JIA 116 (1989) 589-609

A STATE SPACE REPRESENTATION OF THE CHAIN LADDER LINEAR MODEL

By R. J. Verrall, M.A., M.Sc., F.S.S.
(of the City University, London)

[Presented at the Seminar, 'Applications of Mathematics in Insurance, Finance and Actuarial Work', sponsored by the Institute of Mathematics and Its Applications, the Institute of Actuaries, and the Faculty of Actuaries, held at the Institute of Actuaries, 6-7 July 1989.]

1. INTRODUCTION

In a recent paper, Kremer (1982)[2] has shown how the classical chain ladder method for estimating outstanding claims on general insurance business is strongly related to a two-way analysis of variance. It can be argued that the estimation methods in a standard chain ladder analysis are inefficient from a statistical viewpoint and that an analysis of variance is more appropriate. Once the chain ladder method is identified with a standard statistical method, the well-known statistical theory can be used to the advantage of the claims reserver. For a further discussion of the use of main stream statistical theory applied to the least squares estimation of the linear model which is close to the chain ladder method, the reader is referred to Renshaw (1989)[4].

In this paper, the analysis of variance model is used (in a slightly different form from that given in Kremer[2]) as a basis for a method which allows the practitioner to enter prior information or to estimate the parameters dynamically. A Bayesian method is used and the data are analysed recursively. The method uses the Kalman filter: a full specification and discussion of the different modelling possibilities will be given. The Bayesian estimation of the parameters of the analysis of variance model using a non-recursive method has been derived by Verrall (1988)[6], in which paper the theory is extended to include empirical Bayes or credibility theory estimation. A comparison between the state space representation and the credibility type analysis will be made.

The machinery for recursive estimation is based on the Kalman filter, and has been used in a claims reserving context by de Jong & Zehnwirth (1983)[3]. The present paper uses the same basic form of the Kalman filter, but concentrates exclusively on the two-way analysis of variance model which was not discussed by de Jong & Zehnwirth[3]. The various modelling assumptions will be discussed in detail, including the case of distinct parameters and static Bayesian estimation. It is, of course, also possible to relate the parameters to each other recursively and use a dynamic estimation method, and this will also be described and compared with the empirical Bayes method. In order to incorporate prior information into a static model, stochastic input vectors have to be introduced which contain the prior distributions of the new parameters introduced at each stage.

The methods described in this paper should be of use to practitioners who are interested in more sophisticated methods of claims reserving which retain the same basic intuitive appeal as the chain ladder technique. The particular use of the recursive Bayesian estimation method is that it allows the practitioner to incorporate information from other sources such as collateral data sets. The dynamic Kalman filter estimation does not necessarily require prior estimates of the parameters, but it does need the state and observation variances to be specified. There are no variance specifications required for the empirical Bayes method. It will be seen that these last two methods give estimates which are more stable than those from ordinary least-squares estimation.

2. THE MODEL AND PARAMETER ESTIMATION

This section follows Sections 2 and 3 of Kremer[2]. The claims run-off triangle consists of data indexed by two variables: the first represents the year in which the business was written, and the second the delay until a claim is made. Hence X_{ij} represents claims on business written in year i with delay index j and X_{ij} are the incremental claims data, not the cumulative data.

(Note that $j \in \{1,2,3, \ldots\}$ and is an index only—it does not necessarily equal the delay.)

The triangle takes the form

$$\begin{array}{llll} X_{11} & X_{12} & X_{13} & X_{14} \\ X_{21} & X_{22} & X_{23} & \\ X_{31} & X_{32} & & \\ X_{41} & & & \end{array}$$

(Note that there is no loss of generality by considering a triangle—the methods apply equally to other shapes, e.g. a rhombus.)

After the business has been running for t years, the data available are

$$\{X_{ij}: j \leq t-i+1,\ i \leq t\} \tag{2.1}$$

and it is assumed that $X_{ij} > 0$, $\forall i$, $\forall j$.

The model which is applied to the raw data has a multiplicative form:

$$X_{ij} = U_i\, S_j\, R_{ij} \tag{2.2}$$

where $E(R_{ij}) = 1$

U_i is a parameter for row i

and S_j is a parameter for row j.

R_{ij} are random errors: the error term is assumed to be multiplicative.

