Interval Estimation of the Credibility Factor

by Wu-Chyuan Gau, Ashis Gangopadhyay, and Zhongxian Han

ABSTRACT

In this article, we present a Bayesian approach for calculating the credibility factor. Unlike existing methods, a Bayesian approach provides the decision maker with a useful credible interval based on the posterior distribution and the posterior summary statistics of the credibility factor, while most credibility models only provide a point estimate. A simulated example is used to demonstrate the advantages and disadvantages of the Bayesian credibility factor proposed in this article.

KEYWORDS

Credibility, interval estimation

1. Introduction

Credibility modeling is a ratemaking process which allows actuaries to adjust future premiums according to past experience of a risk or group of risks. For instance, Herzog (1994) considered two sets of data. The first is the collection of current observations from the most recent period. The second is the collection of observations from one or more prior periods. Under some credibility approaches, the new rate, C (for claim frequency, claim size, aggregate claim amount, etc.), is calculated by:

$$C = Z \times R + (1 - Z) \times H, \tag{1}$$

where R is the mean of the current observations, H is the mean of the prior observations, and Z is the credibility factor, ranging from 0 to 1. The credibility factor Z denotes the weight assigned to the current data and $(1 - Z)$ is the weight assigned to the prior data. Zero credibility ($Z = 0$) will be assigned to data too small to be used for ratemaking, while full credibility ($Z = 1$) is assigned to fully credible data.

Bailey (1950) showed that the formula $ZR + (1 - Z)H$ can be derived from Bayes' theorem, either by using a Bernoulli-Beta model on the unknown parameter p, or by using a Poisson-Gamma model on the unknown parameter λ. Bailey's work led to the application of Bayesian methodology to credibility theory. Bayesian statistical analysis for a selected model begins by first quantifying the existing state of knowledge and assumptions. These prior inputs are then combined with information from observed data quantified probabilistically through the likelihood function. The mechanism of prior and likelihood combination is Bayes' theorem. In technical terms, the posterior is proportional to the prior and the likelihood, i.e.,

$$\text{posterior} \propto \text{prior} \times \text{likelihood}.$$

However, the prior distributional assumptions in Bailey's models were severely limited. Bühlmann (1967) overcame these limitations and proved that Equation (1) is also a distribution-free credibility formula. The best linear approximation of this formula is found by minimizing $E\{E[\mu(\theta) \mid x_1,\ldots,x_n] - (a + b\bar{X})\}^2$, where $\mu(\theta)$ is the mean of an individual risk (or the hypothetical mean $E(X_i \mid \theta)$), characterized by a risk parameter θ, and $\bar{X} = (x_1 + x_2 + \cdots + x_n)/n$. Additionally, the process variance, $v(\theta)$, is defined as $\mathrm{Var}(X_i \mid \theta)$. Bühlmann and Straub (1970) then formalized the least squares derivation of

$$Z = n/(n + k), \tag{2}$$

where n is the number of trials or exposure units and

$$k = \frac{v}{a}, \tag{3}$$

in which $v = E[v(\theta)]$ and $a = \mathrm{Var}[\mu(\theta)]$. Here, v and a are also known as the expected value of the process variance and the variance of hypothetical mean, respectively. This methodology is called empirical Bayes credibility, although the Bayesian content of this approach has been greatly minimized.

In practice, we have to estimate v and a to determine the credibility factor Z. Naturally, actuaries use unbiased estimators of v and a, denoted as $\hat{v}$ and $\hat{a}$ respectively. When a sample of claims is available, $\hat{v}$ and $\hat{a}$ are then realized. However, actuaries traditionally stop at a point estimate without considering possible variation caused by a random sample. Therefore, we take a Bayesian approach and treat the unknown quantities v and a as random variables. This allows us to estimate v, a, and Z simultaneously and allows the assessment of the credibility factor in terms not only of point estimators, but also of certain characteristics of probability distributions.

To date, Bayesian methodology has been used in various areas within actuarial science. Klugman (1992) provided a Bayesian analysis to credibility theory by carefully choosing a parametric conditional loss distribution for each risk and a parametric prior. To be useful in public discussion, such a prior must be evidence-based in some sense, e.g., be a summary of an expert's opinion on the topic. Bayesian methods provide a natural way to incorporate this prior information, whether statistical or not, in the form of tables or as expert judgement, through the prior distribution of the parameters.

In many cases, Bayesian methods can provide analytic closed-formed solutions for the posterior distribution and the predictive distribution of the variables involved. Then, the inference is carried out directly from these distributions and any of their characteristics and properties. However, if the distribution is not a known type, or if it does not have a closed form, then it is possible to derive approximations by means of Markov chain Monte Carlo (MCMC) simulation methods.

In summary, MCMC algorithms construct the desired posterior distributions of the parameters. Thus, when convergence is reached, it provides a sample of the posterior distribution that can be used for any posterior summary statistics of interest. Details of the technicalities involved in MCMC can be found in Smith and Roberts (1993) or in Gilks, Richardson, and Spiegelhalter (1998). MCMC algorithms have appeared in actuarial literature by de Alba (2002), Carlin (1992), Frees (2003), Gangopadhyay and Gau (2003), Ntzoufras and Dellaportas (2002), Rosenberg and Young (1999), and Scollnik (2001). The advantage of using this procedure is that actuaries can obtain point estimates as well as probability intervals and other summary measures, such as means, variances, and quantiles.

In this article, we provide an alternative method for calculating the credibility factor, particularly the interval estimation of the credibility factor. There is a range of concerns that arise in credibility modeling. We present an effective method to deal with these concerns. Section 2 introduces the credibility problem. Section 3 uses a simulated example to illustrate the basic idea of the Bayesian credibility factor. Section 4 discusses the advantages and disadvantages of the Bayesian credibility factor. Remarks are made in Section 5.

2. Credibility problem

The classical data type in this area involves realizations from the past and present experience of individual policyholders. Suppose there are r different policyholders. We have a claims record in year j, $j = 1,\ldots,n_i$, for policyholder i, $i = 1,\ldots,r$. Therefore, the data can be summarized in the following form,

$$\begin{array}{l} X_{11}, X_{12}, \ldots, X_{1,n_1} \\ X_{21}, X_{22}, \ldots, X_{2,n_2} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ X_{r1}, X_{n2}, \ldots, X_{r,n_r}, \end{array} \tag{4}$$

where X_{ij} can be the losses per exposure unit, the number of claims, or the loss ratio from insurance portfolios. The goal is to estimate the amount or number of claims to be paid on a particular insurance policy in a future coverage period.

