

# THE ESTIMATION ERROR IN THE CHAIN-LADDER RESERVING METHOD: A BAYESIAN APPROACH

BY

ALOIS GISLER

## ABSTRACT

We derive the estimation error in a Bayesian framework and discuss the estimates of Mack [2] and of Buchwalder, Bühlmann, Merz and Wüthrich (BBMW) [1] from a Bayesian point of view.

## 1. INTRODUCTION

In [1] the authors BBMW revisit the chain ladder method on the basis of a time-series' model and suggest a new formula for the estimation error, which differs from the formula in the original paper [2] of Mack. The estimates resulting from the new formula are always greater than the ones resulting from Mack. In [3] Mack, Quarg and Braun make some comments on the approach taken in [1] and make a question mark, whether the new formula is an improvement. Hence there is some controversial discussion on how to estimate the estimation error.

In this contribution we first discuss shortly the different formulae of Mack and BBMW in section 2. The main part is section 3, where we tackle the question of the estimation error in a new way by means of a Bayesian approach. We will consider two Bayesian models with a limiting prior, where the Bayes estimators for the unknown chain ladder factors will coincide with the classical chain ladder estimates  $\hat{f}_j$ , such that the mean square error of this Bayes estimator is comparable with the estimation error of the classical chain ladder method.

By this different approach we hope to contribute to the discussion and to give some further insight into the problem.

Wherever possible and not otherwise stated we will use the same notation as in [1].

## 2. DISCUSSION ON THE ESTIMATES OF BBMW AND MACK

The mean square estimation error of  $X_{k,j}$ ,  $j > K - k + 1$ , is defined as

$$\begin{aligned}
 e_{k,j} &:= E\left[\left(\widehat{X}_{k,j} - E[X_{k,j} | \mathcal{D}_K]\right)^2 \middle| \mathcal{D}_K\right] \\
 &= X_{k, K-k+1}^2 \left( \prod_{l=K-k+1}^{j-1} \widehat{f}_l - \prod_{l=K-k+1}^{j-1} f_l \right)^2.
 \end{aligned}
 \tag{2.1}$$

In the original paper [2] Mack suggests to estimate  $e_{j,k}$  by

$$\widehat{e}_{k,j} = X_{k, K-k+1}^2 \left( \prod_{l=K-k+1}^{j-1} \widehat{f}_l^2 \right) \left( \prod_{l=K-k+1}^{j-1} \frac{\widehat{\sigma}_l^2}{\widehat{f}_l^2} \frac{1}{S_l} \right).
 \tag{2.2}$$

The reasoning behind this formula is well documented in [2].

First we should note that (2.2) takes also into account the information contained in the upper right corner  $\mathcal{D}_K^O$  and that it can therefore also be considered as a conditional approach.

BBMW suggest to estimate  $e_{k,j}$  by

$$\widehat{e}_{k,j} = X_{k, K-k+1}^2 \left( \prod_{l=K-k+1}^{j-1} E[\widehat{f}_l^2 | \mathcal{B}_l] - \prod_{l=K-k+1}^{j-1} \widehat{f}_l^2 \right)
 \tag{2.3}$$

$$= X_{k, K-k+1}^2 \left( \prod_{l=K-k+1}^{j-1} \left( \widehat{f}_l^2 + \frac{\widehat{\sigma}_l^2}{S_l} \right) - \prod_{l=K-k+1}^{j-1} \widehat{f}_l^2 \right),
 \tag{2.4}$$

where in (2.4)  $\sigma_l^2$  was replaced by  $\widehat{\sigma}_l^2$ . The reasoning behind (2.4) is explained in [1], but the arguments are not so obvious and not so easy to follow.

The main objection in [3] to the new approach of BBMW is that, by taking the product

$$\prod_{l=K-k+1}^{j-1} E[\widehat{f}_l^2 | \mathcal{D}_K]$$

in (2.3), this formula disregards the negative correlation between the r.v.  $\widehat{f}_l^2$  given  $\mathcal{B}_{K-k+1}$ . Hence on average over all upper right triangles  $\mathcal{D}_K^O$ , (2.3) has a systematic upward bias, i.e. for  $\widehat{e}_{k,j}$  as in (2.3) we have that  $E[\widehat{e}_{k,j} | \mathcal{B}_{K-k+1}] > E[e_{k,j} | \mathcal{B}_{K-k+1}]$  for all  $f_j$ .

