Closed-Form Distribution of Prediction Uncertainty in Chain Ladder Reserving by Bayesian Approach

Ji Yao, PhD, FIA, CERA

Abstract

Bayesian approach is applied to evaluate the prediction uncertainty in chain ladder reserving. First, the philosophy of the Bayesian approach to prediction uncertainty is introduced and compared with the Frequentist approach. All parameters in the model are then estimated using the Bayesian approach, with multiple types of prior distributions. A closed-from posterior distribution is derived under non-informative and conjugate prior distribution for key parameters in the model. Finally, the theory is illustrated by numerical examples. The paper demonstrates that it is possible to derive closed-form estimates for the prediction uncertainty in chain ladder reserving using the Bayesian approach and that, for certain prior distributions, the estimated uncertainty could be much higher than estimates of uncertainty produced under the Frequentist approach.

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

1. INTRODUCTION

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

The approach taken so far to derive prediction uncertainty is classified as the Frequentist approach, which believes that the truth is fixed and the estimator has a distribution. Typically there are two types of error that leads to prediction uncertainty: the process error and the 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 errors are calculated as the variance of the MLE parameters around their true values.

Paralleling the Frequentist approach, the Bayesian approach is another statistical approach. In the Bayesian approach, the true value of an unknown parameter can be thought of as being a random variable to which a prior probability distribution is assigned. The observed sample data is then synthesized with the prior probability distribution by a likelihood function to give the posterior probability distribution. Statistical measures, such as mean and variance, are derived from the posterior probability distribution.

The debate between the proponents of these two approaches (Frequentist and Bayesian) has lasted for nearly a century without a clear outcome [13]. However, in the context of prediction uncertainty for chain ladder claims reserving, there have been limited studies on the Bayesian approach, to author's knowledge. The 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 in Mack's model [1] is defined as unknown, it is assumed as known in [8] or estimated by MLE and used as a known parameter in [3] and [4]. These approaches are termed as a semi-Bayesian approach in this paper. In [9], although the σ^2 is included in the Bayesian analysis, the author mainly uses simulation techniques, such as the bootstrap method, to estimate the parameter.

The purpose of this paper is two-fold. The first intention is to apply the Bayesian approach in estimation of parameters as well as evaluation of prediction uncertainty. The Bayesian approach has a notorious reputation of making mathematics really difficult and almost always ends up with open-form solutions and simulation. However, it will be shown that, under certain prior assumptions, it is possible to have closed-form solutions.

The second intention of this paper is to provide more evidence into the debate of which formula gives the most accurate estimation of prediction uncertainty [6]-[8]. It might not be fair to compare results from the Frequentist and Bayesian approaches. However, the fact that the Bayesian approach can make assumptions more explicit might help to understand the difference between these approaches.

There are different models and assumptions about stochastic reserving, see for example [3] and [4]. This paper focuses on the Mack's model as one of the most widely used, but the general theory could be applied to other models.

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 the assumptions of the model. Section 4 calculates the prediction uncertainty under the assumptions consistent with the Mack's model. Section 5 estimates the parameters in the model using a Bayesian approach. Numerical examples are presented in Section 6 and finally conclusions are made in Section 7.

2. THE BAYESIAN APRROACH TO PREDICTION UNCERTAINTY

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 variables in the upper left-hand section of the rectangle of $X_{i,j}$ have been observed, as illustrated in (2.1). These variables 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 in the 1st year has fully developed and therefore $X_{i,N}\ (2 \le i \le N)$ are considered the ultimate claim amounts to be estimated.

$$\begin{matrix} x_{1,1} & x_{1,2} & \cdots & \cdots & x_{1,N} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,N-1} & \\ \vdots & & \iddots & & \\ \vdots & \iddots & & & \\ x_{N,1} & & & & \end{matrix} \tag{2.1}$$

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

$$\hat{f}_j \Big| \mathbf{x} = \sum_{i=1}^{N-j} x_{i,j+1} \Big/ \sum_{i=1}^{N-j} x_{i,j} \tag{2.2}$$

and the ultimate claim amount for the i^{th} $(i \geq 2)$ year, denoted as $\hat{x}_{i,N}$, is estimated as

$$\hat{x}_{i,N} \Big| \mathbf{x} = x_{i,N-i+1} \prod_{j=N-i+1}^{N-1} \hat{f}_j \, . \tag{2.3}$$

(2.2) is only one of the common choices to estimate the development factors. Other calculations, such as a straight average of the observed ratios, can be used to come up with development factors in (2.2). It is important to note that (2.2) and (2.3) are deterministic in nature, given the observation $\mathbf{x}$, at least from the Bayesian point of view.

