City Research Online

City, University of London Institutional Repository

Citation: Verrall, R. J. (1989). Stochastic Models for Triangular Tables with Applications to Cohort Data and Claims Reserving. (Unpublished Doctoral thesis, City University London)

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/7407/

Link to published version:

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

The City University

Department of Actuarial Science and Statistics

Stochastic Models for Triangular Tables with Applications to Cohort Data and Claims Reserving

by
Richard John Verrall

Thesis submitted for the degree of Doctor of Philosophy

November 1989

CONTENTS

Acknowledgements 5

Abstract 6

Chapter 1 Introduction

1.0 Subject Matter 7

1.1 Summary of Methods Developed and Related Work 9

Chapter 2 Linear Models and Claims Runoff Triangles

2.0 Introduction 12

2.1 Linear Models 13

2.2 The Chain Ladder Technique as a Linear Model 14

2.2.1 Lemma 20

2.3 Relationship between the estimators of the linear model and the chain ladder model 21

2.3.1 Lemma 22

2.3.2 Lemma 25

2.4 Unbiased Estimation of Reserves, and Variances of Reserves 27

2.4.1 Estimation for Independently, Identically Distributed Data 28

2.4.2 Unbiased Estimation for Claims Runoff Triangles 32

2.4.3 Unbiased Estimates of Total Outstanding Claims 36

2.4.4 Prediction Intervals 37

2.5 Example 39

2.6 Row by Row Linear Models 48

2.7 Example 51

2.8 Maximum Likelihood Estimation 58

2.9 Example 64

Chapter 3 Bayesian Linear Models and Credibility Theory

3.0 Introduction 68

3.1 Bayesian Estimation for Linear Models: Two Stage Models 70

3.1.1 Lemma 71

3.2 Bayesian Estimation for Linear Models: Three Stage Models 73

3.2.1 Lemma 73

3.2.2 Lemma 74

3.2.3 Variance Estimation 75

3.3 Bayesian Estimation for Claims Runoff Triangles 76

3.3.1 Example 77

3.3.2 Example 80
3.4 Bayes Estimates for the Analysis of Variance Model 83
3.4.1 Example 85
3.5 Bayes Estimates for Row Models 92
3.5.1 Example 93
3.6 Empirical Bayes Estimates for the Analysis of Variance Model 97
3.6.1 Theorem 99
3.6.2 Remarks 104
3.6.3 Example 104
3.6.4 Empirical Bayes Assumptions on the Row and Column Parameters 109
3.6.5 Example 110
3.7 Empirical Bayes Estimates for Row Models 113
3.7.1 Theorem 113
3.7.2 Remarks 118
3.7.3 Example 119

Chapter 4 Recursive Bayesian and Dynamic Forecasting
4.0 Introduction 122
4.1 Recursive Linear Models: State Space Representations 125
4.2 The Kalman Filter 127
4.2.1 Theorem 128
4.2.2 Remarks 129
4.2.3 Forecast Values and Variances 129
4.3 Applications of the Kalman Filter to the Chain Ladder Linear Model 130
4.3.1 Example 133
4.3.2 Dynamic Row Parameters 135
4.3.3 Example 136
4.3.4 Dynamic Row and Column Parameters 140
4.3.5 Example 141
4.4 Applications of the Kalman Filter to Row Models 144
4.4.1 Dynamic Estimation for Row Models 146
4.4.2 Bayesian Estimation for Row Models with Distinct Parameters 146
4.4.3 Example 147

Chapter 5 Two-Dimensional Processes
5.0 Introduction 150
5.1 Two-Dimensional State Systems 152
5.1.1 A Two-Dimensional State Space System for the Chain Ladder Linear Model 153
5.1.2 Causality 154

5.1.3 Recursive State Estimation 155
5.1.3.1 Theorem 158
5.1.4 An Example of Two-Dimensional State Space Estimation applied to the Chain Ladder Linear Model 159
5.2 Two-Dimensional Classical Time Series 167
5.2.1 Nearest Neighbour Models 167
5.2.2 Model Selection and Parameter Estimation 170
5.2.2.1 Example 175
5.2.3 Stationarity 176
5.2.4 State Space Representation of Two-Dimensional Classical Time Series 177
5.2.4.1 Theorem 177

Chapter 6 Conclusions
6.0 Introduction 180
6.1 Stability of Predictions 181
6.2 Negative Incremental Values 182
6.3 Reserves 183
6.4 Further Research 183

Appendix 1 185
Appendix 2 187

References 200

Acknowledgements

I wish to express my gratitude to my advisor, Prof. Henry Wynn, and to my colleagues at the City University, in particular Prof. Steven Haberman and Dr. Arthur Renshaw, for many useful discussions in the process of this thesis.

I would also like to thank Prof. Sidney Benjamin of Bacon & Woodrow & Co. who has given me much encouragement in this work. Also, I am grateful to Mr. S. Christofides and his colleagues at the Commercial Union Assurance Company with whom I have had very beneficial discussions of the practical problems of claims reserving during the later stages of this work.

Declaration

Powers of discretion are granted to the University Librarian to allow the thesis to be copied in whole or in part without further reference to the author. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

Abstract

Stochastic models for triangular data are derived and applied to claims reserving data. The standard actuarial technique, the so-called "chain-ladder technique" is given a sound statistical foundation and considered as a linear model. This linear model, the "Chain Ladder Linear Model" is extended to encompass Bayesian, empirical Bayes and dynamic estimation. The empirical Bayes results are given a credibility theory interpretation, and the advantages and disadvantages of the various approaches are highlighted. Finally, the methods are extended to two-dimensional systems and results based on classical time series and Kalman filtering theory are produced.

The empirical Bayes estimation results are very useful, practically, and can be compared to the Kalman filter estimates. They have the advantage that no prior information is required: the Kalman filter method requires the state and observation variances to be specified. For illustration purposes the estimates from the empirical Bayes procedure are used. The empirical Bayes results can also be compared with credibility theory estimators, although they retain the general statistical advantages of the linear modelling approach.

For the classical theory, unbiased estimates of outstanding claims, reserves and variances are derived, and prediction intervals for total outstanding claims are produced. Maximum likelihood theory is utilised to derive the distributions of quantities relating to the column parameters which have actuarial interpretations. The row totals are also considered. Bayesian estimates of similar quantities are derived for the methods based on Bayes theory.

Chapter 1

Introduction

1.0 Subject Matter

This thesis is concerned with the evolution of processes on a lattice. The methods are applicable to a wide range of practical problems, for example in demography, medical statistics and age/period/cohort analysis. However, the application which has stimulated the research is the prediction of outstanding claims on a portfolio of general insurance policies. A complete survey of the literature on claims reserving was carried out by Lyons (1984).

This thesis extends the standard actuarial technique, the "chain ladder technique" (described in Benjamin (1977) and Hossack, Pollard and Zehnwirth (1983)) and derives more sophisticated procedures, bearing in mind that the chain ladder technique is the most widely used method. Therefore, a complete survey of the different methods which have been used will not be repeated here, but the specific papers relevant to this thesis will be considered.

The most important paper on the statistical treatment of claims reserving is Kremer (1982) which noted the connection between the standard actuarial technique and the two-way analysis of variance linear model. It is this linear model, applied to claims reserving, which is examined in detail. This thesis develops the methods of claims reserving, based on linear models, without moving away from the chain ladder method (although the application of the methods to other models will also be considered). Classical maximum likelihood and unbiased estimates are derived. Bayesian and empirical Bayes estimates are derived and the latter interpreted within credibility theory and compared with dynamic estimates. The theory of Bayesian forecasting is extended to two-dimensional systems, and two-dimensional stationary time series are considered. An information criterion for two-dimensional time series is derived, and the state space representation considered.

Most of the examples in this thesis are based on a set of general insurance data, shown in figure 1.1. This data is taken from a paper by Taylor and Ashe (1983) and consists of claims from a portfolio of general insurance policies (the source of the data is not specified).

Figure 1.1

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.

This data set is used as a standard data set to test and compare the methods. The chain ladder technique is used to analyse data from most types of business, and the linear model can also be applied to most types of data. There are some technical difficulties arising in certain circumstances, which are dealt with in chapter 6.

The statistical aspects of the analysis of the data and prediction of reserves are concentrated on, and no distinction is made between the different possible sources of delay. In addition, no distinction is made between 'Incurred' and 'Incurred But Not Reported' (IBNR) Claims, and no 'accounting' methods such as Reid (1978 and Claims Reserving Manual(1989)) are considered.

1.1 Summary of Methods Developed and Related Work

Kremer (1982) showed that the chain ladder technique has the same structure as the two-way analysis of variance linear model. The statistical foundations of the chain ladder technique are studied in detail in this thesis, and the methods of analysis of claims data are extended with particular reference to the two-way analysis of variance linear model. The paper by Kremer was an important first attempt to relate the chain ladder technique to standard statistical methods, but stopped short of using the full machinery of estimation theory. It aimed at producing estimates of outstanding claims which are as close as possible to those from the chain ladder technique. Chapter 2 of this thesis is devoted to a study of the relationship between the chain ladder technique and the two-way analysis of variance linear model and to the derivation of statistically sound estimates of outstanding claims, reserves and 'safe' reserves. Because of its intimate connection with the chain ladder technique, the two-way analysis of variance linear model will also be known as the chain ladder linear model.

The linear model is applied to the logged data, and it is assumed that the data are lognormally distributed. A useful text on the lognormal distribution is Aitchison and Brown (1969). It is necessary to reverse the logarithmic transformation to produce estimates and forecasts on the original scale. Taylor and Ashe (1983) considered unbiased estimation, and derived results for linear transformations. In chapter 2, the results are extended to the (non-linear) logarithmic transformation inherent in the chain ladder linear model. The work of Finney (1942) and Bradu and Mundlak (1970) is drawn on to derive maximum likelihood and unbiased estimates of outstanding claims. Also contained in this chapter is the derivation of the distribution of the parameters of the chain ladder model using maximum likelihood theory. The computer programmes have been implemented in FORTRAN: specimen programmmes are reproduced in appendix 2. The methods of chapter 2 have also been implemented in GLIM (Baker and Nelder (1978)) by Renshaw (1989) and SuperCalc5 by the Commercial Union Assurance Company. See also Renshaw and Verrall (1989). Chapter 2, therefore, contains the classical estimation theory of claims reserving and derives 'safe' reserves using unbiased estimation.

The statistical methods of claims reserving can be extended to accomodate various modelling assumptions. Chapter 3 uses hierarchical Bayes linear models to derive Bayes and empirical Bayes estimates of outstanding claims. The theory of hierarchical Bayes linear models, derived

in, for example, Lindley and Smith (1972) and Smith (1973), has been considered in an insurance context by Klugman (1989). There is an alternative method which imposes a similar structure as the empirical Bayes method, but does not use the linear model. The empirical Bayes estimates can be regarded as 'credibility' estimates, and the alternative method applies the theory of credibility derived in the context of premium rate setting by Jewell (1975) and Buhlmann (1967) to the chain ladder technique. This method, derived by de Vylder (1982), is described in more detail by Goovaerts and Hoogstad (1987). The actuarial approaches of the chain ladder technique and credibility theory are combined to produce estimates of outstanding claims which have a credibility theory interpretation. The statistical advantages of the linear model over the crude chain ladder technique are the same whether a classical or a Bayesian method is used. In a different context, Zeger, See and Diggle (1989) suggest an empirical Bayes approach to the analysis of AIDS data, which also contains a reporting delay.

In chapter 4, another Bayesian method is considered and compared with the empirical Bayes method. This method is based on the Kalman filter which was also used in a claims reserving context by de Jong and Zehnwirth (1983). The causal structure is discussed in chapter 5, and chapter 4 uses the 'wave form' of the Kalman filter which was suggested by de Jong and Zehnwirth and is natural for claims data. J.Q.Smith (1983) also used Bayesian forecasting to analyse claims data, applying dynamic generalised linear models to each row in the triangle. This chapter gives the dynamic form of the chain ladder linear model and compares the results with the empirical Bayes estimates.

In chapter 5, the Bayesian methods based on the Kalman filter are extended to allow dynamic variation in both directions in the triangles. This draws on Givone and Roesser (1972 and 1973), Fanasini and Marchesini (1976), Kung, Levy, Morf and Kailath (1977) and Porter and Aravena (1986). It generalises to two dimensions the Bayesian forecasting models of Harrison and Stevens (1976), although the particular application here is to regression models. The chain ladder technique has been viewed as a time series model by Pater (1989) and Lemaire (1982), and structural time series models have been applied in the context of claims data by Harvey and Fernandes (1989). These use the theory of time series in one dimension. However, cohort data can be viewed as a two-dimensional process and chapter 5 extends the analysis to include two-dimensional time series methods. These were first used in the context of nearest-neighbour

models and agricultural field trials by Bartlett (1968), Bartlett and Besag (1969), Besag (1972) and Whittle (1954). In chapter 5, the use of two-dimensional time series is considered and an information criterion is derived. This is based on the papers by Akaike (1973 and 1976).

Chapter 6 contains the conclusion and discusses the important practical issue of negative incremental values. The stability of predictions is also considered, and some ideas for further research are given.

CHAPTER 2

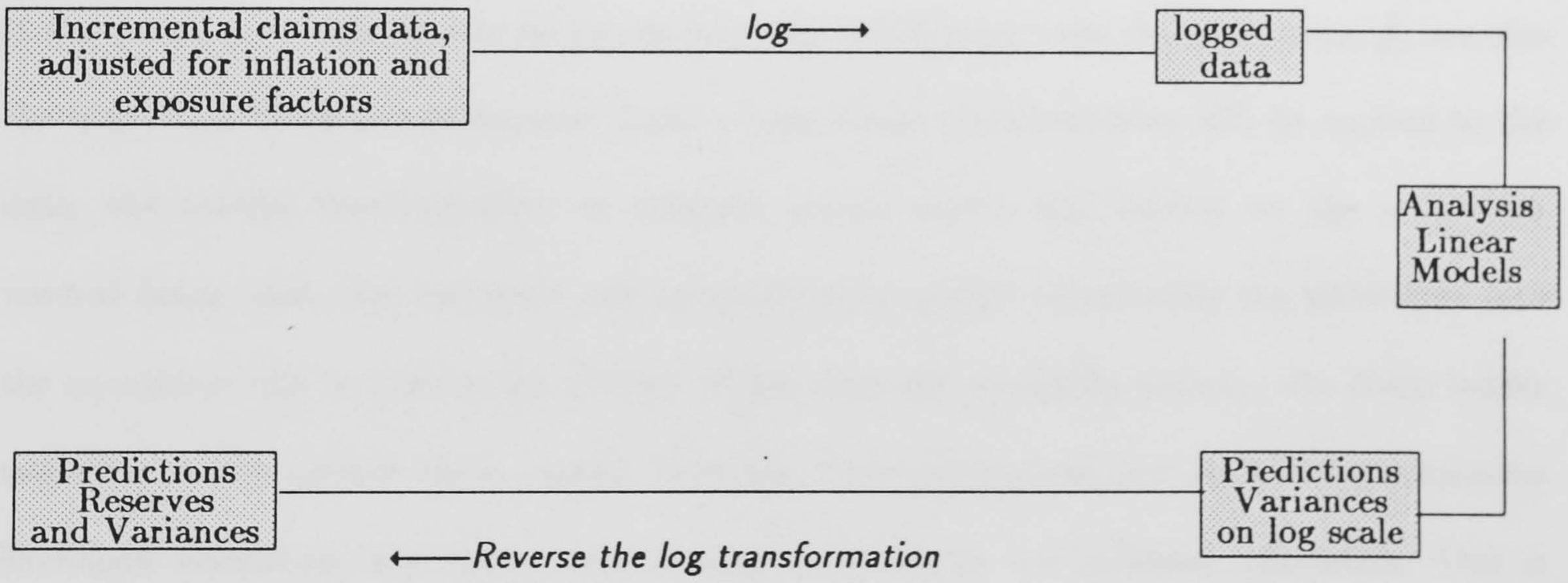
Linear Models and Claims Runoff Triangles

2.0 Introduction

If claims runoff triangles are to be analysed statistically, as a data analysis exercise, it is desirable to express them as linear models. If the claims are analysed using a model for each row, then it may be straightforward to write down a linear model: this is considered in section 2.6. The use of linear models to analyse the data by row can give useful insights into the nature of the data, but it is the linear model which is close to the chain ladder technique which is of greatest interest to actuaries. This linear model, whose connection with the chain ladder technique was first identified by Kremer (1982), is described in sections 2.2 and 2.3.

The data are assumed to be lognormally distributed and is first logged before a linear model is applied. The transformation from the raw data to the logged data is, obviously, straightforward, but the reverse transformation, once the analysis has been carried out, is not simple. This is dealt with in section 2.4. The process is represented in the following diagram:

Figure 2.1



Prediction from linear models when the data are lognormally distributed was first considered by Finney (1942). Finney considered a sample of independently, identically distributed data, and the theory was generalised to a sample of independently, but not necessarily identically distributed data by Bradu and Mundlak (1970). Subsequent papers have considered the properties of the estimators in more detail. Section 2.8 applies maximum likelihood theory to the chain ladder parameters.

2.1 Linear Models

The linear model to be considered is

$$\underline{y} = X\underline{\beta} + \underline{e} \qquad (2.1.1)$$

where $\underline{y}$ is a data vector of length n,

$\underline{\beta}$ is a parameter vector of length p,

X is an (nxp) design matrix

and $\underline{e}$ is an error vector of length n.

$\underline{e}$ is assumed to have mean zero and variance-covariance matrix Σ.

The minimum variance linear unbiased estimators of the parameters, $\underline{\beta}$, are the weighted least-squares estimators, $\hat{\underline{\beta}}$, where

$$\hat{\underline{\beta}} = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}\underline{y}. \qquad (2.1.2)$$

If the errors, $\underline{e}$, are assumed to be jointly normally distributed, then the estimators, $\hat{\underline{\beta}}$, are also the maximum likelihood estimators. Since a logarithmic transformation will be applied to the data, the reverse transformation to estimate actual claims will depend on the estimation method being used. One estimator can be obtained by simply substituting the estimators into the equations: this is used in the lemmas which show the similarity between the chain ladder technique and a certain linear model. However, these estimators, and indeed the maximum likelihood estimators, are biased, and it may be better to use unbiased estimators. This is discussed more fully in section 2.4. If the errors are assumed to be uncorrelated with equal variance then equation (2.1.2) simplifies to

$$\hat{\underline{\beta}} = (X'X)^{-1} X'\underline{y} \qquad (2.1.3)$$

which is a form which will be also be used.

The distributional properties of the maximum likelihood estimators, $\hat{\underline{\beta}}$, are well-known.

Assuming that the errors are independently, identically distributed with variance σ^2,

$$\hat{\underline{\beta}} \sim N(\underline{\beta}, \sigma^2(X'X)^{-1}) \qquad (2.1.4)$$

2.2 The Chain Ladder Technique as a Linear Model

Kremer (1982) showed that the chain ladder technique is very similar to a two-way analysis of variance and investigated the properties of the estimators. This section describes the connection between the actuarial and statistical methods. Assuming a triangular data set (without loss of generality) the cumulative claims data, to which the chain ladder technique is applied, are

$$\{ C_{ij} : i=1,...,t;\ j=1,...,t-i+1\} \qquad (2.2.1)$$

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

$$\{Z_{ij} : i=1,...,t;\ j=1,...,t-i+1\} \qquad (2.2.2)$$

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

$$Z_{i1} = C_{i1}$$

The chain ladder technique is based on the model

$$E(C_{ij}) = \lambda_j C_{i,j-1} \qquad j=2,...,t. \qquad (2.2.3)$$

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

$$\tilde{\lambda}_j = \frac{\sum_{i=1}^{t-j+1} C_{ij}}{\sum_{i=1}^{t-j+1} C_{i,j-1}}. \qquad (2.2.4)$$

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

$$\text{estimate of } E(C_{it}) = \left(\prod_{j=t-i+2}^{t} \tilde{\lambda}_j\right) C_{i,t-i+1} . \qquad (2.2.5)$$

This is illustrated in example 2.5 and the results are compared with those from the analysis of variance method.

The chain ladder technique produces forecasts which have a row effect and a column effect. The column effect is obviously due to the parameters $\{ \lambda_j ; j=2,...,t \}$. There is also a row effect since the estimates for each row depend not only on the parameters $\{ \lambda_j ; j=2,...,t \}$, 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 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:

$$E(Z_{ij}) = U_i S_j \qquad (2.2.6)$$

where U_i is a parameter for row i,

S_j is a parameter for column j

A multiplicative error structure is assumed.

$$\text{Also} \quad \sum_{j=1}^{t} S_j = 1 \qquad (2.2.7)$$

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).

The estimates of U_i will be compared with the estimates of $E(C_{it})$ in equation (2.2.5) and S_j and λ_j will be related to each other.

The analysis of variance estimators are based on the model (2.2.6):

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

and the chain ladder technique is based on the model (2.2.3):

$$E\left(C_{ij} \right) = \lambda_j \, C_{i,j-1}$$

In terms of the models, ignoring for the moment the estimation of the parameters, this simply represents a reparameterisation.

Under the chain ladder model, the expected claim total for business year i is

$$\prod_{j=t-i+2}^{t} \lambda_j \; C_{i,t-i+1} \qquad (2.2.8)$$

and the expected claim amount in development year t-i+2 is

$$\lambda_{t-i+2} \; C_{i,t-i+1} - \; C_{i,t-i+1} \qquad (2.2.9)$$

The equivalent quantities under the multiplicative model, (2.2.6) are

$$U_i \qquad (2.2.10)$$

$$\text{and } U_i \, S_{t-i+2} \qquad (2.2.11)$$

Equating (2.2.8) and (2.2.9) with (2.2.10) and (2.2.11), respectively, gives:

$$S_{t-i+2} \; = \frac{\lambda_{t-i+2} - 1}{\prod_{j=t-i+2}^{t} \lambda_j}.$$

The expected claim amount for development year t-i+3 under each model is:

$$\lambda_{t-i+3}\,\lambda_{t-i+2}\;C_{i,t-i+1} \;-\; \lambda_{t-i+2}\,C_{i,t-i+1} \tag{2.2.12}$$

and $U_i\,S_{t-i+3}$ (2.2.13)

which give

$$S_{t-i+3} \;=\; \frac{\lambda_{t-i+3}-1}{\prod_{j=t-i+3}^{t}\lambda_j}.$$

In general,

$$S_j \;=\; \frac{\lambda_j-1}{\prod_{l=j}^{t}\lambda_l}. \tag{2.2.14}$$

Considering year of business t, the expected total claim amount under each model is:

$$\left[\prod_{j=2}^{t}\lambda_j\right]C_{t1}$$

and U_t .

The claim amount in development year 1, C_{t1} , is modelled by $U_t\,S_1$, and so it can be seen that:

$$S_1 \;=\; \frac{1}{\prod_{l=2}^{t}\lambda_l} \tag{2.2.15}$$

To summarise, the chain ladder model (2.2.3) is equivalent to the multiplicative model given by equation (2.2.6) with the following relationships between the parameters:

$$S_1 = (\prod_{l=2}^{t} \lambda_l)^{-1}$$

$$S_j = (\prod_{l=j}^{t} \lambda_l)^{-1} (\lambda_j - 1)$$

$$U_i = E(C_{it})$$

Equations (2.2.4) and (2.2.5) give the estimators of $\{ \lambda_j ; j=2,...,t \}$ and $E(C_{it})$. 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 (2.2.6), and assuming that the incremental claims are positive:

$$E(Y_{ij}) = \mu + \alpha_i + \beta_j \qquad (2.2.16)$$

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

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.

Kremer (1982) defines μ as the mean of the log U_i's and log S_j's , so that the restriction

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

is imposed.

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

$$\alpha_i = \log U_i - \log U_1 \tag{2.2.17}$$

$$\beta_j = \log S_j - \log S_1 \tag{2.2.18}$$

$$\mu = \log U_1 + \log S_1 \, . \tag{2.2.19}$$

The latter set of assumptions are more appropriate for the more sophisticated techniques which will be introduced in later chapters and so they will also be used in this chapter. It is emphasised that the prediction and estimation of the claims is unaffected by the choice of assumptions.

The assumption that e_{ij} are independently, identically distributed with variance σ^2 will be used, so that the estimators are given by equation (2.1.3).

Now equation (2.2.16) can be written in the form of equation (2.1.1). Suppose, for example, there are three years of data.

Then

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{13} \\ y_{22} \\ y_{31} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{13} \\ e_{22} \\ e_{31} \end{bmatrix} \tag{2.2.20}$$

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

The following lemma, due to Kremer (1982), gives the normal equations for the chain ladder linear model.

2.2.1 Lemma

For n years data, the best linear unbiased estimators of μ, α_i , β_j are the solutions of

$$\hat{\alpha}_i = \frac{1}{\text{t-i+1}} \sum_{j=1}^{t-i+1} (Y_{ij} - \frac{1}{\text{t-j+1}} \sum_{l=1}^{t-j+1} (Y_{lj} - \hat{\alpha}_l)) \qquad (2.2.1.1)$$

(i=2,...,t)

$$\hat{\beta}_j = \frac{1}{\text{t-j+1}} \sum_{i=1}^{t-j+1} (Y_{ij} - \frac{1}{\text{t-i+1}} \sum_{l=1}^{t-i+1} (Y_{il} - \hat{\beta}_l)) \qquad (2.2.1.2)$$

(j=2,...,t)

$$\hat{\mu} = \frac{2}{\text{t(t+1)}} \sum_{i=1}^{t} \sum_{j=1}^{t-i+1} (Y_{ij} - \hat{\alpha}_i - \hat{\beta}_j) \qquad (2.2.1.3)$$

Proof

The normal equations, (2.1.3), are

$$\text{(t-i+1)}\hat{\mu} + \text{(t-i+1)}\hat{\alpha}_i + \sum_{j=2}^{t-i+1} \hat{\beta}_j = \sum_{j=1}^{t-i+1} Y_{ij} \qquad (2.2.1.4)$$

(i=2,...,t)

$$\text{(t-j+1)}\hat{\mu} + \sum_{i=2}^{t-j+1} \hat{\alpha}_i + \text{(t-j+1)}\hat{\beta}_j = \sum_{i=1}^{t-j+1} Y_{ij} \qquad (2.2.1.5)$$

(j=2,...,t)

$$\frac{\text{t(t+1)}}{2} \hat{\mu} + \sum_{i=2}^{t} \text{(t-i+1)}\hat{\alpha}_i + \sum_{j=2}^{t} \text{(t-j+1)}\hat{\beta}_j = \sum_{i=1}^{t} \sum_{j=1}^{t-i+1} Y_{ij} \qquad (2.2.1.6)$$

Noting that $\hat{\alpha}_1 = \hat{\beta}_1 = 0$, equations (2.2.1.6) and (2.2.1.3) are equivalent.

Also equations (2.2.1.4) and (2.2.1.5) can be written as

$$\hat{\alpha}_i = \frac{1}{\text{t-i+1}} \sum_{j=1}^{t-i+1} (Y_{ij} - \hat{\beta}_j) - \hat{\mu} \qquad (2.2.1.7)$$

$$\hat{\beta}_j = \frac{1}{\text{t-j+1}} \sum_{i=1}^{t-j+1} (Y_{ij} - \hat{\alpha}_i) - \hat{\mu}. \qquad (2.2.1.8)$$

Substituting equation (2.2.1.7) into equation (2.2.1.8) and vice versa gives equations (2.2.1.1) and (2.2.1.2).

2.3 Relationship between the estimators of the linear model and the chain ladder model

The previous section derived the relationship between the parameters of the multiplicative model and the chain ladder technique. The parameters are estimated in different ways according to which method is used, and this section is devoted to examining the relationships between the estimators of the parameters.

This section contains two lemmas. The first deals with the estimation of S_j and U_i - the parameters of the multiplicative model - using the chain ladder technique. The second lemma derives the estimators of S_j and U_i using the two-way analysis of variance model. The two sets of estimators are then shown to be similar. Thus, it will be shown 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 chapters 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 insurance data, bearing in mind that the main method in use in the industry is the chain ladder method.

2.3.1 Lemma

If $$S_j = \frac{\lambda_j - 1}{\prod_{l=j}^{t} \lambda_l} \qquad j=2,\ldots,t \qquad (2.3.1.1)$$

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

$$\tilde{\lambda}_j = \frac{\sum_{i=1}^{t-j+1} C_{ij}}{\sum_{i=1}^{t-j+1} C_{ij-1}} \qquad (2.3.1.2)$$

then the estimators of S_j, $\tilde{S}_j$, satisfy the relationship

$$\tilde{S}_j = \frac{\sum_{i=1}^{t-j+1} Y_{ij}}{\sum_{i=1}^{t-j+1} C_{i,t-i+1} \Big/ (1 - \sum_{l=t-i+2}^{t} \tilde{S}_l)} \qquad j=2,\ldots,t. \qquad (2.3.1.3)$$

Also, the estimate of U_i is $\tilde{U}_i$, where

$$\tilde{U}_i = \frac{\sum_{j=1}^{t-i+1} Z_{ij}}{\sum_{j=1}^{t-i+1} \tilde{S}_j}. \qquad (2.3.1.4)$$

Proof

Equations (2.3.1.1) and (2.3.1.2) imply that

$$\tilde{S}_j = \frac{\tilde{\lambda}_j - 1}{\prod_{l=j}^{t} \tilde{\lambda}_j}$$

$$= \frac{\sum_{i=1}^{t-j+1} C_{ij} - \sum_{i=1}^{t-j+1} C_{ij-1}}{\sum_{i=1}^{t-j+1} C_{ij} \prod_{l=j+1}^{t} \tilde{\lambda}_j} . \qquad (2.3.1.5)$$

Now, it can be shown by induction that

$$\sum_{i=1}^{t-j+1} C_{ij} = \sum_{i=1}^{t-j+1} C_{i,t-i+1} \Big/ \prod_{l=j+1}^{t-i+1} \tilde{\lambda}_j \qquad (2.3.1.6)$$

(See Kremer).

Substituting equation (2.3.1.6) into equation (2.3.1.5) gives

$$\tilde{S}_j = \frac{\sum_{i=1}^{t-j+1} Z_{ij}}{\left(\sum_{i=1}^{t-j+1} C_{i,t-i+1} \Big/ \prod_{l=j+1}^{t-i+1} \tilde{\lambda}_l \right) \prod_{l=j+1}^{t} \tilde{\lambda}_l}$$

$$= \frac{\sum_{i=1}^{t-j+1} Z_{ij}}{\sum_{i=1}^{t-j+1} C_{i,t-i+1} \prod_{l=t-i+2}^{t} \tilde{\lambda}_l} . \qquad (2.3.1.7)$$

It can also be shown by induction that

$$\left[\prod_{l=k}^{t} \lambda_l \right]^{-1} = 1 - \sum_{l=k}^{t} S_l .$$

This is true for k=2 by virtue of equation (2.2.15) and the relationship

$$1 - \sum_{l=2}^{t} S_l = S_1 .$$

Suppose it is true for k. Then for k+1:

$$1 - \sum_{l=k+1}^{t} S_l \quad = \quad 1 - \sum_{l=k}^{t} S_l \; + S_k$$

$$= \left[\prod_{l=k}^{t} \lambda_l\right]^{-1} + \frac{\lambda_k - 1}{\prod_{l=k}^{t} \lambda_l}$$

$$= \left[\prod_{l=k+1}^{t} \lambda_l\right]^{-1} . \qquad (2.3.1.8)$$

Hence, by induction, the result holds.

Substituting this result into (2.3.1.7) gives:

$$\tilde{S}_j \;= \frac{\sum_{i=1}^{t-j+1} Z_{ij}}{\sum_{i=1}^{t-j+1} C_{i,t-i+1} \Big/ (1 - \sum_{l=t-i+2}^{t} \tilde{S}_l)} \qquad (2.3.1.9)$$

as required.

Now, since $C_{i,t-i+1} = \sum_{j=1}^{t-i+1} Z_{ij}$

and $\prod_{j=t-i+2}^{t} \tilde{\lambda}_j = (1 - \sum_{j=t-i+2}^{t} \tilde{S}_j)^{-1}$

the estimate of total expected outstanding claims for row i ,

$$C_{i,t-i+1} \prod_{j=t-i+2}^{t} \tilde{\lambda}_j$$

can be written as

$$\frac{\sum_{j=1}^{t-i+1} Z_{ij}}{1 - \sum_{j=t-i+2}^{t} \tilde{S}_j} .$$

This can be written as

$$\frac{\sum_{j=1}^{t-i+1} Z_{ij}}{\sum_{j=1}^{t-i+1} \tilde{S}_j} \qquad (2.3.1.10)$$

since $1 - \sum_{j=t-i+2}^{t} \tilde{S}_j = \sum_{j=1}^{t-i+1} \tilde{S}_j$.

2.3.2 Lemma

Using the estimation method of lemma 2.2.1, an estimate of total expected claims for accident year i, $\hat{U}_i$, is given by

$$\hat{U}_i = \left[\prod_{j=1}^{t-i+1} \frac{Z_{ij}}{w_j}\right]^{\frac{1}{t-i+1}} \cdot \sum_{j=1}^{t} w_j \qquad (2.3.2.1)$$

where $w_j = \dfrac{\left[\prod_{i=1}^{t-j+1} Z_{ij}\right]^{\frac{1}{t-j+1}}}{\left[\prod_{i=1}^{t-j+1} \left[\prod_{l=1}^{t-i+1} Z_{il}\right]^{\frac{1}{t-i+1}} \Big/ \left[\prod_{l=1}^{t-i+1} w_l\right]^{\frac{1}{t-i+1}}\right]^{\frac{1}{t-j+1}}}$. (2.3.2.2)

Further, $\hat{U}_i = \dfrac{\left[\prod_{j=1}^{t-i+1} Z_{ij}\right]^{\frac{1}{t-i+1}}}{\left[\prod_{j=1}^{t-i+1} \hat{S}_j\right]^{\frac{1}{t-i+1}}}$. (2.3.2.3)

This lemma can be used to show that the estimates of expected total outstanding claims for each row have similar forms using each method, and can be expected to behave in similar ways. The estimate of U_i is obtained by "hatting" the parameters in the identity

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

which is derived in the proof of this lemma. 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.