2.1. Variance component models

Dannenburg, Kaas, and Goovaerts (1996) introduced the use of variance component models to the credibility problem. In a variance component model, each cell (policyholder i in year j) consists of a number of contracts m_{ij}, which has been observed over a number of observation periods n_i for contract i. Then, for the ith contract in the portfolio, $i = 1,\ldots,r$, in year $j = 1,\ldots,n_i$, the claim experience is represented by the model

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1,\ldots,r \quad \text{and} \quad j = 1,\ldots,n_i \tag{5}$$

where α_i and ε_{ij} are independent with $E[\alpha_i] = 0$, $\mathrm{Var}[\alpha_i] = a$ and $E[\varepsilon_{ij}] = 0$, $\mathrm{Var}[\varepsilon_{ij}] = \upsilon/m_{ij}$ for all i and j.

In order to determine the credibility factor and the credibility premium, we need to estimate parameters μ, υ, and a. In general, the unknown parameters μ, υ, and a are associated with the structural density, say $\pi(\theta)$, and hence we refer to these as structural parameters. For the Bühlmann and Straub (1970) formulation, the hypothetical mean is defined as $E(X_{ij} \mid \Theta_i = \theta_i) \equiv \mu(\theta_i)$ and the process variance is defined as $\text{Var}(X_{ij} \mid \Theta_i = \theta_i) \equiv \upsilon(\theta_i)/m_{ij}$. Thus, the structural parameters are given by

$$\mu = E[\mu(\Theta_i)], \qquad \upsilon = E[\upsilon(\Theta_i)], \tag{6}$$

and

$$a = \text{Var}[\mu(\Theta_i)]. \tag{7}$$

Consequently, the credibility factor Z_i for each risk is given by

$$Z_i = \frac{m_i}{m_i + k}, \tag{8}$$

where

$$k = \frac{\upsilon}{a} \tag{9}$$

and

$$m_i = \sum_{j=1}^{n_i} m_{ij}, \qquad i = 1,\ldots,r. \tag{10}$$

2.2. Traditional credibility modeling

If estimators of μ, υ, and a are denoted by $\hat{\mu}$, $\hat{\upsilon}$, and $\hat{a}$, respectively, then the resulting credibility premium is given by

$$\hat{P}_i = \hat{Z}_i \bar{X}_i + (1 - \hat{Z}_i)\hat{\mu}, \qquad i = 1,\ldots,r, \tag{11}$$

where

$$\hat{Z}_i = \frac{m_i}{m_i + \hat{k}}, \tag{12}$$

$$\hat{k} = \frac{\hat{\upsilon}}{\hat{a}}, \tag{13}$$

and

$$\bar{X}_i = \frac{\sum_{j=1}^{n_i} m_{ij} X_{ij}}{m_i}, \qquad i = 1,\ldots,r. \tag{14}$$

Today, one of the most widely used methodologies for the choice of $\hat{\mu}$, $\hat{\upsilon}$, and $\hat{a}$ is empirical Bayes parameter estimation. It allows us to use the data at hand to estimate the structural parameters. The resulting estimators are

$$\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^{r} m_i \bar{X}_i}{m}, \tag{15}$$

$$\hat{\upsilon} = \frac{\sum_{i=1}^{r} \sum_{j=1}^{n_i} m_{ij}(X_{ij} - \bar{X}_i)^2}{\sum_{i=1}^{r}(n_i - 1)}, \tag{16}$$

and

$$\hat{a} = \frac{\sum_{i=1}^{r} m_i(\bar{X}_i - \bar{X})^2 - (r-1)\hat{\upsilon}}{m - \dfrac{\sum_{i=1}^{r} m_i^2}{m}}, \tag{17}$$

where $m = \sum_{i=1}^{r} m_i$. Note that Equations (15), (16), and (17) are unbiased estimators of μ, υ, and a, respectively.

However, there are some issues raised from traditional credibility modeling. First, the estimate of a can be negative, a clearly unacceptable value. The second issue is the need for obtaining a measure of the quality of the credibility estimate. The measure of error for Equation (11) depends on a, υ, and the credibility factor Z. None of the standard approaches to credibility analysis provides a method of accounting for this extra variability.

3. A simulated example

In this section, we consider an example based on data simulated from a normal distribution. It is always a concern with this assumption in credibility modeling due to the likelihood of a heavy tail in loss distributions. However, the normal model is still a very useful approach in many problems. For non-normal data, we will recommend a data transformation before using a more sophisticated model. Variable transformations serve a variety of purposes in data analysis, and are used in particular to make distributions more symmetric (or normal), to stabilize variation, and to render relationships between variables more nearly linear. These techniques can be found in Box and Tidwell (1962), Box and Cox (1964), and Klugman (1992).

Based on Equation (5), it is assumed that

1. $\mu = 200$.
2. α_i has a normal distribution with mean 0 and variance 400 (i.e., $a = 400$).
3. ε_{ij} has a normal distribution with mean 0 and variance 2500 (i.e., $\upsilon = 2500$).

Specifically, the data is generated using the following mechanism:

1. $\theta_i \overset{\text{i.i.d.}}{\sim} N(200, 400)$.
2. $X_{ij} \mid \theta_i \overset{\text{i.i.d.}}{\sim} N(\theta_i, 2500)$.

