

G. Taylor, G. McGuire and A. Greenfield

ISBN 0 7340 2898 9

Loss Reserving: Past, Present and Future

by

Greg Taylor
Taylor Fry Consulting Actuaries
and The University of Melbourne

and

Gráinne McGuire and Alan Greenfield
Taylor Fry Consulting Actuaries

RESEARCH PAPER NUMBER 109

SEPTEMBER 2003

Centre for Actuarial Studies
Department of Economics
The University of Melbourne
Victoria 3010
Australia

Loss reserving: past, present and future

Greg Taylor
Taylor Fry Consulting Actuaries & University of Melbourne

Gráinne McGuire
Taylor Fry Consulting Actuaries

Alan Greenfield
Taylor Fry Consulting Actuaries

**An invited lecture to the XXXIVth ASTIN Colloquium,
Berlin, 24-27 August 2003**

Table of Contents

Abstract.....	1
1. Taxonomy of loss reserving models	2
2. Adaptive reserving	9
3. From the present to the future	21
4. Conclusion.....	27
References.....	29

Abstract

The paper reviews the development of loss reserving models over the past, classifying them according to an elementary taxonomy. The taxonomic components include (1) the algebraic structure of the model, (2) the form of its parameter estimation, (3) whether or not it is explicitly stochastic, and (4) whether or not its parameters evolve over time.

Particular attention is given to one of the higher species of model, involving complex structure, optimal estimation, and evolutionary parameters. A generalisation of the Kalman filter is considered as a basis of adaptive loss reserving in this case. Real life numerical examples are provided.

Some implications of this type of data analysis for loss reserving are discussed, with particular reference to the form of data set used. The use of triangular arrays is questioned, and alternatives examined. Again, real life numerical examples are provided.

Keywords: loss reserving, dynamic models, Kalman filter, exponential dispersion family.

1. Taxonomy of loss reserving models

1.1 Taxonomic considerations

The taxonomic classification of loss reserving models was considered by Taylor (1986). The main classification variables recognised there were:

- Whether or not the model was stochastic;
- Model structure (whether a “macro-“ or “micro-model”);
- The model’s dependent variables:
 - o Whether paid or incurred losses;
 - o Whether or not claim counts were modelled;
- Whether or not specific explanatory variables were included.

When Taylor came to research a new book on loss reserving (Taylor, 2000), he found that the literature published in the 1½ decades since the previous book was roughly equal in volume to all the prior literature. A number of new techniques had been introduced.

In short, the situation had changed sufficiently to require a re-consideration of taxonomic questions. These questions were not pursued in the book, but are examined here.

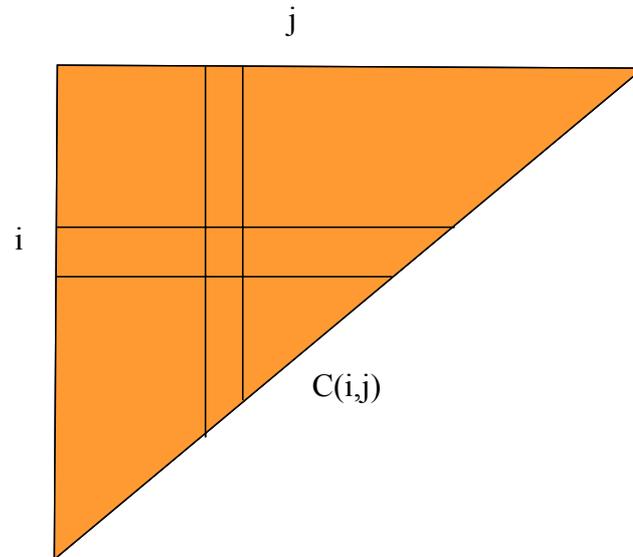
Four dimensions are identified for a re-classification of loss reserving models, namely:

- Stochasticity (as before);
- Dynamism;
- Model structure;
- Parameter estimation.

The space of models may be thought of as a 4-dimensional hyper-cube in these dimensions. Their meaning is now discussed briefly.

For the sake of the discussion, it is assumed that the data being modelled follows the conventional format of a **triangle** of values of some observed claim statistic $C(i,j)$ illustrated in Figure 1.1, where i,j denote accident period and development period respectively.

Figure 1.1
Claims triangle



Stochasticity

For a **stochastic model**, observations are assumed to have the form:

$$C(i,j) = \mu(i,j) + e(i,j) \quad (1.1)$$

where $\mu(i,j)$ is a parameter and $e(i,j)$ an explicitly stochastic error term.

Though it may seem surprising now, not all models have this form. The earlier models, of which the chain ladder was an example, were developed heuristically, and essentially deterministically.

The earliest stochastic models in the literature appear to be those of Kramreiter and Straub (1973) and Hachemeister and Stanard (1975). Others followed, e.g Reid (1978), and by the early 1980's they were appearing rapidly (De Jong and Zehnwirth, 1983; Pollard, 1983; Taylor and Ashe, 1983).

Dynamism

A **dynamic model** is characterised by containing parameters that evolve over time. For example,

$$E[C(i,j)] = \mu(i,j) = f(\beta(i),j) \quad (1.2)$$

where $\beta(i)$ is a parameter vector that evolves as follows:

$$\beta(i) = \beta(i-1) + w(i) \quad (1.3)$$

with $w(i)$ a stochastic perturbation.

Model structure

Whereas stochasticity and dynamism are dichotomous classifications, model structure introduces a spectrum of possibilities, as illustrated in Figure 1.2.

Figure 1.2
Spectrum of model structures



Phenomenological models are those whose structure is not related to aspects of the claim process that have a direct physical meaning. The incurred loss chain ladder model would lie towards this end of the spectrum, with its log-linear structure in terms of somewhat abstract parameters, and its emphasis on age-to-age factors.

Micro-structural models, on the other hand, are concerned with the fine structure of the claims process. An example might be a **statistical case estimation model** (Taylor and Campbell, 2002), i.e. a model of individual claim sizes according to claim-specific covariates such as age of claimant, nature of injury, etc).

This spectrum cannot be regarded as any more than a broad concept. It is difficult to think of a case that lies absolutely at one end or the other. For example, even the chain ladder parameters have some form of physical meaning, such as “proportion of an accident year’s cost paid in a particular development year”, but they are rather remote from the mechanics of the claim payment process itself.

Parameter estimation

There are two main possibilities as to the form of parameter estimation:

- Heuristic;
- Optimal.

Under **optimal estimation** parameters are estimated by optimisation of some statistical criterion, e.g. likelihood.

Heuristic estimation is typical of non-stochastic models; in the absence of a defined stochastic structure, formulation of a suitable objective function for optimisation may be difficult. There are, however, exceptions in which optimal estimation is applied to a non-stochastic structure, e.g. De Vylder (1978).

Prior to Hachemeister and Stanard (1975), the chain ladder model was heuristic in its estimation. These authors placed it on an optimal basis in relation to Poisson claim counts, but it remained heuristic for claim amounts. England and Verrall (2002) placed it on an optimal basis for claim amounts with over-dispersed Poisson distributions. But it remains heuristic for many of its applications.

1.2 Phylogeny

The taxonomy of models defined by the four characteristics described above can be set out in a phylogenetic tree, as in Figure 1.3, where only the more realistic combinations of characteristics are represented. Figures 1.4 to 1.6 mark the main branches of the tree.