The point is that we are interested in the conditional estimation error given the whole triangle  $\mathcal{D}_K$  including the upper right corner  $\mathcal{D}_K^O$ . This conditional estimation error given by (2.1) is a real number, which is unknown to us, because we do not know the true chain ladder factors  $f_j$ . As a consequence, given the whole triangle  $\mathcal{D}_K$  and hence the resulting estimates  $\widehat{f}_j$ , for some values of the underlying  $f_j$  (2.4) will be closer to the true estimation error (2.1) and for others (2.2). The question is, which of the two methods give on average a more accurate result.

But how to define this average? In [3] the authors consider the average over all possible upper right corners  $\mathcal{D}_K^O$  given  $\mathcal{B}_{K-k+1}$ . This is one possible point of view, but not the only one.

Another possibility is to vary the underlying true chain ladder factors  $f_j$  and to take the average over the set of all possible underlying  $f_j$  which might be behind the observed triangle  $\mathcal{D}_K$ . It seems to me that this point of view is even more suited to the conditional situation and hence to the estimation of the conditional estimation error, where the triangle  $\mathcal{D}_K$  is given and where the true factors  $f_j$  are unknown.

However, averaging over the set of possible  $f_j$  can only be done in a strict mathematical way within a Bayesian framework. This will be the subject of the next section.

### 3. DERIVATION OF THE ESTIMATION ERROR IN A BAYESIAN FRAMEWORK

In the Bayesian framework, we assume that the true chain ladder factors of a given triangle are a realisation

$$\mathbf{f}' = (f_1, \dots, f_{J-1})$$

of a random vector

$$\mathbf{F}' = (F_1, \dots, F_{J-1}),$$

and that conditionally, given  $\mathbf{F}$ , the chain ladder assumptions hold true. In a Bayesian model, we further have to specify the family of conditional distributions on the one hand and the a priori distribution on the other hand.

Conditionally given  $\mathbf{F}$  we make the same assumption as (M2) of [1], namely that conditionally  $X_{k,j}$  ( $j = 1, \dots, J$ ) form a Markov chain. To be strict, this is a slightly stronger assumption than the original chain ladder assumption in Mack [2]. In [2] it is assumed that the first and second order moments of  $X_{k,j+1}$  depend only on  $X_{k,j}$  and not on  $X_{k,l}$  for  $l < j$ , but it is not assumed that this property holds true for the whole distribution of  $X_{k,j+1}$ .

In the following two subsections we will consider two different Bayes models. In both models we will consider a limiting case of a prior where the Bayes estimator of the  $F_j$  as well as of  $X_{k,j}$  given the triangle  $\mathcal{D}_K$  coincide with the classical estimators  $\hat{f}_j$  resp.  $\hat{X}_{k,j}$ . Hence the mean square error of this Bayes estimator will be comparable with the estimation error of the classical chain ladder model.

Instead of the random variables  $X_{k,j+1}$  we will rather consider the random variables

$$Y_{k,j} = \frac{X_{k,j+1}}{X_{k,j}}$$

and will denote by

$\mathbf{Y}'_j = (Y_{1,j}, \dots, Y_{K-j,j})$  the random vector and by  
 $\mathbf{y}'_j = (y_{1,j}, \dots, y_{K-j,j})$  a realisation of  $\mathbf{Y}'_j$ .

We will also distinguish between the random variable

$$\widehat{F}_j = \frac{\sum_{k=1}^{K-j} X_{k,j} Y_{k,j}}{S_j}.$$

and the observed chain ladder factor  $\widehat{f}_j$  as a realisation of  $\widehat{F}_j$ .

**3.1. Bayesian model I: normal distribution**

In this model we assume that conditionally, given  $\mathbf{F}$ ,  $Y_{k,j}$  are normally distributed, i.e. we make the following assumptions:

(N1) Conditionally, given  $\mathbf{F}$  and  $\mathcal{B}_j$ , the random variables  $Y_{k,j}$  ( $k = 1, \dots, K-j$ ) are independent and normally distributed with moments

$$E[Y_{k,j} | F_j] = F_j, \quad \text{Var}[Y_{k,j} | F_j, X_{k,j}] = \frac{\sigma_j^2}{X_{k,j}}.$$

(N2)  $F_1, \dots, F_{j-1}$  are independent and uniformly distributed on  $(-a, a)$ .

