Institute of Actuaries of Australia

# Building a reserving robot

Prepared by Gráinne McGuire

Presented to the Institute of Actuaries of Australia

Biennial Convention 23-26 September 2007

Christchurch, New Zealand



The Institute of Actuaries of Australia

Level 7 Challis House 4 Martin Place

Sydney NSW Australia 2000

Telephone: +61 2 9233 3466 Facsimile: +61 2 9233 3446

Email: actuaries@actuaries.asn.au Website: www.actuaries.asn.au

## Abstract

This paper discusses the use of adaptive filters, in particular the GLM adaptive filter, in the construction of a reserving robot – a largely automated system for carrying out the estimation of outstanding claims liabilities at successive valuation time points. Rather than repeating the technical aspects of the GLM filter which are covered in other papers, this paper focuses on some of the practicalities in setting up an initial filtering model. An example of the reserving robot in action is given at the end of the paper.


**Keywords:** adaptive filter, adaptive reserving, Bayesian revision, bootstrap, dynamic generalised linear model, DGLM, loss reserving, model blending, stochastic reserving.

# 1. Introduction

The estimation of outstanding claims liabilities is one of the central problems faced by general insurers. To understand a company's capital position and monitor its ongoing performance, it is necessary to regularly estimate the sum of money required to meet outstanding liabilities. A consequence of this for many companies is that the outstanding claims estimation (or valuation) process essentially operates on a revolving door basis, i.e., the exercise to estimate the outstanding claims liabilities as at the end of a particular quarter draws to a close, only for the valuation of the liabilities as at the end of the next quarter to begin.

Various techniques are used to carry out a valuation. Some examples include the chain ladder in its various forms, payments based models (such as the payments per claim incurred and payments per claims finalised models) and case estimate based models (e.g. the projected case estimates model). A good review of existing methods is given in Taylor (2000).

The following comments may be made about the techniques used in the estimation of outstanding claims liability estimation:

- The models are frequently non-stochastic;
- The models are usually static.

A consequence of the first point is that there is no objective measure of a model's performance in relation to the data. The model fitting process relies on the judgement of the actuary. While judgement is undoubtedly a valuable commodity and a cornerstone of the actuary's trade, it is nonetheless true that two actuaries, exercising judgement, might, very reasonably, come to very different estimates of liabilities.

Static models, whether stochastic or not, are calibrated to the experience to date. They are only applicable to future experience if that experience follows past experience. They have no innate ability to adapt to changing experience. Therefore, valuation upon valuation, an actuary must refit these models to the data from scratch.

Large companies have many lines of business, all requiring separate estimation of the outstanding claims liabilities at each valuation date (frequently successive quarters), a time consuming process. So it would be of considerable value to automate the process.

Consider the features any such automatic process, or reserving robot, would need to have:

- The model must be able to adapt to changing experience without human intervention;
- There should be an objective means of measuring the goodness of fit and performance of the model.

The first requirement is the cornerstone of the reserving robot – experience rarely stays stable so any automatic process must have the flexibility to adapt to changes when sufficient evidence for such changes accrues.

The second requirement enables the reserving model to be tested by the robot to ensure that the fit to the data is adequate. At this stage of computer intelligence, no automatic programme could fully replace actuarial judgement. It is necessary to have tests to flag cases where the experience changes to the extent that intervention may be required to rework the model.

One class of models that satisfy, at least to some extent, the requirements above are dynamic stochastic models. The dynamism means that the models can adapt to change while their stochastic nature means that a series of model diagnostics can be developed to test the performance of the model at each valuation point.

The purpose of this paper is to discuss the use of one type of dynamic stochastic model – adaptive filters – and their potential use in the construction of a reserving robot.

# 2. Adaptive filters

An adaptive filter, as applied to general insurance reserving, is a recursive algorithm that models (filters) the claims experience time period by time period (e.g. by accident period or calendar period). It is adaptive in that the model or filter may change from one period to the next should experience warrant it. In other words, the model parameters may evolve over time.