Figure 1.3
Phylogenetic tree of loss reserving models

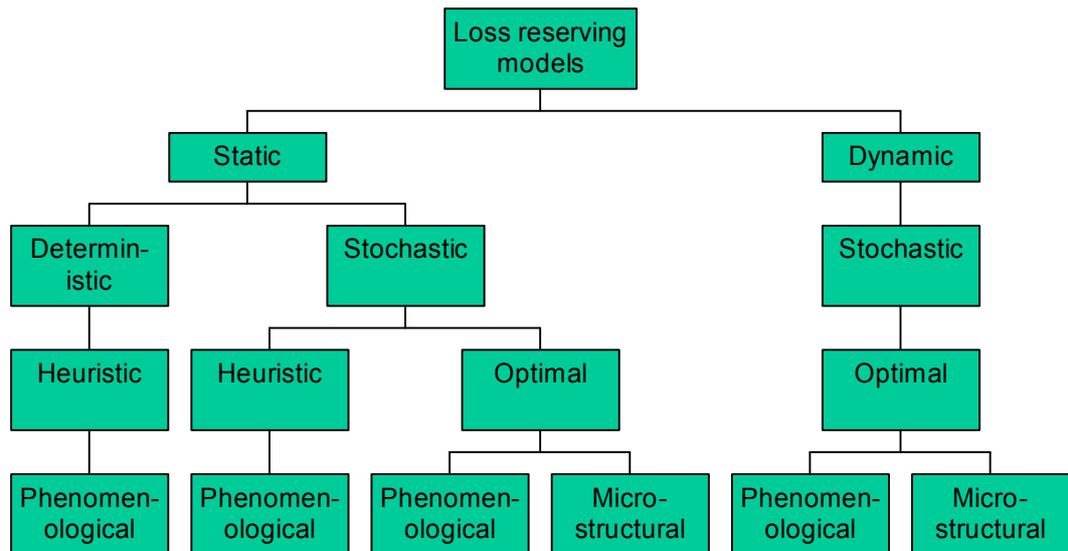


Figure 1.4
Deterministic branch

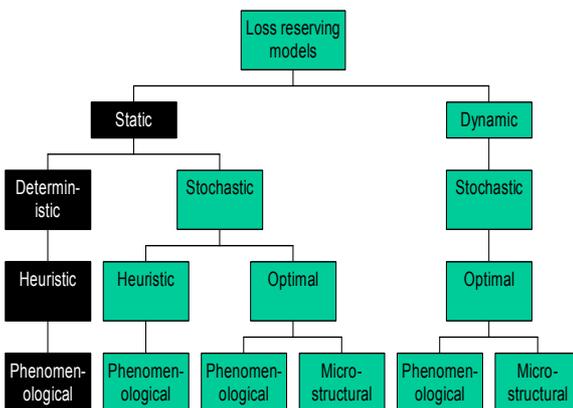


Figure 1.5
Static stochastic branch

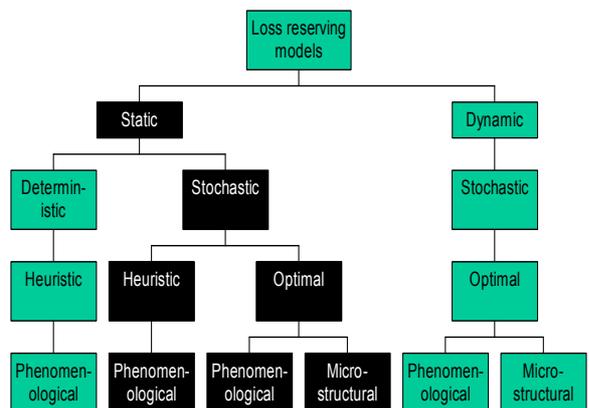
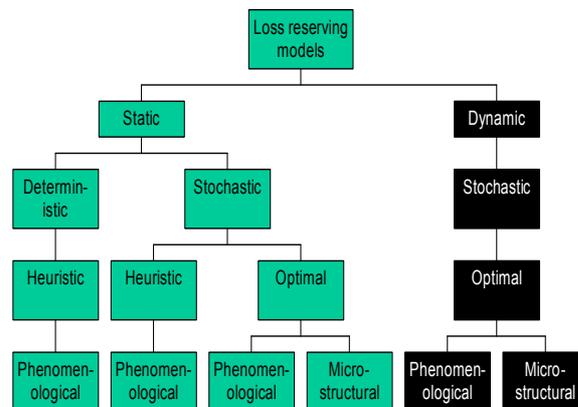


Figure 1.6
Dynamic stochastic branch



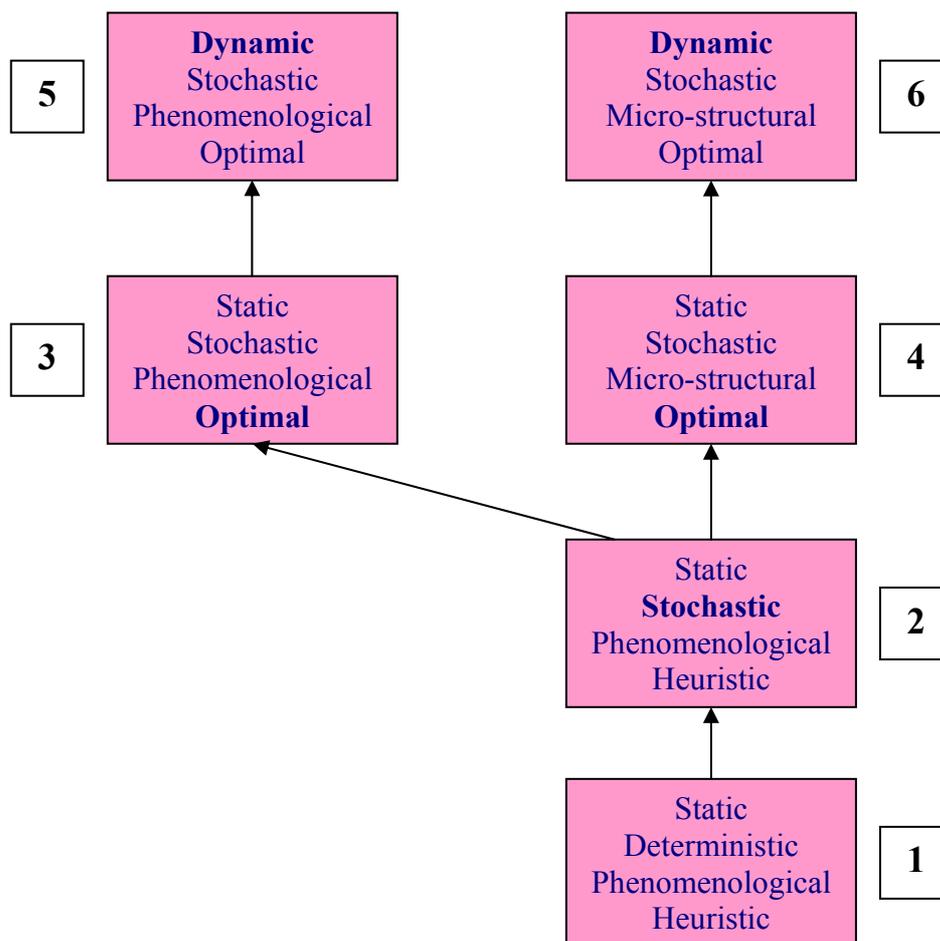
1.3 The Darwinian representation

The Darwinian view of evolution regarded it as a process of progressive improvement from the earliest species to *homo sapiens*. This was the concept of “The Ascent of Man”.

Nowadays, value judgements about the superiority of one species over another are regarded as invalid. One need have no such reservations about the different “species” of loss reserving models, and Figure 1.3 may be organised into a diagrammatic representation of “The Ascent of the Loss Reserving Model”, as in Figure 1.7.

The criterion for the ranking of species here is informal and intuitive. Broadly, however, it might equate to the volume of information encompassed by the model’s output.

Figure 1.7
Darwinian representation of the Loss Reserving phylogeny



The earliest models, most of those up to the late 1970's, fell within **Box 1**. These included methods like:

- The chain ladder (as it was then viewed)
- The separation method (Taylor, 1977)
- The payment per claim finalised model (Fisher and Lange, 1973; Sawkins, 1979).

Any deterministic model may be stochasticised by the addition of error terms. If these terms are left distribution-free, parameter estimation is likely to remain heuristic, as in **Box 2**. An example of such a model is Mack's (1993) version of the stochastic chain ladder.

Alternatively, **optimal parameter estimation** may be applied to the case of distribution-free error terms, as in least squares chain ladder estimation (De Vylder, 1978), to produce a model in **Box 3**.

Optimal parameter estimation may also be employed if an error structure added. Examples are:

- Chain ladder for triangle of Poisson counts (Hachemeister & Stanard, 1975)
- Chain ladder with log normal age-to-age factors (Hertig, 1985)
- Chain ladder with triangle of over-dispersed Poisson cells (England & Verrall, 2002).

Box 4 is similar to Box 3 but with **finer structure** inserted into the model, as in, for example:

- The payments per claim finalised structure of Taylor and Ashe (1983);
- Reid's (1978) model of claim sizes in operational time.

Evolution of parameters over time may be incorporated by means of the Kalman filter (Kalman, 1960). This may be bolted onto many stochastic models, converting Boxes 3 and 4 to 5 and 6 respectively. Examples are:

- Model of distribution of an accident period's claim payments over development periods (De Jong and Zehnwirth, 1983);
- Chain ladder model (England and Verrall, 2002).

The blind application of the Kalman filter can, however, bring some shortcomings with it. This is a subject to be discussed further in Section 2.

2. Adaptive reserving

2.1 Kalman filter

The term **adaptive reserving** will be used to refer to loss reserving based on dynamic models, as defined in Section 1.

The Kalman filter has been noted as a means of converting a static model to a dynamic model. It was introduced into:

- the engineering literature by Kalman (1960);
- the statistical literature by Harrison and Stevens (1976);
- the actuarial literature by De Jong and Zehnwirth (1983).

The model underlying the Kalman filter consists of two equations, called the **system equation** and **observation equation** respectively. The former describes the model's parameter evolution, while the latter describes the model of observations conditional on the parameters. The two equations are as follows.

System equation

$$\beta_{j+1} = G_{j+1} \beta_j + w_{j+1} \quad (2.1)$$

$p \times 1 \quad p \times p \quad p \times 1 \quad p \times 1$

Observation equation

$$Y_j = X_j \beta_j + v_j \quad (2.2)$$

$p \times 1 \quad p \times p \quad p \times 1 \quad p \times 1$

These equations are written in vector and matrix form with dimensions written beneath, and

- Y_j denotes the vector of **observations** made at time j ($=1,2,\dots$)
- β_j denotes the vector of **parameters** at time j
- X_j is the **design matrix** applying to time j
- G_{j+1} is a **transition matrix** governing the evolution of the expected parameter values from one epoch to the next
- w_{j+1} and v_j are stochastic perturbations, each with zero mean, and with

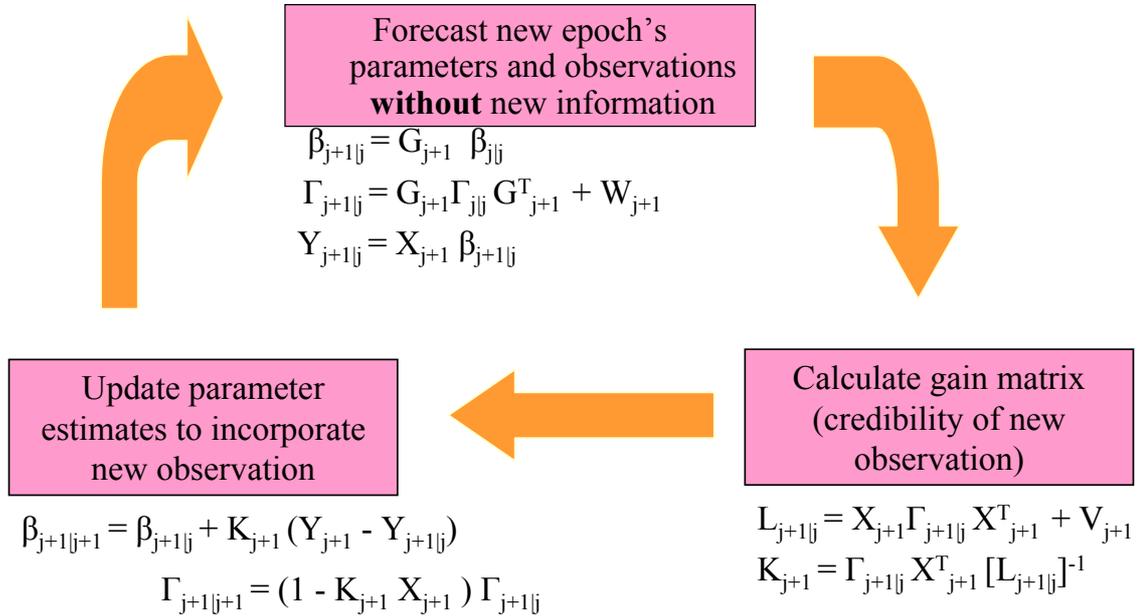
$$V[w_{j+1}] = W_{j+1}, \quad V[v_j] = V_j. \quad (2.3)$$

The Kalman filter is an algorithm for calculating estimates of the β_j . It revises the parameter estimates iteratively over time. Each iteration introduces the additional data Y_j .

