BAYES AND EMPIRICAL BAYES ESTIMATION FOR THE CHAIN LADDER MODEL

BY R. J. VERRALL

City University, London, United Kingdom

ABSTRACT

The subject of predicting outstanding claims on a porfolio of general insurance policies is approached via the theory of hierarchical Bayesian linear models. This is particularly appropriate since the chain ladder technique can be expressed in the form of a linear model. The statistical methods which are applied allow the practitioner to use different modelling assumptions from those implied by a classical formulation, and to arrive at forecasts which have a greater degree of inherent stability. The results can also be used for other linear models. By using a statistical structure, a sound approach to the chain ladder technique can be derived. The Bayesian results allow the input of collateral information in a formal manner. Empirical Bayes results are derived which can be interpreted as credibility estimates. The statistical assumptions which are made in the modelling procedure are clearly set out and can be tested by the practitioner. The results based on the statistical theory form one part of the reserving procedure, and should be followed by expert interpretation and analysis. An illustration of the use of Bayesian and empirical Bayes estimation methods is given.

INTRODUCTION

This paper is concerned with forecasting outstanding claims using the chain ladder model. The methods depend upon the fact (shown by KREMER (1982)) that the chain ladder technique is based upon a *linear model*. The particular linear model which is implied by the chain ladder technique is that of a two-way analysis of variance and it is this model that is main focus of this paper. However, the methodology applies to any linear model which might be used, for example a row-by-row model, and further details of the application of other models may be found in VERRALL (1988). It should be emphasised here that since the Bayes and empirical Bayes results in this paper are applied to the chain ladder model, there will be no consideration of 'tail factors'.

The identification of the model underlying the chain ladder technique as a linear model has made possible the application of statistical methods which have put the estimation procedures within a rigorous framework: for example, unbiased estimation of outstanding claims (and prediction intervals) has been

considered in VERRALL (1989a). The usual method of estimating development factors by calculating ratios of sums of cumulative claims down each column can be regarded as a useful 'rough and ready' method, although the maximum likelihood theorey (see VERRALL (1989b)) has the advantage that second moments can also be found.

Since the chain ladder model is a linear model, the theory of hierarchical linear models can be applied to obtain Bayes and empirical Bayes estimates. The empirical Bayes estimates (which require no prior information) have useful properties which can overcome some of the difficulties which practitioners may encounter when using the chain ladder model. The empirical Bayes estimates have a credibility theory interpretation, and it is interesting to note that de VYLDER (1982) obtained credibility-type estimates by applying the linear empirical Bayes theory directly to the chain ladder technique.

The methods in this paper will be applied, for illustrative purposes, to a particular set of claims data. As this is an illustration of the methods, it does not imply that it is a definitive analysis of the data: in particular no account is taken of inflation.

1. THE CHAIN LADDER LINEAR MODEL

Assuming a triangular data set (without loss of generality) the cumulative claims data, to which the chain ladder technique is applied, is

$$(1.1) \qquad \{C_{ij}: \ i = 1, \ldots, n; \quad j = 1, \ldots, n-i+1\}$$

The differenced data, to which the analysis of variance model is applied, is

$$(1.2) \qquad \{Z_{ij}: \ i = 1, \ldots, n; \quad j = 1, \ldots, n-i+1\}$$

where $\quad Z_{ij} = C_{ij} - C_{i,j-1} \qquad j \geq 2$
$\qquad \quad Z_{i1} = C_{i1}.$

The chain ladder technique is based on the model

$$(1.3) \qquad E(C_{ij}) = \lambda_j C_{i,j-1} \qquad j = 2, \ldots, n.$$

λ_j is estimated by $\tilde{\lambda}_j$, where

$$(1.4) \qquad \tilde{\lambda}_j = \frac{\displaystyle\sum_{i=1}^{n-j+1} C_{ij}}{\displaystyle\sum_{i=1}^{n-j+1} C_{i,j-1}}$$

The expected ultimate loss, $E(C_{in})$, is estimated by multiplying the latest loss, $C_{i,n-i+1}$, by the appropriate estimated λ-values:

$$(1.5) \qquad \text{estimate of } E(C_{in}) = \left(\prod_{j=n-i+2}^{n} \tilde{\lambda}_j \right) C_{i,n-i+1}$$

The chain ladder technique produces forecasts which have a row effect and a column effect. The column effect is obviously due to the parameters $\{\lambda_j; j = 2, \ldots, n\}$. There is also a row effect since the estimate for each of the rows depends not only on the parameters $\{\lambda_j; j = 2, \ldots, n\}$, but also on the row being considered. $C_{i, n-i+1}$ can be interpreted is an estimate of the row effect. This leads to consideration of other models which have row and column effects, in particular the two-way analysis of variance. Assuming that the incremental claims are positive, a natural assumption is that the data have a log-normal distribution and this implies that a logarithmic transformation is appropriate:

$$Y_{ij} = \log Z_{ij}$$

Now if Y_{ij} is assumed to have a normal distribution, Z_{ij} has a log-normal distribution. The following model was shown by KREMER to be similar to the chain ladder model:

(1.6) $$Y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$$

where e_{ij} are assumed to be independent, identically distributed normal disturbances with mean zero and variance σ^2.

The model is familiar as that of a two-way analysis of variance.

The linear model for the whole triangle is

(1.7) $$\underline{Y} = X \underline{\beta} + \underline{e}$$

where $\underline{Y}$ is the vector of observations,
$\overline{X}$ is the design matrix,
$\underline{\beta}$ is the parameter vector
and $\underline{e}$ is an error vector with covariance matrix Σ.

The maximum likelihood estimates of the parameter $\underline{\beta}$ are the solution, $\hat{\underline{\beta}}$, of the equation

(1.8) $$X' \Sigma^{-1} X \underline{\beta} = X' \Sigma^{-1} \underline{y}$$

Suppose, for example, there are three years of data. Then

(1.9) $$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{13} \\ y_{22} \\ y_{31} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{13} \\ e_{22} \\ e_{31} \end{bmatrix}$$

The form of the parameter vector and the design matrix are clear.

KREMER (1982) imposed the following restriction:

$$\sum_{i=1}^{n} \alpha_i = \sum_{j=1}^{n} \beta_j = 0.$$

An alternative assumption is that $\alpha_1 = \beta_1 = 0$, and it is this that is used in this paper.

It can be shown that the representations of the chain ladder model are equivalent (See, for example VERRALL (1989b)). KREMER considered the parameter estimation using the actuarial development factor method in common use and showed that it is similar to the maximum likelihood method applied to the additive model. The statistical methods allow a reassessment of the chain ladder method, and the next section gives a summary of the theory of hierarchical linear models, which can be used to derive Bayes and empirical Bayes estimates.

Denoting the set of data already observed by D, the Bayes estimate of outstanding claims for year of business i is

$$\sum_{j>n-i+1} E(Z_{ij} | D)$$

and the Bayes estimate of the variance is

$$\sum_{j>n-i+1} \left[\mathrm{Var}\,(Z_{ij} | D) + 2 \sum_{k>j} \mathrm{Cov}\,(Z_{ij},\, Z_{ik} | D) \right].$$

The Bayesian estimates are considered in more detail in VERRALL (1989c), where the following lemma is proved. It is assumed in the lemma that variances are known, although in a practical setting they have to be estimated. The substitution of estimates can be justified in the same way as in Section 2.3 below.

