Statistical Methods for the Chain Ladder Technique

by Richard J. Verrall

<u>Abstract</u>

This paper considers the application of loglinear models to claims reserving. The models encompass the chain ladder technique and extend the range of the possible analyses. By bringing the methods within a statistical framework, a coherent strategy for testing goodness of fit and for forecasting outstanding claims is produced. Improvements to the basic chain ladder technique are given which use Bayesian methods.

1. Introduction

Forecasting outstanding claims and setting up suitable reserves to meet these claims is an important part of the business of a general insurance company. Indeed, the published profits of these companies depend not only on the actual claims paid, but on the forecasts of the claims which will have to be paid. It is essential, therefore, that a reliable estimate is available of the reserve to be set aside to cover claims, in order to ensure the financial stability of the company and its profit and loss account. There are a number of methods which have proved useful in practice, one of which is extensively used and is known as the chain ladder technique. In recent years, a statistical framework for analysing this data has been built up, which encompasses the actuarial method, extending and consolidating it. The aim of this paper is to bring together these results and to illustrate how the chain ladder technique can be improved and extended, without altering the basic foundations upon which it has been built. These improvements are designed to overcome two problems with the chain ladder technique. Firstly, that not enough connection is made between the accident years, resulting in an over-parametrised model and unstable forecasts. Secondly, that the development pattern is assumed to be the same for all accident years. No allowance is made by the chain ladder technique for any change in the speed with which claims are settled, or for any other factors which may change the shape of the run-off pattern. Before describing the methods for overcoming these problems, we first define the chain ladder linear model, and show how it can be used to give upper prediction bounds on total outstanding claims.

2. The Data

It will be assumed throughout this paper that the data is in the form of a triangle. It should be emphasised that this is for notational convenience only: there are no problems in extending the methods to other shapes of data. The year in which the policy is written will be called the underwriting year, accident year or year of business. In the years after the policy was written the company may receive claims related to that policy, and these claims are indexed by their business year and the delay. The following data set, which is taken from Taylor and Ashe (1983) will be used for illustrative purposes. The data is given in the form of incremental claims in each delay year.

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

The data may take a slightly different shape if one or more of the corners is truncated, but this paper will consider triangles of data (without loss of generality). The first column will be labelled delay year 1, rather than delay year 0.

Sometimes, the rows are standardised by dividing by a measure of the volume of business, such as the premium income. This is reversed when predictions of outstanding claims are made. For the above triangle the exposure factors are:

610 721 697 621 600 552 543 503 525 420.

The incremental claims relating to business year i and delay year j will be denoted Z_{ij}, so that the set of data observed is

$$\{ Z_{ij} : i=1,...,t;\ j=1,...,t\text{-}i+1\}$$

The statistical approach uses the incremental claims, but the chain ladder technique is applied to the cumulative claims, which are defined by:

$$C_{ij} = \sum_{k=1}^{j} Z_{ik}$$

The problem is to forecast outstanding claims on the basis of past experience. In other words to fill

in the lower right hand triangle of claims. Sometimes it is also useful to extend the forecasts beyond the latest delay year (i.e. to the right of the claims run-off triangle). The standard actuarial technique does not attempt to do this.

3. Linear Models and the Chain Ladder Technique

This paper will concentrate on the chain ladder technique. It its familiar form, this assumes that the cumulative claims for each business year develop similarly by delay year, and estimates development factors as ratios of sums of cumulative claims with the same delay index. Thus the estimate of the development factor for column j is

$$\frac{\sum_{i=1}^{t-j+1} C_{ij}}{\sum_{i=1}^{t-j+1} C_{i,j-1}}. \tag{3.1}$$

The model on which this is based is

$$E\left(C_{ij} \mid C_{i1}, C_{i2}, \ldots, C_{i,j-1} \right) = \lambda_j \, C_{i,j-1} \qquad\qquad j=2,\ldots,t. \tag{3.2}$$

and (3.1) is an estimate of λ_j . It has the advantage that it is relatively straightforward to calculate, but there is no clear basis on which to examine the properties in greater detail. It can be seen as a useful "rough-and-ready" estimation method.

The expected ultimate loss, $E\left(C_{it} \right)$, is estimated by multiplying the latest loss, $C_{i,t-i+1}$, by the appropriate estimated λ-values :

$$\text{estimate of } E\left(C_{it} \right) = \left(\prod_{j=t-i+2}^{t} \tilde{\lambda}_j \right) C_{i,t-i+1} . \tag{3.3}$$

The chain ladder technique produces forecasts which have a row effect and a column effect. The column effect is obviously due to the parameters $\{ \lambda_j \, ; \, j=2,\ldots,t \}$. There is also a row effect since the estimates for each row depend not only on the parameters $\{ \lambda_j \, ; \, j=2,\ldots,t \}$, but also on the row being considered. The latest cumulative claims, $C_{i,t-i+1}$, can be considered as the row effect. This leads to consideration of other models which have row and column effects, in particular the two-way

397

analysis of variance model. The connection is first made with a multiplicative model. This uses the non-cumulative data, Z_{ij} , and models them according to:

$$E \left(Z_{ij} \right) = U_i \, S_j \qquad (3.4)$$

where $\qquad U_i$ is a parameter for row i,

$\qquad\qquad S_j$ is a parameter for column j

A multiplicative error structure is assumed.

$$\text{Also} \qquad \sum_{j=1}^{t} S_j \;=\; 1 \qquad (3.5)$$

S_j is the expected proportion of ultimate claims which occur in the jth development year.

U_i is the expected total ultimate claim amount for business year i (neglecting any tail factor).

Kremer(1982) showed that the following relationships between the parameters hold:

$$S_j \quad = \quad \frac{\lambda_j - 1}{\prod\limits_{l=j}^{t} \lambda_l} \qquad (j \geq 2) \qquad (3.6)$$

$$S_1 \;=\; \frac{1}{\prod\limits_{l=2}^{t} \lambda_l} \qquad (3.7)$$

and $\qquad U_i \;=\; E \left(C_{it} \right).$ $\qquad\qquad (3.8)$

Estimators of { S_i ; i=1,...,t } and { U_j ; j=1,...,t } can be obtained by applying a linear model to the logged incremental claims data. Taking logs of both sides of equation (3.4), and assuming that the incremental claims are positive:

$$E \left(Y_{ij} \right) = \mu + \alpha_i + \beta_j \qquad (3.9)$$

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

and the errors now have an additive structure and are assumed to have mean zero.

The errors will also be assumed to be identically distributed with variance σ^2, although this distributional assumption can be relaxed.

The usual restriction is placed on the parameters to ensure a non-singular design matrix, in this case

$$\alpha_1 = \beta_1 = 0.$$

Now equation (3.9) can be written in the form of a linear model. Suppose, for example, there are three years of data.

Then

$$
\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{13} \\ y_{22} \\ y_{31} \end{bmatrix}
=
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0
\end{bmatrix}
\begin{bmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \beta_2 \\ \beta_3 \end{bmatrix}
+
\begin{bmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{13} \\ e_{22} \\ e_{31} \end{bmatrix}
\qquad (3.10)
$$

Kremer (1982) derived the normal equations for the chain ladder linear model and also examined the relationships between the linear model and the crude chain ladder technique. By reversing the transformation it can be shown that

$$U_i = e^{\alpha_i} e^{\mu} \sum_{j=1}^{i} e^{\beta_j}$$

Kremer showed that if the estimate of U_i is obtained by "hatting" the parameters in the above identity, the result is very similar to that obtained from the chain ladder technique. The resulting estimate of U_i is not the maximum likelihood estimate, neither is it unbiased, but it does serve the purpose of illustrating the similarity between the chain ladder technique and the two-way analysis of variance.

Furthermore, if all the geometric means are replaced by arithmetic means the estimators of the parameters of the models are equivalent. Thus the two estimation methods, the chain ladder method and the linear model, will produce identical results. The structure of the models is identical and the only difference is the estimation technique. It can be argued that the linear model estimates are best

in a statistical sense, but it should be emphasised that in using the linear model instead of the crude chain ladder technique, there are no radical changes.

In general, a loglinear model applied to this data takes the form

$$\underline{y} = X\underline{\beta} + \underline{e}$$

where $\underline{y}$ is the vector of logged incremental claims,

X is the design matrix,

$\underline{\beta}$ is the parameter vector and

$\underline{e}$ is a vector of errors.

Apart from the chain ladder linear model, other models which have been suggested as suitable for claims data include a gamma curve (suggested by Zehnwirth (1985))

$$y_{ij} = \nu_i + \delta_i \log j + \gamma_i \, (\, j - 1 \,) + e_{ij}$$

and an exponential tail (suggested by Ajne (1989)) in which the first few delay years follow the chain ladder model and the later delay years follow an exponential curve.

The statistical treatment facilitates the production of standard errors as well as point forecasts. This is a considerable advantage over the ad hoc methods, and allows 'safe' upper limits on outstanding claims to be set. The statistical analysis is more comprehensive and allows a greater study of the models, their fit to the data and any unusual features in the data. Also, Bayesian methodology can be incorporated to allow the structured input of other information, and to extend the range of the analysis by including empirical Bayes and state space methods. This has beneficial consequences for the stability of the predictions.

