.

$$p\left(f\right) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{\left(f - \mu_0\right)^2}{2\sigma_0^2}\right] \tag{5.8}$$

where μ_0 is the prior knowledge of f and σ_0^2 indicates confidence about the prior knowledge; larger means lower confidence. By this prior, the posterior distribution in (5.2) becomes

$$p\left(f\middle|\mathbf{x}\right) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{K} x_i\left(y_i - f\right)^2 - \frac{1}{2\sigma_0^2}\left(f-\mu_0\right)^2\right]$$

$$\propto \exp\left[-\left(\frac{1}{2\sigma^2}\sum_{i=1}^{K} x_i + \frac{1}{2\sigma_0^2}\right)\left(f - \frac{\frac{\hat{f}}{\sigma^2}\sum_{i=1}^{K} x_i + \frac{\mu_0}{\sigma_0^2}}{\frac{1}{\sigma^2}\sum_{i=1}^{K} x_i + \frac{1}{\sigma_0^2}}\right)^2\right]$$

which shows that posterior distribution is Normal distribution

$$f\middle|\mathbf{x} \sim N\left(\frac{\frac{\hat{f}}{\sigma^2}\sum_{i=1}^{K} x_i + \frac{\mu_0}{\sigma_0^2}}{\frac{1}{\sigma^2}\sum_{i=1}^{K} x_i + \frac{1}{\sigma_0^2}}, \frac{1}{\frac{1}{\sigma^2}\sum_{i=1}^{K} x_i + \frac{1}{\sigma_0^2}}\right).$$

5.2 Unknown σ^2

When the parameter σ^2 is unknown, there are usually three types of prior distributions depending on different philosophy view of prior distribution.

5.2.1 Classical Bayesian Prior

In Classical Bayesian approach, it always tries to use a prior distribution as simple as possible and provide less information as possible. One option would be

$$p\left(f,\sigma^2\right) \propto 1/\sigma^2 . \tag{5.9}$$

which is an improper prior distribution. Substitute this prior distribution into (5.1), the joint posterior distribution becomes

$$p\left(f,\sigma^2\middle|x\right) \propto \left(\sigma^2\right)^{-(K+2)/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{K} x_i\left(y_i - f\right)^2\right] \tag{5.10}$$

As shown in Appendix B, the marginal posterior distribution f is

$$p\left(f\middle|\mathbf{x}\right)\propto\left\{1+\frac{\left(f-\hat{f}\right)^{2}\sum_{i=1}^{K}x_{i}}{\left(K-1\right)s^{2}}\right\}^{-K/2},\tag{5.11}$$

where $\hat{f}$ is defined in (5.4) and

$$s^{2}=\frac{1}{K-1}\sum_{i=1}^{K}x_{i}\left(y_{i}-\hat{f}\right)^{2},\tag{5.12}$$

which is the MLE of variance σ^2 that widely used in [1]-[8] for Frequentist and semi-Bayesian approach. The distribution shown in (5.11) is the standard t-distribution [10] with shift and scale, that is, $\left(f-\hat{f}\right)\Big/\sqrt{s^{2}\Big/\sum_{i=1}^{K}x_{i}}$ has standard t-distribution with degree of freedom $K-1$. So the posterior distribution of f is the t-distribution

$$f\middle|\mathbf{x}\sim t_{K-1}\left(\hat{f},s^{2}\Big/\sum_{i=1}^{K}x_{i}\right).\tag{5.13}$$

By feature of t-distribution, the mean of f is

$$E\left(f\middle|\mathbf{x}\right)=\hat{f}\tag{5.14}$$

and the variance is

$$\operatorname{var}\left(f\middle|\mathbf{x}\right)=\left(\frac{K-1}{K-3}s^{2}\right)\Big/\sum_{i=1}^{K}x_{i}.\tag{5.15}$$

So $\operatorname{var}\left(f\middle|\mathbf{x}\right)$ is not defined for $K\le 3$.