A natural assumption is that the data have a log-normal distribution and this implies that a logarithmic transformation is appropriate:

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

Now if Y_{ij} is assumed to have a normal distribution, X_{ij} has a log-normal distribution. By taking logs of (2.2), the following model is arrived at:

$$Y_{ij}=\mu+\alpha_i+\beta_j+e_{ij} \tag{2.3}$$

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

The parameters have also undergone a logarithmic transformation. Kremer defines μ as the mean of the log U_is and log S_js, so that the restriction

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

An alternative assumption is that $\alpha_1=\beta_1=0$. In this case

$$\alpha_i=\log U_i-\log U_1 \tag{2.4}$$

$$\beta_j=\log S_j-\log S_1 \tag{2.5}$$

$$\mu=\log U_1+\log S_1 \tag{2.6}$$

The following lemma shows how the normal equations turn out for the chain ladder linear model. This lemma gives the classical least squares estimates for the two-way analysis of variance model, and it is used to relate this linear model to the familiar chain ladder method.

2.1 *Lemma*

Based on t years' data, the best linear unbiased estimators of μ, α_i, β_j are the solutions of

$$\hat{\alpha}_i=\frac{1}{t-i+1}\sum_{j=1}^{t-i+1}\left(Y_{ij}-\frac{1}{t-j+1}\sum_{l=1}^{t-j+1}(Y_{lj}-\hat{\alpha}_l)\right) \tag{2.7}$$

$$(i=2,\ldots,t)$$

$$\hat{\beta}_i=\frac{1}{t-j+1}\sum_{j=1}^{t-j+1}\left(Y_{ij}-\frac{1}{t-i+1}\sum_{l=1}^{t-i+1}(Y_{il}-\hat{\beta}_l)\right) \tag{2.8}$$

$$(j=2,\ldots,t)$$

$$\hat{\mu}=\frac{2}{t(t+1)}\sum_{i=1}^{t}\sum_{j=1}^{t-i+1}(Y_{ij}-\hat{\alpha}_i-\hat{\beta}_j) \tag{2.9}$$

with $\hat{\alpha}_1=\hat{\beta}_1=0$.

Proof

The theorem can be proved using the Gauss–Markov theorem. The normal equations are the same as those used in Kremer[2], but the restrictions differ slightly; this gives the parameters a slightly different interpretation.

It is also shown by Kremer that the chain ladder method will produce results which are similar to those produced by the analysis of variance method. The latter has been studied in great depth in the statistical literature, and in the remaining sections the methods will be based on the analysis of variance version of the chain ladder technique. The analysis of variance method has the advantage of a great deal of theoretical background, and this theory will be applied to the insurance data, bearing in mind that the main method in use in the industry is the chain ladder method.

In the comparison of the chain ladder method with the two-way analysis of variance, Kremer reverses the transformation given by (2.4)–(2.6) to obtain the total claims for year of business i as

$$E_i = e^{\mu} e^{\alpha_i} \sum_{j=1}^{n} e^{\beta_j}. \tag{2.10}$$

Up to this point, the change from the chain ladder method to the multiplicative model given by (2.2) is only a reparameterization. Kremer now estimates E_i by substituting the estimates of the parameters into (2.10). Thus

$$\hat{E}_i = e^{\hat{\mu}} e^{\hat{\alpha}_i} \sum_{j=1}^{n} e^{\hat{\beta}_j}. \tag{2.11}$$

While this serves to identify the chain ladder method with the two-way analysis of variance, the estimators obtained are not the maximum likelihood estimators, nor are they unbiased. The unbiased estimates for the classical analysis are derived in Verrall (1989b)[8]. However, since the methods in this paper are based on Bayesian models, the Bayesian estimates will be derived and used.

Since the errors in the linear model are assumed to be jointly normally distributed, it is implicitly assumed that the data, $\{X_{ij}: j \leq t-i+1,\ i \leq t\}$, are lognormally distributed.

The Bayes estimate of a future observation is

$$E(X_{kl} | \{X_{ij}: j \leq t-i+1,\ i \leq t\}) \tag{2.12}$$

and the Bayes estimate of its variance is

$$\mathrm{Var}(X_{kl} | \{X_{ij}: j \leq t-i+1,\ i \leq t\}) \tag{2.13}$$

where X_{kl} is yet to be observed.

For ease of notation, $\{X_{ij}: j \leq t-i+1,\ i \leq t\}$ will be denoted by D.