With reference to the computing aspects, Renshaw (1989) has shown how these models can be implemented in GLIM, and Christofides (1990) has used the spread-sheet package SuperCalc5.

4. Estimation of Reserves, and Variances of Reserves.

It has been shown that the chain ladder model can be considered as a two-way analysis of variance. This linear model, and other linear models, can be used effectively for analysing claims data and producing estimates of expected total outstanding claims for each year of business. The methods have in common the assumption that the data is lognormally distributed, and the linear models are therefore applied to the logged incremental claims rather than the raw incremental claims data. The problem therefore arises of reversing the log transformation to produce estimates on the original scale. It is this problem which is addressed in this section; in particular the unbiasedness of the estimates is considered. This problem was first addressed in Verrall(1991a), in which the following analysis was given.

4.1. Identically Distributed Data

Before considering the claims run-off triangle, consider n independently, identically distributed observations which are lognormally distributed.

i.e. $Z_1 , ... , Z_n$ are independent

and $Z_i \sim$ lognormal.

Suppose also that $E (Z_i) = \theta$.   (4.1)

The aim is to estimate θ and to find the mean square error (or variance, if the estimate is unbiased) of the estimate. One way of proceeding towards the estimation of θ is to take logs of the data and analyse the resulting sample using normal distribution theory. This is an approach which can be generalised to data which is not identically distributed and so is the most appropriate for claims data.

Let $Y_i = \log Z_i$ (i=1,...,n).   (4.2)

Since Z_i has a lognormal distribution, Y_i has a normal distribution.

Suppose $Y_i \sim N (\mu , \sigma^2)$.

Then $\theta = \exp (\mu + \frac{1}{2}\sigma^2)$.   (4.3)

The maximum likelihood estimates of μ and σ^2 are

$$\hat{\mu} = \tfrac{1}{n} \sum_{i=1}^{n} Y_i$$

$$\hat{\sigma}^2 = \tfrac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\mu})^2$$

and the maximum likelihood estimate of θ can be obtained by substituting $\hat{\mu}$ and $\hat{\sigma}^2$ into equation (4.3):

$$\hat{\theta} = \exp (\hat{\mu} + \tfrac{1}{2}\hat{\sigma}^2) \tag{4.4}$$

Finney (1941) showed that the maximum likelihood estimate of θ is biased. In order to correct for the bias, Finney introduced the function $g_m (t)$, where

$$g_m (t) = \sum_{k=0}^{\infty} \frac{m^k (m + 2k)}{m (m + 2) \ldots (m + 2k)} \; \frac{t^k}{k!} \tag{4.5}$$

and m is the degrees of freedom associated with $\hat{\sigma}^2$. In this case $m = n-1$.

It can be shown that an unbiased estimate of θ is $\tilde{\theta}$ where

$$\tilde{\theta} = \exp (\hat{\mu}) \, g_m (\tfrac{1}{2}(1-\tfrac{1}{n}) s^2) \tag{4.6}$$

and $\quad s^2 = \dfrac{n}{n-1} \hat{\sigma}^2$ is an unbiased estimate of σ^2.

One advantage of the use of linear models is that standard errors of the parameter estimates can be produced. These can be used to find standard errors on the original unlogged scale. The variance of $\tilde{\theta}$ is τ^2, where

$$\tau^2 = E (\tilde{\theta}^2) - (E (\tilde{\theta}))^2.$$

An unbiased estimate of $E (\tilde{\theta}^2)$ is obviously $\tilde{\theta}^2$ (since the expectation of this is $E (\tilde{\theta}^2)$) and

$$(E (\tilde{\theta}))^2 \qquad = (\exp (\mu + \tfrac{1}{2} \sigma^2))^2$$

402

$$= \exp (2\mu + \sigma^2).$$

By analogy with the unbiased estimation of θ, an unbiased estimate of

$$\exp (2\mu + \sigma^2)$$

is
$$\exp (2\hat{\mu}) \, g_m ((1-\tfrac{2}{n}) \, s^2).$$

Thus an unbiased estimate of τ^2 is

$$\bar{\tau}^2 \qquad = \exp (2\hat{\mu}) \left[(g_m (\tfrac{1}{2} (1-\tfrac{1}{n}) \, s^2))^2 - g_m ((1-\tfrac{2}{n}) \, s^2) \right]. \qquad (4.7)$$

For comparison purposes, the corresponding maximum likelihood estimates are also found. The maximum likelihood estimate of the variance of the maximum likelihood estimate of θ, $\hat{\theta}$, is

$$\exp (2\hat{\mu} + \tfrac{\hat{\sigma}^2}{n}) \left[\exp (\tfrac{\hat{\sigma}^2}{n}) \left[1 - \tfrac{2\hat{\sigma}^2}{n} \right]^{-\frac{1}{2}(n-1)} - \left[1 - \tfrac{\hat{\sigma}^2}{n} \right]^{-(n-1)} \right]. \quad (4.8)$$

4.2 Unbiased Estimation for Claims Runoff Triangles

A claims runoff triangle consisting of incremental claims (assumed positive) is now considered. It is assumed that the data have been adjusted for inflation and exposure. Z_{ij} is incremental claims in row i, column j.

$$\text{Let } \theta_{ij} \qquad = \; E (Z_{ij}). \qquad\qquad\qquad (4.9)$$

Estimates of θ_{ij} are required along with standard errors of these estimates. In particular, estimates of $\{ \; \theta_{ij} \; : \; i{=}1,...,t \; ; \; j{=}t{-}i{+}2,...,t \; \}$ are required, as these are the estimates of the expected outstanding claims. The row totals of the estimates also have to be considered, as these are the estimates of the expected total outstanding claims for each year of business.

$\{ \; Z_{ij} : i{=}1,...,t; j{=}1,...,t{-}i{+}1 \; \}$ are assumed to be independently, lognormally distributed.

Let $Y_{ij} = \log Z_{ij}$. $\qquad\qquad\qquad$ (4.10)

Then Y_{ij} are independently normally distributed.

Suppose that $\{ Y_{ij} : i=1,...,t; j=1,...,t\text{-}i+1 \}$ are modelled by

$$E\ (\ Y_{ij}\ ) = \underline{X}_{ij}\ \underline{\beta} \qquad\qquad\qquad (4.11)$$

$$\text{Var}\ (\ Y_{ij}\ ) = \sigma^2 \qquad\qquad\qquad (4.12)$$

where $\underline{X}_{ij}$ is a row vector of explanatory variables and $\underline{\beta}$ is a column vector of parameters, both of length p.

The linear model for the whole triangle is

$$E\ (\ \underline{Y}\ ) = X\ \underline{\beta} \qquad\qquad\qquad (4.13)$$

where $\qquad$ X is an (nxp) matrix whose rows are $\underline{X}_{ij}$

and $\qquad$ $\underline{Y}$ is the vector of observations.

n is the number of observations (for a triangular array $n=\frac{1}{2}t(t+1)$), and the errors are assumed to be independently, identically normally distributed.

The expected value of the lognormally distributed data, θ_{ij} , is related to the mean and variance of the normally distributed data by

$$\theta_{ij} = \exp\ (\ \underline{X}_{ij}\ \underline{\beta} + \tfrac{1}{2}\ \sigma^2\ ) \qquad\qquad\qquad (4.14)$$

Thus the maximum likelihood estimate of θ_{ij} is

$$\hat{\theta}_{ij} = \exp\ (\ \underline{X}_{ij}\ \underline{\hat{\beta}} + \tfrac{1}{2}\ \hat{\sigma}^2\ ) \qquad\qquad\qquad (4.15)$$

where $\qquad$ $\underline{\hat{\beta}} = (\ X'X\ )^{-1}\ X'\underline{y}$

and $\qquad$ $\hat{\sigma}^2 = \frac{1}{n}\ (\ \underline{y} - X\underline{\hat{\beta}}\ )'(\ \underline{y} - X\underline{\hat{\beta}}\ )$

The general theory of estimation from linear models when the data is lognormally distributed was

404

considered by Bradu and Mundlak (1970). It can be shown that an unbiased estimate of

$$\exp (\underline{Z} \, \underline{\beta} + a\sigma^2)$$

for any row vector $\underline{Z}$ of length p and scalar a, is

$$\exp (\underline{Z} \, \hat{\underline{\beta}}) \, g_m \left[(a - \tfrac{1}{2} \underline{Z} \, (X'X)^{-1} \, \underline{Z}') \, s^2 \right] \qquad (4.16)$$

where s^2 is an unbiased estimate of σ^2 and m is the number of degrees of freedom associated with s^2.

i.e. $s^2 \qquad = \frac{n}{n-p} \, \hat{\sigma}^2$

$$= \frac{1}{n-p} (\underline{y} - X\hat{\underline{\beta}})'(\underline{y} - X\hat{\underline{\beta}})$$

and m $\quad = n - p.$

Thus an unbiased estimate of θ_{ij} is $\tilde{\theta}_{ij}$, where

$$\tilde{\theta}_{ij} \qquad = \exp (\underline{X}_{ij} \, \hat{\underline{\beta}}) \, g_m \left[\tfrac{1}{2} (1 - \underline{X}_{ij} \, (X'X)^{-1} \, \underline{X}'_{ij}) \, s^2 \right]. \qquad (4.17)$$

Note that $\qquad \mathrm{Var} \, (\, \hat{\underline{\beta}} \,) = \, (X'X)^{-1} \, \sigma^2$

and hence $\qquad \mathrm{Var} \, (\, X_{ij} \, \hat{\underline{\beta}} \,) \, = \underline{X}_{ij} \, (X'X)^{-1} \, \underline{X}'_{ij} \, \sigma^2.$

It can therefore be seen that $\underline{X}_{ij} \, (X'X)^{-1} \, \underline{X}'_{ij} \, s^2$ is an estimate of Var ($X_{ij} \, \hat{\underline{\beta}}$).