The formal statement of the filter requires a little extra notation. Let $Y_{j|k}$ denote the filter's estimate of Y_j on the basis of information up to and including epoch k ; and similarly for other symbols with the subscript $j|k$. Also, let $\Gamma_{j|k}$ denote $V[\beta_{j|k}]$.

Figure 2.1 sets out a single iteration of the Kalman filter.

Figure 2.1
Single iteration of the Kalman filter



To operate the filter, one commences at the top of the diagram with a prior estimate $\beta_{0|0}$ and the associated dispersion $\Gamma_{0|0}$.

The top box generates estimates $\beta_{1|0}$ and $\Gamma_{1|0}$ for the next epoch, still based only on the prior. Working through a complete iteration of the filter introduces the new data vector Y_1 and generates revised estimates $\beta_{1|1}$ and $\Gamma_{1|1}$. The procedure is repeated to generate $\beta_{2|2}$ and $\Gamma_{2|2}$, $\beta_{3|3}$ and $\Gamma_{3|3}$, etc.

The key equation for revising parameter estimates appears under the bottom left box, and is as follows:

$$\beta_{j+1|j+1} = \beta_{j+1|j} + K_{j+1} (Y_{j+1} - Y_{j+1|j}). \quad (2.4)$$

The main characteristics of the formula of note are:

- It is linear in the data vector Y_{j+1} .
- $\beta_{j+1|j+1}$ is the Bayesian revision of the estimate of β_{j+1} in the case that w_{j+1} and v_{j+1} are normally distributed.
- Consequently, K_{j+1} is equivalent to the **credibility** of Y_{j+1} .

2.2 Application of the Kalman filter to loss reserving

Let Y_j denote some vector of loss experience statistics. For example, suppose that $Y_j = (Y_{j1}, Y_{j2}, \dots)^T$ where $Y_{jm} = \log$ [paid losses in (j,m) cell] is and normally distributed. Suppose in addition that

$$E[Y_j] = X_j \beta_j. \quad (2.5)$$

Here paid losses are log normal with a log-linear dependency of expectations on parameters. A specific example, as in De Jong and Zehnwith (1983) or Wright (1990), involves Hoerl curves to represent the distribution of paid losses over development periods m , thus:

$$Y_{jm} = \beta_{j0} + \beta_{j1} \log(m+1) + \beta_{j2} m, \quad m=0,1,2,\text{etc.} \quad (2.6)$$

where $\beta_j = (\beta_{j0}, \beta_{j1}, \beta_{j2})^T$.

The restricted nature of the Kalman filter can cause some difficulties in the formulation of loss reserving models. The linear nature of (2.4) means that the range of parameter estimates is doubly infinite. Hence, by (2.5),

$$Y_{jj} = X_j \beta_{jj} \quad (2.7)$$

also has a doubly infinite range.

This is compatible with log normal Y_j , but not with other distributions. If, for example, Y_j denotes claim count data, negative estimates will be inappropriate.

Distributional difficulties of this sort can sometimes be overcome by the use of transformed data, just as the log normal distribution relies on a log transformation. However, with the transformation comes the need for subsequent inversion, usually with an associated bias correction. The determination of a suitable bias correction, particularly when the distribution of Y_j is only vaguely known, can be problematic.

2.3 Dynamic models with non-normal error terms

Consider the following model as an alternative to that consisting of (2.1) and (2.2), underlying the Kalman filter:

System equation: As in (2.1).

Observation equation: Y_j satisfies a **generalised linear model (GLM)** (see McCullagh and Nelder, 1989) with link function h and linear predictor $X_j \beta_j$, i.e. the distribution of Y_j is chosen from the **exponential dispersion family (EDF)**, and

$$E[Y_j] = h^{-1}(X_j \beta_j). \quad (2.8)$$

When observation equation (2.8) replaces (2.2), a question arises as to how the filter illustrated in Figure 2.1 should be modified.

To answer this question, return to (2.4), and note that it is linear in both the prior estimate $\beta_{j+1|j}$ and the data Y_{j+1} . This leads to the remark, made many times before, that (2.4) may be viewed as a regression of the vector $[Y_{j+1}^T, \beta_{j+1|j}^T]^T$ on β_{j+1} .

Indeed, (2.1) may be stacked with an equation setting out the dispersion of $\beta_{j+1|j}$ as follows:

$$\begin{bmatrix} Y_{j+1} \\ \beta_{j+1|j} \end{bmatrix} = \begin{bmatrix} X_{j+1} \\ 1 \end{bmatrix} \beta_{j+1} + \begin{bmatrix} v_{j+1} \\ u_{j+1} \end{bmatrix} \quad (2.9)$$

where

$$V \begin{bmatrix} v_{j+1} \\ u_{j+1} \end{bmatrix} = \begin{bmatrix} V_{j+1} & 0 \\ 0 & \Gamma_{j+1|j} \end{bmatrix} \quad (2.10)$$

and generally $\Gamma_{j+1|j}$ is not diagonal.

Estimation of β_{j+1} by weighted least squares regression yields the estimate $\beta_{j+1|j+1}$ given by (2.4). This regression is therefore equivalent to the Kalman filter estimation of β_{j+1} .

Moreover, this estimation will be maximum likelihood in the case in which v_{j+1} and w_{j+1} are both normal which, as remarked above, is also the case in which the Kalman filter produces Bayesian revision of the estimate of β_{j+1} .

It is now interesting to consider the case in which the system and observation equations of the present sub-section hold, with v_{j+1} and w_{j+1} from the EDF but not necessarily normal. Then (2.9) is replaced by:

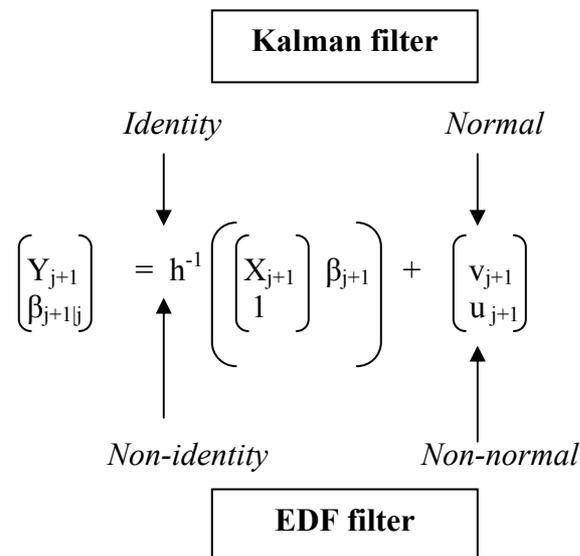
$$\begin{bmatrix} Y_{j+1} \\ \beta_{j+1|j} \end{bmatrix} = h^{-1} \left(\begin{bmatrix} X_{j+1} \\ 1 \end{bmatrix} \beta_{j+1} \right) + \begin{bmatrix} v_{j+1} \\ u_{j+1} \end{bmatrix} \quad (2.11)$$

where the $(n+p)$ -dimensional vector function h is the identity in its last p components, but not necessarily in its first n . Equation (2.10) continues to hold.

By (2.9), the variate u_{j+1} is the centred version of $\beta_{j+1|j}$, whose computation for the present case has not yet been specified. Therefore, it may or may not be from the EDF. Suppose, however, that it may be approximated by a member of the EDF. Then (2.9) holds with the stochastic error terms from the EDF.

Apply the term **EDF filter** to an algorithm that parallels the Kalman filter but applies to a model conforming with (2.11). It is evident that the Kalman filter is a special case. The situation is illustrated by Figure 2.2.

