JIA 117 (1990) 677-731

A STOCHASTIC METHOD FOR CLAIMS RESERVING IN GENERAL INSURANCE

By T. S. Wright, M.A., F.S.S., M.I.S.

ABSTRACT

The paper addresses the problem of estimating future claim payments from the 'run-off' of past claim payments. A model of the claim payment process is postulated. Results from risk theory are applied to give a model for the incremental paid claims data by development period. A fitting method is developed which takes account of the error structure of the data implied by the underlying model of the claim payment process. The application of a similar method to incremental incurred data is considered. A numerical example is given.

KEYWORDS

Claims; General Insurance; Kalman filter; Reserves; Quasi-likelihood

1. INTRODUCTION

The problem addressed in this paper is the estimation of total claim payments to be made in the future under insurance contracts written in the past. Usually, an estimate is required for each past *year of account*. Reserves are set on the basis of such estimation. A stochastic claims reserving method should give both a *best estimate* and a *standard error*: reserves can then be set at any desired degree of prudence by adding a multiple of the standard error to the best estimate. A *year of account* or *origin year* can be either:

(i) an *underwriting year*; that is, the calendar year in which cover commenced, or
(ii) an *accident year*; that is, the calendar year in which the event giving rise to a claim occurred, or
(iii) a *reporting year*, that is, the calendar year in which the event giving rise to a claim was reported.

Whichever definition of origin year is used, claim payments relating to a particular origin year are usually classified according to the delay between the beginning of the origin year and the time of payment. The classification is into discrete intervals such as quarters, half-years, or years; in this paper, *development period* will be used as a generic term. P will denote the number of development periods per year, thus $P=4, 2, 1$, according to whether development periods are quarters, half-years or years. If W represents the origin year ($W=1, 2, \ldots M$) and D the development period ($D=0, 1, 2 \ldots T-1$), data exist for each (W, D) combination satisfying:

$$[D+1+(W-1)\cdot P \leqslant T],$$

where T is the number of complete periods since the beginning of the first year of origin. Other (W, D) combinations correspond to the future.

This paper addresses the usual situation in which the data consist primarily of a single figure for each past (W, D) combination, being the total of all claim payments. This will be denoted Y_{WD}. It is conventional to tabulate such data with one row for each year of origin as illustrated below:

$$
\begin{array}{llllllllllll}
Y_{10} & Y_{11} & . & . & . & . & . & . & . & . & . & Y_{1,T-1} \\
Y_{20} & Y_{21} & . & . & . & . & . & . & . & Y_{2,T-1-P} & & \\
. & . & . & . & . & . & . & . & . & & & \\
. & . & . & . & . & . & . & . & & & & \\
Y_{M-1,0} & Y_{M-1,1} & & & & & & & & & & \\
Y_{M,0} & & & & & & & & & & &
\end{array}
$$

This arrangement is usually referred to as a *run-off triangle*, or *development triangle*. In practice, it is common for some values to be missing, for example:

—the triangle may be truncated at the right, due to a belief that claim payments are negligible beyond a certain development period,
—the top left corner may be missing due to data not having been collated prior to a certain calendar year,
—there may be individual figures missing in the body of the array due to various adverse circumstances,
—the bottom corner of the triangle may be truncated due to termination of business in the class concerned (strictly speaking, this is not missing data).

The method described in this paper does not require any data other than the totals Y_{WD} of paid claims, among which there may be any arrangement of missing values. The only constraints are that there must be some data for at least three different development periods (i.e. three different values of D) and the total number of data-points must not be very small, less than about a dozen. However, estimation is much improved if there is also some measure of *exposure* for each origin year, such as the number of claims reported by the end of development period zero. Similarly, it is desirable, but not essential, to have some prior information on the rate of inflation in claim payments.

In addition to the paid claims triangle, there is frequently a triangle of *incurred* data available. The *cumulative incurred* for a particular origin year is defined to be the total of sums paid to date and estimates of amounts that will be paid on claims which have been reported but not yet settled; that is, *cumulative incurred* is *cumulative paid*, i.e.

$$\sum_{d=0}^{D} Y_{Wd}$$

plus the total of *case estimates* on *claims outstanding*. In this paper, the notation I_{WD} is used to denote the increments, by development period, of this quantity.

These data are often very valuable for reserving purposes, particularly if the *accidents* tend to be reported relatively quickly and the main delay is between reporting and settlement. An extreme case of this is when the origin years are reporting years. However, the value of incurred data clearly depends on the reliability of case estimates, and can approach zero. Because of the great variability in procedures for producing case estimates between different classes of business and insurance companies, it is difficult to formulate a generally applicable stochastic method which takes these data into account. One approach is described in this paper. It is rather more approximate than the method for paid data alone, but it has been found useful on occasions.

The methods described in this paper exist as a computer program and have been used in practice many times.

2. MOTIVATION AND SUMMARY

The application of statistical methods to claims reserving has been characterised in the past by an empirical approach, in which the model used for prediction is determined to a great extent, in each particular application, by the data. Obviously, the data must be used to calibrate the model; the issue raised here is the extent to which this approach is taken.

Recall that a statistical model has a systematic component and a random component; if Z denotes the data (or some transformation thereof), a statistical model can invariably be expressed in the form: $Z = M + E$, where E is a random term with expected value zero, and both the expected value, M, and the variance, $\mathrm{Var}(E)$, of the data (and possibly higher moments) depend on unknown parameters of the model. M is called the *systematic component* of the model, and E (specified by its variance and possibly higher moments) is known as the *random component*. In general, Z, M and E are vectors.

Previously proposed statistical methods have relied on separate calibration by the data of both the systematic and random components of a family of models. The view taken here is that this approach is dangerous when the number of data-points is as small as in the typical run-off triangle (there are typically 20 to 100 data-points). The danger is not just that the parameters of such a model may not all be reliably estimated (this can be quantified), but that there may be too few data-points to indicate that the assumed form of the random component is not valid. The validity of the random component of a model is important, primarily, because the random component dictates how the systematic component should be fitted (i.e. calibrated). The author believes that most previously published statistical claims reserving models have a random component which is unlikely to be valid (the reasons for this belief are given briefly in Appendix 4), and that the invalidity will often not be apparent from post-fit diagnostic checks, because of the small number of data-points involved. In diagnostic tests (and statistical tests generally) a hypothesis is rejected only if there is very clear evidence against it;

such evidence is unlikely to be present with only a few data-points. In technical terms, the *power* of a test generally decreases with sample size. These comments apply equally to formal significance tests and informal procedures, such as the visual examination of residual plots. In either case, the *power* is the probability of realising that the model is not valid when this is actually the case.

The approach adopted here is less empirical in the sense that, rather than beginning by looking at data, the process which gives rise to the data is considered first. A model of the claim payment process is proposed in Section 3. This model dictates the form of the model for the actual data, the incremental paid claims Y_{WD}. In Section 4 it is shown how both the systematic and the random components of the actual data are determined by the assumed model of the underlying claim payment process. This ensures consistency in the model used for the data; whenever the assumptions of the underlying claim payment model are appropriate, then the model for the data is valid in both its systematic and random components. There is no need for a separate validation of the random component, so the problem of having too few data-points to do this reliably does not arise.

Having obtained a complete model for the data, Y_{WD}, from the assumed model of the generating process, a fitting method is developed (Section 5). This method makes use of Quasi-likelihood estimation and the Kalman filter. Several mathematical approximations and asymptotic results are used in developing the fitting method, so it is not obvious how near-optimal the final method is, or whether the standard errors associated with the parameter estimates are valid. To answer these questions, simulations have been carried out in which the true values of the parameters are known and can be compared with the estimates. The results of simulations carried out so far have been favourable. Further details and results of the simulations are not given in this paper.

The final component of the method for paid data is the prediction of future claim payments and standard errors, using the calibrated model. This is dealt with in Section 6.

The same method cannot validly be applied directly to incremental incurred data, I_{WD}, because the model of the generating process which underpins the method is not appropriate in this case. However, if some adjustments are made to the incurred data, then the underlying model can be a good approximation in some cases, so the method can be applied to the adjusted incurred data. The question then arises of how to combine the predictions obtained using incurred data with those obtained using the paid data. This matter is addressed in Section 7.

Sections 3 to 7 are mainly mathematical, dealing with the detailed derivation of the method. An example application to real-life data is given in Section 8. This is followed, in Section 9, by a discussion of some of the issues raised in other sections. Some readers may find the paper more accessible by reading these last two sections before, or in parallel with, Sections 3 to 7.

3. BASIC MODEL OF THE CLAIM PAYMENT PROCESS

3.1 *Delay to Payment (Single Origin Year)*

The paid claims data for a single origin year are made up of a number of individual claim payments. For each of these payments there is a delay between the beginning of the origin year and the payment. These delays are modelled as independent identically distributed (iid) random variables. Thus we can define:

p_D = probability that payment is made in development period D,

where $$\sum_{D=0}^{\infty} p_D = 1.$$

The total number of claim payments for a particular origin year is modelled as a Poisson variable with expected value ε. (ε is referred to in this paper as the *exposure*, and is given by:

ε = (number of units of risk) × (expected number of claims per unit risk) × (expected number of payments per claim)

for classes of business where these terms are meaningful. Note that this usage is not strictly conventional: *exposure* usually means the number of units of risk.)

If N_D = number of payments in development period D, then N_D is also Poisson, with:

$$V(N_D) = E(N_D) = \varepsilon \cdot p_D$$

and the N_D are mutually independent. (Throughout this paper $V(\ )$ is used to indicate the variance of a random variable, and $E(\ )$ the expected value.)

Further modelling assumptions, on the form of the distribution p_D, are made in Section 4.5.

3.2 *Size of Payments*

For most classes of general insurance business the mean severity (after allowing for inflation) is not the same for all development periods. For example, large claims might take longer to settle, on average, than smaller claims: in this case the mean severity increases with delay. The magnitude of a payment is, therefore, modelled as a random variable with mean and variance depending on delay. If X_d represents an inflation adjusted payment made at delay d (where d represents continuous time) the model is:

(i) $E(X_d) = k \cdot d^{\lambda}$ for some constants k and λ, and

(ii) the coefficient of variation of X_d does not depend on d, i.e. $V(X_d) = \rho^2 \cdot E(X_d)^2$ for some constant ρ.

Despite its simplicity, part (i) of this model allows for a wide range of possibilities; the variation of mean severity with delay implied by various values of λ is shown in Figure 3.1. Note that these plots show the *mean severity*, $E(X_d)$.

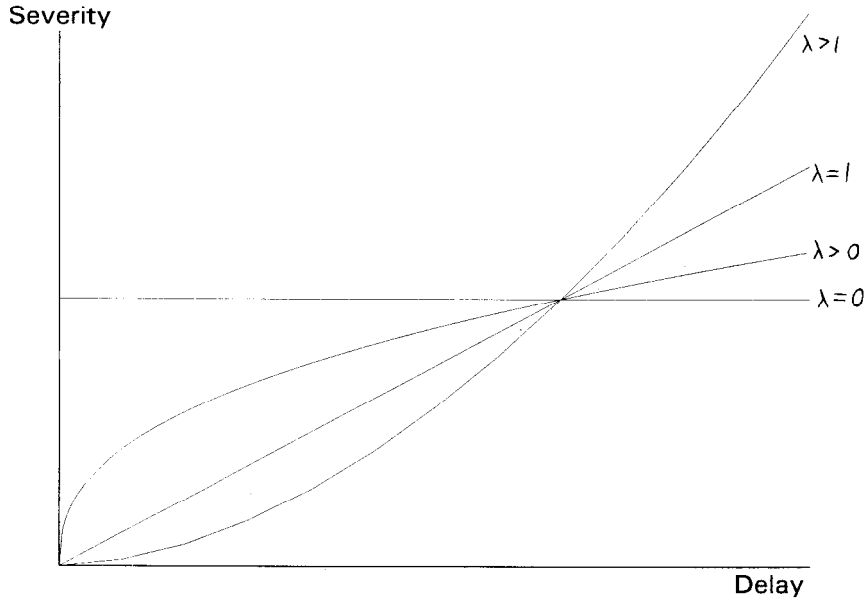


Figure 3.1.

The severity of an individual claim, X_d, is a random variable which (in the absence of reinsurance) can take any value in the range $(0, \infty)$, regardless of the delay, d. Part (ii) of this model is simply an assumption that the percentage variation in individual claim payments is the same for all delays.

Moving to discrete time (delay, d, is measured by development period $D=0, 1, 2 \ldots$), this model is approximated by:

$$\text{(i)} \qquad E(X_D)=k\cdot D'^{\lambda} \tag{1}$$

$$\text{(ii)} \qquad V(X_D)=\rho^2\cdot E(X_D)^2$$

where D' is a function of D, given in Appendix 1.

Individual claim severities are assumed to be mutually independent. Note that no assumption is made about the actual distribution other than that the variance exists.

4. MODEL OF INCREMENTAL PAID DATA

4.1 *Single Origin Year*

Letting Y_D = inflation adjusted incremental paid claims, note that Y_D is the sum of N_D iid payments, X_D. Applying standard results from risk theory we have:

$$E(Y_D) = E(N_D) \cdot E(X_D)$$

$$V(Y_D) = E(N_D) \cdot E(X_D^2) = E(N_D) \cdot [V(X_D) + E(X_D)^2].$$

Hence $$E(Y_D) = \varepsilon \cdot p_D \cdot k \cdot D'^{\lambda}$$

$$V(Y) = \varepsilon \cdot p_D(1 + \rho^2) \cdot k^2 \cdot D'^{2\lambda}$$

and the Y_D are independent for different D. This follows from the mutual independence of the severities, X_D, and the claim numbers, N_D.

4.2 *Several Origin Years*

The entire run-off array is now considered, by introducing subscript W to represent origin year; $W = 1, 2, \ldots$

Y_{WD} is the inflation corrected increment for development period D of origin year W.

The model described above is assumed to hold for each origin year, with k, λ and ρ being the same for all origin years. Thus we have:

$$E(Y_{WD}) = \varepsilon_W \cdot p_{WD} \cdot k \cdot D'^{\lambda}$$

$$V(Y_{WD}) = \varepsilon_W \cdot p_{WD} \cdot (1 + \rho^2) \cdot k^2 \cdot D'^{2\lambda}.$$

4.3 *Including Inflation*

Let ι_T be the average force of claims inflation per annum between payment periods $T-1$ and T. ($T = D + P \cdot W$, where P is the number of periods per annum, usually 1, 2 or 4.)

Equation (1) becomes:

$$E(X_{WD}) = E^{\delta_T} \cdot k \cdot D'^{\lambda}$$

where $$\delta_T = \frac{1}{P} \sum_{t=1}^{T} \iota_t$$

(i.e. e^{δ_T} is the inflation factor from period 0 to period T) so the model for data Y_{WD}, not adjusted for inflation, is:

$$E(Y_{WD}) = \varepsilon_W \cdot p_{WD} \cdot e^{\delta_T} \cdot k \cdot D'^{\lambda}$$

$$V(Y_{WD}) = \varepsilon_W \cdot p_{WD} \cdot (1 + \rho^2) \cdot e^{2\delta_T} \cdot k^2 \cdot D'^{2\lambda}.$$

Writing μ_{WD} for $\varepsilon_W \cdot p_{WD} \cdot e^{\delta_T} \cdot kD'^{\lambda}$ and ϕ_0 for $k(1 + \rho^2)$, we have:

$$E(Y_{WD}) = \mu_{WD}$$

$$V(Y_{WD}) = \phi_0 \cdot e^{\delta_T} \cdot D'^{\lambda} \cdot \mu_{WD}.$$

By earlier assumptions, ϕ_0 is the same for all origin years; it is referred to as the *scale parameter*.

4.4 *Adjusting for Exposure*

Suppose exposure is estimated as $\tilde{\varepsilon}_w$, and define ε'_W by $\varepsilon_w = \tilde{\varepsilon}_W \cdot \varepsilon'_W$ (so ε'_W represents the error in the estimate $\tilde{\varepsilon}_W$).

The *normalised data* are defined by:

$$Y'_{WD} = \frac{Y_{WD}}{\tilde{\varepsilon}_W}.$$

If μ'_{WD} is similarly defined by:

$$\mu'_{WD} = \frac{\mu_{WD}}{\tilde{\varepsilon}_W}$$

then the model becomes:

$$E(Y'_{WD}) = \mu'_{WD}$$

$$V(Y'_{WD}) = \frac{\phi_0}{\tilde{\varepsilon}_W} \cdot e^{\delta_T} \cdot D'^{\lambda} \cdot \mu'_{WD}$$

where

$$\mu'_{WD} = \varepsilon'_W \cdot p_{WD} \cdot e^{\delta_T} \cdot k \cdot D'^{\lambda}.$$

4.5 *Distribution of Payments over Development Periods*

It is convenient to complete the model of the underlying claim payment process of Section 3 at this point, by introducing an assumption about the form of the distribution, p_{WD}, for each origin year, W.

4.5.1 *Stochastic Chain Ladder Model*

One possibility is to make no assumption about the form of this distribution, other than that it is the same for all origin years. Thus we can write $p_{WD} = p_D$.

Note that μ'_{WD} has one factor e^{δ_T} depending on T. If the factors involving D are collected together as e^{β_D} and the remaining factors are collected together as e^{α_W}, we have:

$$\mu'_{WD} = \exp(\alpha_W + \beta_D + \delta_T)$$

where

$$\alpha_W = \ln(\varepsilon'_W \cdot k)$$

$$\beta_D = \ln(p_D \cdot D'^{\lambda}).$$

In this form, it is clear that the systematic part of the model (i.e. the non-random part) is essentially the same as that assumed in the chain ladder method: if the data were corrected for inflation, making the factor $\exp(\delta_T)$ redundant, we would have an *origin year factor*, $\exp(\alpha_W)$, and a *development factor*, $\exp(\beta_D)$. The development factors, $\exp(\beta_D)$, are completely unconstrained by the model, because of the lack of any assumption for the distribution, p_D; the presence of the factor D'^{λ} does not impose any constraint on the βs.