The variance of the unbiased estimate of θ_{ij} , $\tilde{\theta}_{ij}$, is τ_{ij}^2 , where

$$\tau_{ij}^2 \qquad = \, \mathrm{Var} \, (\, \tilde{\theta}_{ij} \,)$$

$$= \, \mathrm{E} \, (\, \tilde{\theta}_{ij}^2 \,) - (\, \mathrm{E} \, (\, \tilde{\theta}_{ij} \,) \,)^2. \qquad (4.18)$$

An unbiased estimate of E ($\tilde{\theta}_{ij}^2$) is $\tilde{\theta}_{ij}^2$ and

$$(\text{E} (\tilde{\theta}_{ij}))^2 \qquad = \theta_{ij}^2$$
$$= \exp (2\underline{X}_{ij} \, \underline{\beta} + \sigma^2).$$

Hence an unbiased estimate of τ_{ij}^2 is $\tilde{\tau}_{ij}^2$, where

$$\tilde{\tau}_{ij}^2 = \exp(2\underline{X}_{ij}\underline{\hat{\beta}})\Big[(g_m \, (\tfrac{1}{2} \, (1 - \underline{X}_{ij} \, (X'X)^{-1} \, \underline{X}_{ij}' \,)s^2))^2 - g_m((1 - 2\underline{X}_{ij} \, (X'X)^{-1} \, \underline{X}_{ij}' \,)s^2) \Big].$$

$$(4.19)$$

4.3. Unbiased Estimates of Total Outstanding Claims

The purpose of the analysis of the claims data is to produce estimates of the expected total outstanding claims, R_i , for each year of business, and the total outstanding claims, R, for the whole triangle.

An unbiased estimate of R_i is $\tilde{R}_i$, where

$$\tilde{R}_i \qquad = \sum_{j=i-i+2}^{t} \tilde{\theta}_{ij} \, . \qquad (4.20)$$

The variance of $\tilde{R}_i$ can be calculated as follows:

$$\text{Var} (\tilde{R}_i) \qquad = \text{Var} \left[\sum_{j=i-i+2}^{t} \tilde{\theta}_{ij} \right]$$

$$= \sum_{j=i-i+2}^{t} \left[\text{Var} (\tilde{\theta}_{ij}) + 2 \sum_{k=j+1}^{t} \text{Cov} (\tilde{\theta}_{ij} , \tilde{\theta}_{ik}) \right] \qquad (4.21)$$

Now

$$\text{Cov} (\tilde{\theta}_{ij} , \tilde{\theta}_{ik}) = \text{E} (\tilde{\theta}_{ij} \tilde{\theta}_{ik}) - \text{E} (\tilde{\theta}_{ij}) \, \text{E} (\tilde{\theta}_{ik}) ,$$

and an unbiased estimate of this can be obtained using the same method as that which was used to find $\tilde{\tau}_{ij}^2$ in section 4.2.

It can be shown that if

$$\tau_{ijk} \qquad = \text{Cov} \left(\; \tilde{\theta}_{ij} \, , \; \tilde{\theta}_{ik} \; \right) \, ,$$

an unbiased estimate of τ_{ijk} is $\tilde{\tau}_{ijk}$, where

$$\tilde{\tau}_{ijk} = \exp((\underline{X}_{ij} + \underline{X}_{ik})\underline{\hat{\beta}}) \Big[\; g_m(\tfrac{1}{2}(1 - \underline{X}_{ij}(X'X)^{-1}\underline{X}'_{ij})s^2) g_m(\tfrac{1}{2}(1 - \underline{X}_{ik}(X'X)^{-1}\underline{X}'_{ik})s^2)$$
$$-g_m((1 - \tfrac{1}{2}(\underline{X}_{ij} + \underline{X}_{ik})(X'X)^{-1}(\underline{X}_{ij} + \underline{X}_{ik}))s^2) \; \Big]$$

(4.22)

Hence an unbiased estimate of Var ($\tilde{R}_i$) is

$$\sum_{j=t-i+2}^{t} \left[\; \tilde{\tau}_{ij}^2 + 2 \sum_{k=j+1}^{t} \tilde{\tau}_{ijk} \; \right]$$

(4.23)

By extending the limits of the summations, the total outstanding claims for the whole triangle can also be considered.

4.4 Prediction Intervals

Having found an unbiased estimate of total outstanding claims, it is now possible to produce a prediction interval for total outstanding claims. The purpose of the analysis so far has been to produce an estimate of total outstanding claims and an estimate of the variance of this estimate. It is often desirable to find a 'safe' value which is unlikely to be exceeded by the total actual claims.

Let $\qquad$ R = total outstanding claims for the whole triangle

and $\qquad$ $\tilde{R}$ be an unbiased estimate of $E(R)$.

Suppose that a 95% upper confidence bound on R is required. i.e. it is required to find a value, k, such that

$$P(R \le \tilde{R} + k) = 0.95$$

(4.24)

i.e. find k such that

$$P(R - \tilde{R} \le k) = 0.95.$$

(4.25)

Since $\tilde{R}$ is an unbiased estimate of $E(R)$,

$$E(\tilde{R}) = E(R) \qquad (4.26)$$

and hence

$$E(R - \tilde{R}) = 0. \qquad (4.27)$$

Also, $\tilde{R}$ is based on past data and is thus independent of R under the assumptions of the model. Thus

$$Var(R - \tilde{R}) = Var(R) + Var(\tilde{R}). \qquad (4.28)$$

In section 4.3, an unbiased estimate of $Var(\tilde{R})$ was derived and it is possible to derive an unbiased estimate of $Var(R)$ using the theory which was used in that section. By independence,

$$Var(R) = \sum_{i=2}^{t} \sum_{j=t-i+2}^{t} Var(Z_{ij}) \qquad (4.29)$$

and an unbiased estimate of $Var(Z_{ij})$ is required. This can be derived as follows, using the method of section 4.2.

Z_{ij} has a lognormal distribution, and the variance of this distribution is given by:

$$Var(Z_{ij}) \qquad = \exp(2\underline{X}_{ij}\underline{\beta} + \sigma^2)(\exp(\sigma^2) - 1)$$

$$= \exp(2\underline{X}_{ij}\underline{\beta} + 2\sigma^2) - \exp(2\underline{X}_{ij}\underline{\beta} + \sigma^2) \qquad (4.30)$$

Hence, applying equation (4.16), an unbiased estimate of $Var(Z_{ij})$ is

$$\exp(2\underline{X}_{ij}\underline{\hat{\beta}})\Big[g_m\big(2(1-\underline{X}_{ij}(X'X)^{-1}\underline{X}'_{ij})s^2\big) - g_m((1-2\underline{X}_{ij}(X'X)^{-1}\underline{X}'_{ij})s^2)\Big]. \qquad (4.31)$$

It is not inappropriate to use a Normal approximation since R and $\tilde{R}$ are, typically, combinations of a reasonably large number of lognormally distributed random variables. Thus a 95% upper bound on total outstanding claims can be found:

408

$$\tilde{R} + 1.645 \sqrt{\text{Var}(R) + \text{Var}(\tilde{R})} \qquad (4.32)$$

$\sqrt{\text{Var}(R) + \text{Var}(\tilde{R})}$ is the root mean square error of prediction.

4.5 Bayesian Estimation for Claims Runoff Triangles

When a method is used which is based on Bayes theory, Bayesian estimators should be used. The Bayesian estimators have a slightly simpler form than the unbiased estimators and so are sometimes used in their place in a classical analysis. When used in a classical analysis, no prior information is assumed.

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. $\qquad \log Z_{kl} \mid \theta \quad \sim \quad N(\theta, \sigma^2)$

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

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

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

and $\qquad \text{Var}(Z_{kl} \mid D) = e^{2m + \sigma^2 + \tau^2}(e^{\sigma^2 + \tau^2} - 1)$

Similar methods can be used to calculate the covariances, total outstanding claims and the variance of the total outstanding claims.

The Bayes estimate of outstanding claims for year of business i is

$$\sum_{j > n-i+1} E(Z_{ij} \mid D) \qquad (4.33)$$

and the Bayes estimate of the variance is

$$\sum_{j > n-i+1} \left[\text{Var}(Z_{ij} \mid D) + 2\sum_{k > j} \text{Cov}(Z_{ij}, Z_{ik} \mid D) \right] \qquad (4.33)$$

409

4.6. Example

This example illustrates and compares the two most basic methods of claims reserving considered in this thesis: the chain ladder method and the two-way analysis of variance. This gives an opportunity to compare the two. For the analysis of variance model, both the unbiased and maximum likelihood estimates of outstanding claims are given. The data used is taken from Taylor and Ashe (1983), and was given in section 2.

The estimates of the parameters in the chain ladder linear model and their standard errors are shown in table 4.1.