Proof

Equations (2.2.17) to (2.2.19) imply that

$$e^{\alpha_i} = \frac{U_i}{U_1} \qquad (2.3.2.4)$$

$$e^{\beta_j} = \frac{S_j}{S_1} \qquad (2.3.2.5)$$

and $e^{\mu} = U_1 S_1$. (2.3.2.6)

Since $\sum_{j=1}^{t} S_j = 1, \quad S_1 = (\sum_{j=1}^{t} e^{\beta_j})^{-1}$.

This, together with equations (2.3.2.4) and (2.3.2.6) give:

$$U_i = e^{\alpha_i} e^{\mu} \sum_{j=1}^{t} e^{\beta_j} . \qquad (2.3.2.7)$$

Now let $w_j = e^{\hat{\beta}_j}$; then equation (2.2.1.2) is equivalent to equation (2.3.2.2).

The BLUE estimate of $\alpha_i + \mu$ is obtained from equation (2.2.1.7). Substituting the estimates of $\alpha_i + \mu$ and β_j into equation (2.3.2.7) gives the estimate of U_i in equation (2.3.2.1).

Now, equation (2.3.2.5) implies that

$$\hat{S}_j = \frac{w_j}{\sum_{l=1}^{t} w_l}$$

and so equations (2.3.2.1) and (2.3.2.2) can be written as

$$\hat{U}_i = \frac{\left[\prod_{j=1}^{t-i+1} Z_{ij}\right]^{\frac{1}{t-i+1}}}{\left[\prod_{j=1}^{t-i+1} \hat{S}_j\right]^{\frac{1}{t-i+1}}}$$

$$\text{and} \quad \hat{S}_j = \frac{\left[\prod_{i=1}^{t-j+1} Z_{ij}\right]^{\frac{1}{t-j+1}}}{\left[\prod_{i=1}^{t-j+1} \left[\prod_{l=1}^{t-i+1} Z_{il}\right]^{\frac{1}{t-i+1}} \Big/ \left[\prod_{l=1}^{t-i+1} \hat{S}_l\right]^{\frac{1}{t-i+1}}\right]^{\frac{1}{t-j+1}}} \quad . \qquad (2.3.2.8)$$

Now, if all the geometric means are replaced by arithmetic means in equation (2.3.2.8), the recurrence relation for the estimators of S_j becomes the same as that in lemma 2.3.1. Similarly the estimators of U_i are equivalent if geometric means are replaced by arithmetic means. Thus the two estimation methods, the chain ladder method and the linear model, will produce similar 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.

2.4 Unbiased 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. It is important that estimates should be unbiased in order that they are aiming at the correct target and do not yield values which consistently under- or over-estimate. It is also important to consider unbiased estimation of the standard errors of the estimates of expected total outstanding claims, in order that some measure of the order of the errors can be attached to the predictions. The procedure for analysing claims data using loglinear models is illustrated by figure 2.1.

The final stage in this procedure - reversing the log transformation - is considered here and unbiased estimates of total outstanding claims are derived. Unbiased estimates of the variances of these estimates are also derived. The theory is applied to some claims data using the analysis of variance linear model and the unbiased estimates compared with some alternatives. In order to make the analysis more easily assimilable, a sample of independently, identically distributed observations is considered first. The theory is then extended to the more general case of independent, but not necessarily identically distributed observations. It is the more general theory which is applicable to claims data.

2.4.1 Estimation for Independently, Identically Distributed Data

Suppose the data consists of n independently, identically distributed observations which are lognormally distributed.

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

and $Z_i \sim$ lognormal.

Suppose also that $E (Z_i) = \theta$. (2.4.1.1)

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 $Y_i = \log Z_i \qquad (i=1,...,n).$ (2.4.1.2)

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

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

Then $\theta = \exp (\mu + \frac{1}{2}\sigma^2)$. (2.4.1.3)

The maximum likelihood estimates of μ and σ^2 are

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

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

and the maximum likelihood estimate of θ can be obtained by substituting $\hat{\mu}$ and $\hat{\sigma}^2$ into equation (2.4.1.3):

$$\hat{\theta} = \exp (\hat{\mu} + \frac{1}{2}\hat{\sigma}^2) \qquad (2.4.1.4)$$

Now $\hat{\mu}$ and $\hat{\sigma}^2$ are independent and

$$\hat{\mu} \sim N (\mu , \frac{\sigma^2}{n})$$

$$n\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$$

The moment generating functions of these distributions are

$$E (\exp (\hat{\mu} t)) = \exp (\mu t + \frac{1}{2} \frac{\sigma^2}{n} t^2)$$

and $\quad E\left(\exp\left(n\frac{\hat{\sigma}^2}{\sigma^2}t\right)\right) = \left[\frac{1}{1-2t}\right]^{\frac{1}{2}(n-1)}.$

Hence $E(\hat{\theta}) = E(\exp(\hat{\mu}))\,E(\exp(\frac{1}{2}\hat{\sigma}^2))$ (by independence)

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

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

$$> \theta.$$

This result, which was derived by Finney (1941), shows that the maximum likelihood estimate of θ is biased.

It can be seen that

$$\lim_{n\to\infty}\exp\left[-\frac{(n-1)}{n}\frac{\sigma^2}{2}\right] = \exp\left(-\frac{\sigma^2}{2}\right)$$

and $\lim_{n\to\infty}\left[1-\frac{\sigma^2}{n}\right]^{-\frac{1}{2}(n-1)} = \exp\left(\frac{\sigma^2}{2}\right)$

so that $\lim_{n\to\infty} E(\hat{\theta}) = \theta$ and $\hat{\theta}$ is asymptotically unbiased. In a claims reserving context, n is generally not large and the bias in the maximum likelihood estimate can be quite serious. This is illustrated in example 2.5.

In order to correct for the bias, Finney introduced the function $g_m(t)$, where

$$g_m(t) = \sum_{k=0}^{\infty}\frac{m^k(m+2k)}{m(m+2)\ldots(m+2k)}\,\frac{t^k}{k!} \qquad (2.4.1.5)$$

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

$$\tilde{\theta} = \exp (\hat{\mu}) \, g_m (\tfrac{1}{2}(1-\tfrac{1}{n}) \, s^2) \qquad (2.4.1.6)$$

and $s^2 = \frac{n}{n-1} \hat{\sigma}^2$ is an unbiased estimate of σ^2. Again it can be seen that

$$\exp (\hat{\mu}) \, g_m (\tfrac{1}{2} (1-\tfrac{1}{n}) \, s^2) \rightarrow \exp (\hat{\mu} + \tfrac{1}{2} \hat{\sigma}^2) \quad \text{as } n \rightarrow \infty.$$

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

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

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

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

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

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

Thus an unbiased estimate of τ^2 is

$$\tilde{\tau}^2 = \exp (2\hat{\mu}) \Big[(g_m (\tfrac{1}{2} (1-\tfrac{1}{n}) \; s^2))^2 - g_m ((1-\tfrac{2}{n}) \; s^2) \Big].$$

The values of the unbiased estimates of θ, $\tilde{\theta}$, and of τ^2, $\tilde{\tau}^2$, for the more general case are illustrated in example 2.5. 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

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

2.4.2 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.

Let $\theta_{ij} = E(Z_{ij})$. (2.4.2.1)

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.

The data available for the estimation of the parameters in the linear model are:

$$\begin{array}{ccccc}
Z_{11} & Z_{12} & \cdots & \cdots & Z_{1t} \\
Z_{21} & Z_{22} & \cdots & Z_{2,t-1} & \\
\vdots & & \iddots & & \\
Z_{t-1,1} & Z_{t-1,2} & & & \\
Z_{t1} & & & &
\end{array}$$

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

Let $Y_{ij} = \log Z_{ij}$. (2.4.2.2)

Then Y_{ij} are independently normally distributed.

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

$$E(Y_{ij}) = \underline{X}_{ij} \underline{\beta} \quad (2.4.2.3)$$

$$Var(Y_{ij}) = \sigma^2 \quad (2.4.2.4)$$

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

This is a linear model and encompasses (among others) the analysis of variance model suggested by Kremer.

The linear model for the whole triangle is

$$E(\underline{Y}) = X \underline{\beta} \quad (2.4.2.5)$$

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

and $\underline{Y}$ 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.

For example, applying the two-way analysis of variance model to a 3x3 triangle gives

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

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

$$\theta_{ij} = \exp\left(\underline{X}_{ij}\, \underline{\beta} + \tfrac{1}{2}\, \sigma^2 \right) \qquad (2.4.2.6)$$

Thus the maximum likelihood estimate of θ_{ij} is

$$\hat{\theta}_{ij} = \exp\left(\underline{X}_{ij}\, \hat{\underline{\beta}} + \tfrac{1}{2}\, \hat{\sigma}^2 \right) \qquad (2.4.2.7)$$

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

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

The general theory of estimation from linear models when the data is lognormally distributed was considered by Bradu and Mundlak (1970). It can be shown that an unbiased estimate of

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

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

$$\exp\left(\underline{Z}\, \hat{\underline{\beta}} \right) g_m\left[\left(a - \tfrac{1}{2}\, \underline{Z}\, (X'X)^{-1}\, \underline{Z}' \right) s^2 \right] \qquad (2.4.2.8)$$

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

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

and $m = n - p$.

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

$$\tilde{\theta}_{ij} = \exp(\underline{X}_{ij}\,\underline{\hat{\beta}})\; g_m\left[\tfrac{1}{2}\,(1 - \underline{X}_{ij}\,(X'X)^{-1}\,\underline{X}'_{ij})\,s^2\right]. \qquad (2.4.2.9)$$

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

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

It can therefore be seen that

$$\underline{X}_{ij}\,(X'X)^{-1}\,\underline{X}'_{ij}\,s^2$$

is an estimate of $\mathrm{Var}(X_{ij}\,\underline{\hat{\beta}})$.

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

$$\tau^2_{ij} = \mathrm{Var}(\tilde{\theta}_{ij})$$

$$= E(\tilde{\theta}^2_{ij}) - (E(\tilde{\theta}_{ij}))^2. \qquad (2.4.2.10)$$

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

$$\begin{aligned} (\, E\,(\,\tilde{\theta}_{ij}\,)\,)^2 \quad &= \quad \theta_{ij}^2 \\ &= \exp\,(\,2\underline{X}_{ij}\,\underline{\beta} + \sigma^2\,). \end{aligned}$$

Hence an unbiased estimate of τ_{ij}^2 , using equations (2.4.2.9) and (2.4.2.8) is $\tilde{\tau}_{ij}^2$, where

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

The results for identically distributed data can be recovered by setting $\underline{X}_{ij} = 1$ (noting that $X'X = n$) and $\underline{\beta} = \mu$.

2.4.3. 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

$$\tilde{R}_i \quad = \sum_{j=t-i+2}^{t} \tilde{\theta}_{ij}\ . \qquad (2.4.3.1)$$

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

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

$$= \sum_{j=t-i+2}^{t}\Big[\ \text{Var}\,(\,\tilde{\theta}_{ij}\,) + 2\ \sum_{k=j+1}^{t} \text{Cov}\,(\ \tilde{\theta}_{ij}\,,\tilde{\theta}_{ik}\,)\ \Big] \qquad (2.4.3.2)$$

Now

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

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 2.4.2.

It can be shown that if

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

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

$$\tilde{\tau}_{ijk} = \exp((\underline{X}_{ij}+\underline{X}_{ik})\hat{\underline{\beta}}) \left[g_m(\tfrac{1}{2}(1-\underline{X}_{ij}\,(X'X)^{-1}\,\underline{X}'_{ij})s^2)\, g_m(\tfrac{1}{2}(1-\underline{X}_{ik}\,(X'X)^{-1}\,\underline{X}'_{ik})s^2) - g_m((1-\tfrac{1}{2}(\underline{X}_{ij}+\underline{X}_{ik})(X'X)^{-1}(\underline{X}_{ij}+\underline{X}_{ik}))s^2) \right] \qquad (2.4.3.3)$$

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

$$\sum_{j=t-i+2}^{t} \left[\tilde{\tau}_{ij}^2 + 2 \sum_{k=j+1}^{t} \tilde{\tau}_{ijk} \right] \qquad (2.4.3.4)$$

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

2.4.4 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 R = total outstanding claims for the whole triangle

and $\tilde{R}$ 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

$$P(R \leq \tilde{R} + k) = 0.95 \qquad (2.4.4.1)$$

i.e. find k such that

$$P(R - \tilde{R} \leq k) = 0.95. \qquad (2.4.4.2)$$

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

$$E(\tilde{R}) = E(R) \qquad (2.4.4.2)$$

and hence

$$E(R - \tilde{R}) = 0. \qquad (2.4.4.3)$$

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

$$Var(R - \tilde{R}) = Var(R) + Var(\tilde{R}). \qquad (2.4.4.4)$$

In section 2.4.3, 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,

$$Var(R) = \sum_{i=2}^{t} \sum_{j=t-i+2}^{t} Var(Z_{ij}) \qquad (2.4.4.5)$$

and an unbiased estimate of $Var(Z_{ij})$ is required. This can be derived as follows, using the method of section 2.4.2.

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

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

$$= \exp(2\underline{X}_{ij}\underline{\beta} + 2\sigma^2) - \exp(2\underline{X}_{ij}\underline{\beta} + \sigma^2) \qquad (2.4.4.6)$$

Hence, applying equation (2.4.2.8), an unbiased estimate of $\mathrm{Var}(Z_{ij})$ is

$$\exp(2\underline{X}_{ij}\hat{\underline{\beta}})\left[g_m\left(2(1-\underline{X}_{ij}(X'X)^{-1}\underline{X}'_{ij})s^2\right) - g_m((1-2\underline{X}_{ij}(X'X)^{-1}\underline{X}'_{ij})s^2)\right]. \qquad (2.4.4.7)$$

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:

$$\tilde{R} + 1.645\sqrt{\mathrm{Var}(R) + \mathrm{Var}(\tilde{R})} \qquad (2.4.4.8)$$

$\sqrt{\mathrm{Var}(R) + \mathrm{Var}(\tilde{R})}$ is the root mean square error of prediction, and an unbiased estimate is used. (It can be compared with the posterior estimate of the variance of the total outstanding claims in the chapters devoted to Bayesian theory.) It is very important to note that the theory of prediction intervals is only valid since unbiased estimators have been used.

2.5 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 results in the example will be compared, in chapter 3, with the Bayes and empirical Bayes results. The data used is that from Taylor and Ashe (1983). The estimates of the parameters in the analysis of variance model and their standard errors are shown in figure 2.2.

Figure 2.2

	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 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). This has consequences for the empirical Bayes estimation method described in chapter 3. It can also be seen that there is an indication that the row parameters follow an increasing trend and this will be remarked on further in chapter 4 when the Kalman filter method is described.

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.

The fitted values for the analysis of variance model are shown in figure 2.3. 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.

Figure 2.3

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

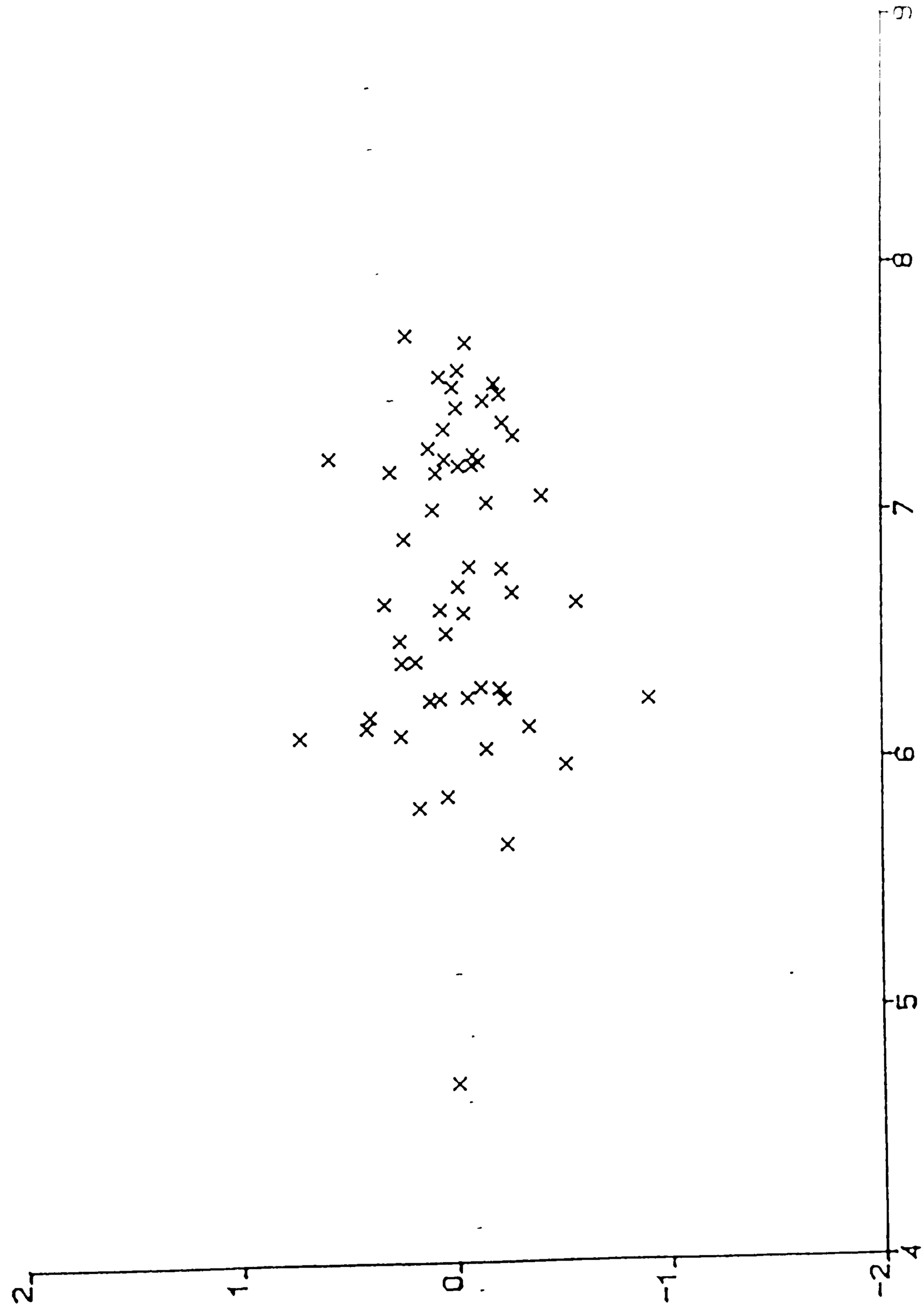
457887 1138894 1170213
359480 1061648 1443370

396651 986582
376686 986608

344014
344014

Figure 2.4 shows a plot of residuals (fitted value − actual value) against fitted value.

Figure 2.4



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 thesis is to extend the possible analyses by using Bayes, empirical Bayes and recursive Bayes methods, rather than to find the most appropriate linear model to fit, and so the residual analysis will be mainly restricted to plots similar to figure 2.4.

Of most interest to practitioners are the predicted outstanding claims for each year of business, which are the row totals of predicted values. Figure 2.5 shows the maximum likelihood predictions of the outstanding claims in the lower triangle, and figure 2.6 shows the unbiased predictions. The method does not produce any predictions for the first row, and each row contains one more predicted value.

Figure 2.5

						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

Figure 2.6

						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.

The total predicted outstanding claims for each year of business (the row totals of the predicted outstanding claims) are shown in figure 2.7. There are three estimates given, the maximum likelihood and unbiased estimates from the analysis of variance model, and the chain ladder estimate.

Figure 2.7

	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

It can be seen that the maximum likelihood estimates differ most significantly from the unbiased estimates in the early and late rows. The estimates for the middle rows are the closest together, 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.

Figure 2.8

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

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 is a 'safe' reserve for this triangle according to the chain ladder linear model using unbiased estimation.

The use of the mean square errors of prediction is discussed further in the concluding chapter.

2.6 Row by Row linear models

An approach which has been found useful by many practitioners is to model the triangle row by row. Sometimes completely separate models are applied to each row, sometimes the same model is applied to each row and sometimes the separate models which are applied to each row are related in some way. The latter approach provides a useful and intuitively appealing compromise between separate models and the same model for each row: the methods for relating the rows together are derived in chapters 3 and 4. In this chapter the methodology which enables the basic models to be fitted, either to each row separately or to all together, will be described.

The data in each row is:

$$\underline{Y}_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{i,t-i+1} \end{bmatrix}$$

and the triangle consists of

$$\underline{Y}_1' \quad \underline{Y}_2' \quad \vdots \quad \underline{Y}_t'$$

The three cases referred to above can, in general, be written as a linear model:

$$\underline{Y}_i = X_i \underline{\beta}_i + \underline{e}_i \qquad (2.6.1)$$

where X_i is the design matrix for row i,

$\underline{\beta}_i$ is the parameter vector, of dimension p, for row i

and $\underline{e}_i$ is a vector of independent errors which are assumed to be normal with mean zero.

One example of the many row models which have been suggested is that used by Zehnwirth (1985). This can be written in the form :

$$Z_{ij} = \kappa_i \, j^{\delta_i} \exp(\gamma_i (j-1)) R_{ij} \qquad (2.6.2)$$

where κ_i , δ_i , γ_i are parameters ($\gamma_i < 0$)

and R_{ij} is a multiplicative error with mean 1.

Taking logs, equation (2.6.2) can be written as a linear model:

$$Y_{ij} = \log \kappa_i + \delta_i \log j + \gamma_i (j-1) + e_{ij}$$

$$= \nu_i + \delta_i \log j + \gamma_i (j-1) + e_{ij} \qquad (2.6.3)$$

where $\nu_i = \log \kappa_i$

and e_{ij} are independent errors with mean zero.

It will also be assumed that the errors are identically distributed, although this is not necessary and should be checked by looking at the residual plots.

Referring back to equation (2.6.1),

$$\underline{\beta}_i = \begin{bmatrix} \nu_i \\ \delta_i \\ \gamma_i \end{bmatrix}$$

and the form of X_i is:

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & \log 2 & 1 \\ 1 & \log 3 & 2 \\ \vdots & \vdots & \vdots \\ 1 & \log(t-i+1) & t-i \end{bmatrix}$$

For the whole triangle, equation (2.6.1) gives the linear model:

$$\underline{Y} = \begin{bmatrix} X_1 & & & & \\ & X_2 & & & \\ & & X_3 & & \\ & & & \ddots & \\ & & & & X_t \end{bmatrix} \begin{bmatrix} \underline{\beta}_1 \\ \vdots \\ \underline{\beta}_t \end{bmatrix} + \underline{e} \qquad (2.6.4)$$

and the least-squares estimates of $\underline{\beta}_1 , \ldots , \underline{\beta}_{t-2}$ can be produced using the standard equations given in section 2.1. The last two rows have fewer observations than parameters and estimation is not possible. In the analysis here, the last two rows have simply been omitted, although it would be possible to assume that

$$\underline{\beta}_t = \underline{\beta}_{t-1} = \underline{\beta}_{t-2}$$

and proceed.

A simpler model is obtained if it is assumed that

$$\underline{\beta}_1 = \underline{\beta}_2 = \ldots = \underline{\beta}_t = \underline{\beta} \text{ (say)} \qquad (2.6.5)$$

$$\text{Then } \underline{Y} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_t \end{bmatrix} \underline{\beta} + \underline{e} \qquad (2.6.6)$$

This is the model which is used in example 2.7.

Between these two models (given by equations (2.6.4) and (2.6.6)) come the models in which the parameters are related in some way. One possibility is to assume that they follow some dynamic evolution and model accordingly; another is to assume that they are independent observations from a common distribution and use a credibility theory approach. Each of these possibilities is described further in subsequent chapters. The methodology of unbiased estimation, which has been derived in relation to the chain ladder linear model above, can be

applied to any loglinear model, including the row models which are under consideration in this section.

2.7 Example

The simpler model given by equation (2.6.5) in which the parameters for each row are assumed to be identical is applied to the Taylor and Ashe data in this example. The model for each row is the gamma curve, given by equation (2.6.2), although it is emphasised that the methods apply to any linear model which is applied to the data row by row. The estimates of the parameters and their standard errors are given in figure 2.9:

Figure 2.9

Parameter	Estimate	Standard Error
ν	6.545	0.120
δ	1.877	0.244
γ	-0.663	0.070

The residual variance is 0.162.

The fitted values, using unbiased estimation, are shown in figure 2.10, figure 2.11 shows a plot of the residuals against fitted values, figure 2.12 shows a plot of residuals against delay and figure 2.13 shows the predicted values for the lower triangle, again using unbiased estimation.

Figure 2.10

457014	868795	957642	846543	663122	480927	330482	218137	139514	86971
357848	766940	610542	482940	527326	574398	146342	139950	227229	67948

540175	1026887	1131902	1000586	783788	568440	390619	257831	164901
352118	884021	933894	1183289	445745	320996	527804	266172	425046

522195	992705	1094224	967279	757698	549518	377616	249249
290507	1001799	926219	1016654	750816	146923	495992	280405

465255	884461	974911	861808	675079	489600	336441
310608	1108250	776189	1562400	272482	352053	206286

449522	854552	941943	832665	652251	473043
443160	693190	991983	769488	504851	470639

413560	786188	866588	766052	600071
396132	937085	847498	805037	705960

406817	773370	852459	753562
440832	847631	1131398	1063269

376849	716399	789662
359480	1061648	1443370

393332	747733
376686	986608

314665
344014

Figure 2.11

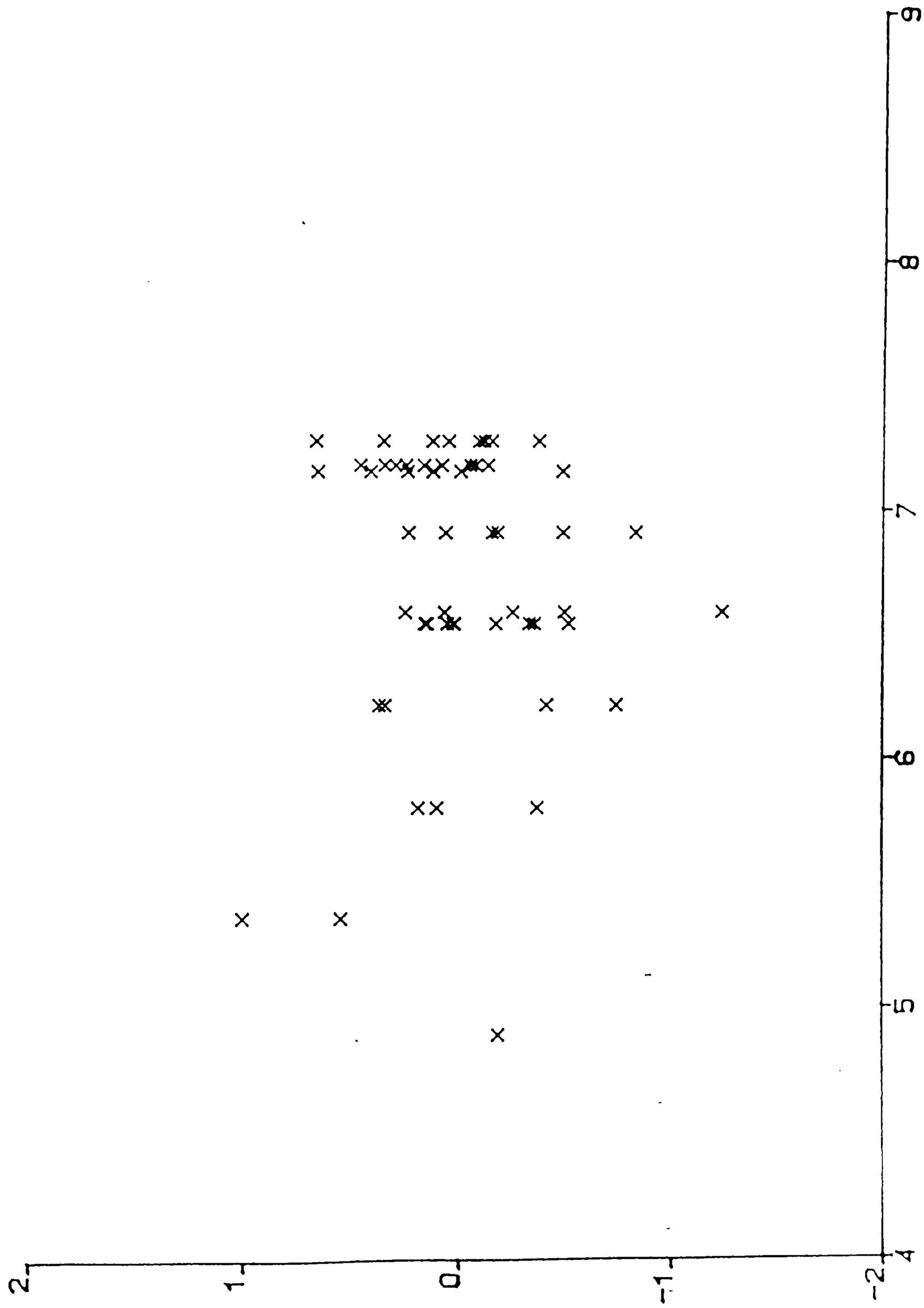


Figure 2.12

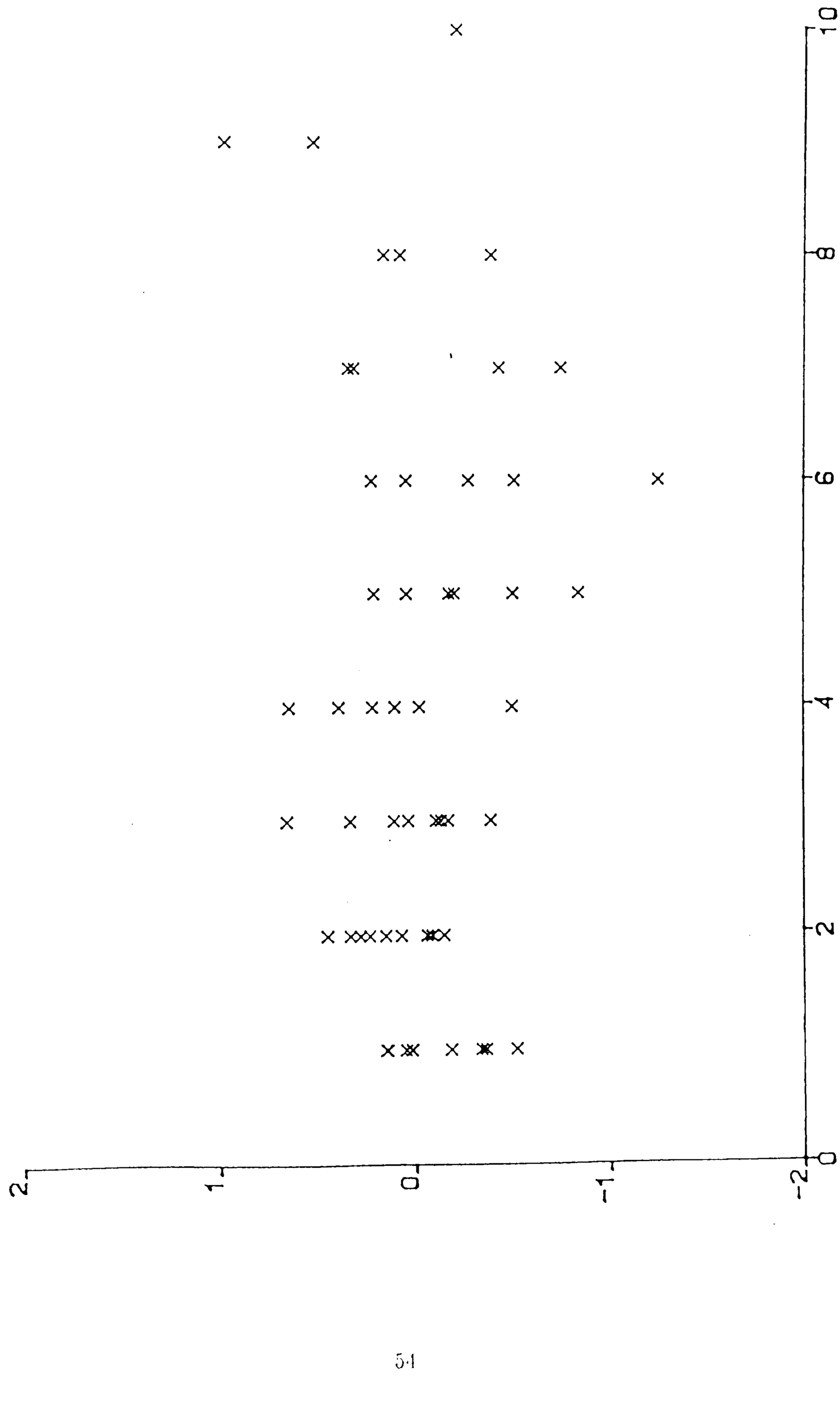
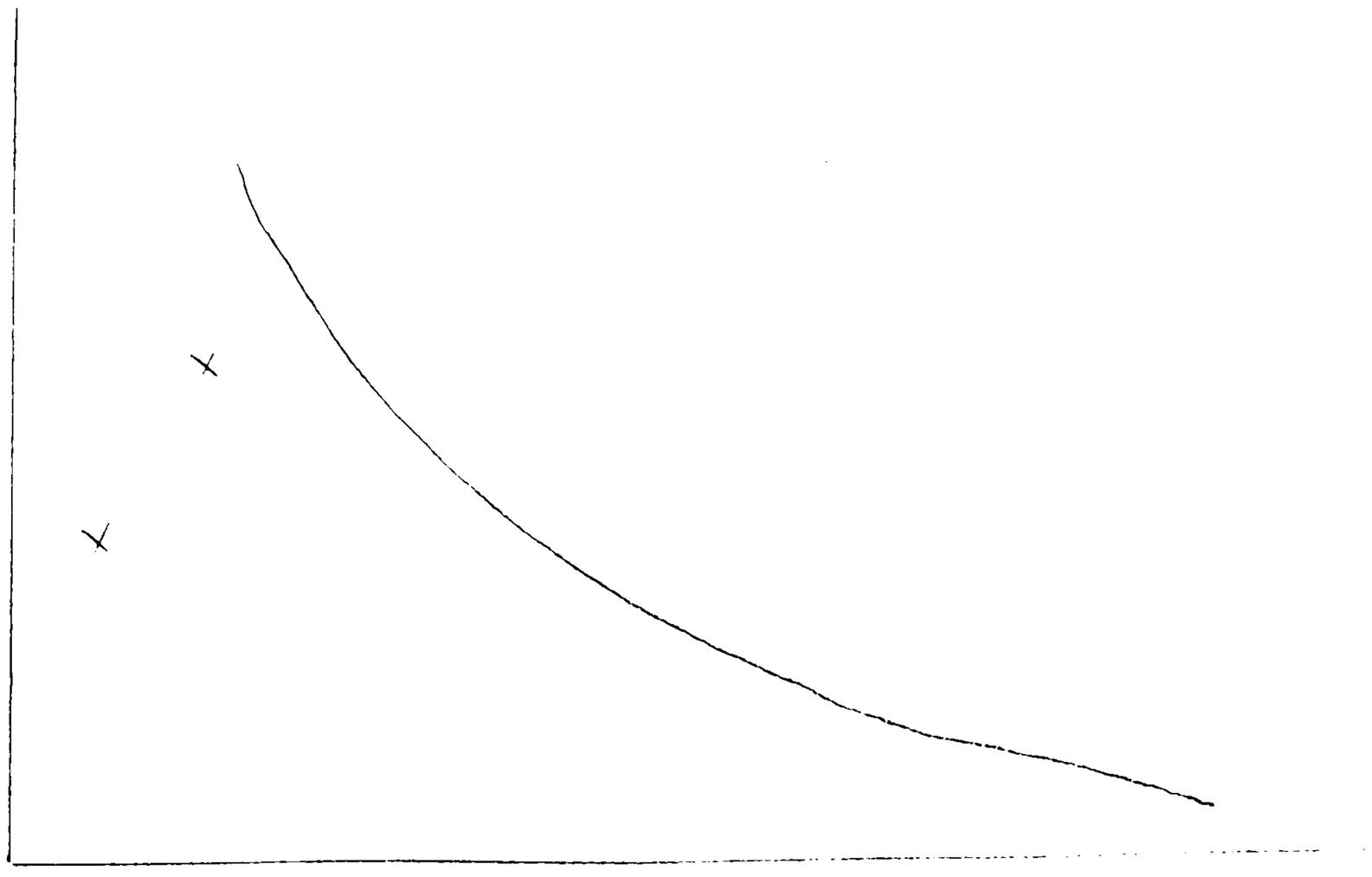


Figure 2.13

						102797
					159412	99375
				222071	142030	88539
			325064	214561	137227	85545
		435200	299059	197396	126249	78702
	590287	428104	294183	194178	124190	77418
	698051	546804	396568	272512	179874	115042	71715
	824200	728582	570719	413913	284431	187741	120074	74852
598186	659360	582865	456575	331130	227545	150193	96059	59882

This is a more rigid model than the analysis of variance model - there are far fewer parameters - and it can be seen from the plot of residuals against delay (figure 2.12) that although the fit is quite good for the later delay years, the first two show some evidence of cyclical deviation. The exponential runoff of the claims is well fitted by the gamma curve (which behaves like an exponential curve for large values of j), but the early years do not fit well. It may be better to replace this linear model by another row by row model which retains the exponential runoff, but which does not impose the gamma curve on the early years. This is illustrated in figure 2.14.