We consider a portfolio of five policyholders with five years of experience. There is only one claim every year for each policyholder. That is $r = 5$, $n_i = 5$, and $m_{ij} = 1$ in Equation (5). Using Equations (8), (9), and (10), we have the true credibility factor

$$Z = \frac{5}{5 + \frac{2,500}{400}} = 0.4444. \tag{18}$$

Table 1. Simulated data

Policyholder	θ_i	Year 1	Year 2	Year 3	Year 4	Year 5
1	207	242	183	237	141	125
2	196	157	181	268	232	220
3	205	219	185	151	261	120
4	215	331	151	239	203	206
5	189	138	213	222	174	189

Table 1 shows simulated data from the above settings, in which $\theta_i = \mu + \alpha_i$. Using Equations (15) to (17), we have $\hat{\mu} = 200$, $\hat{\upsilon} = 2684$, and $\hat{a} = -201$. Traditionally, actuaries will set $\hat{a}$ to be zero when $\hat{a}$ is negative. Thus, the credibility factor is calculated as

$$\hat{Z} = \frac{5}{5 + \frac{2,684}{0}} = 0. \tag{19}$$

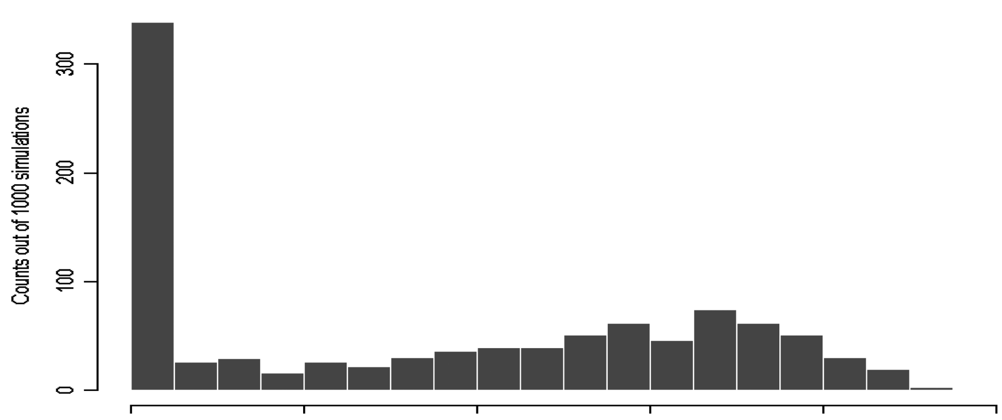
Clearly, this is a situation in which traditional credibility factor analysis does not perform well. Figure 1 shows the histogram of credibility factors based on 1000 simulated portfolios. There are a large number of simulated portfolios (about 300 out of 1000), resulting in a zero credibility factor due to a negative value of $\hat{a}$.

3.1. Bayesian credibility factor

As seen previously, a negative value of $\hat{a}$ is a major drawback of this unbiased estimator. Therefore, we seek other alternatives to estimate the credibility factor Z. The Bayesian method and the MCMC technique can be applied with objective selection of a model structure and prior distributions based on actuarial judgement.

In this section, we will focus on the variance components model introduced in Equation (5). A general introduction to Bayesian inference on variance components can be found in Searle, Casella, and McCulloch (1992). Bayesian cred-

Figure 1. Histogram of credibility factors



ibility factors in actuarial applications are introduced in Appendix A. The benefit of a Bayesian approach is that it provides the decision maker with a posterior distribution of the credibility factor as well as a posterior distribution of the premium.

In actuarial science, both outcomes and predictors are often gathered in a nested or hierarchical fashion (for example, fires within counties within states, employees within companies within industries, and patients within hospitals). Thus, as observed by various researchers in actuarial science, hierarchical models are ideally suited to the insurance business in which single- or multi-stage samples are routinely drawn. A number of examples can be found in Klugman (1992) and Scollnik (2001).

For the simulated example, we use the following hierarchical setting.

1. $X_{ij} \mid \theta_i, \upsilon \sim N(\theta_i, \upsilon)$.
2. $\theta_i = \mu + \alpha_i$.
3. $\theta_i \mid \mu, a \sim N(\mu, a)$.
4. μ is an unknown constant, estimated by $\hat{\mu}$ in Equation (15).
5. $a \sim \text{Gamma}(\alpha_a, \beta_a)$.
6. $\upsilon \sim \text{Gamma}(\alpha_\upsilon, \beta_\upsilon)$.
7. The Bayesian credibility factor is given by Equation (25) in Appendix A.

The Bayesian approach is a powerful formal alternative to deterministic and classical statistical methods when prior information is available. The choice of prior is often presented as an aspect of personal belief. In Appendix B, we present an empirical Bayesian approach for priors of υ and a.

Based on Equations (37) and (38) in Appendix B, we have $\alpha_\upsilon = \sum_{i=1}^{r}(n_i - 1)/2 = 10$ and $\beta_\upsilon = \alpha_\upsilon/\hat{\upsilon}_{\text{init}} = 10/\hat{\upsilon}_{\text{init}}$. The parameters for the prior distribution of a are given by $\alpha_a = (r-1)/2 = 2$ and $\beta_a = \alpha_a/\hat{a}'_{\text{init}} = 2/\hat{a}'_{\text{init}}$ as shown by Equations (39) and (40) in Appendix B.

Using the simulated data in Table 1, we have $\hat{\upsilon}_{\text{init}} = 2684$ and $\hat{a}'_{\text{init}} = 336$. Therefore, $\beta_\upsilon = 10/2684 = 0.0037$ and $\beta_a = 2/336 = 0.0059$. The estimation of the credibility factor Z, implemented with WinBUGS (Spiegelhalter, Thomas, and Gilks, 1996), is summarized in Table 2.

Table 2. Summary statistics for Bayesian credibility factor $\hat{Z}_B$

Mean	2.5%	Median	97.5%
0.2985	0.0511	0.2879	0.6026

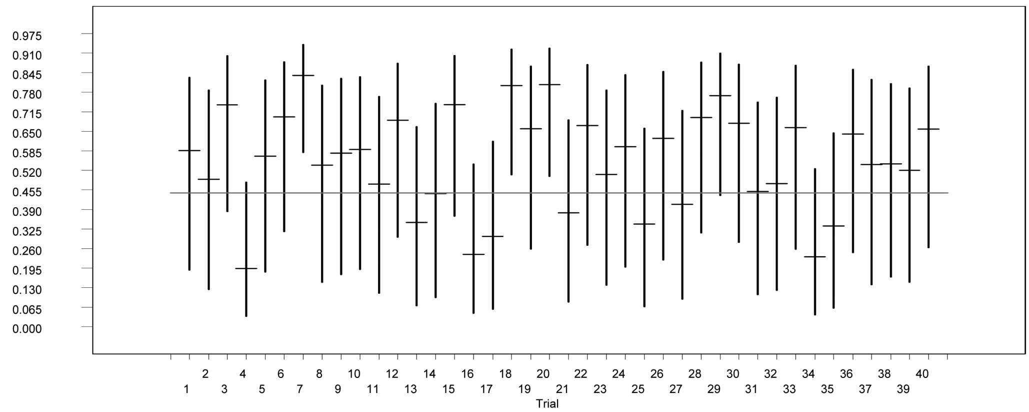
The Bayesian approach suggests that the Bayesian credibility factor, $\hat{Z}_B$, is the mean of the posterior distribution as indicated in Equation (25). Thus, we have $\hat{Z}_B = 0.2985$. Nevertheless, actuaries must recognize a possible variation inherent in the estimation of the credibility factor as shown in Table 2. Clearly, Table 2 gives us a more desirable result compared with the result from Equation (19).