Table 4.1

	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 solely on the design, the standard errors are the same for each row and column

410

parameter. The row parameters are contained within a much smaller range than the column parameters: (0.149, 0.673) compared with (-1.393, 0.965). It can also be seen that there is an indication that the row parameters follow an increasing trend. It is to be expected that the row parameters should be contained within a fairly small range, since the rows are expected to be similar. Any pattern in the row parameters gives an insight into, and depends upon, the particular claims experience. It is thus quite common to observe that the row parameters lie in a small range, but not typical that they follow a trend.

Table 4.2

```
286170  711785  731359  750301  418911  283724  252756  182559  266237  67948
357848  766940  610542  482940  527326  574398  146342  139950  227229  67948


410587  1021245  1049329  1076506  601040  407078  362646  261930  381987
352118   884021   933894  1183289  445745  320996  527804  266172  425046


379337  943516  969461  994572  555294  376094  335044  241994
290507  1001799  926219  1016654  750816  146923  495992  280405


339233   843767  866971   889425  496588  336334  299624
310608  1108250  776189  1562400  272482  352053  206286


378676  941872  967773  992840  554327  375439
443160  693190  991983  769488  504851  470639


389421  968599  995234  1021012   570056
396132  937085  847498   805037   705960


420963  1047052  1075844   1103710
440832   847631  1131398   1063269


457887  1138894  1170213
359480  1061648  1443370


396651  986582
376686  986608


344014
344014
```

The fitted values for the analysis of variance model are shown in table 4.2. These are the unbiased estimates and are shown with the actual observations for comparison. In this figure, and in all similar ones in future, the top entries are the estimates and those underneath are the actual observations.

Table 4.3 shows a plot of residuals (fitted value − actual value) against fitted value.

Table 4.3
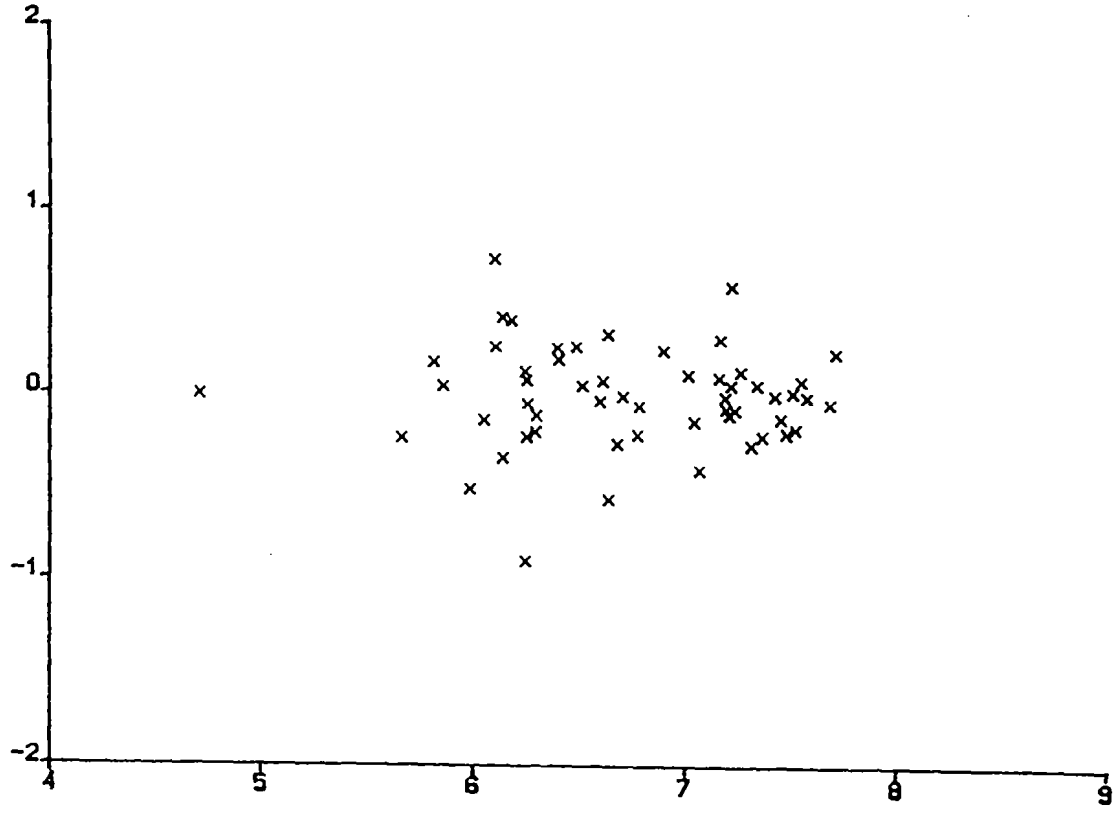
There is no discernible pattern in the residuals, and they seem to be randomly scattered, so there is no cause to question the model on the basis of this plot. Of course, it is possible to look further into a residual analysis and study the plots of residuals against year of business and delay. This is considered further using the GLIM system, by Renshaw (1989). The main purpose of this paper is to extend the possible range of analyses and to consider rigorous statistical estimation, rather than to find the most appropriate linear model to fit, and so the residual analysis will not be commented on further.

Of most interest to practitioners are the predicted outstanding claims for each year of business, which are the row totals of predicted values. Table 4.4 shows the maximum likelihood predictions of the outstanding claims in the lower triangle, and table 4.5 shows the unbiased predictions. The method does not produce any predictions for the first row, and each row contains one more predicted value.

Table 4.4

									101269
								357398	93599
							217465	319835	83761
						335047	243001	357392	93597
					386433	345088	250283	368102	96402
				617309	418743	373941	271209	398880	104462
			1206369	674243	457364	408430	296223	435668	114097
		1026594	1053911	589034	399564	356813	258787	380610	99678
	888831	913640	937951	524224	355600	317554	230313	338732	88710

Table 4.5

								96238
							350362	88841
						215218	313105	79394
					332848	240075	349268	88564
				384305	342028	246696	358900	91006
			613257	415031	369373	266419	387593	98281
		1193906	666126	450811	401216	289387	421005	106752
	1006382	1031734	575643	389575	346716	250077	363813	92248
844677	867203	889047	496032	335695	298762	215487	313486	79483

It can be seen that the maximum likelihood estimates are all higher than the unbiased estimates, as was to be expected.

Table 4.6

	Analysis of Variance		Chain Ladder
Row	Maximum Likelihood	Unbiased	
2	101269	96238	94630
3	450997	439203	464668
4	621061	607717	702101
5	1029037	1010755	965576
6	1446307	1422934	1412202
7	2184544	2149953	2176089
8	3592393	3529202	3897142
9	4164990	4056189	4289473
10	4595556	4339873	4618035

The total predicted outstanding claims for each year of business (the row totals of the predicted outstanding claims) are shown in table 4.6. There are three estimates given, the maximum likelihood and unbiased estimates from the analysis of variance model, and the chain ladder estimate.

It can be seen that the maximum likelihood estimates differ most significantly from the unbiased

which is where the number of observations used in the estimation is the greatest. The maximum likelihood estimate is asymptotically unbiased, and the greater the number of observations used to estimate the parameters, the closer are the two. The chain ladder estimates are sometimes higher and sometimes lower than the analysis of variance estimates. There is nothing significant that can be inferred from the differences. This confirms that the crude chain ladder method is a reasonable 'rough-and-ready' method for calculating outstanding claims, although the more proper method, statistically, is the analysis of variance method (using unbiased estimation).

The total predicted outstanding claims are:

Analysis of	Maximum Likelihood	18186154
Variance	Unbiased	17652064
	Chain Ladder	18619916

The following table shows the unbiased estimates of the total outstanding claims for each year of business, the standard errors of these estimates and the root mean square error of prediction. This table can be used in setting safe reserves, and gives an idea of the likely variation of outstanding . claims.

Table 4.7

Unbiased Estimate	Standard Error	Mean Square Error Of Prediction
96238	35105	47202
439203	108804	163217
607717	127616	182847
1010755	195739	269224
1422934	273082	357593
2149953	429669	538533
3529202	775256	942851
4056189	1052049	1197009
4339873	1534943	1631306

416

The unbiased estimate of total outstanding claims is 17652064 and the root mean square error of prediction is 2759258. Thus a 95% upper bound on total outstanding claims is

$$17652064 + 1.645 \times 2759258 = 22191043$$

This could be regarded as a "safe" reserve for this triangle according to the chain ladder linear model using unbiased estimation.

5. Estimation of the Development Factors

When considering outstanding claims, it is important to use unbiased estimators. However, when comparing several sets of runoff patterns it is simpler to use maximum likelihood theory since unbiasedness is not critical. There are two sets of parameters whose distributions can usefully be found: the development factors, $\{ \lambda_j : j = 2 , \ldots , t \}$, and the proportions of ultimate claims, $\{ S_j : j = 1 , \ldots , t ; \sum_{j=1}^{t} S_j = 1 \}$. It has already been shown that the following relationship between the proportions of ultimate claims and the development factors holds:

$$S_1 = \frac{1}{\sum_{l=2}^{t} \lambda_l} \tag{5.1}$$

$$S_j = \frac{\lambda_j - 1}{\sum_{l=j}^{t} \lambda_l} \qquad (j \geq 2) \tag{5.2}$$

It was also shown by Kremer that the proportions of ultimate claims are related to the column parameters of the linear model as follows

$$S_j = \frac{e^{\beta_j}}{\sum_{l=1}^{t} e^{\beta_l}} \qquad j = 1 , \ldots , t \tag{5.3}$$

where $\beta_1 = 0$ by definition.