Note that the mean of the data, μ'_{WD}, is a function of a linear combination of unknown parameters α_W, β_D, δ_T.

Also, the variance of the data can be written:

$$V(Y'_{WD}) \propto \frac{D'^{\lambda}}{\tilde{\varepsilon}_W} \cdot \exp(\alpha_W + \beta_D + 2\delta_T).$$

Thus, the relative variance is a known function of the same parameters α_W, β_D, δ_T. (If λ is not known, it can be estimated iteratively as described in Section 5.6).

Such a model is known as a *generalised linear model*. Maximum (quasi-) likelihood estimates of the parameters can be obtained for such models using *Fisher's scoring method* (Appendix 2). Predictions and their standard errors can then be obtained in a manner similar to that described in Section 6.

The chain ladder model is not pursued further in this paper, because it suffers from two major limitations (these apply equally whether it is fitted using statistical techniques or in the traditional way):

(i) Estimates cannot be produced beyond a stage of development which has already been observed for at least one year of origin. This is clearly unsatisfactory when the oldest year of origin is not believed to have reached full development.

(ii) The model assumes that the systematic run-off pattern has been the same for all years of origin. This is rarely a plausible assumption in practice.

4.5.2 *The Hoerl Curve Model*

From Section 4.4, the model of the normalised data:

$$Y'_{WD} = \frac{Y_{WD}}{\tilde{\varepsilon}_W},$$

is

$$E(Y'_{WD}) = \mu'_{WD}$$

$$V(Y'_{WD}) = \frac{\phi_0}{\tilde{\varepsilon}_W} \cdot e^{\delta_T} \cdot D'^{\lambda} \cdot \mu'_{WD}$$

where

$$\mu'_{WD} = \varepsilon'_W \cdot p_{WD} \cdot e^{\delta_T} \cdot k \cdot D'^{\lambda}.$$

The following model for the probabilities, p_{WD}, is now proposed:

$$p_{WD} = \alpha_D \cdot \kappa_W \cdot D'^{A_W} \cdot e^{-B_W \cdot D'}$$

for some constants κ_W, A_W and B_W, where α_D and D' are functions of D, P and whether or not the origin years are underwriting years (see Appendix 1).

This model for p_{WD} arises because the delay, d, from accident to payment is likely to have approximately a Gamma distribution (because payment occurs when several successive processes have been completed and each of these

processes is likely to have approximately a negative exponential delay). That is, the pdf of the delay, d (continuous time) is:

$$f(d) = \kappa \cdot d^{A} \cdot e^{-B \cdot d}$$

for some constants κ, A, B (which may depend on W).

α_D and D' are introduced in transferring this model from continuous time to discrete time, as represented by development period, D. For example, consider development year 0 in the case of annual accident-year data. Since accidents can generally occur at any time during a year (not just on 1 January), there is, on average, only about six months available if payment is to be made in the accident year itself (i.e. development year 0), whereas later development years are a full 12 months; hence the default value of 0·5 for alpha in this case. The mean delay for payments made in development year 0 will be about four months in this case, but an approximation of the Gamma pdf by a step function suggests that 0·5 is a more appropriate value for D', given that α is to be 0·5. A complete table of α and D' values, all calculated in a consistent way, is given in Appendix 1.

It is convenient to redefine *normalised data* now as:

$$Y'_{WD} = \frac{Y_{WD}}{\tilde{\varepsilon}_W \cdot \alpha_D}.$$

The model then becomes:

$$E(Y'_{WD}) = \mu'_{WD}$$

$$V(Y'_{WD}) = \frac{\phi_0}{\tilde{\varepsilon}_W \cdot \alpha_D} \cdot e^{\delta_T} \cdot D'^{\lambda} \cdot \mu'_{WD}$$

where

$$\mu'_{WD} = \varepsilon'_W \cdot \frac{p_{WD}}{\alpha_D} \cdot e^{\delta_T} \cdot k \cdot D'^{\lambda}$$

i.e.

$$\mu'_{WD} = \varepsilon'_W \cdot \kappa_W \cdot k \cdot e^{\delta_T} \cdot D'^{(A_W + \lambda)} \cdot e^{-B_W \cdot D'}.$$

To make fitting tractable, it is necessary to assume that past claims inflation has been at a constant rate. (If this is not a plausible assumption, then the data should be adjusted to remove the non-uniform component of inflation which is supposed to exist.) If the uniform force of inflation is denoted by ι we have:

$$\delta_T = (W + D/P) \cdot \iota$$

$$\simeq (W + D'/P) \cdot \iota$$

and the model becomes:

$$V(Y'_{WD}) = \frac{\phi_0}{\tilde{\varepsilon}_W \cdot \alpha_D} \cdot e^{(W + \frac{D'}{P}) \cdot \iota} \cdot D'^{\lambda} \cdot \mu'_{WD}$$

$$\mu'_{WD} = \varepsilon'_W \cdot \kappa_W \cdot k \cdot e^{W \cdot \iota} \cdot D'^{(A_W + \lambda)} \cdot e^{-(B_W - \frac{\iota}{P}) \cdot D'}.$$

This can be written:

$$\mu'_{WD} = e^{\beta_{W1}} \cdot D'^{\beta_{W2}} \cdot e^{-\beta_{W3} \cdot D'}$$

$$V(Y'_{WD}) = \phi_W \cdot \psi_D \cdot \mu'_{WD}$$

where

$$\beta_{W1} = W \cdot \iota + \ln(\varepsilon'_W \cdot \kappa_W \cdot k)$$

$$\beta_{W2} = A_W + \lambda$$

$$\beta_{W3} = B_W - \frac{\iota}{P}$$

$$\phi_W = \frac{\phi_0 \cdot e^{W \cdot \iota}}{\tilde{\varepsilon}_W}$$

$$\psi_D = \frac{D'^{\lambda} \cdot e^{\frac{\iota}{P} \cdot D'}}{\alpha_D}.$$

5. FITTING THE MODEL

5.1 *Independent Fitting for each Origin Year*

If ϕ_W and ψ_D are known, this is a *generalised linear model* (Appendix 2) for we have:

$$Y'_{WD} = \mu'_{WD} + E_{WD}$$

where:

$$\mu'_{WD} = \exp(\boldsymbol{x}_D^T \cdot \boldsymbol{\beta}_W)$$

$$\boldsymbol{x}_D = \begin{bmatrix} 1 \\ \ln D' \\ -D' \end{bmatrix} \qquad \boldsymbol{\beta}_W = \begin{bmatrix} \beta_{W1} \\ \beta_{W2} \\ \beta_{W3} \end{bmatrix}$$

$$E(E_{WD}) = 0$$

$$V(E_{WD}) = \phi_W \cdot \psi_D \cdot \mu'_{WD}$$

and so can be fitted using Fisher's scoring method (Appendix 2) to give parameter estimates, $\hat{\boldsymbol{\beta}}_W$, and corresponding variance–covariance matrices, $\hat{V}_W$, satisfying (approximately):

$$\hat{\boldsymbol{\beta}}_W \sim N(\boldsymbol{\beta}_W, \hat{V}_W).$$

Note that, given values for ϕ_W and ψ_D, the estimates $\hat{\boldsymbol{\beta}}_W$ are mutually independent for different origin years, W.

Of course, ϕ_W and ψ_D are not known initially, but an iterative procedure can be adopted whereby initial values of ϕ_0, ι and λ (the *weight parameters*) are assumed, to give initial values to ϕ_W and ψ_D; the model is then fitted for each origin year in turn, and the results (for all origin years) used to give new estimates of the weight parameters for use in the next fit. (The post-fit estimation of the weight

parameters is detailed in Section 5.6.) This iterative procedure leads to a small measure of dependence between the estimates $\hat{\boldsymbol{\beta}}_W$ for different origin years, but this dependence is clearly very slight and is ignored in what follows.

Estimates $\hat{\boldsymbol{\beta}}_W$ can only be obtained in this way for origin years with at least 3 data-points. The treatment of origin years with less than 3 data-points is described in Section 5.3.

For some origin years with 3 or more data-points, the data may be such that estimates $\hat{\boldsymbol{\beta}}_W$ cannot be obtained by the method of scoring, due to near singularity of the information matrix (Appendix 2). A possible remedy is described in Section 5.4.

5.2 *Transmission of Information across Origin Years*

Clearly, the estimates, $\hat{\boldsymbol{\beta}}_W$, are likely to be very unreliable for origin years with few (but at least 3) data-points. However, an examination of the sources of variation in the true values, $\boldsymbol{\beta}_W$, between origin years, suggests a way of very much improving the reliability. The $\boldsymbol{\beta}_W$ are defined (Section 4.5.2) by:

$$\beta_{W1} = W \cdot \iota + \ln(\varepsilon'_W \cdot \kappa_W \cdot k)$$

$$\beta_{W2} = A_W + \lambda$$

$$\beta_{W3} = B_W - \frac{\iota}{P}$$

where κ_W, A_W, B_W are parameters of the distribution of payments over delay, and ε'_W represents the error in the exposure factors used to normalise the data.

The variation in $\boldsymbol{\beta}_W$ is modelled as follows:

$$\boldsymbol{\beta}_W = \boldsymbol{\beta}_{W-1} + \begin{bmatrix} \iota \\ 0 \\ 0 \end{bmatrix} + \boldsymbol{\omega}'_W$$

where $\boldsymbol{\omega}'_W$ is a random *perturbation*,

with $$V(\boldsymbol{\omega}'_W) = \begin{bmatrix} u_1^2 & 0 & 0 \\ 0 & u_2^2 & 0 \\ 0 & 0 & u_3^2 \end{bmatrix}.$$

(This variance is assumed to be the same for all pairs of origin years; there is usually no reason for expecting it to vary.)

If $\boldsymbol{B}_W$ is defined by:

$$\boldsymbol{B}_W = \begin{bmatrix} \iota \\ \beta_{W1} \\ \beta_{W2} \\ \beta_{W3} \end{bmatrix}$$

we have:

$$\boldsymbol{B}_W = \underline{\boldsymbol{G}}_W \cdot \boldsymbol{B}_{W-1} + \boldsymbol{\omega}_W \qquad (2)$$

where $\underline{G}_W = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ $V(\omega_W) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & u_1^2 & 0 & 0 \\ 0 & 0 & u_2^2 & 0 \\ 0 & 0 & 0 & u_3^2 \end{bmatrix}$.

This can be regarded as the *system equation* of a dynamic linear model (see Appendix 3). The u_i^2 are called *system variances* or *adaptive variances* in what follows.

For years of origin with sufficient data, the estimates, $\hat{\boldsymbol{\beta}}_W$ (obtained by the method of scoring as described in Section 5.1) are regarded as the *observations*, so the *observation equation* is:

$$\hat{\boldsymbol{\beta}}_W = \underline{X}_W \cdot \boldsymbol{B}_W + \boldsymbol{\varepsilon}_W \tag{3}$$

where $\underline{X}_W = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ $V(\boldsymbol{\varepsilon}_W) = \hat{V}_W$.

Equations (2) and (3) specify a dynamic linear model. Therefore, the Kalman filter (Appendix 3) will extract maximum information from the *observations* ($\hat{\boldsymbol{\beta}}_W$, $\hat{V}_W$) to give optimal estimates of the parameters, $\boldsymbol{B}_W$, taking into account the relationship between different origin years, as described by equation (2). However, before applying the Kalman filter it is necessary to:

—include the data for origin years for which independent estimates $\hat{\boldsymbol{\beta}}_W$ cannot be obtained, and
—assign realistic values to the *adaptive variances*, u_i^2.

The first of these matters is dealt with in Section 5.3, the second in Section 5.5.

5.3 *Treatment of Origin Years with only 1 or 2 Data-Points*

For years of origin with insufficient data to fit the model independently (by putting the data for that origin year alone into the scoring method to obtain estimates, $\hat{\boldsymbol{\beta}}_W$) an alternative method of obtaining a linear *observation equation* of the same form as equation (3):

$$\boldsymbol{Y}_W^* = \underline{X}_W \cdot \boldsymbol{B}_W + \varepsilon_W \qquad \varepsilon_W \sim N(0, V_W)$$

(with Y_W^*, $\underline{X}_W$ and V_W all known) is required.

We have:

$$E(Y'_{WD}) = \mu'_{WD}$$
$$V(Y'_{WD}) = \phi_W \cdot \psi_D \cdot \mu'_{WD}.$$

(The quantities on the right are defined in Section 4.5.2.)

The coefficient of variation γ_{WD} of Y'_{WD} is given by:

$$\gamma_{WD}^2 = \frac{V(Y'_{WD})}{\quad}.$$

Hence: $$\gamma^2_{WD} = \frac{\phi_W \cdot \psi_D}{\mu'_{WD}}$$

but $$\phi_W \cdot \psi_D = \frac{\phi_0}{\tilde{\varepsilon}_W \cdot \alpha_D} \cdot e^{\delta T} \cdot D'^{\lambda}$$

and $$\mu'_{WD} = \frac{\varepsilon'_W}{\alpha_D} \cdot p_{WD} \cdot e^{\delta T} \cdot k \cdot D'^{\lambda}$$

hence: $$\gamma^2_{WD} = \frac{\phi_0}{k \cdot \varepsilon_W \cdot p_{WD}} = \frac{(1+\rho^2)}{E(N_{WD})}$$

from which it is seen that the coefficient of variation of the normalised data is invariably much less than 1 for early development periods.

Now consider the random variable, F_{WD}, defined by:

$$F_{WD} = \frac{Y'_{WD}}{\mu'_{WD}}.$$

This has:

$$E(F_{WD}) = 1 \qquad \text{and} \qquad V(F_{WD}) = \gamma^2_{WD}.$$

Also, we have:

$$\ln Y'_{WD} = \ln \mu'_{WD} + \ln F_{WD}$$

which can be written:

$$\ln Y'_{WD} = \underline{X}_W \cdot \boldsymbol{B}_W + \ln F_W \tag{3a}$$

where $$\underline{X}_W = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & \ln D' & -D' \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

If the distribution of Y'_{WD} is approximated by a log-Normal distribution (n.b. this is only for those origin years with only 1 or 2 data-points), we have:

$$\ln F_{WD} \sim N(v_{WD}, \tau^2_{WD}) \text{ for some } v \text{ and } \tau^2$$

but: $$\left.\begin{aligned} E(F_{WD}) &= 1 \\ V(F_{WD}) &= \gamma^2_{WD} \end{aligned}\right\} \Rightarrow \begin{cases} v_{WD} = -\frac{1}{2}\gamma^2_{WD} & (4) \\ \tau^2_{WD} = \ln(1+\gamma^2_{WD}). & (5) \end{cases}$$

We invariably have $\gamma_{WD} \ll 1$ for the values of D concerned here (i.e. $D=0, 1$), so equation (5) implies:

$$\tau_{WD} \ll 1$$

and with equation (4) this implies:

$$|v_{WD}| \ll |\tau_{WD}|.$$

Hence: $$\ln F_{WD} \sim N(0, \tau^2_{WD})$$

to a good approximation, and equation (3a) is of the required form: it will be regarded as the *observation equation* for those origin years with too few points to allow equation (3) to be used. The var-covariance matrix of the error term, $\ln \boldsymbol{F}_W$, is the diagonal matrix of τ^2 values which can be calculated from:

$$\tau^2_{WD} = \ln\left(1 + \frac{\phi_w \cdot \psi_D}{\mu'_{WD}}\right)$$

using parameter estimates $\hat{\beta}_{W-1}$ from the previous origin year to give an estimate of μ'_{WD}.

5.4 *Treatment of Origin Years for which Method of Scoring Fails*

If the data for a certain origin year are such that the model (Section 5.1) cannot be fitted by the method of scoring, then, rather than discarding the data for that origin year, it will often be preferable to assume that the underlying run-off pattern is the same as for an adjacent year (or years) and to apply the method of scoring to several years of origin simultaneously.

For example, suppose W labels an origin year for which fitting fails, then it may be reasonable to combine this with year $W+1$ by assuming:

$$\beta_{w+1,1} = \beta_{W,1} + \iota$$
$$\beta_{w+1,2} = \beta_{W,2}$$
$$\beta_{w+1,3} = \beta_{W,3}.$$

In the notation of Section 5.2, this is:

$$\boldsymbol{B}_{W+1} = \underline{\boldsymbol{G}}_W \cdot \boldsymbol{B}_W.$$

The assumption is that the adaptive variances between years W and $W+1$ are all zero.

The model of Section 5.1 can be rewritten for these two years as:

$$\begin{bmatrix} Y'_{WD} \\ Y'_{W+1,D} \end{bmatrix} = \exp(\underline{\boldsymbol{X}} \cdot \boldsymbol{B}_W) + \text{error}$$

where $$\underline{\boldsymbol{X}} = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & \ln D' & -D' \\ \vdots & \vdots & \vdots & \vdots \\ \hline \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \ln D' & -D' \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$ and $V(\text{error}) = \text{as before.}$

Fitting this by the method of scoring gives the *observation equation*:

$$\hat{\boldsymbol{B}}_W = X_W \cdot B_W + \varepsilon_W$$

where X_W is now the identity matrix and an estimate, $\hat{V}_W$, of $V(\varepsilon_W)$ is given by the method of scoring, as before.

This takes the place of equation (3) for origin year W, and there is clearly no separate observation equation for year $W+1$; both origin years are dealt with by a single call of the Kalman filter.