Similarly, Appendix C shows the marginal distribution of σ^2 is

$$p\left(\sigma^{2}\middle|\mathbf{x}\right)\propto\left(\sigma^{2}\right)^{-(K+1)/2}\exp\left[-\frac{\left(K-1\right)s^{2}}{2\sigma^{2}}\right]\tag{5.16}$$

which indicates that σ^2 has inverse Gamma distribution with parameter $\left(K-1\right)/2$ and

$(K-1)s^2/2$, that is,

$$\sigma^2 \big| \mathbf{x} \sim IG\left((K-1)/2, (K-1)s^2/2\right).$$

So the mean of σ^2 is

$$E\left[\sigma^2 \big| \mathbf{x}\right] = \frac{(K-1)s^2}{2\left[(K-1)/2-1\right]} = \frac{(K-1)}{(K-3)} s^2. \tag{5.17}$$

Similar to $\operatorname{var}\left(f \big| \mathbf{x}\right)$, this is not defined for $K \le 3$.

5.2.2 Modern Bayesian Prior

In modern Bayesian approach, the philosophy is to choose a prior distribution that give convenience for calculation. Typically the conjugate distribution will be used, that is the distribution which makes prior and posterior belong to same distribution family. For the likelihood formation as in (5.1), the conjugate distribution is the Normal-Inverse-Gamma distribution, which is defined as

$$p\left(f, \sigma^2\right) = p\left(f \big| \sigma^2\right) p\left(\sigma^2\right) \tag{5.18}$$

where σ^2 has inverse Gamma distribution

$$\sigma^2 \sim IG\left(\varpi_0/2, \sigma_0^2/2\right)$$

and f has Normal distribution with variance related to σ^2

$$f \big| \sigma^2 \sim N\left(\mu_0, \sigma^2/\eta_0\right).$$

ϖ_0, σ_0^2, μ_0 and η_0 are all parameters that can be chosen based on prior knowledge. In this prior distribution, f are no longer independent of σ^2.

By these prior distributions, the posterior distribution (5.1) becomes

$$p\left(f,\sigma^2\middle|\mathbf{x}\right)\propto\left(\sigma^2\right)^{-K/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{K}x_i\left(f-y_i\right)^2\right]\cdot\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\eta_0\left(f-\mu_0\right)^2}{2\sigma^2}\right)$$

$$\cdot\left(\sigma^2\right)^{-(\varpi_0/2+1)}\exp\left(-\frac{\sigma_0^2}{2\sigma^2}\right)$$

$$\propto\left(\sigma^2\right)^{-(K+\varpi_0+3)/2}\exp\left\{-\frac{1}{2\sigma^2}\left[\eta_0\left(f-\mu_0\right)^2+\sum_{i=1}^{K}x_i\left(f-y_i\right)^2+\sigma_0^2\right]\right\}. \tag{5.19}$$

As shown in Appendix D, the marginal distribution of f is

$$p\left(f\middle|\mathbf{x}\right)\propto\left[1+\frac{\left(f-\mu_K\right)^2\Big/\left(\sigma_K^2/\varpi_K\eta_K\right)}{\varpi_K}\right]^{-(\varpi_K+1)/2} \tag{5.20}$$

where

$$\mu_K=\frac{\eta_0}{\eta_0+\sum_{i=1}^{K}x_i}\mu_0+\frac{\sum_{i=1}^{K}x_i}{\eta_0+\sum_{i=1}^{K}x_i}\hat{f}\,, \tag{5.21}$$

$$\eta_K=\eta_0+\sum_{i=1}^{K}x_i\,, \tag{5.22}$$

$$\varpi_K=\varpi_0+K\,, \tag{5.23}$$

and

$$\sigma_K^2=\sigma_0^2+\left(K-1\right)s^2+\frac{\eta_0}{\eta_K}\left(\hat{f}-\mu_0\right)^2\sum_{i=1}^{K}x_i\,. \tag{5.24}$$

So $\left(f-\mu_K\right)\Big/\sqrt{\sigma_K^2/\varpi_K\eta_K}$ has standard t-distribution with degree of freedom ϖ_K, that is

$$f\middle|\mathbf{x}\sim t_{\varpi_K}\left(\mu_K,\sigma_K^2/\varpi_K\eta_K\right),$$

which giving the mean

$$E\left(f\middle|\mathbf{x}\right)=\mu_K \tag{5.25}$$

and the variance

$$\mathrm{var}\left(f \middle| \mathbf{x} \right) = \frac{\sigma_K^2}{\varpi_K \eta_K} \cdot \frac{\varpi_K}{\varpi_K - 2} = \frac{\sigma_K^2}{\left(\varpi_K - 2 \right) \eta_K}. \tag{5.26}$$