417

Finally, the relationship between the parameters of the chain ladder and linear models was proved in Verrall (1991b):

$$\lambda_j = 1 + \frac{e^{\beta_j}}{\sum_{l=1}^{j-1} e^{\beta_l}} \tag{5.4}$$

The parameters of the additive model can be estimated using maximum likelihood estimation. The variance-covariance matrix of the parameter estimates can be obtained from the Fisher information matrix by differentiating the log-likelihood a second time. Further details of the theory of maximum likelihood which is used in this section can be found in Cox and Hinkley (1974).

Since maximum likelihood estimates are invariant under parameter transformations, the maximum likelihood estimates of the development factors and the proportions of ultimate claims can be obtained by substituting the estimates of $\{\ \beta_j : j = 1 \ , \ldots , t \ ; \beta_1 = 0 \ \}$ into equations (5.3) and (5.4). In addition to the parameter estimates, it is useful to have standard errors of the parameter estimates which can be obtained by maximum likelihood theory. The particular advantage of using maximum likelihood estimation is that the second moments are relatively straightforward to obtain. Denoting the variance-covariance matrix of $\{\ \beta_j : j = 1 \ , \ldots , t \ ; \beta_1 = 0 \ \}$ by $V(\ \underline{\beta}\ )$, the variance-covariance matrix of $\{\ \lambda_j : j = 2 \ , \ldots , t \ \}$ and $\{\ S_j : j = 1 \ , \ldots , t \ ; \sum_{j=1}^{t} S_j = 1 \ \}$ are given by

$$V(\ \underline{\lambda}\ ) = \left(\frac{\partial \underline{\lambda}}{\partial \underline{\beta}}\right) V(\ \underline{\beta}\ ) \left(\frac{\partial \underline{\lambda}}{\partial \underline{\beta}}\right) \tag{5.5}$$

and

$$V(\ \underline{S}\ ) = \left(\frac{\partial \underline{S}}{\partial \underline{\beta}}\right) V(\ \underline{\beta}\ ) \left(\frac{\partial \underline{S}}{\partial \underline{\beta}}\right) \tag{5.6}$$

It is thus necessary to obtain the matrices of the first derivatives of the respective parameter vectors. The (j,k)th element of $\left(\frac{\partial \underline{\lambda}}{\partial \underline{\beta}}\right)$ can be obtained from equation (5.4) and is given by:

$$\frac{\partial \lambda_j}{\partial \beta_k} = \begin{cases} 0 & k > j \\[2em] \dfrac{e^{\beta_j}}{\sum\limits_{l=1}^{j-1} e^{\beta_l}} & k = j \\[2em] \dfrac{e^{\beta_j} e^{\beta_k}}{\left(\sum\limits_{l=1}^{j-1} e^{\beta_l} \right)} & k < j \end{cases}$$

$$= \begin{cases} 0 & k > j \\[1.5em] \lambda_j - 1 & k = j \\[1.5em] -(\lambda_j - 1)(\lambda_k - 1) & k < j \end{cases} \qquad (5.7)$$

Similarly, the (j,k)th element of $\left(\dfrac{\partial S}{\partial \underline{\beta}} \right)$ can be obtained from equation (5.3) and is given by:

$$\frac{\partial S_j}{\partial \beta_k} = \begin{cases} -\dfrac{e^{\beta_j} e^{\beta_k}}{\left(\sum\limits_{l=1}^{t} e^{\beta_l}\right)} & k \neq j \\[6mm] \dfrac{e^{\beta_j}\left(\sum\limits_{l=1}^{t} e^{\beta_l} - e^{\beta_j}\right)}{\left(\sum\limits_{l=1}^{t} e^{\beta_l}\right)} & k = j \end{cases}$$

$$= \begin{cases} -S_j S_k & k \neq j \\[4mm] S_j(1 - S_j) & k = j \end{cases} \tag{5.8}$$

Estimates of the variance-covariance matrices can be obtained by substituting estimates of the parameters into equations (5.7) and (5.8).

A technical note is that the parameter β_1 (which is defined to be zero) has to be included in the matrix of partial derivatives in equation (5.8) since there are n parameters in the vector $\underline{S}$. The variance-covariance matrix of the parameters of the additive model which is obtained from a standard least squares analysis has to augmented to include an extra row and column, all of whose entries are zero. This is not necessary for equation (5.7).

5.1 Example

The method described in section 5 is of use when comparing several different sets of data and therefore a different example will be used than in other sections for illustration purposes. The method is applied to six sets of employers' liability data which have been obtained from the DTI returns. The names of the companies to which the data apply have been suppressed, and it should be commented that this mathematical analysis is only one part of the process by which reserves are set. In particular, the DTI data are gross of reinsurance. The results here should therefore be regarded as a statistical analysis which would give further information to the claims reserver who would use the other information available.

We now consider the parameter estimates for each of the three models in turn. Beginning with the additive model the estimates of the column parameters { β_j : j = 2 , ... , t } and their standard errors are given in the following figure:

COMPANY:

1	2	3	4	5	6
1.796 0.121	1.748 0.148	2.236 0.249	1.846 0.248	1.941 0.201	2.010 0.082
1.848 0.126	1.857 0.155	2.080 0.261	2.260 0.260	2.248 0.211	2.246 0.086
1.669 0.133	1.654 0.163	1.978 0.273	2.159 0.272	2.204 0.221	2.129 0.091
1.413 0.139	1.400 0.171	1.725 0.287	1.986 0.286	1.981 0.232	1.863 0.095
0.994 0.147	1.200 0.180	1.535 0.303	1.535 0.302	1.514 0.245	1.485 0.100
0.615 0.155	0.705 0.190	1.057 0.320	1.235 0.319	0.788 0.259	1.050 0.106
0.415 0.164	0.339 0.201	0.667 0.338	0.644 0.337	0.227 0.274	0.782 0.112
0.038 0.175	0.025 0.215	-0.099 0.360	0.222 0.359	-0.540 0.291	0.234 0.120
-0.812 0.189	-0.407 0.232	-0.300 0.390	0.047 0.388	-0.993 0.315	0.155 0.129
-0.915 0.212	-1.821 0.260	-0.715 0.437	0.382 0.435	-1.311 0.353	-0.324 0.145
-2.513 0.264	-1.492 0.323	-1.708 0.543	-0.896 0.541	-3.206 0.439	-0.304 0.180

Before going on to the parameters which have a physical interpretation, it should be noticed that it is already possible to see some differences between the companies. In particuar, the standard errors

421

of the parameters are larger for some companies (3 and 4) than for others (6). This will be mirrored in the parameter estimates and standard errors of the other models.

Next, consider the chain ladder model. The estimates of the development factors $\{\lambda_j: j = 2, \ldots, t\}$ and their standard errors are given in following table:

COMPANY:

1		2		3		4		5		6	
7.027	0.727	6.742	0.850	10.36	2.327	7.332	1.569	7.963	1.401	8.466	0.616
1.904	0.101	1.950	0.130	1.773	0.181	2.307	0.300	2.189	0.222	2.117	0.086
1.397	0.041	1.398	0.050	1.394	0.084	1.512	0.109	1.520	0.090	1.469	0.033
1.220	0.022	1.221	0.027	1.219	0.046	1.285	0.059	1.274	0.046	1.245	0.017
1.119	0.012	1.148	0.019	1.149	0.032	1.141	0.030	1.135	0.023	1.135	0.009
1.073	0.008	1.079	0.010	1.080	0.018	1.092	0.020	1.057	0.010	1.077	0.006
1.055	0.006	1.051	0.007	1.050	0.012	1.047	0.011	1.031	0.006	1.055	0.004
1.036	0.005	1.035	0.006	1.022	0.006	1.029	0.008	1.014	0.003	1.030	0.003
1.015	0.002	1.022	0.004	1.018	0.005	1.024	0.007	1.009	0.002	1.027	0.003
1.013	0.002	1.005	0.001	1.012	0.004	1.032	0.011	1.006	0.002	1.016	0.002
1.003	0.001	1.007	0.002	1.004	0.002	1.009	0.004	1.001	0.000	1.016	0.003

Finally, consider the multiplicative model. The estimates of the proportions of ultimate claims in each development year $\{ S_j : j = 1, \ldots, t ; \sum_{j=1}^{t} S_j = 1 \}$ and their standard errors are given in the following table:

COMPANY:

1		2		3		4		5		6	
0.032	0.003	0.032	0.004	0.023	0.005	0.021	0.005	0.023	0.004	0.022	0.002
0.196	0.016	0.184	0.019	0.218	0.036	0.135	0.025	0.162	0.023	0.162	0.010
0.206	0.015	0.205	0.019	0.186	0.029	0.204	0.032	0.220	0.027	0.205	0.011
0.172	0.013	0.167	0.015	0.168	0.026	0.184	0.028	0.211	0.025	0.182	0.009
0.133	0.011	0.130	0.013	0.131	0.021	0.155	0.024	0.169	0.022	0.140	0.007
0.088	0.008	0.106	0.011	0.108	0.019	0.099	0.017	0.106	0.015	0.096	0.006
0.060	0.006	0.065	0.007	0.067	0.013	0.073	0.014	0.051	0.008	0.062	0.004
0.049	0.005	0.045	0.006	0.045	0.010	0.040	0.009	0.029	0.005	0.047	0.003
0.034	0.004	0.033	0.005	0.021	0.005	0.027	0.007	0.014	0.003	0.027	0.002
0.014	0.002	0.021	0.004	0.017	0.005	0.022	0.006	0.009	0.002	0.025	0.002
0.013	0.002	0.005	0.001	0.011	0.004	0.031	0.011	0.006	0.002	0.016	0.002
0.003	0.001	0.007	0.002	0.004	0.002	0.009	0.004	0.001	0.000	0.016	0.002

The runoff patterns of the companies can be compared using the two tables above. For example, 1 and 2 seem quite similar, and some of the companies have more runoff in later development years than others. The standard errors can also be compared, with the same conclusions as above.