5.5 *Estimation of Adaptive Variances*

In order to apply the Kalman filter to the dynamic linear model specified by equations (2), (3) and (3a) (Sections 5.2 and 5.3), values must be assigned to the u_i^2 of equation (2). These quantities represent the variance of the random perturbations in the run-off parameters, β_{Wi}, from one origin year to the next.

By fitting Hoerl curves individually for each origin year with sufficient data (as described in Section 5.1), we have independent estimates $\hat{\boldsymbol{\beta}}_W$ and $\hat{V}_W$ satisfying (approximately):

$$\hat{\boldsymbol{\beta}}_W \sim N(\boldsymbol{\beta}_W, \hat{V}_W)$$

(the approximation being better the greater the number of data-points). The variances, u_i^2, can be estimated from the observed variation in these independent estimates, $\hat{\beta}_{Wi}$. However, it is necessary to take account of the observation errors, represented by the $\hat{V}_W$: we are interested in the variance of differences in the true (unknown) run-off parameters $\boldsymbol{\beta}_W$, not the estimates $\hat{\boldsymbol{\beta}}_W$. Intuitively, we wish to look at the estimates $\hat{\boldsymbol{\beta}}_W$, and to 'see through' the error in these estimates, in order to get at the amount of variation in the true values represented by these estimates. This is done as follows:

Consider $i=2$ or 3 initially.

We have: $\beta_{Wi} = \beta_{W-1,i} + \omega_{Wi} \qquad \omega_{Wi} \sim N(0, u_i^2)$

and $\hat{\beta}_{Wi} = \beta_{Wi} + \varepsilon_{Wi} \qquad \varepsilon_{Wi} \sim N(0, \sigma_{Wi}^2)$

where σ_{Wi}^2 is from the leading diagonal of $\hat{V}_W$.

If Δ_W denotes $\hat{\beta}_{Wi} - \hat{\beta}_{W-1,i}$ for $w = 2, 3, \ldots$ then we have:

$$\begin{aligned}\Delta_W &= (\beta_{Wi} + \varepsilon_{Wi}) - (\beta_{W-1,i} + \varepsilon_{W-1,i}) \\ &= \omega_{Wi} + \varepsilon_{Wi} - \varepsilon_{W-1,i}\end{aligned}$$

hence the var-covariance matrix, Σ, of the vector $\boldsymbol{\Delta}$ is given by:

$$\Sigma = \begin{bmatrix} u^2+\sigma_1^2+\sigma_2^2 & -\sigma_2^2 & 0 & 0 \\ -\sigma_2^2 & u^2+\sigma_2^2+\sigma_3^2 & -\sigma_3^2 & 0 \\ 0 & -\sigma_3^2 & u^2+\sigma_3^2+\sigma_4^2 & -\sigma_4^2 \\ 0 & 0 & -\sigma_4^2 & \text{etc.} \end{bmatrix}$$

(subscript i has been dropped)

and we have: $$\Delta \sim N(0, \Sigma)$$

hence: $$\Delta^T \cdot \Sigma^{-1} \cdot \Delta \sim \chi^2_n.$$

(χ^2_n denotes the chi-squared distribution with n degrees of freedom, where n is the dimensionality of Δ, i.e. one less than the number of origin years for which curve-fitting has been carried out.)

The expected value of χ^2_n is n, so an estimate of u^2 is given by solving the equation:

$$\Delta^T \cdot \Sigma^{-1} \cdot \Delta = n \qquad \text{for } u^2.$$

This is a high-order polynomial equation; it is solved numerically. If the observation variances, σ^2_W, are so large compared to the differences, Δ_W, that $\Delta^T \cdot \Sigma^{-1} \cdot \Delta < n$ for all $u^2 > 0$ then it is assumed that $u^2 = 0$.

In the case $i = 1$ we have:

$$\beta_{Wi} = \beta_{W-1,i} + \iota + \omega_{Wi} \qquad \text{and} \qquad \hat{\beta}_{Wi} = \beta_{Wi} + \varepsilon_{Wi}.$$

In this case Δ_W is defined by:

$$\Delta_W = \hat{\beta}_{Wi} - (\hat{\beta}_{W-1,i} + \hat{\iota})$$

where $\hat{\iota}$ is the current estimate of ι, as used in the calculation of ϕ_W and ψ_D. Ultimately (after several iterations of the entire fitting procedure), the estimate of $\hat{\iota}$ is generally very reliable, so the estimation error in $\hat{\iota}$ is ignored here to give:

$$\Delta_W = \omega_{Wi} + \varepsilon_{Wi} - \varepsilon_{W-1,i} \qquad \text{as for } i = 2, 3.$$

Hence the same method as used to estimate u^2_2 and u^2_3 can be used for u^2_1.

5.6 *Post-Fit Estimation of the Weight Parameters (ι, ϕ_0 and λ)*

The Kalman filter gives estimates of the parameters:

$$\hat{\boldsymbol{B}} = \begin{bmatrix} \hat{\boldsymbol{B}}_1 \\ \hat{\boldsymbol{B}}_2 \\ \vdots \\ \hat{\boldsymbol{B}}_W \\ \vdots \end{bmatrix}$$

together with the variance covariance matrix, $\hat{C}$, of these estimates. ($\boldsymbol{B}_W$ is defined in Section 5.2.).

Before using these results to calculate predictions of future payments (as described in Section 6), it is necessary to test whether or not the values of the weight parameters used in the fit are plausible. Post-fit estimates of these quantities are calculated; if any of these differs significantly from the value used for the fit, then the conclusion must be that the value used was incorrect, and the

fit should be repeated using a new value. This is then repeated until none of the post-fit estimates differs significantly from the value used.

The simplest of the three weight parameters to test in this way is the force of inflation, ι. This is one of the parameters of the model; the post-fit estimate is the first element of $\hat{B}$ and the variance of this estimate is the first component of $\hat{C}$.

Post-fit estimates of the other two parameters are based on the standardised residuals, which are defined by:

$$\hat{R}_{WD} = \frac{Y'_{WD} - \hat{\mu}'_{WD}}{(\phi_W \cdot \psi_D \cdot \hat{\mu}'_{WD})^{\frac{1}{2}}}.$$

Suppose $\phi_0^{(0)}$ and $\lambda^{(0)}$ are the values used for the fit. If these are correct, we have approximately:

$$E(\hat{R}_{WD}) = 0 \qquad \text{and} \qquad V(\hat{R}_{WD}) = 1 \qquad \text{for all } W \text{ and } D.$$

For these approximations to be reasonably good, the estimates, $\hat{\mu}'_{WD}$ used in calculating the standardised residuals must be approximately unbiased.

The model (Section 5.1) has:

$$\mu'_{WD} = \exp(\boldsymbol{x}_D^T \cdot \boldsymbol{\beta}_W)$$

where:
$$\boldsymbol{x}_D = \begin{bmatrix} 1 \\ \ln D' \\ -D' \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\beta}_W = \begin{bmatrix} \beta_{W1} \\ \beta_{W2} \\ \beta_{W3} \end{bmatrix}.$$

From the Kalman filter we have:

$$\hat{\boldsymbol{\beta}}_W \sim N(\boldsymbol{\beta}_W, C_W)$$

where C_W is the appropriate sub-matrix of $\hat{C}$.

Hence:
$$\boldsymbol{x}_D^T \cdot \hat{\boldsymbol{\beta}}_W \sim N(\eta_{WD}, \sigma^2_{WD})$$

where:
$$\eta_{WD} = \ln \mu'_{WD} \qquad \sigma^2_{WD} = \boldsymbol{x}_D^T \cdot C_W \cdot \boldsymbol{x}_D.$$

So:
$$E(\exp(\boldsymbol{x}_D^T \cdot \hat{\boldsymbol{\beta}}_W)) = \exp(\eta_{WD} + \tfrac{1}{2}\sigma^2_{WD})$$

and an approximately unbiased estimate of $\mu'_{WD}(= \exp(\eta_{WD}))$ is given by:

$$\hat{\mu}'_{WD} = \exp\{\boldsymbol{x}_D^T \cdot \hat{\boldsymbol{\beta}}_W - \tfrac{1}{2} \cdot \sigma^2_{WD}\}.$$

Now suppose that the value, $\phi_0^{(0)}$, used in the fit is not correct. Since $\phi_W^{\frac{1}{2}}$ in the denominator of $\hat{R}_{WD}$ has a factor $\phi_0^{(0)\frac{1}{2}}$ in place of the correct factor $\phi_0^{\frac{1}{2}}$ we have (to a first approximation):

$$V(\hat{R}_{WD}) = \frac{\phi_0}{\phi_0^{(0)}}$$

so the post-fit estimate is given by:

$$\hat{\phi}_0 = \phi_0^{(0)} \times \frac{1}{(N-P)} \cdot \sum_{WD} \hat{R}_{WD}^2$$

(where N = number of data points, and P = number of parameters in the model).

It is possible to devise post-fit tests and estimation procedures for λ, but these will not in general be very reliable, because of the small sample size (see discussion in Section 2). It is believed that as good a check as any on the value of λ is provided by the conventional scatter-plot of standardised residuals, $\hat{R}_{WD}$, against explanatory variable, which is development period in this case. This plot is examined visually to verify that $V(\hat{R}_{WD}) = 1$ for all D. Suppose, for example, that this plot shows clear 'fanning out', that is $V(\hat{R}_{WD})$ appears to increase with D. This suggests that the factor, ψ_D, in the denominator of $\hat{R}_{WD}$, is not increasing with D as rapidly as it should, which, in turn, suggests that the current value of λ is too small. To put this another way, a larger variance in the tail of the run-off suggests that the data in the tail are made up of a smaller number of larger claims than has been assumed; claim severity appears to increase more rapidly with delay than has been assumed; the value assumed for λ is too small.

It could be argued that this is precisely the type of calibration of the random component of the model which, according to Section 2, should be avoided. However, the parameter, λ, has a concrete interpretation, and, with experience, substantial prior knowledge about its likely value for each class of business can be brought to bear, and the data in each individual analysis will have only a minimal influence (via the residual plot, as described above).

5.7 *Free-Fitting for Early Development Periods*

It often happens that the distribution of claims over development periods differs significantly from that assumed in the Hoerl curve model for p_{WD} (Section 4.5.2) when the default values for α (Appendix 1) are used. If this occurs, it is evident from the plot against development period of the residuals obtained using the Hoerl curve model. In such cases, the model can be used, but without the assumption that α_D takes the values given in Appendix 1; none of the theory in earlier sections is altered. Trials have shown that a satisfactory model is invariably obtained by allowing *free-fitting* for the first f development periods (that is, the default values are not imposed for $\alpha_0 \ldots . \alpha_{f-1}$). (By 'satisfactory model' is meant that the standardised residuals appear to be identically distributed; the data support the modelling assumptions.)

When free-fitting is selected ($f > 0$), the first f data-points for each origin year are excluded in curve fitting for each origin year independently by Fisher's scoring method. When an origin year is met for which too few points remain to allow curve fitting (i.e. less than three points) the latest excluded point (i.e. that for $D = f-1$) is reincluded, using an empirically determined α-value. α_{f-1} is estimated using the data and the fitted curves for all previous origin years, as described below. If necessary, more than one excluded point are reincluded in this way, until there are sufficient points to allow curve fitting (three points).

The *normalised data* are defined (Section 4.5.2) by:

$$Y'_{WD} = \frac{Y_{WD}}{\tilde{\varepsilon}_W \cdot \alpha_D}$$

and the model for the normalised data (Section 5.1) is:

$$Y'_{WD} = \mu'_{WD} + E_{WD}$$

where

$$E(E_{WD}) = 0$$

$$V(E_{WD}) = \phi_W \cdot \psi_D \cdot \mu'_{WD}.$$

If $\hat{\mu}'_{WD}$ is the (unbiased) estimate of μ'_{WD}, obtained from the curve fitted for year W by the method of scoring, then we have:

$$Y'_{WD} = \hat{\mu}'_{WD} + \text{random error}$$

where the random error now contains a component due to the uncertainty in the estimate, $\hat{\mu}'_{WD}$, but its expected value is still zero.

Hence:

$$Y_{WD} = \tilde{\varepsilon}_W \cdot \alpha_D \cdot \hat{\mu}'_{WD} + \text{random error.}$$

It is assumed that, to a first approximation, the random error here has variance proportional to $\tilde{\varepsilon}_W$, hence α_D can be estimated by regression through the origin of the original data, Y_{WD}, on the fitted values $\tilde{\varepsilon}_W \cdot \hat{\mu}'_{WD}$ using $1/\tilde{\varepsilon}_W$ as weights.

That is:

$$\hat{\alpha}_D = \frac{\Sigma(\mu'_{WD} \cdot Y_{WD})}{\Sigma(\tilde{\varepsilon}_W \cdot \hat{\mu}'^2_{WD})}$$

where summation is over all origin years for which curve fitting has already been carried out.

6. PREDICTIONS AND STANDARD ERRORS

6.1 *Bias and Root-Mean-Square Error*

The Kalman filter produces estimates:

$$\hat{\boldsymbol{B}} = \begin{bmatrix} \hat{\boldsymbol{B}}_i \\ \vdots \\ \hat{\boldsymbol{B}}_W \\ \vdots \end{bmatrix}$$

together with an estimated variance–covariance matrix $\hat{C}$, satisfying $\hat{B} \sim N(B, C)$, where the parameter vector $\boldsymbol{B}_W$ is defined by:

$$\boldsymbol{B}_W = \begin{bmatrix} 1 \\ \beta_{W1} \\ \beta_{W2} \\ \beta_{W3} \end{bmatrix}.$$

The model can be written in terms of B_W as:

$$Y'_{WD} = \mu'_{WD} + E_{WD}$$

where

$$E(E_{WD}) = 0$$

$$V(E_{WD}) = \phi_W \cdot \psi_D \cdot \mu'_{WD}$$

$$\mu'_{WD} = \exp(\eta_{WD})$$

$$\eta_{WD} = \boldsymbol{x}_D^T \cdot \boldsymbol{B}_W$$

$$\boldsymbol{x}_D^T = (0, 1, \ln D', -D').$$

Recall that Y'_{WD} is the incremental paid claims data, corrected for α and exposure, but still containing inflation at force ι.

At first sight it might appear that estimates of future incremental claim payments Y_{WD} can be obtained simply by using the systematic part of the model:

$$E(Y_{WD}) = \tilde{\varepsilon}_W \cdot \alpha_D \cdot \exp(\beta_{W1} + \beta_{W2} \cdot \ln D' - \beta_{W3} \cdot D')$$

with the estimates $\hat{\beta}_{W1}$, $\hat{\beta}_{W2}$ and $\hat{\beta}_{W3}$ in place of β_{W1}, β_{W2} and β_{W3}. However, this procedure could introduce bias into the estimation of future values of Y_{WD}, because if β_{W1} or β_{W2} is overestimated, the exponential ensures that $E(Y_{WD})$ is substantially overestimated, whereas if β_{W1} or β_{W2} is underestimated, the underestimation of $E(Y_{WD})$ is not so great. This works the other way around for β_{W3}.

It is shown in Sections 6.2 and 6.3 how the var-covariance matrix, $\hat{C}$, can be used in conjunction with the parameter estimates, $\hat{B}$, to produce approximately unbiased predictions of the incremental paid figures, including the effects of α and exposure, together with root-mean-square errors of these predictions.

If $\hat{\mu}_{WD}$ denotes such a prediction, unbiasedness means:

$$E(Y_{WD} - \hat{\mu}_{WD}) = 0$$

that is, the expected value of the prediction error must be zero. Since $E(Y_{WD}) = \mu_{WD}$, by definition, $\hat{\mu}_{WD}$ is an unbiased prediction if $E(\hat{\mu}_{WD}) = \mu_{WD}$.

The mean-square-error of prediction is

$$E(Y_{WD} - \hat{\mu}_{WD})^2.$$

Note that the mean-square-error of prediction can be considered as the sum of two components, the estimation variance and the process variance:

$$\begin{aligned} E(\hat{\mu}_{WD} - Y_{WD})^2 &= E((\hat{\mu}_{WD} - \mu_{WD}) - (Y_{WD} - \mu_{WD}))^2 \\ &= V(\hat{\mu}_{WD}) + V(Y_{WD}). \end{aligned}$$

This last equality holds because $\hat{\mu}_{WD}$ and Y_{WD} are mutually independent under the model assumptions; $\hat{\mu}_{WD}$ is based on past data, whereas Y_{WD} is a future datapoint.

The term *standard error* means an estimate of the standard deviation of the prediction error. It is calculated as the square root of an estimate of the mean-square-error of prediction.

6.2 *Results in Current Money Terms*

The first step in calculating a prediction $\hat{\mu}_{WD}$ is to calculate the *linear predictor* for the normalised data:

$$\hat{\eta}_{WD} = \boldsymbol{x}_D^T \cdot \hat{\boldsymbol{B}}_W.$$

Clearly, the linear predictor can be calculated for all required (W, D) combinations simultaneously, using:

$$\hat{\boldsymbol{\eta}} = X \cdot \hat{\boldsymbol{B}}$$

where X is a known matrix with one row for each (W, D) combination, e.g.

$$X = \left[\begin{array}{cccc|cccc|c} 0 & 1 & \ln 10 & -10 & & & & & \\ 0 & 1 & \ln 11 & -11 & & 0 & & & 0 \\ 0 & 1 & \ln 12 & -12 & & & & & \\ \hline & & & & 0 & 1 & \ln 8 & -8 & \\ & & & & 0 & 1 & \ln 9 & -9 & \\ & & 0 & & 0 & 1 & \ln 10 & -10 & 0 \\ & & & & 0 & 1 & \ln 11 & -11 & \\ & & & & 0 & 1 & \ln 12 & -12 & \\ \hline & & 0 & & & 0 & & & \text{etc.} \end{array}\right]$$

We have: $\hat{B} \sim N(B, C)$

hence: $\hat{\eta} \sim N(X \cdot B,\ X \cdot C \cdot X^T)$ i.e. $\hat{\eta} \sim N(\eta, \Sigma)$

where: $\Sigma = X \cdot C \cdot X^T.$

An estimate of Σ can be obtained using the Kalman filter output:

$$\hat{\Sigma} = X \cdot \hat{C} \cdot X^T.$$

A subscript, i, is now introduced to label rows of X (each value of i corresponds to a particular (W, D) combination for the future).