2.2 *Lemma*

Suppose that X_{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.
$$\log X_{kl}|\theta \sim N(\theta,\sigma^2)$$
$$\theta|D \sim N(m,\tau^2).$$

Suppose also that σ^2 and τ^2 are known. Then
$$E(X_{kl}|D) = e^{m+\frac{1}{2}\sigma^2+\frac{1}{2}\tau^2}$$
and
$$\text{Var}(X_{kl}|D) = e^{2m+\sigma^2+\tau^2}(e^{\sigma^2+\tau^2}-1).$$

Proof

$$\begin{aligned} E(X_{kl}|D) &= E_{\theta|D}(E(X_{kl}|\theta,D)) \\ &= E_{\theta|D}(e^{\theta+\frac{1}{2}\sigma^2}) \\ &= e^{\frac{1}{2}\sigma^2}E_{\theta|D}(e^{\theta}) \\ &= e^{\frac{1}{2}\sigma^2}e^{m+\frac{1}{2}\tau^2} \qquad \text{using the m.g.f. of the normal distribution} \\ &= e^{m+\frac{1}{2}\sigma^2+\frac{1}{2}\tau^2} \end{aligned}$$

$$\begin{aligned} \text{Var}(X_{kl}|D) &= E_{\theta|D}(\text{Var}(X_{kl}|\theta,D)) + \text{Var}_{\theta|D}(E(X_{kl}|\theta,D)) \\ &= E_{\theta|D}(e^{2\theta+\sigma^2}(e^{\sigma^2}-1)) + \text{Var}_{\theta|D}(e^{\theta+\frac{1}{2}\sigma^2}) \\ &= E_{\theta|D}(e^{2\theta+\sigma^2}(e^{\sigma^2}-1)) + E_{\theta|D}(e^{2\theta+\sigma^2}) - (E_{\theta|D}(e^{\theta+\frac{1}{2}\sigma^2}))^2 \\ &= e^{2\sigma^2}E_{\theta|D}(e^{2\theta}) + (e^{\frac{1}{2}\sigma^2}E_{\theta|D}(e^{\theta}))^2 \\ &= e^{2\sigma^2}e^{2m+2\tau^2} - (e^{\frac{1}{2}\sigma^2}e^{m+\frac{1}{2}\tau^2})^2 \\ &= e^{2m+\tau^2+\sigma^2}(e^{\tau^2+\sigma^2}-1). \end{aligned}$$

Similar methods can be used to calculate the other elements of the covariance matrix, $\text{Cov}(X_{kl},X_{pn}|D)$.

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

$$\sum_{j>n-i+1} E(X_{ij}|D) \tag{2.14}$$

and the Bayes estimate of the variance is

$$\sum_{j>n-i+1} [\text{Var}(X_{ij}|D) + 2\sum_{k>j}\text{Cov}(X_{ij}, X_{ik}|D)]. \tag{2.15}$$

3. RECURSIVE MODELS AND ESTIMATION

In order to consider the Kalman filter 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 run-off triangle are received in the form:

$$X_{1,1}\,,\ \begin{bmatrix} X_{1,2} \\ \\ X_{2,1} \end{bmatrix},\ \begin{bmatrix} X_{1,3} \\ \\ X_{2,2} \\ \\ X_{3,1} \end{bmatrix},\ \ldots, \tag{3.1}$$

and in year t the data which are received are

$$\begin{bmatrix} X_{1,t} \\ X_{2,t-1} \\ \vdots \\ X_{t,1} \end{bmatrix} \tag{3.2}$$

Thus, the direction of propagation of time is along the diagonal:

$$\begin{array}{ccccc} \square & \square & \square & \cdots & \square \\ \square & \cdots & & \square & \\ \vdots & \vdots & & & \\ \vdots & & & \searrow\text{time} & \\ \square & & & & \end{array}$$

A recursive approach must use the data sequentially and must use the data at time t to update the parameter estimates based on the data available before time t.

The data vector at time t is $\boldsymbol{X}_t$, where

$$\boldsymbol{X}_t = \begin{bmatrix} X_{1,t} \\ X_{2,t-1} \\ \vdots \\ X_{t,1} \end{bmatrix}$$

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

$$\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_t, \ldots$$

In this time series, the data vector expands with t: for a triangular set of data,

$$\dim(\boldsymbol{X}_t) = t.$$

If the data are in the shape of a rhombus, which occurs when the early years of business are fully run off, then $\boldsymbol{X}_t$ will reach a point when its dimension does not increase.