Having estimated the ultimate claim amount using (2.2), it is important to know how accurate this estimator is and what the prediction uncertainty is. One measure commonly employed for this purpose is the mean square error (MSE). Although this measure is initially formulated in the Frequentist approach, it can be adjusted to the Bayesian approach and has been widely used to evaluate the prediction uncertainty in [3], [4], [8] and [9]. For each individual year, MSE is defined as

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

and for the 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 $\mathbf{x}$, (2.4) becomes

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

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

$$= \left(\hat{x}_{i,N}\middle|\mathbf{x} - E\left(X_{i,N}\middle|\mathbf{x}\right)\right)^2 + E\left[\left(X_{i,N} - E\left(X_{i,N}\right)\right)^2\middle|\mathbf{x}\right]. \tag{2.5}$$

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

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

which is the case for chain ladder reserving method under the assumptions of [1], MSE of the i^{th} year is further simplified as

$$MSE_i = E\left[\left(X_{i,N} - E\left(X_{i,N}\right)\right)^2\middle|\mathbf{x}\right] = \operatorname{var}\left(X_{i,N}\middle|\mathbf{x}\right), \tag{2.6}$$

If $\hat{x}_{i,N}$ is biased, (2.6) only gives a lower bound of MSE as the second term in (2.5) above represents an additional bias error necessary to calculate the total MSE [10]. Similarly, the minimum MSE for the aggregate ultimate claim amount is

$$MSE_i = \operatorname{var}\left(\sum_{i=2}^{N} X_{i,N}\middle|\mathbf{x}\right). \tag{2.7}$$

A comparison with the Frequentist approach is interesting at this stage. In the Frequentist approach, as explained in [1], MSE is split into two parts, that is

$$MSE_i = \operatorname{var}\left(X_{i,N}\middle|\mathbf{x}\right) + \left(E\left(X_{i,N}\middle|\mathbf{x}\right) - \hat{x}_{i,N}\right)^2. \tag{2.8}$$

Comparing (2.6) with (2.8) suggests that the Bayesian approach misses one term. However, this is not the case because of the different meaning of $\operatorname{var}\left(X_{i,N}\middle|\mathbf{x}\right)$. In the Frequentist approach, $\operatorname{var}\left(X_{i,N}\middle|\mathbf{x}\right)$ is actually the variance of $X_{i,N}$ conditional on the MLE of all parameters. So stringently it is better to express (2.8) in this way

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

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

3. MODEL ASSUMPTIONS

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

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

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

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

Mack's model is claimed to be distribution-free, that is, the results from Mack's model don'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 is often slightly changed to assume that $X_{i,j+1}$ is Normally distributed with mean $f_jX_{i,j}$ and variance $\sigma_j^2X_{i,j}$ [3], [8], that is

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

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

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

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.

The Normal distribution is not the only distribution possible. Moreover, the assumption of normality is not the best from a theoretical standpoint, as the Normal distribution could take negative values while cumulative claims amount usually cannot. However, in common parameterization of the distribution, the probability to take negative value is fairly low. This assumption also provides a mathematically tractable result and was widely used in [3], [4] and [8].

As the distribution of $X_{i,j}$ is defined by parameters $\left(f_j,\sigma_j^2\right)$, the posterior distribution of $\left(f_j,\sigma_j^2\right)$ will be first calculated so that the posterior distribution of $X_{i,j}$ can be evaluated. To simplify further denotation, these vectors are defined