To assess the accuracy of the Bayesian credibility interval estimator, we simulate 40 portfolios from the normal-normal model with $\upsilon = 50^2$ and $a = 20^2$; and we construct a 95% credible interval for each portfolio based on the posterior distribution of the Bayesian credibility factor. These credible intervals are shown in Figure 2, where the horizontal line is the true credibility factor $Z = 0.4444$ and the dash line is the median of the posterior distribution of the Bayesian credibility factor. The true credibility factor falls in the credible interval 37 times out of 40 trials. For a detailed introduction to using WinBUGS in actuarial applications, the reader is referred to Scollnik (2001).

4. Further analysis

In this section, we extend our analysis to the advantages and disadvantages of the Bayesian credibility factor proposed in this article. What makes credibility theory work is that it results in a significant improvement in mean-squared error, even though the resulting credibility pre-

Figure 2. 95% Credible interval of Z



mium given by Equation (11) is a biased estimator. We expect that the biases will cancel out over the entire estimation process.

4.1. Sampling distribution

One advantage of the Bayesian credibility factor approach is its ability to describe the variation inherent in the process of estimating the credibility factor. Traditionally, actuaries use Equation (12) to determine the credibility factor for each policyholder and speak in absolute rather than probabilistic terms. After the credibility factor has been determined, actuaries usually treat it as a known constant.

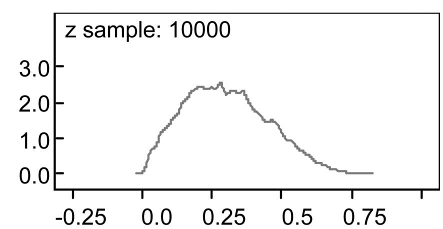
However, the credibility factor estimator itself has a sampling distribution associated with it. That is, a realization of the credibility factor estimator in Equation (12) depends on the sample (or portfolio) drawn from the underlying population. This concept can be seen from Figure 1 and is summarized in Table 3.

Obviously, different simulated portfolios result in different values of the credibility factor. It is known that the impact of variation in the credibility factor estimator diminishes as the amount of experience grows (see Mahler and Dean, Graph 8.6 on page 597, 2001). However, we apply the credibility theory mostly because the data at hand is sparse, and we combine the limited data with other information. This is the concept embedded in Equation (1).

Table 3. Summary statisitcs in 1,000 simulations

(v, a)	Exact k	Exact Z	Simulated Median of $\hat{Z}$	Simulated 95% C.I. of $\hat{Z}$
$(50^2, 20^2)$	6.25	0.4444	0.3565	(0.0000, 0.8553)

Figure 3. Posterior distribution of $\hat{Z}_B$



On the other hand, the Bayesian approach allows us to describe the phenomenon seen in Table 3. From Section 3.1, the Bayesian credibility factor approach suggests the mean of the posterior distribution as the best estimate of the credibility factor, i.e., $\hat{Z}_B = 0.2985$. At the same time, it also suggests that there is possible variation in the estimation process.

For the simulated example in Table 1, the Bayesian credibility factor approach is able to suggest that we are 95% sure that the true credibility factor will fall into the interval from 0.0511 to 0.6026. Additionally, the posterior distribution of $\hat{Z}_B$ for the simulated data in Table 1 can be visualized in Figure 3.

Table 4. Traditional credibility factors and Bayesian credibility factors

Trial	$\hat{Z}$	$SE_{\hat{Z}}$	$\hat{Z}_B$	$SE_{\hat{Z}_B}$	Trial	$\hat{Z}$	$SE_{\hat{Z}}$	$\hat{Z}_B$	$SE_{\hat{Z}_B}$
1	0.51	0.0038	0.54	0.0100	26	0.57	0.0201	0.59	0.0223
2*	0.00	0.1975	0.47	0.0007	27	0.04	0.1620	0.40	0.0024
3	0.74	0.0858	0.71	0.0697	28	0.69	0.0590	0.66	0.0449
4*	0.00	0.1975	0.21	0.0531	29	0.77	0.1064	0.73	0.0843
5	0.50	0.0026	0.54	0.0094	30	0.66	0.0464	0.64	0.0381
6	0.69	0.0585	0.66	0.0476	31	0.19	0.0654	0.43	0.0003
7	0.84	0.1573	0.81	0.1321	32	0.29	0.0253	0.46	0.0002
8	0.42	0.0006	0.51	0.0046	33	0.64	0.0389	0.62	0.0324
9	0.51	0.0039	0.55	0.0109	34*	0.00	0.1975	0.24	0.0400
10	0.53	0.0074	0.56	0.0131	35*	0.00	0.1975	0.33	0.0125
11	0.28	0.0280	0.46	0.0002	36	0.61	0.0268	0.61	0.0274
12	0.67	0.0502	0.65	0.0409	37*	0.00	0.1975	0.51	0.0038
13*	0.00	0.1975	0.35	0.0087	38	0.44	0.0000	0.52	0.0058
14	0.18	0.0706	0.43	0.0003	39	0.39	0.0029	0.50	0.0032
15	0.73	0.0836	0.70	0.0660	40	0.63	0.0349	0.62	0.0296
16*	0.00	0.1975	0.25	0.0379	41*	0.00	0.1975	0.27	0.0321
17*	0.00	0.1975	0.31	0.0192	42	0.80	0.1293	0.75	0.0942
18	0.82	0.1408	0.77	0.1068	43	0.35	0.0084	0.48	0.0012
19	0.63	0.0356	0.62	0.0321	44	0.78	0.1155	0.75	0.0942
20	0.81	0.1328	0.77	0.1085	45*	0.00	0.1975	0.15	0.0853
21*	0.00	0.1975	0.37	0.0048	46	0.55	0.0111	0.57	0.0153
22	0.65	0.0419	0.64	0.0371	47	0.52	0.0059	0.55	0.0101
23	0.38	0.0045	0.49	0.0020	48	0.59	0.0202	0.59	0.0219
24	0.54	0.0101	0.57	0.0150	49*	0.00	0.1975	0.30	0.0213
25*	0.00	0.1975	0.34	0.0099	50	0.23	0.0470	0.44	0.0000
					Mean		0.0882		0.0313

*Zero credibility suggested by traditional credibility estimator

4.2. Traditional credibility factor versus Bayesian credibility factor

Now, we want to demonstrate that the Bayesian credibility process can further improve the mean-squared errors in determining the credibility factor and the credibility premium. To see how this improvement comes about, we use the normal-normal model in Section 3 to simulate 50 trials (or portfolios). For each trial, simulated data is generated the same way as data in Table 1.