Similarly, the marginal posterior distribution of σ^2 is

$$p\left(\sigma^2 \middle| \mathbf{x} \right) \propto \left(\sigma^2 \right)^{-(\varpi_K - 2)/2} \exp\left(-\frac{\sigma_K^2}{2\sigma^2} \right) \tag{5.27}$$

which is proved in Appendix E. (5.27) shows that σ^2 has inverse Gamma distribution with parameter $\varpi_K/2$ and $\sigma_K^2/2$, i.e.

$$\sigma^2 \middle| \mathbf{x} \sim IG\left(\varpi_K/2, \sigma_K^2/2 \right).$$

So the mean is

$$E\left[\sigma^2 \middle| \mathbf{x} \right] = \frac{\sigma_K^2}{\varpi_K - 2}. \tag{5.28}$$

5.2.3 Subjective prior

The third option is to use any distribution that is 'subjectively' chosen based on prior knowledge. One commonly used prior distribution is that f and σ^2 are independent while f has normal distribution and σ^2 has inverse Gamma distribution, that is,

$$p\left(f, \sigma^2 \right) = p\left(f \right) p\left(\sigma^2 \right) \tag{5.29}$$

where

$$f \sim N\left(\mu_0, \varepsilon_0^2 \right)$$

and

$$\sigma^2 \sim IG\left(\varpi_0/2, \sigma_0^2/2 \right).$$

This prior is quite similar to conjugate prior distribution in (5.18) but f and σ^2 are independent. Substitute this into (5.1), the joint posterior distribution is

$$p\left(f,\sigma^2\middle|\mathbf{x}\right)\propto\left(\sigma^2\right)^{-K/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{K}x_i\left(f-y_i\right)^2\right]\cdot\frac{1}{\sqrt{2\pi\varepsilon_0^2}}\exp\left(-\frac{\left(f-\mu_0\right)^2}{2\varepsilon_0^2}\right)$$

$$\cdot\left(\sigma^2\right)^{-(\varpi_0/2+1)}\exp\left(-\frac{\sigma_0^2}{2\sigma^2}\right). \tag{5.30}$$

So the marginal posterior distribution is

$$p\left(f\middle|\mathbf{x}\right)\propto\exp\left[-\frac{\left(f-\mu_0\right)^2}{2\varepsilon_0^2}\right]\int_0^{+\infty}\left(\sigma^2\right)^{-(K+\varpi_0+2)/2}\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{K}x_i\left(f-y_i\right)^2+\sigma_0^2\right]\right\}d\sigma^2$$

$$=\exp\left[-\frac{\left(f-\mu_0\right)^2}{2\varepsilon_0^2}\right]\frac{\Gamma\left[\left(K+\varpi_0\right)/2\right]}{\left\{\frac{1}{2}\left[\sum_{i=1}^{K}x_i\left(f-y_i\right)^2+\sigma_0^2\right]\right\}^{(K+\varpi_0)/2}}$$

$$\propto\exp\left[-\frac{\left(f-\mu_0\right)^2}{2\varepsilon_0^2}\right]\left[\sum_{i=1}^{K}x_i\left(f-y_i\right)^2+\sigma_0^2\right]^{-(K+\varpi_0)/2},$$

which doesn't follow any standard distribution. And the marginal distribution of σ^2 could be calculated in a similar way, which doesn't follow any standard distribution either. However, the mean and variance of f and σ^2 can be calculated by numerical technique based on marginal posterior distribution. This approach isn't developed further in this paper.