Filtering methods fall into the class of Bayesian models. Such models combine prior beliefs (distributions) about parameters together with evidence from data (likelihood) to produce posterior estimates of these parameters. In the case of a filter the prior estimates are those based on data up to the previous period, while the posterior estimates are based on the prior and the current experience period's data. The posterior estimates are modified relative to the prior estimates based on the amount of evidence for change in the data relative to the level of belief in the prior estimates. Thus, filters adapt to changing experience and this makes them suitable models for use in a reserving robot.

## 2.1 The Kalman filter

A well-known adaptive filter is the Kalman filter (Kalman, 1960) which was introduced into 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**

$$\underset{p\times 1}{\beta(s+1|s)} = \underset{p\times p}{\Phi(s+1)}\,\underset{p\times 1}{\beta(s|s)} + \underset{p\times 1}{r(s+1)} \tag{2.1}$$

**Observation equation**

$$\underset{p\times 1}{Y(s+1)} = \underset{n\times p}{X(s+1)}\,\underset{p\times 1}{\beta(s+1|s)} + \underset{p\times 1}{v(s+1)} \tag{2.2}$$

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

- $Y(s+1)$ denotes the vector of observations made at time $s+1$ (=1, 2,…$T$)
- $\beta$(s+1) denotes the vector of parameters at time $s+1$
- $X$(s+1) is the design matrix applying at time $s+1$
- Φ(s+1) is a transition matrix governing the evolution of the expected parameter values from one epoch to the next
- $r$(s+1) and $v$(s+1) are stochastic perturbations, each with zero mean, and with

$$\text{Var}\left[r(s+1)\right] = R(s+1)$$
$$\text{Var}\left[v(s+1)\right] = V(s+1) \qquad (2.3)$$

The formal statement of the filter requires a little extra notation. Let $Y(s|s\text{-}k)$ denote the filter's estimate of $Y(s)$ on the basis of information up to and including epoch $s\text{-}k$; and similarly for other symbols at $s|s\text{-}k$. Also, let $\Theta(s|s\text{-}k)$ denote the estimate of $\text{Var}[\beta(s|s\text{-}k)]$.

Equations (2.4) to (2.10) present the mathematical detail behind the Kalman filter. The process may be split into three steps. A description is given below.

**Step 1: forecast new epoch's parameters and observations *without* new information**

$$\beta(s\,|\,s-1) = \Phi(s)\beta(s-1\,|\,s-1) \qquad (2.4)$$

$$\Theta(s\,|\,s-1) = \Phi(s)\Theta(s-1\,|\,s-1)\Phi^T(s) + R(s) \qquad (2.5)$$

$$\hat{Y}(s\,|\,s-1) = X(s)\beta(s\,|\,s-1) \qquad (2.6)$$

**Step 2: calculate gain matrix (credibility of new observation)**

$$L(s\,|\,s-1) = X(s)\Theta(s\,|\,s-1)X^T(s) + V(s) \qquad (2.7)$$

$$K(s) = \Theta(s\,|\,s-1)X^T(s)\left[L(s\,|\,s-1)\right]^{-1} \qquad (2.8)$$

**Step 3: update parameter estimates to incorporate the new information**

$$\beta(s\,|\,s) = \beta(s\,|\,s-1) + K(s)\left(Y(s) - \hat{Y}(s\,|\,s-1)\right) \qquad (2.9)$$

$$\Theta(s\,|\,s) = \left(1 - K(s)X(s)\right)\Theta(s\,|\,s-1) \qquad (2.10)$$

Equations (2.4) to (2.6) generate forecasts of the new epoch's parameters and observations based on no new information. In a Bayesian framework, these estimates are the prior estimates for that epoch. The gain matrix (credibility of the new observation) is calculated in (2.8). This calculation compares the parameter variances with the process error (i.e. data variance) and calculates the credibility of the latest set of payments on this basis. Finally the parameter estimates are updated in equations (2.9) and (2.10). The process then moves onto the next epoch, starting again with equation (2.4).

