Bayesian Analysis of Insurance Losses Using the Buhlmann-Straub Credibility Model

Abraham J. van der Merwe
University of the Free State, Republic of South Africa, matheeme.sci@mail.uovs.ac.za

Kobus N. Bekker
University of the Free State, South Africa, bekkerkn.sci@mail.uovs.ac.za

Bayesian Analysis of Insurance Losses Using the Bühlmann-Straub Credibility Model

Abraham J. van der Merwe* and Kobus N. Bekker†

Abstract‡

We propose a Bayesian analysis to develop credibility estimates of the well-known Bühlmann-Straub model. We describe simple numerical methods to obtain exact posterior distributions and predictive densities under this model. These distributions are obtained through Monte Carlo simulations that generate independent samples from the joint posterior distribution. Our methods are therefore preferable to methods such as Gibbs sampling, which generates dependent samples from the joint distribution. The methods discussed also can be extended to more complicated credibility models.

Key words and phrases: *Bayesian procedure, credibility theory, Monte Carlo simulation, probability-matching prior, reference priors*

*Abraham J. van der Merwe, Ph.D., is a professor of statistics at the University of the Free State, Republic of South Africa, and is a fellow of the South African Statistical Association. His research interests are in Bayesian applications in mixed linear models, Bayesian calibration, and non-parametric Bayesian estimation.

Dr. van der Merwe's address is: Department of Mathematical Statistics, University of the Free State, PO Box 339, Bloemfontein 9300, SOUTH AFRICA. E-mail: *matheeme.sci@mail.uovs.ac.za*

†Kobus N. Bekker, M.Sc., is a lecturer of statistics and actuarial science at University of the Free State, South Africa, and is pursuing his Ph.D. His dissertation is in the areas of unit-linked products and embedded guarantees; his other research interests include credibility theory, extreme value theory, and Bayesian statistics.

Mr. Bekker's address is: Department of Mathematical Statistics, University of the Free State, PO Box 339, Bloemfontein 9300, SOUTH AFRICA. E-mail: *bekkerkn.sci@mail.uovs.ac.za*

‡The authors are grateful to the editor and referees for their helpful comments that led to a substantial improvement of this paper.

1 Introduction

Let us consider a portfolio of I insurance risks where the i^{th} insured risk is characterized by an unobservable random time homogenous risk parameter Φ_i that influences the occurrence and size of the losses or claims stemming from risk i. We assume that the i^{th} insured is observed for J_i periods (a period may be a month, quarter, year, etc.) and the data consist of the observations Y_{ij} and p_{ij} for risk i in period j, $j = 1, 2, \ldots, J_i$ and for $i = 1, 2, \ldots, I$. The Y_{ij}'s and the p_{ij}'s reflect the i^{th} insured's claims experience (such as average claim amount or claim loss-ratio) and the weight (also called the risk volume), respectively, in period j. In principle, these weights should reflect the total exposure of each insured risk such as the number of claims in one year or the premium volume. A key consideration in the choice of Y_{ij} is that its conditional variance must be inversely proportional to the weight p_{ij}. Following Goulet (1998) we depict the insurance portfolio as in Table 1.

Given the data in Table 1, the insurer's problem is to determine the correct (or credibility) premium to charge each insured risk for period $j + 1$.[1] To determine the correct premium, we will use the well-known Bühlmann–Straub credibility model (Bühlmann and Straub, 1970). The assumptions of the Bühlmann–Straub model are as follows:

(B-S1) $\mathbb{E}\left[Y_{ij}|\Phi_i\right] = \mu(\Phi_i)$, is independent of j (i.e., time invariant);

(B-S2) The vectors $(Y_{i1}, \ldots, Y_{iJ_i}, \Phi_i)$, $i = 1, \ldots, I$ are mutually independent with finite covariance matrix;

(B-S3) The risk parameters $\Phi_1, \ldots, \Phi_I$ are independent and identically distributed; and

(B-S4) Given Φ_i, the i^{th} insured's claims experience is uncorrelated across periods:

$$\mathbb{C}\text{ov}(Y_{ij}, Y_{ik}|\Phi_i) = \begin{cases} \frac{\sigma^2(\Phi_i)}{p_{ij}} & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, \ldots, I$.

[1] This is a standard problem in actuarial credibility theory. There are several approaches to determining this premium using credibility; see, for example, Gerber (1982), Waters (1987), Makov, Smith, and Liu (1996), Dannenburg, Kaas, and Goovaerts (1996), Goulet (1998), and Norberg (2004, pages 398–406) and references therein.

Under the Bühlmann-Straub model, the estimator of the premium is restricted to the class of linear Bayesian estimators. This restriction leads to a credibility premium for the i^{th} insured, $\mathcal{P}_i$, which is given by

$$\mathcal{P}_i = \tilde{Z}_i \overline{Y}_{i.} + (1 - \tilde{Z}_i)\overline{Y}_{..} \tag{1}$$

where $0 \leq \tilde{Z}_i \leq 1$ is the credibility factor, $\overline{Y}_{i.} = \sum_j Y_{ij}/J_i$ is the average claim of the i^{th} insured, and $\overline{Y}_{..} = \sum_i J_i \overline{Y}_{i.} / \sum_i J_i$ is the sample collective mean (based on all of the data).

Table 1
The Basic Insurance Portfolio

Insured Risk	Risk Level	Periodic Observations			Weights		
1	Φ_1	Y_{11}	$\ldots$	Y_{1J_1}	p_{11}	$\ldots$	p_{1J_1}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	Φ_i	Y_{i1}	$\ldots$	Y_{iJ_i}	p_{i1}	$\ldots$	p_{iJ_i}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
I	Φ_I	Y_{I1}	$\ldots$	Y_{IJ_I}	p_{I1}	$\ldots$	p_{IJ_I}

1.1 Our Objectives

We will show that, by using a full Bayesian approach, the credibility premium corresponds to the mean of the posterior distribution of the portfolio's claims. Recall that in the Bühlmann-Straub model the number of periods of experience may be unbalanced across insureds, i.e., J_i depends on i. As a simplification, however, we will consider only the case of a *balanced* claims experience where $J_i = J$ for $i = 1, 2, \ldots, I$. To simplify matters, we assume that the risk level is such that $\mu(\Phi_i) = m + u_i$ and for $i = 1, \ldots, I$ and $j = 1, \ldots, J$,

$$Y_{ij} = m + u_i + \varepsilon_{ij} \tag{2}$$

where m is a global or collective risk level, u_i is a random parameter, and ε_{ij} is a random error term. The random variables u_i and ε_{ij} are sometimes called random effects.[2] We assume the random effects are normally distributed,[3] i.e., $\varepsilon_{ij} \sim \mathrm{N}(0, p_{ij}^{-1}\sigma_1^2)$ and $u_i \sim \mathrm{N}(0, \delta\sigma_1^2)$.

$$\varepsilon_{ij} \sim \mathrm{N}(0, p_{ij}^{-1}\sigma_1^2) \tag{3}$$
$$u_i \sim \mathrm{N}(0, \delta\sigma_1^2). \tag{4}$$

The credibility model as defined in equation (2) can be written in matrix notation as follows

$$\boldsymbol{Y} = m\mathbf{1} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\varepsilon} \tag{5}$$

where

$$\begin{aligned}
\boldsymbol{Y} &= (Y_{11}, \dots, Y_{1J}, Y_{21}, \dots, Y_{2J}, \dots, Y_{I1}, \dots, Y_{IJ})^{\mathsf{T}} \qquad (IJ \times 1)\\
\mathbf{1} &= (1, 1, \dots, 1)^{\mathsf{T}}, \qquad (IJ \times 1)\\
\boldsymbol{u} &= (u_1, u_2, \dots u_I)^{\mathsf{T}}, \qquad (I \times 1)\\
\boldsymbol{\varepsilon} &= (\varepsilon_{11}, \dots, \varepsilon_{1J}, \varepsilon_{21}, \dots, \varepsilon_{2J}, \dots, \varepsilon_{I1}, \dots \varepsilon_{IJ})^{\mathsf{T}}, \quad (IJ \times 1)
\end{aligned}$$

are column vectors (T denotes transpose) and $\boldsymbol{Z}$ is an $IJ \times I$ matrix of 0s and 1s with the ones indicating the insured risk. Also, $\boldsymbol{u}$ and $\boldsymbol{\varepsilon}$ are multivariate normal with $\boldsymbol{u} \sim \mathrm{N}(\mathbf{0}, \delta\sigma_1^2\tilde{\boldsymbol{I}})$ and $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{P}^{-1}\sigma_1^2)$ where $\tilde{\boldsymbol{I}}$ is the identity matrix and $\boldsymbol{P}$ is a diagonal matrix of weights, i.e.,

$$\boldsymbol{P} = \mathrm{Diag}\,(p_{11}, \dots, p_{1J}, p_{21}, \dots, p_{2J}, \dots, p_{I1}, \dots, p_{IJ}], \qquad (IJ \times IJ).$$

For the credibility model of equation (5), the distribution of the data given $m, \boldsymbol{u}$, and σ_1^2 can be written in matrix notation as

$$\begin{aligned}
p\left(\boldsymbol{Y}|m, \boldsymbol{u}, \sigma_1^2\right) &= (2\pi)^{-\frac{1}{2}IJ}\,|P|^{\frac{1}{2}}\left(\frac{1}{\sigma_1^2}\right)^{\frac{1}{2}IJ}\\
&\times \exp\left\{-\frac{1}{2\sigma_1^2}(\boldsymbol{Y} - m\mathbf{1} - \boldsymbol{Z}\boldsymbol{u})^{\mathsf{T}}\boldsymbol{P}(\boldsymbol{Y} - m\mathbf{1} - \boldsymbol{Z}\boldsymbol{u})\right\}.
\end{aligned} \tag{6}$$

[2]The model represented by equation (2) is known in field of the analysis of variance as a one-way random effects model. For more on analysis of variance see, for example, Scheffé (1959) for a classical approach and Box and Tiao (1973) for a Bayesian approach.