6. NUMERICAL EXAMPLE AND RESULTS

The data from Taylor and Ashe [12], which is in Table 1, is used to illustrate the analytical results from previous sections. Four prior distributions are used; they are:

Prior 1: (5.3) with known variance σ^2 equaling s^2 defined in (5.12) and last σ^2 use

Mack's suggestion [1];

Prior 2: (5.9);

Prior 3: (5.18) with parameters $\mu_0 = 0$, $\eta_0 = 0.001$, $\varpi_0 = 0.001$ and $\sigma_0 = 0.001$.

Prior 4: (5.18) with parameter $\mu_0 = 0$, $\eta_0 = 0.001$, $\varpi_0 = 1.001$ and $\sigma_0 = 0.001$;

Prior 1 is the prior used by [3], [4], [8] and served as benchmark in this example. Prior 2-4 are the priors where σ^2 is unknown. Prior 2 give least information about f_j and σ_j^2, which is often called non-informative. Prior 3 is almost non-informative for σ_j^2, but do give more information for f_j compared with Prior 2 beccause the variance of it could be very small when σ_j^2 is small. Prior 4 has same implication for f_j as Prior 3, but it gives more information about σ_j^2.

Table 1. Accumulative claims amount triangle.

i	$j=1$	2	3	4	5	6	7	8	9	10
1	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463
2	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	
3	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315		
4	310608	1418858	2195047	3757447	4029929	4381982	4588268			
5	443160	1136350	2128333	2897821	3402672	3873311				
6	396132	1333217	2180715	2985752	3691712					
7	440832	1288463	2419861	3483130						
8	359480	1421128	2864498							
9	376686	1363294								
10	344014									

First, the mean and variance of parameter f_j and σ_j^2 are calculated. Equation (5.4), (5.14), (5.25) are used to calculate the mean of f_j, which are listed in Table 2. As expected, the mean is very similar between different prior distributions.

Table 2. Results of $E\left(f_j \middle| \mathbf{x}\right)$

j	Prior 1	Prior 2	Prior 3	Prior 4
1	3.4906065	3.4906065	3.4906055	3.4906055
2	1.7473326	1.7473326	1.7473325	1.7473325
3	1.4574128	1.4574128	1.4574127	1.4574127
4	1.1738517	1.1738517	1.1738516	1.1738516
5	1.1038235	1.1038235	1.1038235	1.1038235
6	1.0862694	1.0862694	1.0862693	1.0862693
7	1.0538744	1.0538744	1.0538743	1.0538743
8	1.0765552	1.0765552	1.0765551	1.0765551
9	1.0177247	1.0177247	1.0177245	1.0177245

The variance of f_j is calculated by (5.7), (5.15) and (5.26), which are presented in Table 3. In the case that $\mathrm{var}\left(f_j \middle| \mathbf{x}\right)$ is not defined, it is estimated by multiplying result of Prior 1 with a constant factor. This factor is chosen as the ratio of estimator of this Prior to estimator of Prior 1 at the nearest year where $\mathrm{var}\left(f_j \middle| \mathbf{x}\right)$ is defined. So for Prior 2, the factor is the ratio at year 6, which is 3. For Prior 3, it is the ratio at year 7, which is 2. Table 3 shows that the difference between different prior distributions is quite large, while Prior 4 gives very similar results to Prior 1.

And the mean of σ^2 is calculated by (5.17) and (5.28). For Prior 1, it is an fixed value calculated by (5.12). For Prior 2 and 3, if it is not defined then same approach is used as for $\mathrm{var}\left(f_j \middle| \mathbf{x}\right)$. All results are shown in Table 4, which indicates the difference between prior distributions is also quite large.

Table 3. Results of $\mathrm{var}\left(f_j \middle| \mathbf{x}\right)$.

i	Prior 1	Prior 2	Prior 3	Prior 4
1	0.04817026	0.06422701	0.05504437	0.04816468
2	0.00368120	0.00515367	0.00429406	0.00368071
3	0.00278879	0.00418318	0.00334590	0.00278834
4	0.00082302	0.00137170	0.00102854	0.00082287
5	0.00076441	0.00152882	0.00101890	0.00076424
6	0.00051306	0.00153917	0.00076923	0.00051291
7	0.00003505	0.00010514	0.00007011	0.00003507
8	0.00013466	0.00040399	0.00026932	0.00013466
9	0.00011650	0.00034951	0.00023301	0.00027045