The analysis can be approached from the context of multivariate time series. However, the special relationships between the elements of consecutive data vectors mean that it is simpler to generalize 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 (1989a)[7].

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 analysis of variance model, given by (2.3), 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 $\quad Y_{ij} = \log X_{ij}$.

When the data are handled recursively, the model becomes:

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

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

$$\begin{bmatrix} Y_{1,3} \\ Y_{2,2} \\ Y_{3,1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \alpha_3 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} e_{1,3} \\ e_{2,2} \\ e_{3,1} \end{bmatrix} \tag{3.3}$$

etc.

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

$$\boldsymbol{\theta}_t = \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \vdots \\ \alpha_t \\ \beta_t \end{bmatrix} \tag{3.4}$$

and (3.3) 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

$$\boldsymbol{\theta}_{t+1} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 \end{bmatrix} \boldsymbol{\theta}_t + \begin{bmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \boldsymbol{u}_t \tag{3.5}$$

where $\boldsymbol{u}_t$ 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 (3.5) says 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, $\boldsymbol{u}_t$, and the state vector, $\boldsymbol{\theta}_t$, are independent.

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

$$\begin{aligned} \boldsymbol{Y}_1 &= \boldsymbol{F}_1\boldsymbol{\theta}_1 + \boldsymbol{e}_1 \\ \boldsymbol{Y}_2 &= \boldsymbol{F}_2\boldsymbol{\theta}_2 + \boldsymbol{e}_2 \\ &\vdots \\ \boldsymbol{Y}_t &= \boldsymbol{F}_t\boldsymbol{\theta}_t + \boldsymbol{e}_t \end{aligned} \tag{3.6}$$

where $\boldsymbol{Y}_t = \log \boldsymbol{X}_t$.

Equation (3.6) 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. $\boldsymbol{\theta}_t$ is known as the state vector and is related to $\boldsymbol{\theta}_{t-1}$ 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 $\boldsymbol{\theta}_t$ to $\boldsymbol{\theta}_{t-1}$ 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:

$$\boldsymbol{\theta}_{t+1} = G_t\boldsymbol{\theta}_t + H_t\boldsymbol{u}_t + \boldsymbol{w}_t \tag{3.7}$$

where $\boldsymbol{u}_t$ is a stochastic input vector
and $\boldsymbol{w}_t$ is a disturbance vector

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

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

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

The choices of G_t, W_t and the distribution of $\boldsymbol{u}_t$ govern the dynamics of the system, and some useful cases are now described.

The simplest case is to set $\boldsymbol{u}_t$ and $\boldsymbol{w}_t$ to $\boldsymbol{0}$ for all t. In this case (3.7) becomes:

$$\boldsymbol{\theta}_{t+1} = G_t\boldsymbol{\theta}_t. \tag{3.8}$$

If G_t is chosen such that the parameters at time $t+1$ are the same as the parameters at time t, and the prior distribution of the parameters is vague, (3.8) defines recursive least squares estimation when the parameters are identical for each row and for each column. The case when the new parameters entering at time $t+1$ are distinct from those at time t can be achieved by setting

$$\boldsymbol{\theta}_{t+1} = G_t\boldsymbol{\theta}_t + H_t\boldsymbol{u}_t \tag{3.9}$$

where $\boldsymbol{u}_t$ has the prior distribution of the new parameters. If this prior distribution is vague, least squares estimation with distinct parameters is achieved. Otherwise, Bayesian estimation with distinct parameters results. This is the arrangement which was used in (3.5).

Between the cases of identical and distinct parameters comes dynamic parameter estimation, where the parameters at time $t+1$ are related to, but not necessarily the same as, the parameters at time t. A sequential relationship between the parameters can be achieved by setting

$$\boldsymbol{\theta}_{t+1} = G_t\boldsymbol{\theta}_t + \boldsymbol{w}_t \tag{3.10}$$

where $\boldsymbol{w}_t$ is a disturbance.

This is the form of system equation considered by de Jong and Zehnwirth[3]. The updating of the estimates of the state vector (which contains the parameters), as each new data vector is received is carried out by the Kalman filter.