[3]Klugman (1992) gives a few arguments supporting the normal assumption: (i) analysis is often done on loss ratios, not losses themselves, so that the class-to-class deviations may well be symmetrically distributed; (ii) the normal distribution is easy to work with even when the model includes dependent observations; and (iii) (this is the most compelling argument) the Bayes solution and the credibility (linear Bayes) solution are identical.

Given the data, $p\left(Y|m,\boldsymbol{u},\sigma_{\mathrm{I}}^{2}\right)$ may be regarded as a function of $m,\boldsymbol{u}$, and σ_{I}^{2} and not of $\boldsymbol{Y}$. When so regarded, following Box and Tiao (1973) $p\left(\boldsymbol{Y}|m,\boldsymbol{u},\sigma_{\mathrm{I}}^{2}\right)$ is called the likelihood function of $m,\boldsymbol{u}$, and σ_{I}^{2} and is written as $L\left(m,\boldsymbol{u},\sigma_{\mathrm{I}}^{2}|\boldsymbol{Y}\right)$. The integrated likelihood function is the following I dimensional integral:

$$L\left(m,\sigma_{\mathrm{I}}^{2},\delta|\boldsymbol{Y}\right)=\int_{\mathbb{R}^{I}}L\left(m,\boldsymbol{u},\sigma_{\mathrm{I}}^{2}|\boldsymbol{Y}\right)\left(2\pi\delta\sigma_{\mathrm{I}}^{2}\right)^{-\frac{1}{2}I}\exp\left\{-\frac{1}{2\delta\sigma_{\mathrm{I}}^{2}}\boldsymbol{u}^{\mathsf{T}}\boldsymbol{u}\right\}d\boldsymbol{u},$$

which reduces to

$$\begin{aligned}L(m,\sigma_{\mathrm{I}}^{2},\delta\,\big|\,Y)\propto\left(\tfrac{1}{\sigma_{\mathrm{I}}^{2}}\right)^{\frac{1}{2}IJ}\times\prod_{i=1}^{I}\left(\tfrac{1}{1+p_{i.}\delta}\right)^{\frac{1}{2}}\\ \times\exp\left\{-\tfrac{1}{2\sigma_{\mathrm{I}}^{2}}\left[\sum_{i=1}^{I}\sum_{j=1}^{J}p_{ij}(Y_{ij}-\overline{Y}_{i.})^{2}+\sum_{i=1}^{I}(\overline{Y}_{i.}-m)^{2}\tfrac{p_{i.}}{1+p_{i.}\delta}\right]\right\}\end{aligned}\tag{7}$$

where $p_{i.}=\sum_{j=1}^{J}p_{ij}$ and $\overline{Y}_{i.}=\sum_{j=1}^{J}p_{ij}Y_{ij}/p_{i.}$. The proof of equation (7) is given in van der Merwe and Bekker (2004).

2 Prior and Posterior Distributions

One of the main advantages of the Bayesian approach over the classical statistical approach is that it allows for explicit use of the statistician's prior information on each parameter of interest, thereby giving new insights in problems where classical statistics may fail. In the Bayesian framework, a prior must be specified even when the statistician has no actual prior information. Determination of reasonable non-informative priors in multi-parameter problems is not easy. Common non-informative priors, such as Jeffreys' prior,[4] can have features that have an unexpectedly dramatic effect on the posterior. In recognition of this problem, Berger and Bernardo (1992) proposed using so-called reference priors to develop non-informative priors, while Tibshirani (1989) and Datta and Ghosh (1995) have proposed using so-called probability-matching priors for this purpose. A key feature of Berger and Bernardo's approach is that it permits the reference prior to depend on the parameters of interest and on nuisance parameters.

[4]The Jeffreys' prior is the square root of the determinant of the Fisher information matrix.

2.1 Reference Priors

Suppose the data $\boldsymbol{Y}$ depends on a $k \times 1$ vector of unknown parameters $\boldsymbol{\theta}$. The reference prior method is motivated by the notion of maximizing the expected amount of information about $\boldsymbol{\theta}$ provided by the data $\boldsymbol{Y}$. The expectation is $\mathbb{E}\left[D\left(p\left(\boldsymbol{\theta}|\boldsymbol{Y}\right), p\left(\boldsymbol{\theta}\right)\right)\right]$ where

$$D\left(p\left(\boldsymbol{\theta}|\boldsymbol{Y}\right), p\left(\boldsymbol{\theta}\right)\right) = \int_{\boldsymbol{\theta}} p\left(\boldsymbol{\theta}|\boldsymbol{Y}\right) \log\left(\frac{p\left(\boldsymbol{\theta}|\boldsymbol{Y}\right)}{p\left(\boldsymbol{\theta}\right)}\right) d\boldsymbol{\theta}$$

is the Kullback-Liebler divergence.

The actual reference prior method stems from a modification of the notion of maximizing the expected information provided by the data. Berger and Bernardo (1992) define $\boldsymbol{Z}_t = (\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_t)$ to be a vector containing data from t replications of an experiment. The first step in the reference prior method is to choose a prior distribution to maximize $\mathbb{E}\left[D\left(p\left(\boldsymbol{\theta}|\boldsymbol{Z}_t\right), p\left(\boldsymbol{\theta}\right)\right)\right]$ for each t. The reference prior is then given as the limit of these priors. The algorithm for generating reference priors is described by Berger and Bernardo (1992) and Robert (2001). Only some of the features of the algorithm are described below.

Step 1: Assume that the Fisher information matrix for $\boldsymbol{\theta}$, $F\left(\boldsymbol{\theta}\right)$, exists and is of full rank. Denote $S = F^{-1}\left(\boldsymbol{\theta}\right)$.

Step 2: Separate the parameters into r groups of sizes $n_1, n_2, ..., n_r$ that correspond to their decreasing levels of importance, i.e., $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_{(1)} \vdots \boldsymbol{\theta}_{(2)} \vdots \cdots \vdots \boldsymbol{\theta}_{(r)}\right)$ where $\boldsymbol{\theta}_{(1)} = \left(\theta_1, ..., \theta_{N_1}\right)$, $\boldsymbol{\theta}_{(2)} = \left(\theta_{N_1+1}, ..., \theta_{N_2}\right)$, ..., and $\boldsymbol{\theta}_{(r)} = \left(\theta_{N_{r-1}+1}, ..., \theta_k\right)$ with $N_i = \sum_{j=1}^{i} n_j$ for $j = 1, ..., r$. Note that $\boldsymbol{\theta}_{(1)}$ is the most important and $\boldsymbol{\theta}_{(r)}$ is the least.

Step 3: Define, for $j = 1, ..., r$, $\boldsymbol{\theta}_{[j]} = \left(\boldsymbol{\theta}_{(1)}, ..., \boldsymbol{\theta}_{(j)}\right)$ and $\boldsymbol{\theta}^{[j]} = \left(\boldsymbol{\theta}_{(j+1)}, ..., \boldsymbol{\theta}_{(r)}\right)$ so that $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_{[j]} \vdots \boldsymbol{\theta}^{[j]}\right)$.

Step 4: Decompose the matrix S according to the r groups of sizes $n_1, n_2, ..., n_r$, i.e.,

$$S = \begin{bmatrix} A_{11} & A_{21}^{\mathsf{T}} & \cdots & A_{r1}^{\mathsf{T}} \\ A_{21} & A_{22} & \cdots & A_{r2}^{\mathsf{T}} \\ \vdots & \vdots & & \vdots \\ A_{r1} & A_{r2} & \cdots & A_{rr} \end{bmatrix}$$

where A_{ij} is an $\left(n_i \times n_j\right)$ matrix;

Step 5: Define S_j as the $\left(N_j \times N_j\right)$ matrix consisting of elements from the upper left corner of S with $S_r \equiv S$;

Step 6: Let $H_j \equiv S_j^{-1}$. Then define h_j to be the $\left(n_j \times n_j\right)$ matrix contained in the lower right corner of H_j for $j = 1, \ldots, r$.