We have: $\mu_i' = \exp(\eta_i).$

Therefore, using standard results for the log-normal distribution, we have:

—unbiased estimate of μ_i': $\hat{\mu}_i' = \exp(\hat{\eta}_i - \frac{1}{2}\sigma_{ii})$

—var-covariance of these estimates: $V_{ij}' = \mu_i' \cdot \mu_j' \cdot [\exp(\sigma_{ij}) - 1]$

where σ_{ij} are the elements of $\hat{\Sigma}$.

These quantities are approximated using estimates on the right-hand-side to give:

$$\hat{\mu}'_i = \exp(\hat{\eta}_i - \tfrac{1}{2}\hat{\sigma}_{ii})$$

$$V'_{ij} = \hat{\mu}'_i \cdot \hat{\mu}'_j \cdot [\exp(\hat{\sigma}_{ij}) - 1].$$

These will be denoted $\hat{\boldsymbol{\mu}}'$ and $\hat{\boldsymbol{V}}'$.

It is a simple matter to convert these into estimates, $\hat{\boldsymbol{\mu}}$, of the un-normalised incremental paid figures, together with the estimated variance-covariance matrix $\hat{\boldsymbol{V}}$.

If $\boldsymbol{f}$ is the vector given by $\boldsymbol{f} = (\tilde{\varepsilon}_W \cdot \alpha_D)$, we require:

$$\hat{\boldsymbol{\mu}} = \boldsymbol{f} \times \hat{\boldsymbol{\mu}}' \qquad \text{(dyadic multiplication)}$$

$$\hat{\boldsymbol{V}} = \boldsymbol{f} \cdot \hat{\boldsymbol{V}}' \cdot \boldsymbol{f}^T.$$

Note that the vector $\boldsymbol{f}$ may also contain discount factors and/or factors representing a small difference between future inflation and average past inflation.

$\hat{\boldsymbol{\mu}}$ has one element for each future (W, D) combination. The prediction, $\hat{R}$, of the total for a certain set of (W, D) combinations (e.g. a particular origin year or a particular payment year) can be obtained by constructing the appropriate vector $\boldsymbol{a}$ of 0s and 1s, to give:

$$\hat{R} = \boldsymbol{a}^T \cdot \hat{\boldsymbol{\mu}}$$

with estimated variance:

$$T^2 = \boldsymbol{a}^T \cdot \hat{\boldsymbol{V}} \cdot \boldsymbol{a}.$$

To obtain the mean-square-error of prediction S^2, it is necessary to add the process variance for the future data-points concerned:

$$S^2 = T^2 + \sum_{i \in \boldsymbol{a}} V(Y_i).$$

The model has: $V(Y'_i) = \phi_W \cdot \psi_D \cdot \mu'_i$

but: $Y_i = f_i \cdot Y'_i$

so: $V(Y_i) = f_i^2 \cdot \phi_W \cdot \psi_D \cdot \mu'_i$

$= f_i \cdot \phi_W \cdot \psi_D \cdot \mu_i.$

This can be estimated as:

$$\widehat{V(Y_i)} = f_i \cdot \phi_W \cdot \psi_D \cdot \hat{\mu}_i$$

hence the standard error, S, is given by:

$$S^2 = T^2 + \sum_{i \in \boldsymbol{a}} \widehat{V(Y_i)}.$$

6.3 *Results in Constant Money Terms*

The calculations of Section 6.2 give results in current terms, because inflation, ι, is a component of the parameters β_{W1} and β_{W3} (see Section 4.5.2). (Current terms means estimates of amounts to be paid in 1978 are in 1978 terms, amounts to be paid in 1979 are in 1979 terms, etc.)

If results are required in constant terms, then average past inflation must not be projected into the future. It is not correct simply to remove the estimated inflation, $\hat{\iota}$, from the current-price results by including *discount factors* in f, because this ignores the estimation variance of $\hat{\iota}$ and the covariance between $\hat{\iota}$ and $\hat{\beta}$. It is better to proceed as in Section 6.2, but with X as follows (for example):

$$X = \left[\begin{array}{cccc|cccc|c} -1 & 1 & \ln 10 & -10 & & & & & \\ -2 & 1 & \ln 11 & -11 & & 0 & & & 0 \\ -3 & 1 & \ln 12 & -12 & & & & & \\ \hline & & & & -1 & 1 & \ln 8 & -8 & \\ & & & & -2 & 1 & \ln 9 & -9 & \\ & & 0 & & -3 & 1 & \ln 10 & -10 & 0 \\ & & & & -4 & 1 & \ln 11 & -11 & \\ & & & & -5 & 1 & \ln 12 & -12 & \\ \hline & & 0 & & & & 0 & & \text{etc.} \end{array}\right]$$

7. USE OF INCURRED DATA

7.1 *Modelling Incurred Data*

The value of incurred data lies in the fact that cumulative incurred approaches the same ultimate value as cumulative paid, and since we know what has been paid so far, reserve estimation is equivalent to the estimation of ultimate cumulative paid.

In the examples given in Figures 7.1a and b, the continuous line is cumulative paid, the broken line is cumulative incurred. The difference between the two curves represents cumulative outstanding, that is, the sum of current case estimates.

The second example illustrates the effect of a positive bias in the case estimates. When a claim is reported, outstanding (and therefore incurred) increases by the amount of the case estimate; when a claim is paid, outstanding decreases by the amount of the case estimate, and simultaneously, paid increases by the amount actually paid. If case estimation has a positive bias, the increase in paid is typically less than the decrease in outstanding, so incurred is reduced. The peak of the incurred curve represents the stage of development at which the increments due to new reportings are outweighed by these decrements due to settlements.

There is also the possibility of negative bias in case estimation. This would

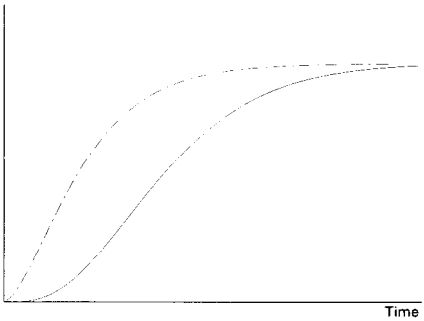


Figure 7.1a.

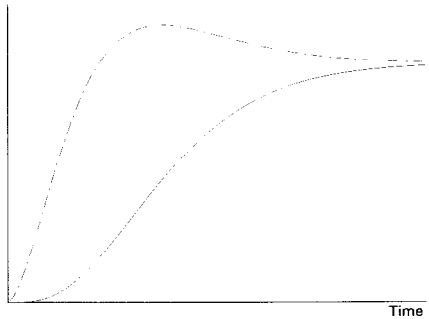


Figure 7.1b.

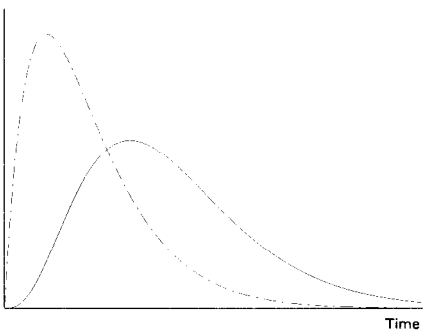


Figure 7.1c.

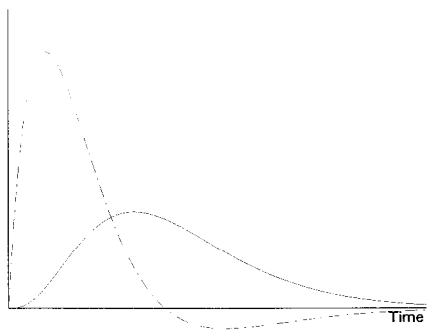


Figure 7.1d.

differ from the first example, in that incurred would approach ultimate paid more slowly, continuing to increase even when no more claims are being reported.

The incremental plots corresponding to these cumulative plots are shown in Figures 7.1c and d.

It is clear from these graphs (Figure 7a–d) that it is not valid in general to apply the Hoerl curve model of Sections 3 to 6 directly to incremental incurred data. It seems reasonable to model the time of reporting as a Gamma-distributed random variable (as was done for the time of settlement in the paid context), but the claim size occurring at this time is now a case-estimate, not a final payment. This case estimate may subsequently vary and will, eventually, be replaced by an actual payment, which may differ from the case estimate; no allowance is made for these possibilities in the paid claims model. However, in some cases, it is possible largely to remove these effects from the incurred data and to apply the method to the adjusted incurred data.

Suppose amendments to case estimates, while a claim is outstanding, have an expected value of zero. Then, if we are dealing with a sufficiently large number of

claims, the total of these amendments can reasonably be ignored. Incremental outstanding, which is denoted Z_D, can thus be broken down into just two components:

$$Z_D = Z_D^{(1)} - Z_D^{(2)}$$

where: $Z_D^{(1)}$ = total of case estimates of claims reported at time D, and

$Z_D^{(2)}$ = total of case estimates of claims settled at time D.

It would be reasonable to apply the method developed for paid increments, Y_D, to the quantity $Z_D^{(1)}$ (or to $Z_D^{(2)}$), if this could be isolated.

Since: $$I_D = Y_D + Z_D$$

we have: $$Z_D^{(1)} = I_D + Z_D^{(2)} - Y_D.$$

Suppose that case estimates are biased by a factor b, that is, for individual claims we have:

$$\text{case-estimate} = b \times \text{amount paid} + \text{random error}$$

where the expected value of the random error is zero.

Again, if we are dealing with sufficiently large numbers of claims, the random error of the aggregate $Z_D^{(2)}$ may be ignored and we have:

$$Z_D^{(2)} = b \cdot Y_D$$

hence: $$Z_D^{(1)} = I_D + (b-1) \cdot Y_D.$$

By the bias assumption, the ultimate cumulative of the $Z_D^{(1)}$ is b times ultimate paid, so in order to estimate ultimate paid, we require:

$$\frac{1}{b} Z_D^{(1)}.$$

This will be denoted I'_D and called adjusted incurred:

$$I'_D = \frac{1}{b} \cdot I_D + \frac{b-1}{b} \cdot Y_D.$$

The method described in Sections 3 to 6 may be applied using this quantity in place of incremental paid. Note that adjusted incurred is a weighted average of the original incremental incurred and the incremental paid data.

The bias factor might, in general, incorporate a component due to bias in claim numbers as well as in claim size. There may be two classes of reported claims: those which lead to a final settlement of around (case-estimate)/b, as already described; and those which lead to a final settlement without any payment. If the proportion settled for no payment is constant over D, this situation can be accommodated by using b to represent the product of the bias in both outstanding severities and outstanding claim numbers. Thus, it may be plausible

for b to take quite large values (e.g. $b=3$), for some classes of business. (Note that claims in this second class do not affect the paid model, as they are outside its scope.)

In general, the bias factor, b, may vary. Variation of b from one origin year to the next clearly causes no complications to the above, because a single origin year was considered. However, it is normally more plausible for b to be constant for each calender year (like the force of claims inflation). Variation of b between calendar years causes a complication, because the claims making up each $Z_D^{(2)}$ will, in general, have been reported in various development years, $d < D$, and so will have been subject to differing amounts of bias. We require an overall bias factor, b_D, such that: $E(b_D\ Y_D - Z_D^{(2)}) = 0$. An approximate value for b_D can be obtained as a weighted average of the bias factors assumed for each of the earlier development years, using the incurred data I_D as weights.

If no specific information is available on the calendar year bias factors, a *trial and error* approach can be used to adjust the incurred data. Plausible values for the bias factors are tried, and these used to determine I'_D, as described above, until two criteria are satisfied simultaneously:

(i) the adjusted incremental incurred data, I'_D, appear to follow a Gamma curve for each origin year, and
(ii) the cumulative paid, Y_D, and adjusted incurred, I'_D, both appear to approach the same ultimate value, for each origin year.

Clearly, both these criteria are most easily judged for well-developed years of origin: if there are no such years, estimation of the bias factors will be poor. It is not vital that criterion (ii) be satisfied; provided the ratio of the two ultimates does not appear to vary significantly between origin years, the paid and incurred results can be combined, as described in Section 7.2.

7.2 *Combining Paid and Incurred Results*

If U_W represents the ultimate paid,
P_W represents paid so far, and
R_W represents the future payments for origin year W,

then we have:

$$U_W = R_W + P_W.$$

From analysing the paid data, we have one set of estimates, $\hat{U}_W^{(1)}$, with standard errors, $S_W^{(1)}$, and from the adjusted incurred data we have a second set, $\hat{U}_W^{(2)}$, with standard errors, $S_W^{(2)}$. The problem now is to combine these two sets of estimates to obtain final estimates, $\hat{U}_W$ (from which final estimates $\hat{R}_W$ can be obtained trivially), and standard errors S_W.

In practice, it sometimes happens that the two sets of estimates differ significantly from one another, that is $(\hat{U}_W^{(1)} - \hat{U}_W^{(2)})$ is large compared to $(S_W^{(1)2} + S_W^{(2)2})^{\frac{1}{2}}$. This indicates that the adjusted incurred data are not tending towards ultimate paid for some reason (which is, perhaps, not surprising, in view

of the approximate nature of the incurred method). To cope with this, a bias factor, k, is hypothesised in the estimation of ultimate paid using adjusted incurred data. Thus we have:

$$\hat{U}_W^{(1)} = U_W + E_W^{(1)}$$

where: $E_W^{(1)}$ has $E(E_W^{(1)}) = 0$ and $V(E_W^{(1)}) = S_W^{(1)2}$

and

$$\hat{U}_W^{(2)} = k \cdot U_W + E_W^{(2)}$$

where: $E_W^{(2)}$ has $E(E_W^{(2)}) = 0$ and $V(E_W^{(2)}) = S_W^{(2)2}$.

To simplify matters, the incurred estimates, $\hat{U}_W^{(2)}$, are taken to be statistically independent of the paid estimates, $\hat{U}_W^{(1)}$. This is not actually true for two reasons. Firstly, the paid data and the incurred data are both generated by the same set of individual claims; and, secondly, the paid data contribute directly to the adjusted incurred data if the case-estimate bias factors are not all one. It would be extremely difficult to calculate the effects of these two factors, but it is believed they will be slight, provided the case-estimate bias factors are not too large.

If $\boldsymbol{T}^{(i)}$ is used to represent the vector of estimates $\hat{U}_W^{(i)}$ for $W = 1$ to M, then rewriting the above in vector notation we have:

$$\boldsymbol{T}^{(1)} = \boldsymbol{U} + \boldsymbol{E}^{(1)} \quad \text{and} \quad \boldsymbol{T}^{(2)} = k \cdot \boldsymbol{U} + \boldsymbol{E}^{(2)}.$$

As well as the variances, $S_W^{(i)2}$, the covariances between origin years of the estimated ultimates can be estimated fairly straightforwardly for each set of results, so we have for $i = 1, 2$: $E(\boldsymbol{E}^{(i)}) = 0$ and $V(\boldsymbol{E}^{(i)}) = \Phi^{(i)}$, where $\Phi^{(i)}$ is a known (estimated) $(M \times M)$ matrix.

The problem is to use the two *observations*, $\boldsymbol{T}^{(1)}$ and $\boldsymbol{T}^{(2)}$ (and their known var-covariance matrices $\Phi^{(1)}$ and $\Phi^{(2)}$), to produce the optimal estimate, $\hat{\boldsymbol{U}}$ (together with its var-covariance matrix). We are not interested in the value of the unknown parameter, k: it is a *nuisance parameter*. An approximate solution to this problem has been found, but the details are not given here.

8. EXAMPLE

8.1 *Analysis of Paid Data*

The data for this example are annual professional indemnity data for accident years 1978 to 1988 inclusive:

2	39	175	347	408	278	295	122	204	112	153
2	31	237	409	592	549	167	247	100	31	
11	213	291	694	473	522	393	284	254		
1	208	568	947	748	807	668	730			
4	375	1349	717	1076	892	1162				
116	622	2174	2101	2255	2832					
5	692	1749	2448	2898						
109	697	2253	3011							
23	509	1693								
19	1473									
277										

The table below gives: (i) the numbers of claims reported during development year zero of each origin year (for example, the figure for 1978 is the number of reports during 1978 of claim events which occurred during 1978), and (ii) the relative exposures (derived from (i) by dividing by 448):

	(i)	(ii)	(iii)
1978	448	1·00	
1979	502	1·12	0·086
1980	583	1·30	0·067
1981	677	1·51	0·039
1982	789	1·76	−0·021
1983	1013	2·26	−0·022
1984	977	2·18	−0·029
1985	1071	2·39	−0·019
1986	1147	2·56	−0·038
1987	1322	2·95	−0·037
1988	1273	2·84	−0·024

The model assumes a uniform rate of claims inflation, so any non-uniform component of inflation believed to be present must be removed from the data before proceeding. Changes in the inflation rate for professional indemnity claims are believed to have reflected changes in RPI inflation. The average force of inflation in the RPI over the period end-of-78 to end-of-88 was 0·074. Subtracting this from the force of RPI inflation for each origin year, gives the variable component of force of inflation in column (iii) above. Removing this element of inflation from the data gives:

2	36	150	287	344	240	262	110	192	109	153
2	27	196	345	511	488	151	232	98	31	
9	176	245	599	420	473	370	277	254		
1	175	490	841	677	759	652	730			
3	323	1198	649	1012	871	1162				
100	552	1968	1976	2201	2832					
4	626	1645	2389	2898						
99	655	2199	3011							
22	497	1693								
19	1473									
277										

The values for the first and last payment years (diagonals of the triangle) have, of course, not been affected by this adjustment.