Lemma

Suppose that Z_{kl} has a lognormal distribution with parameters θ and σ, and that the posterior distribution of θ, given D, is normal with mean m and variance τ^2.

i.e. $\log Z_{kl} | \theta \sim N(\theta, \sigma^2)$

$\quad\quad \theta | D \sim N(m, \tau^2)$

Suppose also that σ^2 and τ^2 are known. Then

$$E(Z_{kl} | D) = e^{m + \frac{1}{2}\sigma^2 + \frac{1}{2}\tau^2}$$

and $\mathrm{Var}\,(Z_{kl} | D) = e^{2m + \sigma^2 + \tau^2} (e^{\sigma^2 + \tau^2} - 1)$

2. HIERARCHICAL LINEAR MODELS

Bayes estimates for the linear model were investigated by LINDLEY and SMITH (1972) and also SMITH (1973). In the actuarial literature, the recent paper by KLUGMAN (1989) has studied the use of hierarchical linear models in a rating context. It has already been seen that many of the models commonly used to analyse claims runoff triangles can be regarded as linear models, and we now analyse these models from a Bayesian point of view. This analysis has two purposes: firstly the practitioner may have some information, from other data for example, which can be used to specify a prior distribution for the parameters in the model and secondly the Bayesian analysis gives rise in a natural way to estimators which have a credibility theory interpretation.

In the first case the prior distribution is set by the practitioner and the usual prior-posterior analysis can be carried out. The models which we are using assume normal (really log-normal) distributions, and so it is only necessary to specify the mean and variance of the prior distribution (which is also normal). For example, if there is a lot evidence to suggest that the row parameters are all 0.1, a normal distribution with mean 0.1 and small variance can be used as prior. If there is not much prior information, the prior variance can be set larger. Indeed, in the limit, as the prior variance becomes large, we revert back to ordinary maximum likelihood estimation of the parameters.

In the second case, we will be using empirical estimation of the prior distributions. Thus the estimation will be empirical Bayes and we will assume that certain of the parameters are exchangeable. The historical requirement that credibility estimators be linear will also be satisfied in a generalised sense and so we could claim to have credibility formulae. The situation has some similarities with credibility estimators of risk premiums in that we can regard the rows in a runoff triangle as a set of risks and proceed as BUHLMANN (1967)—see GOOVAERTS and HOOGSTAD (1987) for a full description of Buhlmann's method. In the case of claims runoff triangles the rows contain different numbers of elements, and there are also the column parameters to contend with. This approach, starting from credibility premiums and working through to a credibility theory for loss runoff triangles was suggested by DE VYLDER (1982)—again see GOOVAERTS and HOOGSTAD (1987) for an exposition of the method. The present method starts from runoff triangles and proceeds to credibility formulae via the linear models. One of the major advantages of the linear model approach is that standard errors of the estimates are also produced. Before considering the application of hierarchical linear models to claims data, some results will be stated, their proof being contained in LINDLEY and SMITH.

2.1. Bayesian estimation for linear models: two stage models

The linear model, described above can be written

$$(2.1) \qquad\qquad \underline{y} \mid \underline{\beta} \sim N(X\underline{\beta}, \Sigma)$$

where y is a data vector of length n,
            β is a parameter vector of length p,
            X is an $(n \times p)$ design matrix
and Σ is an $(n \times n)$ dispersion matrix.

For a Bayesian formulation we need a prior distribution on the parameter vector β. In view of what is to come later in this section we will call this a second stage prior distribution. The second stage prior distribution is

$$\beta \mid \underline{\theta}_1 \sim N(A_1 \underline{\theta}_1, C_1)$$

where θ_1 is a vector of hyperparameters of length p_1,
            A_1 is a $(p \times p_1)$ matrix
and C_1 is a $(p \times p)$ dispersion matrix.

In a straightforward Bayesian analysis, A_1 is an identity matrix, θ_1 contains the prior estimates of the parameters and C_1 is taken as a diagonal matrix with variances inversely proportional to the believed accuracy of the estimates.

Lemma

The posterior distribution of the parameters is

(2.2) $\beta \mid \underline{y} \sim N(B \underline{b}, B)$

where

(2.3) $B^{-1} = X' \Sigma^{-1} X + C_1^{-1}$

and

(2.4) $\underline{b} = X' \Sigma^{-1} \underline{y} + C_1^{-1} A_1 \underline{\theta}_1$

The Bayes estimate of β (assuming a quadratic loss function) is the posterior mean, $\tilde{\beta}$, which is the solution of

$$(X' \Sigma^{-1} X + C_1^{-1}) \tilde{\beta} = X' \Sigma^{-1} \underline{y} + C_1^{-1} A_1 \underline{\theta}_1$$
$$= X' \Sigma^{-1} X \hat{\beta} + C_1^{-1} A_1 \underline{\theta}_1$$

using the normal equations (1.8). $\hat{\beta}$ is the maximum likelihood estimate of β. Then

(2.5) $\tilde{\beta} = (X' \Sigma^{-1} X + C_1^{-1})^{-1} [X' \Sigma^{-1} X \hat{\beta} + C_1^{-1} A_1 \underline{\theta}_1]$

Equation (2.5) shows that the Bayes estimate is a weighted average of the maximum likelihood estimate, $\hat{\beta}$, and the prior mean, $A_1 \theta_1$, with weights depending on the precision of each. It can be written in the usual way as a credibility formula:

$$\tilde{\beta} = z \hat{\beta} + (1 - z) A_1 \underline{\theta}_1$$

where $z = (X' \Sigma^{-1} X + C_1^{-1})^{-1} X' \Sigma^{-1} X$ and can be interpreted as a credibility factor.

As $C_1^{-1} \to 0$, $z \to 1$ and the estimate is based entirely on the data: the estimate is the maximum likelihood estimate.

2.2. Bayesian estimation for linear models: three stage models

In order to get an empirical Bayes formula, it is necessary to have a further prior distribution. The reason for this, and the way in which it is used will become clear in Section 3.

We define a prior distribution for a vector of hyperparameters:

$$(2.6) \qquad \underline{\theta}_1 | \underline{\theta}_2 \sim N(A_2 \underline{\theta}_2, C_2)$$

where $\underline{\theta}_2$ is a p_2-dimensional vector
A_2 is a $(p_1 \times p_2)$-dimensional matrix
C_2 is a $(p_1 \times p_1)$-dimensional matrix.