The updating equations are derived for the most general state system which will be used:

$$Y_t = F_t\theta_t + e_t \tag{3.11}$$

$$\theta_{t+1} = G_t\theta_t + H_t u_t + w_t \tag{3.12}$$

where $e_t \sim N(0, V_t)$,

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

and $w_t \sim N(0, W_t)$

and are independent.

Further e_t, u_t, w_t are sequentially independent.

Suppose $\theta_t|(Y_1, Y_2, \ldots, Y_{t-1}) \sim N(\hat{\theta}_{t|t-1}, C_t)$ (3.13)

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

From (3.11) and (3.12), the distribution of Y_t given information up to time $t-1$ is

$$\hat{Y}_{t|t-1} \sim N(F_t\hat{\theta}_{t|t-1}, F_tC_tF'_t + V_t). \tag{3.14}$$

When the observed value of Y_t is received, the state estimate can be updated to $\hat{\theta}_{t|t}$ and the distribution of the state vector at time t forecast using (3.12).

The recursion is given by the following theorem, a proof of which can be found in (for example) Davis & Vinter (1985)[1].

3.1 *Theorem*

If the system and observation equations are given by (3.11) and (3.12), and the distribution of θ_t given information at time $t-1$ is given by (3.13), then the distribution of the state vector can be updated when Y_t is received using the following recursion:

$$\hat{\theta}_{t+1|t} = G_t\hat{\theta}_{t|t-1} + H_t\hat{u}_t + K_t(Y_t - \hat{Y}_t) \tag{3.15}$$

where $K_t = G_tC_tF'_t(F_tC_tF'_t + V_t)^{-1}$ (3.16)

$$C_{t+1} = G_tC_tG'_t + H_tU_tH'_t - G_tC_tF'_t(F_tC_tF'_t + V_t)^{-1}F_tC_tG'_t + W_t \tag{3.17}$$

and $\hat{Y}_t = F_t\hat{\theta}_{t|t-1}$. (3.18)

4. EXAMPLES

In this section, the models referred to above are applied to the data in Taylor and Ashe (1983)[5]. The state space models are compared with least squares and empirical Bayes models. For each model, the observation equation is the same and is given by (3.6).

The data are

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.

For comparison purposes the results from a static model with no prior information are given. The parameter estimates are the same as those which arise when classical least squares analysis is used, although in a classical estimation problem unbiased predictors might be used (see Verrall (1986b)[(8)]).

The parameter estimates and their standard errors are:

Table 4.1

Parameter	Estimate	Standard Error
μ	6·106	·165
α_2	·194	·161
α_3	·149	·168
α_4	·153	·176
α_5	·299	·186
α_6	·412	·198
α_7	·508	·214
α_8	·673	·239
α_9	·495	·281
α_{10}	·602	·379
β_2	·911	·161
β_3	·939	·168
β_4	·965	·176
β_5	·383	·186
β_6	−·005	·198
β_7	−·118	·214
β_8	−·439	·239
β_9	−·054	·281
β_{10}	−1·393	·379

The fitted values and predicted values are set out in Table 4.2. The actual data values are also shown.

Table 4.2

273714	680804	699807	718428	401531	272374	243232	176409	259453	67948
357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
392716	976794	1004059	1030776	576104	390793	348981	253106	372255	110927
352118	884021	933894	1183289	445745	320996	527804	266172	425046	
362971	902881	928011	952705	532469	361194	322549	233936	379508	102650
290507	1001799	926219	1016654	750816	146923	495992	280405		
324822	807924	830475	852574	476506	323232	288648	228731	340091	91988
310608	1108250	776189	1562400	272482	352053	206286			
362965	902795	927995	952688	532460	361187	351067	256033	380684	102968
443160	693190	991983	769488	504851	470639				
373842	929849	955804	981237	548416	404479	362429	264319	393004	106300
396132	937085	847498	805037	705960					
405100	1007596	1035720	1063280	646825	439787	394066	287392	427310	115580
440832	847631	1131398	1063269						
442462	1100526	1131245	1268870	710402	483014	432799	315640	469311	126940
359480	1061648	1443370							
386545	961445	1090101	1120657	627422	426594	382246	278771	414492	112113
376686	986608								
344014	973601	1001989	1030076	576709	392113	351349	256238	380990	103051
344014									

In Table 4.2, the fitted values have been calculated as suggested by Kremer, but the predicted values use the Bayesian estimation theory of lemma 2.2.