6. Bayesian Linear Models and Credibility Theory

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 of 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 least-squares estimation of the parameters.

In the second case, we will be using empirical priors. 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 considered and 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.

For consistency, the constraints

$$\alpha_1 = \beta_1 = 0$$

424

on the first stage distribution have been retained. This also facilitates the comparison with the recursive approaches such as that based on the Kalman filter. It does, however, introduce a slight degree of assymmetry into the prior distribution and it might be considered more appropriate to use a constraint such as

$$\sum \alpha_i = \sum \beta_j = 0.$$

It is also possible to apply the constraint at the second stage and use the following prior distribution:

$$\alpha_i \sim N(0, \sigma_\alpha^2) \qquad i = 1, \ldots, t.$$

The affect of the exchangeability assumption is similar whichever constraint is used: the estimates are shrunk towards a central value and stability is introduced. The amount of shrinkage is greatest where the number of observations is small.

6.1 Bayes Estimates for the Analysis of Variance Model

In this section the use of two-stage Bayesian linear models which assume that there is some prior information is described. A prior distribution can be written down using the prior knowledge. It was shown earlier that the chain ladder linear model can be written as a linear model in the form:

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

and the prior information is quantified in the prior distribution on $\underline{\beta}$

$$\underline{\beta} | \underline{\theta}_l \qquad \sim \qquad N(A_1 \underline{\theta}_l, C_1)$$

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:

425

$$\underline{\beta}\,|\underline{\theta}_1 \quad \sim \quad N(\underline{\theta}_1,\, C_1) \qquad (6.1)$$

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

The Bayes estimate of the parameter vector is the solution, $\tilde{\underline{\beta}}$, of

$$(\sigma^{-2} X'X \; + \qquad C_1^{-1}\,)\,\tilde{\underline{\beta}} \;\; = \;\; \sigma^{-2}\,X'X\,\hat{\underline{\beta}} + C_1^{-1}\,\underline{\theta}_1 \qquad (6.2)$$

and the variance-covariance matrix of the estimates is

$$\mathrm{Var}(\tilde{\underline{\beta}}) \qquad = \; [\,\sigma^{-2}\,X'X + C_1^{-1}\,]^{-1} \qquad (6.3)$$

The equation for $\tilde{\underline{\beta}}$ (3.4.2) can be written as a credibility formula:

$$\tilde{\underline{\beta}} \qquad = \; z\,\hat{\underline{\beta}} \; + \; (1-z)\,\underline{\theta}_1 \qquad (6.4)$$

where $z \; = \; (\sigma^{-2} X'X + C_1^{-1})^{-1}\,\sigma^{-2} X'X$ is the credibility factor.

It is interesting to note that the credibility factor has been generalised into a credibility matrix, since z is a (pxp) matrix. There will be a subtle dependence of the elements in the Bayes estimator $\tilde{\underline{\beta}}$ on each of the elements in the least squares estimator. It is not possible to write a credibility formula separately for each factor in the form

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

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

$$s^2 \qquad = \; (\underline{y} - X\tilde{\underline{\beta}})'(\underline{y} - X\tilde{\underline{\beta}})/(n+2) \qquad (6.5)$$

Thus the equations which give the Bayes estimates are (6.2), with σ^2 replaced by s^2, and (6.5).

426

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

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

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

6.2 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 a three-stage Bayesian model described in Verrall (1990) to derive empirical Bayes estimates for the 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

$$\underline{y}|\underline{\beta} \qquad \sim \qquad N(X\underline{\beta} \; , \; \sigma^2 I) \tag{6.6}$$

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 $\alpha_2, \; \alpha_3, \; ... \; , \alpha_t$: they are

assumed to be independent observations from a common distribution. For the overall mean, μ, and the column parameters β_2, β_3, ... ,β_t , the same distributional assumptions as for ordinary maximum likelihood estimation will be used. Thus at the second stage

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

and take $\sigma_\mu^{-2} \to 0$ and $\sigma_\beta^{-2} \to 0$. \hfill (6.7)

ψ 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 ψ. Since $\sigma_\mu^{-2} \to 0$ and $\quad \sigma_\beta^{-2} \to 0$, 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.

The Bayes estimate of $\underline{\beta}$, $\underline{\tilde{\beta}}$, is given by

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

\hfill (6.8)

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

and has a credibility interpretation.

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 μ, β_2, ... , β_t involves the estimates of α_2, ... ,α_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

$$s^2 \quad = \quad \frac{\nu\lambda+(\underline{y}-X\underline{\tilde{\beta}})'(\underline{y}-X\underline{\tilde{\beta}})}{n+\nu+2} \tag{6.9}$$

$$s_\alpha^2 \quad = \frac{\nu_\alpha\lambda_\alpha+\sum_{i=2}^{t}(\tilde{\alpha}_i-\bar{\alpha}.)^2}{t+\nu_\alpha+1} \tag{6.10}$$

where ν, λ, ν_α and λ_α are set by the prior distribution of the variances. The derivation of these formulae, and the discusion of the prior parameter values is given in Lindley and Smith (1972).

Again, the estimates are obtained by iterating between (6.8) and (6.9),(6.10). 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.

6.3. Example

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 6.1:

Table 6.1

	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 graph given in section 7.3 which shows a plot of the maximum likelihood and empirical Bayes estimates of the row parameters, togther with the estimates from the dynamic model discussed in section 7.

Table 6.2 shows the row totals and their standard errors. For comparison purposes, the Bayes estimates with no prior assumptions are also given.

Table 6.2

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

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 benn quite considerably reduced, reflecting the reduction in the estimates of the row 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.

7. State Space Models

The previous section described the empirical Bayes framework in which it is assumed that the row parameters have the same prior mean. The advantage of this assumption is the connection made between the accident years. The chain ladder technique suffers from over-parameterisation which is a result of the accident years being regarded as cimpletely separate. The empirical Bayes assumption is one way of overcoming this. Another way of tackling this problem, and in some ways a superior way, is to use a state space approach. This method assumes a recursive connection between the rows, rather than the static assumption made by the empirical Bayes method that all the rows are similar. The state space model assumes that each accident year is similar to the previous one. Just how similar can be governed by the choice of a parameter variance. Section 7.1 describes the state space approach to the chain ladder linear model.

Another problem with the chain ladder technique is, paradoxically, that it makes too much connection between the accident years. It does this by assuming that the shape of the run-off is the same for all accident years: the same development parameters are used. It is also possible to relax this assumption, and details of this are given in section 7.2.

7.1 A state space representation of the chain ladder linear model.

In order to consider the state space model and dynamic estimation methods, it is necessary to set up the two-way analysis of variance model in a recursive form. This takes advantage of the natural causality of the data. The data which makes up the claims runoff triangle are received in the form :

$$Z_{1,1}, \begin{bmatrix} Z_{1,2} \\ Z_{2,1} \end{bmatrix}, \begin{bmatrix} Z_{1,3} \\ Z_{2,2} \\ Z_{3,1} \end{bmatrix}, \cdots, \tag{7.1}$$

and in year t the data which are received are

$$\begin{bmatrix} Z_{1,t} \\ Z_{2,t-1} \\ \vdots \\ Z_{t,1} \end{bmatrix} \tag{7.2}$$

The set of data vectors which together make up the whole triangle form a time series:

432

$$Z_1 , Z_2 , \ldots , Z_t , \ldots$$

In this time series, the data vector expands with t: for a triangular set of data, dim (Z_t) = t.

If the data are in the shape of a rhombus, which occurs when the early years of business are fully run off, then Z_t will reach a point when its dimension does not increase.

The analysis can be approached from in the context of multivariate time series. However, the special relationships between the elements of consecutive data vectors mean that it is simpler to generalise the theory of classical and Bayesian time series to two-dimensional processes. For a fuller discussion of the use of classical time series, the reader is referred to Verrall (1989).

There are two methods for calculating the forecast values and their standard errors. The simplest is to use the final parameter estimates and variance-covariance matrix as would be the case in a standard least-squares analysis. The more proper method calculates one-step-ahead , two-step-ahead , ... , (t-1)-steps-ahead forecasts at time t and their variance-covariance matrices. However, since the recursive approaches do not store covariances between, for example, the one-step-ahead and the (t-1)-step-ahead forecasts, the calculation of the variances of the forecasts causes problems. For this reason the first method will be used.