Remarks:

- The time series' model considered in [1] with the additional assumption, that the  $\varepsilon_{k,j}$  are normally distributed, assumes that (conditionally given  $\mathbf{F} = \mathbf{f}$ ) the triangle  $\mathcal{D}_K$  is generated by

$$X_{k,j+1} = f_j X_{k,j} + \sigma_j \sqrt{X_{k,j}} \varepsilon_{k,j+1},$$

where

$$\varepsilon_{k,j} \text{ are independent and } \varepsilon_{k,j} \sim \mathcal{N}(0,1).$$

Assumption (N1) is slightly more general, because the r.v.  $\varepsilon_{k,j+1}$  ( $k = 1, \dots, K-j$ ) are assumed to be only conditionally independent given the observations  $X_{k,j}$  in column  $j$ .

- Assumption (N1) could be generalised by allowing the  $\sigma_j$  to depend on the observations  $X_{k,j}$  in column  $j$ .
- Because of the normality assumption, some of the  $X_{k,j}$  could become negative. A negative value of  $X_{k,j}$  contradicts the chain ladder assumption, since the conditional variance of  $X_{k,j+1}$  would then become negative. Therefore,

from a theoretical point of view, this model is a bit questionable as it is the time-series' model in [1]. We consider it here to get a direct Bayesian counterpart to the time series' model of [1] with normally distributed  $\varepsilon_{k,j}$ . In the next subsection we will consider a model which has no longer this theoretical deficiency. However, as we will see, it will have other drawbacks.

We now derive the Bayes estimator and its mean square error. On the condition  $\mathbf{F} = \mathbf{f}$  and given  $\mathcal{B}_1$ , the joint density of  $\mathbf{Y}_1, \dots, \mathbf{Y}_{J-1}$  is given by

$$u_{\mathbf{f}}(\mathbf{y}_1, \dots, \mathbf{y}_{J-1}) = \prod_{j=1}^{J-1} \prod_{k=1}^{K-j} \left\{ \frac{1}{\sqrt{2\pi\sigma_j^2 x_{k,j}^{-1}}} e^{-\frac{1}{2} \frac{(y_{k,j} - f_j)^2}{\sigma_j^2 x_{k,j}^{-1}}} \right\} \tag{3.1}$$

where  $x_{k,j} = X_{k,1} \cdot y_{k,1} \cdot \dots \cdot y_{k,j-1}$ .

From (3.1) we obtain for the joint posterior distribution of  $\mathbf{F}$  given the triangle  $\mathcal{D}_K$

$$u_{\mathcal{D}_K}(\mathbf{f}) \propto \prod_{j=1}^{J-1} e^{-\frac{1}{2} \sum_{k=1}^{K-j} \frac{(Y_{k,j} - f_j)^2}{\sigma_j^2 x_{k,j}^{-1}}} \text{ for } f_j \in (-a, a). \tag{3.2}$$

The exponent can be written as

$$\begin{aligned} \sum_{k=1}^{j-1} \frac{(Y_{k,j} - f_j)^2}{\sigma_j^2 X_{k,j}^{-1}} &= \frac{S_j}{\sigma_j^2} \sum_{k=1}^{K-j} \frac{X_{k,j}}{S_j} (Y_{k,j} - f_j)^2 \\ &= \frac{S_j}{\sigma_j^2} \sum_{k=1}^{K-j} \frac{X_{k,j}}{S_j} (Y_{k,j}^2 - 2 Y_{k,j} f_j + f_j^2) \\ &= \frac{S_j}{\sigma_j^2} \left( (f_j - \hat{f}_j)^2 + h(\mathcal{D}_K) \right), \end{aligned} \tag{3.3}$$

where  $h(\mathcal{D}_K)$  depends only on the observations  $\mathcal{D}_K$  and not on the unknown parameters  $f_j$ . Inserting (3.3) into (3.2) yields

$$u_{\mathcal{D}_K}(\mathbf{f}) \propto \prod_{j=1}^{J-1} e^{-\frac{1}{2} \frac{(f_j - \hat{f}_j)^2}{\sigma_j^2 S_j^{-1}}} \text{ for } f_j \in (-a, a). \tag{3.4}$$