Figure 2.2
Kalman filter and EDF filter

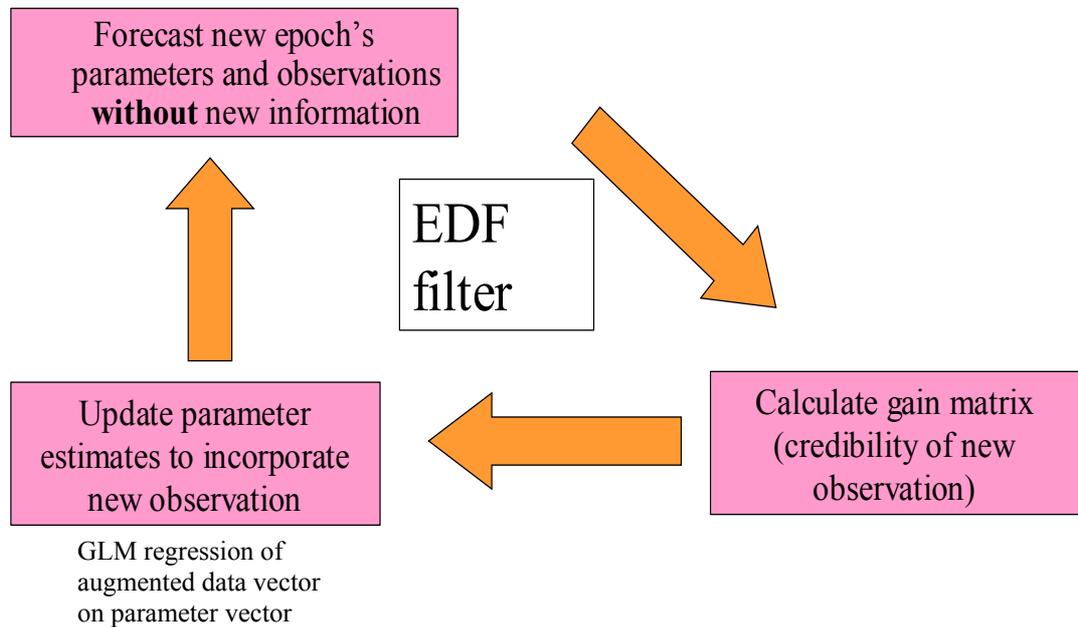


Note that v_{j+1} and u_{j+1} may be from different sub-families of the EDF, e.g. v_{j+1} Poisson and u_{j+1} gamma. For simplicity, however, consider first the case in which both come from the same sub-family P_1 , e.g. v_{j+1} and u_{j+1} both gamma.

As previously remarked in connection with the Kalman filter, $\beta_{j+1|j+1}$ is obtained by weighted least squares regression of the vector $[Y_{j+1}^T, \beta_{j+1|j}^T]^T$ on β_{j+1} . In the case of the EDF filter, it may be obtained by GLM regression with link function h and error term from P_1 .

Figure 2.3 generalises Figure 2.1 to accommodate the EDF filter. To reduce this to the special case of the Kalman filter, replace the GLM regression mentioned at the foot of the diagram by least squares regression.

Figure 2.3
EDF filter

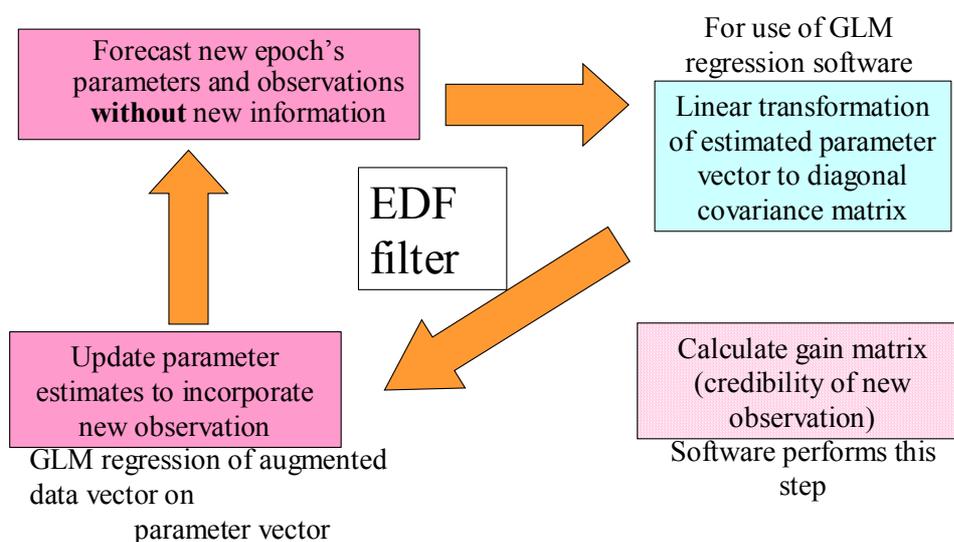


It is usual in GLM applications for all components of the error vector to be chosen from the same sub-family of the EDF, as was assumed above. This assumption is certainly made within GLM software packages such as SAS GENMOD, GLIM, EMBLEM, etc. It is not essential, however.

The same packages assume that all components of the error vector are stochastically independent. If they are to be used to carry out the regression stage of Figure 2.3, then it will be necessary first to diagonalise the covariance matrix $\Gamma_{j+1|j}$ appearing in (2.10). The full details of the regression will be given in a forthcoming paper (Taylor and McGuire, 2003).

With this step included in the procedure, the EDF filter may be represented diagrammatically as in Figure 2.4. The bottom right box in this diagram is shown as stippled, and by-passed by the process. This reflects the fact that this step is implicit in the regression when it is carried out by a software package.

Figure 2.4
EDF filter implemented with GLM software



There is a substantial literature dealing with the properties of estimates produced by the EDF filter for the simple special case of 1-dimensional observations ($n=1$). When the prior at $j=0$ is suitably conjugate to the EDF sub-family from which the distribution of data points is taken, the estimates are Bayes (Jewell, 1974; Nelder and Verrall, 1997; Landsman and Makov, 1998).

For $n>1$, the parameter estimates produced by the EDF filter will **not** be Bayes in general. They can, however, be shown to approximately Bayes.

A further conjecture is that they may be derivable as a form of **stochastic approximation**. Landsman and Makov (1999, 2003) have demonstrated a stochastic approximation interpretation of credibility formulas, which, like Kalman filter parameter estimates, are also linear Bayes.

The properties of EDF filter estimates are to be studied in the forthcoming paper (Taylor and McGuire, 2003) referred to above.

2.4 Numerical examples

2.4.1 Example 1: workers compensation payments

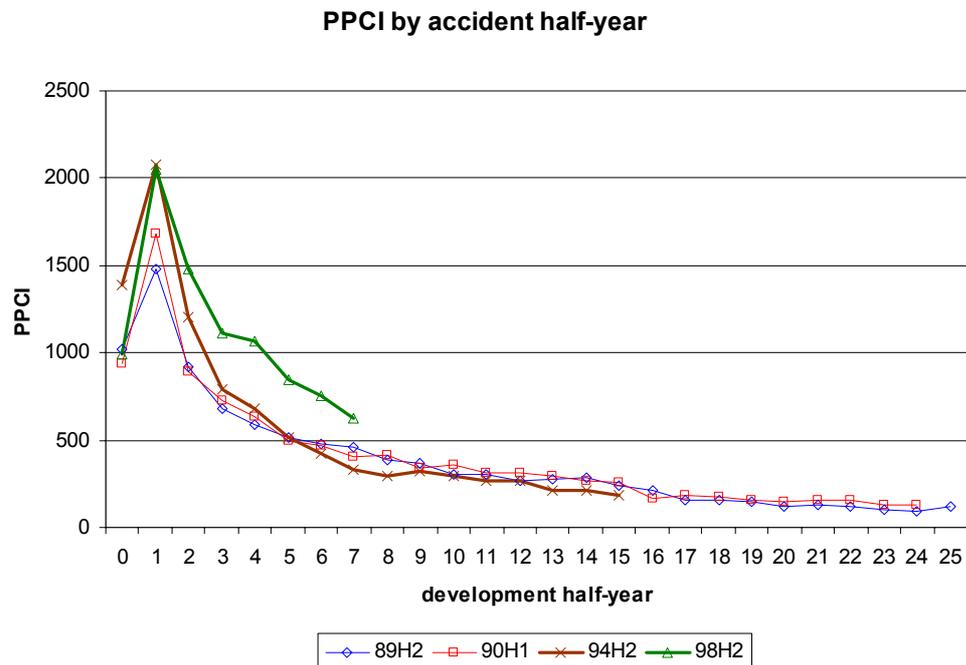
This is an example of a workers compensation portfolio, in which benefits are dominated by payments of weekly compensation. Paid loss experience has been corrected for past inflation and recorded according to accident half-years and development half-years.

The data are represented in a triangle of **payments per claim incurred** (PPCI), i.e. paid losses in each cell have been divided by the number of claims estimated to have been incurred in the relevant accident half-year.

Simple inspection of the data triangle reveals that there have been gradual but systematic shifts in PPCI from one past accident period to the next. Figure 2.5 illustrates this, displaying the PPCI of four accident half-years dispersed in time. The evident features in these data are:

- The steady upward movement of PPCI at the low development half-years;
- The dramatic upward movement in later accident periods of PPCI in the mid-range of development half-years.

Figure 2.5
PPCI by development half-year for various accident half-years



These changes are modelled using an EDF filter with log link and gamma error. Data vectors relate to accident half-years, so that the progress of the filter is as illustrated in Figure 2.6.

Figure 2.6
Diagrammatic illustration of application of filter



The observation equation takes the generic form contained in (2.11), with Y_j denoting the vector of PPCIs from accident half-year j , and the linear response term $X_j \beta_j$ has m -th component

$$\beta_{j0} + \beta_{j1} (m+1) + \beta_{j2} / (m+1) + \beta_{j3} / (m+1)^2 + \beta_{j4} \delta_{m0}, \quad m=0,1,2,\text{etc.} \quad (2.12)$$

and with δ denoting the Kronecker delta.

Figures 2.7 to 2.13 display data and filtered results for selected accident half-years, showing how the models fitted to individual accident periods track the data.