We first compare the traditional credibility factor $\hat{Z}$ given by Equation (12) to the Bayesian credibility factor $\hat{Z}_B$ suggested by Equation (25) in Appendix A. We use squared error as our criterion. The results are shown in Table 4, in which, $SE_{\hat{Z}}$ is calculated as $(\hat{Z}-0.4444)^2$ and $SE_{\hat{Z}_B}$ is equal to $(\hat{Z}_B-0.4444)^2$. Note that the normal-normal simulation in Section 3 has a true credibility factor of 0.4444. For data in Table 1, we have $SE_{\hat{Z}}=(0-0.4444)^2=0.1975$ and $SE_{\hat{Z}_B}=(0.2985-0.4444)^2=0.0213$. This is shown in trial 49 of Table 4. The interesting part is the significant improvement to mean-squared errors that result from the Bayesian credibility factor.

Our next task is to compare the traditional credibility premium $\hat{P}_i$ provided by Equation (11) to the Bayesian credibility premium $\hat{P}_i^B$, where

$$\hat{P}_i^B=\hat{Z}_i^B\bar{X}_i+(1-\hat{Z}_i^B)\hat{\mu},\qquad i=1,\ldots,r. \tag{20}$$

For each simulated trial, we have $\Theta=(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5)$ which represents the true underlying individual premiums. For example, Θ is given by $(197,201,202,208,207)$ for the simulated data in Table 1. The sum of squared errors for the traditional credibility premium $\hat{P}_i$ is defined as

$$\sum_{i=1}^{5}(\hat{P}_i-\theta_i)^2. \tag{21}$$

The sum of squared errors for the Bayesian credibility premium $\hat{P}_i^B$ is defined as

$$\sum_{i=1}^{5}(\hat{P}_i^B - \theta_i)^2. \qquad (22)$$

For the simulated data in Table 1, we have $\hat{P}_1 = \hat{P}_2 = \hat{P}_3 = \hat{P}_4 = \hat{P}_5 = \bar{X} = 200$, since $\hat{Z} = 0$. Meanwhile, the Bayesian credibility premium is determined as follows.

$$\begin{aligned}\hat{P}_1^B &= 0.2985 * \bar{X}_1 + (1 - 0.2985) * 200 \\ &= 0.2985 * 185.53 + (1 - 0.2985) * 200 \\ &= 195.35.\end{aligned}$$

Similarly, we have $\hat{P}_2^B = 203.17$, $\hat{P}_3^B = 195.87$, $\hat{P}_4^B = 207.41$, and $\hat{P}_5^B = 195.85$. Thus, the sum of squared errors for $\hat{P}_i$, $i = 1,\ldots 5$, is then calculated as

$$\begin{aligned}&(200 - 207)^2 + (200 - 196)^2 + (200 - 205)^2 \\ &\quad + (200 - 215)^2 + (200 - 189)^2 = 466.\end{aligned}$$

And the sum of squared errors for $\hat{P}_i^B$, $i = 1,\ldots,5$, is equal to 386. Table 5 shows results for all 50 trials. We also include the individual sample means $\bar{X}_i$, $i = 1,\ldots r$ as a benchmark. They are known as the maximum likelihood estimates and are always unbiased. We see that the Bayesian credibility premium has the smallest mean-squared error.

4.3. Applications

In this section, we show one application of interval estimation of the credibility factor. Focusing on the simulated data in Table 1, we have $\hat{P}_1^B = 195.35$, $\hat{P}_2^B = 203.17$, $\hat{P}_3^B = 195.87$, $\hat{P}_4^B = 207.41$, and $\hat{P}_5^B = 195.85$. They represent future premiums without considering possible variation caused by the credibility factor.

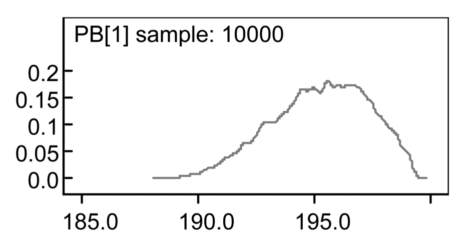
To see the impact of the variation inherent in the estimation of a credibility factor, we use Equations (24) and (26) in Appendix A to determine the posterior distribution of $\hat{P}_1^B$. For illustrative purpose, we only consider the impact

Table 5. A comparison of the sample mean (MLE), the traditional credibility premium, and the Bayesian credibility premium

Trial	$\bar{X}_i$ SSE	$\hat{P}_i$ SSE	$\hat{P}_i^B$ SSE	Trial	$\bar{X}_i$ SSE	$\hat{P}_i$ SSE	$\hat{P}_i^B$ SSE
1	4473	3040	3100	26	3667	2106	2121
2*	4597	4087	3992	27	2542	5335	3643
3	2624	1370	1289	28	4201	1897	1734
4*	830	1286	1042	29	3046	1150	950
5	1839	843	845	30	896	635	654
6	846	262	265	31	3636	3807	3581
7	4321	3089	2871	32	1015	1314	877
8	3881	1686	1877	33	806	1142	1182
9	1643	661	660	34*	1008	1614	1270
10	1873	302	343	35*	3051	684	1004
11	2472	2786	2314	36	1634	911	912
12	2338	944	908	37*	1566	1392	952
13*	1601	330	453	38	2086	790	777
14	2346	1699	1448	39	1559	1133	1116
15	3352	1523	1383	40	3206	1198	1146
16*	1150	745	628	41*	1534	1428	1129
17*	5459	5191	4969	42	3415	1674	1338
18	4695	2727	2341	43	1166	268	332
19	4648	2532	2496	44	2252	1265	1196
20	3827	2299	2082	45*	198	1106	884
21*	3217	3821	3103	46	4258	2290	2337
22	3472	2958	2957	47	2003	697	718
23	1938	618	654	48	8258	5856	5879
24	711	251	244	49*	1142	466	386
25*	1958	160	351	50	1818	1340	1172
				Mean	2601	1734	1598