The chain ladder linear model takes the following form when three years' data have been received:

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

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

When the data are handled recursively, the model becomes:

$$Y_{1,1} = \mu + e_{1,1}$$

$$\begin{bmatrix} Y_{1,2} \\ Y_{2,1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_{1,2} \\ e_{2,1} \end{bmatrix}$$

$$\begin{bmatrix} Y_{1,3} \\ Y_{2,2} \\ Y_{3,1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \alpha_3 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} e_{1,3} \\ e_{2,2} \\ e_{3,1} \end{bmatrix} \qquad (7.3)$$

etc.

In general, the state vector at time t is defined by:

$$\underline{\theta}_t = \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \vdots \\ \alpha_t \\ \beta_t \end{bmatrix} \qquad (7.4)$$

and equation (7.3) is called the observation equation. The state vector at time t is related to the state vector at time t-1 by the system equation. A recursive version of the chain ladder method is achieved by defining the system equation matrices as

$$\underline{\theta}_{t+1} = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 \\ 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 \end{bmatrix} \underline{\theta}_t + \begin{bmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \underline{u}_t \qquad (7.5)$$

where $\underline{u}_t$ contains the prior distribution of $\begin{bmatrix} \alpha_{t+1} \\ \beta_{t+1} \end{bmatrix}$.

The new parameters at time t+1 are $\begin{bmatrix} \alpha_{t+1} \\ \beta_{t+1} \end{bmatrix}$ and equation (7.5) implies that the existing parameters are unchanged, while the new parameters are treated as stochastic inputs.

If the variance of the errors, e_{ij} , is known and vague priors are used for the parameters, this method gives exactly the same results as ordinary least-squares estimation. It has the advantage that the data can be handled recursively. Also, it gives a method of implementing Bayesian estimation on some or all of the parameters. It has been assumed that the prior estimates of the parameters are uncorrelated: in other words that the stochastic input vector, $\underline{u}_t$, and the state vector, $\underline{\theta}_t$, are independent.

The equations above are an example of a state space system; a more general form is now considered. The models for $\underline{Y}_1$, $\underline{Y}_2$, ... , $\underline{Y}_t$, ... , which together make up the triangle can be written as

$$
\begin{aligned}
\underline{Y}_1 &= F_1\underline{\theta}_1 + \underline{e}_1 \\
\underline{Y}_2 &= F_2\underline{\theta}_2 + \underline{e}_2 \\
&\vdots \\
\underline{Y}_t &= F_t\underline{\theta}_t + \underline{e}_t
\end{aligned} \tag{7.6}
$$

where $\quad \underline{Y}_t = \log \underline{Z}_t$

Equation (7.6) is an observation equation and forms one part of a state system to which the Kalman filter can be applied in order to obtain recursive estimates of the parameters. $\underline{\theta}_t$ is known as the state vector and is related to $\underline{\theta}_{t-1}$ by the system equation. The observation equation and the system equation together make up the state space representation of the analysis of variance model.

The system equation relates $\underline{\theta}_t$ to $\underline{\theta}_{t-1}$ and defines how the state vector evolves with time. Thus, the time evolution of the system is defined on the state vector and the observation vector is then related to the state vector by the observation equation. There are many choices of system equation, the most general being:

$$
\underline{\theta}_{t+1} = G_t \underline{\theta}_t + H_t \underline{u}_t + \underline{w}_t \tag{7.7}
$$

where $\quad \underline{u}_t$ is a stochastic input vector

and $\quad \underline{w}_t$ is a disturbance vector

and the distributions of $\underline{u}_t$ and $\underline{w}_t$ are:

$$\underline{u}_t \sim N(\hat{\underline{u}}_t, U_t)$$

$$\underline{w}_t \sim N(\underline{0}, W_t)$$

The choices of G_t, W_t and the distribution of $\underline{u}_t$ govern the dynamics of the system.

Suppose $\quad \underline{\theta}_t \mid (\underline{Y}_1, \underline{Y}_2, \dots, \underline{Y}_{t-1}) \quad \sim \quad N(\hat{\underline{\theta}}_{t|t-1}, C_t).$ \hfill (7.8)

i.e. the distribution of the parameters, based on the data up to time $t-1$ is normal with mean $\hat{\underline{\theta}}_{t|t-1}$ and variance-covariance matrix C_t.

From equations (7.6) and (7.7), the distribution of $\underline{Y}_t$ given information up to time $t-1$ is

$$\hat{\underline{Y}}_{t|t-1} \sim N(F_t \hat{\underline{\theta}}_{t|t-1}, F_t C_t F_t' + V_t) \tag{7.9}$$

When the observed value of $\underline{Y}_t$ is received, the state estimate can be updated to $\hat{\underline{\theta}}_{t|t}$ and the distribution of the state vector at time t forecast using equation (7.8).

The recursion is given by the following equations, a proof of which can be found in (for example) Davis & Vinter (1985). If the system and observation equations are given by equations (7.6) and (7.7), and the distribution of $\underline{\theta}_t$ given information at time t-1 is given by (7.9), then the distribution of the state vector can be updated when $\underline{Y}_t$ is received using the following recursion:

$$\hat{\underline{\theta}}_{t+1|t} = G_t \hat{\underline{\theta}}_{t|t-1} + H_t \hat{\underline{u}}_t + K_t(\underline{Y}_t - \hat{\underline{Y}}_t) \tag{7.10}$$

where $\quad K_t = G_t C_t F_t'(F_t C_t F_t' + V_t)^{-1}$ \hfill (7.11)

$$C_{t+1} = G_t C_t G_t' + H_t U_t H_t' - G_t C_t F_t'(F_t C_t F_t' + V_t)^{-1} F_t C_t G_t' + W_t \tag{7.12}$$

and $\quad \hat{\underline{Y}}_t = F_t \hat{\underline{\theta}}_{t|t-1}$ \hfill (7.13)

7.3 Dynamic Estimation of the Row Parameters

A model which applies dynamic estimation to the row parameters has the following system equation:

$$\underline{\theta}_{t+1} = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ 0 & \dots & 1 & 0 \\ 0 & \dots & \dots & 0 \end{bmatrix} \underline{\theta}_t + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ 1 \end{bmatrix} u_t + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{bmatrix} w_t$$

where u_t has the prior distribution of β_{t+1}

and w_t is a disturbance term.

Thus the new row parameter, α_{t+1} , is related to α_t by:

$$\alpha_{t+1} = \alpha_t + w_t \tag{7.14}$$

and a sophisticated smoothing method is produced.

The row parameters are related recursively and the column parameters are left as they were if their · prior distribution is vague (although the estimates change because of the change in the estimation of the row parameters). The state variance is set as 0.0289, in order to compare with the empirical Bayes procedure. The practitioner is free to choose this variance as he sees fit: the larger the variance, the less connection is made between the rows. It is also possible to let this variance depend on t, and thus allow the amount of smoothing to be controlled by the perceived changes in the claims experience. It is also possible to obtain an estimate of this variance from the data, using maximum likelihood estimation. In the case in which the variance is set as 0.0289, the parameter estimates are as follows:

Table 7.1

| Parameter	Estimate	Standard Error
μ	6.119	0.163
α_2	0.187	0.151
α_3	0.170	0.148
α_4	0.196	0.152
α_5	0.296	0.158
α_6	0.396	0.164
α_7	0.482	0.171
α_8	0.550	0.183
α_9	0.536	0.202
α_{10}	0.546	0.238
β_2	0.906	0.158
β_3	0.940	0.165
β_4	0.951	0.173
β_5	0.364	0.183
β_6	-0.028	0.195
β_7	-0.145	0.212
β_8	-0.457	0.236
β_9	-0.062	0.278
β_{10}	-1.406	0.378

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

Table 7.2

Row	Predicted Outstanding Claims	Standard Error
2	109955	59278
3	491787	187134
4	686441	206954
5	1076957	277762
6	1486991	347441
7	2217311	491998
8	3309887	744931
9	4545466	1048855
10	4591188	1169469

The predicted overall total outstanding claims is 18515984 and the standard error of this estimate is 2660211. The standard error is lower than that when no prior knowledge is assumed because of the recursive relationship between the parameters. The effect of the Kalman filter on the parameter estimates will be illustrated by a graph, but it is interesting to compare the results with the empirical Bayes approach.