Figure 2.7
Initiation of the filter

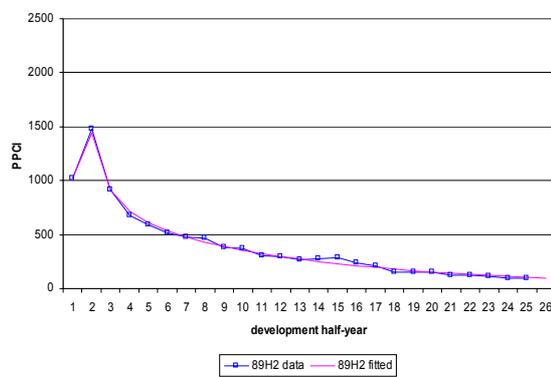


Figure 2.8
The next accident half-year's data

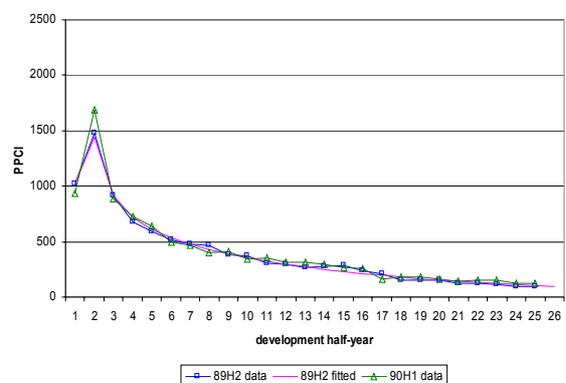


Figure 2.9
Filtered results

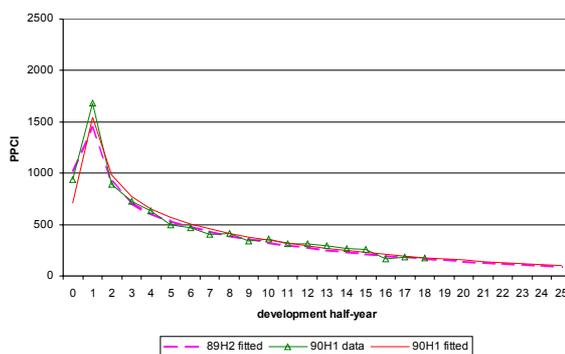


Figure 2.10
More filtered results

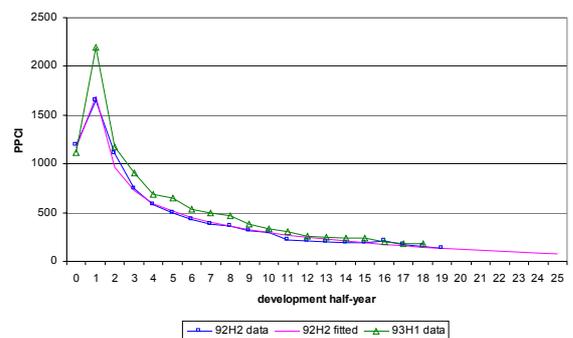


Figure 2.11
More filtered results

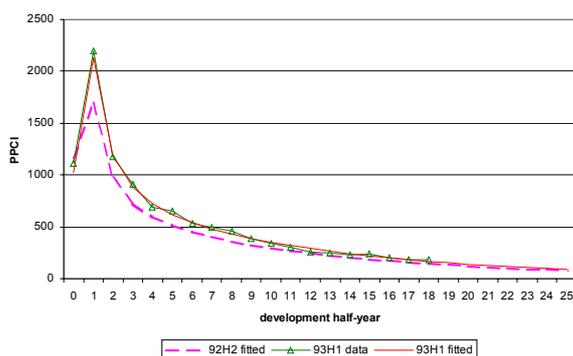


Figure 2.12
More filtered results

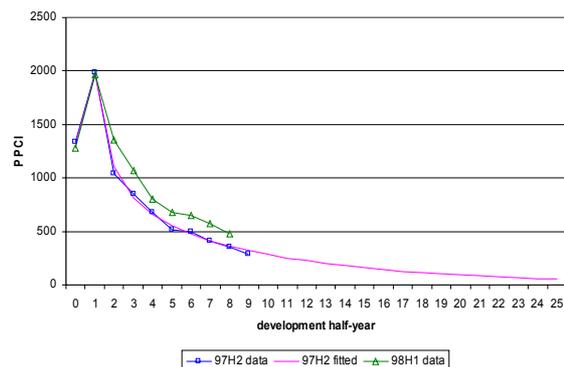
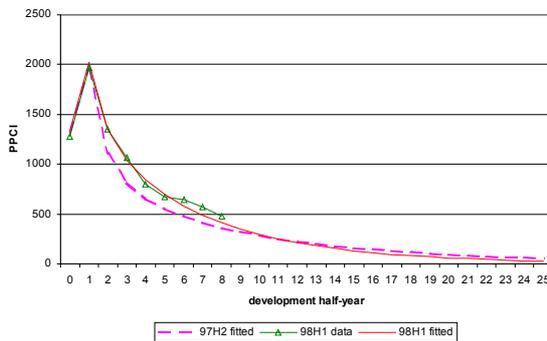


Figure 2.13
More filtered results



2.4.2 Example 2: motor vehicle bodily injury claim closure rates

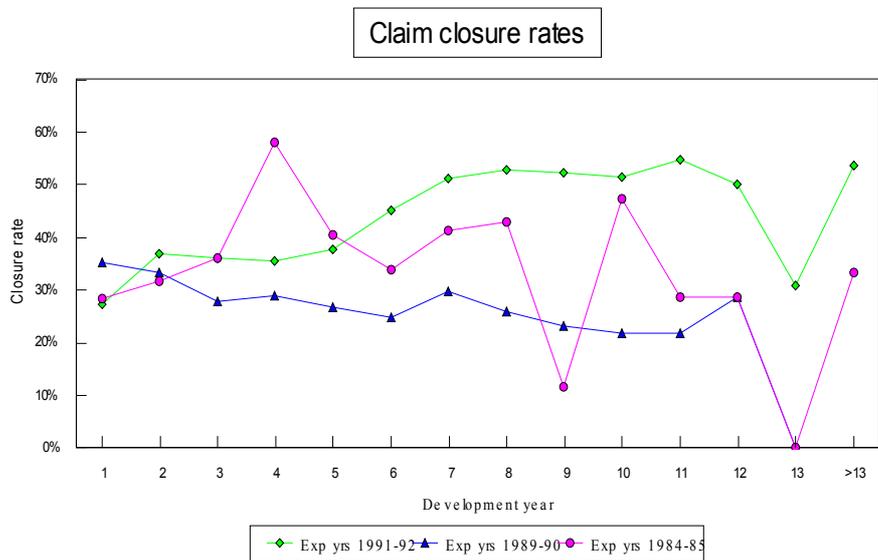
This example is taken directly from Chapter 10 of Taylor (2000), and uses the data set that underlies the numerical examples from that book.

The data are represented in a triangle of **claim closure rates** in which accident and development periods are both annual. As in the reference, the closure rate in any cell is calculated as:

$$\frac{\text{Number of claims closed in the cell}}{\text{Number open at the start} + \frac{1}{3} \times \text{number newly reported in cell.}}$$

Inspection of the data triangle reveals that closure rates are relatively flat over development years, but are subject to upward or downward shocks that tend to affect whole experience years (diagonals). Figure 2.14 illustrates, displaying closure rates for three separate experience years.

Figure 2.14
Claim closures rates for different experience years

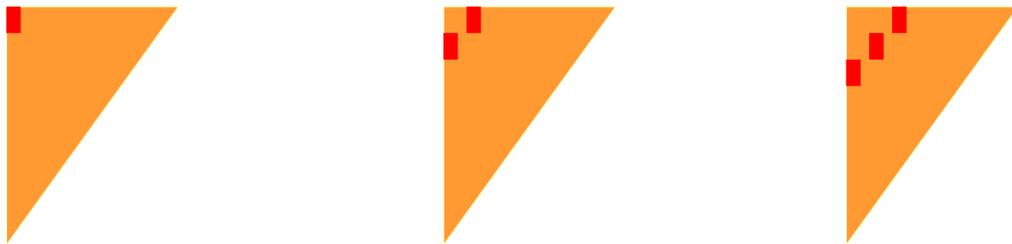


It is evident that numbers-based closure rates should be modelled with a discrete error structure, such as binomial. However, in the present case, most of the numbers of closures underlying the closure rates are sufficiently large that a normal approximation suffices.

Hence, the EDF filter is applied to the rates with identity link, normal error, and weights equal to the denominators of the rates. This choice of link and error structure reduces the EDF filter to the Kalman filter.

Since the shocks hit diagonals, the diagonals of the data triangle are taken as the data vectors, so that the progress of the filter is as illustrated in Figure 2.15.

Figure 2.15
Diagrammatic illustration of application of filter



Let C_{jm} denote the closure rate for accident year j and development year m . The model takes the form:

$$C_{jm} = \beta_{j0} + \beta_{j1} / (m+1) + \beta_{j2} / (m+1)^2 + \gamma_k \delta_{j+m,k}, \quad m=0,1,2,\text{etc.} \quad (2.13)$$

where γ_k denotes the shock to experience year k .

When filtering takes place along diagonals in the manner illustrated, each data vector Y_k (now subscripted by experience year k , instead of j) takes the form:

$$Y_k^T = [C_{0k}, C_{1,k-1}, C_{2,k-2}, \dots, C_{k0}]. \quad (2.14)$$

The vector of parameters involved in this is the stacked vector

$$[\beta_0^T, \beta_1^T, \beta_2^T, \dots, \beta_k^T, \gamma_k]^T, \quad (2.15)$$

where β_j^T denotes $[\beta_{j0}^T, \beta_{j1}^T, \beta_{j2}^T]^T$. That is, at the iteration of the filter that incorporates data diagonal k the parameter vector includes the parameters associated with **all rows** up to row k .

This defines the observation equation for this example. This form of model was introduced by Verrall (1989). Full details of its application to the present example are given in Chapter 10 of Taylor (2000).

The system equation (2.1) takes the form:

$$\beta_{j+1} = \beta_j + w_{j+1}^{(\beta)} \quad (2.16)$$

$$\gamma_{k+1} = \gamma_k + w_{k+1}^{(\gamma)} \quad (2.17)$$

Figures 2.16 to 2.19 display data and filtered results for selected experience years. Each graph displays the closure rates of specific experience years, plotted against development year.