*Zero credibility suggested by traditional credibility estimator

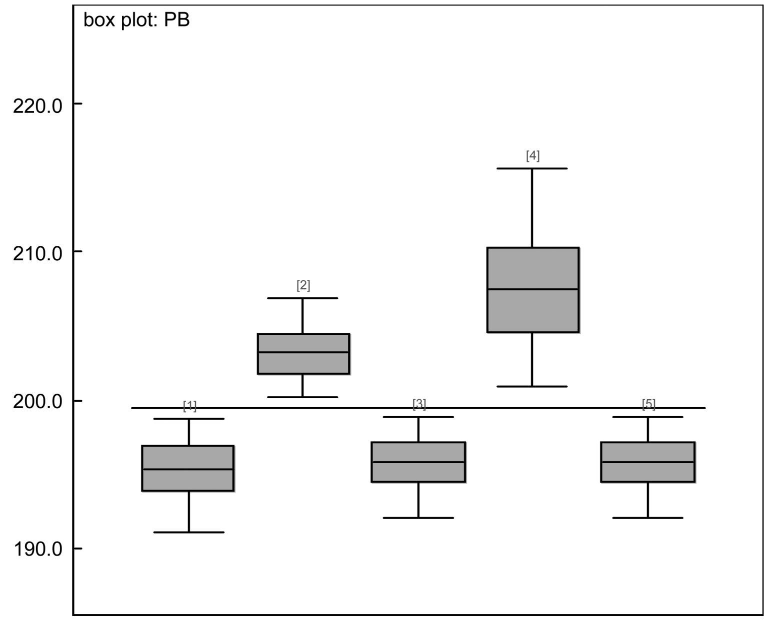
Figure 4. Variation of premium caused by variation of credibility factor



of the variation of the credibility factor on future premiums. Thus, μ_i and μ in Equation (24) are replaced by their maximum likelihood estimates $\bar{X}_1 = 185.53$ and $\bar{X} = 200$, respectively. Figure 4 shows the posterior distribution of $\hat{P}_1^B$. As we can see, there is variation in the premium caused by variation in the credibility factor.

Figure 5 shows the impact of variation in the credibility factor to future premiums for policyholders considered in Table 1. The horizon-

Figure 5. Box plot for $\hat{P}_i^B$, $i = 1,\ldots,5$ with consideration of the variation in the credibility factor



tal line at 200 represents the value of $\bar{X}$. The box plots represent posterior distributions of $\hat{P}_i^B$, $i = 1,\ldots,5$. It gives us a good visual representation of the variability of the premium caused by variation in the credibility factor. The median is shown as a horizontal line within the box. With this information, actuaries can decide whether or not to take an action on the volatility caused by the credibility factor estimation process.

4.4. Disadvantage

The disadvantage of using the proposed Bayesian credibility factor approach shown in Appendices A and B is that we drop the subtraction term in Equation (30). The impact of this action is that we systematically shift the prior distribution of $\hat{a}$ to the right by the amount of $\hat{v}/n$. However, as seen in Table 4, the resulting Bayesian credibility factors are not materially different from those traditional credibility factors with non-zero values. We believe that it is worthwhile to be biased in the prior distribution of $\hat{a}$.

Alternatively, we can also use Equations (28) and (33) for a non-negative $\hat{a}$ to have a more precise prior distribution. Table 6 shows a small sample of trials with this adjustment. We only list trials with relatively large estimates in Table 6, because they contribute relatively large variation in the total squared errors.

Table 6. Bayesian credibility factor using Equations (29) and (34)

Trial	$\hat{Z}$	$\hat{Z}_B$	Adjust $\hat{Z}_B$
3	0.74	0.71	0.59
6	0.69	0.66	0.58
7	0.84	0.81	0.79
12	0.67	0.65	0.55
15	0.73	0.70	0.64
18	0.82	0.77	0.76
20	0.81	0.77	0.75
28	0.69	0.66	0.58
29	0.77	0.73	0.69
42	0.80	0.75	0.72
44	0.78	0.75	0.71

5. Remarks

In this article, we have attempted to explore a range of concerns that arise in credibility modeling. As in any statistical estimation problem, the goal is to estimate the value of an unknown quantity, such as a credibility factor or a future premium in the actuarial field. There are different techniques being used for estimation of a credibility factor or a future premium. Nevertheless, the objectives remain the same. We want to use the sample information to estimate the parameters of interest and to assess the reliability of the estimate.

The variability of a credibility estimator $\hat{Z}$ can cause misclassification of and inaccuracy in the future premium. The benefit of the Bayesian credibility factor approach is that it provides the decision maker an interval estimate of a credibility factor for assessing the accuracy of its point estimator, while the traditional credibility factor approach only provides a point estimator.

Additionally, we use an empirical approach to determine priors for v and a shown in Appendix B. As described earlier, Bayesian statistical methods not only incorporate available prior information either from experts or previous data, but allow the knowledge in these and subsequent data to accumulate in the determination of the parameter values. To avoid negative estimates of a, we replace Equation (28) with Equation (34). Alternatively, one might consider shrinking Equation (28) for the same purpose. That is, we can consider the shrinkage estimates where we use Equation (28) if it is positive. If Equation (28) is negative, we could replace it with 1/2 of the value from Equation (34). However, it is difficult to say if this will work without further investigation.

When credibility theory is applied to ratemaking, $\hat{v}$ and $\hat{a}$, given by Equations (16) and (17), respectively, contain the least information available to the actuary. Even if the components of the prior distributions are not set at their optimal values, the Bayesian credibility factor and the Bayesian credibility premium are still likely to produce better results. Overall, the methodology is easy and straightforward. We believe that this model is a good alternative to credibility modeling.

Acknowledgments

The authors gratefully acknowledge the referees for their extensive comments and suggestions that led to significant improvement of the paper. The authors would also like to thank the editors for their help throughout the review process.

References

Bailey, A., "Credibility Procedures," *Proceedings of the Casualty Actuarial Society* 37, 1950, pp. 7–23, 94–115.