We now have the following lemma:

Lemma

Suppose that

$$
\begin{aligned}
\underline{y} | \underline{\beta} &\sim N(X \underline{\beta}, \Sigma) \\
\underline{\beta} | \underline{\theta}_1 &\sim N(A_1 \underline{\theta}_1, C_1) \\
\underline{\theta}_1 | \underline{\theta}_2 &\sim N(A_2 \underline{\theta}_2, C_2)
\end{aligned}
$$

Then the posterior distribution of $\underline{\beta}$ is

$$(2.7) \qquad \underline{\beta} | \underline{y}, \underline{\theta}_2 \sim N(D \underline{d}, D)$$

where

$$(2.8) \qquad D^{-1} = X' \Sigma^{-1} X + [C_1 + A_1 C_2 A_1']^{-1}$$

and

$$(2.9) \qquad \underline{d} = X' \Sigma^{-1} \underline{y} + [C_1 + A_1 C_2 A_1']^{-1} A_1 A_2 \underline{\theta}_2$$

Again, it can be seen that the Bayesian estimate (the posterior mean) is a weighted average of the maximum likelihood estimate and the prior mean:

$$(2.10) \quad \tilde{\underline{\beta}} = [X' \Sigma^{-1} X + (C_1 + A_1 C_2 A_1')^{-1}]^{-1} [X' \Sigma^{-1} X \hat{\underline{\beta}} + (C_1 + A_1 C_2 A_1')^{-1} A_1 A_2 \underline{\theta}_2]$$

This can be viewed as a credibility formula with credibility factor

$$[X' \Sigma^{-1} X + (C_1 + A_1 C_2 A_1')^{-1}]^{-1} X' \Sigma^{-1} X$$

It can be seen that the weight given to the maximum likelihood estimate depends on $X' \Sigma^{-1} X$, which is the inverse of the dispersion matrix of $\hat{\underline{\beta}}$ and the weight given to the prior mean depends on its dispersion matrix.

When the three-stage Bayesian model is used, a vague third-stage prior distributon will be used. In other words, C_2^{-1} will always be 0. In this case, the results are as given by the following lemma:

Lemma

Suppose

$$\begin{aligned}
\underline{y} \,|\, \underline{\beta} \quad &\sim N(X\underline{\beta}, \Sigma) \\
\underline{\beta} \,|\, \underline{\theta}_1 \quad &\sim N(A_1 \underline{\theta}_1, C_1) \\
\underline{\theta}_1 \,|\, \underline{\theta}_2 \quad &\sim N(A_2 \underline{\theta}_2, C_2)
\end{aligned}$$

and $C_2^{-1} = 0$.

Then the posterior distribution of $\underline{\beta} \,|\, \underline{y}$ is

$$\underline{\beta} \,|\, \underline{y} \sim N(D_o \underline{d}_o, D_o)$$

where

(2.11) $D_o^{-1} = X' \Sigma^{-1} X + C_1^{-1} - C_1^{-1} A_1 (A_1' C_1^{-1} A_1)^{-1} A_1' C_1^{-1}$

and

(2.12) $\underline{d}_o = X' \Sigma^{-1} \underline{y}$

This is the form which we will use for empirical Bayes estimation of the parameters. These estimators have an interpretation in credibility theory similar to the estimators used in premium setting by BUHLMANN. The credibility interpretation will be left until we consider the models which are used for claims analysis.

2.3. Variance estimation

In the preceding sections it has been assumed that the variances and covariances are known. This can be an unrealistic assumption: for example, when using an empirical Bayes approach, the dispersion matrix of $\underline{y} \,|\, \underline{\beta}$ is not known. It is thus necessary to have a method of estimating the variance-covariance matrices. Unfortunately, the usual procedure of putting a prior on the variances and integrating does not lead to tractable solutions and it is necessary to make a simplifying approximation.

Following LINDLEY and SMITH (1972), the modes of the joint distribution are used to estimate the parameters and the variances. Of course, since normal distributions are being used, this gives the same estimators for the parameters as already derived. It also produces variance estimates which are functions of the parameter estimates. An iterative procedure is used, iterating between the

parameter estimates assuming that the variances are known and the variance estimates assuming the parameters are known. Each case is different and the procedure will be described in each of the applications. Convergence is not found to cause any problems.

3. APPLICATION OF HIERARCHICAL LINEAR MODELS TO THE CHAIN LADDER MODEL

The chain ladder model can be considered as a linear model and the results of Section 2 applied. The results consist of two parts: Bayesian and empirical Bayes theory, corresponding to two and three-stage linear models.

3.1. Bayesian estimation for the chain ladder linear model

If there is some prior information, it may be possible to quantify this as a distribution which can then be used in a prior-posterior analysis.

Suppose that the prior information is quantified in the following prior distribution on $\underline{\beta}$:

$$(3.1) \qquad \underline{\beta} \mid \underline{\theta}_1 \sim N(A_1 \underline{\theta}_1, C_1)$$

A situation which may occur is that there are similar sets of data available which give information on the individual parameters. In this case A_1 can be taken as an identity matrix, the prior estimates can be put into $\underline{\theta}_1$ and their variances into C_1. In many cases C_1 will be a diagonal matrix of variances, although it is not necessary that the covariances are zero. In this case, the prior distribution becomes:

$$(3.2) \qquad \underline{\beta} \mid \underline{\theta}_1 \sim N(\underline{\theta}_1, C_1)$$

Assuming that the errors are independent, $\Sigma = \sigma^2 I_n$ where I_n is a square identity matrix of dimension $(n \times n)$.

From the lemma in Section 2.1 it can be seen that the Bayes estimate of the parameter vector is the solution, $\tilde{\beta}$, of

$$(3.3) \qquad (\sigma^{-2} X' X + C_1^{-1}) \tilde{\underline{\beta}} = \sigma^{-2} X' X \hat{\underline{\beta}} + C_1^{-1} \underline{\theta}_1$$

and the variance-covariance matrix of the estimates is

$$(3.4) \qquad \mathrm{Var}(\tilde{\underline{\beta}}) = [\sigma^{-2} X' X + C_1^{-1}]^{-1}$$

The equation for $\tilde{\underline{\beta}}$ can be written as a credibility formula:

$$(3.5) \qquad \tilde{\underline{\beta}} = z \hat{\underline{\beta}} + (1 - z) \underline{\theta}_1$$

where $z = (\sigma^{-2} X' X + C_1^{-1})^{-1} \sigma^{-2} X' X$ can be considered as a credibility factor.

It is interesting to note that the credibility factor has been generalised into a credibility matrix, since z is a $(p \times p)$ matrix. There will be subtle dependence of the elements in the Bayes estimator $\tilde{\underline{\beta}}$ on each of the elements in the maximum

likelihood estimator. It is not possible to write a credibility formula separately for each factor in the form

$$\tilde{\alpha}_i = z\hat{\alpha}_i + (1-z)\theta_i$$

To estimate the variance σ^2, the modal procedure described is used. The estimate of σ^2 is s^2, where

(3.6) $$s^2 = (\underline{y} - X\underline{\tilde{\beta}})'(\underline{y} - X\underline{\tilde{\beta}})/(n+2)$$

Thus the equations which give the Bayes estimates are (3.3), with σ^2 replaced by s^2, and (3.6).

The procedure begins with $s^2 = 0$ and iterates between the solutions of

$$(s^{-2}X'X + C_1^{-1})\underline{\tilde{\beta}} = s^{-2}X'X\underline{\hat{\beta}} + C_1^{-1}\underline{\theta}_1$$

and

$$s^2 = (\underline{y} - X\underline{\tilde{\beta}})'(\underline{y} - X\underline{\tilde{\beta}})/(n+2)$$

3.2. Empirical Bayes estimates for the chain ladder linear model

The previous section described the use of a two-stage conventional Bayesian model to analyse claims data. This section uses the three-stage Bayesian model described in Section 2.2 to derived empirical Bayes estimates for the analysis of variance (chain ladder) model. This method uses an improper prior distribution at the third stage for the row parameters and improper priors at the second stage for the overall mean and the column parameters. This means that for the overall mean and the column parameters the same assumptions are made as for the maximum likelihood estimators.