Table 4. Results of $E\left(\sigma_j^2 \middle| \mathbf{x}\right)$.

i	Prior 1	Prior 2	Prior 3	Prior 4
1	160280.327	213707.103	183153.093	160261.818
2	37736.855	52831.597	44019.503	37731.901
3	41965.213	62947.820	50348.611	41958.574
4	15182.903	25304.838	18974.230	15180.142
5	13731.324	27462.648	18302.737	13728.197
6	8185.772	24557.315	12273.111	8183.437
7	446.617	1339.850	893.451	446.949
8	1147.366	3442.098	2294.732	1147.379
9	446.617	1339.850	893.233	1036.763

Then the MSE can be calculated. First, three different recursive formulas by Mack (4.16), BBMW/Murphy (4.17) and Bayesian approach (4.9) are compared under Prior 1, which are presented in Table 5. The results are exactly matched to results in [1], [3], [4] and [10], which shows that Bayesian approach by Prior 1 is same as Frequentist approach [8]. The results also show that the difference between three approaches is quite small, to the maximum of 0.01% of reserving amount.

Table 5. MSE by Frequentist and Bayesian approach under Prior 1.

Year	Mack	Murphy/BBMW	Bayesian
2	75,535	75,535	75,535
3	121,699	121,700	121,703
4	133,549	133,551	133,556
5	261,406	261,412	261,436
6	411,010	411,028	411,111
7	558,317	558,356	558,544
8	875,328	875,430	875,921
9	971,258	971,385	972,234
10	1,363,155	1,363,385	1,365,456
Total	2,447,095	2,447,618	2,449,345
Total MSE in %	13.10%	13.10%	13.11%

Finally, the MSE under four different prior distributions are calculated in Table 6. The MSN under the non-informative prior distribution, i.e. Prior 2, is about 38% larger than that under Prior 1 or the MSE of the Frequentist approach, which shows that the MSE is greatly underestimated if the variance in assumed known or fixed.

The MSE is about 3% different between Prior 1 and Prior 4 although the parameters estimation in Table 2-4 is very similar between these two prior distributions. This indicated that MSE is quite sensitive to parameters in the tail.

Table 6. MSE of different prior distributions.

Year	Prior 1	Prior 2	Prior 3	Prior 4
2	75,535	130,831	106,823	115,086
3	121,703	210,810	172,120	149,104
4	133,556	231,348	188,890	158,383
5	261,436	452,921	332,284	273,259
6	411,111	641,245	495,957	419,342
7	558,544	816,905	655,425	565,685
8	875,921	1,184,204	995,294	882,037
9	972,234	1,259,424	1,085,789	976,334
10	1,365,456	1,664,613	1,488,920	1,367,860
Total	2,449,345	3,383,619	2,830,505	2,527,166
Total MSE in %	13.11%	18.11%	15.15%	13.53%

7. CONCLUSIONS

The general Bayesian approach to prediction error is first explained and compared with Frequentist approach. The key difference is that Bayesian approach gives posterior distributions of the unknown parameters, rather than point estimations as by Frequentist approach. Because the process uncertainty in Frequentist approach is calculated using point estimations, there is no corresponding term in Bayesian approach: all uncertainty is caught in the posterior distribution of ultimate claim amount based on the posterior distribution of parameters. Due to this different philosophy, it has been shown that the total prediction error of Bayesian approach is larger than that of Frequentist approach. However, this difference should be quite small in most real-life situation.

In parameter estimation, the Bayesian approach also takes a different approach. The analytical distributions of f and σ^2 are derived for different types of prior distribution under Mack's model. Under non-informative and conjugate prior distribution, the posterior distribution of development factor f has standard t-distribution and σ^2 has inverse Gamma distribution. It has also been shown that if the parameter σ^2 is considered as known and fixed, the Bayesian approach gives same result as Frequentist approach. This indicates that the widely used Frequentist approach underestimate the prediction error because it neglects the uncertainty of σ^2. This has a great impact on the parameters estimation in tail and increases the aggregate prediction error by 38% based on Taylor and Ashe's data.