Figure 2.14



The alternative linear model implied by this is:

$$Y_{ij} = \begin{cases} \alpha & \text{if } j=1 \\ \beta & \text{if } j=2 \\ \gamma e^{-\gamma j} & \text{if } j>2 \end{cases}$$

This has been used by Swedish actuaries (see Claims Reserving Manual vol.2).

Returning to the gamma curve, figure 2.15 shows the unbiased estimates of total outstanding claims for each year of business together with the maximum likelihood estimates and the crude chain ladder estimates.

Figure 2.15

Row	Maximum Likelihood	Unbiased	Chain Ladder
2	104202	102797	94630
3	261211	258787	464668
4	455258	452640	702101
5	764741	762398	965576
6	1138039	1136606	1412202
7	1708762	1708361	2176089
8	2279985	2280566	3897142
9	3202840	3204512	4289473
10	3159459	3161796	4618035

The unbiased estimates of total outstanding claims for each year of business and the root mean square errors of prediction are shown in figure 2.16.

Figure 2.16

Unbiased Estimate	Mean Square Error Of Prediction
102797	46345
258787	88072
452640	130660
762398	194920
1136606	263932
1708361	366839
2280566	457833
3204512	601910
3161796	551115

The unbiased estimate of total outstanding claims is 13068463 and the root mean square error of prediction is 1342038. Thus a 95% upper bound on total outstanding claims is

$$13068463 + 1.645 \times 1342038 = 15276116$$

It can be seen that the predicted outstanding claims are lower than the predicted outstanding claims using the analysis of variance model in example 2.5.

This is in part due to the fact that only three parameters are used and the standard errors of these parameters are smaller than the standard errors of the parameters of the analysis of variance model. In that model each parameter has a different number of data points from which it is estimated. In particular, the last row parameter has only one data value which can be used in its estimation and each previous row parameter has one more data point which can be used. Thus the standard errors, which are a function of the design matrix, are larger and this contributes to the estimates.

As previously stated, the main purpose of this thesis is to derive more sophisticated estimation

procedures rather than compare individual linear models. It is therefore more relevant to compare this example with the Bayesian methods of the later chapters when they are applied to the same row by row linear model.

2.8 Maximum Likelihood Estimation

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:

$$S_1 = \frac{1}{\sum_{l=2}^{t} \lambda_l} \qquad (2.8.1)$$

$$S_j = \frac{\lambda_j - 1}{\sum_{l=j}^{t} \lambda_l} \qquad (j \geq 2) \qquad (2.8.2)$$

It was also shown, in theorem 2.3.2, that the proportions of ultimate claims are related to the column parameters of the linear model as follows

$$S_j = \frac{e^{\beta_j}}{\sum_{l=1}^{t} e^{\beta_l}} \qquad j = 1 , \ldots , t \qquad (2.8.3)$$

where $\beta_1 = 0$ by definition.

Finally, the relationship between the parameters of the chain ladder and linear models is given in the following theorem.

Theorem 2.8.1

Suppose that the claims are modelled by equations (2.2.3) and (2.2.16). Then the following relationships between the parameters hold:

$$\lambda_j = 1 + \frac{e^{\beta_j}}{\sum_{l=1}^{j-1} e^{\beta_l}} \qquad (2.8.1.1)$$

Proof

In order to show that equation (2.8.1.1) holds, it is first necessary to establish the following relationship between the parameters of the chain ladder and multiplicative models:

$$\lambda_j = \frac{1 - \sum_{l=j+1}^{t} S_l}{1 - \sum_{l=j}^{t} S_l} \qquad (j \neq t) \qquad (2.8.1.2)$$

$$\lambda_t = \frac{1}{1 - S_t} \qquad (2.8.1.3)$$

This can be proved by induction. Consider first λ_t . From equation (2.8.2),

$$S_t = \frac{\lambda_t - 1}{\lambda_t}.$$

Thus $\lambda_t - S_t \lambda_t = 1$

and $\lambda_t = \frac{1}{1 - S_t}$.

Now suppose that equation (2.8.1.2) holds for $\lambda_t , \ldots , \lambda_{j+1}$. From equation (2.8.2),

$$S_j = \frac{\lambda_j - 1}{\lambda_j \prod_{l=j+1}^{t} \lambda_l}$$

$$= \frac{\lambda_j - 1}{\lambda_j} \cdot \left(1 - \sum_{l=j+1}^{t} S_l \right) \quad \text{using the assumption of the induction.}$$

Thus $1 - \sum_{l=j+1}^{t} S_l = \lambda_j \left(1 - \sum_{l=j+1}^{t} S_l - S_j \right)$.

and $\lambda_j = \frac{1 - \sum_{l=j+1}^{t} S_l}{1 - \sum_{l=j}^{t} S_l}$

which completes the proof of equation (2.8.1.2) by induction.

Substituting for S_j from equation (2.8.3) into equations (2.8.1.2) and (2.8.1.3) gives:

$$\lambda_j = \frac{1 - \sum_{l=j+1}^{t} \left(\frac{e^{\beta_l}}{\sum_{k=1}^{t} e^{\beta_k}} \right)}{1 - \sum_{l=j}^{t} \left(\frac{e^{\beta_l}}{\sum_{k=1}^{t} e^{\beta_k}} \right)}$$

$$= \frac{\sum_{l=1}^{t} e^{\beta_l} - \sum_{l=j+1}^{t} e^{\beta_l}}{\sum_{l=1}^{t} e^{\beta_l} - \sum_{l=j}^{t} e^{\beta_l}}$$

$$= \frac{\sum_{l=1}^{j} e^{\beta_l}}{\sum_{l=1}^{j-1} e^{\beta_l}}$$

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

and $\lambda_t = \dfrac{1}{1 - \dfrac{e^{\beta_t}}{\sum_{l=1}^{t} e^{\beta_l}}}$

$$= \frac{\sum_{l=1}^{t} e^{\beta_l}}{\sum_{l=1}^{t-1} e^{\beta_l}}$$

$$= 1 + \frac{e^{\beta_t}}{\sum_{l=1}^{t-1} e^{\beta_l}}$$

which completes the proof of the theorem.

The parameters of the additive model can be estimated by least squares, which is equivalent to maximum likelihood estimation if the errors are assumed to be independently, normally distributed. 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 (2.8.3) and (2.8.4). 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

$$V (\underline{\lambda}) = \left(\frac{\partial \underline{\lambda}}{\partial \underline{\beta}} \right)' V (\underline{\beta}) \left(\frac{\partial \underline{\lambda}}{\partial \underline{\beta}} \right) \qquad (2.8.4)$$

$$\text{and } V (\underline{S}) = \left(\frac{\partial \underline{S}}{\partial \underline{\beta}} \right)' V (\underline{\beta}) \left(\frac{\partial \underline{S}}{\partial \underline{\beta}} \right) \qquad (2.8.5)$$

It is thus necessary to obtain the matrices of the first derivatives of the respective parameter vectors.

The (j,k)th element of $\left(\frac{\partial \lambda}{\partial \underline{\beta}}\right)$ can be obtained from equation (2.8.1.1) and is given by:

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

$$= \begin{cases} 0 & k > j \\ \lambda_j - 1 & k = j \\ -(\lambda_j - 1)(\lambda_k - 1) & k < j \end{cases} \qquad (2.8.6)$$

Similarly, the (j,k)th element of $\left(\frac{\partial S}{\partial \underline{\beta}}\right)$ can be obtained from equation (2.8.3) and is given by:

$$\frac{\partial S_j}{\partial \beta_k} = \begin{cases} -\dfrac{e^{\beta_j} e^{\beta_k}}{\left(\sum_{l=1}^{t} e^{\beta_l}\right)^2} & k \neq j \\ \dfrac{e^{\beta_j}\left(\sum_{l=1}^{t} e^{\beta_l} - e^{\beta_j}\right)}{\left(\sum_{l=1}^{t} e^{\beta_l}\right)^2} & k = j \end{cases}$$

$$= \begin{cases} -S_j S_k & k \neq j \\ S_j (1 - S_j) & k = j \end{cases} \qquad (2.8.7)$$

Estimates of the variance-covariance matrices can be obtained by substituting estimates of the parameters into equations (2.8.4) and (2.8.5).

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 (2.8.5) 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 (2.8.4).

2.9. Example

The method described in section 2.8 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 $\{ \beta_j : j = 2 , \ldots , t \}$ 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 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.

CHAPTER 3

Bayesian Linear Models and Credibility Theory

3.0 Introduction

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

$$\alpha_1 = \beta_1 = 0$$

on the first stage distribution have been retained. This also facilitates the comparison with the recursive approaches described in later chapters. 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.

3.1 Bayesian Estimation for Linear Models: Two Stage Models

This section will follow Lindley and Smith (1972). The linear model, described in chapter 2, which contains the models under consideration can be written

$$\underline{y}|\underline{\beta} \quad \sim \quad N(\, X\underline{\beta}\,,\, \Sigma\,) \qquad (3.1.1)$$

where $\underline{y}$ is a data vector of length n,

$\underline{\beta}$ is a parameter vector of length p,

X is an (nxp) design matrix

and Σ is an (nxn) dispersion matrix.

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

$$\underline{\beta}|\underline{\theta}_1 \quad \sim \quad N(\, A_1\underline{\theta}_1\,,\, C_1\,) \qquad (3.1.2)$$

where $\underline{\theta}_1$ is a vector of hyperparameters of length p_1,

A_1 is a (pxp_1) matrix

and C_1 is a (pxp) dispersion matrix.

In a straightforward Bayesian analysis, A_1 is an identity matrix, $\underline{\theta}_1$ contains the prior estimates of the parameters and C_1 is taken as a diagonal matrix with variances inversely proportional to the believed accuracy of the estimates.

3.1.1 Lemma

The posterior distribution of the parameters is

$$\underline{\beta}|\underline{y} \quad \sim \quad N(\, B\underline{b}\,,\, B\,) \tag{3.1.1.1}$$

where $B^{-1} = X'\Sigma^{-1}X + C_1^{-1}$ (3.1.1.2)

and $\underline{b} \quad = X'\Sigma^{-1}\underline{y} + C_1^{-1}A_1\underline{\theta}_1$ (3.1.1.3)

Proof

Bayes theorem states that the posterior density of $\underline{\beta}|\underline{y}$, $f(\,\underline{\beta}\mid\underline{y}\,)$ is related to the prior density of $\underline{\beta}$, $f(\underline{\beta})$ and the conditional density of $\underline{y}$, $f(\,\underline{y}\mid\underline{\beta}\,)$, by

$$f\,(\underline{\beta}\mid\underline{y}) \qquad \propto \qquad f(\,\underline{y}\mid\underline{\beta}\,)\,f(\underline{\beta}\,)$$

Hence $f(\underline{\beta}|\underline{y}) \quad \propto \quad \exp\{-\frac{1}{2}[\,(\underline{y}-X\underline{\beta})'\Sigma^{-1}(\underline{y}-X\underline{\beta}) + (\underline{\beta}-A_1\underline{\theta}_1)'\,C_1^{-1}(\underline{\beta}-A_1\underline{\theta}_1)]\}$

$$\propto \quad \exp\{-\tfrac{1}{2}[\underline{\beta}'X'\Sigma^{-1}X\underline{\beta} - 2\underline{\beta}'X'\Sigma^{-1}\underline{y} + \underline{\beta}'C_1^{-1}\ \beta - 2\underline{\beta}'C_1^{-1}A_1\underline{\theta}_1]\}$$

$$= \quad \exp\{-\tfrac{1}{2}[\underline{\beta}'B^{-1}\underline{\beta} - 2\underline{\beta}'\underline{b}]\ \}$$

Completing the square we have

$$f(\underline{\beta}\,|\,\underline{y}) \quad \propto \quad \exp\{-\tfrac{1}{2}[\,(\underline{\beta}-B\underline{b})'B^{-1}\ (\underline{\beta}-B\underline{b})]\}$$

which is the density of a normal distribution with mean $B\underline{b}$ and variance B.

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

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

$$= \quad X'\Sigma^{-1}X\hat{\underline{\beta}} + C_1^{-1}A_1\underline{\theta}_1$$

using the normal equations (2.1.2).

$\hat{\underline{\beta}}$ is the least-squares estimate of $\underline{\beta}$.

Then

$$\tilde{\underline{\beta}} \quad = \quad (X'\Sigma^{-1}X + C_1^{-1})^{-1}\,[X'\Sigma^{-1}X\hat{\underline{\beta}} + C_1^{-1}A_1\underline{\theta}_1] \qquad (3.1.1.4)$$

Equation (3.1.1.4) shows that the Bayes estimate is a weighted average of the least-squares estimate, $\hat{\underline{\beta}}$, and the prior mean, $A_1\underline{\theta}_1$, with weights proportional to the precision of each. It can be written in the usual way as a credibility formula:

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

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

As $C_1^{-1} \rightarrow 0$, $z \rightarrow 1$ and the estimate is based entirely on the data: the estimate is the least-squares estimate.

In least squares estimation theory it is necessary to have an extra restriction since $X'\Sigma^{-1}X$ is often singular. For example, in the chain-ladder model we need $\sum\alpha_i = \sum\beta_j = 0$.

In the Bayesian analysis this is not necessary since

$$X'\Sigma^{-1}X + C_1^{-1}$$

is , in general, non-singular if $C_1^{-1} \rightarrow 0$. For comparison purposes the restraint will be retained in all cases.

The alternative constraint, and the one which will be used in all future cases, is

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

3.2 Bayesian Estimation for Linear Models : Three Stage Models

In order to get an empirical Bayes formula, it is necessary to have a further prior distribution. The reason for this, and the way in which it is used will become clear in section 3.6. We define a prior distribution for a vector of hyperparameters:

$$\underline{\theta}_1 \mid \underline{\theta}_2 \quad \sim \quad N(A_2\underline{\theta}_2, C_2) \tag{3.2.1}$$

where $\underline{\theta}_2$ is a p_2 - dimensional vector

A_2 is a $(p_1 \times p_2)$ - dimensional matrix

C_2 is a $(p_2 \times p_2)$ - dimensional matrix.

We now have the following lemma :

3.2.1 Lemma

Suppose that

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

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

$$\underline{\beta}|\underline{y}, \underline{\theta}_2 \quad \sim \quad N(D\underline{d}, D) \tag{3.2.1.1}$$

$$\text{where } D^{-1} = X'\Sigma^{-1}X + [C_1 + A_1 C_2 A_1']^{-1} \tag{3.2.1.2}$$

$$\text{and } \underline{d} = X'\Sigma^{-1}\underline{y} + [C_1 + A_1C_2A_1']^{-1} A_1A_2\underline{\theta}_2 \tag{3.2.1.3}$$

Proof

$$\underline{\beta}|\underline{\theta}_1,\underline{\theta}_2 \quad \sim \quad N(A_1A_2\underline{\theta}_2\,, C_1 + A_1C_2A_1')$$

Hence, using lemma 3.1.1, the posterior distribution of $\underline{\beta}$ is $N(D\underline{d}\,, D)$ where D and $\underline{d}$ are given by equations (3.2.1.2) and (3.2.1.3).

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

$$\tilde{\underline{\beta}} = [X'\Sigma^{-1}X + (C_1 + A_1C_2A_1')^{-1}]^{-1}\,[X'\Sigma^{-1}X\hat{\underline{\beta}} + (C_1 + A_1C_2A_1')^{-1}A_1A_2\underline{\theta}_2] \tag{3.2.2}$$

This is a credibility formula with credibility factor

$$[X'\Sigma^{-1}X + (C_1 + A_1C_2A_1')^{-1}]^{-1}\,X'\Sigma^{-1}X$$

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

When the three-stage Bayesian model is used, a vague third-stage prior distribution will be used. In other words, C_2^{-1} will always be 0. Since it is not easy to see what happens to the posterior variance when $C_2^{-1} \to 0$ with repect to the usual norm and equation (3.2.1.2) is used, the following matrix identity is used:

$$[C_1 + A_1C_2A_1']^{-1} = C_1^{-1} - C_1^{-1}A_1(A_1'C_1^{-1}A_1 + C_2^{-1})^{-1}A_1'C_1^{-1}$$

Thus

$$D^{-1} = X'\Sigma^{-1}X + C_1^{-1} - C_1^{-1}A_1(A_1'C_1^{-1}A_1 + C_2^{-1})^{-1}A_1'C_1^{-1} \tag{3.2.3}$$

The proof of the following lemma is now clear.

3.2.2 Lemma

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

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

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

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

where $D_o^{-1} = X'\Sigma^{-1}X + C_1^{-1} - C_1^{-1}A_1(A_1'C_1^{-1}A_1)^{-1}A_1'C_1^{-1}$ (3.2.2.1)

and $\underline{d}_o = X'\Sigma^{-1}\underline{y}$ (3.2.2.2)

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

3.2.3 Variance Estimation

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

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

procedure will be described in each of the applications.

3.3 Bayesian Estimation for Claims Runoff Triangles

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

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

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

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

Proof

$$E(Z_{kl} \mid D) = E_{\theta|D}(E(Z_{kl} \mid \theta, D))$$

$$= E_{\theta|D}(e^{\theta+\frac{1}{2}\sigma^2})$$

$$= e^{\frac{1}{2}\sigma^2} E_{\theta|D}(e^{\theta})$$

$$= e^{\frac{1}{2}\sigma^2} e^{m+\frac{1}{2}\tau^2}$$ using the m.g.f. of the normal distribution

$$= e^{m+\frac{1}{2}\sigma^2+\frac{1}{2}\tau^2}$$

$$\text{Var}\left(Z_{kl} \mid D \right) = E_{\theta|D}\left(\text{Var}\left(Z_{kl} \mid \theta , D \right) \right) + \text{Var}_{\theta|D}\left(E\left(Z_{kl} \mid \theta , D \right) \right)$$

$$= E_{\theta|D}\left(e^{2\theta+\sigma^2}\left(e^{\sigma^2} - 1 \right) \right) + \text{Var}_{\theta|D}\left(e^{\theta+\frac{1}{2}\sigma^2} \right)$$

$$= E_{\theta|D}\left(e^{2\theta+\sigma^2}\left(e^{\sigma^2} - 1 \right) \right) + E_{\theta|D}\left(e^{2\theta+\sigma^2} \right) - \left(E_{\theta|D}\left(e^{\theta+\frac{1}{2}\sigma^2} \right) \right)^2$$

$$= e^{2\sigma^2} E_{\theta|D}\left(e^{2\theta} \right) + \left(e^{\frac{1}{2}\sigma^2} E_{\theta|D}\left(e^{\theta} \right) \right)^2$$

$$= e^{2\sigma^2} e^{2m+2\tau^2} - \left(e^{\frac{1}{2}\sigma^2} e^{m+\frac{1}{2}\tau^2} \right)^2$$

$$= e^{2m+\tau^2+\sigma^2}\left(e^{\tau^2+\sigma^2} - 1 \right)$$

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

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

$$\sum_{j>n-i+1} E\left(Z_{ij} \mid D \right) \qquad (3.3.1)$$

and the Bayes estimate of the variance is

$$\sum_{j>n-i+1} \left[\text{Var}\left(Z_{ij} \mid D \right) + 2 \sum_{k>j} \text{Cov}\left(Z_{ij} , Z_{ik} \mid D \right) \right] \qquad (3.3.2)$$

3.3.1 Example

The Bayesian estimation method gives different results from the least-squares estimation described in chapter 2, even if no prior information is assumed. In order to assess the affect of

the prior assumptions, it is useful to have the results of a Bayesian analysis when no prior information is used, and these are given in this example. The parameter estimates are the same as those in the least-squares analysis, given in figure 2.2, and so are not repeated here. The fitted values are shown in figure 3.1 and the predicted outstanding claims are shown in figure 3.2. These can be compared with the results in figures 2.3, 2.6 and 2.7.

Figure 3.1

277449 690094 709643 729030 407880 277109 248060 180660 267966 72013
357848 766940 610542 482940 527326 574398 146342 139950 227229 67948

398075 990124 1018171 1045987 585212 397587 355909 259205 384468
352118 884021 933894 1183289 445745 320996 527804 266172 425046

368073 915501 941434 967154 541106 367622 329085 239669
290507 1001799 926219 1016654 750816 146923 495992 280405

329616 819847 843071 866103 484570 329212 294701

310608 1108250 776189 1562400 272482 352053 206286

368704 917069 943047 968811 542034 368252
443160 693190 991983 769488 504851 470639

380341 946016 972814 999391 559142
396132 937085 847498 805037 705960

413142 1027599 1056708 1085577
440832 847631 1131398 1063269

453123 1127043 1158969

359480 1061648 1443370

399227 992989
376686 986608

364596
344014

Figure 3.2

							110927
						379508	102650
					228731	340091	91988
				351067	256033	380684	102968
			404479	362429	264319	393004	106300
		646825	439787	394066	287392	427310	115580
	1268870	710402	483014	432799	315640	469311	126940
1090101	1120657	627422	426594	382246	278771	414492	112113
973601	1001989	1030076	576709	392113	351349	256238	380990	103051

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

Figure 3.3

Predicted Outstanding Claims	Standard Error
110927	60216
482157	189896
660810	210040
1090752	304721
1530532	401125
2310959	601536
3806976	1056660
4452396	1375446
5066116	2049337

The Bayesian estimate of total outstanding claims is 19511632 and the Bayesian estimate of the standard error of total outstanding claims is 3194056. As before, a normal distribution can be used to find an upper 95% confidence bound on total outstanding claims:

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

This is a little higher than the unbiased estimate, reflecting the upward bias of the Bayesian estimate.

3.3.2 Example

This example gives the results from the row model in example 2.7, using the Bayesian estimates. The parameter estimates are unchanged and are given in figure 2.8. Figure 3.4 shows the fitted values, together with the actual observations, and figure 3.5 the predicted outstanding claims.

Figure 3.4

424619 803551 886061 783212 613300 444817 306000 202511 130117 81676
357848 766940 610542 482940 527326 574398 146342 139950 227229 67948

501886 949771 1047295 925731 724900 525759 361682 239361 153794
352118 884021 933894 1183289 445745 320996 527804 266172 425046

485179 918156 1012434 894916 700770 508258 349642 231393
290507 1001799 926219 1016654 750816 146923 495992 280405

432276 818041 902039 797336 624359 452838 311518
310608 1108250 776189 1562400 272482 352053 206286

417658 790378 871536 770372 603246 437525
443160 693190 991983 769488 504851 470639

384245 727148 801813 708743 554986
396132 937085 847498 805037 705960

377980 715292 788740 697187
440832 847631 1131398 1063269

350137 662600 730637
359480 1061648 1443370

365451 691581
376686 986608

292361
344014

Figure 3.5

						106553
					162966	103007
				224939	145197	91775
			327529	217332	140287	88671
		437548	301327	199946	129064	81578
	593410	430414	296414	196686	126959	80248
	702233	549697	398708	274579	182197	117607	74336
	829253	732947	573739	416147	286588	190166	122751	77587
601399	663402	586358	458991	332917	229271	152133	98201	62070

Figure 3.6 shows the predicted outstanding claims for each year of business together with the Bayes estimate of the standard error of the total outstanding claims.

Figure 3.6

Predicted Outstanding Claims	Standard Error
106553	49779
265973	92808
461911	135618
773820	200088
1149463	269032
1724132	372619
2299357	464456
3229178	610154
3184742	558099

The Bayes estimate of total outstanding claims is 13195129 and the standard error is 2163719.

3.4 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 in chapter 2 that the chain ladder linear model can be written as a linear model in the form:

$$\underline{Y}|\underline{\beta} \quad \sim \quad N(\, X\underline{\beta}\,,\, \Sigma)$$

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

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

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

$$\underline{\beta}\,|\underline{\theta}_1 \quad \sim \quad N(\,\underline{\theta}_1,\, C_1\,) \qquad (3.4.1)$$

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

From lemma 3.1.1 it can be seen that the Bayes estimate of the parameter vector is the solution, $\tilde{\underline{\beta}}$, of

$$(\,\sigma^{-2}X'X \;+\; C_1^{-1}\,)\,\tilde{\underline{\beta}} \;=\; \sigma^{-2}\,X'X\,\hat{\underline{\beta}} + C_1^{-1}\,\underline{\theta}_1 \qquad (3.4.2)$$

and the variance-covariance matrix of the estimates is

$$\mathrm{Var}(\tilde{\underline{\beta}}) \;=\; [\,\sigma^{-2}\,X'X + C_1^{-1}\,]^{-1} \qquad (3.4.3)$$

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

$$\tilde{\underline{\beta}} \quad = \quad z\,\hat{\underline{\beta}} \; + \; (1 - z)\,\underline{\theta}_1 \qquad (3.4.4)$$

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

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

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

$$s^2 \quad = \quad (\underline{y} - X\tilde{\underline{\beta}})'(\underline{y} - X\tilde{\underline{\beta}})/(n+2) \qquad (3.4.5)$$

Thus the equations which give the Bayes estimates are (3.4.2), with σ^2 replaced by s^2, and (3.4.5).

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

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

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

3.4.1 Example

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

To illustrate the method, the Taylor and Ashe data is reanalysed assuming that there is other information which should be taken into account. This information can be summarised in the following prior distribution on each row parameter:

$$\alpha_i \sim N(0.3, 0.05) \qquad (3.4.1.1)$$

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

Thus equation (3.4.1) can be written in this case as:

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

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

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

and L is large (in the limit infinite).

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

The Bayes estimates of the parameters and their standard errors are shown in figure 3.7:

Figure 3.7

	Bayes Estimate	Least Squares Estimate	Standard Error Of Bayes Estimate
Overall mean	6.178	6.106	0.125
Row Parameters	0.202	0.194	0.120
	0.168	0.149	0.123
	0.172	0.153	0.127
	0.276	0.299	0.133
	0.349	0.412	0.139
	0.400	0.508	0.147
	0.475	0.673	0.157
	0.360	0.495	0.172
	0.367	0.602	0.192
Column Parameters	0.893	0.911	0.161
	0.910	0.939	0.168
	0.915	0.965	0.175
	0.318	0.383	0.184
	−0.081	−0.005	0.196
	−0.201	−0.118	0.212
	−0.520	−0.439	0.236
	−0.129	−0.054	0.277
	−1.465	−1.393	0.370

These parameter estimates can be compared with the least-squares estimates from example 2.5. Perhaps it is easier to see how the prior assumptions have affected the estimates by looking at

figures 3.8 and 3.9. These show plots of the least-squares and Bayes estimates of the row and column parameters.