The row parameters are assumed to be independent samples from a common distribution—of course, they are unobservable, but this is the underlying assumption. A similar assumption is made in credibility theory. When premiums are calculated using credibility theory, a risk parameter is assigned to each risk and these are assumed to be independently, identically distributed. The set of risks is known as a collective, and the distribution from which the risk parameters is drawn is known as the structure of the collective. The situation in the claims reserving case is similar for the row parameters, but is complicated by the column parameters. The estimators produced will combine information from each row with information from the triangle as a whole. The prior distribution (i.e. the second stage distribution) is estimated from the data, and hence the estimators have an empirical Bayes interpretation.

The linear model for the chain ladder method is

(3.7) $$\underline{y}|\underline{\beta} \sim N(X\underline{\beta}, \sigma^2 I)$$

and the constraint $\alpha_1 = \beta_1 = 0$ will be used.

The errors have been assumed to be independently, identically distributed. X is as defined in the first section.

As in credibility theory, a structure is put onto the row parameters $\alpha_2, \alpha_3, \ldots, \alpha_t$: they are assumed to be independent observations from a common distribution. For the overall mean, μ, and the column parameters $\beta_2, \beta_3, \ldots, \beta_t$, the same distributional assumptions as for ordinary maximum likelihood estimation will be used. Thus at the second stage

(3.8)

$$\underline{\beta} \mid w, \theta, \underline{\xi} \sim N \left(\begin{bmatrix} 1 & 0 & \cdots & & 0 \\ 0 & 1 & 0 & & \vdots \\ \vdots & & 1 & & \vdots \\ \vdots & \vdots & & \vdots & \\ 0 & & 1 & & 0 \\ \vdots & & 0 & 1 & \\ \vdots & \vdots & & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} w \\ \theta \\ \xi_2 \\ \vdots \\ \xi_t \end{bmatrix}, \begin{bmatrix} \sigma_\mu^2 & & & & & & \\ & \sigma_\alpha^2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_\alpha^2 & & & \\ & & & & \sigma_\beta^2 & & \\ & & & & & \ddots & \\ & & & & & & \sigma_\beta^2 \end{bmatrix} \right)$$

and take $\sigma_\mu^{-2} \to 0$ and $\sigma_\beta^{-2} \to 0$.

In terms of equation (3.1),

$$\underline{\theta}_1 = \begin{bmatrix} w \\ \theta \\ \xi_2 \\ \vdots \\ \xi_t \end{bmatrix} \quad \text{and} \quad A_1 = \begin{bmatrix} 1 & 0 & \cdots & & 0 \\ 0 & 1 & 0 & & \vdots \\ \vdots & & 1 & & \vdots \\ \vdots & \vdots & & \vdots & \\ 0 & & 1 & 0 & \\ \vdots & & 0 & 1 & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}$$

θ is the mean of the common distribution of the row parameters $\alpha_2, \ldots, \alpha_t$.

Although the assumptions on the estimation of μ and $\beta_2, \ldots, \beta_t$ are the same as for the maximum likelihood estimation, the estimators produced will not be the same because of the treatment of the row parameters.

A vague prior distribution (a third-stage distribution) is used for θ. Since $\sigma_\mu^{-2} \to 0$ and $\sigma_\beta^{-2} \to 0$, a third-stage distribution is not needed for w and $\xi_2, \ldots, \xi_t$. Hence a combination of two-stage and three-stage models is used.

Theorem

If $\underline{y} | \underline{\beta} \sim N(X\underline{\beta}, \sigma^2 I)$

where $\underline{\beta} = (\mu, \alpha_2, \ldots, \alpha_t, \beta_2, \ldots, \beta_t)'$
and $\underline{\beta} | w, \theta, \underline{\xi} \sim N(A_1 \underline{\theta}_1, C_1)$.

$A_1, \underline{\theta}_1, C_1$ are given above, $\sigma_\mu^{-2} = 0$, $\sigma_\beta^{-2} = 0$ and a vague prior at the third stage is used.
    Then the Bayes estimates of $\underline{\beta}$, $\tilde{\underline{\beta}}$, is given by

(3.9)

$$
\tilde{\underline{\beta}} = \left(\sigma^{-2} X'X + \begin{bmatrix} 0 & & & & & \\ & \sigma_\alpha^{-2} & & & & \\ & & \ddots & & & \\ & & & \sigma_\alpha^{-2} & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix} \right)^{-1} \left(\sigma^{-2} X'X\hat{\underline{\beta}} + \begin{bmatrix} 0 & & & & & \\ & \sigma_\alpha^{-2} & & & & \\ & & \ddots & & & \\ & & & \sigma_\alpha^{-2} & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \tilde{\alpha}. \\ \vdots \\ \tilde{\alpha}. \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right)
$$

where $\tilde{\alpha}. = \dfrac{1}{t-1} \displaystyle\sum_{i=2}^{t} \tilde{\alpha}_i$

and has a credibility interpretation.
    The proof of this theorem is given in the appendix.
    It can be seen that the empirical Bayes estimates of the row parameters are in the general form of credibility estimates: they are the weighted average of the maximum likelihood estimates and the (weighted) average of the estimates from all the rows. The situation is complicated by the fact that $X'X$ is not a diagonal or block-diagonal matrix, so that the estimation of $\mu, \beta_2, \ldots, \beta_t$ involves the estimates of $\alpha_2, \ldots, \alpha_t$ and vice versa. This is entirely natural since changing the estimates of the row parameters obviously forces changes in the other estimates. However, it can be seen that the form of the estimates is the same as the form of credibility estimates. They are the weighted average of the maximum likelihood estimates and the (weighted) average of the estimates to which the credibility theory type assumptions have been applied. The weights depend on the precision of the estimates.
    As before, the variances σ^2 and σ_α^2 are replaced by modal estimates s^2 and s_α^2, which are given by

$$(3.10) \qquad s^2 = \frac{v\lambda + (\underline{y} - X\underline{\tilde{\beta}})'\,(\underline{y} - X\underline{\tilde{\beta}})}{n + v + 2}$$

$$(3.11) \qquad s_\alpha^2 = \frac{v_\alpha \lambda_\alpha + \sum_{i=2}^{t} (\tilde{\alpha}_i - \tilde{\alpha}.)^2}{t + v_\alpha + 1}$$

where v, λ, v_α and λ_α are set by the prior distribution of the variances. The derivation of these formulae, and the discussion of the prior parameter values is given in LINDLEY and SMITH (1972).

Again, the estimates are obtained by iterating between (3.9) and (3.10), (3.11). This is illustrated in the example.

The empirical Bayes assumptions could also be applied to the column parameters, although this is of little practical use.

4. EXAMPLE

For illustration purposes, a set of data is now analysed using the methods in this paper. This is an example of the way in which these methods can be applied, but it is not implied that this is a complete analysis of the data, or that the models used are optimal. The data is taken from a paper by TAYLOR and ASHE (1983) and consists of claims from a portfolio of general insurance policies (the source of the data is not specified).