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

and

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

4. CALCULATION OF PREDICTION ERROR

To calculate (2.6), the first step of the 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 the assumptions of model. Assuming independence in (3.1) and (3.2), this probability is

$$\begin{aligned}
p\left(\mathbf{x} \middle| \mathbf{f},\boldsymbol{\sigma}^2\right) &= \prod_{i=1}^{N} p\left(x_{i,1}, x_{i,2}, \ldots, x_{i,N-i+1} \middle| \mathbf{f},\boldsymbol{\sigma}^2\right) \\
&= \prod_{i=1}^{N}\left[p\left(x_{i,1} \middle| \mathbf{f},\boldsymbol{\sigma}^2\right) \prod_{j=2}^{N-i+1} p\left(x_{i,j} \middle| x_{i,1}, x_{i,2}, \ldots, x_{i,j-1}, \mathbf{f},\boldsymbol{\sigma}^2\right)\right] \\
&= \left[\prod_{i=1}^{N} p\left(x_{i,1} \middle| \mathbf{f},\boldsymbol{\sigma}^2\right)\right] \prod_{j=2}^{N}\left\{\prod_{i=1}^{N-j+1} p\left(x_{i,j} \middle| x_{i,j-1}, x_{i,j-2} \ldots, x_{i,1}, \mathbf{f},\boldsymbol{\sigma}^2\right)\right\} \\
&\propto \prod_{j=2}^{N}\left\{\prod_{i=1}^{N-j+1}\left\{\frac{1}{\sqrt{2\pi\left(\sigma_{j-1}^2 x_{i,j-1}\right)}} \exp\left[-\frac{\left(x_{i,j} - f_{j-1}x_{i,j-1}\right)^2}{2\left(\sigma_{j-1}^2 x_{i,j-1}\right)}\right]\right\}\right\} \\
&\propto \prod_{j=1}^{N-1}\left\{\prod_{i=1}^{N-j}\left\{\frac{1}{\sqrt{\sigma_j^2}} \exp\left[-\frac{\left(y_{i,j} - f_j\right)^2}{2\left(\sigma_j^2 / x_{i,j}\right)}\right]\right\}\right\}.
\end{aligned} \tag{4.2}$$

There are several options for the prior distribution $p\left(\mathbf{f},\boldsymbol{\sigma}^2\right)$, which will be discussed in detail in Section 5. By definition, $p\left(\mathbf{f},\boldsymbol{\sigma}^2\right)$ is a multi-dimensional distribution and generally there is no guarantee of independency between pairs $\left(f_j, \sigma_j^2\right)$. However, in the Bayesian theory, any appropriate distribution can be chosen as prior distribution, so it is reasonable to assume that the chosen prior distribution have the feature of independency, i.e., any pair $\left(f_j, \sigma_j^2\right)$ is independent to other pair, so that

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

Note that the non-informative prior distributions used in [3], [4] and [8] satisfy this assumption and all prior distributions used in Section 5 meet this criteria as well. Substituting (4.2) and (4.3) into (4.1), results in:

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

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

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

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

The second step of the 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

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

and similarly

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

In cases where the integration in (4.6) and (4.7) cannot be performed analytically, numerical techniques have to be used to calculate the posterior marginal distribution. This is where the Bayesian approach becomes tricky and has to resort to simulation techniques. However, as will be shown in section 5, these two integrations could give closed-form distribution under certain prior distributions, which gives interesting standard statistical distributions.

Having derived the marginal posterior distribution, the final step is to calculate the variance in (2.6). In this paper, this is done in a recursive way. Because any pair $\left(f_j,\sigma_j^2\right)\middle|\mathbf{x}$

is independent of another pair $\left(f_k,\sigma_k^2\right)\big|\mathbf{x}$ $\left(j\neq k\right)$, $\left(f_j,\sigma_j^2\right)\big|\mathbf{x}$ is also independent of $X_{i,k+1}\big|\mathbf{x}$ if $j\neq k$. Using this independence, the mean of $X_{i,j+1}\big|\mathbf{x}$ (for $j\geq i$) is