The filter starts with prior estimates for $\beta(0|0)$ and the associated dispersion $\Theta(0|0)$.

## 2.2 The GLM filter

Although a very flexible tool, the Kalman filter requires the assumption of normally distributed data. In the context of loss reserving, this is generally dealt with by assuming that the variables of interest (claim sizes, payments in a specific period etc) are log-normally distributed. Although a common assumption in the actuarial field (see, e.g, England and Verrall, 2002), this is an approach not without its problems. The main difficulty is the requirement for a bias correction which results from modelling a transformation of the data (in this case the log transform), the determination of which, particularly when the distribution of the data is only approximately known, can be problematic. A second drawback is the restriction to one possible distribution for the data. Modelling claim counts, for example, would not be easily done with either a normal or log normal distribution.

An alternative is the GLM filter, derived by Taylor (2008), with some practical applications discussed in McGuire and Taylor (2007). This is an extension of the Kalman filter to certain members of the Exponential Dispersion Family of distributions (Nelder and Wedderburn, 1972; McCullough and Nelder, 1989). The family includes common distributions like the Normal, Poisson, Gamma and Inverse Gaussian. The Gamma distribution, and to a lesser extent, the Inverse Gaussian distribution, are both candidates for the modelling of long-tailed, strictly positive variables. The Poisson distribution is a natural choice for claim counts. Applied within the framework of a filter, this would mean updating a generalised linear model (McCullough and Nelder, 1989) rather than a normal linear model (the Kalman filter).

Although Taylor (2008) gives a general form of the analytical filter, he notes that it is only tractable in a limited number of cases, depending on both the distribution used and the link function applied. However, tractable cases include the gamma distribution/log link and the Poisson distribution/log link; the former is useful for claim size modelling while the latter may be used for claim count or finalisation numbers estimation. An additional tractable case is the normal distribution/identity link which corresponds exactly to the Kalman filter.

For the GLM filter, the system equation remains unchanged from that in (2.1). However, the observation equation becomes

$$\underset{p\times 1}{Y(s+1)} = h^{-1}\left(\underset{n\times p}{X(s+1)}\,\underset{p\times 1}{\beta(s+1|s)}\right) + \underset{p\times 1}{v(s+1)} \tag{2.11}$$

where $h^{-1}$ is the inverse link function and $\upsilon(s+1)$ is not necessarily normal. In the cases discussed in this paper, $h^{-1}$ is the exponential function, while $\upsilon(s+1)$ is gamma or Poisson distributed.

The equations underlying the Analytical GLM filter follow the Kalman filter, in principle, but are considerably more complex and are not reproduced here. The equations for the Gamma error/log link case are outlined in Appendix A to this paper. For more complete information, the reader is referred to Taylor (2008) and McGuire and Taylor (2007).

The philosophy behind the GLM filter is the same as that behind the Kalman filter in that prior estimates of an epoch's parameters and observations (either based on user input if the first epoch, or data from previous epochs if not) are calculated. Then a quantity similar to the credibility gain matrix is calculated, based on the epochs's observations, using which the parameter estimates are updated to take account of these data.

Further, like the Kalman filter it is an analytical procedure, consisting of a set of equations. Therefore it is a quick procedure, taking negligible computing time. It is an appropriate tool to use in the construction of a reserving robot.

# 3. Data

The data set used in Section 4 of this paper is that used throughout Taylor (2000), which is drawn from an Australian Motor Bodily Injury Portfolio from 1978 to 1995. The data are adjusted for past economic inflation as per Taylor (2000). Data from accident years 1980 to 1995 is used in this paper. For convenience, the data used in this paper are replicated in Appendix B. This consists of claim payments (adjusted for past economic inflation) and estimates of ultimate claim numbers by accident year.