Step 7: Define the $\left(n_j \times N_{j-1}\right)$ matrix $B_j = \left[\begin{array}{cccc} A_{j1} & A_{j2} & \cdots & A_{j\,j-1} \end{array}\right]$, for $j = 2, \ldots, r$, of sizes $\left(n_j \times N_{j-1}\right)$.

Step 8: It is straightforward to verify that for $j = 2, \ldots, r$

$$h_j = \left[A_{jj} - B_j H_{j-1} B_j^{\mathsf{T}}\right]^{-1}$$

and

$$H_j = \left[\begin{array}{cc} H_{j-1} + H_{j-1} B_j^{\mathsf{T}} h_j B_j H_{j-1} & -H_{j-1} B_j^{\mathsf{T}} h_j \\ -h_j B_j H_{j-1} & h_j \end{array}\right].$$

Step 9: Iteratively calculate $H_2, \ldots, H_r$, and hence $h_2, \ldots, h_r$ to obtain the ordered reference priors under asymptotic normality.

According to Bernardo (1998), the derivation of the ordered reference prior is greatly simplified if the $h_j(\boldsymbol{\theta})$ terms depend only on $\boldsymbol{\theta}_{[j]}$, and not on $\boldsymbol{\theta}^{[j]}$, then:

$$p^l(\boldsymbol{\theta}) = \prod_{j=1}^{m} \frac{\left|h_j(\boldsymbol{\theta})\right|^{\frac{1}{2}}}{\int \left|h_j(\boldsymbol{\theta})\right|^{\frac{1}{2}} d\boldsymbol{\theta}_{[j]}}.$$

Often some of the integrals appearing in the algorithm are not defined. Berger and Bernardo (1992) then propose to derive the reference prior for compact subsets $\boldsymbol{\theta}^l$ of $\boldsymbol{\theta}$ and to consider the limit of the corresponding reference priors as l tends to infinity and $\boldsymbol{\theta}^l$ tends to $\boldsymbol{\theta}$. In general, the resulting limits do not depend on the choice of sequence of the compact sets.

The Bühlmann-Straub model, where we are concerned with the three parameters m, σ_1^2, and δ, represents a typical situation where reference priors had been shown to be very promising; see Ye (1995) and Yang and Chen (1995). As in the case of the Jeffreys' prior, the reference prior method is derived from the Fisher information matrix. Berger and Bernardo (1992) recommended the reference prior be based on having each parameter in its own group, i.e. having each conditional reference prior be only one-dimensional. The notation $\{\theta_1, \theta_2, \theta_3\}$ means that the parameter θ_1 is most important and θ_3 is the least important.[5] Only the reference prior for the group ordering $\{m, \delta, \sigma_1^2\}$ will be derived. The reference priors for other group orderings can be computed in a similar fashion.

The Fisher information matrix and its inverse for the group ordering $\{m, \delta, \sigma_1^2\}$ are given below for the Bühlmann–Straub model with $\boldsymbol{Y} = m\mathbf{1} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\varepsilon}$ where $\boldsymbol{u} \sim \mathrm{N}(\mathbf{0}, \delta\sigma_1^2\tilde{I})$, $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{P}^{-1}\sigma_1^2)$. (See van der Merwe and Bekker (2004) for the derivation of these matrices.)

$$\boldsymbol{F} = \begin{bmatrix} \frac{1}{\sigma_1^2}\sum_{i=1}^{I}\frac{p_{i.}}{1+p_{i.}\delta} & 0 & 0 \\ 0 & \frac{1}{2}\sum_{i=1}^{I}\frac{p_{i.}^2}{(1+p_{i.}\delta)^2} & \frac{1}{2\sigma_1^2}\sum_{i=1}^{I}\frac{p_{i.}}{1+p_{i.}\delta} \\ 0 & \frac{1}{2\sigma_1^2}\sum_{i=1}^{I}\frac{p_{i.}}{1+p_{i.}\delta} & \frac{IJ}{2}\left(\frac{1}{\sigma_1^2}\right)^2 \end{bmatrix} \tag{8}$$

and its inverse is

$$\boldsymbol{F}^{-1} = \begin{bmatrix} \sigma_1^2\left(\sum_{i=1}^{I}\frac{p_{i.}}{1+p_{i.}\delta}\right)^{-1} & 0 & 0 \\ 0 & \frac{IJ}{2|H|}\left(\frac{1}{\sigma_1^2}\right)^2 & -\frac{1}{2|H|\sigma_1^2}\sum_{i=1}^{I}\frac{p_{i.}}{1+p_{i.}\delta} \\ 0 & -\frac{1}{2|H|\sigma_1^2}\sum_{i=1}^{I}\frac{p_{i.}}{1+p_{i.}\delta} & \frac{1}{2|H|}\sum_{i=1}^{I}\frac{p_{i.}^2}{(1+p_{i.}\delta)^2} \end{bmatrix} \tag{9}$$

where the determinant $|H|$ is equal to

$$|H| = \frac{IJ}{4\left(\sigma_1^2\right)^2}\sum_{i=1}^{I}\frac{p_{i.}^2}{(1+p_{i.}\delta)^2} - \frac{1}{4\left(\sigma_1^2\right)^2}\left(\sum_{i=1}^{I}\frac{p_{i.}}{1+p_{i.}\delta}\right)^2. \tag{10}$$

Here $r = 3$, $n_1 = n_2 = n_3 = 1$, $N_1 = 1$, $N_2 = 2$, and $N_3 = 3$. For $i, j = 1, 2, 3$, let F_{ij} and F^{ij} denote the $(i,j)^{\text{th}}$ element of $\boldsymbol{F}$ and $\boldsymbol{F}^{-1}$ as defined in equations (8) and (9), respectively. The matrices h_j, $j = 1, 2, 3$ are needed to obtain the reference prior. Now,

[5] In this terminology, Jeffreys' prior is also a reference prior, arising when all the parameters are treated as a single group.

$$h_1 = F_{11} - \begin{bmatrix} F_{12} & F_{13} \end{bmatrix} \begin{bmatrix} F_{22} & F_{23} \\ F_{32} & F_{33} \end{bmatrix}^{-1} \begin{bmatrix} F_{21} \\ F_{31} \end{bmatrix}$$
$$= \frac{1}{\sigma_1^2} \sum_{i=1}^{I} \frac{p_{i.}}{1 + p_{i.}\delta}.$$

Further, as

$$H = \begin{bmatrix} F^{11} & F^{12} \\ F^{21} & F^{22} \end{bmatrix}^{-1} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}$$

and

$$H = \begin{bmatrix} \frac{1}{\sigma_1^2} \sum_{i=1}^{I} \frac{p_{i.}}{1+p_{i.}\delta} & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^{I} \frac{p_{i.}^2}{(1+p_{i.}\delta)^2} \end{bmatrix} - \left[\frac{IJ}{2} \left(\frac{1}{\sigma_1^2} \right)^2 \right]^{-1} \begin{bmatrix} 0 \\ \frac{1}{2\sigma_1^2} \sum_{i=1}^{I} \frac{p_{i.}}{1+p_{i.}\delta} \end{bmatrix} \begin{bmatrix} 0 & \frac{1}{2\sigma_1^2} \sum_{i=1}^{I} \frac{p_{i.}}{1+p_{i.}\delta} \end{bmatrix}$$

which means that $H_{22} = F_{22} - \frac{1}{F_{33}} F_{23} F_{32} = h_2$. Therefore

$$h_2 = \frac{1}{2} \sum_{i=1}^{I} \frac{p_{i.}^2}{(1 + p_{i.}\delta)^2} - \frac{1}{2IJ} \left(\sum_{i=1}^{I} \frac{p_{i.}}{1 + p_{i.}\delta} \right)^2 \quad \text{and}$$
$$h_3 = F_{33} = \frac{IJ}{2} \left(\frac{1}{\sigma_1^2} \right)^2.$$

It follows that

$$p(m) \propto h_1^{\frac{1}{2}} = 1 \quad \text{because it does not contain } m;$$

$$p(\delta | m) \propto h_2^{\frac{1}{2}} = \left\{ \sum_{i=1}^{I} \frac{p_{i.}^2}{(1 + p_{i.}\delta)^2} - \frac{1}{IJ} \left(\sum_{i=1}^{I} \frac{p_{i.}}{1 + p_{i.}\delta} \right)^2 \right\}^{\frac{1}{2}}, \quad \text{and}$$

$$p\left(\sigma_1^2 | m, \delta\right) \propto h_3^{\frac{1}{2}} = \frac{1}{\sigma_1^2}.$$

Notice that the $h_j(\boldsymbol{\theta})$ terms depend only on $\boldsymbol{\theta}_{[j]}$, which permits factorization (Bernardo, 1998), and not on $\boldsymbol{\theta}^{[j]}$.