Figure 3.8

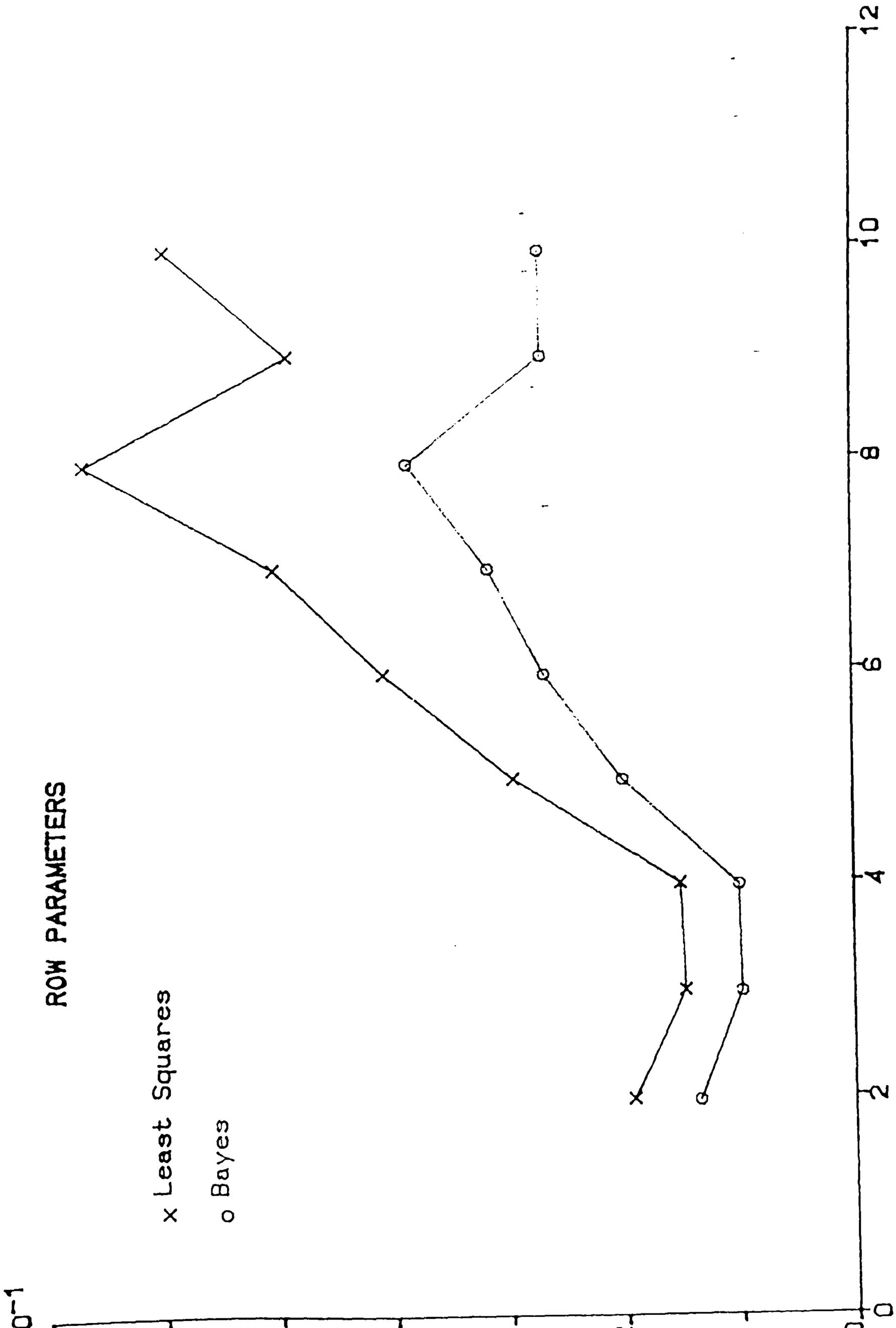
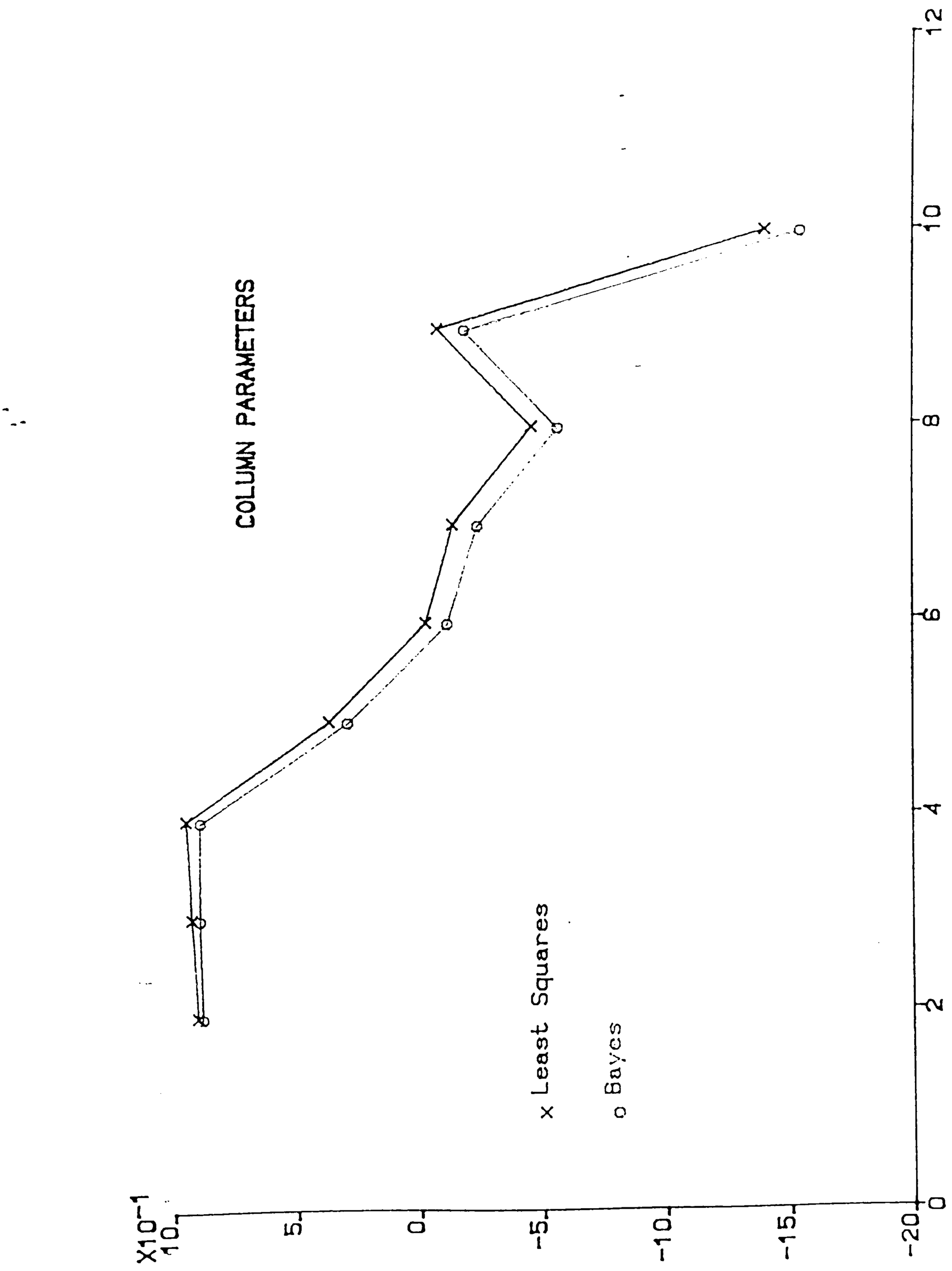


Figure 3.9


COLUMN PARAMETERS
x Least Squares
o Bayes
X10⁻¹
10
5
0
-5
-10
-15
-20
0
2
4
6
8
10
12

The situation is complicated by the interaction of the parameter estimates and the effect of the Bayes prior distribution is not, in general, straightforward. It can be seen from the plot of the row parameters, figure 3.8, that the general effect is to draw these estimates towards the prior mean : these estimates certainly show less variation, and the row parameter estimates are a lot lower. The variation of the column parameters (looking at figure 3.9) remains about the same, but their values have altered slightly. This is entirely due to the change of prior assumptions on the row parameters and the interaction between the row and column parameters. The estimate of the overall mean has changed for the same reason.

It can also be seen, by comparing figures 3.7 and 2.2 that the standard errors of the parameter estimates have been reduced. This is due to the extra prior information, and is an example of the usual property of Bayes estimates. That this should be the case is clear from equations (2.1.4) and (3.4.3).

The fitted values are shown in figure 3.10 (with the observed values).

Figure 3.10

296336 724378 737686 741606 409126 274988 244664 178781 266871 72202
357848 766940 610542 482940 527326 574398 146342 139950 227229 67948

430213 1051361 1070410 1075816 593327 398653 354525 258877 385940
352118 884021 933894 1183289 445745 320996 527804 266172 425046

402069 982569 1000359 1005399 554483 372548 331302 241911
290507 1001799 926219 1016654 750816 146923 495992 280405

359925 879570 895487 899990 496344 333480 296554
310608 1108250 776189 1562400 272482 352053 206286

386236 943859 960930 965753 532606 357839
443160 693190 991983 769488 504851 470639

382292 934212 951100 955864 527147
396132 937085 847498 805037 705960

396316 968471 985969 990898
440832 847631 1131398 1063269

396284 968382 985867
359480 1061648 1443370

369365 902588
376686 986608

298643
344014

The predicted outstanding claims are shown in figure 3.11 and the row totals and their standard errors in figure 3.12. Also given in figure 3.12 are the row totals when there is no prior information, taken from figure 3.3.

Figure 3.11

								385439	111748
							231078	345129	93517
						339352	248041	370448	100365
					377571	335993	245581	366755	99352
				582502	391565	348440	254671	380312	103010
			1056177	582716	391702	348554	254746	380401	103018
		979747	985010	543442	365295	325047	237555	354704	96039
	778495	792825	797067	439740	295579	263002	192199	286951	77671

Figure 3.12

Bayesian Estimates		Standard Error
With Prior	No Prior	With Prior
111748	110927	60516
489893	482157	191702
669724	660810	207990
1058206	1090752	282991
1425252	1530532	348013
2060499	2310959	482661
3117315	3806976	745547
3886838	4452396	936372
3923530	5066116	982585

The Bayesian estimate of overall total outstanding claims is 16743004 and the Bayesian

estimate of the standard error of total outstanding claims is 1995669.

These can be compared with the results for the analysis in which no prior information is assumed (example 3.3). Looking first at the standard errors of the estimates of outstanding claims for each year of business, it can be seen that these have been quite considerably reduced. This is because the prior information (which was fairly precise) gave tight estimates of the parameters. The effect of the prior estimates on the estimates of outstanding claims is clear: the estimates of the total outstanding claims for the later rows have been drawn down due to the effect of the prior mean.

3.5 Bayes Estimates for Row Models

The linear model which is applied to the data row by row, derived in section 2.3 is

$$\underline{y}|\underline{\beta} \quad \sim \quad N(X\underline{\beta}, \sigma^2 I) \qquad (3.5.1)$$

Again, it is assumed that the errors are independently, identically distributed with variance σ^2.

The matrix X consists of row vectors X_1, X_2, ...,X_t :

$$X \quad = \quad \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_t \end{bmatrix}$$

(assuming there are t rows in the triangle).

As in section 3.4, the prior distribution of the parameters is

$$\underline{\beta}\,|\,\underline{\theta}_1 \quad \sim \quad N(\underline{\theta}_1, C_1) \tag{3.5.2}$$

where $\underline{\theta}_1$ contains the prior estimates of the parameters and C_1 (which is often diagonal) is their variance-covariance matrix.

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

$$(s^{-2}X'X + C_1^{-1})\tilde{\underline{\beta}} \quad = \quad s^{-2}X'X\hat{\underline{\beta}} \quad + \quad C_1^{-1}\underline{\theta}_1 \tag{3.5.3}$$

where $\hat{\underline{\beta}}$ is the least squares estimate of $\underline{\beta}$

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

This iterative procedure is illustrated in the following example.

3.5.1 Example

Example 2.7 found the least-squares estimates of the parameters of the gamma model. It is now possible to incorporate prior belief about the parameter values in the estimates. For example, suppose that there is prior knowledge which can be quantified in the following prior distribution on the parameter vector $\underline{\beta}$:

$$\underline{\beta} \quad \sim \quad N\left[\begin{bmatrix} 0 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} L & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.1 \end{bmatrix}\right]$$

where L is large.

Thus the prior distributions of the three parameters are:

$$\delta \sim N(2, 0.1)$$

$$\gamma \sim N(-1, 0.1)$$

and ν has a vague prior distribution, so that it has been assumed that there is no prior knowledge about ν. The parameters are assumed to be independent based on the prior information.

The Bayes estimates of the parameters are shown in figure 3.13, along with their standard errors:

Figure 3.13

Parameter	Bayes Estimate	Least-Squares Estimate	Bayes Standard Error
ν	6.523	6.545	0.108
δ	1.955	1.877	0.191
γ	−0.686	−0.663	0.056

The residual variance is 0.162.

It can be seen that the parameter estimates have been drawn towards the prior means.

The fitted values are shown in figure 3.14.

Figure 3.14

415330	810830	901911	796849	620557	446191	303625	198446	125769	77796
357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
490906	958374	1066030	941850	733478	527383	358874	234557	148655	
352118	884021	933894	1183289	445745	320996	527804	266172	425046	
474565	926473	1030545	910498	709063	509828	346928	226749		
290507	1001799	926219	1016654	750816	146923	495992	280405		
422819	825451	918175	811219	631747	454237	309100			
310608	1108250	776189	1562400	272482	352053	206286			
408521	797538	887126	783786	610384	438876				
443160	693190	991983	769488	504851	470639				
375839	733735	816156	721083	561553					
396132	937085	847498	805037	705960					
369712	721771	802849	709327						
440832	847631	1131398	1063269						
342477	668602	743707							
359480	1061648	1443370							
357456	697845								
376686	986608								
285965									
344014									

Figure 3.15 shows the predicted values for the lower triangle.

Figure 3.15

								101191
							157324	97822
						220367	140169	87156
					325032	212915	135429	84209
				438944	299029	195882	124595	77472
			600335	431787	294154	192688	122564	76209
		714174	556112	399980	272485	178494	113535	70595
	843797	745410	580435	417474	284403	186301	118501	73682
606902	675038	596328	464348	333979	227522	149041	94801	58946

Figure 3.16 shows the total predicted outstanding claims for each year of business, together with their standard errors.

Figure 3.16

Bayesian Estimates		Standard Error
With Prior	No Prior	With Prior
101191	106553	46485
255146	265973	87558
447692	461911	130072
757585	773820	195379
1135922	1149463	266844
1717737	1724132	373595
2305374	2299357	468117
3250003	3229178	615614
3206904	3184742	562754

The Bayesian estimate of total outstanding claims is 13177554 and the Bayesian estimate of the standard error of total outstanding claims is 2165948.

These values and the figures in the above table can be compared with the values in figure 3.6. In this case the fairly imprecise prior distribution which is close to the distribution of the parameters implied by the data has had little effect on the estimates and predictions.

3.6 Empirical Bayes Estimates for the Analysis of Variance Model

Section 3.4 described the use of a two-stage conventional Bayesian model to analyse claims data. This section uses the three-stage Bayesian model described in section 3.2 to derive empirical Bayes estimates for the analysis of variance (chain ladder) model. This method uses an improper prior distribution at the third stage for the row parameters and improper priors at the second stage for the overall mean and the column parameters. This means that for the overall mean and the column parameters the same assumptions are made as for the least squares 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 (section 2.2) is

$$\underline{Y}|\underline{\beta} \quad \sim \quad N(X\underline{\beta}\ ,\ \sigma^2 I) \qquad (3.6.1)$$

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 section 2.2.

As in credibility theory, a structure is put onto the row parameters $\alpha_2, \alpha_3, \ldots, \alpha_t$: they are

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

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

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

In terms of equation (3.1.2),

$$\underline{\theta}_1 = \begin{bmatrix} \mathrm{w} \\ \theta \\ \xi_2 \\ \vdots \\ \xi_t \end{bmatrix} \qquad \text{and} \qquad \mathrm{A}_1 = \begin{bmatrix} 1 & 0 & \dots & & 0 \\ 0 & 1 & 0 & & \vdots \\ \vdots & 1 & \vdots & & \\ \vdots & \vdots & \vdots & & \\ 0 & 1 & 0 & & \\ \vdots & 0 & 1 & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & \dots & \dots & 0 & 1 \end{bmatrix}$$

θ is the mean of the common distribution of the row parameters α_2, ... ,α_t.

Although the assumptions on the estimation of μ and β_2, ... ,β_t are the same as for the least squares estimation, the estimators produced will not be the same because of the treatment of the row parameters.

A vague prior distribution (a third-stage distribution) is used for θ. Since $\sigma_\mu^{-2} \rightarrow 0$ and

$\sigma_\beta^{-2} \rightarrow 0$, 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.

3.6.1 Theorem

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

where X is as given in equation (2.2.20) and $\underline{\beta} = (\mu, \alpha_2 , \ldots , \alpha_t , \beta_2 , \ldots , \beta_t)'$

and $\underline{\beta} | w , \theta , \xi \sim N (A_1 \underline{\theta}_1 , C_1)$

where A_1 , $\underline{\theta}_1$, C_1 are given in (3.6.2), $\sigma_\mu^{-2} = 0$, $\sigma_\beta^{-2} = 0$ and a vague prior at the third stage is used, then the Bayes estimate of $\underline{\beta}$, $\tilde{\underline{\beta}}$, is given by

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

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

and has a credibility interpretation.

Proof

The posterior distribution of $\underline{\beta} \mid \underline{y}$ is given by lemma 3.2.2 :

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

where $D_o^{-1} = \sigma^{-2} X'X + C_1^{-1} - C_1^{-1} A_1 (A_1' C_1^{-1} A_1)^{-1} A_1' C_1^{-1}$ (3.6.1.1)

and $\underline{d}_o = \sigma^{-2} X' \underline{y}$ (3.6.1.2)

$$C_1^{-1}A_1 = \begin{bmatrix} \sigma_\mu^{-2} & & & & \\ & \sigma_\alpha^{-2} & & & \\ & & \ddots & & \\ & & & \sigma_\beta^{-2} & \\ & & & & \ddots \\ & & & & & \sigma_\beta^{-2} \end{bmatrix} \begin{bmatrix} 1 & 0 & \dots & & 0 \\ 0 & 1 & 0 & & \vdots \\ \vdots & 1 & \vdots & & \\ \vdots & \vdots & \vdots & & \\ 0 & 1 & 0 & & \\ \vdots & 0 & 1 & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & \dots & \dots & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_\mu^{-2} & & & & \\ & \sigma_\alpha^{-2} & & & \\ & \vdots & & & \\ & \sigma_\alpha^{-2} & & & \\ & & \sigma_\beta^{-2} & & \\ & & & \ddots & \\ & & & & \sigma_\beta^{-2} \end{bmatrix} \tag{3.6.1.3}$$

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

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

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

$$\text{Hence } (A_1' C_1^{-1} A_1)^{-1} = \begin{bmatrix} \sigma_\mu^2 & & & & \\ & (t-1)^{-1}\sigma_\alpha^2 & & & \\ & & \sigma_\beta^2 & & \\ & & & \ddots & \\ & & & & \sigma_\beta^2 \end{bmatrix} \qquad (3.6.1.4)$$

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

$$= \begin{bmatrix} 1 & & & & \\ & (t-1)^{-1} \dots (t-1)^{-1} & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix} \qquad (3.6.1.5)$$

Using (3.6.1.3) and (3.6.1.5) :

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

$$= \begin{bmatrix} \sigma_\mu^{-2} & & & & \\ & (t-1)^{-1}\sigma_\alpha^{-2} & \dots & (t-1)^{-1}\sigma_\alpha^{-2} & \\ & \vdots & \ddots & & \\ & (t-1)^{-1}\sigma_\alpha^{-2} & \dots & (t-1)^{-1}\sigma_\alpha^{-2} & \\ & & & & \sigma_\beta^{-2} \\ & & & & & \ddots \\ & & & & & & \sigma_\beta^{-2} \end{bmatrix} \qquad (3.6.1.6)$$

$$= \begin{bmatrix} \sigma_{\mu}^{-2} & & \\ & (t-1)^{-1}\sigma_{\alpha}^{-2}\mathrm{J}_{t-1} & \\ & & \sigma_{\beta}^{-2}\mathrm{I}_{t-1} \end{bmatrix} \qquad (3.6.1.7)$$

where J_n is a square (nxn) matrix, all of whose elements are 1

and I_n is an (nxn) identity matrix.

Substituting (3.6.1.7) into (3.6.1.1) and (3.6.1.2) gives the Bayes posterior mean, $\tilde{\underline{\beta}}$, as the solution of

$$\left(\sigma^{-2}\mathrm{X}'\mathrm{X}+\begin{bmatrix} \sigma_{\mu}^{-2} & & & & \\ & \sigma_{\alpha}^{-2} & & & \\ & & \ddots & & \\ & & & \sigma_{\beta}^{-2} & \\ & & & & \ddots \\ & & & & & \sigma_{\beta}^{-2} \end{bmatrix} - \begin{bmatrix} \sigma_{\mu}^{-2} & & \\ & (t-1)^{-1}\sigma_{\alpha}^{-2}\mathrm{J}_{t-1} & \\ & & \sigma_{\beta}^{-2}\mathrm{I}_{t-1} \end{bmatrix}\right)\tilde{\underline{\beta}} = \sigma^{-2}\mathrm{X}'\underline{\mathrm{Y}} \quad (3.6.1.8)$$

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

$$\left(\sigma^{-2}\mathrm{X}'\mathrm{X}+\begin{bmatrix} 0 & & & & \\ & \sigma_{\alpha}^{-2} & & & \\ & & \ddots & & \\ & & & \sigma_{\alpha}^{-2} & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix} - \begin{bmatrix} 0 & & & \\ & (t-1)^{-1}\sigma_{\alpha}^{-2}\mathrm{J}_{t-1} & & \\ & & 0 & \\ & & & \ddots \\ & & & & 0 \end{bmatrix}\right)\tilde{\underline{\beta}} = \sigma^{-2}\mathrm{X}'\underline{\mathrm{Y}} \quad (3.6.1.9)$$

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

$$\tilde{\underline{\beta}} = (\tilde{\mu}\ ,\ \tilde{\alpha}_2\ ,\ \ldots\ ,\ \tilde{\alpha}_t\ ,\ \tilde{\beta}_2\ ,\ \ldots\ ,\ \tilde{\beta}_t\ )' \qquad (3.6.1.10)$$

There should be no confusion between the vector of parameters $\underline{\beta}$ and the column parameters β_2 , ... , β_t which are components of $\underline{\beta}$.

Equation (3.6.1.9) can be rewritten as

$$\left(\sigma^{-2}X'X + \begin{bmatrix} 0 & & & & & \\ & \sigma_{\alpha}^{-2} & & & & \\ & & \ddots & & & \\ & & & \sigma_{\alpha}^{-2} & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix}\right)\underline{\tilde{\beta}} - \sigma_{\alpha}^{-2}\begin{bmatrix} 0 \\ \tilde{\alpha} \\ \vdots \\ \tilde{\alpha} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \sigma^{-2}X'\underline{Y} \qquad (3.6.1.11)$$

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

Hence

$$\underline{\tilde{\beta}} = \left(\sigma^{-2}X'X + \begin{bmatrix} 0 & & & & & \\ & \sigma_{\alpha}^{-2} & & & & \\ & & \ddots & & & \\ & & & \sigma_{\alpha}^{-2} & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix}\right)^{-1}\left(\sigma^{-2}X'X\underline{\hat{\beta}} + \begin{bmatrix} 0 & & & & & \\ & \sigma_{\alpha}^{-2} & & & & \\ & & \ddots & & & \\ & & & \sigma_{\alpha}^{-2} & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix}\begin{bmatrix} 0 \\ \tilde{\alpha} \\ \vdots \\ \tilde{\alpha} \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right) \qquad (3.6.1.12)$$

3.6.2 Remarks

From equation (3.6.1.12) 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 least squares 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 least squares estimates and the (weighted) average of the estimates to which the credibility thoery type assumptions have been applied. The weights depend in a natural way 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

$$s^2 = \frac{\nu\lambda+(\underline{y}-X\underline{\tilde{\beta}})'(\underline{y}-X\underline{\tilde{\beta}})}{n+\nu+2} \qquad (3.6.2.1)$$

$$s_\alpha^2 = \frac{\nu_\alpha\lambda_\alpha+\sum_{i=2}^{t}(\tilde{\alpha}_i-\tilde{\alpha})^2}{t+\nu_\alpha+1} \qquad (3.6.2.2)$$

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

Again, the estimates are obtained by iterating between (3.6.1.12) and (3.6.2.1),(3.6.2.2). This is illustrated in the following example.

3.6.3 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 results can be compared with the least-squares results given in example 2.5 and the Bayesian estimates with no prior information in example

3.3.1.

The estimates of the parameters and their standard errors are shown in figure 3.17:

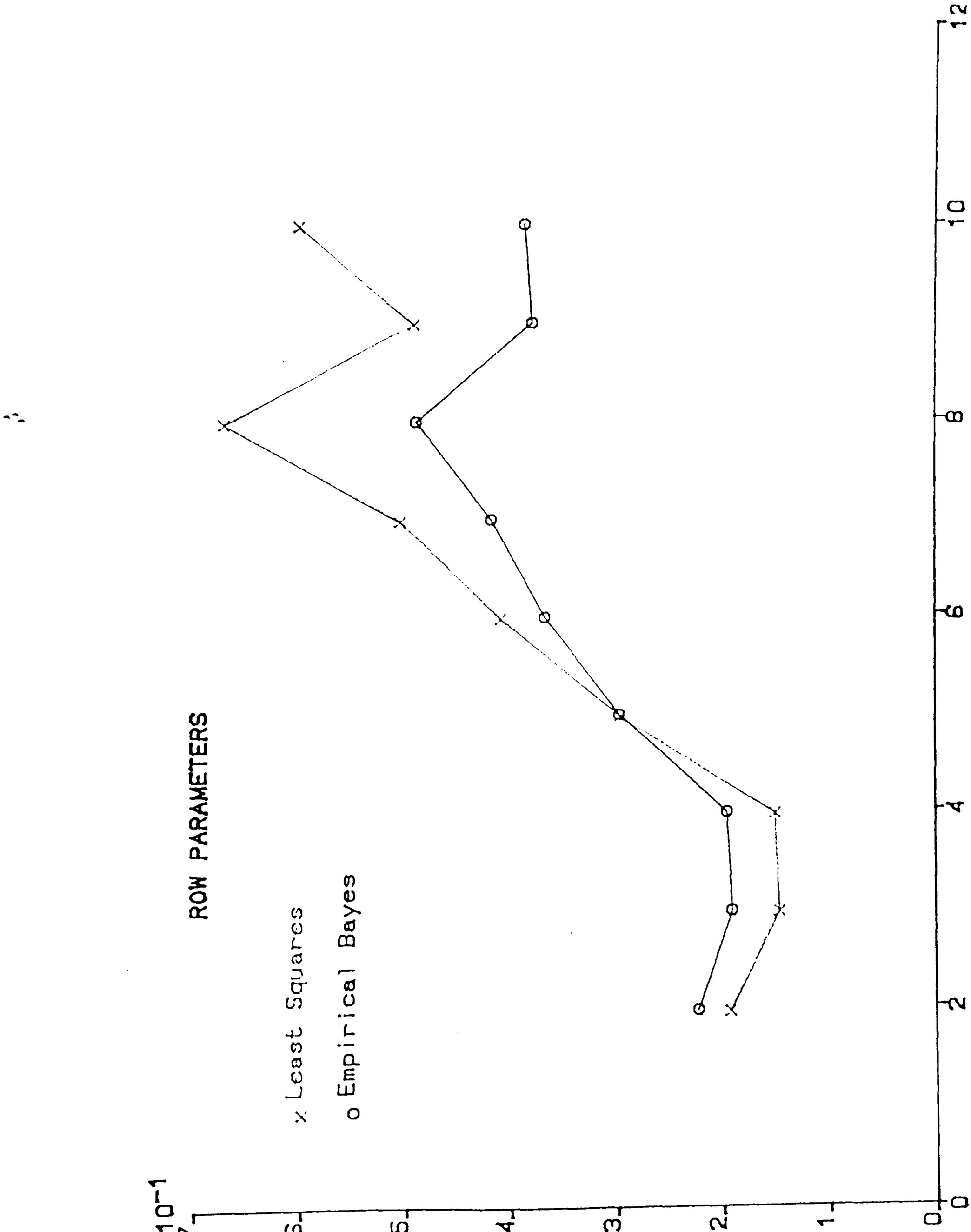
Figure 3.17

	Empirical Bayes Estimate	Least Squares 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 (shown in figure 2.2) have increased, while the higher ones have decreased. This can be seen more clearly from the following graph (figure 3.18) which shows a plot of the least-squares and empirical Bayes estimates of the row parameters.

Figure 3.18


ROW PARAMETERS
x Least Squares
o Empirical Bayes
X10-1
7
6
5
4
3
2
1
0
0
2
4
6
8
10
12

The fitted values are shown in figure 3.19.

Figure 3.19

290487 709948 722855 726022 400256 268865 239120 174750 260890 70607
357848 766940 610542 482940 527326 574398 146342 139950 227229 67948

429454 1049667 1068833 1073607 591935 397668 353728 258565 386181
352118 884021 933894 1183289 445745 320996 527804 266172 425046

402047 982684 1000631 1005104 554168 372298 331164 242073
290507 1001799 926219 1016654 750816 146923 495992 280405

360147 880273 896353 900363 496420 333504 296657
310608 1108250 776189 1562400 272482 352053 206286

385648 942608 959829 964125 531578 357124
443160 693190 991983 769488 504851 470639

380989 931222 948238 952485 525162
396132 937085 847498 805037 705960

394246 963628 981239 985637
440832 847631 1131398 1063269

392849 960218 977769
359480 1061648 1443370

367802 899000
376686 986608

297090
344014

The following figure shows the predicted outstanding claims for the lower triangle, and can be compared with figure 3.2.

Figure 3.20

					109448
				377062	102506
			225936	337865	91854
		330835	241997	361888	98389
	367359	326920	239136	357615	97231
565796	380243	338389	247527	370169	100649
1022573	563964	379014	337297	246731	368985	100332
953224	957737	528209	354987	315917	231094	345609	93982
756346	770356	774008	426882	286892	255320	186770	279330	75965

Figure 3.21 shows the row totals and their standard errors. For comparison purposes, the Bayes estimates with no prior assumptions (from figure 3.3) are also given.

Figure 3.21

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 been 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 least-squares 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.

3.6.4 Empirical Bayes Assumptions on the Row and Column Parameters

It is also possible to assume that the column parameters are exchangeable. The three-stage distribution in this case is

$$\underline{y} \mid \underline{\beta} \quad \sim \quad N(X\underline{\beta} , \sigma^2 I) \tag{3.6.4.1}$$

$$\underline{\beta} \mid w,\theta,\xi \quad \sim \quad N\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \vdots \\ \vdots & \vdots & \\ & 1 & 0 \\ & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} w \\ \theta \\ \xi \end{bmatrix}, \begin{bmatrix} \sigma^2_\mu & & & & & \\ & \sigma^2_\alpha & & & & \\ & & \ddots & & & \\ & & & \sigma^2_\alpha & & \\ & & & & \sigma^2_\beta & \\ & & & & & \ddots \\ & & & & & & \sigma^2_\beta \end{bmatrix} \right) \tag{3.6.4.2}$$

Again $\sigma_\mu^{-2} \rightarrow 0$ and a vague prior at the third stage for θ and ξ is used.

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

$$\tilde{\underline{\beta}} = \left(\sigma^{-2}X'X + \begin{bmatrix} 0 & & & & & \\ & \sigma_\alpha^{-2} & & & & \\ & & \ddots & & & \\ & & & \sigma_\alpha^{-2} & & \\ & & & & \sigma_\beta^{-2} & \\ & & & & & \ddots \\ & & & & & & \sigma_\beta^{-2} \end{bmatrix} \right)^{-1} \left(\sigma^{-1}X'X\hat{\underline{\beta}} + \begin{bmatrix} 0 & & & & & \\ & \sigma_\alpha^{-2} & & & & \\ & & \ddots & & & \\ & & & \sigma_\alpha^{-2} & & \\ & & & & \sigma_\beta^{-2} & \\ & & & & & \ddots \\ & & & & & & \sigma_\beta^{-2} \end{bmatrix} \begin{bmatrix} 0 \\ \tilde{\alpha} \\ \vdots \\ \tilde{\alpha} \\ \tilde{\beta} \\ \vdots \\ \tilde{\beta} \end{bmatrix} \right) \tag{3.6.4.3}$$

where $\tilde{\alpha} = \frac{1}{t-1} \sum_{i=2}^{t} \tilde{\alpha}_i$ and $\tilde{\beta} = \frac{1}{t-1} \sum_{j=2}^{t} \tilde{\beta}_j$

3.6.5 Example

The estimates of the parameters and their standard errors are shown in figure 3.22:

Figure 3.22

	Empirical Bayes Estimate	Least Squares Estimate	Standard Error Of Bayes Estimate
Overall Mean	6.122	6.106	0.131
Row Parameters	0.254	0.194	0.125
	0.225	0.149	0.129
	0.235	0.153	0.134
	0.341	0.299	0.139
	0.415	0.412	0.144
	0.466	0.508	0.151
	0.537	0.673	0.159
	0.424	0.495	0.171
	0.429	0.602	0.186
Column Parameters	0.878	0.911	0.129
	0.894	0.939	0.133
	0.896	0.965	0.139
	0.317	0.383	0.146
	-0.066	-0.005	0.156
	-0.175	-0.118	0.168
	-0.464	-0.439	0.187
	-0.081	-0.054	0.218
	-1.168	-1.393	0.286

The estimate of the variance of the distribution of the row parameters is 0.0296 and the estimate of the variance of the distribution of the column parameters is 0.4135.

The fitted values are shown in figure 3.23.

Figure 3.23

280471 674775 685943 688035 385698 263190 236407 177439 261560 89451
357848 766940 610542 482940 527326 574398 146342 139950 227229 67948

427007 1027429 1044516 1047788 587423 400887 360145 270370 398696
352118 884021 933894 1183289 445745 320996 527804 266172 425046

400971 964794 980844 983921 551620 376456 338199 253898
290507 1001799 926219 1016654 750816 146923 495992 280405

360990 868604 883057 885831 496629 338929 304488
310608 1108250 776189 1562400 272482 352053 206286

387869 933289 948821 951805 533619 364174
443160 693190 991983 769488 504851 470639

384279 924660 940052 943012 528691
396132 937085 847498 805037 705960

398147 958039 973991 977061
440832 847631 1131398 1063269

396276 953550 969431
359480 1061648 1443370

370066 890495
376686 986608

298160
344014

The predicted outstanding claims are shown in figure 3.24 and the row totals, with standard errors, in figure 3.25.

Figure 3.24

142697

390587 134051

238273 351747 120726

340896 256082 378044 129756

375880 337829 253781 374652 128597

570618 389553 350120 263016 388293 133285

1013058 568117 387848 348590 261870 386610 132714

943218 946423 530753 362342 325670 244655 361204 124001

747719 760352 762942 427861 292101 262542 197235 291204 99979

Figure 3.25

Empirical Bayes Estimates	Bayes No Prior	Standard Error of Empirical Bayes
142697	110927	59015
524638	482157	157033
710746	660810	172221
1104778	1090752	233349
1470739	1530532	285575
2094885	2310959	390731
3098807	3806976	587274
3838265	4452396	733542
3841936	5066116	763092

The empirical Bayes estimate of total outstanding claims is 16827488 and the empirical Bayes estimate of the standard error is 1346017.

3.7 Empirical Bayes Estimates for Row Models

In this section the theory of section 3.2 will be applied to the row models described in section 2.6. Because each row is treated separately the estimators which arise will have a complete credibility interpretation: they will combine the estimators for each row with those of all the rows in the triangle.

Assuming that there are t rows in the triangle, the ith row can be modelled by

$$\underline{Y}_i \quad \sim \quad N(X_i \underline{\beta}_i , \sigma_i^2 I_{t-i+1}) \qquad (3.7.1)$$

where X_i is the design matrix

$\underline{\beta}_i$ is the vector of parameters

and the errors are assumed to be independent but not necessarily identically distributed. The errors are in each row are assumed to be identically distributed with variance σ_i^2.

The parameters in each row are assumed to be independently, identically distributed with the following distribution:

$$\underline{\beta}_i \quad \sim \quad N(\underline{\theta} , C) \qquad (3.7.2)$$

This is the same assumption as is made in calculating credibility premiums. The parameters play the role of risk parameters and the distribution (3.7.2) is the equivalent of the structure of the risks.