In Section 5, the data used are long-tailed claims. The data are summarised as follows:

- Total amount of claim payments for year of accident $i$ ($i$=1, 2, …, 14) and development quarter $j$ ($j$=1, 2, …, 53)
- Total number of claims reported for year of accident $i$ and development quarter $j$
- Total number of claim closures for year of accident $i$ and development quarter $j$
- Total amount of case estimates for year of accident $i$ and development quarter $j$

The data were split into two groups by legal jurisdiction, and each of the above summaries was available for each of these two groups.

Ultimate numbers of incurred claims by accident year were also available; these had been estimated in a separate modelling exercise.

The payments and case estimate data were adjusted for past economic inflation in line with an Australian wage earning index.

Some notation is defined here. Let

- $C_{ij}^L$ = claim payments in development period $j$ for accident period $i$ for jurisdictional grouping $L$ ($L$=1, 2 for the bodily injury data and absent for the motor data);
- $N_{ij}^L$ = number of claims reported in development period $j$ for accident period $i$ for jurisdictional grouping $L$;
- $F_{ij}^L$ = number of claim closures in development period $j$ for accident period $i$ for jurisdictional grouping $L$;
- $E_{ij}^L$ = case estimates at end of development period $j$ for accident period $i$ for jurisdictional grouping $L$;
- $U_i^L$ = estimated ultimate number of claims incurred in accident period $i$ for jurisdictional grouping $L$;
- $O_{ij}^L$ = number of claims open for accident period $i$, at start of development period $j$ for jurisdictional grouping $L$.

# 4. Constructing the robot

In this section the process of setting up the reserving robot is illustrated using a PPCI model based on the motor bodily injury payments data from 1980 to 1995. The steps involved in this process are:

- Gather the relevant data;
- Select the error distribution and link function;
- Make assumptions on the process error associated with these data;
- Select appropriate basis functions for modelling the data;
- Initialise the parameter estimates;
- Make assumptions on the variance within the parameter estimates;
- Apply the filter;
- Examine model diagnostics to ensure the fit is adequate.

Once these steps have been carried out, the robot will be programmed.

## 4.1 Data

For this example, a PPCI (Payments per Claim Incurred ) model is used where

$PPCI_{ij} = C_{ij} / U_i$ .

The payments used are gross payments.

## 4.2 Selecting the error distribution and link function

As discussed by Taylor (2008) there is a limited number of error and link combinations for which an analytical version of the GLM filter exists. Such cases include:

- Normal error, identity link (which is identical to the Kalman filter);
- Gamma error, log link;
- Gamma error, reciprocal link;
- Poisson error, log link.

For PPCI data, possible choices would be a gamma distribution or the log-normal model (meaning that the Kalman filter would be applied to log(PPCI)). As indicated in Section 2.1, there are difficulties associated with the log transform required to use the Kalman filter, namely the bias correction. Thus the Gamma distribution has been selected here. To guarantee strictly positive PPCI values, a log link is used.

## 4.3 Process error

The process error is the variance associated with the data. All else being equal, the more variable the data, the greater support needed for the filtered parameter estimates to change significantly from one period to the next since there is a good chance that changes result from noise rather than a genuine shift in experience. Conversely if the process error is low, then changes should be followed by the filter since these are likely to be genuine experience effects. Process error may vary by development period or even by accident and development period.

It is necessary to make assumptions regarding the levels of process error in the data. For the gamma distribution, this may be done by making assumptions for the coefficient of variation relating to each cell. It is reasonable to expect that the coefficient of variation might vary by development period only. Guidance may be sought from the data on the levels of variation seen.

For the motor data used here, the mean and standard deviation of the observed PPCI values are calculated by development year, from which the coefficient of variation is calculated. This process is illustrated in Figure 4.1.

Looking at the filter from a Bayesian viewpoint, the process error assumptions represent information about the data or likelihood. Thus, their estimation on the basis of the data itself is reasonable.