The values which are of most interest when comparing the methods are the row totals and overall totals. In the following examples, the fitted values and predicted values will be omitted.

Table 4.3

Row	Predicted Outstanding Claims	Bayes Standard Error
2	110927	60216
3	482157	189896
4	660810	210040
5	1090752	304721
6	1530532	401125
7	2310959	601536
8	3806976	1056660
9	4452396	1375446
10	5066116	2049337

The row totals and their standard errors are given in Table 4.3. The predicted overall total outstanding claims is 19511632 and the standard error of this estimate is 3194056. It is justifiable to use a normal approximation in this case since the total is a sum of over 40 random variables. Thus an approximate 95% upper bound on the total outstanding claims is

$$19511632 + 1{\cdot}645 \times 3194056 = 24765854$$

4.1 *Static Estimation*

Firstly, the recursive Bayes estimation model is considered. The state equation is given by (3.5), and is

$$\boldsymbol{\theta}_{t+1} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 \end{bmatrix} \boldsymbol{\theta}_t + \begin{bmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \boldsymbol{u}_t$$

Suppose that there is prior information which suggests that the prior distribution of the row parameters has mean ·3 and variance ·05, but that there is no prior information about the other parameters.

Table 4.4 shows the parameter estimates and their standard errors. It can be seen, by comparison with Table 4.1, that the estimates of the row parameters

Table 4.4

Parameter	Estimate	Standard Error
μ	6·177	·123
α_2	·200	·118
α_3	·166	·121
α_4	·170	·126
α_5	·275	·131
α_6	·349	·137
α_7	·402	·145
α_8	·479	·156
α_9	·362	·170
α_{10}	·369	·191
β_2	·893	·158
β_3	·911	·164
β_4	·915	·171
β_5	·320	·180
β_6	−·080	·191
β_7	−·199	·207
β_8	−·518	·231
β_9	−·128	·271
β_{10}	−1·464	·362

have generally been drawn towards the prior mean. For example the estimate of α_9 is now ·362, compared with the estimate of ·495 when there was no prior information. It should also be noted that the estimates of the column parameters have only changed as a result of the change in the estimates of the row parameters.

The row totals and their standard errors are given in Table 4.5. The predicted overall total outstanding claims in this case is 16770131 and the standard error of this estimate is 1953764. Table 4.5 should be compared with Table 4.3. The effect of the prior distribution can be seen clearly in the predicted outstanding claims for each row: the earlier ones have increased and the later ones decreased. The prior distribution has also affected the standard errors of the predicted outstanding claims for each year of business.

Table 4.5

Row	Predicted Outstanding Claims	Bayes Standard Error
2	131413	69379
3	596992	228013
4	812958	246552
5	1157631	302566
6	1450615	346525
7	1992212	457084
8	2796170	655781
9	3911910	926180
10	3920230	968366

4.2 *Dynamic Estimation of the Row Parameters*

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

$$\theta_{t+1} = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & & 1 \\ 0 & \dots & & 1 & 0 \\ 0 & \dots & & \dots & 0 \end{bmatrix} \theta_t + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ 1 \end{bmatrix} u_t + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{bmatrix} w_t$$

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:

$$\alpha_{t+1} = \alpha_t + w_t \tag{4.1}$$

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 ·0289, for reasons which will become clear.

In this case the parameter estimates are as in Table 4.6.

Table 4.6

Parameter	Estimate	Standard Error
μ	6·119	·163
α_2	·187	·151
α_3	·170	·148
α_4	·196	·152
α_5	·296	·158
α_6	·396	·164
α_7	·482	·171
α_8	·550	·183
α_9	·536	·202
α_{10}	·546	·238
β_2	·906	·158
β_3	·940	·165
β_4	·951	·173
β_5	·364	·183
β_6	−·028	·195
β_7	−·145	·212
β_8	−·457	·236
β_9	−·062	·278
β_{10}	−1·406	·378

The row totals and their standard errors are given in Table 4.7.

Table 4.7

Row	Predicted Outstanding Claims	Bayes 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 Figure 4.1, but it is interesting to compare the results with another estimation method. This is an empirical Bayes approach and was derived by Verrall (1988)[6]. The empirical Bayes approach assumes that the row parameters are independent observations from a common distribution,

i.e. $$\alpha_i \sim N(\theta,\sigma_\alpha^2). \tag{4.2}$$

This can be compared with (4.1) which can be written as

$$\alpha_{i+1} \sim N(\alpha_i,w_\alpha). \tag{4.3}$$

Equation (4.3) assumes a recursive relationship, while (4.2) simply assumes that the row parameters are all independent observations of the same random variable.