Box, G. E. P., and D. R. Cox, "An Analysis of Transformations," *Journal of the Royal Statistical Society: Series B*, 26, 1964, pp. 211–252.

Box, G. E. P., and P. W. Tidwell, "Transformation of the Independent Variables," *Technometrics* 4, 1962, pp. 531–550.

Bühlmann, H., "Experience Rating and Credibility," *ASTIN Bulletin* 4, 1967, pp. 199–207.

Bühlmann, H., and E. Straub, "Glaubwürdigkeit für Schadensätze (Credibility for Loss Ratio)," *Mitteilungen der Vereinigung Schweizerischer Versicherungs-Mathematiker* 70, 1970, pp. 111–133.

Carlin, B. P., "State Space Modeling of Non-standard Actuarial Time Series," *Insurance, Mathematics, and Economics* 11, 1992, pp. 209–222.

Dannenburg, D. R., R. Kaas, and M. J. Goovaerts, *Practical Actuarial Credibility Models*, University of Amsterdam, The Netherlands: Institute of Actuarial Science and Econometrics, 1996.

de Alba, E., "Bayesian Estimation of Outstanding Claim Reserves," *North American Actuarial Journal* 6:4, 2002, pp. 1–20.

Frees, E. W., "Multivariate Credibility For Aggregate Loss Models," *North American Actuarial Journal* 7:1, 2003, pp. 13–37.

Gangopadhyay, A. and W. Gau, "Credibility Modeling via Spline Nonparametric Regression," *Casualty Actuarial Society Forum*, Winter 2003, pp. 215–252.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Boca Raton, FL: Chapman and Hall, 1998.

Herzog, T. N., *Introduction to Credibility Theory*, Winstead, CT: ACTEX, 1994.

Klugman, S. A., *Bayesian Statistics in Actuarial Science with Emphasis on Credibility*, Boston: Kluwer, 1992.

Mahler, H. C., and C. G. Dean, "Credibility," in *Foundations of Casualty Actuarial Science* (4th ed.), Arlington, VA: Casualty Actuarial Society, 2001, pp. 485–659.

Ntzoufras, I., and P. Dellaportas, "Bayesian Modeling of Outstanding Liabilities Incorporating Claim Count Uncertainty," *North American Actuarial Journal* 6:1, 2002, pp. 113–136.

Rosenberg, M. A., and V. R. Young, "A Bayesian Approach to Understanding Time Series Data," *North American Actuarial Journal* 3:2, 1999, pp. 130–143.

Scollnik, D. P. M., "Actuarial Modeling with MCMC and BUGS," *North American Actuarial Journal* 5:2, 2001, pp. 96–124.

Searle, S., G. Casella, and C. E. McCulloch, *Variance Components*, New York: Wiley, 1992.

Smith, A. F. M., and G. O. Roberts, "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society: Series B*, 55, 1993, pp. 3–23.

Spiegelhalter, D. J., A. Thomas, N. G. Best, and W. R. Gilks, *BUGS 0.5: Bayesian Inference using Gibbs Sampling Manual (Version II)*, Cambridge: MRC Biostatistics Unit, 1996.

Appendix A. Bayesian credibility factor in actuarial applications

For the hierarchical setting in Section 3.1, let

$$V = \{\mu, \alpha_1, \ldots \alpha_r, \upsilon, a\}.$$

We use the notation $V\backslash\nu$ to represent all other parameters except ν in the parameter set V. Let $p(\nu \mid V\backslash\nu)$ be the full conditional distribution of $\nu \mid V\backslash\nu$. For a detailed example of the derivation of the full conditional distribution, the reader is referred to Scollnik (2001). Therefore, we have the following sampling scheme.

1. Initialize μ, $\alpha_1, \ldots, \alpha_r$, υ, and a.
2. Generate μ from

$$p(\mu \mid V\backslash\mu) \propto \prod_{i=1}^{r}\prod_{j=1}^{n_i} \exp\left[-\frac{1}{2\upsilon}(x_{ij}-\mu-\alpha_i)^2\right].$$

3. Generate α_i from

$$p(\alpha_i \mid V\backslash\alpha_i) \propto \exp\left(-\frac{1}{2a}\alpha_i^2\right) \times \exp\left[-\frac{1}{2\upsilon}\sum_{j=1}^{n_i}(x_{ij}-\mu-\alpha_i)^2\right],$$

for $\quad i = 1, \ldots, r$.

4. Generate a from

$$p(a \mid V\backslash a) \propto \frac{1}{a^{(r/2)+1}} \exp\left(-\frac{1}{2a}\sum_{i=1}^{r}\alpha_i^2\right).$$

5. Generate υ from

$$p(\upsilon \mid V\backslash\upsilon) \propto \frac{1}{\upsilon^{(N/2)+1}} \times \exp\left[-\frac{1}{2\upsilon}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij}-\mu-\alpha_i)^2\right],$$

where $N = \sum_{i=1}^{r} n_i$.

6. Set the credibility factor as

$$Z_i^B = \frac{m_i}{m_i + \dfrac{\upsilon}{a}}, \qquad i = 1, \ldots, r. \tag{23}$$

7. The credibility premium is defined as

$$P_i^B = Z_i^B \times \mu_i + (1 - Z_i^B) \times \mu, \qquad i = 1, \ldots, r. \tag{24}$$

In the sampling scheme, steps 1 through 7 are repeated. For the hierarchical setting introduced in Section 3.1, μ is estimated by $\hat{\mu}$ in Equation (15). The Gibbs sampling is carried out in two stages. The first stage is a burn-in period. At the end of this stage, it is assumed the iterations have converged to draws from the posterior distribution. The second stage is called the sampling stage, which is used to estimate the posterior means. After a sufficiently long burn-in period, we have a sample of the credibility factor $Z_{i,l}^B$, $l = 1, \ldots, N$. Similarly, we have a sample of the credibility premium P_i^l, $l = 1, \ldots, N$.