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

and the second central moment is

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

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

$$\begin{aligned}
&\mathrm{var}\left(X_{i,j+1}\big|\mathbf{x}\right)=E\left(X_{i,j+1}^2\big|\mathbf{x}\right)-E^2\left(X_{i,j+1}\big|\mathbf{x}\right)\\
&=E\left(f_j^2\big|\mathbf{x}\right)E\left(X_{i,j}^2\big|\mathbf{x}\right)+E\left(\sigma_j^2\big|\mathbf{x}\right)E\left(X_{i,j}\big|\mathbf{x}\right)-E^2\left(f_j\big|\mathbf{x}\right)E^2\left(X_{i,j}\big|\mathbf{x}\right)\\
&=\mathrm{var}\left(f_j\big|\mathbf{x}\right)E^2\left(X_{i,j}\big|\mathbf{x}\right)+E\left(f_j^2\big|\mathbf{x}\right)\mathrm{var}\left(X_{i,j}\big|\mathbf{x}\right)+E\left(\sigma_j^2\big|\mathbf{x}\right)E\left(X_{i,j}\big|\mathbf{x}\right).
\end{aligned} \tag{4.9}$$

The value of $E\left(f_j\big|\mathbf{x}\right)$, $\mathrm{var}\left(f_j\big|\mathbf{x}\right)$ and $\mathrm{var}\left(\sigma_j^2\big|\mathbf{x}\right)$ can be calculated from the posterior distribution in (4.6) and (4.7).

A boundary condition is needed to calculate (4.9) properly. For the first term $X_{i,N-i+1}$ in the recursive formula, because

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

its mean is

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

and its variance is

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

So by recursive formula (4.8), (4.9) and boundary condition (4.10), (4.11), MSE in (2.6) 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 the Frequentist and Bayesian approaches. In the Frequentist approach, the MLE $\hat{\sigma}_j^2$ is used while in the Bayesian approach the mean of $\sigma_j^2\big|\mathbf{x}$ is used. As will be shown in Section 5, this difference is very large when there are few data points available, such as at the tail of reserving triangle.

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

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

Substitute (4.12) into (4.9) and (4.9) becomes

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

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

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

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

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

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

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

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

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

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

To calculate the variance of the aggregate claim amount in (2.7), 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

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

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

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

and

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

respectively.
The recursive formula for 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 for variance is

$$\begin{aligned}
&\operatorname{var}\left(Z_{j+1} \middle| \mathbf{x} \right) = E\left(Z_{j+1}^2 \middle| \mathbf{x} \right) - E^2\left(Z_{j+1} \middle| \mathbf{x} \right) \\
&= E\left(f_j^2 \middle| \mathbf{x} \right) E\left(Z_j^2 \middle| \mathbf{x} \right) + E\left(\sigma_j^2 \middle| \mathbf{x} \right) E\left(Z_j \middle| \mathbf{x} \right) - E^2\left(f_j \middle| \mathbf{x} \right) E^2\left(Z_j \middle| \mathbf{x} \right) \\
&= \operatorname{var}\left(f_j \middle| \mathbf{x} \right) E^2\left(Z_j \middle| \mathbf{x} \right) + E\left(f_j^2 \middle| \mathbf{x} \right) \operatorname{var}\left(Z_j \middle| \mathbf{x} \right) + E\left(\sigma_j^2 \middle| \mathbf{x} \right) E\left(Z_j \middle| \mathbf{x} \right),
\end{aligned} \tag{4.21}$$

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

5. PARAMETER ESTIMATION

As shown in last section, the posterior distributions for each pair of parameters $\left(f_j, \sigma_j^2 \right)$ can be calculated individually and the posterior distributions in (4.5) have similar forms for different j's. To make the notation in further analysis more concise, the analysis in this section focuses on the term

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

with K replacing $N-j$ in (4.5). In this paper, f is always assumed unknown, while σ^2 could be known or unknown.

5.1 Known σ^2

For completeness and in order to make the comparison, this section includes a brief analysis of the case when σ^2 is known, even though that was already been done in [8]. With known σ^2, (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

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

By substituting (5.3) into (5.2), there is

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

where

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

So the posterior distribution of f is a Normal distribution

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

and the mean is

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

and the variance is

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

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

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

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

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

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

which shows that posterior distribution is Normal distribution

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

5.2 Unknown σ^2

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

5.2.1 Non-informative Prior

In a non-informative prior approach, the intention is to use a prior distribution as simple as possible, which provides the smallest amount of information. One option would be

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

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

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

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 is widely used in [1]-[8] for the 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 the standard t -distribution with $K-1$ degrees of freedom. 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 the 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 $K\leq 3$.