The third stage distribution is a non-informative prior and so lemma 3.2.2 can be used.

3.7.1 Theorem

If $\underline{Y}_i \sim N(X_i \underline{\beta}_i, \sigma_i^2 I_{t-i+1})$

and $\underline{\beta}_i \sim N(\underline{\theta}, C)$

and a vague prior at the third stage is used, then the Bayes estimates of the row parameters in row i are $\tilde{\underline{\beta}}_i$, where

$$\tilde{\underline{\beta}}_i = (\sigma_i^{-2}\mathrm{X}_i'\,\mathrm{X}_i + \mathrm{C}^{-1})^{-1}\left(\sigma_i^{-2}\mathrm{X}_i'\mathrm{X}_i\hat{\underline{\beta}}_i + \mathrm{C}^{-1}(\sum_{i=1}^{t}\mathrm{W}_i)^{-1}\sum_{i=1}^{t}\mathrm{W}_i\hat{\underline{\beta}}_i\right)$$

and $\mathrm{W}_i = (\mathrm{X}_i'\mathrm{X}_i\sigma_i^{-2} + \mathrm{C}^{-1})^{-1}\mathrm{X}_i'\mathrm{X}_i\sigma_i^{-2}$

Proof

The linear model for the whole triangle can be written as

$$\underline{\mathrm{Y}} \sim \mathrm{N}\left(\mathrm{X}\begin{bmatrix}\underline{\beta}_1\\ \underline{\beta}_2\\ \vdots\\ \underline{\beta}_t\end{bmatrix}, \quad \Sigma\right) \qquad (3.7.1.1)$$

where $\mathrm{X} = \begin{bmatrix}\mathrm{X}_1 & & & \\ & \mathrm{X}_2 & & \\ & & \ddots & \\ & & & \mathrm{X}_t\end{bmatrix}$

and $\Sigma = \begin{bmatrix}\sigma_1^2\mathrm{I}_t & & & \\ & \sigma_2^2\mathrm{I}_{t-1} & & \\ & & \ddots & \\ & & & \sigma_t^2\end{bmatrix}$

The second stage prior distribution is (from equation (3.7.2)):

$$\begin{bmatrix}\underline{\beta}_1\\ \underline{\beta}_2\\ \vdots\\ \underline{\beta}_t\end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix}\mathrm{I}_p\\ \mathrm{I}_p\\ \vdots\\ \mathrm{I}_p\end{bmatrix}\underline{\theta}, \begin{bmatrix}\mathrm{C} & & & \\ & \mathrm{C} & & \\ & & \ddots & \\ & & & \mathrm{C}\end{bmatrix}\right) \qquad (3.7.1.2)$$

and it has been assumed that the parameter vector for each row has dimension p.

The third stage prior distribution is non-informative and hence, from lemma 3.2.2

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

with D_o^{-1} and $\underline{d}_o$ given by equations (3.2.2.1) and (3.2.2.2).

$$\text{Here} \quad A_1 = \begin{bmatrix} I_p \\ I_p \\ \vdots \\ I_p \end{bmatrix} \qquad \text{and} \qquad C_1 = \begin{bmatrix} C & & & \\ & C & & \\ & & \ddots & \\ & & & C \end{bmatrix}$$

$$\text{Hence} \quad C_1^{-1} A_1 \quad = \begin{bmatrix} C^{-1} & & & \\ & C^{-1} & & \\ & & \ddots & \\ & & & C^{-1} \end{bmatrix} \begin{bmatrix} I_p \\ I_p \\ \vdots \\ I_p \end{bmatrix} \quad = \begin{bmatrix} C^{-1} \\ C^{-1} \\ \vdots \\ C^{-1} \end{bmatrix}$$

$$\text{and} \quad A_1' C_1^{-1} A_1 \quad = \begin{bmatrix} I_p & I_p & \dots & I_p \end{bmatrix} \begin{bmatrix} C^{-1} \\ C^{-1} \\ \vdots \\ C^{-1} \end{bmatrix}$$

$$= tC^{-1}$$

$$C_1^{-1} A_1 (A_1' C_1^{-1} A_1)^{-1} A_1' C_1^{-1} \quad = \begin{bmatrix} C^{-1} \\ C^{-1} \\ \vdots \\ C^{-1} \end{bmatrix} t^{-1} \begin{bmatrix} C^{-1} & C^{-1} & \dots & C^{-1} \end{bmatrix}$$

$$= t^{-1} \begin{bmatrix} C^{-1} & C^{-1} & \dots & C^{-1} \\ \vdots & & & \\ C^{-1} & \dots & & C^{-1} \end{bmatrix}$$

$$\text{Hence } D_o^{-1} = \begin{bmatrix} \sigma_1^{-2} X_1' X_1 & & \\ & \ddots & \\ & & \sigma_t^{-2} X_t' X_t \end{bmatrix} + \begin{bmatrix} C^{-1} & & \\ & C^{-1} & \\ & & \ddots \\ & & & C^{-1} \end{bmatrix} - t^{-1} \begin{bmatrix} C^{-1} & C^{-1} & \dots & C^{-1} \\ \vdots & & & \\ C^{-1} & \dots & & C^{-1} \end{bmatrix}$$

(using equation (3.2.2.1)).

And from equation (3.2.2.2), the Bayes estimates, $\tilde{\underline{\beta}}_i$, are the solutions of

$$\left(\begin{bmatrix} \sigma_1^{-2} X_1' X_1 + C^{-1} & & \\ & \ddots & \\ & & \sigma_t^{-2} X_t' X_t + C^{-1} \end{bmatrix} - t^{-1} \begin{bmatrix} C^{-1} & C^{-1} & \dots & C^{-1} \\ \vdots & & & \\ C^{-1} & \dots & & C^{-1} \end{bmatrix} \right) \begin{bmatrix} \tilde{\underline{\beta}}_1 \\ \vdots \\ \tilde{\underline{\beta}}_t \end{bmatrix} = \begin{bmatrix} \sigma_1^{-2} X_1 \underline{Y}_1 \\ \vdots \\ \sigma_t^{-2} X_t \underline{Y}_t \end{bmatrix}$$

(3.7.1.3)

In this case the ith row can be considered because the first matrix in equation (3.7.1.3) is diagonal. This gives the Bayes estimate of the parameters in the ith row:

$$(\sigma_i^{-2} X_i' X_i + C^{-1}) \tilde{\underline{\beta}}_i - C^{-1} \tilde{\underline{\beta}} = \sigma_i^{-2} X_i \underline{Y}_i \qquad (3.7.1.4)$$

where $\tilde{\underline{\beta}} = t^{-1} \sum_{i=1}^{t} \tilde{\underline{\beta}}_i$

The least squares estimate of $\tilde{\underline{\beta}}_i$ is the solution of

$$X_i \underline{y}_i = X_i' X_i \hat{\underline{\beta}}_i$$

and so we can rewrite equation (3.7.1.4) as

$$\tilde{\underline{\beta}}_i = (\sigma_i^{-2} X_i' X_i + C^{-1})^{-1} (\sigma_i^{-2} X_i' X_i \hat{\underline{\beta}}_i + C^{-1} \tilde{\underline{\beta}}) \qquad (3.7.1.5)$$

$\tilde{\underline{\beta}}_i$ is a weighted average of the least squares estimate for the ith row and the (weighted) average of the estimates from the whole triangle. It is thus a credibility estimator, and it is possible to go one step further by summing both sides of equation (3.7.1.5) with respect to i and solving for $\tilde{\underline{\beta}}$.

Setting $W_i = (\sigma_i^{-2} X_i' X_i + C^{-1})^{-1} \sigma_i^{-2} X_i' X_i$, equation (3.7.1.5) can be written as

$$\tilde{\underline{\beta}}_i = W_i \hat{\underline{\beta}}_i + (1 - W_i) \tilde{\underline{\beta}}$$

Summing with resect to i :

$$t\tilde{\underline{\beta}} = \sum_{i=1}^{t} W_i \hat{\underline{\beta}}_i + \tilde{\underline{\beta}} \sum_{i=1}^{t} (1 - W_i)$$

$$\text{Hence } \tilde{\underline{\beta}} \sum_{i=1}^{t} W_i = \sum_{i=1}^{t} W_i \hat{\underline{\beta}}_i$$

$$\text{and } \tilde{\underline{\beta}} = \frac{\sum_{i=1}^{t} W_i \hat{\underline{\beta}}_i}{\sum_{i=1}^{t} W_i}$$

Therefore

$$\tilde{\underline{\beta}}_i = (\sigma_i^{-2} X_i' X_i + C^{-1})^{-1} \left(\sigma_i^{-2} X_i' X_i \hat{\underline{\beta}}_i + C^{-1} (\sum_{i=1}^{t} W_i)^{-1} \sum_{i=1}^{t} W_i \hat{\underline{\beta}}_i \right) \qquad (3.7.1.6)$$

3.7.2 Remarks

Equation (3.7.1.6) can be rewritten as:

$$\tilde{\underline{\beta}}_i \;=\; W_i\,\hat{\underline{\beta}}_i \;+\; (\,I_p - W_i\,)\;\frac{\sum_{i=1}^{t} W_i\,\hat{\underline{\beta}}_i}{\sum_{i=1}^{t} W_i} \qquad (3.7.2.1)$$

$\tilde{\underline{\beta}}_i$ is clearly a weighted average of the least squares estimate for each row and the (weighted) average of the least squares estimates for all the rows in the triangle. It can therefore be regarded as a credibility estimator, with credibility factor W_i .

Note that W_i increases as $\sigma_i^2\,(\,X_i' X_i\,)^{-1}$ decreases: as the variance of the observations in the ith row decreases. W_i is inversely proportional to the within row variance: thus more weight is given to the row estimate if the data in the row has a small variance.

The estimates of the variances , C and $\sigma_1^2, \sigma_2^2, \ldots, \sigma_t^2$ are $\hat{C}$ and $s_1^2, s_2^2, \ldots, s_t^2$ where

$$\hat{C} = [\,R + \sum_{i=1}^{t} (\tilde{\underline{\beta}}_i - \tilde{\underline{\beta}}\;)(\tilde{\underline{\beta}}_i - \tilde{\underline{\beta}}\;)'\,]/(\,t-p-1) \qquad (3.7.2.2)$$

$$s_i^2 = (\underline{y} - X_i\,\tilde{\underline{\beta}}_i\,)'(\underline{y} - X_i\,\tilde{\underline{\beta}}_i\,)/(t-i+3) \qquad (3.7.2.3)$$

where R is a diagonal matrix with small positive entries. Starting from $\hat{C}^{-1} = 0$, the Bayes estimates are obtained by iterating between equation (3.7.1.6) and equations (3.7.2.2),(3.7.2.3) until convergence.

The preceding theory is the more general case in which the row variances are assumed to be heterogeneous. It is, however, more common to assume a homogeneous variance structure with $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma^2$. In this case the estimate of σ^2 is s^2 , where

$$s^2 = (\underline{y} - X\,\tilde{\underline{\beta}}\,)'(\underline{y} - X\,\tilde{\underline{\beta}}\,)/(n+2) \qquad (3.7.2.4)$$

and the estimate of σ_i^2 in equation (3.7.1.6) is s^2, $\forall$ i. The procedure is to iterate between equations (3.7.1.6) and (3.7.2.2) , (3.7.2.4) and is illustrated in the following example.

3.7.3 Example

Since the empirical Bayes estimates use the least-squares estimates, it is not possible to use the last two rows, in which there are fewer data points than parameters. The following table gives the empirical Bayes and least squares estimates of the parameters for the first eight rows of the Taylor and Ashe data.

Figure 3.26

Empirical Bayes Estimates			Least Squares Estimates		
6.374	1.748	-0.617	6.440	1.522	-0.559
6.313	1.711	-0.534	6.388	1.318	-0.419
6.386	1.833	-0.632	6.202	2.160	-0.709
6.544	1.949	-0.742	6.288	2.978	-1.072
6.512	1.859	-0.649	6.583	1.749	-0.625
6.554	1.896	-0.617	6.612	1.972	-0.671
6.590	1.924	-0.558	6.691	1.709	-0.487
6.623	1.991	-0.465	6.572	2.697	-0.786

The fitted values are shown in figure 3.27.

Figure 3.27

357583 648069 710330 633695 505015 374745 264716 180373 119566 77555
357848 766940 610542 482940 527326 574398 146342 139950 227229 67948

397922 763655 895855 859134 737785 590834 450890 332164 238196
352118 884021 933894 1183289 445745 320996 527804 266172 425046

413701 783942 876664 789962 632403 469764 331388 225100
290507 1001799 926219 1016654 750816 146923 495992 280405

431676 793688 832937 694815 511085 347185 223244
310608 1108250 776189 1562400 272482 352053 206286

403738 765408 849991 758297 600007 440062
443160 693190 991983 769488 504851 470639

387523 778569 906580 844361 695834
396132 937085 847498 805037 705960

395102 858215 1071633 1066738
440832 847631 1131398 1063269

378315 944349 1329376
359480 1061648 1443370

Figure 3.28 shows the predicted values.

Figure 3.28

					172819
				154028	101250
			143529	87971	53032
		320303	220707	149211	99661
	558348	418330	305623	219986	157126
994745	846596	695730	561029	448637	358381
1588480	1656196	1636008	1573415	1497895	1426995	137068

Figure 3.29

Empirical Bayes Estimates	Bayes No Prior	Empirical Bayes Standard Error
172819	110927	45100
255278	482157	76600
284532	660810	97279
789882	1090752	318687
1659413	1530532	812772
3905118	2310959	2529433
10749675	3806976	11295723

The empirical Bayes estimate of total outstanding claims is 17816720 and the estimate of the standard error of total outstanding claims is 11609088.

There is very little information in the last row for which estimation is possible, and so the standard error of outstanding claims is quite large. This model has to be used with care, and it may be necessary to group the later rows together. The empirical Bayes method has improved the properties of the estimators for the later rows, but they are still problematic.

CHAPTER 4

Recursive Bayesian and Dynamic Estimation

4.0 Introduction

In this chapter, as in earlier chapters, it will be assumed (without loss of generality) that the data is in the form of a triangle. After t years, the data available is:

$$
\begin{array}{ccccccc}
y_{1,1} & y_{1,2} & y_{1,3} & \cdots & \cdots & y_{1,t} \\
y_{2,1} & y_{2,2} & \cdots & \cdots & y_{2,t-1} & \\
\vdots & \vdots & & & & \\
\vdots & & & & & \\
y_{t,1} & & & & &
\end{array}
$$

Figure 4.1

This chapter is concerned with the use of recursive and dynamic forecasting methods and examines the use of methods similar to those advocated by Harrison and Stevens(1976). Attention will be restricted to linear models, in particular those described in chapter 2, and the results will be compared with the empirical Bayes results in chapter 3.

The recursive forecasting procedures to be described update the parameter estimates as each year's data are received. In the first year, the only information available is the claims on the first year's business on which the delay index is 1, ie $y_{1,1}$. In the second year, $y_{1,2}$ and $y_{2,1}$ are received, being the claims on the first two years' business with delay indices 1 and 2. In the third year $y_{1,3}$, $y_{2,2}$ and $y_{3,1}$ are received, and so on.

The data which makes up the claims runoff triangle, figure 4.1, is received in the form :

$$y_{1,1}\ ,\ \begin{bmatrix} y_{1,2} \\ y_{2,1} \end{bmatrix}\ ,\ \begin{bmatrix} y_{1,3} \\ y_{2,2} \\ y_{3,1} \end{bmatrix}\ ,\ \cdots\ , \tag{4.0.1}$$

and in year t the data which is received is

$$\begin{bmatrix} y_{1,t} \\ y_{2,t-1} \\ \vdots \\ y_{t,1} \end{bmatrix} \tag{4.0.2}$$

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

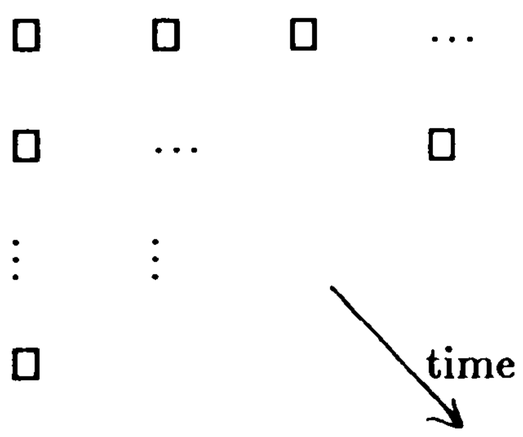


Figure 4.2

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

The data vector at time t is $\underline{y}_t$, where

$$\underline{y}_t = \begin{bmatrix} y_{1,t} \\ y_{2,t-1} \\ \vdots \\ y_{t,1} \end{bmatrix} \tag{4.0.3}.$$

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

$$\underline{Y}_1 , \underline{Y}_2 , \cdots , \underline{Y}_t , \cdots \qquad (4.0.4).$$

Note that the triangle has been broken down in a different way from the row method in chapters 2 and 3.

In this time series, the data vector expands with t: in the case of a tringular set of data,

$$\dim (\underline{y}_t) = t \qquad (4.0.5).$$

[Note: the definition of $\underline{y}_t$ is different from that implied by equation (2.6.1)].

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

The problem can be seen 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. This development is left mainly until chapter 5, although similar techniques are used in this chapter. In the present chapter, it will be shown how to analyse the claims data recursively using the linear models discussed in previous chapters. These static models will be generalised to one-dimensional dynamic models in this chapter and further generalised to two-dimensional dynamic models in chapter 5.

For the chain ladder linear model, a model which leads to estimates having similarities to those arising from the empirical Bayes estimation methods of chapter 3 will be described. This model will assume a similar structure on the row parameters, although the relationship between them will be sequential.

For row models, the methods discussed will include recursive least-squares and static and dynamic Bayesian estimation. Again the dynamic Bayesian estimation method will be generalised in chapter 5.

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 in chapters 2

and 3. The more proper method calculates one-step-ahead , two-steps-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.

4.1 Recursive Linear Models : State Space Representations

A linear model for the complete data set takes the form

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

For notational reasons this is rewritten as

$$\underline{y} = F\underline{\theta} + \underline{e} \qquad (4.1.1)$$

where $\underline{\theta}$ is the parameter vector

and F is the design matrix.

Equation (4.1.1) is the linear model for the whole triangle. The linear models for $\underline{y}_1$, $\underline{y}_2$, ... , $\underline{y}_t$, ... , which together make up the triangle can be written as

$$\begin{aligned}
\underline{y}_1 &= F_1\underline{\theta}_1 + \underline{e}_1 \\
\underline{y}_2 &= F_2\underline{\theta}_2 + \underline{e}_2 \\
&\vdots \\
\underline{y}_t &= F_t\underline{\theta}_t + \underline{e}_t \qquad (4.1.2)
\end{aligned}$$

$\underline{y}_t$ is defined in equation (4.0.3) and the precise forms of the parameter vector at time t, $\underline{\theta}_t$, and the design matrix at time t, F_t , depend on the model being considered. Later sections of this chapter give the forms of $\underline{\theta}_t$ and F_t for the chain ladder linear model and a row model. Equation (4.1.2) 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 linear model. There are several ways of using this representation, which are described in the later sections of this chapter.

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:

$$\underline{\theta}_{t+1} = G_t \underline{\theta}_t + H_t \underline{u}_t + \underline{w}_t \qquad (4.1.3)$$

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

and $\underline{w}_t$ 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, and some useful cases are now described.

The simplest case is to set $\underline{u}_t$ and $\underline{w}_t$ to $\underline{0}$ for all t. In this case equation (4.1.3) becomes:

$$\underline{\theta}_{t+1} = G_t \underline{\theta}_t \qquad (4.1.4)$$

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 is vague, equation (4.1.4) defines recursive least squares estimation.

The case when the new parameters entering at time t+1 are distinct from those at time t can be achieved by setting

$$\underline{\theta}_{t+1} = G_t \underline{\theta}_t + H_t \underline{u}_t \qquad (4.1.5)$$

where $\underline{u}_t$ 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.

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

$$\underline{\theta}_{t+1} = G_t \underline{\theta}_t + \underline{w}_t \qquad (4.1.6)$$

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

The specific forms of F_t and G_t for the analysis of variance linear model and row models are given in sections 4.3 and 4.5.

4.2 The Kalman Filter

The updating of the estimates of the state vector (which contains the parameters) as each new data vector is received is done by the Kalman filter. It was first investigated by Kalman (1963). This section represents a special case in which the dimensions of the state vector and observation vector increase with t. In the actuarial literature, the use of the Kalman filter to analyse certain models for claims runoff triangles was first suggested by De Jong and Zehnwirth (1983).

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

chapter:

$$\underline{y}_t = F_t \underline{\theta}_t + \underline{e}_t \qquad (4.2.1)$$

$$\underline{\theta}_{t+1} = G_t \underline{\theta}_t + H_t \underline{u}_t + \underline{w}_t \qquad (4.2.2)$$

where $\underline{e}_t \sim N(\underline{0}, V_t)$,

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

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

and are independent.

Further $\underline{e}_t$,$\underline{u}_t$, $\underline{w}_t$ are sequentially independent.

Suppose $\underline{\theta}_t \mid (\underline{y}_1, \underline{y}_2, \cdots, \underline{y}_{t-1}) \sim N(\hat{\underline{\theta}}_{t|t-1}, C_t)$. (4.2.3)

From equations (4.2.3) and (4.2.1), the distribution of $\underline{y}_t$ given information up to time t-1 is

$$\hat{\underline{y}}_{t|t-1} \sim N(F_t \hat{\underline{\theta}}_{t|t-1}, F_t C_t F_t' + V_t) \qquad (4.2.4)$$

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 (4.2.2).

This iterative procedure, known as the Kalman filter, is proved in the Appendix. The recursion is given by the following theorem:

4.2.1 Theorem

If the system and observation equations are given by equations (4.2.1) and (4.2.2), and the distribution of $\underline{\theta}_t$ given information at time t-1 is given by equation (4.2.3), then the distribution can be updated when $\underline{y}_t$ is received using the following recursion:

$$\hat{\underline{\theta}}_{t+1|t} = G_t \hat{\underline{\theta}}_{t|t-1} + H_t \hat{\underline{u}}_t + K_t (\underline{y}_t - \hat{\underline{y}}_t) \qquad (4.2.1.1)$$

where $K_t = G_t C_t F_t' (F_t C_t F_t' + V_t)^{-1}$ (4.2.1.2)

$$\text{and } C_{t+1} = G_t C_t G_t' + H_t U_t H_t' - G_t C_t F_t' (F_t C_t F_t' + V_t)^{-1} F_t C_t G_t' + W_t \qquad (4.2.1.3)$$

Proof

The proof is given in appendix 1.

4.2.2 Remarks

$\hat{\underline{\theta}}_{t+1|t}$ is the forecast of $\underline{\theta}_{t+1}$, given all information up to time t, and C_{t+1} is the variance-covariance matrix of this estimate. The one-step-ahead forecast of $\underline{Y}_{t+1}$, assuming a squared-error loss function, is:

$$\hat{\underline{Y}}_{t+1|t} = F_{t+1}\hat{\underline{\theta}}_{t+1|t} + H_{t+1}\,\hat{\underline{u}}_{t+1} \qquad (4.2.2.1)$$

and its variance-covariance matrix is:

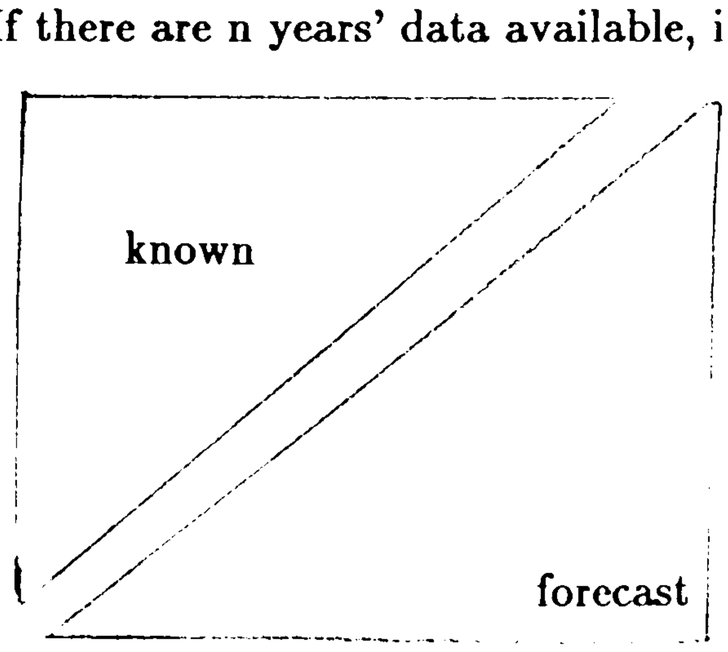
$$\text{Var} (\hat{\underline{Y}}_{t+1|t}) = F_{t+1}C_{t+1}F_{t+1}' + H_{t+1}U_{t+1}H_{t+1}' + V_t \qquad (4.2.2.2)$$

The n-step-ahead forecast of $\underline{\theta}_{t+n}$ can be obtained by repeated application of equations (4.2.2.1):

$$\hat{\underline{\theta}}_{t+n|t} = G_{t+n-1}\,\hat{\underline{\theta}}_{t+n-1|t} + H_{t+n-1}\,\hat{\underline{u}}_{t+n-1} \qquad (4.2.2.3)$$

4.2.3 Forecast Values and Variances

If there are n years' data available, it is usual to forecast the values in the lower triangle:



This is the aim of the chain ladder method and of the techniques in chapter 3. The Kalman filter can produce forecasts for the lower triangle by using the one-step-ahead, two-step-ahead, ... , (t-1)-steps-ahead forecast values and their variances given by equations (4.2.2.1) to (4.2.2.3). It can go further and produce estimates beyond the lower triangle if the prior distributions of the input vectors are specified.

However, the variances of the row totals (the forecasts of total outstanding claims for each year of business) present great problems since it is necessary to calculate

$$\mathrm{Cov}\ (\ \hat{Y}_{ij}\ ,\ \hat{Y}_{ik}\ )$$

for all relevant values of j and k. This entails storing the variances and covariances of all the state vectors for the lower triangle, and not just the old and up-dated values. It thus destroys the recursive nature of the Kalman filter, and for this reason the predicted values and their variances are calculated in the same way as in chapter 3: the i-steps-ahead forecasts are not used. The parameters are estimated using the Kalman filter up to the most recent year and future values are forecast using the structure of the model. For most of the models discussed the results would be identical, the only difference occuring when there are disturbance terms in the state equation relating to the parameters used in the estimation.

4.3 Applications of the Kalman Filter to the Chain Ladder Linear Model

The chain ladder linear model for the whole triangle is (for 3 years' data):

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{13} \\ y_{22} \\ y_{31} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \alpha_3 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{13} \\ e_{22} \\ e_{31} \end{bmatrix} \qquad (4.3.1)$$

The form of the design matrix F is clear. The state vector is:

$$\underline{\theta} = \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \alpha_3 \\ \beta_3 \end{bmatrix} \tag{4.3.2}$$

The first few observation equations, when the data is handled recursively, are:

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

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

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

In general, the state vector at time t is:

$$\underline{\theta}_t = \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \vdots \\ \alpha_t \\ \beta_t \end{bmatrix} \tag{4.3.3}$$

A recursive version of the chain ladder method is achieved by defining the system equation matrices as

$$\underline{\theta}_{t+1} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 \end{bmatrix} \underline{\theta}_t + \begin{bmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \underline{u}_t \qquad (4.3.4)$$

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 (4.3.4) says that the existing parameters are unchanged, while the new parameters are treated as stochastic inputs.

If the variance of the errors, e_{ij} , are 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 and can be seen as an alternative to the method of section 3.4. 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. A similar assumption, which is not unreasonable and is often made in Bayes estimation methods, was made in example 3.4.1. The method is illustrated in example 4.3.1.

There are two other variations on the chain ladder linear model which are of use practically. Since the chain ladder linear model is under consideration, the observation equation remains unaltered and only the system equation needs to be changed. The variations can be compared with the empirical Bayes estimation methods in chapter 3. They are both estimation methods which impose structure on the parameters. The empirical Bayes estimation method is based on the assumption that the parameters are independently, identically distributed: the dynamic estimation method assumes that the parameters are related in a similar way, but imposes a sequential structure as well.

4.3.1 Example

Suppose that there is prior information which suggests that the prior distribution of the row parameters has mean 0.3 and variance 0.05, but that there is no prior information about the other parameters.

The following table shows the parameter estimates and their standard errors:

Figure 4.4

	Bayes Estimate	Least Squares Estimate	Standard Error Of Bayes Estimate
Overall Mean	6.177	6.106	0.123
Row Parameters	0.200	0.194	0.118
	0.166	0.149	0.121
	0.170	0.153	0.126
	0.275	0.299	0.131
	0.349	0.412	0.137
	0.402	0.508	0.145
	0.479	0.673	0.156
	0.362	0.495	0.170
	0.369	0.602	0.191
Column Parameters	0.893	0.911	0.158
	0.911	0.939	0.164
	0.915	0.965	0.171
	0.320	0.383	0.180
	−0.080	−0.005	0.191
	−0.199	−0.118	0.207
	−0.518	−0.439	0.231
	−0.128	−0.054	0.271
	−1.464	−1.393	0.362

It can be seen, by comparison with figure 2.2, that the estimates of the row parameters have generally be drawn towards the prior mean. For example, the estimate of α_9 is now 0.362, compared with the estimate of 0.495 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 of predicted outstanding claims and their standard errors are given in the following table:

Figure 4.5

Bayesian Estimates		Standard Error
With Prior	No Prior	With Prior
131413	110927	69379
596992	482157	228013
812958	660810	246552
1157631	1090752	302566
1450615	1530532	346525
1992212	2310959	457084
2796170	3806976	655781
3911910	4452396	926180
3920230	5066116	968366

The predicted overall total outstanding claims in this case is 16770131 and the standard error of this estimate is 1953764.

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.

4.3.2 Dynamic Row Parameters

Firstly the dynamic estimation method will be applied to the row parameters, leaving the column parameters as they were (although the estimates will change because of the change in the estimation of the row parameters). The system equation is

$$\underline{\theta}_{t+1} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ 0 & \dots & 1 & 0 \\ 0 & \dots & \dots & 0 \end{bmatrix} \underline{\theta}_t + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ 1 \end{bmatrix} \underline{u}_t + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{bmatrix} \underline{w}_t \qquad (4.3.2.1)$$

where $\underline{u}_t$ has the prior distribution of β_{t+1}

and $\underline{w}_t$ is a disturbance term.

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

$$\alpha_{t+1} = \alpha_t + w_t \qquad (4.3.2.2)$$

and a sophisticated smoothing method is produced. Equation (4.3.2.2) can be written as

$$\alpha_{t+1} | \alpha_t \sim N(0, W_t) \qquad (4.3.2.3)$$

and can be compared with the empirical Bayes assumption that

$$\alpha_i \sim N(\theta, \sigma_\alpha^2) \quad \forall i \qquad (4.3.2.4)$$

(a vague prior distribution is used for θ).

Example 4.3.3 compares the empirical Bayes estimates with the dynamic estimates using the state variance

$$W_t = \sigma_\alpha^2$$

In fact, σ_α^2 is estimated from the data in the empirical Bayes theory, and it is this estimate which has been used as the system variance. It is not to be inferred that the state variance is the same as the variance of the parameters in the empirical Bayes theory.

4.3.3 Example

The row parameters are related recursively and the column parameters are left as they were when 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, which was the estimate of the variance of the row parameters in the empirical Bayes method of section 3.6.3. In this case the parameter estimates are as follows:

Figure 4.6

	Bayes Estimate	Least Squares Estimate	Standard Error Of Bayes Estimate
Overall Mean	6.119	6.106	0.163
Row Parameters	0.187	0.194	0.151
	0.170	0.149	0.148
	0.196	0.153	0.152
	0.296	0.299	0.158
	0.396	0.412	0.164
	0.482	0.508	0.171
	0.550	0.673	0.183
	0.536	0.495	0.202
	0.546	0.602	0.238
Column Parameters	0.906	0.911	0.158
	0.940	0.939	0.165
	0.951	0.965	0.173
	0.346	0.383	0.183
	−0.028	−0.005	0.195
	−0.145	−0.118	0.212
	−0.457	−0.439	0.236
	−0.062	−0.054	0.278
	−1.406	−1.393	0.378

The fitted values (with the actual values) are shown in figure 4.7 and the predicted outstanding claims in figure 4.8.