Figure 2.16
Adding the 1991 data diagonal

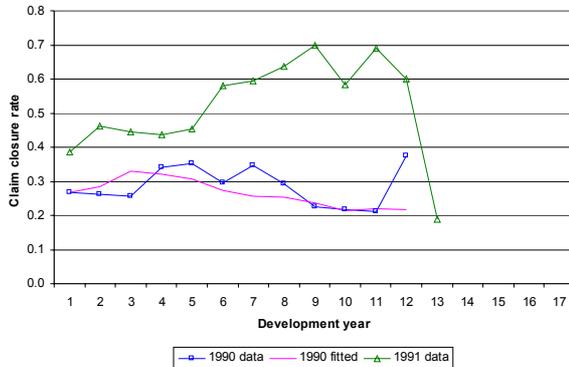


Figure 2.17
Fitting the 1991 diagonal

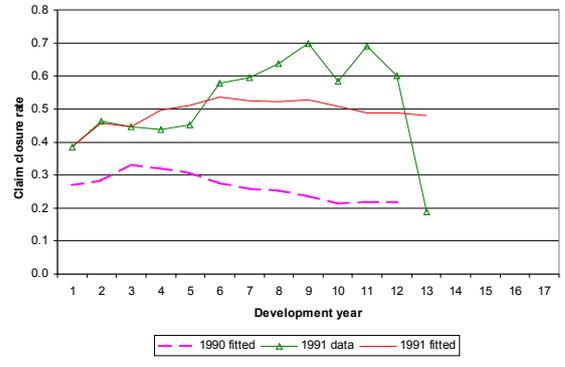


Figure 2.18
Adding the 1992 data diagonal

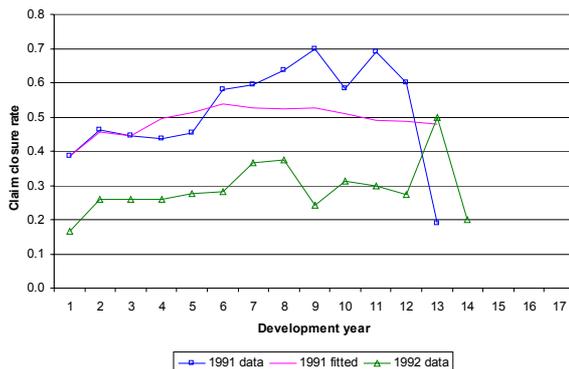
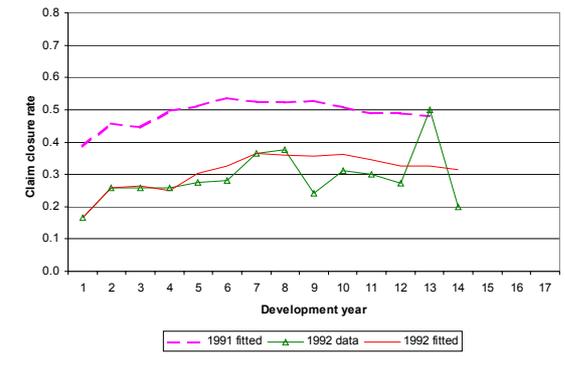


Figure 2.19
Fitting the 1992 diagonal



The examples deal with the relatively high closure rates of 1991, and the much lower ones of 1992. The filtered results follow the data closely in their general level, i.e. there is little smoothing of the diagonal effects. There is, however, considerable smoothing across development years.

Note that the fitted curves shown in these figures are not quite smooth. This is because each represents a diagonal, hence a cross-section of accident years. Consecutive points along one of these curves represent different accident years, and therefore, according to (2.16), are subject to different β parameters in (2.13).

3. From the present to the future

3.1 Loss experience triangles

All discussion of loss reserving models to this point has been framed within the context of the triangle set out in Figure 1.1. This format of data is fundamental to the loss reserving literature. Indeed, the literature contains little else.

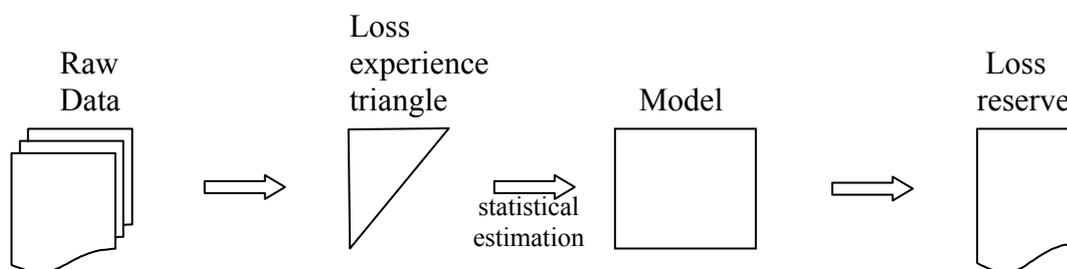
The raw data underlying the triangle is, of course, much more extensive, relating as it does to individual claims. Typically, the raw loss data might comprise:

- A unit record claim header file, containing static information about the claim, such as claim number, name of claimant, date of occurrence, date of notification, etc.
- A unit record claim transaction file, containing a record for each payment made, each change of claim status (e.g. from open to closed), each change of estimate of incurred loss, etc.

These are the real data. The triangle is a summary, whose origins are very much driven by the computational restrictions of a bygone era.

The position of the triangle in the general schema of the construction of a loss reserve is as illustrated in Figure 3.1.

Figure 3.1
General schema for construction of a loss reserve



The first stage of this procedure is essentially one of simplification, in which a data set of many dimensions is reduced to one of much lower dimensionality.

The next stage, labelled “statistical estimation”, may be sophisticated or not, depending on where the loss reserving model fits into the “Model structure” and “Parameter estimation” dimensions of the phylogeny discussed in Section 1. If, however, the model is such that this stage involves a high degree of complexity, questions may well arise about the appropriateness of a pre-processing simplification stage.

The data summary stage of the loss reserving procedure, and even the loss experience triangle, is indeed integral to some of the more phenomenological model structures. Moreover, it may well be a useful data exploration tool at the stage at which one seeks to gain a basic understanding of the data in order to formulate a model.

However, it is difficult to avoid the conclusion that one of its major motivations is the compression of data to meet the computational needs of a bygone era. Nowadays, when it is quite customary to use statistical software to process large data sets using sophisticated models, it seems natural to think of loss reserving models adapted to unit record claim data sets.

Indeed, one can imagine future generations of students, educated on the basis of such models, finding the compression of data into a triangle quite artificial. This theme is pursued further by means of another numerical example.

3.2 Example 3: a non-triangular numerical example

This example deals with another motor vehicle bodily injury portfolio. The data set comprises unit record data on all claims closed at non-zero cost, each record including:

- accident quarter
- closure quarter
- operational time at closure
- cost of claim.

Here, all payments contributing to the cost of the claim are inflation corrected on the basis of some agreed index. The operational time at closure of claim X is, as usual, equal to the proportion of claims from X's accident quarter closed before it.

Consistently with the previous example, let

j = accident quarter
 m = development quarter
 $k = j+m$ = experience quarter.

In addition, let

t = operational time at closure
 $C(t,j,k)$ = claim cost.

Assume that $C(t,j,k)$ is subject to a quasi-likelihood from the EDF with

$$V[C] = \varphi \{E[C]\}^p \quad (3.1)$$

and $p=2.3$. This makes claim size substantially longer tailed than gamma ($p=2$).

The model will take the general form:

$$E[C(t,j,k)] = \exp[f(t,j) + g(t,k)] \quad (3.2)$$

for functions $f(t,j)$ and $g(t,k)$ whose form is yet to be specified. First, however, note their purposes.

The function $f(t,j)$ is primarily a description of the pattern of claim size by operational time. Generally, the higher the value of t , the greater the expected claim size. The dependency on accident quarter recognises a discrete change in the form of the model at $j=j_0$ on account of changes to legislation affecting claim sizes. That is, $f(t,j)$ takes one form for $j \leq j_0$, and another for $j > j_0$.

The function $f(t,j)$ depends on a sub-vector $\beta^{(f)}$ of the full parameter vector β which evolves according to (2.16), i.e. with changes from one **accident period** to the next.

The function $g(t,k)$ is primarily a description of superimposed inflation through its dependency on k , e.g. $g(t,k) = \gamma k$ would represent superimposed inflation at a constant rate of $e^\gamma - 1$ per quarter. The dependency on t allows the rate of superimposed inflation to vary with operational time, reflecting the fact that in this portfolio claims of small to medium injury severity tend to inflate rapidly whereas high severity claims do not.

The function $g(t,k)$ is given the more specific form:

$$g(t,k) = k h(t) \quad (3.3)$$

where the appearance of k as a multiplier indicates the superimposed inflation effect, and

$$h(t) = X^{(h)} \beta^{(h)} \quad (3.4)$$

the superscript (h) indicating that the design relates to just the function $h(t)$ in (3.3) and (3.4) and $\beta^{(h)}$ is another sub-vector of the full parameter vector β .

Let $\beta_k^{(h)}$ denote the dynamic form of this sub-vector relating to experience period k . Its dynamics are as follows:

$$\beta_k^{(h)} = \beta_{k-1}^{(h)} + \varepsilon_k^{(h)} \quad (3.5)$$

with the $\varepsilon_k^{(h)}$ stochastically independent error terms.

Note the difference between (3.5) and (2.17). In the latter the diagonal effect is subject to upward and downward shocks, whereas in the former the **rate** of superimposed inflation $h(t)$ receives the shocks. This means that the diagonal effect will evolve more smoothly than in the earlier example.

The form (3.5) is also better adapted than (2.17) to a conventional understanding of superimposed inflation, whose rate is precisely the first difference of the diagonal effect in the linear response in a log-linear model (as here).

Note also the difference in evolution of $\beta^{(f)}$ and $\beta^{(h)}$ in that one evolves over rows, the other over diagonals.

This example illustrates a number of aspects of the power of the EDF filter, specifically:

- it models a large data set with a variety of structural complexities;
- it incorporates a long tailed distribution of individual claim size;
- it does so dynamically, with a subset of the parameters evolving over accident periods and another subset over experience periods;
- all of this can be expressed in concise and relatively simple mathematical form.