The reference prior for the group ordering $\left\{m, \delta, \sigma_{\mathrm{I}}^2\right\}$ is therefore given by

$$p_{R_1}\left(m, \delta, \sigma_{\mathrm{I}}^2\right) = p(m)\, p(\delta|m)\, p\left(\sigma_{\mathrm{I}}^2|\delta, m\right)$$
$$= \frac{1}{\sigma_{\mathrm{I}}^2}\left\{\sum_{i=1}^{I} \frac{p_{i.}^2}{(1+p_{i.}\delta)^2} - \frac{1}{IJ}\left(\sum_{i=1}^{I} \frac{p_{i.}}{1+p_{i.}\delta}\right)^2\right\}^{\frac{1}{2}}. \quad (11)$$

This prior is independent of the limits of the compact subsets and yields a proper posterior distribution. As will be seen, $p_{R_1}\left(m, \delta, \sigma_{\mathrm{I}}^2\right)$ also satisfies the probability-matching criterion.

It turns out that for the Bühlmann-Straub model, the reference prior for the group orderings $\left\{m, \delta, \sigma_{\mathrm{I}}^2\right\}$, $\left\{\delta, m, \sigma_{\mathrm{I}}^2\right\}$, $\left\{\delta, \sigma_{\mathrm{I}}^2, m\right\}$ is given by equation (11), while for the group orderings $\left\{m, \sigma_{\mathrm{I}}^2, \delta\right\}$, $\left\{\sigma_{\mathrm{I}}^2, m, \delta\right\}$, $\left\{\sigma_{\mathrm{I}}^2, \delta, m\right\}$ the reference prior is given by

$$p_{R_2}(m, \sigma_{\mathrm{I}}^2, \delta) \propto \sigma_{\mathrm{I}}^{-2}\left\{\sum_{i=1}^{I} \frac{p_{i.}^2}{(1+p_{i.}\delta)^2}\right\}^{\frac{1}{2}}. \quad (12)$$

2.2 Probability-Matching Priors

The reference prior algorithm is but one way to obtain useful non-informative priors. Another type of non-informative prior is the probability-matching prior. This prior has good frequentist properties, i.e. properties that hold on the average (in $\boldsymbol{Y}$) rather than conditional on $\boldsymbol{Y}$. Two reasons for using probability-matching priors are that they provide a method for constructing accurate frequentist intervals, and that they could be potentially useful for comparative purposes in a Bayesian analysis. Also, Berger states (in Wolpert, 2004) that frequentist reasoning will play an important role in finally obtaining good general objective priors for model selection. Indeed, some statisticians argue that frequency calculations are an important part of applied Bayesian statistics. [See Rubin (1984) for an example.]

There are two methods for generating probability-matching priors due to Tibshirani (1989) and to Datta and Ghosh (1995). Tibshirani (1989) generated probability-matching priors by transforming the model parameters so that the (single) parameter of interest is orthogonal to the other parameters. The prior distribution is then taken to be proportional to the square root of the upper left element of the information matrix in the new parameterization.

Datta and Ghosh (1995) provided a different solution to the problem of finding probability-matching priors. They derived the differential equation that a prior must satisfy if the posterior probability of a one-sided credibility interval for a parametric function and its frequentist probability agree up to $O(n^{-1})$ where n is the sample size.

The exact definition of Datta and Ghosh (1995) is as follows: Suppose $Y_1, ..., Y_n$ are independently and identically distributed with density $f(y, \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \ldots \theta_k)^\top$ is a k-dimensional vector of parameters and the parameter of interest is $t(\boldsymbol{\theta})$, which is a real-valued twice continuously differentiable parametric function. Consider a prior density for $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$, which matches frequentist and posterior probability for $t(\boldsymbol{\theta})$ as follows: For $-\infty < z < \infty$

$$\mathbb{P}_{\boldsymbol{\theta}}\left[n^{\frac{1}{2}}\left(t(\boldsymbol{\theta}) - t\left(\hat{\boldsymbol{\theta}}\right)\right)\frac{1}{\tau} \le z\right] = \mathbb{P}_{p(\boldsymbol{\theta})}\left[n^{\frac{1}{2}}\left(t(\boldsymbol{\theta}) - t\left(\hat{\boldsymbol{\theta}}\right)\right)\frac{1}{\tau} \le z \,|\, \boldsymbol{Y}\right] + O_p\left(n^{-1}\right)$$

where $\hat{\boldsymbol{\theta}}$ is the posterior mode or maximum likelihood estimator of $\boldsymbol{\theta}$, τ^2 is the asymptotic posterior variance of $n^{1/2}\left[t(\boldsymbol{\theta}) - t\left(\hat{\boldsymbol{\theta}}\right)\right]$ up to $O_p(n^{-1/2})$, $\mathbb{P}_{\boldsymbol{\theta}}(\cdot)$ is the joint probability measure of $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ under $\boldsymbol{\theta}$, and $\mathbb{P}_{P(\boldsymbol{\theta})}(\cdot|\boldsymbol{Y})$ is the posterior probability measure of $\boldsymbol{\theta}$ under the prior $p(\boldsymbol{\theta})$. According to Datta and Ghosh, such a prior may be sought in an attempt to reconcile a frequentist and Bayesian approach or to find (in some cases validate) a non-informative prior, or to construct frequentist confidence sets.

Let

$$\nabla_t(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial}{\partial\theta_1}t(\boldsymbol{\theta}) & \cdots & \frac{\partial}{\partial\theta_k}t(\boldsymbol{\theta}) \end{bmatrix}^\top$$

and

$$\eta(\boldsymbol{\theta}) = \frac{F^{-1}(\boldsymbol{\theta})\nabla_t(\boldsymbol{\theta})}{\sqrt{\nabla_t^\top(\boldsymbol{\theta})F^{-1}(\boldsymbol{\theta})\nabla_t(\boldsymbol{\theta})}} = \begin{bmatrix} \eta_1(\boldsymbol{\theta}) & \cdots & \eta_k(\boldsymbol{\theta}) \end{bmatrix}^\top$$

where $F(\boldsymbol{\theta})$ is the Fisher information matrix and $F^{-1}(\boldsymbol{\theta})$ is its inverse. It is evident that $\eta^\top(\boldsymbol{\theta}) F(\boldsymbol{\theta}) \eta(\boldsymbol{\theta}) = 1$ for all $\boldsymbol{\theta}$. Datta and Ghosh proved that the agreement between the posterior probability and the frequentist probability holds if and only if

$$\sum_{\alpha=1}^{k} \frac{\partial}{\partial\theta_\alpha} \{\eta_\alpha(\boldsymbol{\theta})p(\boldsymbol{\theta})\} = 0. \tag{13}$$

Henceforth $p(\boldsymbol{\theta})$ is the probability-matching prior for $\boldsymbol{\theta}$, the vector of unknown parameters.

The method of Datta and Ghosh (1995) provides a necessary and sufficient condition that a prior distribution must satisfy in order to have the probability-matching property. They pointed out that their method is more general than Tibshirani's, but will yield equivalent results when the parameter of interest is defined to be the first parameter in an orthogonal parameterization.

In the case of the Bühlmann-Straub model, we are interested in the probability-matching prior for δ. Let $\boldsymbol{\theta} = [m, \delta, \sigma_1^2]^\top$ and $t(\boldsymbol{\theta}) = \delta$, then

$$\frac{\partial t(\boldsymbol{\theta})}{\partial\theta} = \frac{\partial t(\boldsymbol{\theta})}{\partial\sigma_1^2} = 0, \quad \frac{\partial t(\boldsymbol{\theta})}{\partial\delta} = 1, \quad \nabla_t^\top(\boldsymbol{\theta}) = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \quad \text{and}$$

$$\nabla_t^\top(\boldsymbol{\theta})F^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} 0 & \frac{IJ}{2|H|}\left(\frac{1}{\sigma_1^2}\right)^2 & -\frac{1}{2|H|\sigma_1^2}\sum_{i=1}^{I}\frac{p_{i.}}{1+p_{i.}\delta} \end{bmatrix}.$$

Further,

$$\sqrt{\nabla_t^\top(\boldsymbol{\theta})F^{-1}(\boldsymbol{\theta})\nabla_t(\boldsymbol{\theta})} = \sqrt{\frac{IJ}{2\,|H|}\left(\frac{1}{\sigma_1^2}\right)^2} \quad \text{and}$$

$$\begin{aligned}\frac{\nabla_t^\top(\boldsymbol{\theta})F^{-1}(\boldsymbol{\theta})}{\sqrt{\nabla_t^\top(\boldsymbol{\theta})F^{-1}(\boldsymbol{\theta})\nabla_t(\boldsymbol{\theta})}} &= \begin{bmatrix} 0 & \frac{(IJ)^{\frac{1}{2}}}{\sqrt{2}\sqrt{|H|}}\frac{1}{\sigma_1^2} & -\frac{(IJ)^{-\frac{1}{2}}}{\sqrt{2}\sqrt{|H|}}\sum_{i=1}^{I}\frac{p_{i.}}{1+p_{i.}\delta} \end{bmatrix} \\ &= \begin{bmatrix} \eta_1(\boldsymbol{\theta}) & \eta_2(\boldsymbol{\theta}) & \eta_3(\boldsymbol{\theta}) \end{bmatrix}.\end{aligned}$$

The prior $p(\boldsymbol{\theta}) = p(m, \delta, \sigma_1^2)$ is a probability-matching prior if the differential equation (13) is satisfied.