Now we make the prior non informative by considering the limiting case  $a \rightarrow \infty$ . Then we obtain immediately from (3.4) that the posterior density of  $\mathbf{F}$  is given by

$$u_{\mathcal{D}_K}(\mathbf{f}) \propto \prod_{j=1}^{J-1} \left\{ \frac{1}{\sqrt{2\pi\sigma_j^2 S_j^{-1}}} e^{-\frac{1}{2} \frac{(f_j - \hat{f}_j)^2}{\sigma_j^2 S_j^{-1}}} \right\} f_j \in \mathbb{R}. \tag{3.5}$$

From (3.5) we see that a posteriori, given the observed triangle  $\mathcal{D}_K$ , the random variables  $F_1, \dots, F_{j-1}$  are independent and normally distributed with

$$E[F_i | \mathcal{D}_K] = \hat{f}_j,$$

$$\text{Var}[F_i | \mathcal{D}_K] = \frac{\sigma_j^2}{S_j}.$$

Hence the Bayes estimator of  $X_{k,j}$  is

$$X_{k,j}^{\text{Bayes}} = X_{k,K-k+1} \cdot E \left[ \prod_{l=K-k+1}^{j-1} F_l | \mathcal{D}_K \right] = X_{k,K-k+1} \prod_{l=K-k+1}^{j-1} \hat{f}_l = \hat{X}_{k,j},$$

and for the mean square error of the Bayes estimator we obtain

$$E \left[ \left( \hat{X}_{k,j} - X_{k,j} \right)^2 | \mathcal{D}_K \right] =$$

$$= X_{k,K-k+1}^2 \cdot E \left[ \left( \prod_{l=K-k+1}^{j-1} \hat{f}_l - \prod_{l=K-k+1}^{j-1} F_l \right)^2 | \mathcal{D}_K \right] \tag{3.6}$$

$$= X_{k,K-k+1}^2 \cdot \left( \prod_{l=K-k+1}^{j-1} \hat{f}_l^2 - 2 \prod_{l=K-k+1}^{j-1} E[F_l | \mathcal{D}_K] \cdot \hat{f}_l + \prod_{l=K-k+1}^{j-1} E[F_l^2 | \mathcal{D}_K] \right)$$

$$= X_{k,K-k+1}^2 \cdot \left( \prod_{l=K-k+1}^{j-1} E[F_l^2 | \mathcal{D}_K] - \prod_{l=K-k+1}^{j-1} \hat{f}_l^2 \right)$$

$$= X_{k,K-k+1}^2 \cdot \left( \prod_{l=K-k+1}^{j-1} \left( \hat{f}_l^2 + \frac{\sigma_l^2}{S_l} \right) - \prod_{l=K-k+1}^{j-1} \hat{f}_l^2 \right). \tag{3.7}$$

By replacing  $\sigma_l^2$  in (3.7) by  $\hat{\sigma}_l^2$  we obtain the BMW formula (2.4).

The Mack formula (2.2) is also obtained in the Bayesian framework by the following Taylor-approximation:

$$\prod_{l=K-k+1}^{j-1} \hat{f}_l - \prod_{l=K-k+1}^{j-1} F_l \simeq \sum_{l=K-k+1}^{j-1} (\hat{f}_l - F_l) \frac{\partial}{\partial F_l} \prod_{l=K-k+1}^{j-1} F_l \bigg|_{\hat{f}_l}$$

$$= \sum_{l=K-k+1}^{j-1} (\hat{f}_l - F_l) \prod_{\substack{h=K-k+1 \\ h \neq l}}^{j-1} \hat{f}_h. \tag{3.8}$$

By inserting (3.8) into (3.6) we obtain

$$\begin{aligned}
 \hat{e}_{k,j} &\simeq X_{k,K-k+1}^2 E \left[ \left( \sum_{l=K-k+1}^{j-1} (\hat{f}_l - F_l) \prod_{\substack{h=K-k+1 \\ h \neq l}}^{j-1} \hat{f}_h \right)^2 \middle| \mathcal{D}_K \right] \\
 &= X_{k,K-k+1}^2 \left( \prod_{l=K-k+1}^{j-1} \hat{f}_l \right)^2 E \left[ \left( \sum_{l=K-k+1}^{j-1} \frac{(\hat{f}_l - F_l)}{\hat{f}_l} \right)^2 \middle| \mathcal{D}_K \right] \\
 &= X_{k,K-k+1}^2 \left( \prod_{l=K-k+1}^{j-1} \hat{f}_l \right)^2 \sum_{l=K-k+1}^{j-1} E \left[ \frac{(\hat{f}_l - F_l)^2}{\hat{f}_l^2} \middle| \mathcal{D}_K \right] \tag{3.9}
 \end{aligned}$$