As to the last point, the model is applied to data which, if put in triangular form, would form a 35x35 triangle, i.e. with 630 quarterly data cells. However, with some tens of thousands of claims and these labelled by operational time rather than development quarter, the data input has much greater dimension than this. Yet it is adequately modelled with 10 parameters:

- 7 describing the dependency on operational time of closure ($f(t,j)$ in (3.2)); and
- 3 describing superimposed inflation ($g(t,k)$ in (3.2)).

In the application of the filter to this structure, each data vector relates to a new diagonal of claim closure, but consists of the vector of sizes of individual claims closed in the quarter, rather than a summary of these into development quarters or the like.

Figures 3.1 to 3.4 illustrate the model fit of the filter as it stood at a couple of different closure quarters. In each of these plots:

- There is a comparison of actual and model claim sizes by development quarter.
- For purposes of comparison, the actual claim size (appearing in the plots as “data”) is the average observed claim size in the development quarter. Model claim size is the fitted claim size at the average operational time in that development quarter.
- All claims in the plot are subject to the same superimposed inflation parameters $\beta^{(h)}_k$ since they have the same value of k though, as seen in (3.3), superimposed inflation depends on operational time.

Figure 3.1
Adding the December 1997 quarter diagonal

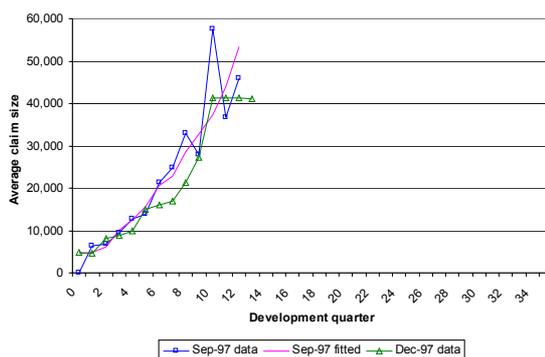


Figure 3.2
Fitting the December 1997 quarter diagonal

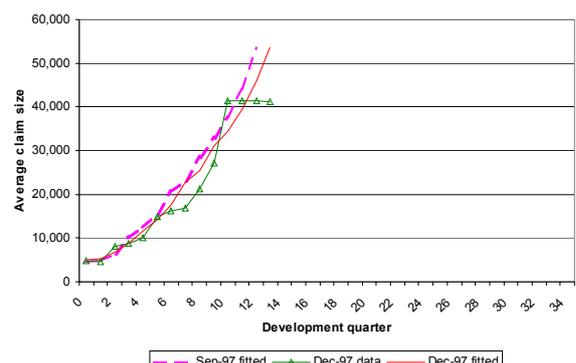


Figure 3.3
Adding the December 2000 quarter diagonal

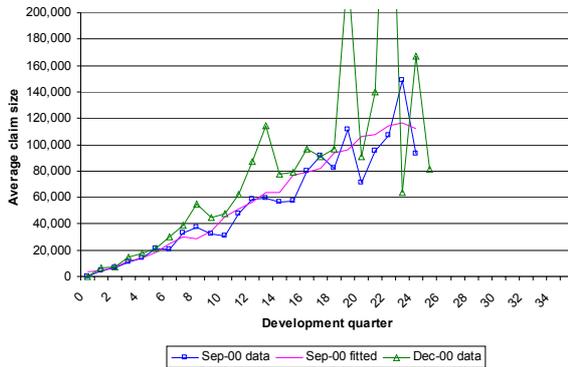
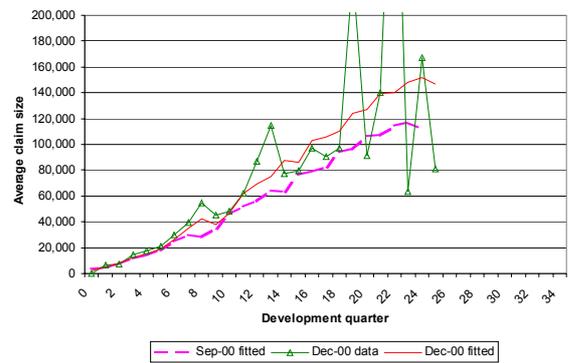


Figure 3.4
Fitting the December 2000 quarter diagonal



It is interesting to examine the trends in the filtered estimates of superimposed inflation parameters. The rate of superimposed inflation is modelled as a piecewise linear function of operational time with a peak at operational time 10 (%).

Other analysis had suggested:

- Low superimposed inflation up to September 2000;
- A much higher rate thereafter; and
- A further increase in the rate after March 2002.

Figures 3.5 to 3.10 track the shift in the estimated rate of superimposed inflation over the critical period from June 2000 to March 2003. Each plot shows the filtered estimates over three successive quarters. For the most part, successive plots overlap by one quarter.

The general upward drift is visible over the course of the plots. For example, from the June 2000 to the March 2003 quarters, the estimated rate of superimposed inflation:

- Changes from about 6% p.a. to 7% p.a. at operational time 10% (its peak);
- Changes from about -4% p.a. to +3% p.a. at operational time 50%.

Figure 3.5
Filtered estimates of superimposed inflation June to December 2000 quarters

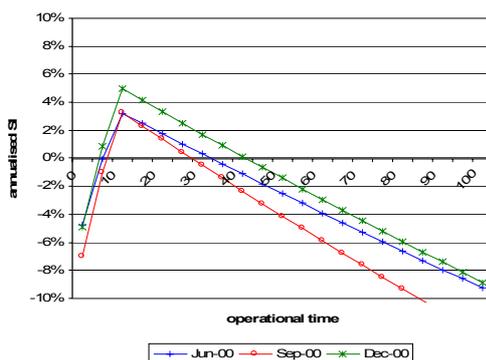


Figure 3.6
Filtered estimates of superimposed inflation December 2000 to June 2001 quarters

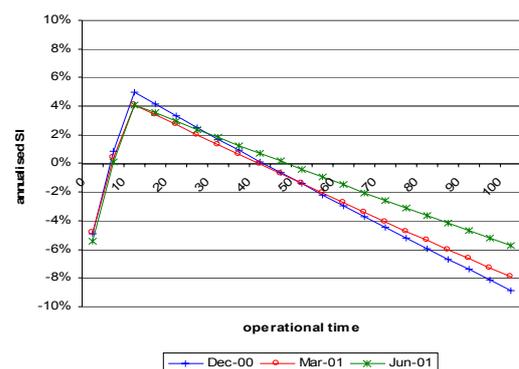


Figure 3.7
 Filtered estimates of superimposed inflation
 June to December 2001 quarters

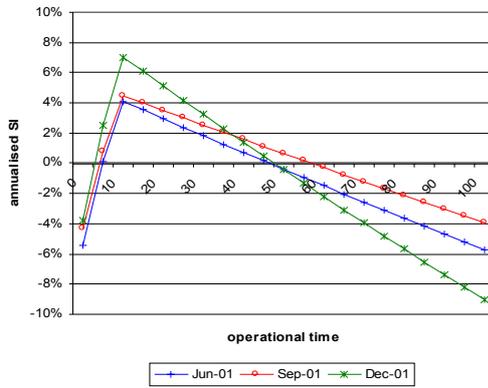


Figure 3.8
 Filtered estimates of superimposed inflation
 December 2001 to June 2002 quarters

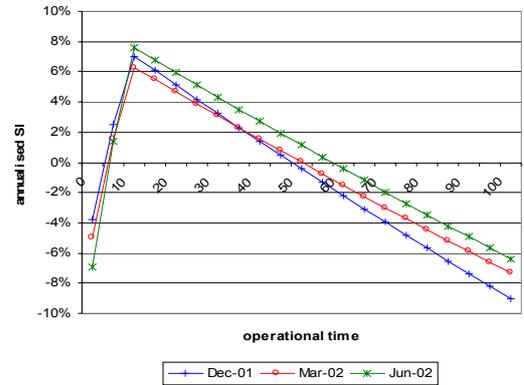


Figure 3.9
 Filtered estimates of superimposed inflation
 March to September 2002 quarters

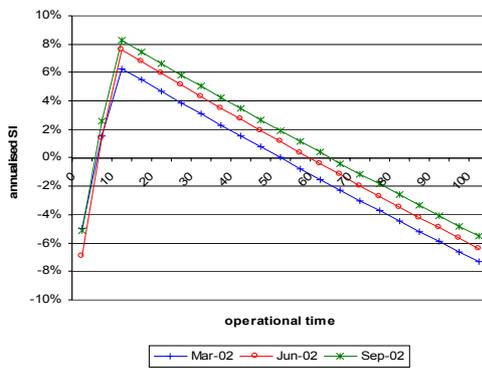
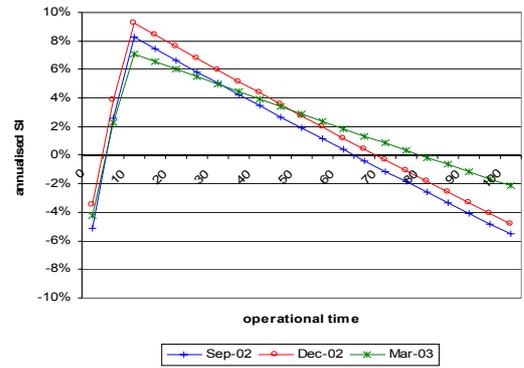


Figure 3.10
 Filtered estimates of superimposed inflation
 September 2002 to March 2003 quarters



4. Conclusion

While triangular data formats may be useful tools for data exploration, there is no apparent reason why they should be retained as the structures for data analysis. It may often be preferable to model unit record claim data, or at least a considerably more comprehensive data structure than the conventional triangle.