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

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

So the mean of σ^2 is

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

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

5.2.2 Conjugate Prior

In the conjugate prior approach, the philosophy is to choose a prior distribution that provides convenience of 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

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

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

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

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 is 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

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

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

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

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

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

which gives the mean

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

and the variance

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

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 Other priors

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 but is still a closed-form distribution. The marginal distribution of σ^2 could be calculated in a similar way, but it does not give a closed-form result. However, the mean and variance of f and σ^2 can be calculated by numerical technique based on marginal posterior distribution. This approach is not 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.

Table 1. Cumulative claims amount triangle.

i	$j=1$	2	3	4	5	6	7	8	9	10
1	357,848	1,124,788	1,735,330	2,218,270	2,745,596	3,319,994	3,466,336	3,606,286	3,833,515	3,901,463
2	352,118	1,236,139	2,170,033	3,353,322	3,799,067	4,120,063	4,647,867	4,914,039	5,339,085	
3	290,507	1,292,306	2,218,525	3,235,179	3,985,995	4,132,918	4,628,910	4,909,315		
4	310,608	1,418,858	2,195,047	3,757,447	4,029,929	4,381,982	4,588,268			
5	443,160	1,136,350	2,128,333	2,897,821	3,402,672	3,873,311				
6	396,132	1,333,217	2,180,715	2,985,752	3,691,712					
7	440,832	1,288,463	2,419,861	3,483,130						
8	359,480	1,421,128	2,864,498							
9	376,686	1,363,294								
10	344,014									

Four prior distributions are used; they are:

Prior 1: (5.3) with known variance σ^2 equaling s^2 defined in (5.12). For the last variance of σ_9^2, the formula does not work as there is only one observation of development factor, a common issue in the Frequentist approach as well. The σ_9^2 is estimated according to Mack's suggestion in [1] as

$$\sigma_9^2 = \min\left(\sigma_8^4/\sigma_7^2, \min\left(\sigma_7^2, \sigma_8^2\right)\right);$$

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], and [8] and served as benchmark in this example. Prior 2-4 are the priors where σ^2 is unknown. Prior 2 gives the 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 it does give more information for f_j compared with Prior 2 because the variance of f_j could be very small when σ_j^2 is small. Prior 4 has same implication for f_j as Prior 3, and it gives more information about σ_j^2.

It is important to note that the chosen parameters uniquely define the prior distribution. However, that does not necessarily guarantee that statistical measures of the distribution, such as the mean and variance, exist. For example, for a non-informative prior, it is common to have infinite mean or variance.

First, the mean and variance of parameters f_j and σ_j^2 are calculated. Equations (5.4), (5.14), and (5.25) are used to calculate the mean of f_j, shown in Table 2. As expected, the mean is very similar among the 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 using (5.7), (5.15) and (5.26), and is presented in Table 3. In the tail of the triangle, the formula might not work--a similar issue when estimating σ_9^2. (5.15) and (5.26) do not work when the number of observation is small, which does not mean that the variance does not exist but that there is not enough information to estimate it under a non-informative prior. In such case, the approach suggested in [8] is used: the variance is estimated by multiplying the result of Prior 1 with a constant factor.

The multiplicative factor is chosen, subjectively, as the ratio of estimator of this Prior to the estimator of Prior 1 at the nearest year where $\operatorname{var}\left(f_j \middle| \mathbf{x}\right)$ can be estimated. So for Prior 2, the factor is the ratio at year 6, which is 3. For Prior 3, it is the ratio at year 8, which is 2. $\operatorname{var}\left(f_j \middle| \mathbf{x}\right)$ calculated by these factor are highlighted in Italic in the Table 3. Table 3 shows that the differences in the variance between different prior distributions are quite large, while Prior 4 gives very similar results to Prior 1 except that last term of $\operatorname{var}\left(f_9 \middle| \mathbf{x}\right)$. This is because at the extreme tail of triangle, the observed information is not enough to estimate the variance and the estimation largely depends on the prior information. As stronger prior is assumed in Prior 1, so the variance is lower.

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

j	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

The mean of σ^2 is calculated using (5.17) and (5.28). For Prior 1, it is a fixed value given by (5.12). For Prior 2 and 3, if the formula does not work in the tail of triangle, the same approach - multiplying results for Prior 1 by a factor - as for $\operatorname{var}\left(f_j \middle| \mathbf{x}\right)$ is used. All results are shown in Table 4, which indicates the difference between prior distributions is also quite large.