If we take $p(\boldsymbol{\theta}) = \sqrt{|H|}$, then

$$\frac{\partial\{\eta_1(\boldsymbol{\theta})p(\boldsymbol{\theta})\}}{\partial m} = 0,$$

$$\frac{\partial\{\eta_2(\boldsymbol{\theta})p(\boldsymbol{\theta})\}}{\partial \delta} = \frac{\partial}{\partial \delta}\left\{\frac{(IJ)^{\frac{1}{2}}}{\sqrt{2}\sigma_{\mathrm{I}}^2}\right\} = 0, \quad \text{and}$$

$$\frac{\partial\{\eta_3(\boldsymbol{\theta})p(\boldsymbol{\theta})\}}{\partial \sigma_{\mathrm{I}}^2} = -\frac{\partial}{\partial \sigma_{\mathrm{I}}^2}\left\{\frac{(IJ)^{-\frac{1}{2}}}{\sqrt{2}}\sum_{i=1}^{I}\frac{p_{i.}}{1+p_{i.}\delta}\right\} = 0\,.$$

Thus the differential equation is satisfied and the probability-matching prior is given by

$$p_{R_3}(m,\delta,\sigma_{\mathrm{I}}^2) \propto \sigma_{\mathrm{I}}^{-2}\left\{\sum_{i=1}^{I}\frac{p_{i.}^2}{(1+p_{i.}\delta)^2} - \frac{1}{IJ}\left(\sum_{i=1}^{I}\frac{p_{i.}}{1+p_{i.}\delta}\right)^2\right\}^{\frac{1}{2}}, \tag{14}$$

which is identical to equation (11).

2.3 Posterior and Predictive Densities

Posterior and predictive densities are needed to make inferences about the unknown parameters and to predict future observations. For the linear model $\boldsymbol{Y} = m\boldsymbol{1} + Z\boldsymbol{u} + \boldsymbol{\varepsilon}$, where $\boldsymbol{u} \sim \mathrm{N}(\boldsymbol{0}, \delta\sigma_{\mathrm{I}}^2\tilde{I})$, and $\boldsymbol{\varepsilon} \sim \mathrm{N}(\boldsymbol{0}, P^{-1}\sigma_{\mathrm{I}}^2)$, van der Merwe and Bekker (2004) proved the following posterior and predictive densities:

1. The posterior distribution of m given σ_{I}^2 and δ, is normal with mean

$$\mathbb{E}\left(m\,|\,\boldsymbol{Y}, \sigma_{\mathrm{I}}^2, \delta\right) = \frac{\sum_{i=1}^{I}\overline{Y}_{i.}\frac{p_{i.}}{1+p_{i.}\delta}}{\sum_{i=1}^{I}\frac{p_{i.}}{1+p_{i.}\delta}} = \overline{Y}_{..} \tag{15}$$

and variance

$$\mathbb{V}\mathrm{ar}\left(m\,|\,\boldsymbol{Y}, \sigma_{\mathrm{I}}^2, \delta\right) = \left(\sum_{i=1}^{I}\frac{p_{i.}}{\sigma_{\mathrm{I}}^2(1+p_{i.}\delta)}\right)^{-1}. \tag{16}$$

2. The joint posterior distribution of σ_1^2 and δ is given by

$$p(\sigma_1^2, \delta \,|\, Y) = p(\sigma_1^2 \,|\, \delta, Y) p(\delta \,|\, Y)$$

where

$$p(\sigma_1^2 \,|\, \delta, Y) = C_1 \left(\frac{1}{\sigma_1^2}\right)^{\frac{1}{2}(IJ+1)} e^{-\frac{v(\delta, Y)}{2\sigma_1^2}}, \tag{17}$$

which is an inverse gamma probability density function, with

$$v(\delta, Y) = \sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij}(Y_{ij} - \overline{Y}_{i.})^2 + \sum_{i=1}^{I} \frac{p_{i.}(\overline{Y}_{i.} - \overline{Y}_{..})^2}{1 + p_{i.}\delta} \tag{18}$$

$$C_1 = \frac{1}{\Gamma\left(\frac{1}{2}(IJ-1)\right)} \left(\frac{2}{v(\delta, Y)}\right)^{-\frac{1}{2}(IJ-1)} \quad \text{and}$$

$$p(\delta \,|\, Y) = C_2 \left\{ \prod_{i=1}^{I} \left(\frac{1}{1 + p_{i.}\delta}\right)^{\frac{1}{2}} \right\} \left(\sum_{i=1}^{I} \frac{p_{i.}}{1 + p_{i.}\delta} \right)^{-\frac{1}{2}} \times (v(\delta, Y))^{-\frac{1}{2}(IJ-1)} p(\delta) \tag{19}$$

where C_2 is the normalizing constant and $p(\delta)$ is the prior (reference, probability-matching prior, or any other prior) distribution for δ.

3. The posterior distribution of $m_i = m + u_i$ for $i = 1, \ldots, I$, given δ is a Student t-distribution with $IJ - 1$ degrees of freedom, mean

$$\mathbb{E}(m_i \,|\, Y, \delta) = \frac{p_{i.}\delta}{1 + p_{i.}\delta} \overline{Y}_{i.} + \frac{1}{1 + p_{i.}\delta} \overline{Y}_{..} \tag{20}$$

and variance

$$\mathbb{V}\text{ar}(m_i \,|\, Y, \delta) = \left(\frac{1}{1 + p_{i.}\delta}\right) \left\{ \delta + \frac{1}{1 + p_{i.}\delta} \left(\sum_{i=1}^{I} \frac{p_{i.}}{1 + p_{i.}\delta} \right)^{-1} \right\} \times \left(\frac{1}{IJ-3}\right) v(\delta, Y). \tag{21}$$

4. The predictive probability density function of the mean of q future observations from the i^{th} group (q future claims from the i^{th} risk), given δ, is a Student-t distribution with $IJ-1$ degrees of freedom, mean

$$\mathbb{E}(\overline{\tilde{Y}}\,|\,\boldsymbol{Y},\delta) = \frac{p_{i.}\delta}{1+p_{i.}\delta}\overline{Y}_{i.} + \frac{1}{1+p_{i.}\delta}\overline{Y}_{..} \tag{22}$$

and variance

$$\mathbb{V}\mathrm{ar}(\overline{\tilde{Y}}\,|\,\boldsymbol{Y},\delta) = \left\{\frac{1}{q} + \frac{1}{1+p_{i.}\delta}\left[\delta + \frac{1}{1+p_{i.}\delta}\left(\sum_{i=1}^{I}\frac{p_{i.}}{1+p_{i.}\delta}\right)^{-1}\right]\right\} \times \left(\frac{1}{IJ-3}\right)v(\delta,\boldsymbol{Y}). \tag{23}$$

where $\overline{\tilde{Y}} = \sum_{j=1}^{q}\tilde{Y}_{ij}/q$, and $\tilde{Y}_{i1},\tilde{Y}_{i2},\ldots,\tilde{Y}_{iq}$ are the future claims from the i^{th} insured risk.

5. The predictive probability density function of the mean of q future claims from a new or arbitrary insured risk, given δ, is a Student t-distribution with $IJ-1$ degrees of freedom, mean

$$\mathbb{E}(\overline{Y}^*\,|\,\boldsymbol{Y},\delta) = \overline{Y}_{..} \tag{24}$$

and variance

$$\mathbb{V}\mathrm{ar}(\overline{Y}^*\,|\,\boldsymbol{Y},\delta) = \left\{\frac{1}{q} + \delta + \left(\sum_{i=1}^{I}\frac{p_{i.}}{1+p_{i.}\delta}\right)^{-1}\right\} \times \left(\frac{1}{IJ-3}\right)v(\delta,\boldsymbol{Y}) \tag{25}$$

where $\overline{Y}^* = \frac{1}{q}\sum_{j=1}^{q}Y_j^*$, and $Y_1^*,Y_2^*,\ldots,Y_q^*$ are the future claims.