A line plot of these data, further adjusted for exposure (the figures are divided by the appropriate figure from column (ii) of the table above), is shown in Figure 8.1a.

This plot shows the typical Hoerl curve type shape of the run-off. It also shows that the later the origin year (i.e. the shorter the curve), the higher the level of the curve. This is because of the uniform component of claims inflation; as this is a component of the model, no attempt is made to remove it from the data.

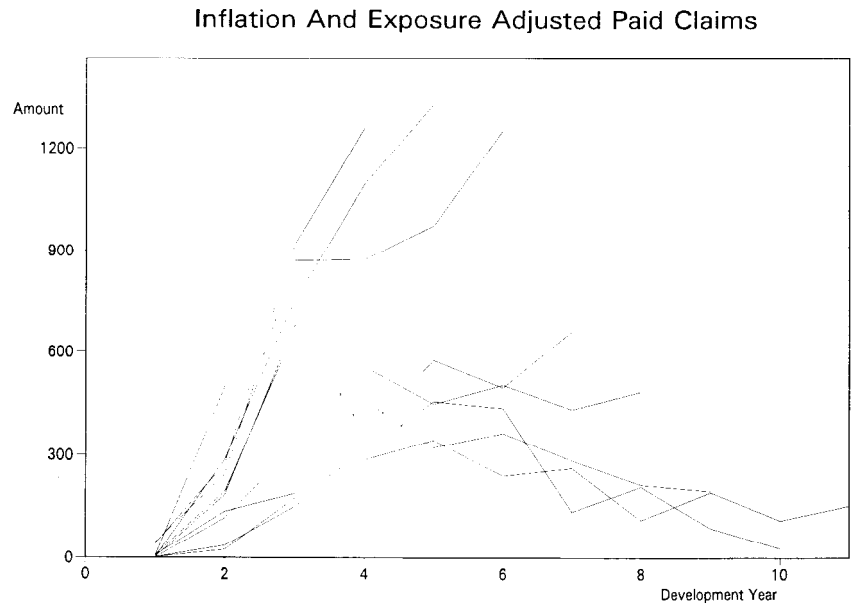


Figure 8.1a.

The first stage of model calibration is to use Fisher's scoring method to fit the Hoerl curve model of Section 5.1 to the data for each of origin years 1978 to 1986 separately. Since ϕ_W is constant throughout each origin year, this first stage can be accomplished given values of ψ_D. To calculate values for ψ_D it is necessary to assume values for λ and the constant force of inflation, ι; the values 0 and 0·18 are used initially. After fitting the model for each origin year, the residuals for all nine origin years together are used in the obvious way to estimate the scale parameter as $\phi_0 = 9{\cdot}573$. This value is used to calculate the ϕ_W, which are used to scale the inverse information matrices, to give the variance–covariance matrices $\hat{V}_W$.

The next stage is to use the method described in Section 5.5, to estimate the variance of the *perturbations* in the run-off parameters from one origin year to the next. These calculations give the following adaptive standard deviations u_i:

Beta-1	Beta-2	Beta-3
0·072	0·087	0·000

The final stage of fitting is to apply the Kalman filter to all origin years (including the last two) as described in Sections 5.2 and 5.3. For the present example, the Kalman filter is initialised using uninformative priors for the beta parameters, and a prior estimate of 0·18, with a standard error of 0·06, for the constant force of claims inflation.

The posterior estimate of the constant force of inflation (which is one of the model parameters estimated through the Kalman filter) is 0·20, with a standard error of 0·03. As this differs somewhat from the value used to calculate the ψ_D values which were used to fit the model, the other parameter estimates are disregarded. The estimate $\hat{\imath}=0{\cdot}20$ is used to recalculate the ψ_D values, and the entire fitting procedure is repeated.

As before, Hoerl curve fitting is successful for each of the first nine origin years (so Section 5.4 is not needed). The scale parameter is estimated as 8·25 and the adaptive standard deviations as:

Beta-1	Beta-2	Beta-3
0·036	0·102	0·000

The substantial decrease in the adaptive standard deviation for beta-1 should not be surprising: the value of $\imath$ was previously too small, so part of the trend increase in the beta-1 values was being falsely attributed to random perturbation (see end of Section 5.5).

The Kalman filter is applied using the same prior distributions as before. This time the posterior estimate of $\imath$ is 0·201, with a standard error of 0·027: the difference between this and the value of 0·200 used for calculating ψ_D and ϕ_W is very insignificant.

The post-fit estimate of the scale parameter, calculated as described in Section 5.6, is 8·89, which differs from the value of 8·25 used in fitting. Therefore, fitting is repeated using this new value. Note, however, that there is no need to repeat the first stage of the fitting process; the curves fitted for individual origin years do not depend on the value of the scale parameter. The second stage of fitting; the calculation of the adaptive standard deviations (Section 5.5), does depend on the value of the scale parameter (through the var-covariance matrices, $\hat{V}_W$), but the initial estimate of the scale parameter is the most appropriate value for this purpose. Thus, it is only the final stage of fitting, the Kalman filter, which needs to be redone, using the new estimate of the scale parameter; in two further iterations the estimate of the scale parameter converges to 8·94.

The parameter estimates are then as follows:

	Beta-1	Standard Error	Beta-2	Standard Error	Beta-3	Standard Error
1978	4·803	0·151	2·448	0·250	0·623	0·071
1979	5·006	0·130	2·415	0·238	0·623	0·071
1980	5·212	0·113	2·389	0·229	0·623	0·071
1981	5·425	0·099	2·437	0·217	0·623	0·071
1982	5·638	0·091	2·458	0·205	0·623	0·071
1983	5·857	0·089	2·591	0·192	0·623	0·071
1984	6·047	0·093	2·589	0·185	0·623	0·071
1985	6·231	0·102	2·555	0·187	0·623	0·071
1986	6·414	0·117	2·524	0·203	0·623	0·071
1987	6·614	0·136	2·552	0·222	0·623	0·071
1988	6·816	0·159	2·544	0·242	0·623	0·071

The variation in the true values of beta-1 from one origin year to the next is made up of two components. There is a trend increase of 0·20 (the force of past claims inflation), and there is random perturbation with a standard deviation of about 0·036. There is additional variation in the estimates displayed above because of estimation error: the magnitude of this is indicated by the standard errors to the right of the estimates. The values of beta-2 are subject only to genuine random variation, with a standard deviation of 0·102, and estimation error with a standard deviation given in the table. For beta-3, the data showed no evidence of any variation in the underlying value between origin years (the adaptive standard deviation was estimated to be zero). Therefore, the value was assumed to be the same for all origin years, and has been estimated as 0·623, the standard error of estimation being 0·071.

The plots of standardised residuals against both origin year and development year are shown in Figure 8.1b. These are consistent with the hypotheses $E(\hat{R}_{WD})=0$ and $V(\hat{R}_{WD})=1$ for all W and D, (see Section 5.6). In particular, there is no evidence of heteroscedasticity (varying variance) with respect to development year, so the value of 0 for λ is empirically satisfactory. There is also no evidence that the mean of the residuals differs from zero for the early development years: the default values of α have provided a satisfactory fit and free-fitting as described in Section 5.7 is not necessary in this case.

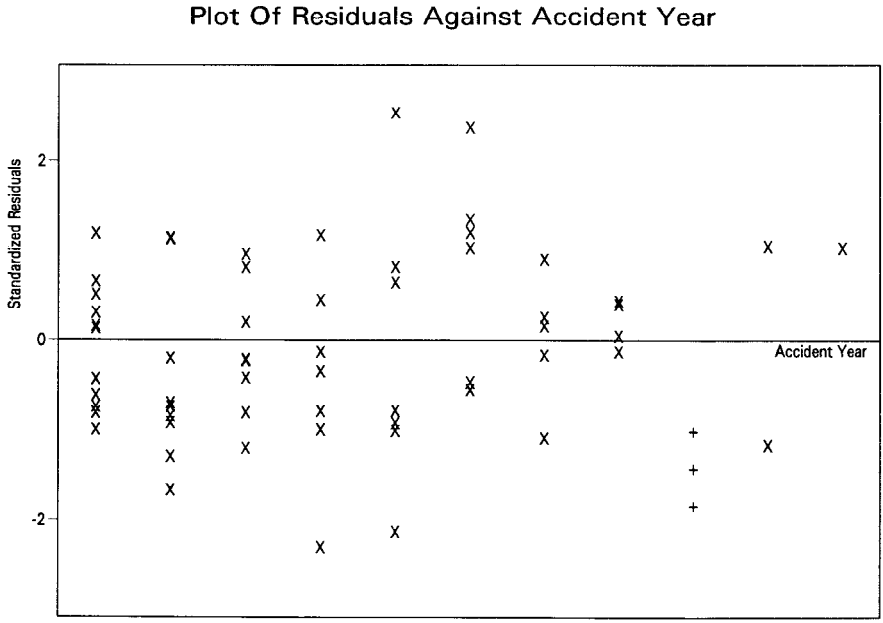
At first sight, the fact that all three residuals are negative for origin year 1986 may cause some concern, but, in fact, this is not significant. The same sort of occurrence is very common in purely random scatter-plots.

Figure 8.1c shows plots of observed and fitted values for two of the origin years.

Results in current terms, calculated as described in Section 6.2, with projection to the end of the twentieth development year, are given below. These are not discounted, and the projected force of inflation for the future is the estimated past value of 0·20. (Discounted results could easily have been obtained: see Section 6.2.)

	$\hat{R}$	S	P	$\hat{U}$
1978	118·81	139·45	2135·00	2253·81
1979	232·82	203·38	2365·00	2597·82
1980	475·59	306·71	3135·00	3610·59
1981	1131·35	521·10	4677·00	5808·35
1982	2449·38	875·60	5575·00	8024·38
1983	7165·96	1950·25	10100·00	17265·96
1984	10867·05	2775·95	7792·00	18659·05
1985	16753·35	4216·08	6070·00	22823·35
1986	23861·52	6516·97	2225·00	26086·52
1987	38378·76	11719·82	1492·00	39870·76
1988	45315·88	15638·14	277·00	45592·88
Totals:	146750·48	35827·14	45843·00	192593·48

$\hat{R}$ is the total of projected future increments for each origin year, S is the


Plot Of Residuals Against Accident Year
Standardized Residuals
2
0
-2
Accident Year

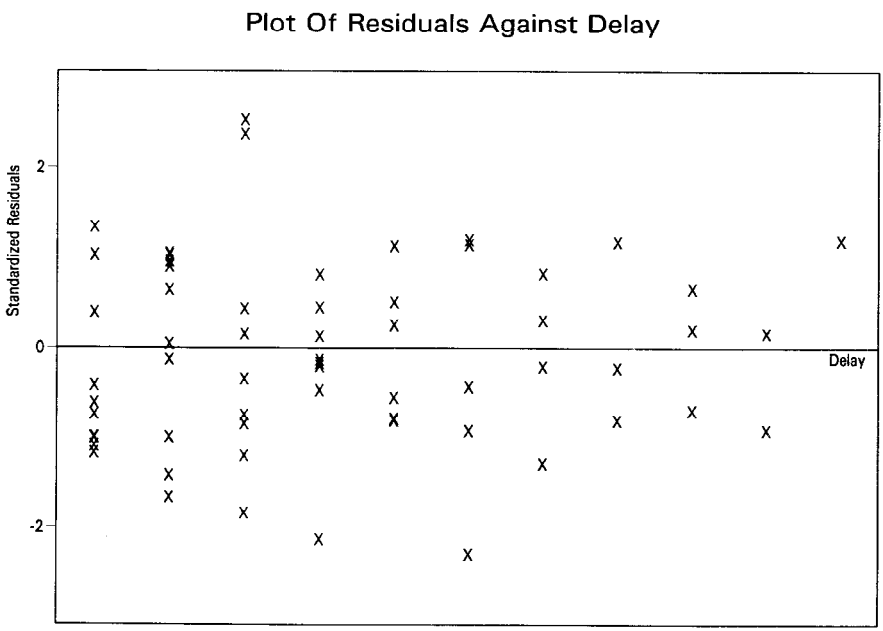
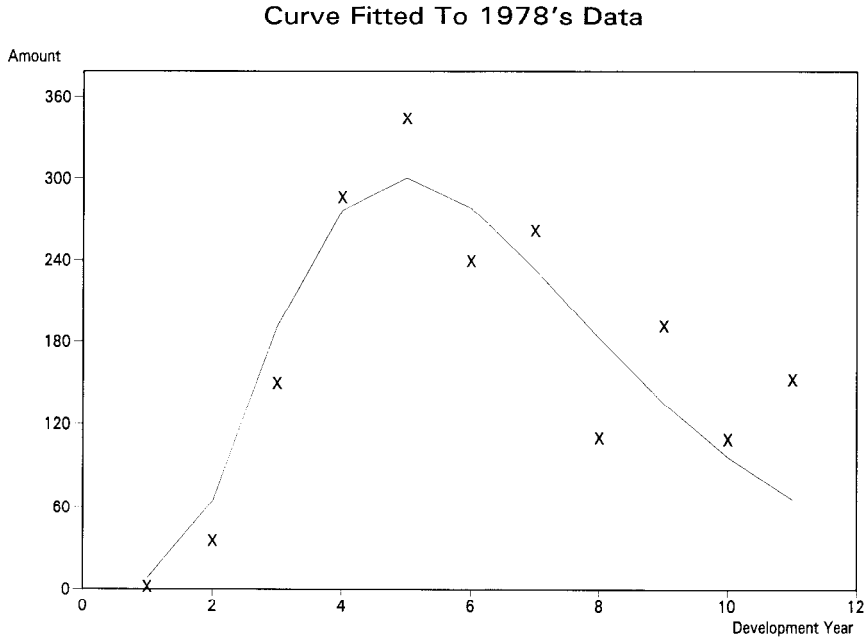

Plot Of Residuals Against Delay
Standardized Residuals
2
0
-2
Delay

Figure 8.1b.


Curve Fitted To 1978's Data
Amount
360
300
240
180
120
60
0
0
2
4
6
8
10
12
Development Year

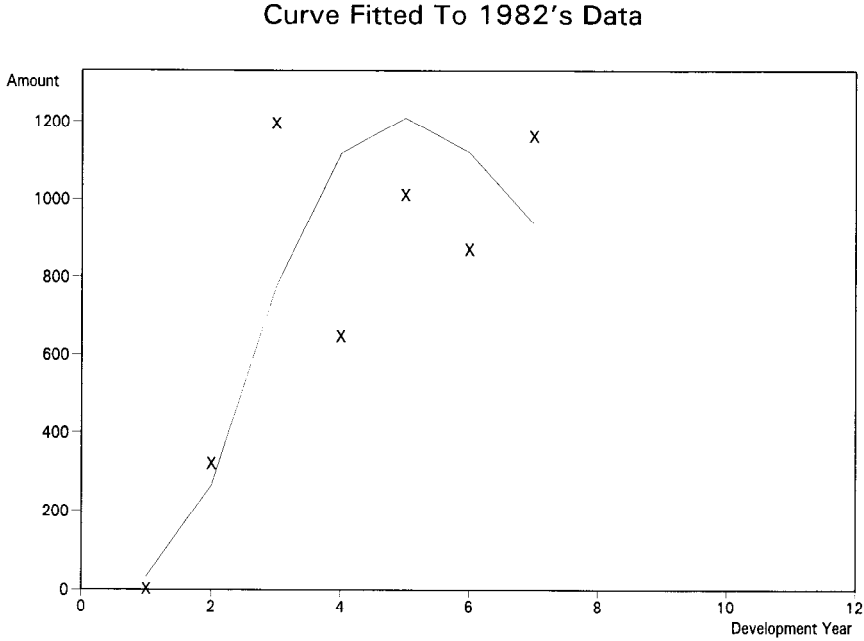

Curve Fitted To 1982's Data
Amount
1200
1000
800
600
400
200
0
0
2
4
6
8
10
12
Development Year

Figure 8.1c.

standard error of this total, P is the sum of the original paid data, and $\hat{U}$ is the estimated ultimate of cumulative paid, that is $\hat{U}=\hat{R}+P$. For each of the later origin years (those with substantial future development, R) the standard error, S, is about 35% of the estimate $\hat{R}$ itself. However, the estimated total for all origin years combined has a standard error of only about 25%. This improvement is due to partial cancellation of the random variation when origin years are combined.