$$= X_{k,K-k+1}^2 \left( \prod_{l=K-k+1}^{j-1} \hat{f}_l \right)^2 \sum_{l=K-k+1}^{j-1} \frac{\sigma_l^2}{\hat{f}_l^2} \frac{1}{S_l}, \tag{3.10}$$

which is identical to (2.2) if we replace  $\sigma_j^2$  in (3.10) by  $\hat{\sigma}_j^2$ .

### 3.2. Bayesian model II: Gamma distribution

In this more realistic model, where the  $X_{k,j}$  are always strictly positive, we assume that conditionally, given  $\mathbf{F}$ , the observed chain ladder factors are Gamma-distributed, i.e. we make the following assumptions:

- (G1) Conditionally, given  $\mathbf{F}$  and  $\mathcal{B}_j$ , the random variables  $Y_{k,j}$  ( $k = 1, \dots, K-j$ ) are independent and Gamma-distributed with moments

$$E[Y_{k,j}|F_j] = F_j, \tag{3.11}$$

$$\text{Var}[Y_{k,j}|F_j, X_{k,j}] = \frac{\tau_j^2 F_j^2}{X_{k,j}}, \tag{3.12}$$

where  $\tau_j^2$  are given positive constants.

- (G2)  $\{F_1, \dots, F_{j-1}\} = \{\Theta_1^{-1}, \dots, \Theta_{j-1}^{-1}\}$ , where the random variables  $\Theta_1, \dots, \Theta_{j-1}$  are independent and uniformly distributed on  $(a^{-1}, a)$ , where  $a \in \mathbb{R}^+$ .

Remarks:

- Conditionally, given  $\mathbf{F} = \mathbf{f}$ , the chain ladder assumptions are fulfilled with  $\sigma_j^2 = \tau_j^2 f_j^2$ .
- The difficulty of this model is that  $\sigma_j^2$  depends on the parameter  $f_j$  to be estimated. Contrary to the normal model it is therefore not possible here that  $E[F_l^2|\mathcal{D}_K]$  in the Bayesian model is identical to  $E[\hat{f}_l^2|\mathcal{B}_l]$  in the classical model, since the latter depends on the parameter  $f_l$  to be estimated. We can therefore not expect to get out of this model exactly the BMW formula (2.4).

- Assumption (G1) can also be written as a time-series' model in the following way: conditionally, given  $\mathbf{F} = \mathbf{f}$ , the triangle  $\mathcal{D}_K$  is generated by the process

$$X_{k,j+1} = \tilde{G}_{k,j} \cdot X_{k,j},$$

where conditionally, given the observations in column  $j$ , the r.v.  $\tilde{G}_{k,j}$  ( $k = 1, \dots, K-j$ ) are independent and Gamma distributed with

$$E[\tilde{G}_{k,j}] = f_j,$$

$$\text{Var}[\tilde{G}_{k,j}|X_{k,j}] = \frac{\tau_j^2 \cdot f_j^2}{X_{k,j}}.$$

- It follows from assumption (G1) that on the condition  $\Theta = \vartheta$ ,  $Y_{k,j}$  ( $k = 1, \dots, K-j$ ) are Gamma-distributed with

$$\text{shape parameter } \gamma_{k,j} = X_{k,j} \tau_j^{-2} \text{ and} \tag{3.13}$$

$$\text{scale parameter } c_{k,j} = X_{k,j} \tau_j^{-2} \vartheta_j, \tag{3.14}$$

i.e. the conditional density of  $Y_{k,j}$  is given by

$$u_{\vartheta_j}(y) = \frac{(c_{k,j})^{\gamma_{k,j}}}{\Gamma(\gamma_{k,j})} y^{\gamma_{k,j}-1} e^{-c_{k,j}y}.$$

- Note also that

$$\text{Var}[\hat{F}_j | \Theta_j, \mathcal{B}_j] = \frac{\tau_j^2 F_j^2}{S_j}.$$

- Assumption (G2) is equivalent to the assumption that the density of the  $F_j$  is given by

$$u(f) = \frac{a^2 - 1}{a} f^{-2}.$$

- It might look a bit strange that we have introduced  $\Theta_j$  in assumption (G2). The reason is just mathematical feasibility.