Equations (20) and (22) can be written in the form $\tilde{Z}_i\overline{Y}_{i.} + (1-\tilde{Z}_i)\overline{Y}_{..}$, which means that the posterior mean of m_i (posterior mean of a group or company) and also the mean of the predictive density for that company is equal to the credibility estimator of the Bühlmann-Straub model.

3 Monte Carlo Simulation

The usual approach to the problem of predicting linear combinations of fixed and random effects ($m + u_i$, $i = 1, \ldots, I$), when the variance components are unknown is to estimate the unknown variance components and then proceed as if these estimates were the true values of the variance components. Patterson and Thompson (1971) and Harville (1974) have developed a method based on the maximum likelihood principle to derive unbiased estimates of the unknown variance components. This method is called restricted maximum likelihood estimation (REML). Substituting the REML estimates yields the empirical Bayes solution to the random effects model.[6]

In our opinion, there are several problems with simply substituting REML estimates for actual values:

1. The properties of REML estimators are hard to assess.

2. Sampling errors are generally ignored in the subsequent analysis. Therefore, the variance of the prediction error will generally be underestimated.

3. Depending upon the size and characteristics of the data, point estimators of variance components can be volatile. For certain values of the variance component estimators, the predictors obtained by substituting these values in the best linear unbiased predictor are intuitively unappealing.

An alternative approach to the empirical Bayesian approach is the fuller, more involved Bayesian approach, which, according to Harville (1990) and Gianola and Foulley (1990), can be used to devise prediction procedures that are more sensible, from both a Bayesian and frequentist perspective, than those in current use.

In many Bayesian problems marginal posterior distributions are used to make appropriate inferences. Technical difficulties that arise in the calculation of the marginal posterior densities needed in Bayesian inference, however, have long served as a practical impediment to the wider application of Bayesian methods. The main technical difficulties arise from the evaluation of high order multidimensional integrals. In the last few years, there have been a number of advances in the numerical integration and analytic approximation techniques for such calculations.

[6]For an example of this, see Chapter 8 on credibility theory in the ActEd Study Materials (2002) for Subject 106. *ActEd Study Materials 2002 Examinations. Subject 106 Course Notes* Oxford, United Kingdom: The Actuarial Education Company, 2002.

Implementation of these approaches typically requires access to high speed computers and sophisticated numerical or analytic approximation expertise and software.

In response to this limitation, Gelfand and Smith (1991), Gelfand et al. (1990), Carlin, Gelfand, and Smith (1992), and Gelfand, Smith, and Lee (1992) have applied Markov Chain Monte Carlo (MCMC) procedures, and more specifically the Gibbs sampler to the evaluation of these integrals.[7] The Gibbs sampler is an adaptive Monte Carlo integration technique. The typical objective of the sampler is to collect a sufficiently large number of parameter realizations from conditional posterior densities in order to obtain accurate estimates of the marginal posterior densities. The principal requirement of the sampler is that all conditional densities must be available in the sense that they can generate random variables.

The Gibbs sampler is appealing for its general applicability and ease of implementation. The burden of proof, however, is shifted to monitoring stochastic convergence and the mixing of the Markov chain. To date the monitoring only can be assessed with convergence diagnostics; see Robert and Casella (1999) and Jones and Hobert (2001). As pointed out by Gelfand (2002), "in general, convergence can never be assessed, as comparison can be made only between different iterations of one chain or between different observed chains, but never with the true stationary distribution." Because of this problem, researchers are interested in generating samples that are perfectly distributed as the stationary distribution of the Markov chain; see Green and Murdoch (1999) and Casella, Lavine, and Robert (2001). Unfortunately, generating samples that are perfectly distributed is currently feasible only for limited low-dimensional problems, and the cost of obtaining n samples is far greater than that of the usual MCMC, because essentially the entire algorithm must be repeated n times (Skare, Bølviken, and Holden, 2003).

We will now describe a simple algorithm to obtain the exact posterior and predictive densities for the Bühlmann-Straub model. These densities are obtained through Monte Carlo simulations where *independent* samples are obtained.[8] Conditional posterior densities of the form $p(\sigma_1^2 \mid \delta, \boldsymbol{Y})p(\delta \mid \boldsymbol{Y})$ or $p(m_i \mid \delta, \boldsymbol{Y})p(\delta \mid \boldsymbol{Y})$ and predictive densities such as $p(\tilde{\overline{Y}} \mid \delta, \boldsymbol{Y})p(\delta \mid \boldsymbol{Y})$ or $p(\overline{Y}^* \mid \delta, \boldsymbol{Y})p(\delta \mid \boldsymbol{Y})$ are used to sim-

[7] The Gibbs sampler is implicit in the work of Hastings (1970) and was made popular in the image-processing context by Geman and Geman (1984).

[8] This algorithm may be preferable to Gibbs sampling, which generates dependent samples from the joint distribution.

ulate the unconditional posterior and predictive densities. These densities can be obtained in the following way:

Step 1: By using the rejection method (Rice, 1995, page 91), an observation is generated from $p(\delta \mid Y)$ (equation (19)).

Step 2: Given δ, the conditional posterior and predictive densities $p(m_i \mid \delta, Y)$, $p(\tilde{\overline{Y}} \mid \delta, Y)$, and $p(\overline{Y}^* \mid \delta, Y)$ are Student t-distributions, while $p(\sigma_1^2 \mid \delta, Y)$ is an inverse gamma distribution.

These steps are repeated n times to get a sample of size n. Using a Rao-Blackwell argument (Gelfand and Smith, 1991) density estimates of the unconditional densities are obtained by averaging the conditional densities over the n repetitions.

4 Illustrative Examples

For these examples we set the sample size as $n = 10000$.

Example 1

We will first apply the Bayesian simulation procedure to the simple data set given on page 46 of Chapter 8 on Credibility Theory in the ActEd 106 Actuarial Study Guide, 2002. Table 2 shows the data for an international insurer's fire portfolio for a five year period. The data consist of X_{ij}, which is the aggregate claim amount and p_{ij}, which is the volume—aggregate claim amounts and volume are expressed in appropriate units. The claims per unit volume is $Y_{ij} = X_{ij}/p_{ij}$. For example, $Y_{11} = 48/12 = 4.00$ and $Y_{45} = 71/10 = 7.10$. Given the data for the past five years and the current volume p_{i6}, the insurer's problem is to determine the credibility premium for year 6 for each country.

The posterior density of δ is

$$p(\delta \mid Y) \propto \left\{ \prod_{i=1}^{I} \left(\frac{1}{1 + p_{i.}\delta} \right)^{\frac{1}{2}} \right\} \left(\sum_{i=1}^{I} \frac{p_{i.}}{1 + p_{i.}\delta} \right)^{-\frac{1}{2}} (v(\delta, Y))^{-\frac{1}{2}(IJ-1)} p(\delta) \tag{26}$$

Figure 1 shows $p(\delta \mid Y)$ for four different priors:

Table 2
An International Insurer's Annual Fire Portfolio (X_{ij}, p_{ij})
Aggregate Claims (X_{ij}) and Volume (p_{ij})

Country	Year (j)					
(i)	1	2	3	4	5	p_{i6}
1	(£48, 12)	(£53, 15)	(£42, 13)	(£50, 16)	(£59, 10)	20
2	(£64, 20)	(£71, 14)	(£64, 22)	(£73, 15)	(£70, 30)	25
3	(£85, 5)	(£54, 8)	(£76, 6)	(£65, 12)	(£90, 4)	10
4	(£44, 22)	(£52, 35)	(£69, 30)	(£55, 16)	(£71, 10)	12

Source: ActEd Study Materials 2002 Examinations. Subject 106 Course Notes. Oxford, United Kingdom: The Actuarial Education Company, 2002.

$$p_1(\delta) = \left\{ \sum_{i=1}^{I} \frac{p_{i.}^2}{(1 + p_{i.}\delta)^2} - \frac{1}{IJ} \left(\sum_{i=1}^{I} \frac{p_{i.}}{1 + p_{i.}\delta} \right)^2 \right\}^{\frac{1}{2}} \tag{27}$$

$$p_2(\delta) = \left\{ \sum_{i=1}^{I} \frac{p_{i.}^2}{(1 + p_{i.}\delta)^2} \right\}^{\frac{1}{2}} \tag{28}$$

$$p_A(\delta) = \left\{ \prod_{i=1}^{I} (1 + p_{i.}\delta) \right\}^{-\frac{1}{I}} \tag{29}$$

$$p_B(\delta) = \left\{ \prod_{i=1}^{I} \prod_{j=1}^{J} \frac{1}{(1 + p_{ij}\delta)} \right\}^{\frac{2}{IJ}} . \tag{30}$$

Note that $p_1(\delta)$ and $p_2(\delta)$ are the two reference priors defined in equations (11) and (12), while $p_A(\delta)$ and $p_B(\delta)$ are two priors motivated by Klugman [1992, page 133, equations (8.26) and (8.27)].