The same estimates are given below, broken down by payment year:

	$\hat{R}$	S
1989	18308·45	2049·88
1990	22089·53	3205·79
1991	22871·03	4327·32
1992	20949·33	4972·34
1993	17528·01	5064·77
1994	13703·20	4734·28
1995	10165·60	4160·69
1996	7232·51	3499·06
1997	4972·47	2855·82
1998	3321·30	2289·60
1999	2164·26	1823·09
2000	1379·98	1454·74
2001	862·55	1170·11
2002	528·84	950·56
2003	315·67	775·55
2004	183·29	633·44
2005	101·97	511·97
2006	52·50	401·04
2007	9·97	271·09

Results, in 1988 terms, calculated as described in Section 6.3, are given below:

	$\hat{R}$	S	P	$\hat{U}$
1978	74·87	82·65	5306·30	5381·16
1979	145·05	118·23	5347·61	5492·67
1980	292·49	174·77	6155·12	6447·61
1981	682·72	289·20	7694·25	8376·97
1982	1447·58	473·77	8484·68	9932·27
1983	4081·20	1011·39	13568·37	17649·57
1984	5976·39	1390·98	9425·68	15402·07
1985	8819·93	2005·40	6857·22	15677·15
1986	11804·87	2816·86	2333·22	14138·10
1987	17031·91	4307·06	1496·22	18528·13
1988	17185·70	4715·19	277·00	17462·70
Totals:	67542·72	13249·70	66945·68	134488·40

Here, P is the sum of the original increments, each inflated to 1988 terms, using the estimated force of claims inflation 0·20. These results can be useful for comparing the experience of different origin years in real terms. The variation in the estimated ultimate, $\hat{U}$, between origin years is mainly because of varying exposure—1983 appears to have been a particularly bad year, even after allowing

for exposure (this is also clear from the original data). The standard errors are lower in percentage terms than in the current price results; in those results there was an additional element of uncertainty because of the projection of an uncertain inflation rate into the future. The current price results could also be broken down by payment year if required.

It is interesting to examine the sensitivity of the results to the run-off horizon. The table below gives the total prediction in current terms:

Number of Development Years	$\hat{R}$	S
20	146750	35827
22	146831	35898
24	146857	35924
26	146865	35933
28	146867	35936
30	146868	35937

There is clearly no need to project any further: only 0·1% of future payments are expected to occur between development years 20 and 30, and S increases by only 0·3% over this interval.

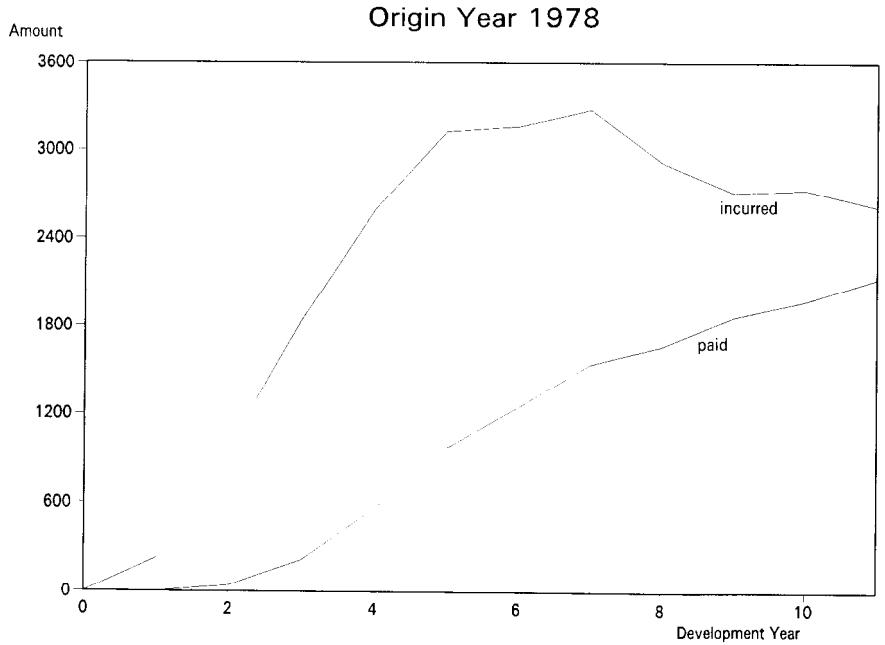
The actual reserve should be $\hat{R} + C \times S$, where C is a factor chosen to reflect the required level of prudence. In the present example, the distribution of the standardised residuals is approximately Normal. This implies that the prediction errors are also approximately Normal, and the factor, C, can be chosen on this basis. For example, $C = 1{\cdot}28$ gives an upper 90% confidence limit for the total of future payments; there is only a 1 in 10 chance that the total for all origin years will exceed 193000.

8.2 *Analysis of Incurred Data*

The incremental incurred data is given below for each of the origin years 1978 to 1988:

229	737	882	746	533	36	114	−356	−202	19	−117
118	1163	851	846	−94	231	−189	58	−73	55	
48	1143	1744	983	652	−522	−33	111	−103		
337	2035	2022	1664	211	−65	121	−415			
485	4054	2462	202	751	162	569				
994	6993	5117	1283	767	741					
1461	7297	6822	1114	503						
2101	8119	7107	2574							
915	8963	6614								
1423	13142									
6011										

Cumulative plots of these data on the same axes as the cumulative paid data are shown in Figure 8.2a for two of the origin years. Cumulative incurred appears to be approaching the same ultimate as cumulative paid, but not monotonically; incurred decreases in the tail when case-estimates are being replaced by actual payments. This suggests that there is a positive bias in case estimates.


Origin Year 1978
Amount
3600
3000
2400
1800
1200
600
0
0
2
4
6
8
10
Development Year
incurred
paid

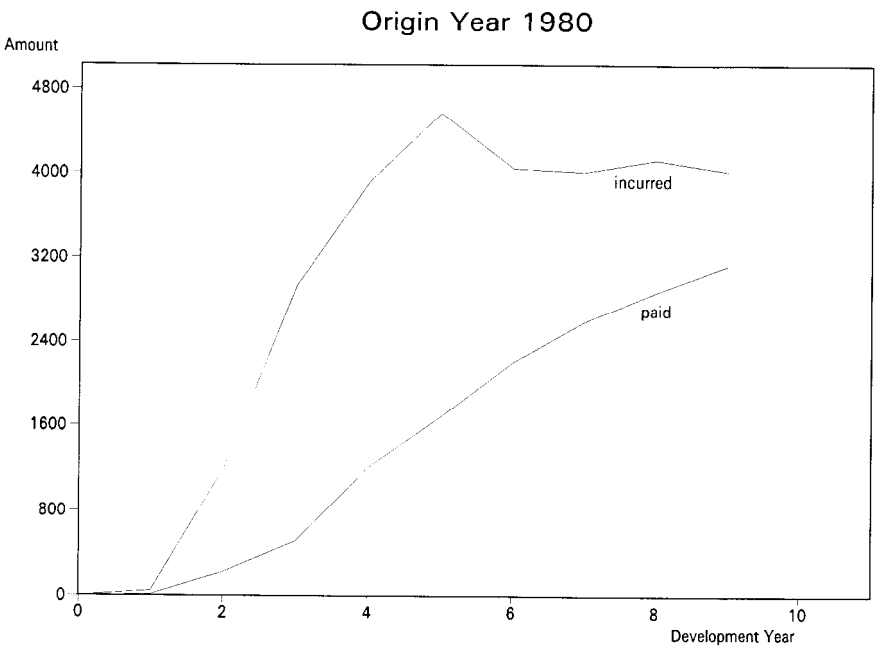
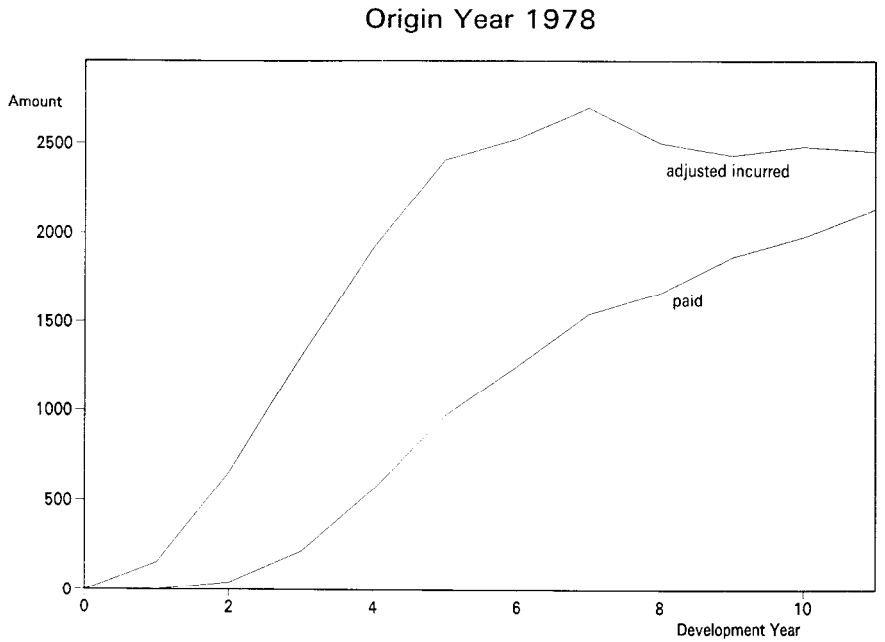

Origin Year 1980
Amount
4800
4000
3200
2400
1600
800
0
0
2
4
6
8
10
Development Year
incurred
paid

Figure 8.2a.


Origin Year 1978
Amount
2500
2000
1500
1000
500
0
adjusted incurred
paid
0
2
4
6
8
10
Development Year

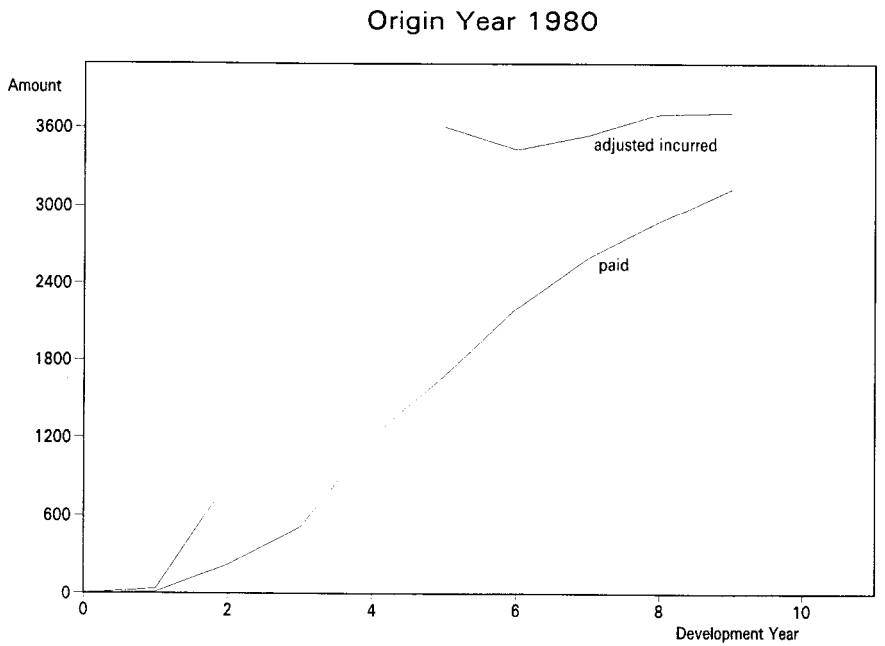

Origin Year 1980
Amount
3600
3000
2400
1800
1200
600
0
adjusted incurred
paid
0
2
4
6
8
10
Development Year

Figure 8.2b.

By trial and error, it is found that if the incurred data are adjusted on the assumption of a bias factor of 1·5 for all calendar years (as described in Section 7.1), then the adjusted data appear to be consistent with the Hoerl curve model for each origin year except 1978: adjusted cumulative plots are shown in Figure 8.2b. The incurred data for 1978 appear to follow a pattern different from those for all other origin years: it is, therefore, not very useful and is discarded. (As 1978 is sufficiently well developed for the paid data alone to give a reliable estimate, it is unlikely that any useful information is lost by discarding the incurred data. It is quite common for the incurred data for the first one or two origin years in a triangle to contain little useful information; it seems that case estimates are often poor initially due to the lack of past experience.)

The adjusted incurred data are given below for origin years 1979 to 1988 inclusive:

79	786	646	700	135	337	−70	121	−15	47
36	833	1260	887	592	−174	109	169	16	
225	1426	1537	1425	390	226	303	−33		
325	2828	2091	374	859	405	767			
701	4869	4136	1556	1263	1438				
976	5095	5131	1559	1301					
1437	5645	5489	2720						
618	6145	4974							
955	9252								
4100									

The methods of Sections 3 to 6 can be applied to these data. Initial values of 0 and 0·201 are taken for λ and ι. The value 0·201 is the estimate previously obtained from the paid data. Fisher's scoring method can then be applied to fit a Hoerl curve to each origin year separately, and the results used, as described in Section 5.5, to estimate the adaptive standard deviations. These estimates are all zero, indicating that the underlying run-off pattern of the adjusted incurred data is the same for all origin years 1979 to 1988. The Kalman filter could now be applied as before, but, as the adaptive standard deviations are all zero, there is no need to use the Kalman filter to link the origin years; a single Hoerl curve (with allowance for inflation) can be fitted to all origin years simultaneously, as described in Section 5.4. (This procedure is preferable to the use of the Kalman filter, because it avoids the Normal approximation for quasi-likelihood estimates obtained from small numbers of data-points.) The Kalman filter is used only to combine the curve-fitting results with the prior information. As for paid, no prior information is assumed for the beta parameters. However, the prior distribution for the force of inflation is taken as the posterior distribution from the analysis of paid data; the estimate is 0·201, with a standard error of 0·027.

The scale parameter converges to 48·5 in three iterations of the Kalman filter. The posterior estimate of the force of inflation is then 0·193, with a standard error 0·018. The value of 0·201, used to calculate ψ_D for the initial curve fits, does not differ significantly from the posterior estimate, so there is no need to refit using the new value. The other parameter estimates are given below:

	Beta-1	Standard Error	Beta-2	Standard Error	Beta-3	Standard Error
1978	7·514	0·221	1·884	0·275	1·201	0·160
1979	7·707	0·212	1·884	0·275	1·201	0·160
1980	7·900	0·205	1·884	0·275	1·201	0·160
1981	8·093	0·199	1·884	0·275	1·201	0·160
1982	8·287	0·195	1·884	0·275	1·201	0·160
1983	8·480	0·192	1·884	0·275	1·201	0·160
1984	8·673	0·191	1·884	0·275	1·201	0·160
1985	8·866	0·192	1·884	0·275	1·201	0·160
1986	9·059	0·194	1·884	0·275	1·201	0·160
1987	9·253	0·198	1·884	0·275	1·201	0·160
1988	9·446	0·204	1·884	0·275	1·201	0·160

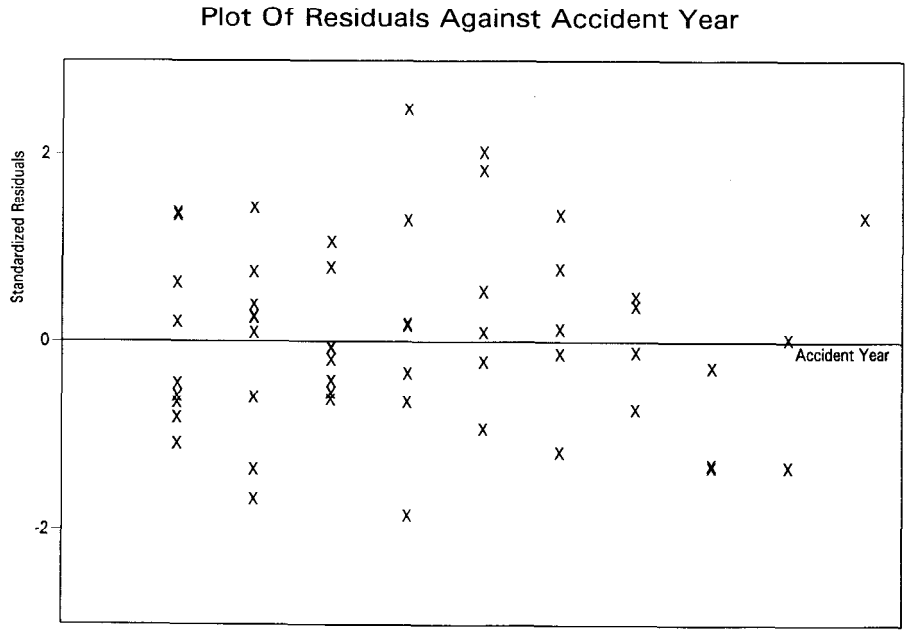
The residual plots (Figure 8.2c) and fitted-value plots (Figure 8.2d) are satisfactory, so final results may be calculated as described in Section 6.2. These are given below for a development horizon of 16 years:

	$\hat{R}$	S	P	$\hat{U}$
1979	402·19	26·73	2365·00	2767·19
1980	598·13	55·87	3135·00	3733·13
1981	843·31	114·90	4677·00	5520·31
1982	2158·96	233·27	5575·00	7733·96
1983	4217·64	493·11	10100·00	14317·64
1984	7317·05	875·91	7792·00	15109·05
1985	12479·31	1614·91	6070·00	18549·31
1986	18460·72	2791·88	2225·00	20685·72
1987	31430·08	4819·52	1492·00	32922·08
1988	41065·50	6843·68	277·00	41342·50
Totals:	118872·82	—	45843·00	164715·82

Here, $\hat{U}$ is the projected ultimate of the cumulative adjusted incurred data, S is the standard error of $\hat{U}$, P is again the total of the original paid data, and $\hat{R}$ is $\hat{U}$ minus P (thus S is also the standard error of $\hat{R}$). The standard error of the total for all origin years has not been calculated.