The following graph shows the parameter estimates for three cases: the model with no prior knowledge, the empirical Bayes model and the state space model. It can be seen from the graph that the state space model and empirical Bayes estimates have both smoothed the estimates of the row parameters to a certain degree. The empirical Bayes estimates have been drawn towards the overall estimate, with the amount of change depending on the data through the variation in each row and between the rows. The differences in the estimates of the row parameters has affected the estimates of outstanding claims. The standard errors have been reduced because the estimation has involved more of the data for each parameter. This is a beneficial affect of any of the Bayesian methods.
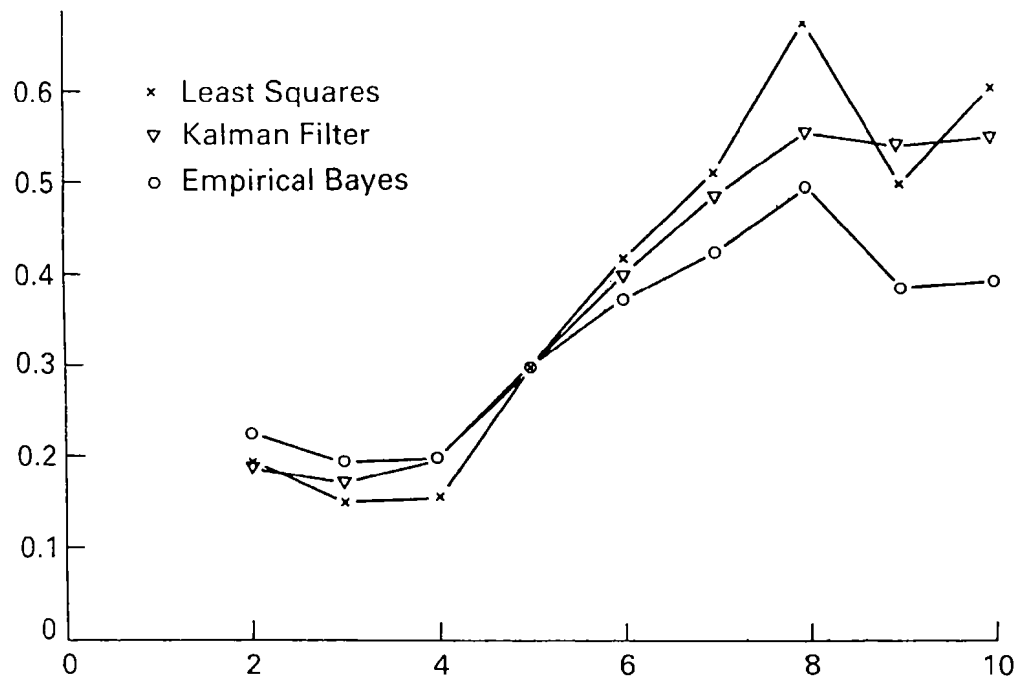
7.4 Dynamic estimation of the development factors

It is well-known that the chain ladder technique assumes that the shape of the run-off curve is the same for each accident year, since the same development factors are used. However, it is doubtful whether this is justified in practice. It is likely that there will be a similarity between the run-offs in successive accident years, and it is possible to formulate a state space model to allow this without imposing an identical shape for each year. The basic chain ladder linear model is

$$E (Y_{ij}) = \mu + \alpha_i + \beta_j \qquad (7.15)$$

Allowing the devolpment factors to be completely separate for each accident year would lead to the following model:

$$E (Y_{ij}) = \mu + \alpha_i + \beta_{ij} \qquad (7.16)$$

We would expect the parameters β_{ij} to be similar for succesive values of i and so we impose the model

$$\beta_{i+1,j} = \beta_{ij} + \text{stochastic disturbance} \qquad (7.16)$$

The variance of the stochastic disturbance can be treated in much the same way as for the row parameters in section 7.3. We can now allow the shape of the run-off to vary from accident year to accident year by the choice of the variance of this stochastic disturbance. If it is zero, the run-off pattern is the same in each accident year and as it increases, the connection becomes less significant. We can allow the variance to depend on t and input a large value for one time point if it is believed that there has been a sudden change in the run-off pattern.

To illustrate the effect of this model, we analyse the data given in section 2, with the variance in equation (7.16) taken as 0.01. The main interest in this case is the effect on the run-off pattern, and so table 7.3 gives just the column parameters, β_{ij}.

Table 7.3

Column parameters from model with the same run-off in each row (from table 7.1):

| 0.906 | 0.940 | 0.951 | 0.364 | -0.028 | -0.145 | -0.457 | -0.062 | -1.406 |

Column parameters from model with the same dynamic run-off pattern:

0.925	0.886	0.914	0.383	0.025	-0.175	-0.479	-0.074	-1.413
0.917	0.895	0.945	0.361	-0.035	-0.135	-0.460	-0.063
0.920	0.907	0.964	0.361	-0.080	-0.130	-0.447
0.918	0.920	0.980	0.332	-0.050	-0.161
0.895	0.942	0.951	0.352	-0.026
0.894	0.960	0.940	0.375
0.890	0.990	0.944
0.898	1.014
0.897

This illustration shows how changes in the run-off pattern can be observed. For example, the first column parameter is generally decreasing and the second one is increasing.

8. Conclusions

This paper has explored the various models which are available within the framework of the chain ladder linear model. It is envisaged that the practitioner will find all of these of use. The following points are of particular note.

Firstly, any of the Bayesian methods will improve upon the least squares (or uninformative prior) approach on the basis of parameter stability. This is because more information is used in estimating each parameter. For example, in the least squares case, there is only one data point from which to estimate the last row parameter; the Bayesian methods use the data from the other rows as well. To illustrate the affect of this consider a change in the data point in the last row from its present value of 344014 to 544014. The following table shows the predicted outstanding claims for each row from the different models. The first column shows the original results with no prior information.

Table 8.1

Row	Original Results			Revised Results		
	No prior Information	Dynamic Estimation	Empirical Bayes	No prior Information	Dynamic Estimation	Empirical Bayes
2	110927	109955	109448	110927	109958	110094
3	482157	491787	479568	482157	491822	481329
4	660810	686441	655656	660810	686637	657998
5	1090752	1076957	1033109	1090752	1078058	1039692
6	1530532	1486991	1388261	1530532	1491978	1400466
7	2310959	2217311	2002772	2310959	2239482	2024720
8	3806976	3309887	3018896	3806976	3399256	3063229
9	4452396	4545466	3780759	4452396	4847221	3819051
10	5066116	4591188	3811869	8011412	5261069	4411270

The last row prediction using no prior information has changed in proportion with the change in the data point. The other methods have dampened down this change because they use more information in the estimation of the parameter. They therefore exhibit greater predictor stability.

It is important to realise that the results must be used correctly. For example, it is often not necessary to produce a 95% upper confidence bound (a 'safe' reserve) on oustnanding claims for each row, but only for the whole triangle, although the 'safe' reserve for the whole triangle may be allocated among the rows. This is important since it can be seen that the standard errors for each row are, in general, relatively large. The standard error of the overall total is more reasonable. To extend this further, the practitioner may be required to set a 'safe' reserve for the whole company, rather than for each triangle; this would reduce the relative size of the standard error still further.

There are now a number of Bayesian methods which are available to the claims reserver, all of which have particular advantages over the classical estimation method. The chain ladder linear model represents a great step forward from the crude chain ladder technique and has opened the way to more sophisticated estimation methods.

9. References

B. Ajne (1989) *Exponential Run-off* Claims Reserving Manual, Vol.2, Institute of Actuaries.

D.Bradu and Y.Mundlak (1970) Estimation in Lognormal Linear Models. JASA Vol 6?, pp 198-211

H.Buhlmann (1967) *Experience Rating and credibility.* Astin Bulletin, Vol.4. No.3, pp 199-207.

S.Christofides (1990) *Regression Models based on Log-incremental Payments* Claims Reserving Manual, Institute of Actuaries, London.

D.R.Cox and D.V.Hinkley (1974) *Theoretical Statistics.* Chapman and Hall, London.

M.H.A.Davis & R.B.Vinter (1985) *Stochastic Modelling and Control.* Chapman & Hall.

D.J.Finney (1941) On the Distribution of a Variate whose logarithm is Normally Distributed. JRSS Suppl. 7 pp 155-61

Goovaerts and Hoogstad (1987) *Credibility Theory.* Surveys of Actuarial Studies , No.4.

S.Klugman (1989) *Credibility for Classification Ratemaking via the Hierarchical Normal Linear Model.* Proc. of Casualty Actuarial Society, Vol. 74, pp 272-321.

E.Kremer (1982) *IBNR-Claims and the Two-Way Model of ANOVA.* Scand. Act. J., Vol.1, pp 47-55.

D.V.Lindley and A.F.M.Smith (1972) *Bayes Estimates for the Linear Model (with Discussion)* JRSS, Series B, Vol. 34, No.1 pp 1-41.

A.Renshaw (1989) *Chain Ladder and Interactive Modelling* J.I.A. Vol.116, pp 559-587.

A.F.M.Smith (1973) *A General Bayesian Linear Model.* JRSS, Series B, Vol.35, No.1, pp67-75.

G.C.Taylor and F.R.Ashe (1983)   *Second Moments of Estimates of Outstanding Claims.* J. of Econometrics, Vol.23, pp 37-61.

R.J.Verrall (1989) *Modelling Claims Runoff Triangles with Two-Dimensional Time Series,* Scand. Act. J, pp 129-138.

R.J.Verrall (1990) *Bayes and Empirical Bayes Estimation for the Chain Ladder Model* Astin Bulletin, Vol.20, No2, pp217-243.

R.J.Verrall (1991a) *On the Unbiased Estimation of Reserves from Loglinear Models* Insurance: Mathematics and Economics, Vol. 10 pp75-80.

R.J.Verrall (1991b) *Chain Ladder and Maximum Likelihood* J.I.A. Vol.118, pp 489-499.

F.de Vylder (1982) *Estimation of IBNR claims by Credibility Theory.* Insurance: Mathematics and Economics 1, pp 35-40.

B.Zehnwirth (1985) Interactive Claims Reserving Forecasting System, Benhar Nominees Pty Ltd., Tunawarra, NSW, Australia.