The posteriors $p(\delta \mid Y)$ generated by the two reference priors are almost indistinguishable for all practical purposes. The largest discrepancy is in the case of prior $p_B(\delta)$. The choice of any one of these priors, however, does not influence the posterior distributions of m_i (for $i = 1, \ldots, 4$) or the predictive densities of future claims that much. Therefore, we will use $p_1(\delta)$, as it is both a reference and a probability-matching prior.

Table 3 shows the credibility estimates (risk premiums per unit volume) for the four countries using a full Bayesian approach versus the

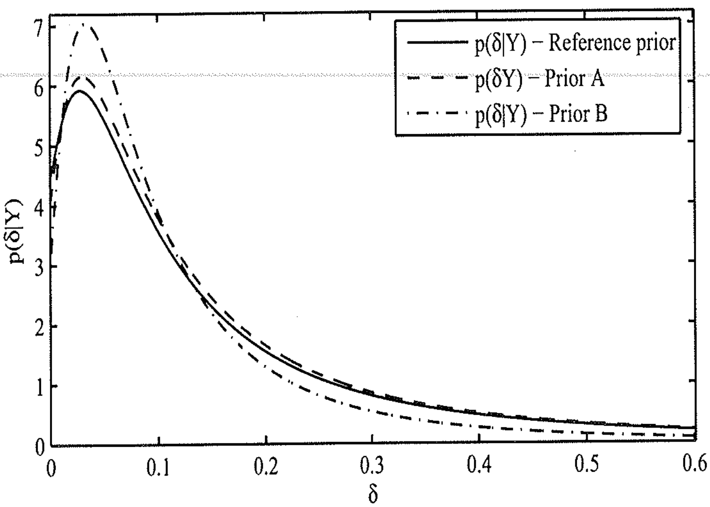

Figure 1: Posterior Probability Density Function of δ for Various Prior Distributions

empirical Bayes approach (EBCT risk premiums given on pages 61–63 of ActEd 106 Actuarial Study Guide, which is cited in Table 2).

Table 3
Credibility Estimates for the Four Countries

Country (i)	Empirical Bayes Estimate	Full Bayesian Analysis Posterior Mode	Full Bayesian Analysis Posterior Mean
1	3.851	3.9750	3.9762
2	3.468	3.5750	3.5668
3	8.504	9.2250	8.8427
4	2.750	2.8250	2.8562

As the volume for Country 4 for the coming year is 12 units, the fully Bayesian risk premium is: $2.8250 \times 12 = 33.9$ compared to the empirical Bayes premium of 33. The reason for the large difference between $\overline{Y}_{3.} = 10.571$ and the risk premium for Country 3 (8.504 for the empirical Bayes and 8.8427 for the full Bayesian procedure) is the small exposure [small amount of business ($p_{3.} = 35$)] associated with Country 3 over the years that results in large uncertainty with respect to estimation and prediction. The credibility factor for Country 3 is therefore quite small, which means that the mean of Country 3 ($\overline{Y}_{3.}$) will be closer to the overall mean ($\overline{Y}_{..}$) than in the case of the other countries. By comparing the two procedures, we see that there is not much of a difference between them. The fully Bayesian approach has, however, some additional advantages over the empirical Bayes analysis:

1. The Bayesian practitioner does not need to commit to only a point estimate of σ_{I}^2, δ, and the credibility estimator $\widetilde{Z}_i$. Credibility intervals and predictive densities can be easily obtained.

2. Uncertainty about the true values of σ_{I}^2 and δ is formally incorporated into the analysis through the choice of an appropriate prior distribution.

3. The Bayesian approach provides a set of widely applicable and mathematically tractable tools, often more tailored to the requirement of users than the corresponding frequentist tools.

Table 4 shows the means and credibility intervals for m_i, as well as the prediction intervals for future observations.

Table 4
Means and Credibility Intervals for m_i

		90% Credibility Interval Limits		90% Prediction Interval Limits	
i	$\mathbb{E}[m_i \mid Y]$	Lower	Upper	Lower	Upper
1	3.9762	1.8895	6.0745	0.7492	8.7214
2	3.5668	1.8495	5.3185	1.0874	6.0690
3	8.8427	5.0235	12.1850	4.0438	13.3000
4	2.8562	1.2415	4.5505	0.4870	5.2672

Example 2

This example is from Dannenburg, Kaas, and Goovaerts (1996, page 34). Our results will be illustrated by an application to collective automobile insurance data. Consider a portfolio consisting of nine fleets of cars that has been observed for a period of ten years. The relevant data associated with fleet i ($i = 1, \ldots, 9$) in year j ($j = 1, \ldots, 10$) are represented, as before, by the random variable Y_{ij}, which is an average taken over p_{ij} cars. We assume the data are consistent with the Bühlmann-Straub assumptions. Table 5 displays the data Y_{ij} with the number of cars p_{ij} in parentheses. The total observed risk exposure in the portfolio is 1510 years.

Figure 2 shows the posterior densities of δ [equation (26)] for the four priors [equations (27) to (30)], The four posterior distributions are more symmetrical and nearer to each other than the corresponding distributions illustrated in Figure 1 because the number of risks ($I = 9$) in this example is more than the four in Example 1. This means that the between-group variance is based on more degrees of freedom and δ can be more accurately estimated. The reference posteriors are again indistinguishable for all practical purposes.

Table 6 shows the credibility estimates for the nine fleets calculated using both a full Bayesian analysis and empirical Bayes. From Table 6 it is evident that there is little difference between the two methods. But as in the case of Example 1, the fully Bayesian estimate tends to shrink less than the empirical Bayes procedure. This means that the credibility factors for the fully Bayesian method are in general larger.

Table 7 shows the credibility estimates and credibility intervals for m_i using the full Bayesian approach, as well as the prediction intervals for the average of q future claims. The means of the posterior distri-

Table 5
Annual Average Claims in Fleet with Number of Car Units in Parentheses

i	Year 1	Year 2	Year 3	Year 4	Year 5
1	540 (44)	514 (50)	576 (56)	483 (58)	481 (58)
2	99 (20)	103 (20)	163 (24)	126 (32)	0 (28)
3	0 (8)	400 (6)	1042 (10)	313 (6)	0 (8)
4	275 (22)	278 (22)	430 (18)	196 (20)	667 (12)
5	543 (26)	984 (24)	727 (22)	562 (18)	722 (20)
6	0 (6)	0 (8)	0 (6)	645 (6)	833 (2)
7	333 (18)	404 (20)	400 (20)	361 (16)	588 (18)
8	494 (16)	133 (16)	735 (14)	519 (16)	1000 (14)
9	1667 (6)	313 (6)	556 (4)	769 (2)	1818 (4)

i	Year 6	Year 7	Year 8	Year 9	Year 10
1	493 (56)	438 (54)	588 (52)	541 (52)	441 (46)
2	219 (28)	370 (28)	273 (22)	155 (26)	275 (22)
3	833 (4)	0 (6)	0 (4)	0 (4)	0 (4)
4	185 (10)	517 (12)	204 (10)	323 (6)	968 (6)
5	610 (16)	794 (12)	299 (14)	580 (14)	488 (8)
6	0 (4)	0 (2)	769 (2)	0 (2)	0 (2)
7	349 (18)	435 (14)	476 (12)	635 (12)	556 (10)
8	641 (16)	339 (12)	513 (8)	227 (8)	244 (8)
9	0 (2)	1429 (4)	0 (2)	0 (4)	0 (2)

bution of m_i and the predictive distribution of the average of q future claim $\overline{\tilde{Y}}_{i,q}$ are exactly the same [equations (20) and (22)] but the 90% predictive interval for $\overline{\tilde{Y}}_{i,q}$ is much wider than the corresponding credibility interval for m_i. This fact illustrates the uncertainty associated with the prediction of future values. If we compare $\mathbb{V}\mathrm{ar}(m_i|Y,\sigma_1^2,\delta)$, equation (21), with $\mathbb{V}\mathrm{ar}(\overline{\tilde{Y}}_{i,q}|Y,\sigma_1^2,\delta)$ it is also evident that the latter variance is much larger.