357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
352118	884021	933894	1183289	445745	320996	527804	266172	425046	
290507	1001799	926219	1016654	750816	146923	495992	280405		
310608	1108250	776189	1562400	272482	352053	206286			
443160	693190	991983	769488	504851	470639				
396132	937085	847498	805037	705960					
440832	847631	1131398	1063269						
359480	1061648	1443370							
376686	986608								
344014									

with exposure factors

610 721 697 621 600 552 543 503 525 420

The exposures for each year of business are divided into the claims data before the analysis is carried out.

The example is divided into 3 sections. The first analyses the data asssuming no prior information, the second uses a prior distribution and a Bayesian analysis and the last uses the empirical Bayes estimation method.

4.1. No prior information

In order to assess the affect of the prior assumptions, it is useful to have the results of a Bayesian analysis when no prior information is used, and these are given in this example. The parameter estimates are the same as those obtained from a maximum likelihood analysis, and are given in Table 1.

TABLE 1

	Estimate	Standard error
Overall mean	6.106	0.165
Row parameters	0.194	0.161
	0.149	0.168
	0.153	0.176
	0.299	0.186
	0.412	0.198
	0.508	0.214
	0.673	0.239
	0.495	0.281
	0.602	0.379
Column parameters	0.911	0.161
	0.939	0.168
	0.965	0.176
	0.383	0.186
	-0.005	0.198
	-0.118	0.214
	-0.439	0.239
	-0.054	0.281
	-1.393	0.379

The standard errors are obtained from the estimates of the variance-covariance matrix of the parameter estimates:

$$(X'X)^{-1}\hat{\sigma}^2$$

where $\hat{\sigma}^2$ is the estimate of the residual variance. For this example, $\hat{\sigma}^2 = 0.116$.

Since the data is in the form of a triangle (there are the same number of rows and columns) and the matrix X is based on the design, the standard errors are the same for each row and column parameter.

The following table shows the Bayesian estimates of total outstanding claims for each year of business, together with the Bayesian estimate of the standard error of these.

TABLE 2

Predicted outstanding claims	Standard error
110927	60216
482157	189896
660810	210040
1090752	304721
1530532	401125
2310959	601536
3806976	1056660
4452396	1375446
5066116	2049337

When no prior information is assumed, the estimate of total outstanding claims is 19511632 and the estimate of the standard error of total outstanding claims is 3194056. A normal distribution approximation can be used to find an upper 95% confidence bound on total outstanding claims:

$$19\,511\,632 + 1.645 \times 3\,194\,056 = 24\,765\,854$$

4.2. Bayesian analysis with prior information

It is now possible to use standard Bayes theory to enter prior beliefs about the parameter values. These are entered in the usual way, as a prior distribution. This distribution can arise in several ways: a practitioner may be able to give information from past experience as to the values he expects for the parameters, or there may be sets of similar data which can be used to set up the prior distributions. The Institute of Actuaries General Insurance study group on claims runoff patterns has been carrying out extensive analyses of claims data, and one of their proposals is to produce tables of standard patterns. Tables of this sort could be used in the Bayesian analysis.

To illustrate the method, the TAYLOR and ASHE data is reanalysed assuming that there is other information which should be taken into account. Suppose, for example, that there is information which suggests that the following prior distribution is appropriate for each row parameter:

(4.1) $$\alpha_i \sim N(0.3, \ 0.05).$$

In practice, there will be specific reasons for choosing a particular prior distribution: the choice here is, in a sense, arbitrary.

The distribution of the overall mean, μ, and the distributions of the column parameters, $(\beta_j; \ j = 2, \ldots, n)$, remain unchanged. Also, it is assumed that the row parameters are independent (based on the prior information).

Thus equation (3.1) can be written in this case as:

$$\underline{\beta} \,|\, \underline{\theta}_1 \sim N(\underline{\theta}_1, C_1)$$

$$\text{where } \underline{\theta}_1 = \begin{bmatrix} 0 \\ 0.3 \\ 0.3 \\ \vdots \\ 0.3 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\text{and } C_1 = \begin{bmatrix} L & & & & & & \\ & 0.05 & & & & & \\ & & \ddots & & & & \\ & & & 0.05 & & & \\ & & & & L & & \\ & & & & & \ddots & \\ & & & & & & L \end{bmatrix}$$

and L is large (in the limit, infinite).

The assumption is that there no prior information on the overall mean and the column parameters, and hence the same distribution is used as would be used for maximum likelihood analysis for these parameters. This distribution is a vague prior in which the variance is very large indicating that there is no prior knowledge. The prior distribution of the row parameters, not being vague and hence different from the maximum likelihood analysis, affects the estimates of these parameters and has consequences for the estimates of the other parameters.

The Bayes estimates of the parameters and their standard errors are shown in Table 3.

These parameter estimates can be compared with estimates when there is no prior knowledge assumed, given in Table 1. Perhaps it is easier to see how the prior assumptions have affected the estimates by looking at Figures 1 and 2. These show plots of the Bayes estimates of the row and column parameters with and without prior information.

The situation is complicated by the interaction of the parameter estimates and the effect of the Bayes prior distribution is not, in general, straightforward. It can be seen from the plot of the row parameters, Figure 1, that the general effect is to draw these estimates towards the prior mean: these estimates certainly show less variation, and the row parameter estimates are a lot lower. The variation of the column parameters (looking at Fig. 2) remain about the same, but their values have altered slightly. This is entirely due to the change of