Figure 4.7

280871 695378 719436 727701 405244 274091 244522 179767 268833 72013
357848 766940 610542 482940 527326 574398 146342 139950 227229 67948

399348 988704 1022913 1034665 576191 389720 347699 255691 382930
352118 884021 933894 1183289 445745 320996 527804 266172 425046

379495 939553 972064 983237 547558 370368 330481 243178
290507 1001799 926219 1016654 750816 146923 495992 280405

346861 858764 888488 898718 500515 338601 302306
310608 1108250 776189 1562400 272482 352053 206286

370416 917100 948874 959863 534661 361894
443160 693190 991983 769488 504851 470639

376568 932394 964797 976176 544043
396132 937085 847498 805037 705960

403956 1000400 1035470 1048310
440832 847631 1131398 1063269

400937 993409 1029002
359480 1061648 1443370

413320 1025208
376686 986608

336409
344014

Figure 4.8

						109955
					387238	104549
				236325	354494	95622
			343098	252735	378935	102189
		390931	349303	257203	385581	103974
	620155	420042	375175	276218	414069	111653
	1105771	616673	417538	372891	274524	411525	110966
1127356	1142248	636820	431128	385011	283443	424891	114570
886202	919094	931049	519029	351372	313783	231004	346283	93373

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

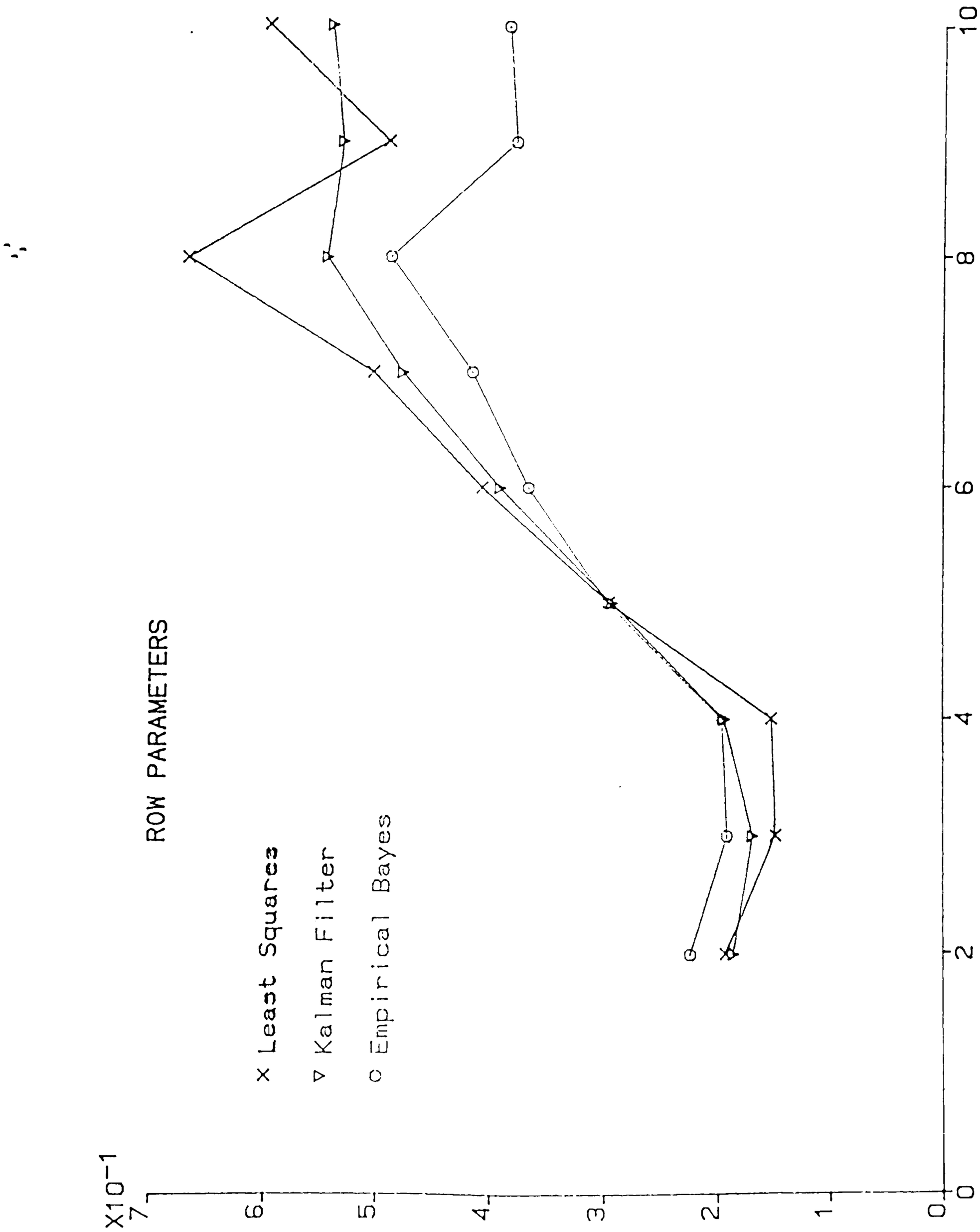
Figure 4.9

Bayesian Estimates		Standard Error
With Prior	No Prior	With Prior
109955	110927	59278
491787	482157	187134
686441	660810	206954
1076957	1090752	277762
1486991	1530532	347441
2217311	2310959	491998
3309887	3806976	744931
4545466	4452396	1048855
4591188	5066116	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 the following graph, which also

shows the empirical Bayes estimates from example 3.6.3. It can be seen that the empirical Bayes estimates are all pulled towards a central value, while the Kalman filter has smoothed the variation in the estimates. The difference between the two methods is clear: the empirical Bayes method assumes that the row parameters have a common mean, while the Kalman filter imposes a recursive relationship.

Figure 4.10



4.3.4 Dynamic Row and Column Parameters

The dynamic estimation method can be extended to the column parameters to give a method which has similarities with the empirical Bayes method of section (3.6.4). The system equation becomes

$$\underline{\theta}_{t+1} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ 0 & \dots & 1 & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix} \underline{\theta}_t + \begin{bmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \underline{w}_t \qquad (4.3.4.1)$$

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

Thus both the new parameters, α_{t+1} and β_{t+1} , are related to the previous parameters and dynamic estimation is used for each. The empirical Bayes theory assumes that α_i and β_j are independent and

$$\alpha_i \sim N(\theta, \sigma_\alpha^2)$$

$$\beta_j \sim N(\xi, \sigma_\beta^2)$$

where θ,ξ have vague prior distributions.

Example 4.3.4 compares the empirical Bayes with the dynamic Bayes estimates using the following system variance-covariance matrix:

$$W_t = \begin{bmatrix} \sigma_\alpha^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix}$$

Again, σ_α^2 and σ_β^2 are replaced by the estimates from the empirical Bayes analysis.

4.3.5 Example

This model was applied with state variance for the row parameters 0.0296 and for the column parameters 0.4135. Again, these were the figures obtained from the empirical Bayes method this time as obtained in section 3.6.4. The parameters estimates are as follows:

Figure 4.11

	Bayes Estimate	Least Squares Estimate	Standard Error Of Bayes Estimate
Overall Mean	6.102	6.106	0.163
Row Parameters	0.211	0.194	0.150
	0.186	0.149	0.148
	0.212	0.153	0.152
	0.313	0.299	0.158
	0.414	0.412	0.164
	0.502	0.508	0.171
	0.569	0.673	0.183
	0.553	0.495	0.202
	0.564	0.602	0.239
Column Parameters	0.908	0.911	0.157
	0.939	0.939	0.162
	0.929	0.965	0.170
	0.373	0.383	0.179
	−0.012	−0.005	0.189
	−0.151	−0.118	0.204
	−0.411	−0.439	0.224
	−0.215	−0.054	0.256
	−1.132	−1.393	0.342

Figure 4.12 shows the fitted values, and figure 4.13 the predicted outstanding claims.

Figure 4.12

276241 684518 706407 700184 401947 273582 238639 184557 225716 92117
357848 766940 610542 482940 527326 574398 146342 139950 227229 67948

402630 997709 1029613 1020545 585856 398766 347861 269134 329849
352118 884021 933894 1183289 445745 320996 527804 266172 425046

379071 939332 969371 960838 551588 375458 327585 253633
290507 1001799 926219 1016654 750816 146923 495992 280405

346663 859030 886510 878724 504475 343449 299848
310608 1108250 776189 1562400 272482 352053 206286

370496 918110 947509 939252 539324 367387
443160 693190 991983 769488 504851 470639

377404 935291 965339 957136 549908
396132 937085 847498 805037 705960

405386 1004836 1037423 1029245
440832 847631 1131398 1063269

401924 996766 1029863
359480 1061648 1443370

413762 1027306
376686 986608

336813
344014

Figure 4.13

						143834
					330264	135583
				246548	302525	124102
			340516	263890	323649	132738
		397610	347381	269099	329988	135329
	627610	427793	373607	289376	354836	145517
1084427	623432	424793	370939	287295	352280	144467
1126688	1118681	642922	438019	382470	296222	363223	148955
888158	918705	911987	524086	357044	311761	241456	296070	121416

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

Figure 4.14

Bayesian Estimates		Standard Error
With Prior	No Prior	With Prior
143834	110927	72675
465847	482157	166438
673175	660810	194229
1060794	1090752	266228
1479407	1530532	339755
2218738	2310959	487975
3287633	3806976	735669
4517179	4452396	1040596
4570683	5066116	1167068

The predicted overall total outstanding claims is 18417296 and the standard error of this estimate is 2627190.

4.4 Applications of the Kalman Filter to Row Models

The row models described in chapter 2 have explanatory variables which are dependent only on the delay index and not on the year of business. Hence

$$y_{ij} = \underline{X}_j \underline{\beta}_i + e_{ij} \qquad (4.4.1)$$

where $\underline{X}_j$ is a row vector whose dimension is the same as the dimension of $\underline{\beta}_i$. The vector $\underline{X}_j$ is not the same as the matrix X_j in equation (2.6.1).

Considering the data which is received at time t, the following linear model is applied:

$$\underline{y}_t = \begin{bmatrix} y_{1,t} \\ y_{2,t-1} \\ \vdots \\ y_{t,1} \end{bmatrix} = \begin{bmatrix} \underline{X}_t\underline{\beta}_1 \\ \underline{X}_{t-1}\underline{\beta}_2 \\ \vdots \\ \underline{X}_1\underline{\beta}_t \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t-1} \\ \vdots \\ e_{t,1} \end{bmatrix} \qquad (4.4.2).$$

$$\underline{y}_t = \begin{bmatrix} \underline{X}_t & & \\ & \underline{X}_{t-1} & \\ & & \ddots \\ & & \underline{X}_1 \end{bmatrix} \begin{bmatrix} \underline{\beta}_1 \\ \underline{\beta}_2 \\ \vdots \\ \underline{\beta}_t \end{bmatrix} + \underline{e}_t \qquad (4.4.3)$$

where $\underline{e}_t = \begin{bmatrix} e_{1,t} \\ e_{2,t-1} \\ \vdots \\ e_{1,t} \end{bmatrix}$

$$\text{ie} \qquad \underline{y}_t = F_t \underline{\theta}_t + \underline{e}_t \qquad (4.4.4)$$

where $F_t = \begin{bmatrix} \underline{X}_t & & \\ & \underline{X}_{t-1} & \\ & & \ddots \\ & & \underline{X}_1 \end{bmatrix}$

$$\text{and} \quad \underline{\theta}_t = \begin{bmatrix} \underline{\beta}_1 \\ \underline{\beta}_2 \\ \vdots \\ \underline{\beta}_t \end{bmatrix}$$

Note that dim ($\underline{\beta}_i$) is constant for all i.

The system equation is

$$\underline{\theta}_{t+1} = G_t \, \underline{\theta}_t + H_t \, \underline{u}_t + \underline{w}_t \qquad (4.4.5)$$

$$\text{If} \quad G_t = \begin{bmatrix} I & & & \\ & I & & \\ & & \ddots & \\ & & & I \\ & & & I \end{bmatrix}$$

where I is an identity matrix whose dimension is the same as the dimension of the parameter vector $\underline{\beta}_i$, and the system equation (4.4.5) is simplified to

$$\underline{\theta}_{t+1} = G_t \, \underline{\theta}_t \qquad (4.4.6)$$

then the estimation is recursive static Bayesian estimation with identical row parameters. If the prior distribution of $\underline{\beta}_1$ is vague then the Kalman filter gives the same results as recursive least-squares estimation with known variance.

4.4.1 Dynamic Estimation for Row Models

The static model in section 4.4 can be generalised to a dynamic model by altering (4.4.6) to

$$\underline{\theta}_{t+1} = G_t \underline{\theta}_t + \begin{bmatrix} \underline{0} \\ \underline{0} \\ \vdots \\ \underline{0} \\ \underline{w}_t \end{bmatrix} \qquad (4.4.1.1)$$

This is the situation which was considered in De Jong and Zehnwirth (1982).

The Kalman filter then gives dynamic estimation of the parameters, with

$$\underline{\beta}_{t+1} = \underline{\beta}_t + \underline{w}_t \qquad (4.4.1.2)$$

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

Again, this can be compared with the empirical Bayes estimation method in chapter 3. In that chapter it was assumed that

$$\underline{\beta}_i \sim N(\underline{\xi}, \Sigma).$$

In example (4.4.3), the dynamic estimation method is compared to the empirical Bayes estimation method, using $W_t = \hat{\Sigma}$.

4.4.2 Bayesian Estimation for Row Models with Distinct Parameters

The observation equation for all row models is given by equation (4.4.4) with F_t and $\underline{\theta}_t$ suitably defined. Section 4.4 illustrated the recursive estimation of the row parameters when they are assumed to be identical and this was extended in section 4.4.1 to dynamic estimation. This section goes a step further and assumes that the parameters in each row are distinct. The system equation under these assumptions is

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

where $G_t = \begin{bmatrix} I & & & \\ & I & & \\ & & \ddots & \\ & & & I \\ O & \cdots & & O \end{bmatrix}$ and $H_t = \begin{bmatrix} O \\ O \\ \vdots \\ I \end{bmatrix}$

O is a matrix of zeros.

$\underline{u}_t$ has the distribution of the new parameter vector and this distribution governs the estimation method being used. If it is a vague distribution then classical least-squares estimation is used, otherwise Bayesian estimation is used.

4.4.3 Example

Again, the model of section 2.6 is considered in this example. As was mentioned earlier, the row models are unsatisfactory in many cases due to the paucity of data in the last few rows. Because of this the methods are only illustrated here for the dynamic model of section 4.4.1. The state variances are taken from the empirical Bayes results of chapter 3.

Figure 4.15

Kalman Filter Estimates			Least Squares Estimates		
6.328	1.704	-0.596	6.440	1.522	-0.559
6.291	1.747	-0.543	6.388	1.318	-0.419
6.357	1.868	-0.638	6.202	2.160	-0.709
6.493	1.966	-0.737	6.288	2.978	-1.072
6.535	1.975	-0.704	6.583	1.749	-0.625
6.584	2.001	-0.673	6.612	1.972	-0.671
6.621	2.038	-0.617	6.691	1.709	-0.487
6.648	2.099	-0.527	6.572	2.697	-0.786

The fitted values are shown in figure 4.16.

Figure 4.16

341493 613111 674155 606458 488733 367390 263227 182084 122619 80844
357848 766940 610542 482940 527326 574398 146342 139950 227229 67948

389180 759282 896117 860903 738877 590493 449244 329688 235383
352118 884021 933894 1183289 445745 320996 527804 266172 425046

401667 774469 872540 788908 632290 469562 330840 224294
290507 1001799 926219 1016654 750816 146923 495992 280405

410191 767074 814802 686592 509611 349079 226232
310608 1108250 776189 1562400 272482 352053 206286

413300 803135 884281 771578 592680 420002
443160 693190 991983 769488 504851 470639

399397 815486 936281 849244 676969
396132 937085 847498 805037 705960

407683 902965 1112778 1078707
440832 847631 1131398 1063269

387744 980533 1355722
359480 1061648 1443370

Figure 4.17 shows the predicted values.

Figure 4.17

								178466
							160696	105019
						153884	94747	57313
					309490	204627	132475	84602
				550856	397070	279008	193270	133015
			1025391	845936	675990	532756	419074	331781
		1658531	1691678	1651640	1592551	1546076	1530746	1559629
	1582684	1914732	2167576	2431715	2791360	3344256	4237589	5732263
942365	1521425	2133926	2936156	4197804	6440130	10816831	20158917	42083218

Figure 4.18

Kalman Filter Estimates	Bayes No Prior	Kalman Filter Standard Error
178466	110927	75253
265714	482157	99995
305944	660810	119366
731194	1090752	317911
1553218	1530532	809896
3830928	2310959	2741090
11230852	3806976	15686239

The predicted outstanding claims are shown in the above table, and can be compared with the empirical Bayes estimates in table 3.29. It can be seen that the later rows are still problematic and if this model is to be used in practise, they would have to be grouped together.

CHAPTER 5

Two-Dimensional Processes

5.0 Introduction

Chapter 4 described the use of dynamic forecasting methods in which, for example, the parameters of the (i+1)th row are related to the parameters of the ith row. In this case, the direction of the dynamic variation is vertical, from row to row, and the parameters are assumed to be constant along the rows. This is illustrated in the following figure.

$\underline{\beta}_1 \quad \underline{\beta}_1 \quad \underline{\beta}_1 \quad \cdots \quad \cdots \quad \underline{\beta}_1$

$\underline{\beta}_2 \quad \underline{\beta}_2 \quad \cdots \quad \cdots \quad \underline{\beta}_2$

$\vdots \quad \vdots$

$\vdots$

$\underline{\beta}_t$

Figure 5.1

where $\underline{\beta}_{j+1} = \underline{\beta}_j + \underline{w}_j$ (5.0.1)

Equation (5.0.1) is the state equation, and it is assumed that there is no variation along the rows. This is not always the case and other dynamic forecasting methods have been used, for example J.Q.Smith (1983) models the claims run off triangle using generalised Bayesian forecasting methods which allow the parameters to vary dynamically along each row. The triangle is considered one row at a time and the parameters are related as shown in figure 5.2.

$\underline{\beta}_1 \quad \underline{\beta}_2 \quad \underline{\beta}_3 \quad \cdots \quad \underline{\beta}_{t-i+1}$

Figure 5.2

where $\underline{\beta}_{j+1} = \underline{\beta}_j + \underline{w}_j$ (5.0.2)

is the system equation.

Other authors have also suggested the use of models which are in the form of figure 5.2 and, when fitting a dynamic model to each row, sometimes attempt to relate the rows to each other. For example, it is possible to set the first parameters of the ith row equal to the last parameters of the (i-1)th row, but this destroys the recursive nature of the process. This method can only be used if there is no more data arriving to be processed, or if the whole estimation procedure is carried out each time a new data vector is received.

For the whole triangle, the form of the model in figure 5.2, with the parameters varying dynamically along each row but not connected from row to row, is illustrated in figure 5.3.

$$
\begin{array}{cccccc}
\underline{\beta}_{1,1} & \underline{\beta}_{1,2} & \underline{\beta}_{1,3} & \cdots & \cdots & \underline{\beta}_{1,t} \\
\underline{\beta}_{2,1} & \underline{\beta}_{2,2} & \cdots & \cdots & \underline{\beta}_{2,t-1} & \\
\vdots & \vdots & & & & \\
\vdots & & & & & \\
\underline{\beta}_{t,1} & & & & &
\end{array}
$$

Figure 5.3

where $\underline{\beta}_{i,j+1} = \underline{\beta}_{i,j} + \underline{w}_{i,j}$ (5.0.3)

is the system equation.

The other direction of variation, as in figure 5.1 can also be represented by figure 5.3. In this case the system equation is

$$\underline{\beta}_{i+1,j} = \underline{\beta}_{i,j} + \underline{w}_{i,j} \qquad (5.0.4)$$

It should also be noted that this is a slight generalisation of the state system represented by figure 5.1 and equation (5.0.1).

This chapter is devoted to the problem of reconciling these two state recursions. In doing this, the theory of dynamic estimation and Bayesian time series is extended to two dimensions. The application of the methodology derived to the analysis of two dimensional classical time series will also be considered, together with the generalisation of other methods of analysis of classical time series.

Two-dimensional state space systems have been used in circuit theory and control theory. In particular, the work of Givone and Roesser in the control theory literature motivates much of this chapter. Two-dimensional state space systems have been used to analyse multi-dimensional linear iterative circuits and in image restoration.

The state space systems and filtering procedures are adapted and extended in this chapter to a stochastic, time series setting.

5.1 Two-Dimensional State Systems

The state equations for a two-dimensional system are

$$\underline{\beta}^{h}_{i+1,j} = A_1 \underline{\beta}^{h}_{i,j} + A_2 \underline{\beta}^{v}_{i,j} + \underline{w}^{h}_{i,j} \qquad (5.1.1)$$

$$\underline{\beta}^{v}_{i,j+1} = A_3 \underline{\beta}^{h}_{i,j} + A_4 \underline{\beta}^{v}_{i,j} + \underline{w}^{v}_{i,j} \qquad (5.1.2)$$

$\underline{\beta}^{h}_{i,j}$ and $\underline{\beta}^{v}_{i,j}$ are defined for all values of i , j , s , t ($i \geq 1$, $j \geq 1$) , and are the horizontal and vertical components of the state vector. The state vector has been split into two parts, each of which is updated in the horizontal or vertical direction. Since there are two components of the state vector, the observation equation needs altering from its one-dimensional form.

The datum at (i,j) is modelled by

$$y_{i,j} = \underline{B}_1 \underline{\beta}^{h}_{i,j} + \underline{B}_2 \underline{\beta}^{v}_{i,j} + e_{i,j} \qquad (5.1.3)$$

When equal weight is given to each component of the state vector, the model will be known as

a balanced model. In this case

$$\underline{B}_1 = \underline{B}_2$$

Equations (5.1.1) to (5.1.3) are the generalisation of state space forms to two-dimensional lattices. In order to apply the Kalman filter and obtain recursive estimates, it is necessary to take advantage of the nature of the inter-dependences between the data. Before doing this, the state space system for the chain ladder linear model is described. The chain ladder linear model is examined exclusively, although the methods apply to any linear model, or indeed to any Bayesian time series model as described in Harrison and Stevens (1976).

5.1.1 A Two-Dimensional State System for the Chain Ladder Linear Model

In order to illustrate the two dimensional system, consider the chain ladder linear model, which is given by equation (5.1.1.1):

$$y_{i,j} = \mu + \alpha_i + \beta_j + e_{i,j} \qquad (5.1.1.1)$$

The restriction $\alpha_1 = \beta_1 = 0$ is generalised to

$$\alpha^h_{1,j} = \alpha^v_{1,j} = 0 \quad \forall\ j$$

$$\beta^h_{i,1} = \beta^v_{i,1} = 0 \quad \forall\ i$$

The horizontal and vertical components of the state are:

$$\begin{bmatrix} \mu^h_{i,j} \\ \alpha^h_{i,j} \\ \beta^h_{i,j} \end{bmatrix} \text{ and } \begin{bmatrix} \mu^v_{i,j} \\ \alpha^v_{i,j} \\ \beta^v_{i,j} \end{bmatrix}$$

Using a balanced model, and thus giving equal weight to the horizontal and vertical

components of the state vector, the observation equations for the first few data points are:

$$y_{1,1} = \tfrac{1}{2}\begin{bmatrix} \mu^h_{1,1} \\ 0 \\ 0 \end{bmatrix} + \tfrac{1}{2}\begin{bmatrix} \mu^v_{1,1} \\ 0 \\ 0 \end{bmatrix} + e_{1,1}$$

$$y_{1,2} = \tfrac{1}{2}\begin{bmatrix} \mu^h_{1,2} \\ 0 \\ \beta^h_{1,2} \end{bmatrix} + \tfrac{1}{2}\begin{bmatrix} \mu^v_{1,2} \\ 0 \\ \beta^v_{1,2} \end{bmatrix} + e_{1,2}$$

$$y_{2,1} = \tfrac{1}{2}\begin{bmatrix} \mu^h_{2,1} \\ \alpha^h_{2,1} \\ 0 \end{bmatrix} + \tfrac{1}{2}\begin{bmatrix} \mu^v_{2,1} \\ \alpha^v_{2,1} \\ 0 \end{bmatrix} + e_{2,1}$$

$$y_{2,2} = \tfrac{1}{2}\begin{bmatrix} \mu^h_{2,2} \\ \alpha^h_{2,2} \\ \beta^h_{2,2} \end{bmatrix} + \tfrac{1}{2}\begin{bmatrix} \mu^v_{2,2} \\ \alpha^v_{2,2} \\ \beta^v_{2,2} \end{bmatrix} + e_{2,2}$$

etc.

The state equations are exactly equations (5.1.1) and (5.1.2) and it is clear that these can only apply away from the boundaries of the system. At the boundaries, stochastic (or deterministic) inputs are needed: these are dealt with in detail in section (5.1.5).

5.1.2 Causality

In a two-dimensional state system, the question of causality has to be addressed. In a one-dimensional system the causality is usually clear: the time dependence is straightforward. In the two-dimensional case, the main possibilities are half-plane and quarter-plane causality.

Consider a lattice whose points are referenced by

$$\{ (i, j) : i \in Z^+ , j \in Z^+ \}.$$

Quarter-plane causality corresponds to the following ordering of the lattice:

$$(i,j) \leq (k,l) \Leftrightarrow i \leq k \text{ and } k \leq l$$

$$(i,j) = (k,l) \Leftrightarrow i = k \text{ and } k = l$$

$$(i,j) < (k,l) \Leftrightarrow (i,j) \leq (k,l) \text{ and } (i,j) \neq (k,l)$$

Under the conditions of quarter-plane causality, the estimates of the states at (i,j) can only depend on the data $y_{k,l}$ for $(k,l) \leq (i,j)$. This is a restrictive restraint and it has a cost on the variance of the estimators. The more natural causality for claims run-off triangles is half-plane causality. This is defined by the following ordering of the lattice:

$$(i,j) \leq (k,l) \Leftrightarrow i + j \leq k + l$$

$$(i,j) = (k,l) \Leftrightarrow i + j = k + l$$

$$(i,j) < (k,l) \Leftrightarrow i + j < k + l$$

It was implicitly assumed in chapter 4 that the above half-plane (or wave) causality applied, and it will also be used in this chapter. The consideration of quarter-plane causality is contained in Porter and Aravena (1986).

5.1.3 Recursive State Estimation

In order to estimate the state vector, the state equations (5.1.1) and (5.1.2) have to be rewritten in much the same way as the state equations were rewritten in chapter 4. The data vector at time t is

$$\underline{y}_t = \begin{bmatrix} y_{1,t} \\ y_{2,t-1} \\ \vdots \\ y_{t,1} \end{bmatrix} \qquad (5.1.3.1)$$

and time propagates along the diagonal. The state along the diagonal is defined as

$$\underline{\theta}_t = \begin{bmatrix} \underline{\beta}^v_{1,t} \\ \underline{\beta}^h_{2,t-1} \\ \underline{\beta}^v_{2,t-1} \\ \vdots \\ \underline{\beta}^v_{t-1,2} \\ \underline{\beta}^h_{t,1} \end{bmatrix}$$

In this state vector, all the components of the state vectors for the data in the data vector are present except $\underline{\beta}^h_{1,t}$ and $\underline{\beta}^v_{t,1}$. These are treated as stochastic input vectors since $\underline{\beta}^h_{0,t}$, $\underline{\beta}^v_{0,t}$, $\underline{\beta}^h_{t,0}$ and $\underline{\beta}^v_{t,0}$ are not defined.

The state at time t+1 is related to the state at time t by the system equation, which in this case is

$$\begin{bmatrix} \underline{\beta}^v_{1,t+1} \\ \underline{\beta}^h_{2,t} \\ \underline{\beta}^v_{2,t} \\ \vdots \\ \underline{\beta}^v_{t,2} \\ \underline{\beta}^h_{t+1,1} \end{bmatrix} = \begin{bmatrix} A_4 & & & & & & \\ A_2 & & & & & & \\ & A_3 & A_4 & & & & \\ & A_1 & A_2 & & & & \\ & & & \ddots & & & \\ & & & & A_3 & A_4 & \\ & & & & A_1 & A_2 & \\ & & & & & & A_3 \\ & & & & & & A_1 \end{bmatrix} \begin{bmatrix} \underline{\beta}^v_{1,t} \\ \underline{\beta}^h_{2,t-1} \\ \underline{\beta}^v_{2,t-1} \\ \vdots \\ \underline{\beta}^v_{t-1,2} \\ \underline{\beta}^h_{t,1} \end{bmatrix} + \begin{bmatrix} A_3 & 0 \\ A_1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & A_4 \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} \underline{\beta}^h_{1,t} \\ \underline{\beta}^v_{t,1} \end{bmatrix} + \begin{bmatrix} \underline{w}^v_{1,t+1} \\ \underline{w}^h_{2,t} \\ \underline{w}^v_{2,t} \\ \vdots \\ \underline{w}^v_{t,2} \\ \underline{w}^h_{t+1,1} \end{bmatrix}$$

(5.1.3.2)

In terms of the general state equation, given by equation (4.2.2),

$$G_t = \begin{bmatrix} A_4 & & & & & & & \\ A_2 & & & & & & & \\ & A_3 & A_4 & & & & & \\ & A_1 & A_2 & & & & & \\ & & & \ddots & & & & \\ & & & & A_3 & A_4 & & \\ & & & & A_1 & A_2 & & \\ & & & & & & A_3 & \\ & & & & & & A_1 & \end{bmatrix}, \quad H_t = \begin{bmatrix} A_3 & 0 \\ A_1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & A_4 \\ 0 & A_2 \end{bmatrix}$$

and the stochastic input vector, $\underline{u}_t$, is $\begin{bmatrix} \underline{\beta}^h_{1,t} \\ \underline{\beta}^v_{t,1} \end{bmatrix}$

The observation equation, which relates the data vector to the state vector, is

$$\begin{bmatrix} y_{1,t} \\ y_{2,t-1} \\ \vdots \\ y_{t,1} \end{bmatrix} = \begin{bmatrix} \underline{B}_2 & & & & & & \\ \underline{B}_1 & \underline{B}_2 & & & & & \\ & & \underline{B}_1 & \underline{B}_2 & & & \\ & & & & \ddots & & \\ & & & & \underline{B}_1 & \underline{B}_2 & \\ & & & & & & \underline{B}_1 \end{bmatrix} \begin{bmatrix} \underline{\beta}^v_{1,t} \\ \underline{\beta}^h_{2,t-1} \\ \underline{\beta}^v_{2,t-1} \\ \vdots \\ \underline{\beta}^v_{t-1,2} \\ \underline{\beta}^h_{t,1} \end{bmatrix} + \begin{bmatrix} \underline{B}_1 & 0 \\ 0 & \\ \vdots & \vdots \\ 0 & 0 \\ 0 & \underline{B}_2 \end{bmatrix} \begin{bmatrix} \underline{\beta}^h_{1,t} \\ \underline{\beta}^v_{t,1} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t-1} \\ \vdots \\ e_{t,1} \end{bmatrix} \tag{5.1.3.3}$$

which is similar to the general observation equation, (4.2.1), with

$$F_t = \begin{bmatrix} \underline{B}_2 & & & & & \\ & \underline{B}_1 & \underline{B}_2 & & & \\ & & & \underline{B}_1 & \underline{B}_2 & \\ & & & & & \underline{B}_1 \end{bmatrix}$$

The stochastic input vector is also present in the observation equation, which is generalised to:

$$\underline{y}_t = F_t \underline{\theta}_t + E_t \underline{u}_t + \underline{e}_t \qquad (5.1.3.4)$$

where $E_t = \begin{bmatrix} \underline{B}_1 & 0 \\ 0 & \\ \vdots & \vdots \\ 0 & 0 \\ 0 & \underline{B}_2 \end{bmatrix}$.

When the observed value of $\underline{y}_t$ is received, the estimate of the state vector is updated using the following theorem.

5.1.3.1 Theorem

Suppose that the observation equation and system equation are

$$\underline{y}_t = F_t \underline{\theta}_t + E_t \underline{u}_t + \underline{e}_t$$

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

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

$$\underline{e}_t \sim N(\underline{0}, V_t)$$

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

and are independent.

Suppose also that the estimate of $\underline{\theta}_t$ based on $\underline{y}_{t-1}, \underline{y}_{t-2}, \ldots, \underline{y}_1$ and other collateral

information is $\hat{\underline{\theta}}_{t|t-1}$ and its variance-covariance matrix is C_{t-1} .

Then the updated estimate of the state vector and the variance-covariance matrix of the estimate, once $\underline{y}_t$ has been received are given by equations (5.1.3.1.1) and (5.1.3.1.2).

$$\hat{\underline{\theta}}_{t+1|t} = G_t \hat{\underline{\theta}}_{t|t-1} + H_t \underline{\hat{u}}_t + K_t (\underline{y}_t - \hat{\underline{y}}_t) \qquad (5.1.3.1.1)$$

$$C_{t+1}=G_tC_tG_t'+H_tU_tH_t'-(G_tC_tF_t'+H_tU_tE_t')(F_tC_tF_t'+V_t)^{-1}(F_tC_tG_t'+H_tU_tE_t')+W_t \qquad (5.1.3.1.2)$$

$$\text{where } K_t= (G_tC_tF_t' + H_tU_tE_t') (F_tC_tF_t' + E_tU_tE_t' + V_t)^{-1} \qquad (5.1.3.1.3)$$

$$\text{and } \hat{\underline{y}}_t = F_t \hat{\underline{\theta}}_{t|t-1} + E_t \underline{\hat{u}}_t \qquad (5.1.3.1.4)$$

Proof

See Appendix 1.

5.1.4 An Example of Two-Dimensional State Estimation applied to the Chain Ladder Linear Model

The two-dimensional state estimation procedure can be illustrated by considering the chain ladder linear model. The treatment of the boundaries can be simplified by assuming that a one-dimensional recursive state structure applies. The procedure is illustrated by considering a simple dynamic model. The two-dimensional process may be more useful in situations which have a simpler structure (just one parameter: a Harrison and Stevens steady model, for example), and the example here is for illustration purposes. The overall mean and the column parameters are estimated statically and the dynamic, two-dimensional estimation is applied only to the row parameters. This can be compared to the one-dimensional recursive estimation procedure applied to the row parameters which was described in section 4.3.2.