This highlights the problem of parameter estimations in the tail of triangle. To have a reasonably reliable estimation for σ^2, there are at least 4 years' of experience if there is no prior knowledge. If data is not available, prior knowledge must be used to stabilize parameters estimation.

Appendix A. Prove of Equation (4.19)

It will be proved in recursive approach. By the assumptions of the model from (3.2), there is

$$X_{N-1,j+1}\left|\left(X_{N-1,1},...,X_{N-1,j}\right)\sim N\left(f_jX_{N-1,j},\sigma_j^2X_{N-1,j}\right)\right.$$

and

$$X_{N,j+1}\left|\left(X_{N,1},...,X_{N,j}\right)\sim N\left(f_jX_{N,j},\sigma_j^2X_{N,j}\right)\right..$$

So the distribution of $\left(X_{N-1,j+1}+X_{N,j+1}\right)\left|\left(X_{N-1,1},...,X_{N-1,j},X_{N,1},...,X_{N,j}\right)\right.$ is

$$p\left\{\left(X_{N-1,j+1}+X_{N,j+1}\right)=x\left|\left(X_{N-1,1},...,X_{N-1,j},X_{N,1},...,X_{N,j}\right)\right.\right\}$$

$$=\int_{-\infty}^{+\infty}p\left\{X_{N-1,j+1}=t\left|\left(X_{N-1,1},...,X_{N-1,j}\right)\right.\right\}\cdot p\left\{X_{N,j+1}=x-t\left|\left(X_{N,1},...,X_{N,j}\right)\right.\right\}dt$$

$$=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi\sigma_j^2X_{N-1,j}}}\exp\left[-\frac{\left(t-f_jX_{N-1,j}\right)^2}{2\sigma_j^2X_{N-1,j}}\right]\cdot\frac{1}{\sqrt{2\pi\sigma_j^2X_{N,j}}}\exp\left[-\frac{\left(x-t-f_jX_{N,j}\right)^2}{2\sigma_j^2X_{N,j}}\right]dt$$

$$= \frac{1}{\sqrt{2\pi\sigma_j^2\left(X_{N-1,j}+X_{N,j}\right)}}\exp\left[-\frac{\left(x-f_j\left(X_{N-1,j}+X_{N,j}\right)\right)^2}{2\sigma_j^2\left(X_{N-1,j}+X_{N,j}\right)}\right],$$

which shows that it is Normal distributed with mean $f_j\left(X_{N-1,j}+X_{N,j}\right)$ and variance $\sigma_j^2\left(X_{N-1,j}+X_{N,j}\right)$. Recursively, $X_{N-2,j+1}$, $X_{N-3,j+1}$, …, $X_{N-j+1,j+1}$ can be put into summation and the sum $\sum_{i=N-j+1}^{N} X_{i,j+1}$ is Normal distribution with mean $f_j \sum_{i=N-j+1}^{N} X_{i,j}$ and variance $\sigma_j^2 \sum_{i=N-j+1}^{N} X_{i,j}$. So by the definition of Z_{j+1} in (4.18), there is

$$Z_{j+1} = x_{N-j,j+1} + \sum_{i=N-j+1}^{N} X_{i,j+1} \sim N\left(f_j \sum_{i=N-j+1}^{N} X_{i,j} + x_{N-j,j+1}, \sigma_j^2 \sum_{i=N-j+1}^{N} X_{i,j}\right)$$

$$\sim N\left(f_j Z_j + x_{N-j,j+1}, \sigma_j^2 Z_j\right). \qquad \square$$

Appendix B. Prove of Equation (5.11)

By substituting (5.10) into (4.6), the posterior distribution of f is

$$p\left(f\,|\,\mathbf{x}\right) \propto \int_0^{+\infty}\left(\sigma^2\right)^{-(K+2)/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{K} x_i\left(y_i-f\right)^2\right]d\sigma^2$$

$$= \frac{\Gamma\left(K/2\right)}{\left[\frac{1}{2}\sum_{i=1}^{K} x_i\left(y_i-f\right)^2\right]^{K/2}}$$

$$\propto \left[\sum_{i=1}^{K} x_i\left(y_i-f\right)^2\right]^{-K/2}$$

$$= \left[\sum_{i=1}^{K} x_i\left(f-\hat{f}\right)^2 + \sum_{i=1}^{K} x_i\left(y_i-\hat{f}\right)^2\right]^{-K/2}$$

$$\propto \left\{ 1 + \frac{\left(f - \hat{f} \right)^2 \sum_{i=1}^{K} x_i}{(K-1) s^2} \right\}^{-K/2} . \qquad \square$$