TABLE 3

| | Estimate | | Standard error of Bayes estimate with prior |
	With prior	No prior	
Overall mean	6.178	6.106	0.125
Row parameters	0.202	0.194	0.120
	0.168	0.149	0.123
	0.172	0.153	0.127
	0.276	0.299	0.133
	0.349	0.412	0.139
	0.400	0.508	0.147
	0.475	0.673	0.157
	0.360	0.495	0.172
	0.367	0.602	0.192
Column parameters	0.893	0.911	0.161
	0.910	0.939	0.168
	0.915	0.965	0.175
	0.318	0.383	0.184
	− 0.081	− 0.005	0.196
	− 0.201	− 0.118	0.212
	− 0.520	− 0.439	0.236
	− 0.129	− 0.054	0.277
	− 1.465	− 1.393	0.370
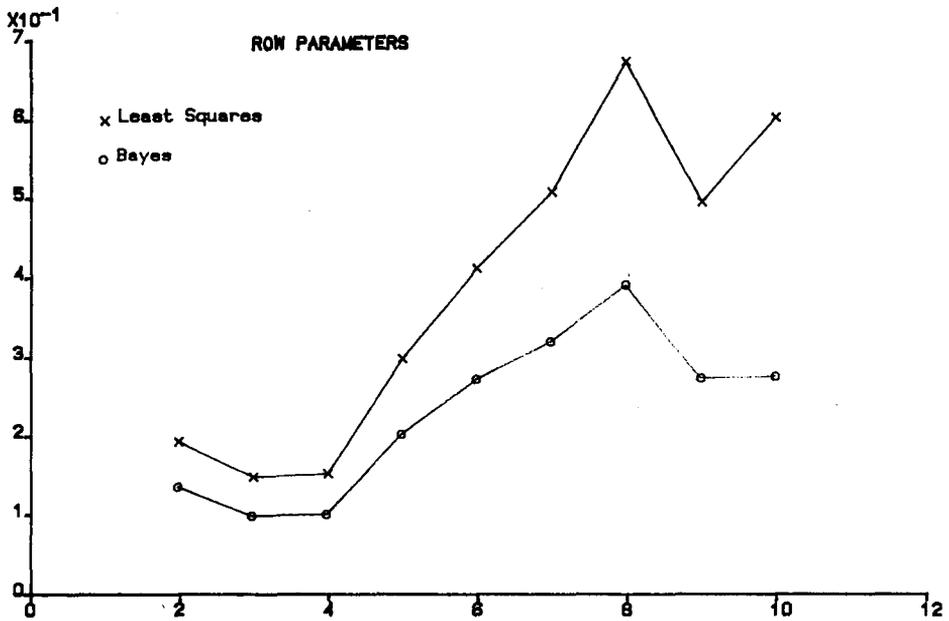
FIGURE 1.
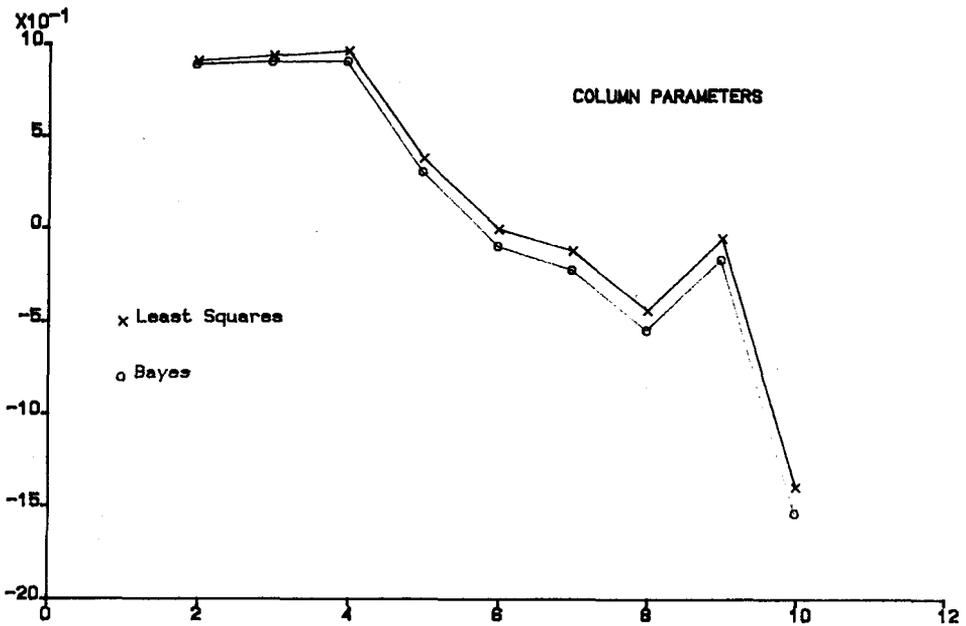
FIGURE 2.

prior assumptions on the row parameters and the interaction between the row and column parameters. The estimates of the overall mean has changed for the same reason.

It can also be seen, by comparing Tables 3 and 1 that the standard errors of the parameter estimates have been reduced. This is due to the extra prior information, and is an example of the usual property of Bayes estimates.

The row totals and their standard errors are given in the following table:

TABLE 4

| Bayesian estimates | | Standard error |
With prior	No prior	with prior
111748	110927	60516
489893	482157	191702
669724	660810	207990
1058206	1090752	282991
1425252	1530532	348013
2060499	2310959	482661
3117315	3806976	745547
3886838	4452396	936372
3923530	5066116	982585

The Bayesian estimate of overall total outstanding claims is 16743004 and the Bayesian estimate of the standard error of total outstanding claims is 1995669.

These can be compared with the results for the analysis in which no prior information is assumed. Looking first at the standard errors of the estimates of outstanding claims for each year of business, it can be seen that these have been quite considerably reduced. This is because the prior information (which was fairly precise) gave tight estimates of the parameters. The effect of the prior estimates on the estimates of outstanding claims is clear: the estimates of the total outstanding claims for the later rows have been drawn down due to the effect of the prior mean.

4.3. Empirical Bayes assumptions on the row parameters

To illustrate the effect of the assumptions made in the empirical Bayes theory, namely that the row parameters are independent observations from a common distribution, the TAYLOR and ASHE data is reanalysed in this example.

The estimates of the parameters and their standard errors are shown in Table 5:

TABLE 5

	Empirical Bayes estimate	No prior estimate	Standard error of Bayes estimate
Overall mean	6.157	6.106	0.131
Row parameters	0.225	0.194	0.124
	0.193	0.149	0.129
	0.198	0.153	0.133
	0.300	0.299	0.138
	0.371	0.412	0.144
	0.421	0.508	0.150
	0.493	0.673	0.159
	0.383	0.495	0.170
	0.391	0.602	0.185
Column parameters	0.893	0.911	0.128
	0.911	0.939	0.133
	0.915	0.965	0.139
	0.319	0.383	0.147
	−0.080	−0.005	0.156
	−0.199	−0.118	0.170
	−0.515	−0.439	0.190
	−0.120	−0.054	0.224
	−1.444	−1.393	0.306

The estimate of the variance of the row parameter distribution is 0.0289.

The empirical Bayes assumptions have been applied to the row parameters only. The effect of these assumptions is that the row parameters have been drawn towards a central point (a weighted average). The lower row parameter estimates have increased, while the higher ones have decreased. This can be seen more clearly from the following graph (Fig. 3) which shows a plot of the maximum likelihood and empirical Bayes estimates of the row parameters.
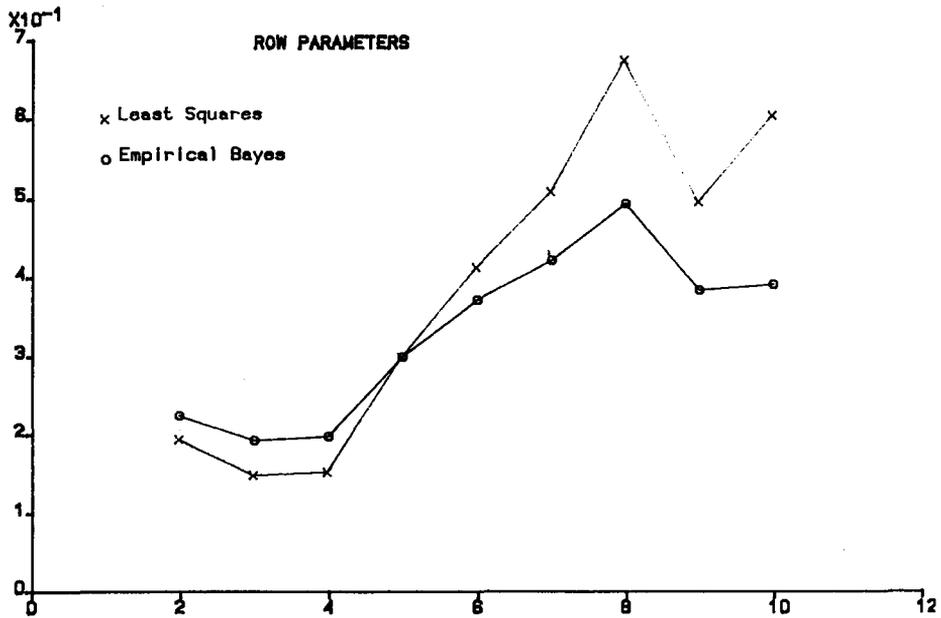
FIGURE 3.