Combining these results with the paid results to the end of twenty development years, as outlined in Section 7.2, yields the following final estimates:

	$\hat{R}$	S	P	$\hat{U}$
1978	118·81	139·45	2135·00	2253·81
1979	323·03	25·77	2365·00	2688·03
1980	492·79	53·47	3135·00	3627·79
1981	707·79	109·19	4677·00	5384·79
1982	1973·73	219·49	5575·00	7548·73
1983	4006·51	465·42	10100·00	14106·51
1984	7234·53	813·91	7792·00	15026·53
1985	12542·21	1470·94	6070·00	18612·21
1986	18763·79	2505·03	2225·00	20988·79
1987	31590·29	4349·66	1492·00	33082·29
1988	40733·85	6120·88	277·00	41010·85
Totals:	118466·33	—	45843·00	164309·33


Plot Of Residuals Against Accident Year
Standardized Residuals
2
0
-2
Accident Year

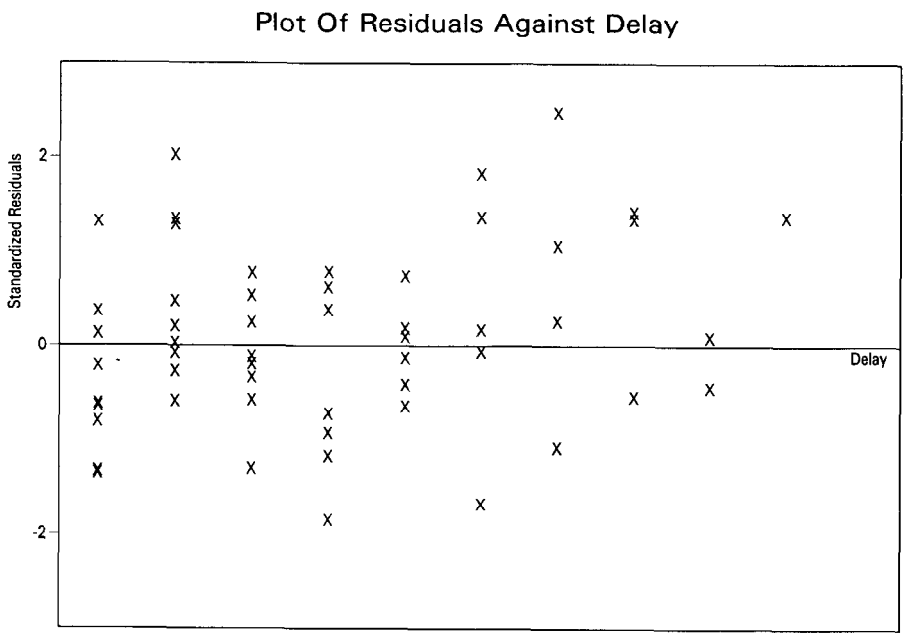
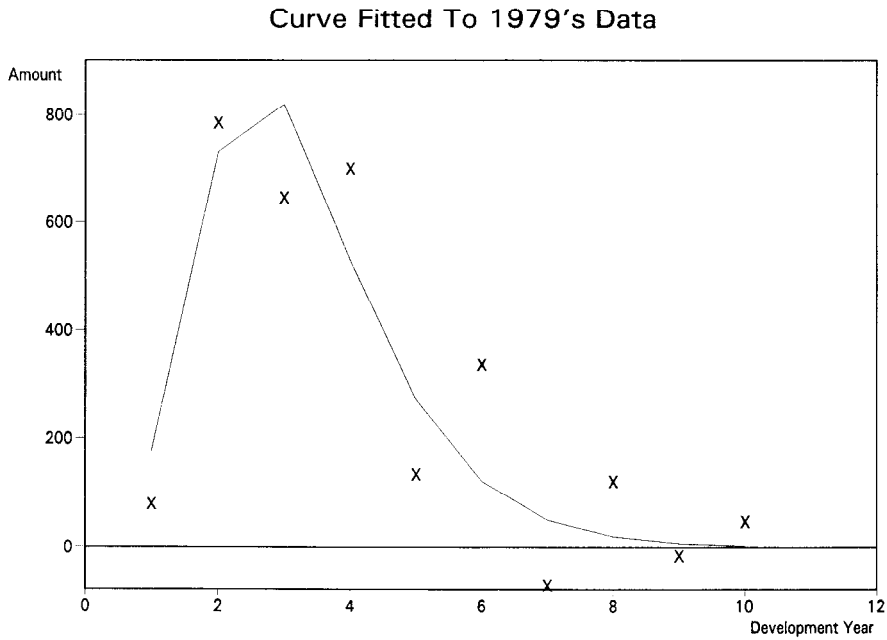

Plot Of Residuals Against Delay
Standardized Residuals
2
0
-2
Delay

Figure 8.2c.


Curve Fitted To 1979's Data
Amount
800
600
400
200
0
0
2
4
6
8
10
12
Development Year

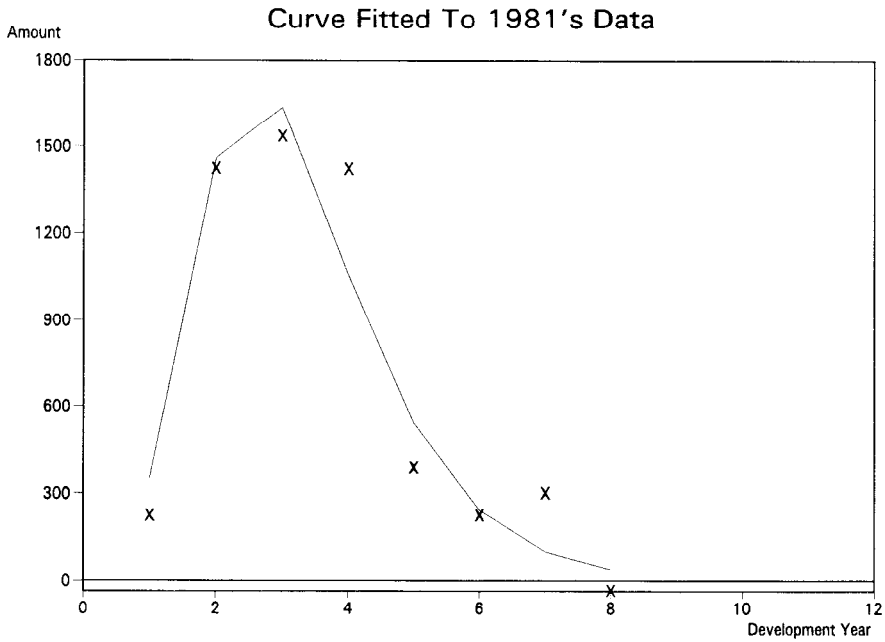

Curve Fitted To 1981's Data
Amount
1800
1500
1200
900
600
300
0
0
2
4
6
8
10
12
Development Year

Figure 8.2d.

These estimates, $\hat{R}$, are substantially less for the later years of origin than those based on the paid data alone. However, the paid estimates did not purport to be very reliable; the differences between the two sets of estimates are no more than about one standard error of the paid estimates.

9. CONCLUDING REMARKS

The main strengths of the method proposed in this paper when compared to other stochastic claims reserving methods are believed to be:

(a) that it requires only data that are normally readily available,
(b) that the model for these data is based entirely (both systematic and random components) on a model of the generating process, and
(c) that it involves no specific assumptions about the distribution of the data or of individual claim severities.

The ideas that the process which generates claims run-off data should be considered in stochastic reserving methods, and that results from risk theory are likely to be relevant, have previously been discussed in general terms by Hayne[6].

The estimation method developed in Section 5 of this paper makes use of Fisher's scoring method and the Kalman filter. Fisher's scoring method is the algorithm used in the well-known statistical package GLIM. The mechanics are given in Appendix 2. For a derivation, Dobson[4] is recommended, whilst McCullagh & Nelder[9] gives the theory in more depth, including a full account of quasi-likelihood.

The Kalman filter is now almost commonplace in claims reserving. Its application in this field appears to have been first suggested by de Jong & Zehnwirth[3], but, in the author's opinion, the best general account remains that of Harrison & Stevens[5]. The Kalman filter is, in essence, no more than the recursive use of Bayes' theorem. In the present context, prior information for the run-off of an origin year is provided by the run-off already estimated for the previous origin year. Bayes' theorem is used to combine this prior information with the data for the origin year in question to give the posterior estimate of the run-off. This, in turn, is taken as the prior estimate for the next origin year.

The assignment of values for the system variances in a dynamic linear model has been considered a problem area in the past in applications of the Kalman filter. The approach adopted in this paper is: first to use each set of observations alone to estimate the corresponding system state, and then to use these estimates to estimate the system variance (Section 5.5). This approach arises naturally in the present application, but, in fact, could be used quite generally in any application of the Kalman filter. The properties of the particular estimation procedure proposed in Section 5.5 have not yet been fully explored; some other procedure may be more appropriate in general.

In previous applications of the Kalman filter in the claims reserving context, the system variances are not necessarily the same between all pairs of origin years

(e.g. Zehnwirth[13]). They are sometimes reduced between later pairs of origin years, if this can be done without adversely affecting the fit, the 'justification' being that the standard errors of the final results diminish as a consequence. This could equally well be done using the present method, but is not recommended. In the example of Section 8, the quality of fit is hardly reduced if all adaptive variances for 1987 to 1988 are set to zero. This is equivalent to assuming that the underlying run-off pattern will be the same for both these origin years. The calculated standard error diminishes, because there are now three data-points for the estimation of this single curve, instead of two for the 87 curve and only one for the 88 curve. However, the calculated value is not valid, because the assumption cannot be justified. The values for the adaptive variances calculated as in Section 5.5 are estimates of the actual amount of variation in the underlying run-off pattern from one origin year to the next, that has occurred in the past. ('Underlying' means after allowing for random variation in individual data-points.) It is usually unjustifiable to assume that the run-off pattern will be more stable in the future than it has been in the past, so the estimated system variances should be used, even if the resulting standard errors are disappointingly large.

If there is no exposure information available the method can still be used. In such a case we have $\tilde{\varepsilon}_W = 1$ for all W, and so the error, ε'_W, represents the whole of the unknown exposure, ε_W (see Section 4.4 for notation). This is subsumed into beta-1, which, in Section 5.2, is assumed to follow a random walk. Thus, the effect of not correcting for exposure is to increase the system variance for beta-1. This, in turn, is likely to increase the standard errors of the reserve estimates, especially for the latest years of origin. This is not a flaw in the method, it is a manifestation of the obvious fact that, with only one or two data-points and not much idea about how the exposure relates to that for better developed years, it is impossible to estimate future payments with any reliability.

Stochastic models for paid claims have previously been proposed with both forms of the systematic component given in Section 4.5. Kremer[8], and several more recent papers (e.g. Renshaw[10] and Verrall[12]) have considered a stochastic version of the Chain Ladder model, whilst Zehnwirth[13] has proposed a stochastic Hoerl curve model. In all these cases, the random component is assumed to be additive and Normally distributed after taking logs of the data.

This assumption allows ordinary linear regression to be used for fitting the model (via the Kalman filter in the case of Verrall[12] and Zehnwirth[13]). However, the present author believes this assumption for the error component of incremental paid claims data to be untenable, that it can lead to very misleading results, and that its invalidity is quite unlikely to be detected using diagnostic tests (see Section 2 and Appendix 4).

The first of the problems of the Chain Ladder model, mentioned in Section 4.5.1 (namely that it does not help with prediction beyond a stage of development already observed) can be overcome by fitting a curve to the development factors estimated from the data. Development factors for stages of development not yet observed can then be estimated by extrapolation (see Craighead[2] and

Sherman[11] for example). However, this procedure is not statistically efficient; if there are reasons to expect the run-off pattern to follow a certain curve, then this model should be invoked at the outset. By first fitting the Chain Ladder model, the data are reduced to a single figure (or perhaps two, using stochastic methods) for each development period; valuable information on the shape of the curve could be lost in the process. Also, it would be difficult to calculate valid standard errors of estimates obtained by projecting development factors in this way.

The procedure described in Section 5.7, of using α values estimated from the early origin years, in modelling the latest origin years, is likely to lead to an overstatement of the reliability of the final results (because uncertainty in these empirical α values is not reflected in the final standard errors). However, it is believed that this effect will be slight, unless f is unusually large or the data are unusually variable.

A recent paper by Jewell[7] considered in detail, and without approximations, the problem dealt with in this paper by the α factors and by using D' instead of D (namely that claim events continue to occur during the first one or more development periods). The approximate solution to this problem adopted in this paper is believed to be adequate when the *exposure interval* is small compared to the length of the run-off. A more exact treatment of this aspect, such as that in Jewell, may be desirable for short-tailed lines of business. However, Jewell has so far considered only claim numbers, not severities, and it is not immediately obvious how to incorporate his work into a practicable claims reserving method.

The assumptions in the model of the claim payment process from which the methods in this paper have been derived are as follows:

(i) that the number of claim payments for each origin year is a Poisson variate,
(ii) that delay to payment is Gamma-distributed,
(iii) that the mean size of payments increases as some power function of delay,
(iv) that the coefficient of variation of severities is the same for all delays, and
(v) that payment delays and sizes are all mutually independent.

Assumption (i) is quite likely to be violated; there may be *contagion*, as described by Beard *et al.*[1]. In property insurance, for example, positive contagion may be caused by weather conditions; we would expect the variance of the number of claims to exceed the mean. Assumption (i) is not a vital assumption, however; any effects of contagion will either be offset by the use of a random measure of exposure (such as the number of claims reported in development year zero), or will be incorporated in the system variance of beta-1. The distribution of the number, N_D, of claim payments falling in development year D is not subject to these contagion effects; the assumed Poisson distribution can be regarded as the usual approximation of the multinomial.

Assumptions (ii), (iii) and (iv) are capable of direct testing. This has not yet been done, but these assumptions have been accepted as plausible by actuaries practising in general insurance who are familiar with this work.

Assumption (v) is known to be false, because separate payments relating to the same claim are obviously not independent. However, the number of claims is invariably large enough for this assumption to provide a good approximation.

As well as investigating the degree of truth of these assumptions, some further work needs to be done to investigate the sensitivity of the method to the truth of the assumptions. This could best be done by simulation.

The method for incurred data, described in Section 7, is not expected to be very widely applicable. For most datasets it is not as easy as in the example of Section 8.2 to find case-estimate bias factors which have the desired effect when used to adjust the incurred data. When these can be found, a range of values will usually do. Unfortunately, in such cases, the final results (estimated ultimates) are often sensitive to the values used. The method can only be used satisfactorily if either: bias factors are known from sources other than the data; or the criteria are satisfied only by small values of the bias factors.

It would be much more satisfactory to have a method which could be applied to both paid and incurred data simultaneously, taking into account the fact that they are both generated by the same set of claims. The development of such a method which is generally applicable is extremely difficult. Modern theory of stochastic processes looks quite promising here.

10. ACKNOWLEDGEMENTS

I would like to acknowledge many valuable comments made during the development of this work by Andrew English, F.I.A., Andrew Smith, B.A., Professor Sidney Benjamin, F.I.A., F.I.S. and Allan Kaufman, F.C.A.S.

REFERENCES

(1) BEARD, R. E., PENTIKAINEN, T. & PESONEN, E. (1969). *Risk Theory—The Stochastic Basis of Insurance*. (3rd Edition). Chapman & Hall

(2) CRAIGHEAD, D. H. (1979). Some Aspects of the London Reinsurance Market in World-Wide Short-Term Business. *J.I.A.* **106,** 286.

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

(4) DOBSON, A. J. (1983). *An Introduction to Statistical Modelling*. Chapman & Hall.

(5) HARRISON, P. J. & STEPHENS, C. F. (1976). Bayesian Forecasting. *Journal of the Royal Statistical Society* (B) **38,** 205.

(6) HAYNE, R. (1989). Application of Collective Risk Theory to Estimate Variability in Loss Reserves. *P.C.A.S.*

(7) JEWELL, W. S. (1989). Predicting IBNYR Events and Delays. *ASTIN Bulletin*, **19,** 25–56.

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

(9) MCCULLAGH, P. & NELDER, J. A. (1983). *Generalised Linear Models*. Chapman & Hall.

(10) RENSHAW, A. E. (1989). Chain Ladder and Interactive Modelling (Claims Reserving and GLIM). *J.I.A.* **116,** 559–587.

(11) SHERMAN, R. E. (1984). Extrapolating, Smoothing and Interpolating Development Factors. *P.C.A.S.* **LXXI,** 122–192.

(12) VERRALL, R. J. (1989). A State Space Representation of the Chain Ladder Linear Model. *J.I.A.* **116,** 589–609.

(13) ZEHNWIRTH, B. (1985). *ICRFS Version 4 Manual and Users Guide*. Benhar Nominees Pty Ltd, Turramurra, N.S.W., Australia.

APPENDIX 1

DEFAULT VALUES FOR D′ AND α

A1.1 Accident Year Data

	D	α	D'
Annual	0	1/2	1/2
	$\geqslant 1$	1	D
Six-monthly	0	1/4	1/2
	1	3/4	5/6
	$\geqslant 2$	1	$D-1/2$
Quarterly	0	1/8	1/2
	1	3/8	5/6
	2	5/8	13/10
	3	7/8	25/14
	$\geqslant 4$	1	$D-3/2$

A1.2 Underwriting Year Data

	D	α	D'
Annual	0	1/6	1/2
	1	5/6	7/10
	$\geqslant 2$	1	$D-1/2$
Six-monthly	0	1/24	1/2
	1	7/24	9/14
	2	17/24	33/34
	3	23/24	73/46
	$\geqslant 4$	1	$D-3/2$
Quarterly	0	1/96	1/2
	1	7/96	9/14
	2	19/96	35/38
	3	37/96	91/74
	4	59/96	187/118
	5	77/96	323/154
	6	89/96	489/178
	7	95/96	673/190
	$\geqslant 8$	1	$D-7/2$

A1.3 Similar values can be obtained for monthly data.