Appendix C. Prove of Equation (5.16)

By substituting (5.10) into (4.7), the posterior distribution is

$$p\left(\sigma^2 \middle| \mathbf{x}\right) \propto \int_{-\infty}^{+\infty} \left(\sigma^2\right)^{-(K+2)/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{K} x_i \left(y_i - f \right)^2 \right] df$$

$$= \left(\sigma^2\right)^{-(K+2)/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{K} x_i \left(y_i - \hat{f} \right)^2 \right] \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{K} x_i \left(f - \hat{f} \right)^2 \right] df$$

$$= \left(\sigma^2\right)^{-(K+2)/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{K} x_i \left(y_i - \hat{f} \right)^2 \right] \sqrt{\sum_{i=1}^{K} x_i \Big/ 2\pi\sigma^2}$$

$$\propto \left(\sigma^2\right)^{-(K+1)/2} \exp\left[-\frac{(K-1) s^2}{2\sigma^2} \right]. \qquad \square$$

Appendix D. Prove of Equation (5.20)

Substituting (5.19) into (4.6), there is

$$p\left(f \middle| \mathbf{x} \right) \propto \int_{0}^{+\infty} \left(\sigma^2\right)^{-(K+\varpi_0+3)/2} \exp\left\{ -\frac{1}{2\sigma^2} \left[\eta_0 \left(f - \mu_0 \right)^2 + \sum_{i=1}^{K} x_i \left(f - y_i \right)^2 + \sigma_0^2 \right] \right\} d\sigma^2$$

$$= \frac{\Gamma\left[\left(K + \varpi_0 + 1 \right) / 2 \right]}{\left\{ \frac{1}{2} \left[\eta_0 \left(f - \mu_0 \right)^2 + \sum_{i=1}^{K} x_i \left(f - y_i \right)^2 + \sigma_0^2 \right] \right\}^{(K+\varpi_0+1)/2}}$$

$$\propto \left[\eta_0 \left(f - \mu_0 \right)^2 + \sum_{i=1}^{K} x_i \left(f - y_i \right)^2 + \sigma_0^2 \right]^{-(K+\varpi_0+1)/2}$$

$$= \left[\eta_0 \left(f - \mu_0 \right)^2 + \sum_{i=1}^{K} x_i \left(f - \hat{f} \right)^2 + \sum_{i=1}^{K} x_i \left(y_i - \hat{f} \right)^2 + \sigma_0^2 \right]^{-(K+\varpi_0+1)/2}$$

$$=\left[\left(\eta_0+\sum_{i=1}^{K}x_i\right)f^2-2\left(\eta_0\mu_0+\hat{f}\sum_{i=1}^{K}x_i\right)f+\left(\eta_0\mu_0^2+\hat{f}^2\sum_{i=1}^{K}x_i\right)+\sum_{i=1}^{K}x_i\left(y_i-\hat{f}\right)^2+\sigma_0^2\right]^{-(K+\varpi_0+1)/2}$$

$$=\left[\left(\eta_0+\sum_{i=1}^{K}x_i\right)\left[f-\frac{\left(\eta_0\mu_0+\hat{f}\sum_{i=1}^{K}x_i\right)}{\left(\eta_0+\sum_{i=1}^{K}x_i\right)}\right]^2+\frac{\eta_0\left(\hat{f}-\mu_0\right)^2\sum_{i=1}^{K}x_i}{\left(\eta_0+\sum_{i=1}^{K}x_i\right)}+\sum_{i=1}^{K}x_i\left(y_i-\hat{f}\right)^2+\sigma_0^2\right]^{-(K+\varpi_0+1)/2}$$

$$=\left[\eta_K\left(f-\mu_K\right)^2+\frac{\eta_0\left(\hat{f}-\mu_0\right)^2\sum_{i=1}^{K}x_i}{\eta_K}+\left(K-1\right)s^2+\sigma_0^2\right]^{-(\varpi_K+1)/2}$$

$$=\left[1+\frac{\left(f-\mu_K\right)^2\big/\left(\sigma_K^2/\varpi_K\eta_K\right)}{\varpi_K}\right]^{-(\varpi_K+1)/2} \qquad \square$$

where μ_K, η_K, ϖ_K and σ_K^2 are defined in (5.21)-(5.24).