Section 3 gives an example to illustrate this point. The example is in many ways typical of loss reserving exercises involving large data sets. The main characteristics are that:

- The volume of data is sufficient that a number of subtleties in the algebraic structure to be estimated become manifest, e.g.
 - Dependency of the rate of superimposed inflation on some variables, rather than just a flat rate;
 - Changes in the relation between the dependent variable and the measure of development time (operational time in the example), possibly deriving from legislative change, or other specific change to the external claims environment;
- With the manifest nature of the fine structure of the data comes the likelihood that its underlying parameters will shift over time.

The first characteristic motivates a detailed model structure such as a GLM. The second requires that time variation be incorporated in that GLM. One can, of course, include time-based interactions directly in a GLM. This is, however, rather different from converting it to a dynamic model by randomisation of parameters.

On the one hand, the inclusion of time-based interactions in a GLM implies a model in which the time-dependent changes in model structure are assumed **known with certainty**, and only their parameters remain to be estimated.

Shifts in underlying parameters will initially be ignored by such a model. Over time, the user will observe the failure of the model to fit more recent data, and will revise the model's algebraic structure. This will create a fundamental discontinuity in the sequence of estimates, such as loss reserves, flowing from the model.

On the other hand, the dynamic models considered in the foregoing sections are of a Bayesian nature, and so result in more gradual recognition of underlying change. The degree of recognition will relate directly to the extent of the accumulated evidence of change. This will lead in turn to smooth changes in parameter estimates and in any derivatives of these estimates, such as loss reserves.

Dynamic models may take various forms. The Kalman filter is an extremely powerful model, but is essentially adapted to the general linear model, in which there is a linear relation between explanatory variable and response, and all stochastic terms are normally distributed.

This may not always be adequate for loss reserving problems, which tend to involve non-linearities and non-normal errors. For this reason the extension of the Kalman filter to the EDF filter in Section 2 may be useful in the loss reserving context.

References

- De Jong, P. and Zehnwirth, B. (1983). Claims reserving state space models and the Kalman filter. **Journal of the Institute of Actuaries**, 110, 157-181.
- De Vylder, F. (1978). Estimation of IBNR claims by least squares. **Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker**, 78, 249-254.
- England, P.D. and Verrall, R.J. (2002). Stochastic claims reserving in general insurance. **British Actuarial Journal**, 8, 443-518.
- Fisher, W.H. and Lange, J.T. (1973). Loss reserve testing: a report year approach. **Proceedings of the Casualty Actuarial Society**, 60, 189-207.
- Hachemeister, C.A. and Stanard, J.N. (1975). IBNR claims count estimation with static lag functions. Paper presented to the 12th Astin Colloquium, Portimao, Portugal, 1975.
- Harrison, P.J. and Stevens, C.F. (1976). Bayesian forecasting. **Journal of the Royal Statistical Society**, Series B, 38, 205-247.
- Hertig, J. (1985). A statistical approach to the IBNR-reserves in marine reinsurance. **ASTIN Bulletin**, 15, 171-183.
- Jewell, W.S. (1974). The credible distribution. **Astin Bulletin**, 7, 237-269.
- Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. **Journal of Basic Engineering**, 82, 340-345.
- Kramreiter, H. and Straub, E. (1973). On the calculation of IBNR reserves II. **Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker**, 73, 177-190.
- Landsman, Z. and Makov, U. (1998). Exponential dispersion models and credibility. **Scandinavian Actuarial Journal**, 89-96.
- Landsman, Z. and Makov, U. (1999). On stochastic approximation and credibility. **Scandinavian Actuarial Journal**, 15-31.
- Landsman, Z. and Makov, U. (2003). Sequential quasi credibility for scale dispersion models. **Scandinavian Actuarial Journal**, 119-135.
- McCullagh P. and Nelder J.A. (1989). **Generalised Linear Models** (3rd Edition). Chapman and Hall, New York.
- Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates, **Astin Bulletin**, 23, 213-225.
- Nelder, J.A. and Verrall, R.J. (1997). Credibility theory and generalised linear models. **Astin Bulletin**, 27, 370-384.

Pollard, J.H. (1983). Outstanding claims provisions: a distribution-free statistical approach. **Journal of the Institute of Actuaries**, 109, 417-433. Also appears in **Institute of Actuaries of Australia General Insurance Seminar**, 3, 88-110.

Reid, D.H. (1978). Claim reserves in general insurance. **Journal of the Institute of Actuaries**, 105, 211-296.

Sawkins, R.W. (1979). Methods of analysing claim payments in general insurance. **Transactions of the Institute of Actuaries of Australia**, 435-519.

Taylor, G.C. (1977). Separation of inflation and other effects from the distribution of non-life insurance claim delays. **Astin Bulletin**, 9, 217-230.

Taylor, G.C. (1986). **Claim Reserving in Non-life Insurance**. North-Holland, Amsterdam.

Taylor, G.C. (2000). **Loss reserving – an actuarial perspective**. Kluwer Academic Publishers, Boston.

Taylor, G.C. and Ashe, F.R. (1983). Second moments of estimates of outstanding claims. **Journal of Econometrics**, 23, 37-61.

Taylor, G.C. and Campbell, M. (2002). Statistical Case Estimation. Research Paper No. 104, Centre for Actuarial Studies, University of Melbourne.

Verrall, R.J. (1989). A State Space Representation of the Chain Ladder Linear Model. **Journal of the Institute of Actuaries**, 116, 589-609.

Wright, T.S. (1990). A stochastic method for claims reserving in general insurance. **Journal of the Institute of Actuaries**, 117, 677-731.

RECENT ADDITIONS TO THE RESEARCH PAPER SERIES

No.	Date	Subject	Author
84	FEBRUARY 2001	DISCRETE TIME RISK MODELS UNDER STOCHASTIC FORCES OF INTEREST	Jun Cai
85	FEBRUARY 2001	MODERN LANDMARKS IN ACTUARIAL SCIENCE Inaugural Professorial Address	David C M Dickson
86	JUNE 2001	LUNDBERG INEQUALITIES FOR RENEWAL EQUATIONS	Gordon E Willmot Jun Cai X Sheldon Lin
87	SEPTEMBER 2001	VOLATILITY, BETA AND RETURN WAS THERE EVER A MEANINGFUL RELATIONSHIP?	Richard Fitzherbert
88	NOVEMBER 2001	EXPLICIT, FINITE TIME RUIN PROBABILITIES FOR DISCRETE, DEPENDENT CLAIMS	Zvetan G Ignatov Vladimir K Kaishev Rossen S Krachunov
89	NOVEMBER 2001	ON THE DISTRIBUTION OF THE DEFICIT AT RUIN WHEN CLAIMS ARE PHASE-TYPE	Steve Drekić David C M Dickson David A Stanford Gordon E Willmot
90	NOVEMBER 2001	THE INTEGRATED SQUARE-ROOT PROCESS	Daniel Dufresne
91	NOVEMBER 2001	ON THE EXPECTED DISCOUNTED PENALTY FUNCTION AT RUIN OF A SURPLUS PROCESS WITH INTEREST	Jun Cai David C M Dickson
92	JANUARY 2002	CHAIN LADDER BIAS	Greg Taylor
93	JANUARY 2002	FURTHER OBSERVATIONS ON CHAIN LADDER BIAS	Greg Taylor
94	JANUARY 2002	A GENERAL CLASS OF RISK MODELS	Daniel Dufresne
95	JANUARY 2002	THE DISTRIBUTION OF THE TIME TO RUIN IN THE CLASSICAL RISK MODEL	David C M Dickson Howard R Waters
96	MAY 2002	A NOTE ON THE MAXIMUM SEVERITY OF RUIN AND RELATED PROBLEMS	David C M Dickson
97	JUNE 2002	UPPER BOUNDS FOR ULTIMATE RUIN PROBABILITIES IN THE SPARRE ANDERSEN MODEL WITH INTEREST	Jun Cai David C M Dickson
98	JUNE 2002	CONTINUOUS COMPOUNDING, VOLATILITY AND THE EQUITY PREMIUM	Richard Fitzherbert

No.	Date	Subject	Author
99	JUNE 2002	THE DEFICIT AT RUIN IN THE STATIONARY RENEWAL RISK MODEL	Gordon E Willmot David C M Dickson Steve Drekić David A Stanford
100	AUGUST 2002	ASIAN AND BASKET ASYMPTOTICS	Daniel Dufresne
101	AUGUST 2002	RUIN PROBABILITIES WITH A MARKOV CHAIN INTEREST MODEL	Jun Cai David C M Dickson
102	AUGUST 2002	THE GERBER-SHIU DISCOUNTED PENALTY FUNCTION IN THE STATIONARY RENEWAL RISK MODEL	Gordon E Willmot David C M Dickson
103	NOVEMBER 2002	INITIAL CAPITAL AND MARGINS REQUIRED TO SECURE A JAPANESE LIFE INSURANCE POLICY PORTFOLIO UNDER VARIABLE INTEREST RATES	Manabu Sato David C M Dickson Richard M Fitzherbert
104	NOVEMBER 2002	STATISTICAL CASE ESTIMATION	Greg Taylor Mireille Campbell
105	MARCH 2003	ACTUARIAL PRACTICE AND CONTROL: OBJECTIVES AND CAPABILITIES	Julian D Gribble
106	MARCH 2003	THE IDENTIFICATION AND MEASUREMENT OF SPECULATIVE RISK	Richard Fitzherbert
107	AUGUST 2003	SOME OPTIMAL DIVIDENDS PROBLEMS	David C M Dickson Howard R Waters
108	AUGUST 2003	THE JOINT DISTRIBUTION OF THE SURPLUS PRIOR TO RUIN AND THE DEFICIT AT RUIN IN SOME SPARRE ANDERSEN MODELS	David C M Dickson Steve Drekić
109	SEPTEMBER 2003	LOSS RESERVING: PAST, PRESENT AND FUTURE	Greg Taylor Gráinne McGuire Alan Greenfield

A complete list of papers can be found at
<http://www.economics.unimelb.edu.au/actwww/papers.html>