We derive again the Bayes estimator of the unknown chain ladder factors  $F_j$  as well as of  $X_{k,j}$  and the mean square error of these estimators. On the condition  $\Theta = \vartheta$  and  $\mathcal{B}_1$ , the joint density of  $\{\mathbf{Y}_1, \dots, \mathbf{Y}_{J-1}\}$  is given by

$$u_{\vartheta}(\mathbf{y}_1, \dots, \mathbf{y}_{J-1}) = \prod_{j=1}^{J-1} \prod_{k=1}^{K-j} \frac{(x_{k,j} \tau_j^{-2} \vartheta_j)^{x_{k,j} \tau_j^{-2}}}{\Gamma(x_{k,j} \tau_j^{-2})} y_{k,j}^{x_{k,j} \tau_j^{-2} - 1} e^{-x_{k,j} \tau_j^{-2} \vartheta_j y_{k,j}}, \tag{3.15}$$

where  $x_{k,j} = X_{k,1} \cdot y_{k,1} \cdot \dots \cdot y_{k,j-1}$ .

Denote by  $u_{\mathcal{D}_K}(\vartheta_1, \dots, \vartheta_{J-1})$  the posterior density of  $\Theta_1, \dots, \Theta_{J-1}$  given the observed triangle  $\mathcal{D}_K$ . From  $x_{k,j} y_{k,j} = X_{k,j+1}$  and (3.15) follows that

$$u_{\mathcal{D}_K}(\vartheta_1, \dots, \vartheta_{J-1}) \propto \prod_{j=1}^{J-1} \vartheta_j^{\sum_{k=1}^{K-j} X_{k,j} \tau_j^{-2}} e^{-\sum_{k=1}^{K-j} X_{k,j} \tau_j^{-2} \vartheta_j} \text{ for } \vartheta_j \in (a^{-1}, a), \quad (3.16)$$

where terms not depending on  $\vartheta_j (j = 1, \dots, J)$  are included in the normalising constant and therefore omitted on the right hand side of the equation.

Now we consider the limiting case for  $a \rightarrow \infty$ . Because  $S_j = \sum_{k=1}^{K-j} X_{k,j}$ , we then obtain from (3.16) that

$$u_{\mathcal{D}_K}(\vartheta_1, \dots, \vartheta_{J-1}) \propto \prod_{j=1}^{J-1} \vartheta_j^{S_j \tau_j^{-2}} e^{-\sum_{k=1}^{K-j} X_{k,j+1} \tau_j^{-2} \vartheta_j} \text{ for } \vartheta_j \in \mathbb{R}^+. \quad (3.17)$$

From (3.17) follows that a posterior, given the observed triangle  $\mathcal{D}_K$ ,

- the random variables  $\Theta_1, \dots, \Theta_{J-1}$  are independent,
- the  $\Theta_j$  are Gamma distributed with

$$\text{shape parameter } \gamma_j = S_j \tau_j^{-2} + 1 \text{ and}$$

$$\text{scale parameter } c_j = \sum_{k=1}^{K-j} X_{k,j+1} \tau_j^{-2}.$$

Remark:

- If we take an exponential distribution with density

$$u(\vartheta) \propto \lambda e^{-\lambda \vartheta}$$

as a prior for  $\Theta_j$  and if we consider the limiting case  $\lambda \rightarrow 0$ , we obtain the same a posteriori distribution as in (3.17).

For the first and second order moments we get from (3.17)

$$\begin{aligned} E[F_j | \mathcal{D}_K] &= \frac{\left(\sum_{k=1}^{K-j} X_{k,j+1} \tau_j^{-2}\right)^{S_j \tau_j^{-2} + 1}}{\Gamma(S_j \tau_j^{-2} + 1)} \int_0^\infty \vartheta_j^{-1} \vartheta_j^{S_j \tau_j^{-2}} e^{-\sum_{k=1}^{K-j} X_{k,j+1} \tau_j^{-2} \vartheta_j} \\ &= \frac{\left(\sum_{k=1}^{K-j} X_{k,j+1} \tau_j^{-2}\right)^{S_j \tau_j^{-2} + 1}}{\Gamma(S_j \tau_j^{-2} + 1)} \frac{\Gamma(S_j \tau_j^{-2})}{\left(\sum_{k=1}^{K-j} X_{k,j+1} \tau_j^{-2}\right)^{S_j \tau_j^{-2}}} \\ &= \frac{\sum_{k=1}^{K-j} X_{k,j+1}}{S_j} \\ &= \hat{f}_j, \end{aligned}$$