Appendix E. Prove of Equation (5.27)

Substitute (5.19) into (4.7), the posterior distribution of σ^2 is

$$p\left(\sigma^2\middle|\mathbf{x}\right)\propto\int_{-\infty}^{+\infty}\left(\sigma^2\right)^{-(K+\varpi_0+1)/2+1}\exp\left\{-\frac{1}{2\sigma^2}\left[\eta_0\left(f-\mu_0\right)^2+\sum_{i=1}^{K}x_i\left(f-y_i\right)^2+\sigma_0^2\right]\right\}df$$

$$=\left(\sigma^2\right)^{-(\varpi_K-1)/2}\int_{-\infty}^{+\infty}\exp\left\{-\frac{1}{2\sigma^2}\left[\eta_0\left(f-\mu_0\right)^2+\sum_{i=1}^{K}x_i\left(f-\hat{f}\right)^2+\sum_{i=1}^{K}x_i\left(\hat{f}-y_i\right)^2+\sigma_0^2\right]\right\}df$$

$$=\left(\sigma^2\right)^{-(\varpi_K-1)/2}\int_{-\infty}^{+\infty}\exp\left\{-\frac{1}{2\sigma^2}\left[\eta_K\left(f-\mu_K\right)^2+\sigma_K^2\right]\right\}df$$

$$=\left(\sigma^2\right)^{-(\varpi_K-1)/2}\sqrt{\frac{2\pi\sigma^2}{\eta_K}}\exp\left(-\frac{\sigma_K^2}{2\sigma^2}\right)$$

$$\propto\left(\sigma^2\right)^{-(\varpi_K-2)/2}\exp\left(-\frac{\sigma_K^2}{2\sigma^2}\right). \qquad \square$$

REFERENCES

[1] Mack T., “Distribution-free calculation of the standard error of chain-ladder reserve estimates,” *ASTIN Bulletin*, **23(2)**, 213-225.

[2] Murphy D. M., “Unbiased loss development factors,” *PCAS* **81**, 154-222.

[3] England P. D. and Verrall R. J., “Stochastic claims reserving in general insurance,” *British Actuarial Journal*, **8**, 443-544.

[4] England P. D. and Verrall R. J., “Predictive distributions of outstanding liabilities in general insurance,” *Annual of Actuarial Study*, **1**, 221-270.

[5] Buchwalder M., Buhlmann H., Herz M., and Wuthrich M. V., “The mean square error of prediction in the chain ladder reserving method”, *ASTIN Bulletin*, **36(2)**, 521-542.

[6] Mack T., Quarg G., and Braun C., “The mean square error of prediction in the chain ladder reserving method – a comment,” *ASTIN Bulletin*, **36(2)**, 543-552.

[7] Venter G., “Discussion of mean square error of prediction in the chain-ladder reserving method,” *ASTIn Bulletin*, **36(2)**, 566-572.

[8] Gisler A., “The estimation error in the chain-ladder reserving method: a Bayesian approach,” *ASTIN Bulletin*, **36(2)**, 554-565.

[9] Scollnik D. P. M., “Bayesian reserving models inspired by chain ladder methods and implemented using WinBUGS,”

[10] Murphy D. M., “Chain ladder reserve risk estimators,” *CAS E-Forum Summer*, **2007**.

[11] Patel J. K., Kapadia C. H., and Owen D.B., “Handbook of statistical distributions,” *Marcel Dekker, Inc*, **1976**.

[12] Taylor G. and Ashe G., “Second moments of estimates of outstanding claims,” *Journal of Econometrics*, **23**, 37-61.