Table 6 shows the row totals and their standard errors. For comparison purposes, the Bayes estimates with no prior assumptions (from Table 2) are also given.

The empirical Bayes estimate of total outstanding claims is 16280338 and the estimate of the standard error of total outstanding claims is 1313997.

The empirical Bayes standard errors are lower than the estimates with no prior information.

The estimates of total outstanding claims for the later rows have been quite considerably reduced, reflecting the reduction in the estimates of the row

TABLE 6

Empirical Bayes estimates	Bayes no prior	Empirical Bayes standard error
109448	110927	46963
479568	482157	148617
655656	660810	162104
1033109	1090752	220459
1388261	1530532	270730
2002772	2310959	374041
3018896	3806976	572899
3780759	4452396	720836
3811869	5066116	752593

parameters. The empirical Bayes procedure has thus given less weight to the estimates of the parameters from the later years: it has allowed that the rise in the maximum likelihood parameter estimates from row to row may be due to random variation. As more data becomes available, and there is more evidence in favour of either of these possibilities, this may, or may not, be revised.

5. CONCLUSIONS

The use of linear models to forecast outstanding claims enables the practitioner to apply results from statistical theory and hence arrive at a sound approach which retains the basic structure of the chain ladder model. Two estimation methods have been given here, which will be useful in different circumstances. Bayesian estimation allows the structured input of prior information, while empirical Bayes estimation allows more connection to be made between the factors in the model. The estimates are inherently more stable than those produced by a maximum likelihood analysis (or an analysis in which no prior information is assumed). An example which demonstrates the stability of the estimates is given in VERRALL (1989c).

It should be emphasised again that the method in this paper can be applied to any loglinear model, not just the chain ladder linear model which is the specific subject of this paper.

The incremental claims are assumed to be positive: the treatment of negative values is the subject of further research. A simple first approach is to add a suitable constant to the data: this is often sufficient.

The information obtained from a statistical analysis can be of use in the claims reserving process, although it has to be emphasised that this is only one part of the procedure: the practitioner will take into account other factors (such as inflation) and use his own skill and judgement.

REFERENCES

BUHLMANN, H. (1967) Experience Rating and credibility. *ASTIN Bulletin*, Vol. 4, No. 3, 199–207.
GOOVAERTS and HOOGSTAD (1987) Credibility Theory. *Surveys of Actuarial Studies*, No. 4.
KLUGMAN, S. (1989) Credibility for Classification Ratemaking via the Hierarchical Normal Linear Model. *Proc. of Casualty Actuarial Society*, Vol. 74, 272–321.
KREMER, E. (1982) IBNR-Claims and the Two-Way Model of ANOVA. *Scand. Act. J.*, Vol. 1, 47–55.
LINDLEY D. V. and SMITH, A. F. M. (1972) Bayes Estimates for the Linear Model (with Discussion) *JRSS, Series B*, Vol. 34, No. 1, 1–41.
SMITH, A. F. M. (1973) A General Bayesian Linear Model. *JRSS, Series B*, Vol. 35, No. 1, 67–75.
TAYLOR, G. C. and ASHE, F. R. (1983) Second Moments of Estimates of Outstanding Claims. *J. of Econometrics*, Vol. 23, 37–61.
VERRALL, R. J. (1988) Bayesian Linear Models and the Claims Run-off Triangle. *Actuarial Research Paper*, No. 7, The City University, London.
VERRALL, R. J. (1989a) On the Unbiased Estimation of Reserves from Loglinear Models. *Insurance: Mathematics and Economics*, to appear.

VERRALL, R.J. (1989b) Chain Ladder and Maximum Likelihood. Under review.
VERRALL, R.J. (1989c) A State Space Representation of the Chain Ladder Linear Model. Applications of Mathematics in Insurance, Finance and Accounting Work; joint meeting of the Institute and Faculty of Actuaries and the Institute of Mathematics and its Applications. *Journal of the Institute of Actuaries*, **Vol. 116, Part. 111,** 589–610.
DE VYLDER, F. (1982) Estimation of IBNR claims by Credibility Theory. *Insurance: Mathematics and Economics,* **1,** 35–40.

R.J. VERRALL
Department of Actuarial Science and Statistics, City University,
Northampton Square
London EC1V OHB

APPENDIX
PROOF OF THEOREM IN SECTION 3.2

The posterior distribution of $\underline{\beta}|\underline{y}$ is given by the second lemma in Section 2.2:

$$\underline{\beta}|\underline{y} \sim N(D_o\,\underline{d}_o, D_o)$$

where

(A1) $\qquad D_o^{-1} = \sigma^{-2}X'X + C_1^{-1} - C_1^{-1}A_1(A_1'C_1^{-1}A_1)^{-1}A_1'C_1^{-1}$

and

(A.2) $\qquad\qquad\qquad\qquad \underline{d}_o = \sigma^{-2}X'\underline{y}$

(A3) $\quad C_1^{-1}A_1 =$

$$
\begin{bmatrix}
\sigma_\mu^{-2} & & & & & & \\
 & \sigma_\alpha^{-2} & & & & & \\
 & & \ddots & & & & \\
 & & & \sigma_\alpha^{-2} & & & \\
 & & & & \sigma_\beta^{-2} & & \\
 & & & & & \ddots & \\
 & & & & & & \sigma_\beta^{-2}
\end{bmatrix}
\begin{bmatrix}
1 & 0 & \dots & & 0 \\
0 & 1 & 0 & & \vdots \\
\vdots & & 1 & & \vdots \\
\vdots & \vdots & \vdots & & \\
0 & & 1 & 0 & \\
\vdots & & 0 & 1 & \vdots \\
\vdots & \vdots & & \ddots & 0 \\
0 & \dots & \dots & 0 & 1
\end{bmatrix}
$$

$$
=
\begin{bmatrix}
\sigma_\mu^{-2} & & & & & \\
 & \sigma_\alpha^{-2} & & & & \\
 & & \vdots & & & \\
 & & \sigma_\alpha^{-2} & & & \\
 & & & \sigma_\beta^{-2} & & \\
 & & & & \ddots & \\
 & & & & & \sigma_\beta^{-2}
\end{bmatrix}
$$

and, since C_1 is diagonal, $A_1' C_1^{-1} = (C_1^{-1} A_1)' = \begin{bmatrix} \sigma_\mu^{-2} & & & & & \\ & \sigma_\alpha^{-2} & \cdots & \sigma_\alpha^{-2} & & \\ & & & & \sigma_\beta^{-2} & \\ & & & & & \ddots \\ & & & & & & \sigma_\beta^{-2} \end{bmatrix}$

$$A_1' \, C_1^{-1} A_1 \quad = \begin{bmatrix} 1 & & & & \\ & 1 & \cdots & 1 & \\ & & & 1 & \\ & & & & \ddots \\ & & & & & 1 \end{bmatrix} \begin{bmatrix} \sigma_\mu^{-2} & & & & & \\ & \sigma_\alpha^{-2} & & & & \\ & \vdots & & & & \\ & \sigma_\alpha^{-2} & & & & \\ & & \sigma_\beta^{-2} & & \\ & & & \ddots & \\ & & & & \sigma_\beta^{-2} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_\mu^{-2} & & & & \\ & (t-1)\sigma_\alpha^{-2} & & & \\ & & \sigma_\beta^{-2} & & \\ & & & \ddots & \\ & & & & \sigma_\beta^{-2} \end{bmatrix}$$