and by an analogous calculation

$$\begin{aligned}
 E[F_j^2 | \mathcal{D}_K] &= \begin{cases} \frac{\sum_{k=1}^{K-j} X_{k,j+1}}{S_j} \frac{\sum_{k=1}^{K-j} X_{k,j+1} \tau_j^{-2}}{S_j \tau_j^{-2} - 1} & \text{if } S_j > \tau_j^2, \\ \infty & \text{otherwise} \end{cases} \\
 &= \begin{cases} \hat{f}_j^2 \left( 1 + \frac{\tau_j^2}{S_j - \tau_j^2} \right) & \text{if } S_j > \tau_j^2, \\ \infty & \text{otherwise} \end{cases}
 \end{aligned}$$

Hence the Bayes estimator of  $X_{k,j}$  is given by

$$X_{k,j}^{\text{Bayes}} = X_{k, K-k+1} \cdot E \left[ \prod_{l=K-k+1}^{j-1} F_l | \mathcal{D}_K \right] = X_{k, K-k+1} \prod_{l=K-k+1}^{j-1} \hat{f}_l = \hat{X}_{k,j},$$

and for the mean square error of the Bayes estimator we obtain

$$\begin{aligned}
 E \left[ \left( \hat{X}_{k,j} - X_{k,j} \right)^2 | \mathcal{D}_K \right] &= \\
 &= X_{k, K-k+1}^2 \cdot E \left[ \left( \hat{f}_{K-k+1} \cdot \dots \cdot \hat{f}_{j-1} - F_{K-k+1} \cdot \dots \cdot F_{j-1} \right)^2 | \mathcal{D}_K \right] \\
 &= X_{k, K-k+1}^2 \cdot \left( \prod_{l=K-k+1}^{j-1} E[F_l^2 | \mathcal{D}_K] - \prod_{l=K-k+1}^{j-1} \hat{f}_l^2 \right) \tag{3.18} \\
 &= \begin{cases} X_{k, K-k+1}^2 \cdot \left( \prod_{l=K-k+1}^{j-1} \hat{f}_l^2 \left( 1 + \frac{\tau_l^2}{S_l - \tau_l^2} \right) - \prod_{l=K-k+1}^{j-1} \hat{f}_l^2 \right) & \text{if } S_l > \tau_l^2 \text{ for all } l. \\ \infty & \text{otherwise} \end{cases}
 \end{aligned}$$

Formula (3.18) is not directly comparable to (2.4) because  $\tau_l^2$  is not the same parameter as  $\sigma_l^2$ . Indeed the difficulty with this model to be comparable with the classical formula is that  $\sigma_l^2 = \tau_l^2 F_l^2$  and thus depends itself on the parameter to be estimated. A natural estimator for  $\tau_l^2$  is

$$\hat{\tau}_l^2 = \frac{\hat{\sigma}_l^2}{\hat{f}_l^2}.$$

If we replace  $\tau_l^2$  in (3.18) with  $\hat{\tau}_l^2$  we obtain

$$\hat{e}_{k,j} = \begin{cases} X_{k,K-k+1}^2 \cdot \left( \prod_{l=K-k+1}^{j-1} \left( \hat{f}_l^2 + \frac{\hat{\sigma}_l^2}{S_l - \frac{\hat{\sigma}_l^2}{\hat{f}_l^2}} \right) - \prod_{l=K-k+1}^{j-1} \hat{f}_l^2 \right) & \text{if } S_l > \frac{\hat{\sigma}_l^2}{\hat{f}_l^2} \text{ for all } l, \\ \infty & \text{otherwise} \end{cases} \tag{3.19}$$

(3.19) is not exactly the same as the BMW formula (2.4) and yields higher values than (2.4). The resulting estimates are approximately the same in the case where  $S_l > \frac{\hat{\sigma}_l^2}{\hat{f}_l^2}$  for all  $l$ .