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

j	Prior 1	Prior 2	Prior 3	Prior 4
1	160,280.327	213,707.103	183,153.093	160,261.818
2	37,736.855	52,831.597	44,019.503	37,731.901
3	41,965.213	62,947.820	50,348.611	41,958.574
4	15,182.903	25,304.838	18,974.230	15,180.142
5	13,731.324	27,462.648	18,302.737	13,728.197
6	8,185.772	24,557.315	1,2273.111	8,183.437
7	446.617	*1,339.850*	893.451	446.949
8	1,147.366	*3,442.098*	*2,294.732*	1,147.379
9	446.617	*1,339.850*	*893.233*	1,036.763

Then the MSE can be calculated. First, the recursive formulas by Mack (4.16) and BBMW/Murphy (4.17) are compared to the Bayesian approach (4.9) under Prior 1, with results presented in Table 5. The results are exactly matched to results in [1], [3], [4] and [10], which shows that the Bayesian approach under Prior 1 is very similar to the Frequentist approach with a difference of 0.01% in reserve amount.

Table 5. MSE by Frequentist and Bayesian approaches 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 MSE 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 is assumed known or fixed.

The MSE is about a 3% different between Prior 1 and Prior 4 although the parameters estimated in Table 2-4 are very similar between these two prior distributions. This indicates 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 evaluate prediction uncertainty is first explained and compared with the Frequentist approach. The key difference is that the Bayesian approach evaluates the posterior distributions of unknown parameters, rather than point estimates as in the Frequentist approach. Due to this different philosophy, it has been shown that the total prediction uncertainty of the Bayesian approach is different from that of the Frequentist approach, and under certain assumptions the Bayesian approach gives a higher estimate.

In parameter estimation, the Bayesian approach also takes a different approach. Closed-form distributions for f and σ^2 are derived for several prior distributions in Mack's model, which is one of the key results of this paper. It is shown that under non-informative and conjugate prior distribution, the posterior distribution of development factor f is the standard t-distribution while σ^2 has inverse Gamma distribution. For some other prior distributions, it is possible to derive a closed-form distribution which doesn't match any standard statistical distribution. It is also shown that if the parameter σ^2 is considered known and fixed, which is a very strong prior distribution assumption, the Bayesian approach gives the same result as the Frequentist approach. This indicates that the widely used Frequentist approach could underestimate the prediction uncertainty because it doesn't full reflect the uncertainty of σ^2.

The numerical results based on Taylor and Ashe data [12] are presented to confirm these conclusions. The Bayesian approach with strong prior distribution gives essentially the same results as Mack's and Murphy/BBMW's results. However, the prior distribution has a significant impact on the prediction uncertainty: a non-informative prior could increase aggregate prediction uncertainty by as much as 38%. Most of the difference comes from $\operatorname{var}\left(f_j \middle| \mathbf{x}\right)$ and $E\left(\sigma_j^2 \middle| \mathbf{x}\right)$. This highlights the problem of parameter estimations in chain ladder method: with no prior knowledge, the estimation of development factor could be very volatile.

Acknowledgment

The author would like to thank reviewers Vladimir Shander and Nancy Arico for their comments and suggestions, which greatly improve readability of the paper.

Appendix A. Proof of Equation (4.19)

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

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

and

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

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

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

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

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

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

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

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

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

Appendix B. Proof 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. Proof 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. Proof of Equation (5.20)

Substituting (5.19) into (4.6), there is

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

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

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

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

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

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

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

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

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

Appendix E. Proof 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.
[13] Bayarri M. J. and Berger J. O., "The Interplay of Bayesian and Frequentist Analysis," *Statistical Science* Vol. **19**, No. 1 (Feb., 2004), pp. 58-80.

Biography of the Author

Dr. Ji Yao is a manager of European Actuarial Services at Ernst & Young LLP. Since he graduated with a PhD in mathematics and statistics in 2005, he mainly practices in general insurance and specializes in statistical modeling and pricing. He qualified as a Fellow of Institute of Actuaries in 2008. He has also participated in various research projects, including reserving and Solvency II, and is a frequent speaker at international actuarial conferences.