The empirical Bayes parameter estimates are as in Table 4.8. The method also produces an estimate of the variance of the distribution of the row parameters, which in this case is ·0289. This value was used in the dynamic estimation method, as the state variance of the row parameters.

Table 4.8

Parameter	Estimate	Standard Error
μ	6·157	·131
α_2	·225	·124
α_3	·193	·129
α_4	·198	·133
α_5	·300	·138
α_6	·371	·144
α_7	·421	·150
α_8	·493	·159
α_9	·383	·170
α_{10}	·391	·185
β_2	·893	·128
β_3	·911	·133
β_4	·915	·139
β_5	·319	·147
β_6	−·080	·156
β_7	−·199	·170
β_8	−·515	·190
β_9	−·120	·224
β_{10}	−1·444	·306

The row totals and their standard errors are given in Table 4.9. The predicted overall total outstanding claims is 16280338 and the standard error of this estimate is 1313997.

Figure 4.1 shows the parameter estimates for the three cases above. It can be seen from the graph that the Kalman filter 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

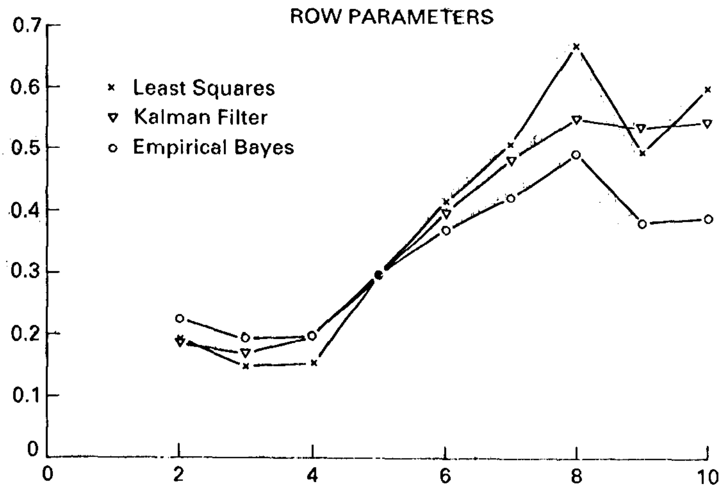

Figure 4.1

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 have affected the estimates of outstanding claims, as illustrated by Tables 4.7 and 4.9. The standard errors have been reduced because the estimation has involved more of the data for each parameter. This is a beneficial effect of any of the Bayesian methods.

Table 4.9

Row	Predicted Outstanding Claims	Bayes Standard Error
2	109448	46963
3	479568	148617
4	655656	162104
5	1033109	220459
6	1388261	270730
7	2002772	374041
8	3018896	572899
9	3780759	720836
10	3811869	752593

4.3 *Dynamic Estimation of the Row and Column Parameters*

The dynamic estimation method can be extended to the column parameters. The system equation becomes:

$$\theta_{t+1} = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ 0 & \dots & 1 & 0 & \\ 0 & \dots & 0 & 1 & \end{bmatrix} \theta_t + \begin{bmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} w_t$$

where $\quad w_t \sim N(0, W_t)$

The latter model is the closest to those considered in de Jong and Zehnwirth[3], although the analysis of variance method was not discussed.

This model was applied with state variance for the row parameters ·0296 and for the column parameters ·4135. Again, these were the figures obtained from the empirical Bayes method which will be described later in this section. The parameters estimates are as in Table 4.10.

Table 4.10

Parameter	Estimate	Standard Error
μ	6·102	·163
α_2	·211	·150
α_3	·186	·148
α_4	·212	·152
α_5	·313	·158
α_6	·414	·164
α_7	·502	·171
α_8	·569	·183
α_9	·553	·202
α_{10}	·564	·239
β_2	·908	·157
β_3	·939	·162
β_4	·929	·170
β_5	·373	·179
β_6	−·012	·189
β_7	−·151	·204
β_8	−·411	·224
β_9	−·215	·256
β_{10}	−1·132	·342

The row totals and their standard errors are given in Table 4.11. The predicted overall total outstanding claims is 18417296 and the standard error of this estimate is 2627190.