### 3.3. Summary and Conclusions

- The estimation error resulting in the Bayesian normal model coincides with the formula of BMW. In this normal model the  $X_{k,j}$  could become negative, which is problematic from a theoretical point of view. In the Gamma Bayesian model we have obtained a formula, which is similar but not fully identical to the BMW formula. Hence we could not find a model which doubtlessly confirmed the new approach (2.4) of BMW. Nevertheless, by comparing the results with BMW and Mack, the results seem to be rather in favour of the formula of BMW.
- One question and criticism in [3] is, whether it is allowed and whether it makes sense to take the product

$$E \left[ \prod_{l=K-k+1}^{j-1} \hat{f}_l^2 \mid \mathcal{B}_l \right]$$

in (2.3). From a Bayesian point of view, the answer to this question is definitely positive the reason being that in the Bayesian context the random variables  $F_j$  are a posteriori independent and that therefore

$$E \left[ \prod_{l=K-k+1}^{j-1} F_l^2 \mid \mathcal{D}_K \right] = \prod_{l=K-k+1}^{j-1} E[F_l^2 \mid \mathcal{D}_K].$$

The result that the r.v.  $F_j$  are a posteriori independent is not a specific of the two considered Bayesian models. It results essentially from the natural model assumptions that, on the condition  $\mathbf{F} = \mathbf{f}$ ,  $X_{k,j}$  ( $j = 1, \dots, J$ ) form a Markov process and that the r.v.  $F_j$  are a priori independent. The posterior independency property will therefore also be encountered in other Bayesian models with the above natural assumptions.

- The question, whether one should consider the average over the set of all possible true factors  $f_j$  or the average over all possible upper right corners

$\mathcal{D}_K^O$  given  $\mathcal{B}_{K-k+1}$ , is rather a question of philosophy than mathematics. The first definition of average is in my view probably better suited to the problem of estimating the conditional estimation error. Then, based on the results above, I would give preference to the formula of BMW. If one takes the average over all possible upper right corners  $\mathcal{D}_K^O$  given  $\mathcal{B}_{K-k+1}$  then the formula of Mack should be preferred, because BMW has then an upward bias whereas Mack does not. In the classical as well as in the Bayesian framework, the Mack formula is obtained by a first order Taylor approximation to BMW. Therefore the difference between BMW and Mack will mostly be rather small and not essential for practical purposes. Finally, what is called the conditional approach in the paper by BMW, is a third way to look at the problem. This approach can be described as follows: each of the r.v.  $Y_{k,j}$  is associated with a weight  $w_{k,j}$  and the variance of  $Y_{k,j}$  is inversely proportional to  $w_{k,j}$ ; the r.v.  $Y_{k,j}$  are then considered as independent and the weights are taken from the observed triangle, i.e.  $w_{k,j} = X_{k,j} = \text{observed value in the triangle}$ .

- We have focused on the problem of the estimation error, since the discussion on this part of the mean square prediction error is controversial. We should however remark here that, surprisingly, the estimate of the process variance in the Bayesian framework would differ and would be greater than the estimate suggested by both, Mack and BMW. If we consider the Bayesian approach as the stringent conditional approach, then the estimate of the process variance would have to be changed too.

#### ACKNOWLEDGEMENT

I would like to thank Thomas Mack for many fruitful discussions which have helped me to better understand many features and which have very much improved the quality of this discussion contribution. I would also like to thank the authors BMW of [1] for their comments on an earlier version and for having brought up a problem which might also be relevant in other areas.

#### REFERENCES

- [1] BUCHWALDER M., BÜHLMANN H., MERZ M., WÜTHRICH M.V. (2006) The Mean Square Error of Prediction in the Chain Ladder Reserving Method, *ASTIN-Bulletin* **00(0)**, 000-000.
- [2] MACK T. (1993) Distribution-free calculation of the standard error of Chain ladder reserve estimates. *ASTIN Bulletin* **23(2)**, 213-225.
- [3] MACK T., QUARG G., BRAUN Ch. (2006) The Mean Square Error of Prediction in the Chain Ladder Reserving Method – a Comment. *ASTIN Bulletin* **00(0)**, 000-000.

ALOIS GISLER  
 Winterthur Insurance Company  
 Römerstrasse 17  
 Box 357 – CH-8401 Winterthur  
 Switzerland