The approach to the stochastic input vectors described in the previous section will be used, so that there is only one row parameter for the first column and second row. The state vector at time t is

$$\underline{\theta}_t = \begin{bmatrix} \mu \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_t \\ \alpha_{2,t} \\ \alpha^h_{3,t-1} \\ \alpha^v_{3,t-1} \\ \alpha^h_{4,t-2} \\ \alpha^v_{4,t-2} \\ \vdots \\ \alpha_{t,1} \end{bmatrix}$$

The two-dimensional state system, given by equations (5.1.1) and (5.1.2), is simplified to:

$$\alpha^h_{i+1,j} = \alpha^h_{i,j} + w^h_{i,j} \qquad (i \geq 2\ ,\ j \geq 1) \qquad (5.1.4.1)$$

$$\alpha^v_{i,j+1} = \alpha^v_{i,j} + w^v_{i,j} \qquad (i \geq 1\ ,\ j \geq 1) \qquad (5.1.4.2)$$

The row parameters for the first row are defined to be zero. Along the boundaries of the region in which the row parameters are non-zero,

$$\alpha^h_{2,t} = \alpha^v_{2,t} \qquad \forall\ t \qquad (5.1.4.3)$$

$$\alpha^v_{t,1} = \alpha^h_{t,1} \qquad \forall\ t \qquad (5.1.4.4)$$

This is a generalisation of the "steady model" described by Harrison and Stevens. A balanced model will be used.

The first few recursions, in wave form, are:

$$y_{1,1} = 2\,\mu + e_{1,1} \tag{5.1.4.5}$$

$$\begin{bmatrix} \mu \\ \beta_2 \\ \alpha_{2,1} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \mu + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \underline{u}_2 \tag{5.1.4.6}$$

$$\begin{bmatrix} y_{1,2} \\ y_{2,1} \end{bmatrix} = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix} \begin{bmatrix} \mu \\ \beta_2 \\ \alpha_{2,1} \end{bmatrix} + \begin{bmatrix} e_{1,2} \\ e_{2,1} \end{bmatrix} \tag{5.1.4.7}$$

$$\begin{bmatrix} \mu \\ \beta_2 \\ \beta_3 \\ \alpha_{2,2} \\ \alpha_{3,1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \beta_2 \\ \alpha_{2,1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} u_3 + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} w_{2,1}^{v} \\ w_{2,1}^{h} \end{bmatrix} \tag{5.1.4.8}$$

$$\begin{bmatrix} y_{1,3} \\ y_{2,2} \\ y_{3,1} \end{bmatrix} = \begin{bmatrix} 2 & 0 & 2 & 0 & 0 \\ 2 & 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} \mu \\ \beta_2 \\ \beta_3 \\ \alpha_{2,2} \\ \alpha_{3,1} \end{bmatrix} + \begin{bmatrix} e_{1,3} \\ e_{2,2} \\ e_{3,1} \end{bmatrix} \tag{5.1.4.9}$$

$$\begin{bmatrix} \mu \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \alpha_{2,3} \\ \alpha^{h}_{3,2} \\ \alpha^{v}_{3,2} \\ \alpha_{4,1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \beta_2 \\ \beta_3 \\ \alpha_{2,2} \\ \alpha_{3,1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} u_4 + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} w^{v}_{2,2} \\ w^{h}_{2,2} \\ w^{h}_{3,1} \\ w^{v}_{3,1} \end{bmatrix}$$

(5.1.4.10)

$$\begin{bmatrix} y_{1,4} \\ y_{2,3} \\ y_{3,2} \\ y_{4,1} \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 2 & 0 & 2 & 0 & 2 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 & 1 & 1 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} \mu \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \alpha_{2,3} \\ \alpha^{h}_{3,2} \\ \alpha^{v}_{3,2} \\ \alpha_{4,1} \end{bmatrix} + \begin{bmatrix} e_{1,4} \\ e_{2,3} \\ e_{3,2} \\ e_{4,1} \end{bmatrix}$$

(5.1.4.11)

Some properties of this two-dimensional system can be studied which indicate the general behaviour of such systems. Firstly, if the state disturbances are defined to be zero and the initial row parameters to be zero, the one-way analysis of variance is reproduced. The same results are obtained as from the least-squares analysis on the model

$$y_{ij} = \mu + \beta_j + e_{ij} .$$

This is the non-dynamic limit of the dynamic model and the static estimation results are recaptured, as is to be expected. Because of the definitions of the stochastic input vectors, it is not possible to recapture precisely the one-dimensional estimation results. This is entirely due to the treatment of the boundaries: if the stochastic input vectors are retained at each point on the boundary of the lattice, the results can be reproduced.

To illustrate the behaviour of the estimation system under different values of the horizontal and vertical state variances, the following cases will be examined in more detail:

(i) Suppose that the horizontal state variance is set to zero.

$$\text{i.e.} \quad \alpha^h_{i+1,j} = \alpha^h_{i,j} + w^h_{i,j} \qquad \text{where } w^h_{i,j} \sim N(0, 0)$$

This means that the horizontal state estimates should be identical.

Suppose also that the vertical state variance is large, so that there is no statistical connection between successive vertical state estimates.

The estimates of the horizontal row parameters are given in figure 5.7, and the estimates of the vertical row parameters in figure 5.8.

Figure 5.7

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-0.092	-0.013	0.129	0.364	-0.168	-0.375	0.558	0.238	0.230	
-0.171	-0.013	0.129	0.364	-0.168	-0.375	0.558	0.238		
-0.080	-0.013	0.129	0.364	-0.168	-0.375	0.558			
0.115	-0.013	0.129	0.364	-0.168	-0.375				
0.101	-0.013	0.129	0.364	-0.168					
0.162	-0.013	0.129	0.364						
0.099	-0.013	0.129							
0.101	-0.013								
0.167									

Figure 5.8

0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-0.092 -0.013 0.129 0.364 -0.168 -0.375 0.558 0.238 0.230

-0.171 0.146 0.155 0.247 0.388 -1.122 0.529 0.324

-0.080 0.363 0.093 0.792 -0.510 -0.133 -0.232

0.115 -0.072 0.373 0.118 0.141 0.192

0.101 0.313 0.299 0.246 0.559

0.162 0.229 0.604 0.541

0.099 0.531 0.924

0.101 0.414

0.167

Some points to note are that, as a result of the analysis of variance model, the row parameters are always zero along the first row. Also, the horizontal parameters down the first column are defined to be equal to the vertical parameters down the first column. In the remaining columns, the horizontal parameters are identical, reflecting the specification of the state variances.

(ii) Similar results are obtained for the vertical row estimates when the vertical state variance is set to zero. The results in the tables below are obtained by setting the vertical state variance to zero, with the horizontal state variance large. Figure 5.9 shows the horizontal state estimates and figure 5.10 the vertical state estimates.

Figure 5.9

0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-0.092 -0.013 0.129 0.364 -0.168 -0.375 0.558 0.238 0.230

-0.171 0.305 0.454 0.782 0.391 -1.326 1.258 0.733

-0.080 0.430 0.302 1.236 -0.598 -0.428 0.405

0.115 -0.200 0.387 0.367 -0.142 -0.298

0.101 0.200 0.327 0.510 0.291

0.162 0.054 0.571 0.743

0.099 0.419 0.955

0.101 0.301

0.167

Figure 5.10

0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-0.092 -0.013 0.129 0.364 -0.168 -0.375 0.558 0.238 0.230

-0.171 -0.171 -0.171 -0.171 -0.171 -0.171 -0.171 -0.171

-0.080 -0.080 -0.080 -0.080 -0.080 -0.080 -0.080

0.115 0.115 0.115 0.115 0.115 0.115

0.101 0.101 0.101 0.101 0.101

0.162 0.162 0.162 0.162

0.099 0.099 0.099

0.101 0.101

0.167

Concentrating on the estimates of the vertical row parameters, it can be seen that from the third row onwards the estimates are identical (reflecting the state variance specifications). The first row parameters are defined to be zero, and the second row parameters are defined to be equal to the corresponding horizontal estimates.

(iii) A conflict may arise when the specifications of cases (i) and (ii), above, are combined. If the horizontal and vertical state variances are both set to zero the estimates of the row parameters should be identical. The results of this specification are shown in figures 5.11 (the horizontal parameters) and 5.12 (the vertical parameters).

Figure 5.11

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-0.092	-0.080	-0.002	0.105	0.071	0.067	0.062	0.105	0.135	
-0.080	-0.002	0.105	0.071	0.067	0.062	0.105	0.135		
-0.002	0.105	0.071	0.067	0.062	0.105	0.135			
0.105	0.071	0.067	0.062	0.105	0.135				
0.071	0.067	0.062	0.105	0.135					
0.067	0.062	0.105	0.135						
0.062	0.105	0.135							
0.105	0.135								
0.135									

Figure 5.12

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-0.092	-0.080	-0.002	0.105	0.071	0.067	0.062	0.105	0.135	
-0.080	-0.002	0.105	0.071	0.067	0.062	0.105	0.135		
-0.002	0.105	0.071	0.067	0.062	0.105	0.135			
0.105	0.071	0.067	0.062	0.105	0.135				
0.071	0.067	0.062	0.105	0.135					
0.067	0.062	0.105	0.135						
0.062	0.105	0.135							
0.105	0.135								
0.135									

It can be seen that the estimates are identical along the diagonals: the estimates of all the parameters at each time point are the same.

5.2 Two-Dimensional Classical Time Series

The model of the form

$$y_{i,j} = \gamma_j y_{i,j-1} + \text{error} \qquad (5.2.1)$$

has been discussed in earlier chapters. This chapter has been concerned with generalising one-dimensional processes to two-dimensional lattices, and it is appropriate to consider two-dimensional classical time series.

The recursion given by equation (5.2.1) is in the same form as equation (5.0.3) and it is also possible to define a recursion in the same form as equation (5.0.4). Of course, since these are recursions for the data itself and not states, horizontal and vertical components will not be defined. The structure of a two-dimensional time series, or a simple example of it, is given by equation (5.2.2).

$$y_{i,j} = \gamma_j y_{i,j-1} + \eta_i y_{i-1,j} + \text{error} \qquad (5.2.2)$$

Equation (5.2.2) is in the form of a nearest-neighbour model, which is a generalisation of Box and Jenkins' autoregressive models to two-dimensions.

5.2.1 Nearest-Neighbour Models

Nearest-neighbour models have been used in analysing agricultural field trials, the early refences being Whittle (1954), Bartlett (1968) and Besag (1972). In particular, the correlation structure of the class of stationary, unilateral, linear autoregressives defined on a plane lattice has been derived and such results have been applied in the context of measurements at discrete

points in a continuous plane. The situation in a claims reserving context is slightly different in that it is a purely discrete problem. However, the results were derived from a time series point of view and still hold even though they were not applied in a time series context. It will be seen that the stationarity requirements are rather restrictive for forecasting.

One of the simplest models, which is related to that given by equation (5.2.2), takes the form

$$y_{i,j} = \alpha\, y_{i-1,j} + \beta\, y_{i,j-1} + e_{i,j} \qquad (5.2.1.1)$$

where $e_{i,j}$ is an error whose distribution is usually assumed to be normal.

The structure of the dependence is shown in figure 5.13

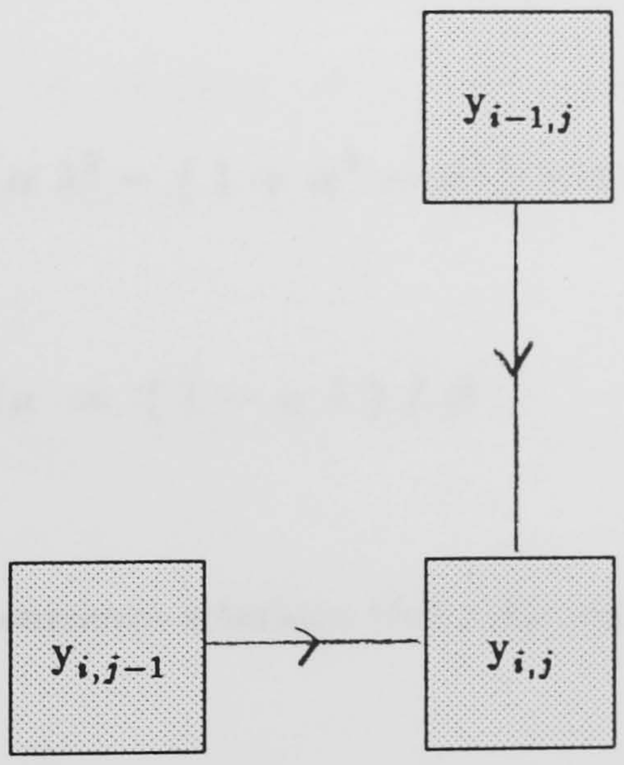


Figure 5.13

This model can be regarded as a two-dimensional generalisation of an AR(1) time series model.

In Box and Jenkins time series, the autocorrelation structure of the data plays an important part in model selection. For nearest-neighbour models, the autocovariances are defined as

$$\rho_{s,t} = E\left(y_{i,j}\, y_{i+s,j+t} \right) \qquad (5.2.1.2)$$

It is also assumed that the errors are uncorrelated, and that

$$E (e_{i,j} \, y_{i+s,j+t}) = 0 \quad , \forall \, i \, , j \, , s \, , t \qquad (5.2.1.3)$$

For the model defined by equation (5.2.1.1), the autocovariance is

$$E (y_{i,j} \, y_{i+s,j+t}) = \alpha \, E (y_{i-1,j} \, y_{i+s,j+t}) + \beta \, E (y_{i,j-1} \, y_{i+s,j+t})$$

$$\text{i.e.} \quad \rho_{s,t} = \alpha \, \rho_{s+1,t} + \beta \, \rho_{s,t+1} \qquad (5.2.1.4)$$

Equation (5.2.1.4) can be solved to give

$$\rho_{s,t} = \lambda^s \, \mu^t$$

where $\alpha \, \lambda^2 - (1 + \alpha^2 - \beta^2) \, \lambda + \alpha = 0$

$$\text{and} \, \mu = (1 - \alpha \, \lambda) \, / \, \beta$$

The recurrence relations (5.2.1.4), together with the boundary conditions

$$\rho_{s,0} = \lambda^s \quad \text{and} \quad \rho_{0,t} = \lambda^{-t} \quad (s > 0 \, , t > 0)$$

give the theoretical autocovariances.

If there are n years' data available, in the form of a triangle, there are

$$N = \frac{n (n + 1)}{2}$$

observations. The empirical autocovariance of lag s,t is

$$C_{s,t} = \frac{1}{N} \sum_{i,j} y_{i,j}\, y_{i+s,j+t} \qquad (5.2.1.5)$$

The empirical and theoretical autocovariances can be compared when selecting a model.

The model given by equation (5.2.1.1) is one example of a nearest-neighbour model, and can be generalised to the following:

$$y_{i,j} = \sum_{\{s,t\}} \alpha_{s,t}\, y_{i-s,j-t} + e_{i,j} \qquad (5.2.1.6)$$

where { s , t } is a set of duplets, for which $\alpha_{s,t}$ are non-zero.

5.2.2 Model Selection and Parameter Estimation

This section examines maximum likelihood estimation of the parameters and extends the theory of Akaike's Information Criterion (AIC) to two dimensional processes. The general nearest-neighbour model, given by equation (5.2.1.6) can be written as

$$\sum_{\{s,t\}} \alpha'_{s,t}\, y_{i-s,j-t} = e_{i,j} \qquad (5.2.2.1)$$

where $\alpha'_{0,0} = 1$

and $\alpha'_{s,t} = -\alpha_{s,t}$ for other values if (s,t).

The distributional assumption is that the errors, $e_{i,j}$, are independently, identically distributed with mean zero and variance C, and that $y_{i-s,j-t}$ and $e_{i,j}$ are independent ($(s,t) \neq (0,0)$).

The likelihood based on

$$\underline{Y} = \{ y_{i,j} , y_{i-1,j} , \cdots \}$$

can then be written as

$$l\,(\,\underline{\alpha}\,,\,C\mid \underline{y}\,)\;=\;\frac{1}{\sqrt{2\pi C}}\exp\{\,-\,\frac{1}{2C}\,(\sum_{\{s,t\}}\alpha'_{s,t}\,y_{i-s,j-t}\,)^2\,\}\qquad(5.2.2.2)$$

where $\underline{\alpha}$ is the set of parameters $\{\,\alpha'_{s,t}\,\}$.

The log-likelihood is

$$L\,(\,\underline{\alpha}\,,\,C\mid \underline{y}\,)\;=\;-\tfrac{1}{2}\log 2\pi C\;-\;\frac{1}{2C}\,(\sum_{\{s,t\}}\alpha'_{s,t}\,y_{i-s,j-t}\,)^2\qquad(5.2.2.3)$$

Maximum likelihood estimates of the parameters can be obtained by maximising the log-likelihood with respect to $\underline{\alpha}$ and C.

Two examples of maximum likelihood estimation are now given:

(a) consider a lag(1,0) model

$$y_{i,j}\;=\;\alpha_{10}\,y_{i-1,j}\;+e_{i,j}$$

The maximum likelihood estimate of α_{10} is

$$\hat{\alpha}_{10}=\frac{\sum_{i=2}^{n}\sum_{j=1}^{n}y_{i,j}\,y_{i-1,j}}{\sum_{i=2}^{n}\sum_{j=1}^{n}y_{i-1,j}^2}$$

(a) consider a lag(1,1) model

$$y_{i,j}\;=\;\alpha_{10}\,y_{i-1,j}\;+\alpha_{01}\,y_{i,j-1}+e_{i,j}$$

The maximum likelihood estimates of α_{10} and α_{01} are

$$\hat{\alpha}_{10} = \frac{\left(\sum_{i,j=2}^{n} y_{i,j}\, y_{i-1,j}\right)\left(\sum_{i,j=2}^{n} y_{i,j-1}^2\right) - \left(\sum_{i,j=2}^{n} y_{i,j-1}\, y_{i-1,j}\right)\left(\sum_{i,j=2}^{n} y_{i,j-1}\, y_{i,j}\right)}{\left(\sum_{i,j=2}^{n} y_{i,j-1}^2\right)\left(\sum_{i,j=2}^{n} y_{i-1,j}^2\right) - \left(\sum_{i,j=2}^{n} y_{i,j-1}\, y_{i-1,j}\right)^2}$$

$$\hat{\alpha}_{01} = \frac{\left(\sum_{i,j=2}^{n} y_{i,j-1}\, y_{i,j}\right)\left(\sum_{i,j=2}^{n} y_{i-1,j}^2\right) - \left(\sum_{i,j=2}^{n} y_{i,j-1}\, y_{i-1,j}\right)\left(\sum_{i,j=2}^{n} y_{i-1,j}\, y_{i,j}\right)}{\left(\sum_{i,j=2}^{n} y_{i,j-1}^2\right)\left(\sum_{i,j=2}^{n} y_{i-1,j}^2\right) - \left(\sum_{i,j=2}^{n} y_{i,j-1}\, y_{i-1,j}\right)^2}$$

Instead of doing this, the information criterion method of Akaike will be used, as it simplifies the calculations. Akaike's Information Criterion has been widely applied to one-dimensional ARMA models, and it can be extended to nearest-neighbour type models. AIC is defined as

$$-2 \log (\text{maximum likelihood}) + 2 (\text{number of parameters}) \qquad (5.2.2.4)$$

The log-likelihood for a single observation is given by equation (5.2.2.3), and the log-likelihood for the whole sample is

$$N \times L (\underline{\alpha} , C \mid \underline{y}). \qquad (5.2.2.5)$$

The maximum likelihood estimates, $\underline{\hat{\alpha}}$ and $\hat{C}$, are such that

$$L (\underline{\hat{\alpha}} , \hat{C} \mid \underline{y}) = \max_{\underline{\alpha}, C} L (\underline{\alpha} , C \mid \underline{y}). \qquad (5.2.2.6)$$

Hence AIC can be written as

$$-2\ \mathrm{N}\ \mathrm{x}\ \mathrm{L}\ (\ \underline{\hat{\alpha}}\ ,\ \hat{\mathrm{C}}\ |\ \underline{\mathrm{y}}\ )\ +\ 2\ (\text{number of parameters}\ ). \qquad (5.2.2.7)$$

The first term decreases as the number of parameters increases, and the second term increases. There is thus a trade-off between fit and parsimony: a good model will have a low value of AIC. In order to calculate the maximum log-likelihood value, the maximum likelihood estimates could be found and substituted into the log-likelihood. Instead of doing this, it is simpler to estimate the limit of the time average of the expected log-likelihood and maximise that. This yields estimates with the same characteristics as the ordinary maximum likelihood estimates. Akaike (1976) derives the results for a one-dimensional ARMA model and the analogous results for a nearest-neighbour model are now derived.

The log-likelihood is

$$\begin{aligned}
\mathrm{L}\ (\ \underline{\alpha}\ ,\ \mathrm{C}\ |\ \underline{\mathrm{y}}\ )\ &=\ -\tfrac{1}{2}\log 2\pi\mathrm{C}\ -\ \frac{1}{2\mathrm{C}}\ (\sum_{\{s,t\}}\ \alpha'_{s,t}\ \mathrm{y}_{i-s,j-t}\ )^2 \\
&=\ -\tfrac{1}{2}\log 2\pi\mathrm{C}\ -\ \frac{1}{2\mathrm{C}}\sum_{\{s,t\}}\sum_{\{k,l\}}\ \alpha'_{s,t}\ \alpha'_{k,l}\ \mathrm{y}_{i-s,j-t}\ \mathrm{y}_{i-k,j-l}
\end{aligned} \qquad (5.2.2.8)$$

The expected value of the log-likelihood is

$$-\tfrac{1}{2}\log 2\pi\mathrm{C}\ -\ \frac{1}{2\mathrm{C}}\sum_{\{s,t\}}\sum_{\{k,l\}}\ \alpha'_{s,t}\ \alpha'_{k,l}\ \rho_{k-s,l-t}\ . \qquad (5.2.2.9)$$

The theoretical autocovariances are unknown, but they can be replaced by their sample estimates:

$$-\tfrac{1}{2}\log 2\pi C \; - \frac{1}{2C}\sum_{\{s,t\}}\sum_{\{k,l\}} \alpha'_{s,t}\,\alpha'_{k,l}\,C_{k-s,l-t}\,. \qquad (5.2.2.10)$$

Maximising (5.2.2.10) with respect to $\alpha'_{s,t}$ and C gives the relations (5.2.2.11), which can be solved for $\hat{\alpha}'_{s,t}$ and $\hat{C}$.

$\sum_{\{k,l\}} \hat{\alpha}'_{k,l}\,C_{k-s,l-t} = 0$ over all combinations of (s,t) except (0,0).

(5.2.2.11)

$$\hat{C} = \sum_{\{k,l\}} \hat{\alpha}'_{k,l}\,C_{k,l}$$

[$\hat{\alpha}'_{0,0} = 1$, by definition]

The estimate of the maximum log-likelihood is then

$$-\tfrac{1}{2}\log 2\pi\hat{C}$$

and it is asymptotically justifiable to use an information criterion defined by

$$\text{AIC} \;= 2\,N\,(\,-\tfrac{1}{2}\log 2\pi\hat{C}\,) + 2\,(\text{number of parameters}\,).$$

This equivalent to using

$$\text{AIC} \;=\; -N\log\hat{C} \;+ 2\,(\text{number of parameters}\,) \qquad (5.2.2.12)$$

(neglecting N log 2π , which is common to all models.) An example of the method is given in the following (artificial) example.

5.2.2.1 Example

Data was simulated from the model

$$y_{i,j} = 0.6\, y_{i-1,j} - 0.2\, y_{i-2,j} + 0.1\, y_{i,j-1}$$

The methods described in section 5.2.2 were applied to calculate AIC for some possible models and the results are tabulated below. In the table, (s , t) indicates that the term in $y_{i-s,j-t}$ was included in the model.

Terms Included	AIC
(1,0)	91.9
(0,1)	330.7
(2,0)	349.0
(1,0) (0,1)	81.9
(1,0) (2,0)	38.2
(1,0) (0,1) (1,1)	79.6
(1,0) (0,1) (2,0)	30.7
(1,0) (0,1) (1,1) (2,0)	34.2

Thus the model selected from these, using AIC is

$$y_{i,j} = \alpha_{1,0}\, y_{i-1,j} + \alpha_{0,1}\, y_{i,j-1} + \alpha_{2,0}\, y_{i-2,j} \cdot$$

The approximate maximum likelihood estimates of $\alpha_{1,0}$, $\alpha_{0,1}$ and $\alpha_{2,0}$, using equation (5.2.2.11), are

$$\hat{\alpha}_{1,0} = 0.53$$

$$\hat{\alpha}_{0,1} = 0.080$$

$$\hat{\alpha}_{2,0} = -0.21$$

and the model which would have been estimated from the data is

$$y_{i,j} = 0.53\, y_{i-1,j} - 0.21\, y_{i-2,j} + 0.080\, y_{i,j-1}$$

5.2.3 Stationarity

Nearest neighbour models of the type given by equation (5.2.1.6) can be applied to any data which is in a suitable form. However, it is necessary first that the data are stationary if the estimation techniques described in the previous sections are to be used. The stationarity requirements are that

$$E\,(\,y_{i,j}\,) \text{ and } \mathrm{Var}\,(\,y_{i,j}\,)$$

do not depend on (i , j) and that

$$E\,(\,y_{i,j}\,y_{i+s,j+t}\,)$$

depends only on (s , t), not on (i , j).

Claims data are not usually stationary in this sense, and so have to be pre-processed before AIC can be applied. When forecasting the expected outstanding claims therefore, the present theory of classical time series has relatively limited use. However, they can be effective for searching for correlations between the residuals and can be applied to the residuals after one of the models described earlier has been fitted. Further research is needed on methods of reducing the data to stationarity.

5.2.4 State Space Representation of Two-Dimensional Classical Time Series

The state space representation of one-dimensional classical time series has proved effective and useful in the analysis of ordinary time series. All autoregressive processes can be represented in state space form (see, for example, Davis and Vinter (1985)). It is possible to represent some two-dimensional time series in state space form, although a general formulation has not been found. The following theorem proves the interchangeability of a state space model and a two-dimensional time series model with (1,0), (0,1) and (1,1) lag terms.

5.2.4.1 Theorem

Consider the model

$$y_{i,j} = \alpha_{10} y_{i-1,j} + \alpha_{01} y_{i,j-1} + \alpha_{11} y_{i-1,j-1} + e_{i,j} \qquad (5.2.4.1.1)$$

where $e_{i,j}$ are independent, identically distributed white noises.

There exists some two-dimensional state process such that

$$\begin{bmatrix} x^h_{i+1,j} \\ x^v_{i,j+1} \end{bmatrix} = \begin{bmatrix} \alpha_{10} & \alpha_{11}+\alpha_{10}\alpha_{01} \\ 1 & \alpha_{01} \end{bmatrix} \begin{bmatrix} x^h_{i,j} \\ x^v_{i,j} \end{bmatrix} + \begin{bmatrix} 0 & \alpha_{10} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e_{i+1,j} \\ e_{i,j+1} \end{bmatrix} \qquad (5.2.4.1.2)$$

$$y_{i,j} = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x^h_{i,j} \\ x^v_{i,j} \end{bmatrix} \qquad (5.2.4.1.3)$$

Proof

The two-dimensional autoregression given by equation (5.2.4.1.1) can be written as

$$\begin{aligned} y_{i,j} - \alpha_{01} y_{i,j-1} &= \alpha_{10} y_{i-1,j} + \alpha_{11} y_{i-1,j-1} + e_{i,j} \\ &= \alpha_{10} (y_{i-1,j} - \alpha_{01} y_{i-1,j-1}) + (\alpha_{11} + \alpha_{10}\alpha_{01}) y_{i-1,j-1} + e_{i,j} \end{aligned} \qquad (5.2.4.1.4)$$

Let $D(i,j)$ represent { all information before (i,j) }

i.e. $D(i,j) = \{ y_{k,l} : (k,l) < (i,j) \}$ (5.2.4.1.5)

Let F and G represent backward shift operators such that

$$F\, y_{i,j} = y_{i-1,j}$$

$$G\, y_{i,j} = y_{i,j-1} \quad (5.2.4.1.6)$$

From equation (5.2.4.1.4),

$$(y_{i,j+1} - \alpha_{01} y_{i,j}) - \alpha_{10} (y_{i-1,j+1} - \alpha_{01} y_{i-1,j}) = (\alpha_{11} + \alpha_{10}\alpha_{01}) y_{i-1,j} + e_{i,j+1}$$

i.e. $(1 - \alpha_{10} F) (y_{i,j+1} - \alpha_{01} y_{i,j}) = (\alpha_{11} + \alpha_{10}\alpha_{01}) y_{i-1,j} + e_{i,j+1}$

Therefore

$$y_{i,j+1} - \alpha_{01} y_{i,j} = (1 - \alpha_{10} F)^{-1} (\alpha_{11} + \alpha_{10}\alpha_{01}) y_{i-1,j} + (1 - \alpha_{10} F)^{-1} e_{i,j+1} \quad (5.2.4.1.7)$$

Now, from equation (5.2.4.1.4),

$$y_{i+1,j+1} - \alpha_{01} y_{i+1,j} = \alpha_{10} (y_{i,j+1} - \alpha_{01} y_{i,j}) + (\alpha_{11} + \alpha_{10}\alpha_{01}) y_{i,j} + e_{i+1,j+1}$$

and since $e_{i+1,j+1}$ has expectation zero, the expectation of $y_{i+1,j+1} - \alpha_{01} y_{i+1,j}$, given $D(i+1,j+1)$ is

$$E [y_{i+1,j+1} - \alpha_{01} y_{i+1,j} \mid D(i+1,j+1)]$$

$$= \alpha_{10} E [(y_{i,j+1} - \alpha_{01} y_{i,j}) \mid D(i+1,j+1)] + (\alpha_{11} + \alpha_{10}\alpha_{01}) y_{i,j} \quad (5.2.4.1.8)$$

From equation (5.2.4.1.7),

$$E [y_{i,j+1} - \alpha_{01} y_{i,j} \mid D(i+1,j+1)]$$

$$= (1 - \alpha_{10} F)^{-1} (\alpha_{11} + \alpha_{10}\alpha_{01}) y_{i-1,j} + e_{i,j+1} + \alpha_{10}\, e_{i-1,j+1} + \cdots$$

and

$$E [y_{i,j+1} - \alpha_{01} y_{i,j} \mid D(i , j+1)]$$

$$= (1 - \alpha_{10} F)^{-1} (\alpha_{11} + \alpha_{10}\alpha_{01}) y_{i-1,j} + \alpha_{10}\, e_{i-1,j+1} + \alpha_{10}^2\, e_{i-2,j+1} + \cdots$$

Therefore

$$E [y_{i+1,j+1} - \alpha_{01} y_{i+1,j} \mid D(i+1 , j+1)] = \alpha_{10}\, E [y_{i,j+1} - \alpha_{01} y_{i,j} \mid D(i , j+1)]$$
$$+ \alpha_{10}\, e_{i,j+1} + (\alpha_{11} + \alpha_{10}\alpha_{01}) y_{i,j}$$
(5.2.4.1.9)

Let $x_{i,j}^h = E [y_{i,j+1} - \alpha_{01} y_{i,j} \mid D(i , j+1)]$

and $x_{i,j}^v = y_{i,j}$ (5.2.4.1.10)

Equation (5.2.4.1.9) can be written as

$$x_{i+1,j}^h = \alpha_{10}\, x_{i,j}^h + (\alpha_{11} + \alpha_{10}\alpha_{01}) x_{i,j}^v + \alpha_{10}\, e_{i,j+1}$$ (5.2.4.1.11)

Also

$$y_{i,j+1} = E [y_{i,j+1} - \alpha_{01} y_{i,j} \mid D(i , j+1)] + \alpha_{01} y_{i,j} + e_{i,j+1}$$

which can be written as

$$x_{i,j+1}^v = x_{i,j}^h + \alpha_{01}\, x_{i,j}^v + e_{i,j+1}$$ (5.2.4.1.12)

Equations (5.2.4.1.10) to (5.2.4.1.12) together give the state space representation of the system given by equations (5.2.4.1.2) and (5.2.4.1.3).

It should be noted that this representation is not unique.

CHAPTER 6

Conclusions

6.0 Introduction

In conclusion, some practical aspects of claims reserving have to be considered. These are the stability of the predictions, the consideration of negative incremental claims and the use of the predictions, their standard errors and the 'safe' reserves in practice. These subjects are discussed in sections 6.1 to 6.3.

The empirical Bayes Method has been compared with the dynamic forecasting method which uses the Kalman filter. An important point is that the Kalman filter requires the state variances to be specified before the analysis begins. In the examples in chapter 4, 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. This requirement to specify variances is a disadvantage for the Kalman filter and practitioners will find the empirical Bayes method distinctly easier to use, even if it may be thought that the recursive relationship between the parameters is appropriate. This point is discussed further in section 6.4, where some ideas for further research are given.

It should be noted that the methods in this thesis can be applied to any models for claims reserving which can be written in the form of a linear model. This includes, for example, the separation method.

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.

A useful extension of the usual analyses which are carried out on claims data is the more detailed consideration of the development factors and the proportions of ultimate claims: the parameters of the chain ladder and multiplicative models. The maximum likelihood theory developed in this thesis has been used in a practical setting by the Institute of Actuaries Working Party on Claims Reserving (1989).

6.1 Stability of Predictions

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, or the prior information. 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 6.1

Original Results			Revised Results		
No prior Information	Dynamic Estimation	Empirical Bayes	No prior Information	Dynamic Estimation	Empirical Bayes
110927	109955	109448	110927	109958	110094
482157	491787	479568	482157	491822	481329
660810	686441	655656	660810	686637	657998
1090752	1076957	1033109	1090752	1078058	1039692
1530532	1486991	1388261	1530532	1491978	1400466
2310959	2217311	2002772	2310959	2239482	2024720
3806976	3309887	3018896	3806976	3399256	3063229
4452396	4545466	3780759	4452396	4847221	3819051
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.

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.

6.2 Negative Incremental Claims

The incremental claims, $C_{i,j}$, have been logged before the linear models have been applied, and the implicit assumption that the data are lognormally distributed requires the data to be positive. Because of the nature of some lines of business, and the practice of using IBNR claims, negative values do arise and there has to be a sensible way to deal with them.