Table 4.11

Row	Predicted Outstanding Claims	Bayes Standard Error
2	143834	72675
3	465847	166438
4	673175	194229
5	1060794	266228
6	1479407	339755
7	2218738	487975
8	3287633	735669
9	4517179	1040596
10	4570683	1167068

Again, the results can be compared with the empirical Bayes results. In this case the row parameters are assumed to be independent observations from a common distribution, as before, and the column parameters also are;

i.e. $$\alpha_i \sim N(\theta, \sigma_\alpha^2) \tag{4.4}$$

and $$\beta_i \sim N(\eta, \sigma_\beta^2). \tag{4.5}$$

The parameters estimates are as in Table 4.12.

Table 4.12

Parameter	Estimate	Standard Error
μ	6·122	·131
α_2	·254	·125
α_3	·225	·129
α_4	·235	·134
α_5	·341	·139
α_6	·415	·144
α_7	·466	·151
α_8	·537	·159
α_9	·424	·171
α_{10}	·429	·186
β_2	·878	·129
β_3	·894	·133
β_4	·896	·139
β_5	·317	·146
β_6	−·066	·156
β_7	−·175	·168
β_8	−·464	·187
β_9	−·081	·218
β_{10}	−1·168	·286

The estimate of the variance of the distribution of the row parameters is ·0296, and the estimate of the variance of the column parameters is ·4135.

The row totals and their standard errors are given in Table 4.13. The predicted overall total outstanding claims is 16827488 and the standard error of this estimate is 1346017.

Table 4.13

Row	Predicted Outstanding Claims	Bayes Standard Error
2	142697	59015
3	465847	157033
4	710746	172221
5	1104778	233349
6	1470739	285575
7	2094885	390731
8	3098807	587274
9	3838265	733542
10	3841936	763092

5. CONCLUSIONS

This paper has attempted to show how Bayesian methods can be applied to the chain ladder linear model. The Kalman filter and a state space approach have been concentrated upon and some other possibilities illustrated. 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 datum point from which to estimate the last row parameter; the Bayesian methods use the datum from the other rows as well. To illustrate the effect of this, consider a change in the datum point in the last row from its present value of 344014 to 544014. Table 5.1 shows the predicted outstanding claims for each row from the different models. The first column shows the original results with no prior information.

Table 5.1

	Original Results			Revised Results		
Row	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 datum 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.

The Kalman filter requires the state variances to be specified before the analysis begins. In the examples above, figures from the empirical Bayes analysis have been used although these may not always be appropriate, or available. There is no such requirement for the empirical Bayes method, which estimates the variances from the data.

The empirical Bayes estimation method has the advantage that no prior information is needed: the 'prior' distribution is estimated from the data (hence the term 'empirical'). The predictions from the Kalman filter estimates have been obtained using the same method as in the other analyses, although it may be more proper to use the n-step-ahead forecasts. Either of the Bayesian methods has the advantage of predictor stability over the ordinary chain-ladder linear model. It is important to realize 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 outstanding 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.

6. REFERENCES

(1) DAVIS, M. H. A. & VINTER, R. B. (1985) *Stochastic Modelling and Control*. Chapman & Hall.

(2) KREMER, E. (1982) IBNR-Claims and the Two-Way Model of ANOVA. *Scand. Act. J.* **1**, 47–55.

(3) DE JONG, P. & ZEHNWIRTH, B. (1983) Claims Reserving, State-space Models and The Kalman Filter. *J.I.A.* **110**, 157–182.

(4) RENSHAW, A. E. (1989) Chain Ladder and Interactive Modelling. *J.I.A.* **116**, 559.

(5) TAYLOR, G. C. & ASHE, F. R. (1983) Second Moments of Estimates of Outstanding Claims. *J. of Econometrics*, **23**, 37–61.

(6) VERRALL, R. J. (1988) Bayesian Linear Models and the Claims Run-off Triangle. Actuarial Research Paper No 7, The City University, London.

(7) VERRALL, R. J. (1989a) Modelling Claims Runoff Triangles with Two-Dimensional Time Series. *Scand. Act. J.* (to appear).

(8) VERRALL, R. J. (1989b) On the Unbiased Estimation of Reserves from Loglinear Models. (Under review).