APPENDIX 2

GENERALISED LINEAR MODELS AND METHOD OF SCORING

A2.1 Suppose we have n independent observations, Y_i. If μ_i and σ_i^2 denote the mean and variance of the ith observation, we can write:

$$Y_i = \mu_i + E_i$$

where: $$E(E_i) = 0 \qquad \text{and} \qquad V(E_i) = \sigma_i^2.$$

A *generalised linear model* is a model which relates the means and variances (μ_i, σ_i^2) to a number, p, of parameters, β_j $(p < n)$, through equations of the form:

$$\mu_i = h(\boldsymbol{x}_i^T \cdot \boldsymbol{\beta}) \qquad \sigma_i^2 = v_i(\boldsymbol{\beta})$$

where $\boldsymbol{x}_i$ is a p-vector of known coefficients, and h and v_i are known functions (differentiable, etc.).

A2.2 There is an iterative algorithm known as *Fisher's scoring method* (or just *the method of scoring*), which gives a sequence of estimates, $\hat{\boldsymbol{\beta}}^{(r)}$, of the parameters $\boldsymbol{\beta}'$. ($\hat{\boldsymbol{\beta}}^{(r)}$ denotes the estimates given by the rth iteration.) We often find that this sequence of estimates converges; $\hat{\boldsymbol{\beta}}$ will denote the limit. It can be shown that, if the distributions of the data, Y_i, are from one of the exponential families of distributions (e.g. the Normal family, or the Gamma family), then $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimate of the parameters $\boldsymbol{\beta}$. The algorithm is as follows:

$$\hat{\boldsymbol{\beta}}^{(r+1)} = (X^T \cdot W^{(r)} \cdot X)^{-1} \cdot X^T \cdot W^{(r)} \cdot z^{(r)}$$

where: $$z^{(r)} = X \cdot \hat{\boldsymbol{\beta}}^{(r)} + \frac{Y - h(X \cdot \hat{\boldsymbol{\beta}}^{(r)})}{h'(X \cdot \hat{\boldsymbol{\beta}}^{(r)})} \qquad (n\text{-vector})$$

$$W^{(r)} = \operatorname{diag}\left\{\frac{[h'(X \cdot \hat{\boldsymbol{\beta}}^{(r)})]^2}{v_i(\hat{\boldsymbol{\beta}}^{(r)})}\right\} \qquad (\text{diagonal } n \times n \text{ matrix})$$

$$X = \begin{bmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_n^T \end{bmatrix} \qquad (n \times p \text{ matrix of known coefficients}).$$

It can also be shown that, in the limit, as the sample size, n, tends to infinity, we have:

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{B}, \boldsymbol{V})$$

and a good estimate of V is the inverse of the *information matrix*:

$$\hat{V} = (X^T \cdot W^{(\infty)} \cdot X)^{-1}.$$

If the distributions of the data, Y_i, do not belong to an exponential family, then, if the method of scoring converges, the estimate, $\hat{\boldsymbol{\beta}}$, so obtained, is known as a *quasi-likelihood estimate*. The asymptotic distribution of $\hat{\boldsymbol{\beta}}$ (as the sample size tends to infinity) remains as given above; this justifies application of the method of scoring, regardless of the distribution of the data.

A2.3 In both the models of Section 4.5, we have:

(i) $h(-) = \exp(-)$ and hence $h'(-) = \exp(-)$ also,

(ii) $$V_i(\boldsymbol{\beta}) = \frac{\phi_0}{\tilde{\varepsilon}_W} \cdot e^{\delta T} \cdot D^{\prime\lambda} \cdot h(\boldsymbol{x}_i^T \cdot \boldsymbol{\beta})$$

$$= \tau_i(\lambda) \cdot h(\boldsymbol{x}_i^T \cdot \boldsymbol{\beta}) \qquad \text{(where } \tau_i \text{ is a known function of } \lambda\text{)}.$$

Hence the method of scoring becomes:

$$\hat{\boldsymbol{\beta}}^{(r+1)} = (X^T \cdot W^{(r)} \cdot X)^{-1} \cdot X^T \cdot W^{(r)} \cdot z^{(r)}$$

where:

$$z^{(r)} = X \cdot \hat{\boldsymbol{\beta}}^{(r)} + \frac{Y - \exp(X \cdot \hat{\boldsymbol{\beta}}^{(r)})}{\exp(X \cdot \hat{\boldsymbol{\beta}}^{(r)})}$$

$$W^{(r)} = \operatorname{diag}\left\{\frac{\exp(X \cdot \hat{\boldsymbol{\beta}}^{(r)})}{\tau_i(\lambda)}\right\}.$$

In the fitting procedure, described in Section 5, the method of scoring is initially applied for each origin year, W, separately, and the asymptotic Normal distribution is used to approximate the distribution of the resulting estimates $\hat{\boldsymbol{\beta}}_W$. The later origin years may have very few data points, so the Normal approximation may be poor. However, simulations suggest that the overall accuracy is acceptable.

APPENDIX 3

DYNAMIC LINEAR MODELS AND THE KALMAN FILTER

A3.1 Suppose we have a sequence of data-points, Z_t, where t is the indexing variable for the sequence. (In many applications t represents time.) In general, each data-point, Z_t, is a vector. Suppose an ordinary linear model is believed to be appropriate for each data-point, Z_t:

$$Z_t \sim N(X_t \cdot \boldsymbol{\beta}_t, V_t) \tag{A.1}$$

where: $\boldsymbol{\beta}_t$ is the vector of unknown parameters to be estimated,
X_t is the matrix of known explanatory variables, and
V_t is the assumed variance-covariance matrix for the random component of the data.

Suppose, also, that the unknown parameters are believed to be related for different values of t, as follows:

$$\boldsymbol{\beta}_t \sim N(G_t \cdot \boldsymbol{\beta}_{t-1}, U_t) \tag{A.2}$$

where: G_t is a known matrix, and
U_t is the assumed variance–covariance matrix of the difference between $\boldsymbol{\beta}_t$ and $G_t\boldsymbol{\beta}_{t-1}$.

Thus we have a model specified by two equations: (A.1) is known as the observation equation; (A.2) is known as the system equation. Such a model is known as a dynamic linear model (DLM).

A3.2 The Kalman filter is an algorithm which gives optimal estimates of the parameters, $\boldsymbol{\beta}_t$, of a dynamic linear model. It is a recursive algorithm; one iteration is necessary for each t. Each iteration starts with a probability distribution describing the state of knowledge about the parameters, $\boldsymbol{\beta}_t$, prior to analysis of the data Z_t. The algorithm optimally combines the data, Z_t, with this prior distribution to give a posterior distribution describing the state of knowledge about $\boldsymbol{\beta}_t$ after inclusion of the information contained in Z_t. This posterior distribution for $\boldsymbol{\beta}_t$, together with the system equation (equation (A.2)), gives the prior distribution for $\boldsymbol{\beta}_{t+1}$ for use in the next iteration. For the first iteration, the prior distribution for $\boldsymbol{\beta}_1$ must be obtained from external sources. If there is no prior information about $\boldsymbol{\beta}_1$ the algorithm is initialised using an *uninformative prior*: i.e. the prior distribution has a very large variance.

A3.3 The Kalman filter is given below. It can be shown that this follows from Bayes theorem. The mean and variance-covariance of a distribution for $\boldsymbol{\beta}_t$ are denoted respectively by m_t and C_t, with an additional subscript, $-$ or $+$, to distinguish between prior and posterior distributions.

(i) Suppose the prior distribution for $\boldsymbol{\beta}_t$ is $N(m_{t-}, C_{t-})$.

(ii) The prior estimate of Z_t (the one-step-ahead forecast) is given by:

$$\hat{Z}_t = X_t \cdot m_{t-}. \qquad \text{(from equation A.1)}$$

(iii) The variance–covariance of this prior estimate, $\hat{Z}_t$, is given by:

$$S_t = X_t \cdot C_{t-} \cdot X_t^T + V_t. \qquad \text{(from (i), (ii) and equation (A.1))}$$

(iv) The Kalman gain matrix is given by:

$$K_t = C_{t-} \cdot X_t^T \cdot S_t^{-1}.$$

(v) When the data, Z_t, become available, the prediction error is given by:

$$e_t = Z_t - \hat{Z}_t.$$

(vi) The posterior distribution of $\boldsymbol{\beta}_t$ is $N(m_{t+}, C_{t+})$ where:

$$m_{t+} = m_{t-} + K_t \cdot e_t$$

$$C_{t+} = C_{t-} - C_{t-} \cdot X_t^T \cdot K_t^T.$$

(vii) The prior distribution of $\boldsymbol{\beta}_{t+1}$ is $N(m_{(t+1)-}, C_{(t+1)-})$ where:

$$m_{(t+1)-} = G_{(t+1)} \cdot m_{t+}$$

$$C_{(t+1)-} = G_{(t+1)} \cdot C_{t+} \cdot G_{(t+1)}^T + u_{t+1} \qquad \text{(from equation (A.2)).}$$

A3.4 The direct application of the Kalman filter to a dynamic linear model gives optimal estimates of each set of parameters, $\boldsymbol{\beta}_t$, based on the data received up to and including time t. This is appropriate in control systems where t is real time and the best estimate of the state of the system (represented by $\boldsymbol{\beta}_t$) is required at each time, t, in order to decide on control actions to be made at time t. In such applications, the past states of the system ($\boldsymbol{\beta}_s$ for $s < t$) are not of interest. However, in claims reserving, at each point in time, reserves are required for all past origin years, so we require not only the best estimates of the most recent run-off parameters, but also the best estimates of all past run-off parameters. For this reason, it is appropriate to use the *smoothed Kalman filter*, in which estimates of parameters for past years of origin are affected by data received for later years of origin. Obviously, there can be no causal effect in this direction on the true parameter values, but data for later origin years allow us to improve the estimates of the values for past years, by virtue of the assumed system equation.

There is nothing essentially new in the smoothed Kalman filter: it is simply a matter of rewriting the original model in a different way to give another dynamic linear model, and then applying the ordinary Kalman filter to this.

If $\boldsymbol{\beta}_t^*$ is defined by:

$$\boldsymbol{\beta}_t^* = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_t \end{bmatrix}$$

(so $\boldsymbol{\beta}_t^*$ is a vector of length $t \cdot p$, if each $\boldsymbol{\beta}_t$ is of length p) then the observation equation (equation (A.1)) can be written:

$$Z_t \sim N(X_t^* \cdot \boldsymbol{\beta}_t^*, V_t)$$

where X_t^* is a matrix with $t \cdot p$ columns, the last p columns being X_t and the other columns being zero.

The system equation (equation (A.2)) can be written:

$$\boldsymbol{\beta}_t^* \sim N(G_t^* \cdot \boldsymbol{\beta}_{t-1}^*, U_t^*)$$

where: $G_t^* = \left[\begin{array}{c|c} \mathrm{I} & \\ \hline 0 & G_t \end{array}\right]$ and $U_t^* = \left[\begin{array}{c|c} 0 & \\ \hline 0 & U_t \end{array}\right]$.

This is of the same form as the original model, and so the Kalman filter can be applied to give optimal estimates of $\boldsymbol{\beta}_t^*$ based on all the data up to and including Z_t. But $\boldsymbol{\beta}_t^*$ contains all present and past parameter values, $\boldsymbol{\beta}_t$. In the claims reserving context each iteration brings in data for another origin year to give the best estimates of the run-off parameters for that origin year and all past origin years.

APPENDIX 4

CRITIQUE OF ERROR STRUCTURE IN OTHER STOCHASTIC CLAIMS RESERVING MODELS

A4.1 The model proposed in this paper can be written:

$$Y_D = \mu_D + E_D$$

where: Y_D is incremental paid claims,
μ_D is the systematic component of Y_D, and
E_D is the random component of Y_D with $E(E_D) = 0$

and we have: $\mu_D = \exp(x_D^T \cdot \beta)$ for some known vector x_D

and $$V(E_D) = \phi \cdot \psi_D \cdot \mu_D$$

where ϕ is an unknown constant, but the ψ_D are known.

No further assumptions are made about the distribution of E_D.

A4.2 The stochastic models used in Kremer[8], Renshaw[10], Verrall[12] and Zehnwirth[13] are all of the form:

$$Y_D = \mu_D \cdot \exp(E_D)$$

where: $\mu_D = \exp(\boldsymbol{x}_D^T \cdot \boldsymbol{\beta})$ for some known vector, $\boldsymbol{x}_D$

and E_D is assumed to be Normally distributed, with $V(E_D) = \sigma_D^2$, say.

Taking logs of the model equation gives:

$$\ln Y_D = \ln \mu_D + E_D$$

i.e.: $$\ln Y_D = x_D^T \cdot \beta + E_D.$$

Hence ordinary linear regression (weighted least squares) can be used to fit such models (this is done via the Kalman filter in Verrall[12] and Zehnwirth[13]).

NB: (1) An assumption is generally made about the form of the variance function σ_D^2, but it contains unknown parameters which are estimated from the data.
(2) In a model of this form, μ_D is not the expected value of Y_D as we have; $E(Y_D) = \mu_D \cdot E(\exp(E_D)) = \mu_D \cdot \exp(\frac{1}{2}\sigma_D^2)$.
(3) Such models do not allow for negative values of Y_D, although these are not uncommon.

A4.3 It is well known that the variance of a sum of independent random variables is the sum of their variances. Since each data-point, Y_D, is the sum of all claim payments made in development year D, both the variance and the expected

value of Y_D are proportional to the expected number of payments. We should, therefore, expect $V(Y_D) \propto E(Y_D)$. If individual payments are all mutually independent, and if their distribution does not depend on D, then there is no approximation here; we have exactly $V(Y_D) = \phi \cdot E(Y_D)$ for some constant, ϕ. A numerical example may make this clearer.

Suppose:

(i) the expected number of payments is:
1000 in development year 1
4000 in development year 2.

(ii) the expected size of payments is £10 for both these development years.

Then we have:

$$E(Y_1) = £10{,}000$$
$$E(Y_2) = £40{,}000$$

and if $\phi = 100$:

$$SD(Y_1) = £1000$$
$$SD(Y_2) = £2000$$

(where $SD(\ )$ means standard deviation, i.e. $\sqrt{V(\ )}$).

Thus we have 10% random variation in Y_1, but only 5% random variation in Y_2; with more individual payments in Y_2 there is increased opportunity for random variation in individual payments to cancel out.

A4.4 From A4.1 we see that the model proposed in this paper has:

$$V(Y_D) = \phi \cdot \psi_D \cdot E(Y_D).$$

The factor, ψ_D, appears in order to take account of two complications not considered in A4.3:

(i) inflation may cause the payment severities to depend on D, and
(ii) payment severities in real terms may depend on D.

ψ_D is given by $\psi_D = D'^{\lambda} \cdot e^{\iota \cdot D'/P}$ (the denominator α_D is not necessary here because we are dealing with un-normalised data, Y_D). If effect (i) is absent we have $\iota = 0$ and if effect (ii) is absent we have $\lambda = 0$, so if both are absent we have $\psi_D \equiv 1$, which gives $V(Y_D) = \phi \cdot E(Y_D)$ as in A4.3.

A4.5 Now consider models of the form:

$$Y_D = \mu_D \cdot \exp(E_D) \qquad \text{with} \qquad V(E_D) = \sigma_D^2.$$

Suppose $\sigma_D^2 \ll 1$ (*):

then $E_D \ll 1$ with high probability, so $\exp(E_D) \simeq 1 + E_D$ and, approximately, we have:

$$Y_D = \mu_D \cdot (1 + E_D)$$
$$= \mu_D + E_D \cdot \mu_D.$$

Thus we have an additive error term, $E_D \cdot \mu_D$, with $E(E_D \cdot \mu_D) = 0$ and $V(E_D \cdot \mu_D) = \sigma_D^2 \cdot \mu_D^2$.

Clearly, for this variance to be of the form $\phi \cdot \mu_D$, as argued in A4.3, we must have: $\sigma_D^2 = \phi/\mu_D$. Thus, the reasoning of A4.3 can be satisfied provided μ_D is sufficiently large for all D to satisfy the condition (*) above. However, in general, $\mu_D \to 0$ as $D \to 0$ or $D \to \infty$, so (*) is violated for both small and large values of D, and some other form for σ_D^2 would be appropriate in these ranges. Clearly, in order to use approximately 'correct' weights (according to A4.3) in the regression, the form of the function, σ_D^2, would have to be complex; and this is before allowing for the possibility that the claim severity distribution may vary with delay.

The fitting method proposed for such models is generally as follows:

(i) Initially take $\sigma_D^2 = \sigma^2$ (the same for all D) and fit the model $\ln Y_D = \boldsymbol{x}_D^T \cdot \boldsymbol{\beta} + E_D$ using ordinary linear regression.
(ii) Examine the dependence of the residuals on D (heteroscedasticity), and use this to give a second approximation to the form of σ_D^2.
(iii) Refit the model using weighted linear regression, as dictated by the latest approximation of σ_D^2.

Steps (ii) and (iii) are then iterated until there is no obvious heteroscedasticity remaining in the residuals.

The chances of finding the 'correct' profile (according to A4.3) for σ_D^2 in this way are fairly remote, given the complex form of this function and the small sample size (typically about 50 points, of which there are very few for large values of D because of the triangular shape). The approach generally adopted for step (ii) is to postulate quite a simple functional form and to use the residuals for calibration. For example, Zehnwirth[13] assumes:

$$\sigma_D^2 = \sigma^2 \cdot D^\delta$$

and uses the residuals to estimate σ^2 and δ. It can be shown that this form for σ_D^2 cannot possibly give anything like 'correct' variances for all values of D simultaneously.

Note that the question of correctness of the assumed variances is not of merely academic interest; they determine the weights given to the data-points in fitting any model. The reasonableness of the variance assumptions is, therefore, crucial for the validity of the estimates.