Firstly, it may be possible to infer from the data the reason for specific negative values: for example a large cumulative value at one delay period may be offset by an adjustment in the following delay period. If it is thought that these values are not indicative of an intrinsic structure in the data, but are due to some accounting phenomenon (for example), then it may be appropriate to adjust the data to avoid negative values arising.

Secondly, it is often sufficient to add a suitable constant to all the data values to ensure that they are positive before the analysis and then to subtract the same constant from all fitted and forecast values. This is the approach which has been used, at present, when any data sets have negative values.

The third, and most sophisticated, method is to model the data according to the following distribution:

$$f(C_{i,j}) = \alpha_{i,j}\, f(C_{i,j} \mid C_{i,j} \leq 0) + (1 - \alpha_{i,j})\, f(C_{i,j} \mid C_{i,j} > 0) \qquad (6.3.1)$$

The positive data incremental claims can be analysed using the linear models as above. The negative incremental claims can also be analysed in the same way, or a suitable negative weight can be chosen, depending on which is more appropriate in the light of the data. The probability that the incremental claim at position (i , j) can also be estimated: a model could be applied, or it could be assumed that

$$\alpha_{i,j} = \alpha \qquad \forall\ i\ ,\ \forall\ j. \tag{6.3.2}$$

and α estimated.

The forecast or fitted value is then a weighted average:

$$\hat{C}_{i,j} = \hat{\alpha}_{i,j}\ (\ \hat{C}_{i,j} \mid C_{i,j} \leq 0\ ) + (\ 1 - \hat{\alpha}_{i,j}\ )\ (\ \hat{C}_{i,j} \mid C_{i,j} > 0\ ) \tag{6.3.3}$$

Generally, the second method should suffice, unless the data is very difficult.

6.3 Reserves

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 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.

6.4 Further Research

There are many extensions of the work in this thesis. The classical least squares theory has been implemented in GLIM by Renshaw (1989) and in SuperCalc5 by the Commercial Union Assurance Company. It would be desirable to implement the Bayesian methods, and in particular the results of chapter 3, in these packages.

The consideration of negative incremental claims needs further attention, and some method of estimation without taking logs, but retaining the basic overall structure is needed. There is also scope for examining whether the variation is sufficiently modelled in the lognormal distributional assumptions.

The models in this thesis have assumed that the data is lognormally distributed. On the basis

of this assumption, a rigorous and statistically sound theory of claims reserving has been developed, but it would be desirable to consider the underlying assumptions in more detail using a large data set.

It is also possible to model numbers of claims using a Poisson model. The methods of this thesis could be applied to such a model.

The empirical Bayes results have a credibilty theory interpretation and can be compared with the results of de Vylder (1982). It would be interesting to compare empirical results from the two methods. It would also be possible to produce recursive empirical Bayes results which could be compared with recursive credibility estimates.

It has been stated that the Kalman filter estimation method has the disadvantage that the state variances have to be specified in advance. One way to get over this problem would be to use a classical estimation method as suggested by Harvey — see for example Harvey and Fernandes (1989). This would mean either reanalysing the data each time a new diagonal is received, or assuming that the estimates from the early data are appropriate throughout.

The time series models of chapter 5 have great scope for further work. In particular, the problem of stationarity needs further attention. It is also possible to apply the two-dimensional Kalman filter theory to the Bayesian forecasting models in Harrison and Stevens (1976) other than the regression models which are the subject of this thesis.

The results of this thesis can be used to model Age/Period/Cohort data, and a full study of the methods of Osmond and Gardner (1982), Osmond, Gardner and Acheson (1982) and Walter, Miller and Lee (1976) is justified. It would also be possible to apply the results to demographic theory, a recent paper which uses similar models to those used here being Raeside (1988). The application of the results to the epidemiological theory in the paper by Zeger, See and Diggle (1989) would also be possible.

Appendix 1

Theorem

Consider the state space system defined by the following observation and system equations:

$$\underline{y}_t = F_t \underline{\theta}_t + E_t \underline{u}_t + \underline{e}_t \qquad \text{(A1.1)}$$

$$\underline{\theta}_{t+1} = G_t \underline{\theta}_t + H_t \underline{u}_t + \underline{w}_t \qquad \text{(A1.2)}$$

where $\underline{e}_t \sim N(\underline{0}, V_t)$,

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

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

and are independent. Further $\underline{e}_t$,$\underline{u}_t$, $\underline{w}_t$ are sequentially independent.

Suppose $\underline{\theta}_t \mid (\underline{y}_1, \underline{y}_2, \cdots, \underline{y}_{t-1}) \sim N(\hat{\underline{\theta}}_{t|t-1}, C_t)$. (A1.3)

From equations (A1.3) and (A1.1), the distribution of $\underline{y}_t$ given information up to time t-1 is

$$\hat{\underline{y}}_{t|t-1} \sim N(F_t \hat{\underline{\theta}}_{t|t-1}, F_t C_t F_t' + V_t) \qquad \text{(A1.4)}$$

Then

$$\hat{\underline{\theta}}_{t+1|t} = G_t \hat{\underline{\theta}}_{t|t-1} + H_t \hat{\underline{u}}_t + K_t (\underline{y}_t - \hat{\underline{y}}_t) \qquad \text{(A1.5)}$$

where

$$K_t = (G_t C_t F_t' + H_t U_t E_t')(F_t C_t F_t' + E_t U_t E_t' + V_t)^{-1} \qquad \text{(A1.6)}$$

and

$$C_{t+1} = G_t C_t G_t' + H_t U_t H_t' - (G_t C_t F_t' + H_t U_t E_t')(F_t C_t F_t' + V_t)^{-1} (G_t C_t F_t' + H_t U_t E_t') + W_t \qquad \text{(A1.7)}$$

Proof

From equations (A1.2) and (A1.5),

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

$$= G_t(\underline{\theta}_t - \underline{\hat{\theta}}_{t|t-1}) - K_t(F_t\underline{\theta}_t + E_t\underline{u}_t + \underline{e}_t - F_t\underline{\hat{\theta}}_{t|t-1} - E_t\underline{\hat{u}}_t) + H_t(\underline{u}_t - \underline{\hat{u}}_t) + \underline{w}_t$$

(A1.8)

(using equation (A1.1))

Hence the covariance matrix of $\underline{\hat{\theta}}_{t+1|t}$, C_{t+1} , satisfies the following equation:

$$C_{t+1} = (G_t - K_tF_t)C_t(G_t - K_tF_t)' + (H_t - K_tE_t)U_t(H_t - K_tE_t)' + K_tV_tK_t' + W_t$$

(A1.9)

Differentiating equation (A1.9) with respect to K_t and equating to zero gives

$$(G_t - K_tF_t)C_tF_t' + (H_t - K_tE_t)U_tE_t' - K_tV_t = 0 \qquad \text{(A1.10)}$$

Equation (A1.10) can be solved for K_t in the form given by equation (A1.6), and substituting for K_t into equation (A1.9) gives the Riccati equations, (A1.7):

$$C_{t+1} = G_tC_tG_t' - 2K_tF_tC_tG_t' + K_tF_tC_tF_t'K_t' + H_tU_tH_t' - 2K_tE_tU_tH_t' + K_tE_tU_tE_t'K_t' + K_tV_tK_t' + W_t$$

$$= G_tC_tG_t' + H_tU_tH_t' - 2K_t(F_tC_tG_t' + E_tU_tH_t') + K_t(F_tC_tF_t' + E_tU_tE_t' + V_t)K_t' + W_t$$

$$= G_tC_tG_t' + H_tU_tH_t' - K_t(F_tC_tG_t' + E_tU_tH_t') + W_t$$

$$= G_tC_tG_t' + H_tU_tH_t' - (G_tC_tF_t' + H_tU_tE_t')(F_tC_tF_t' + V_t)^{-1}(G_tC_tF_t' + H_tU_tE_t') + W_t$$

Appendix 2

This appendix contains examples of the computer programmes which have been written to implement the methods in this thesis. The first illustrates the Standard Chain Ladder Linear Model, giving the unbiased and maximum likelihood estimates. This programme has to be used with a vague prior distribution in its present form, although it is straightforward to change the estimates to Bayesian estimates, as given in chapter 3. The Bayesian version of this programme, which is identical up to the estimation of outstanding claims has been used for the results in chapter 3. The second programme is an example of the empirical Bayes estimation methods.

The least squares results have been checked with the results produced by GLIM, see Renshaw (1989).

```
c  This programme implements the Chain Ladder Linear Model, and gives
c  the Unbiased and Maximium Likelihood Estimates of Outstanding Claims.
c  Least squares results are obtained with a vague prior, and it is also
c  possible to enter Prior Information, althuogh the Prediction Equations
c  have to be altered.
      double precision y(150,150),x(150,150),d(150,150),dummy(150,150)
     &,d1(150),wk1(150),wk2(150),d2(150,150),z(150)
     &,bstar(150,150),yhat(150,150),b(150,150)
     &,bstar1(150,150),y1(150,150),p(150),bhat1(150)
     &,var,xn,X02AAF,d3(150,150),theta(150),c(150,150)
     &,tot(150),v(150,150),rowvar(150),finf(150),xnp
     &,mean(150,150),varmean(150,150),resid(150,150),yfit(150,150)
     &,mean1(150,150),varmean1(150,150),tot1(150),rowvar1(150)
     &,yhat1(150,150),var1(150,150)
      open (01,file='claims')
      open (02,file='result')
      open (03,file='plotdata')
      rewind 01
      read(01,*) ni,nj
      read(01,*) crit
      np=ni+nj-1
      read(01,*) ((y1(i,j),j=1,min(nj,ni-i+1)),i=1,ni)
      read (01,*) (p(i),i=1,ni)
      read (01,*) (finf(i),i=1,ni)
      read(01,*) (theta(i),i=1,np)
      do 5 i=1,np
       do 4 j=1,np
```

```
      c(i,j)=0.0
4     continue
      read(01,*) c(i,i)
5     continue
      k=0
      do 7 i=1,ni
      do 7 j=1,min(nj,ni-i+1)
       k=k+1
       y(k,1)=log(y1(i,j))-log(p(i))+log(finf(i+j-1))
7     continue
      n=k
      m=n-np
      xn=real(n)
      k=0
      do 30 i=1,ni
      do 30 j=1,min(nj,ni-i+1)
       k=k+1
       x(k,1)=1.0
       if (i.eq.1) goto 11
       x(k,i)=1.0
11     continue
       if (j.eq.1) goto 20
       x(k,ni+j-1)=1.0
20     continue
30    continue
      call prod1(x,x,dummy,n,np,np)
      do 32 i=1,np
      do 32 j=1,np
       d2(i,j)=dummy(i,j)
32    continue
      call prod1(x,y,d,n,np,1)
      do 34 i=1,np
       d1(i)=d(i,1)
34    continue
      call F04ASF(d2,150,d1,np,bhat1,wk1,wk2,ifail)
      do 38 i=1,np
       bstar(i,1)=bhat1(i)
38    continue
40    var=0.0
      call prod2(x,bstar,yhat,n,np,1)
```

```
      k=0
      do 50 i=1,ni
      do 50 j=1,min(nj,ni-i+1)
       k=k+1
       var=var+(y(k,1)-yhat(k,1))*(y(k,1)-yhat(k,1))
50    continue
      xnp=real(np)
      s=var/xn
      var=var/(xn-xnp)
      do 60 i=1,np
       do 59 j=1,np
        d3(i,j)=dummy(i,j)/var
59     continue
       d3(i,i)=d3(i,i)+1/c(i,i)
60    continue
      ifail=0
      call F01ACF(np,X02AAF(IT),d3,150,b,150,z,l,ifail)
      print *,ifail
      call square (d3,d2,np)
      do 80 i=1,np
       d3(i,1)=d(i,1)/var+theta(i)/c(i,i)
80    continue
      call prod2 (d2,d3,bstar1,np,np,1)
      ncrit=0
      do 130 i=1,np
       if (abs(bstar(i,1)-bstar1(i,1)).ge.crit) then
       ncrit =1.0
       endif
       bstar(i,1)=bstar1(i,1)
130    continue
      if (ncrit .eq.1) goto  40
      do 140 i=1,np
       write (02,9997) bstar(i,1),bhat1(i),sqrt(d2(i,i))
140    continue
      write (02,*) var
      nrow=0
      k1=1
      do 300 i=1,ni
       tot(i)=0.0
       tot1(i)=0.0
```

```
      do 250 j=1,nj
       yhat(i,j)=bstar(1,1)
       if (i .eq. 1) goto 200
       yhat(i,j)=yhat(i,j)+bstar(i,1)
200     if (j.eq.1) goto 210
       yhat(i,j)=yhat(i,j)+bstar(ni+j-1,1)
210     continue
       do 240 k=1,nj
        v(j,k)=d2(1,1)
        if (i.eq.1) goto 224
        v(j,k)=v(j,k)+d2(i,i)+2*d2(1,i)
        if (j.eq.1) goto 222
        v(j,k)=v(j,k)+d2(i,ni+j-1)
222      continue
        if (k.eq.1) goto 224
        v(j,k)=v(j,k)+d2(i,ni+k-1)
224      continue
        if (j.eq.1) goto 226
        v(j,k)=v(j,k)+d2(1,ni+j-1)
        if (k.eq.1) goto 226
        v(j,k)=v(j,k)+d2(ni+j-1,ni+k-1)
226      continue
       if (k.eq.1) goto 228
       v(j,k)=v(j,k)+d2(1,ni+k-1)
228      continue
240     continue
      yfit(i,j)=yhat(i,j)
      if (j.gt.min(nj,ni-i+1)) goto 241
      resid(i,j)=y(k1,1)-yfit(i,j)
      k1=k1+1
241     continue
      mean(i,j)=exp(yhat(i,j)+0.5*s)
      tfunc=(m+1)*(var-v(j,j))/(2*m)
      call g(m,tfunc,xg)
      mean1(i,j)=exp(yhat(i,j))*xg
      tfunc=4*tfunc
      call g(m,tfunc,xg)
      var1(i,j)=xg
      tfunc=(m+1)*(var-2*v(j,j))/m
      call g(m,tfunc,xg)
```

```
      var1(i,j)=var1(i,j)-xg
      var1(i,j)=var1(i,j)*exp(2*yhat(i,j))
250   continue
     do 255 j=1,nj
     do 255 k=1,nj
     v1=exp(yhat(i,j)+yhat(i,k)+0.5*(v(j,j)+v(k,k)))
     varmean(j,k)=(1-2*s/xn)**(0.5*(xnp-xn))
     varmean(j,k)=varmean(j,k)*exp(v(j,k))
     varmean(j,k)=varmean(j,k)-(1-s/xn)**(xnp-xn)
     varmean(j,k)=varmean(j,k)*v1
     tfunc=(m+1)*(var-v(k,k))/(2*m)
     call g(m,tfunc,xg)
     varmean1(j,k)=mean1(i,j)*exp(yhat(i,k))*xg
     tfunc=(m+1)*(var-v(j,k)-0.5*(v(j,j)+v(k,k)))/m
     call g(m,tfunc,xg)
     varmean1(j,k)=varmean1(j,k)-exp(yhat(i,j)+yhat(i,k))*xg
255   continue
      if (i.le.ni-nj+1) goto 299
      rowvar(i)=0.0
      rowvar1(i)=0.0
      do 280 j=nj-nrow,nj
       tot(i)=tot(i)+mean(i,j)
       tot1(i)=tot1(i)+mean1(i,j)
       if (j.eq.nj) goto 261
       do 260 k=j+1,nj
     rowvar(i)=rowvar(i)+2*varmean(j,k)
     rowvar1(i)=rowvar1(i)+2*varmean1(j,k)
260     continue
261     continue
     rowvar(i)=rowvar(i)+varmean(j,j)
     rowvar1(i)=rowvar1(i)+varmean1(j,j)+var1(i,j)
280   continue
     nrow=nrow+1
     tot(i)=tot(i)*p(i)
     tot1(i)=tot1(i)*p(i)
     rowvar(i)=rowvar(i)*p(i)*p(i)
     rowvar1(i)=rowvar1(i)*p(i)*p(i)
299   do 300 j=1,nj
      yhat(i,j)=mean(i,j)*p(i)
      yhat1(i,j)=mean1(i,j)*p(i)
```

```
300   continue
      do 310 i=1,ni
       write (02,9999) (yhat(i,j),j=1,nj)
310    continue
       do 311 i=1,ni
       write (02,9999) (yhat1(i,j),j=1,nj)
311   continue
      do 320 i=1,ni
       write (02,9998) tot(i),sqrt(rowvar(i))
320   continue
      do 321 i=1,ni
       write (02,9998) tot1(i),sqrt(rowvar1(i))
321   continue
      write(03,*) k1-1
      do 350 i=1,ni
      do 350 j=1,min(nj,ni-i+1)
      write(03,9990) yfit(i,j),resid(i,j)
350   continue
9990   format(1x,2f10.6)
9997   format(3x,3(f6.3,3x))
9998  format(1x,3(f12.0,1x))
9999  format(1x,10(f10.0,1x))
      stop
      end
      subroutine g(m1,t1,xg1)
      term=1
      sum=1
      k=1
1      term=term*m1*m1*t1/((m1+2*k-2)*(m1+1)*k)
      sum=sum+term
      k=k+1
      if (abs(term) .lt. 0.0002) goto 2
      goto 1
2      xg1=sum
      return
      end
         subroutine prod1(a1,a2,a3,n1,n2,n3)
         double precision a1(150,150),a2(150,150),a3(150,150)
         do 2 i=1,n2
         do 2 j=1,n3
```

```
      a3(i,j)=0.0
      do 1 k=1,n1
       a3(i,j)=a3(i,j)+a1(k,i)*a2(k,j)
1      continue
2     continue
      return
      end
      subroutine prod2(a1,a2,a3,n1,n2,n3)
      double precision a1(150,150),a2(150,150),a3(150,150)
      do 2 i=1,n1
      do 2 j=1,n3
       a3(i,j)=0.0
       do 1 k=1,n2
        a3(i,j)=a3(i,j)+a1(i,k)*a2(k,j)
1      continue
2     continue
      return
      end
      subroutine square(a1,b1,n1)
      double precision a1(150,150),b1(150,150)
      do 10 i=1,n1
       do 8 j=1,i
        b1(i,j)=a1(i+1,j)
8      continue
10     continue
      do 20 i=1,n1-1
       do 18 j=i+1,n1
        b1(i,j)=b1(j,i)
18     continue
20     continue
      return
      end
```

```
c This programme is an empirical Bayes version of the chain
c ladder linear model
      double precision y(150,150),x(150,150),d(150,150),dummy(150,150)
     &, d1(150),wk1(150),wk2(150),d2(150,150)
     &,bstar(150,150),yhat(150,150),b(150,150)
     &,bstar1(150),y1(150,150),p(150),bhat1(150)
     &,var,varalp,xni,xn,salp,ssalp,X02AAF
     &,wks1(150),wks2(150),aa(150,150),z(150)
     &,v(12,12,12,12),tot(150),rowvar(150),finf(150)
     & ,ymax(150,150),totmax(150),resid(150,150),yfit(150,150)
     & ,yhat1(150,150)
      open (01,file='claims')
      open (02,file='result')
      open (03,file='plotdata')
      rewind 01
      read(01,*) ni,nj
      read(01,*) crit
      np=ni+nj-1
      read(01,*) ((y1(i,j),j=1,min(nj,ni-i+1)),i=1,ni)
      read (01,*) (p(i),i=1,ni)
      read (01,*) (finf(i),i=1,ni)
      k=0
      do 27 i=1,ni
      do 27 j=1,min(nj,ni-i+1)
      k=k+1
      y(k,1)=log(y1(i,j))-log(p(i))+log(finf(i+j-1))
27    continue
      n=k
      xn=real(n)
      k=0
      do 30 i=1,ni
      do 30 j=1,min(nj,ni-i+1)
      k=k+1
      x(k,1)=1.0
      if (i.eq.1) goto 11
      x(k,i)=1.0
11    continue
      if (j.eq.1) goto 20
      x(k,ni+j-1)=1.0
20    continue
```

```
30   continue
     call prod1(x,x,dummy,n,np,np)
     do 32 i=1,np
     do 32 j=1,np
     d2(i,j)=dummy(i,j)
32    continue
      call prod1(x,y,d,n,np,1)
      do 34 i=1,np
      d1(i)=d(i,1)
34    continue
     call F04ASF(d2,150,d1,np,bhat1,wk1,wk2,ifail)
     xni=real(ni)
      do 50 i=1,np
      do 49 j=1,np
       b(i,j)=0.0
49    continue
      bstar(i,1)=bhat1(i)
50    continue
      do 51 i=2,ni
      do 51 j=2,ni
      if (i.eq.j) then
      b(i,j)=b(i,j)+(xni-2)/(xni-1)
       else
       b(i,j)=b(i,j)-1/(xni-1)
      endif
 51    continue
52       salp=0.0
         ssalp=0.0
      do 60 i=2,ni
      salp=salp+bstar(i,1)
      ssalp=ssalp+bstar(i,1)*bstar(i,1)
 60    continue
      varalp=(.2+ssalp-salp*salp/(xni-1))/(xni+.1)
      call prod2(x,bstar,yhat,n,np,1)
       var=0.0
       k=0
       do 100 i=1,ni
       do 100 j=1,min(nj,ni-i+1)
       k=k+1
       var=var+(y(k,1)-yhat(k,1))*(y(k,1)-yhat(k,1))/(xn+2)
```

```
100     continue
        do 110 i=1,np
        do 110 j=1,np
        d(i,j)=b(i,j)*var/varalp
        d(i,j)=d(i,j)+dummy(i,j)
110     continue
        call F04ATF(d,150,d1,np,bstar1,aa,150,wks1,
     &  wks2,ifail)
        ncrit=0
        do 130 i=1,np
        if (abs(bstar(i,1)-bstar1(i)).ge.crit) then
        ncrit =1.0
        endif
        bstar(i,1)=bstar1(i)
130     continue
        if (ncrit .eq.1) goto  52
        call F01ACF(np,X02AAF(IT),d,150,b,150,z,l,ifail)
        call square(d,dummy,np)
        do 135 i=1,np
        do 135 j=1,np
        dummy(i,j)=dummy(i,j)*var
135     continue
        do 140 i=1,np
        write (02,9997) bstar(i,1),bhat1(i),sqrt(dummy(i,i))
140     continue
        write (02,*) "varalp=", varalp
      nrow=0
      k1=1
      do 160 i=1,ni
      do 160 l=1,ni
      do 160 j=1,nj
      do 160 k=1,nj
      v(i,j,l,k)=dummy(1,1)
      if (i.eq.1) goto 150
      v(i,j,l,k)=v(i,j,l,k)+dummy(1,i)
      if (l.eq.1) goto 150
      v(i,j,l,k)=v(i,j,l,k)+dummy(i,l)
150   if (l.eq.1) goto 152
      v(i,j,l,k)=v(i,j,l,k)+dummy(1,l)
      if (j.eq.1) goto 152
```

```
      v(i,j,l,k)=v(i,j,l,k)+dummy(l,ni+j-1)
152   if (k.eq.1) goto 154
      v(i,j,l,k)=v(i,j,l,k)+dummy(1,ni+k-1)
      if (i.eq.1) goto 154
      v(i,j,l,k)=v(i,j,l,k)+dummy(i,ni+k-1)
154    if (j.eq.1) goto 156
      v(i,j,l,k)=v(i,j,l,k)+dummy(1,ni+j-1)
      if (k.eq.1) goto 156
      v(i,j,l,k)=v(i,j,l,k)+dummy(ni+j-1,ni+k-1)
156    continue
      if ((i.eq.l).and.(j.eq.k).and.(j.gt.min(nj,ni-
     &i+1))) then
      v(i,j,l,k)=v(i,j,l,k)+var
      endif
160    continue
          do 300 i=1,ni
          tot(i)=0.0
          totmax(i)=0.0
          do 250 j=1,nj
          yhat(i,j)=bstar(1,1)
          if (i .eq. 1) goto 200
          yhat(i,j)=yhat(i,j)+bstar(i,1)
200           if (j.eq.1) goto 210
          yhat(i,j)=yhat(i,j)+bstar(ni+j-1,1)
210        continue
          yfit(i,j)=yhat(i,j)
          if (j.gt. min(nj,ni-i+1)) goto 241
          resid(i,j)=y(k1,1)-yfit(i,j)
          k1=k1+1
241        continue
          yhat(i,j)=yhat(i,j)+0.5*v(i,j,i,j)
          yhat(i,j)=exp(yhat(i,j))
          if (j.gt. min(nj,ni-i+1)) goto 247
          ymax(i,j)=exp(yfit(i,j)-log(finf(i+j-1)))*p(i)
247        continue
250        continue
            if (i.le.ni-nj+1) goto 299
            rowvar(i)=0.0
            do 280 j=nj-nrow,nj
            tot(i)=tot(i)+yhat(i,j)
```

```
            totmax(i)=totmax(i)+ymax(i,j)
            if (j.eq.nj) goto 261
            do 260 k=j+1,nj
      rowvar(i)=rowvar(i)+2*yhat(i,k)*(exp(v(i,j,i,k))-1)*yhat(i,j)
260           continue
261           continue
          rowvar(i)=rowvar(i)+yhat(i,j)*yhat(i,j)*(exp(v(i,j,i,j))-1)
280         continue
     nrow=nrow+1
           tot(i)=tot(i)*p(i)
         rowvar(i)=rowvar(i)*p(i)*p(i)
299        do 300 j=1,nj
           yhat1(i,j)=yhat(i,j)*p(i)
300       continue
         do 310 i=1,ni
         write (02,9999) (yhat1(i,j),j=1,nj)
310       continue
         do 320 i=1,ni
         write (02,9998) tot(i),sqrt(rowvar(i)),totmax(i)
320       continue
            write(03,*)k1-1
            do 350 i=1,ni
            do 350 j=1,min(nj,ni-i+1)
            write(03,*) yfit(i,j),resid(i,j)
350          continue
       do 355 i=ni-nj+2,ni
       gtot=gtot+tot(i)
       vgtot=vgtot+rowvar(i)
355    continue
       do 400 i=ni-nj+2,ni-1
       do 400 j=ni-i+2,nj
       do 400 l=i+1,ni
       do 400 k=ni-l+2,nj
       vgtot=vgtot+2*yhat(i,j)*yhat(l,k)*(exp(v(i,j,l,k))-1)
400    continue
       write (02,9991) gtot,sqrt(vgtot)
9991   format(1x,2(f15.0,1x))
9997       format(1x,3(f6.3,3x))
9998       format(1x,3(f12.0,1x))
9999       format(1x,10(f10.0,1x))
```

```
      stop
      end
      subroutine prod1(a1,a2,a3,n1,n2,n3)
      double precision a1(150,150),a2(150,150),a3(150,150)
      do 2 i=1,n2
      do 2 j=1,n3
      a3(i,j)=0.0
      do 1 k=1,n1
      a3(i,j)=a3(i,j)+a1(k,i)*a2(k,j)
    1 continue
    2 continue
      return
      end
      subroutine prod2(a1,a2,a3,n1,n2,n3)
      double precision a1(150,150),a2(150,150),a3(150,150)
      do 2 i=1,n1
      do 2 j=1,n3
      a3(i,j)=0.0
      do 1 k=1,n2
      a3(i,j)=a3(i,j)+a1(i,k)*a2(k,j)
    1 continue
    2 continue
      return
      end
      subroutine square(a1,b1,n1)
      double precision a1(150,150),b1(150,150)
      do 10 i=1,n1
      do 8 j=1,i
      b1(i,j)=a1(i+1,j)
    8 continue
   10 continue
      do 20 i=1,n1-1
      do 18 j=i+1,n1
      b1(i,j)=b1(j,i)
   18 continue
   20 continue
      return
      end
```

References

J.Aitchison and J.A.C.Brown (1969) *The Log Normal Distribution*. Cambridge University Press.

H.Akaike (1973) *Information Theory and an extension of the Maximum Likelihood Principle.* In 2nd. Int. Symposium on Inference Theory (eds.B.N.Petrov and F.Csali), pp267-81, Akademaii-Kiado, Budapest.

H.Akaike (1976) *Canonical Correlation Analysos of Time Series and the Use of an Information Criterion.* In system Identification: Advances and Case Studies (eds. R.K.Mehra and D.G.Lainotis),pp27-96 Academic Press, New York.

R.J.Baker and J.A.Nelder (1978) *The GLIM System*, Release 3, Numerical Algorithms Group, Oxford.

M.S.Bartlett (1968) *A Further Note on Nearest Neighbour Models.* JRSS, Series A, Vol.131, pp 579-580

M.S.Bartlett and J.E.Besag (1969) *Correlation Properties of some Nearest-Neighbour Models.* Bull. Int. Statis. Inst., Vol. 43, Book 2, pp 191-193.

B.Benjamin (1977) *General Insurance*, Heinemann, London.

J.E.Besag (1972) *On the Correlation Structure of some Two-Dimensional Stationary Processes.* Biometrika, Vol. 59, pp 43-48.

G.E.P.Box and G.M.Jenkins (1976) *Time Series Analysis: Forecasting and Control.* San Francisco, Holden Day.

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.

Claims Reserving Manual, Vol.2 (ed. S.Benjamin and R.Verrall) Institute of Actuaries (1989).

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.

E.Fancusini and G.Marchesini (1976) *State Space Realisation Theory of 2-D Filters.* IEEE trans. on Automatic Control, Vol.AC-21, pp484-492.

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

D.D.Givone and R.P.Roesser (1972) *Multidimensional Linear Iterative Circuits — General Properties.* IEEE Transactions on Computers, Vol. c-21, No.10, pp 1067-1073.

D.D.Givone and R.P.Roesser (1973) *Minimisation Of Multidimensional Linear Iterative Circuits.* IEEE Transactions on Computers, Vol. c-22, No.7, pp 673-677.

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

P.J.Harrison and C.F.Stevens (1976) *Bayesian Forecasting (with Discussion)*, JRSS, Series B, Vol. 38, No. 3, pp 205-247.

A.C.Harvey and C.Fernandes (1989) *Time Series Models for Random Sums.* Applications of Mathematics in Insurance, Finance and Accounting Work; joint meeting of the Institute and Faculty of Actuaries and the Institute of Mathematics and its Applications.

I.Hossack, J.H.Pollard and B.Zehnwirth (1983) *Introductory Statistics with Applications in General Insurance*, Cambridge University Press.

Institute of Actuaries Working Party on Claims Reserving (1989).

W.S.Jewell (1975) *The Use of Collateral Data in Credibility Theory: a Hierarchical Model* RM 75-24, International Institute for Applied Systems Analysis, Scloss Laxenberg, Austria; also published in Girnale dell'Instituto Italiano degli Attuari, Vol. 38, pp1-16.

P.de Jong & B.Zehnwirth (1983) *Claims Reserving, State-space Models and The Kalman Filter* J.I.A. Vol. 110 Part 1 pp 157-182.

R.E.Kalman(1963) *New Methods in Wiener Filtering Theory.* In proceedings of the 1st Symposium on Engineering Applications of Random Function Theory and Probability (eds. J.L.Bagdanoff and F.Kozin) New York, Wiley.

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.

S.Kung, B.Levy, M.Morf and T.Kailath (1977) *New Results in 2-D System Theory, Part II*. Proc. IEEE, Vol.65, No.6, pp 945-961.

J.Lemaire (1982) *Claims Reserving in Liability Insurance*, Journal of Forecasting, Vol.1, pp 303-318.

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.

G.E.Lyons (1984) *Outstanding Claims Reserves.* Presented to the Actuaries Students' Society.

C.Osmond and M.J.Gardner (1982) *Age, Period and Cohort Models Applied to Cancer Mortality Rates.* Statistics in Medicine, Vol. 1, pp 245-259.

C.Osmond, M.J.Gardner and E.D.Acheson (1982) *Analysis of Trends in Cancer Mortality in England and Wales during* 1951-80 *Separating Changes Associated with Period of Birth and Period of Death.* British Medical Journal, Vol.284, pp1005-1008.

R.Pater (1989) *The Run-off-Triangle: Least Squares — Against Chainladder Estimations.*

W.A.Porter and J.L.Aravena (1986) *State Estimation in Discrete m-D Systems* IEEE trans. on Automatic Control, Vol. AC-31, No.3, pp280-3.

R.Raeside (1988) *The Use of Sigmoids in Modelling and Forecasting Human Populations* JRSS, Series A, Vol. 151, Part 3, pp 499-513

D.H.Reid (1978) *Claim Reserves in General Insurance* JIA, Vol.105, part 3, pp 211-296.

A.Renshaw (1989) *Chain Ladder and Interactive Modelling* Applications of Mathematics in Insurance, Finance and Accounting Work; joint meeting of the Institute and Faculty of Actuaries and the Institute of Mathematics and its Applications. To appear in the Journal of the Institute of Actuaries.

A.Renshaw and R.Verrall (1989) *Claims Reserving: Statistical Treatment of the Chain Ladder Technique.* Notes for a Short Course at the City University.

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

J.Q.Smith (1983) *Forecasting Accident Claims for an Assurance Company.* The Statistician, Vol.32, pp 109-115.

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

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

R.J.Verrall (1989b) *On the Unbiased Estimation of Reserves from Loglinear Models* Under review.

R.J.Verrall (1989c) *Chain Ladder and Maximum Likelihood* Under review.

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

S.D.Walter, C.T.Miller and J.A.H.Lee (1976) *The Use of Age-Specific Mean Cohort Slopes in the Analysis of Epidemiological Incidence and Mortality Data.* JRSS, Series A, Vol.139, Part 2, pp 227-245.

P.Whittle (1954) *On Stationary Processes in the Plane.* Biometrika, Vol. 41, pp 434-449

S.L.Zeger, L.-C.See and P.J.Diggle (1989) *Statistical Methods for Monitoring the Aids Epidemic* Statistics in Medicine, Vol. 8, pp 3-21.

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