In closing, Example 1 (small data set) was mainly used for illustrative purposes. It therefore does not matter what procedure (Bayesian, empirical Bayes, or frequentist) is used, a large amount of uncertainty will always be associated with the estimation of parameters and the pre-

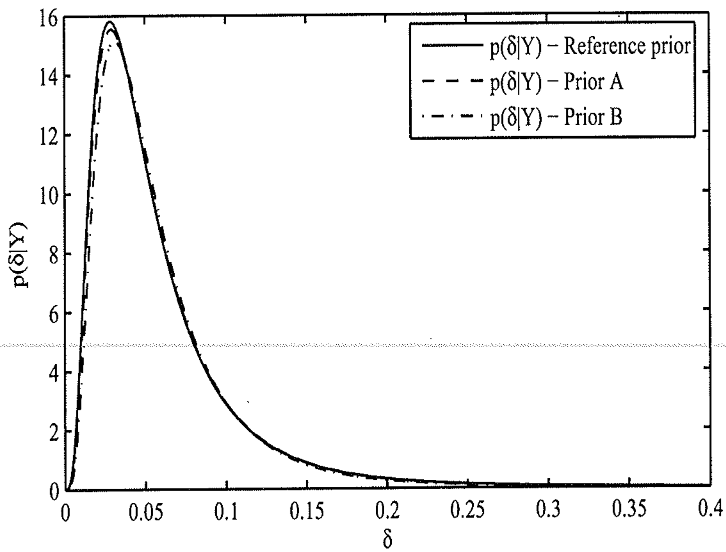

Figure 2: Posterior Probability Density Function of δ

Table 6
Credibility Premiums for the Automobile Insurance Data

Fleet (i)	1	2	3	4	5	6	7	8	9
$\overline{Y}_{i.}$	509	178	301	360	654	177	441	506	795
EBE	506	204	344	373	626	283	441	495	646
FBE	506	202	339	372	626	271	440	494	655

Notes: EBE = Emperical Bayes estimate; and FBE = Full Bayesian Estimate

Table 7
Means and Credibility Intervals for m_i

i	$\mathbb{E}[m_i\|Y]$	90% Credibility Interval Limits		90% Prediction Interval Limits	
		Lower	Upper	Lower	Upper
1	506	446	565	421	589
2	202	115	291	81	324
3	339	180	493	105	574
4	372	261	481	209	531
5	626	522	728	477	771
6	271	77	455	0 (-5)	559
7	440	337	544	288	592
8	494	381	609	326	660
9	655	456	866	348	961

diction of future values. In the case of Example 2, it is clear that there is large variation within and across groups. This is also the reason for the large prediction intervals illustrated in Table 7. The reason for the extremely large credibility and prediction intervals for fleets 3, 6, and 9, is that the experience of these fleets is limited.

One might possibly argue that some of the credibility and prediction intervals in the examples are so wide that they may appear useless for practical purposes and ask what the actuary should actually do in such situations. One possible solution is to obtain more data with many more groups and more observations per group. Larger samples will in general give smaller credibility intervals. Another possible solution is to assign proper priors with small variances or to assign priors on a restricted parameter space to the unknown parameters. The assignment of proper priors to the parameters must be justifiable from a practical point of view. In conclusion, it might be easy to obtain small Bayesian intervals but the question is whether the posterior and frequentist probabilities of these intervals will be the same. This is one of the reasons why the probability-matching prior (14) is used.

References

Antonio, K. and Beirlant, J. *Applications of Generalized Linear Mixed Models in Actuarial Statistics.* Technical Report. Leuven, Belgium: Actuarial Science Department, Katholieke Universiteit Leuven, 2005.

Berger, J.O. and Bernardo, J.M. "On the Development of Reference Priors." In *Bayesian Statistics IV* (eds. J.M. Bernardo, J.O. Berger, A.P. David, and A.F.M. Smith). Oxford, United Kingdom: Oxford University Press, 1992: 35–70.

Bernardo, J.M. *Bayesian Reference Analysis. A Postgraduate Tutorial Course*, Department of Mathematics, University of Valencia, Valencia, Spain, 1998.

Box, G.E.P. and Tiao, G.C. *Bayesian Inference in Statistical Analysis.* Reading, MA: Addison-Wesley, 1973.

Bühlmann, H. and Straub, E. "Glaubwürdigkeit für Schadensätze." *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* (1970): 111–133.

Carlin, B.P., Gelfand, A.E., and Smith, A.F.M. "Hierarchical Bayes Analysis of Change Point Problems." *Applied Statistics* 41, no. 2 (1992): 389–405.

Casella, G., Lavine, M., and Robert, C.P. "Explaining the Perfect Sampler." *American Statistician* 55 (2001): 299–305.

Dannenburg, D.R., Kaas, R., and Goovaerts, M.J. *Practical Actuarial Credibility Models.* Leuven, Belgium: Ceuterick, 1996.

Datta, G.S. and Ghosh, M. "On Priors Providing Frequentist Validity of Bayesian Inference." *Biometrika* 82 (1995): 37–45.

Gelfand, A.E., Hills, S.E., Racine-Poon, A., and Smith, A.F.M. "Illustration of Bayesian Inference in Normal Data Models using Gibbs Sampling." *Journal of the American Statistical Association* 85 (1990): 972–985.

Gelfand, A.E. and Smith, A.F.M. "Gibbs Sampling for Marginal Posterior Expectations." *Communications in Statistics Theory and Methods* 20, nos. 5 and 6 (1991): 1747–1766.

Gelfand, A.E., Smith, A.F.M., and Lee, T.M. "Bayesian Analysis of Constrained Parameters and Truncated Data Problems using Gibbs Sampling." *Journal of the American Statistical Association* 87 (1992): 523–532.

Gelfand, A.E. "Gibbs Sampling." In *Statistics in the 21st Century* (eds. A.E. Raftery, M.A. Tanner, and M.T. Wells). Boca Raton, FL: Chapman and Hall / CRC, 2002.

Geman, S. and Geman, D. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6 (1984): 721–741.

Gerber, H.U. "An Un-Bayesed Approach to Credibility." *Insurance: Mathematics and Economics* 1 (1982): 271–276.

Goulet, V. "Principles and Applications of Credibility Theory." *Journal of Actuarial Practice* 6 (1998): 5–62.

Gianola, D. and Foulley, J.L. "Variance Estimation from Integrated Likelihoods (VEIL)." *Genetics Selection Evolution* 22 (1990): 403–417.

Green, P.J. and Murdoch, D.J. "Exact Sampling for Bayesian Inference: Towards General Purpose Algorithms (with Discussion)." In *Bayesian Statistics 6* (eds. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith). Oxford, UK: Oxford University Press, 1999: 301–321.

Harville, D.A. "Bayesian Inference for Variance Components Using only Error Contrasts." *Biometrika* 61 (1974): 383–385.

Harville, D.A. "BLUP (Best Linear Unbiased Prediction and Beyond)." In *Advances of Statistical Methods for Genetic Improvement of Livestock* (eds. D. Gianola and K. Hammond). New York, NY: Springer-Verlag, 1990: 15–40.

Hastings, W.K. "Monte Carlo Sample Methods using Markov Chain and Their Applications." *Biometrika* 57 (1970): 97–109.

Jones, G.L. and Hobert, J.P. "Honest Exploration of Intractable Probability Distribution via Markov Chain Monte Carlo." *Statistical Science* 16 (2001): 312–334.

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

Makov, U.E., Smith, A.F.M., and Liu, Y.H. "Bayesian Methods in Actuarial Science." *Statistician* 45 (1996): 503-515.

Norberg, R. "Credibility Theory." In *Encyclopedia of Actuarial Science: Volume 1* (eds. J.L. Teugels and B. Sundt). New York: Wiley and Sons, 2004.

Patterson, H.D. and Thompson, R. "Recovery of Inter-block Information when Block Sizes are Unequal." *Biometrika* 58 (1971): 545–554.

Rice, J.A. *Mathematical Statistics and Data Analysis.* California: Duxbury Press, 1995.

Robert, C.P. *The Bayesian Choice,* Second Edition. New York, NY: Springer, 2001.

Robert, C.P. and Casella, G. *Monte Carlo Statistical Methods.* New York, NY: Springer, 1999.

Rubin, D.B. "Bayesian Justifiable and Relevant Frequency Calculation for the Applied Statistician." *The Annals of Statistics* 12 (1984): 1151–1172.

Scheffé, H. *The Analysis of Variance.* New York, NY: John Wiley and Sons, 1959.

Skare, Ø., Bølviken, E., and Holden, L. "Improved Sampling-Importance Resampling and Reduced Bias Importance Sampling." *Scandinavian Journal of Statistics* 30 (2003): 719–737

Tibshirani, R. "Non-informative Priors for One Parameter of Many." *Biometrika* 76, no. 3 (1989): 604–608.

Van der Merwe, A.J. and Bekker, K.N. *A Bayesian Approach to the Bühlmann-Straub Credibility Model.* Technical Report No. 336. Bloemfontein, South Africa: Department of Mathematical Statistics, University of the Free State, 2004.

Waters, H.R. *An Introduction to Credibility Theory.* London, United Kingdom: Institute of Actuaries, 1987.

Wolpert, R.L. "A Conversation with James O Berger." *Statistical Science* 91, 1 (2004): 205–218

Yang, R. and Chen, M.H. "Bayesian Analysis for Random Coefficient Regression Models using Non-informative Priors." *Journal of Multivariate Analysis* 55 (1995): 283–280.

Ye, K. "Selection of the Reference Priors for a Balanced Random Effects Model." *Test* 4 (1995): 1–17.