(A4)   Hence $(A_1' \, C_1^{-1} A_1)^{-1} = \begin{bmatrix} \sigma_\mu^{2} & & & & \\ & (t-1)^{-1}\sigma_\alpha^{2} & & & \\ & & \sigma_\beta^{2} & & \\ & & & \ddots & \\ & & & & \sigma_\beta^{2} \end{bmatrix}$

$$(A5) \quad (A_1' C_1^{-1} A_1)^{-1} A_1' C_1^{-1} = \begin{bmatrix} \sigma_\mu^2 & & & & \\ & (t-1)^{-1}\sigma_\alpha^2 & & & \\ & & \sigma_\beta^2 & & \\ & & & \ddots & \\ & & & & \sigma_\beta^2 \end{bmatrix} \begin{bmatrix} \sigma_\mu^{-2} & & & & \cdot \\ & \sigma_\alpha^{-2} \dots \sigma_\alpha^{-2} & & & \\ & & \sigma_\beta^{-2} & & \\ & & & \ddots & \\ & & & & \sigma_\beta^{-2} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & & & & \\ & (t-1)^{-1} \dots (t-1)^{-1} & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}$$

Using (A3) and (A5):

$$C_1^{-1} A_1 (A_1' C_1^{-1} A_1)^{-1} A_1' C_1^{-1} = \begin{bmatrix} \sigma_\mu^{-2} & & & & & \\ & \sigma_\alpha^{-2} & & & & \\ & \vdots & & & & \\ & \sigma_\alpha^{-2} & & & & \\ & & \sigma_\beta^{-2} & & & \\ & & & \ddots & & \\ & & & & \sigma_\beta^{-2} \end{bmatrix} \begin{bmatrix} 1 & & & & \\ & (t-1)^{-1} \dots (t-1)^{-1} & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}$$

$$(A6) \quad = \begin{bmatrix} \sigma_\mu^{-2} & & & & & \\ & (t-1)^{-1}\sigma_\alpha^{-2} & \dots & (t-1)^{-1}\sigma_\alpha^{-2} & & \\ & \vdots & \ddots & & & \\ & (t-1)^{-1}\sigma_\alpha^{-2} & \dots & (t-1)^{-1}\sigma_\alpha^{-2} & & \\ & & & & \sigma_\beta^{-2} & \\ & & & & & \ddots & \\ & & & & & & \sigma_\beta^{-2} \end{bmatrix}$$

$$(A7) \qquad = \begin{bmatrix} \sigma_\mu^{-2} & & \\ & (t-1)^{-1}\sigma_\alpha^{-2}J_{t-1} & \\ & & \sigma_\beta^{-2}I_{t-1} \end{bmatrix}$$

where J_n is a square $(n \times n)$ matrix, all of whose elements are 1
and I_n is an $(n \times n)$ identity matrix.

Substituting (A7) into (A1) and (A2) gives the Bayes posterior mean, $\tilde{\underline{\beta}}$, as the solution of

$$(A8) \qquad \left(\sigma^{-2}X'X + \begin{bmatrix} \sigma_\mu^{-2} & & & & & & \\ & \sigma_\alpha^{-2} & & & & & \\ & & \ddots & & & & \\ & & & \sigma_\alpha^{-2} & & & \\ & & & & \sigma_\beta^{-2} & & \\ & & & & & \ddots & \\ & & & & & & \sigma_\beta^{-2} \end{bmatrix} - \begin{bmatrix} \sigma_\mu^{-2} & & \\ & (t-1)^{-1}\sigma_\alpha^{-2}J_{t-1} & \\ & & \sigma_\beta^{-2}I_{t-1} \end{bmatrix} \right) \tilde{\underline{\beta}} = \sigma^{-2}X'\underline{y}$$

Now putting $\sigma_\mu^{-2} = 0$ and $\sigma_\beta^{-2} = 0$:

$$(A9) \qquad \left(\sigma^{-2}X'X + \begin{bmatrix} 0 & & & & & & \\ & \sigma_\alpha^{-2} & & & & & \\ & & \ddots & & & & \\ & & & \sigma_\alpha^{-2} & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix} - \begin{bmatrix} 0 & & & \\ & (t-1)^{-1}\sigma_\alpha^{-2}J_{t-1} & & \\ & & 0 & \\ & & & \ddots \\ & & & & 0 \end{bmatrix} \right) \tilde{\underline{\beta}} = \sigma^{-2}X'\underline{y}$$

If the individual Bayes solutions are called $\tilde{\mu}, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_t, \tilde{\beta}_1, \ldots, \tilde{\beta}_t$, then

$$(A10) \qquad \tilde{\underline{\beta}} = (\tilde{\mu}, \tilde{\alpha}_2, \ldots, \alpha_t, \tilde{\beta}_2, \ldots, \tilde{\beta}_t)'$$

There should be no confusion between the vector of parameters $\underline{\beta}$ and the column parameters $\beta_2, \ldots, \beta_t$ which are components of $\underline{\beta}$.

Equation (A9) can be rewritten as

$$
(A11) \quad \left(\sigma^{-2} X' X + \begin{bmatrix} 0 & & & & & & \\ & \sigma_\alpha^{-2} & & & & & \\ & & \ddots & & & & \\ & & & \sigma_\alpha^{-2} & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix} \right) \tilde{\underline{\beta}} - \sigma_\alpha^{-2} \begin{bmatrix} 0 \\ \tilde{\alpha}. \\ \vdots \\ \tilde{\alpha}. \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \sigma^{-2} X' \underline{y}
$$

where $\tilde{\alpha}. = \dfrac{1}{t-1} \displaystyle\sum_{i=1}^{t} \tilde{\alpha}_i$

Hence

$$
(A12) \quad \tilde{\underline{\beta}} = \left(\sigma^{-2} X' X + \begin{bmatrix} 0 & & & & & & \\ & \sigma_\alpha^{-2} & & & & & \\ & & \ddots & & & & \\ & & & \sigma_\alpha^{-2} & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix} \right)^{-1}
$$

$$
\left(\sigma^{-2} X' X \hat{\underline{\beta}} + \begin{bmatrix} 0 & & & & & & \\ & \sigma_\alpha^{-2} & & & & & \\ & & \ddots & & & & \\ & & & \sigma_\alpha^{-2} & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \tilde{\alpha}. \\ \vdots \\ \tilde{\alpha}. \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right)
$$