Thus, the Bayesian credibility factor is given by

$$\hat{Z}_i^B = \frac{\sum_{l=1}^{N} Z_{i,l}^B}{N}, \qquad i = 1, \ldots, r, \tag{25}$$

and the Bayesian credibility premium is given by

$$\hat{P}_i^B = \frac{\sum_{l=1}^{N} P_{i,l}}{N}, \qquad i = 1, \ldots, r. \tag{26}$$

For the simulated data in Table 1, the MCMC sampling scheme described above can be implemented in WinBUGS as below. The code itself is fairly self-explanatory. For a comprehensive overview of how to implement actuarial models with MCMC and WinBUGS, the reader is referred to Scollnik (2001), in which monitoring and assessing the convergence of the MCMC simulation is also clearly addressed.

```
CODE for SIMULATED DATA in TABLE 1
model;
{
### Step 1 ###
  for( i in 1 : 5 ) {
    for( j in 1 : 5 ) {
      y[i , j] ~dnorm(theta[i],tau.v)
    }
  }
for( i in 1 : 5 ) {
    theta[i] ¡- mu.hat+alpha[i]
  }
### Step 2 ###
  mu.hat ¡- mean(y[,])
### Step 3 ###
  for( i in 1 : 5 ) {
    alpha[i] ~dnorm(0, tau.a)
  }
### Step 4 ###
  a ~dgamma(2,0.0059)
  tau.a ¡- 1 / a
### Step 5 ###
  v ~dgamma(10,0.0037)
  tau.v ¡- 1 / v
### Step 6 ###
  k ¡- v / a
  z ¡- 5 / (5 + k)
### Step 7 ###
  for( i in 1 : 5 ) {
    PB[i] ¡- z*mean(y[i,])+(1-z)* mean(y[,])
  }
}
```

Appendix B. Empirical Bayesian approach to priors

The prior distribution usually presents a summary of external evidence about the quantities of interest. To be useful in public discussion, such a prior must be evidence-based in some sense. Here, we use an empirical approach to determine priors for υ and a. Recall in Equations (16) and (17), we have

$$\hat{\upsilon} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X}_i)^2}{\sum_{i=1}^{r}(n_i-1)} \tag{27}$$

and

$$\hat{a} = \frac{\sum_{i=1}^{r} m_i(\bar{X}_i-\bar{X})^2-(r-1)\hat{\upsilon}}{m-\dfrac{\sum_{i=1}^{r} m_i^2}{m}}. \tag{28}$$

For the basic model, $m_{ij}=1$ for $1\le i\le r, 1\le j\le n$, Equation (5) provides a balanced one-way variance component model. Hence, we have

$$\hat{\upsilon} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n}(X_{ij}-\bar{X}_i)^2}{r(n-1)} \tag{29}$$

and

$$\hat{a} = \frac{\sum_{i=1}^{r}(\bar{X}_i-\bar{X})^2}{r-1}-\frac{\hat{\upsilon}}{n}. \tag{30}$$

To obtain empirical prior distributions on these parameters, we need to estimate the moments of these estimators (i.e., method of moments). Since $E(\hat{\upsilon})=\upsilon$ and $E(\hat{a})=a$, estimates themselves can be used as estimates of the first moments. Meanwhile, we have $r(n-1)\hat{\upsilon}/\upsilon$ coming from a χ^2-distribution with $r(n-1)$ degrees of freedom. Thus, we can obtain the estimated variance of $\hat{\upsilon}$ as

$$\hat{\mathrm{Var}}(\hat{\upsilon}) = \frac{2\hat{\upsilon}_{\mathrm{init}}^2}{\sum_{i=1}^{r}(n-1)}, \tag{31}$$

where $\hat{\upsilon}_{\mathrm{init}}^2$ is the initial value obtained from Equation (29).

As for $\hat{a}$, there is no identifiable distribution for this estimator. Nevertheless, the variance of $\hat{a}$ can be derived as

$$\mathrm{Var}(\hat{a}) = \frac{2}{r-1}\left(a+\frac{\upsilon}{n}\right)^2+\frac{2}{(m-r)n^2}\upsilon^2. \tag{32}$$

An unbiased estimator of this variance is given by

$$\hat{\mathrm{Var}}(\hat{a}) = \frac{2}{r+1}\left[\frac{\sum_{i=1}^{r}(\bar{X}_i-\bar{X})^2}{r-1}\right]^2+\frac{2}{(m-r)n^2}\hat{\upsilon}^2. \tag{33}$$

Then, parameters of the prior distribution $\pi_2(a)$ can be determined by matching its mean and variance with $\hat{a}$ and Equation (33) respectively.

However, this approach does not eliminate the issue of negative values for $\hat{a}$. Thus, we use an approximation for the variance of $\hat{a}$. To avoid the negativity in the estimation of a, we drop the subtraction term in Equation (30), i.e.,

$$\hat{a'} = \frac{\sum_{i=1}^{r}(\bar{X}_i - \bar{X})^2}{r-1}. \tag{34}$$

Note that $\hat{a}$ converges in distribution to $\hat{a'}$ as $n \to \infty$. Suppose $E(\hat{a'}) = a'$. Then, $(r-1)\hat{a'}/a'$ has a χ^2-distribution with $(r-1)$ degrees of freedom. Thus, we obtain the approximated variances of $\hat{a}$ by

$$\hat{\mathrm{Var}}(\hat{a}) = \frac{2\hat{a'}^2_{\mathrm{init}}}{r-1} \tag{35}$$

where $\hat{a'}^2_{\mathrm{init}}$ is the initial value obtained by Equation (34).

For example, if we adopt Gamma(α, β) as our prior distribution, we can quickly determine α and β by matching the moments. The equations to find parameters for the prior distribution of v are given by

$$\alpha_v/\beta_v = \hat{v}_{\mathrm{init}}, \qquad \alpha_v/\beta_v^2 = \frac{2\hat{v}^2_{\mathrm{init}}}{\sum_{i=1}^{r}(n-1)} \tag{36}$$

with solutions,

$$\alpha_v = \frac{\sum_{i=1}^{r}(n-1)}{2} \tag{37}$$

and

$$\beta_v = \frac{\alpha_v}{\hat{v}_{\mathrm{init}}}. \tag{38}$$

Similarly, parameters with ad hoc approximation for the prior distribution of a are given by

$$\alpha_a = \frac{r-1}{2} \tag{39}$$

and

$$\beta_a = \frac{\alpha_a}{\hat{a'}_{\mathrm{init}}}. \tag{40}$$

In the Bühlmann-Straub setting, the data are not balanced, however, and there are unequal numbers of observations in the subclasses. The mean squares are no longer independent and they do not have χ^2-distributions. The variance of weighted sums of squares needs to be approximated and usually is much more complicated than balanced data. We recommend simply using Equations (27) and (34) to match the first moment of the prior distributions.