**Figure 4.1 Selection of coefficients of variation**


## 4.4 Selection of basis functions

In the particular example of the PPCI model for the motor bodily injury data, the PPCI by development year are modelled for each accident period. Thus it is necessary to choose appropriate basis functions to accommodate the pattern of average claim payments.

One possibility for PPCI values is to use a Hoerl Curve (De Jong and Zehnwirth, 1983; Wright, 1990). This has the following form:

$$\mu_{ij} = \{\exp \beta_0 + \beta_1 (j-1) + \beta_2 \log j\}, \qquad j = 1, 2, \ldots \qquad (4.1)$$

where $\mu_{ij}$ represents the mean value of $PPCI_{ij}$ in cell $(i,j)$.

For the data here, the Hoerl curve above is used with one modification in that an additional basis function is used to separately model the first development year. Therefore the model applied is

$$\mu_{ij} = \{\exp \beta_0 + \beta_1 (j-1) + \beta_2 \log(j) + \beta_3 \mathrm{I}(j=1)\}, \qquad j = 1, 2, \ldots \qquad (4.2)$$

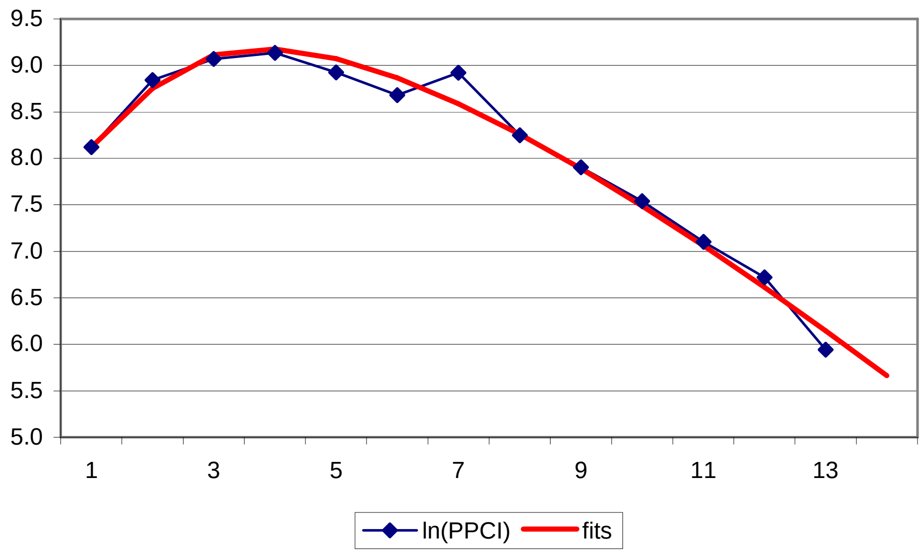
with basis functions:

- 1;
- *j-1*;
- log(*j*);
- I(*j*=1) where I(condition) is the indicator function, being 1 if the condition is true and 0 otherwise.

## 4.5 Initialisation of parameter estimates

The filter requires initial values for the parameters [$\beta(0|0)$] associated with each of the basis functions. Although, in principle, any values could be input since the filter will eventually adapt to the experience, there is generally insufficient experience to allow this adaptive process to run its course if the starting parameters are very wrong. Therefore, it is best to start with reasonable choices.

**Figure 4.2 Initialisation of parameter estimates**




One way of selecting the initial parameters is to fit the basis functions to the first accident year in isolation, and use these values as the starting point for the filter. Here, however, it is observed that experience in accident years 1980 –

1982 is somewhat different to that in later years in early to mid development years. Hence, the basis functions were fitted to the log of the average PPCI values for accident years 1983-1985 using unweighted least squares regression (see Figure 4.2 and Table 4.1) and these values used as the starting point. This means that the filter will poorly fit early to mid development periods for the early accident years. However it is the estimates for the later development periods only that are of interest for these accident periods; here the model forecasts are more reasonable.

**Table 4.1 Parameter initialisation using unweighted linear regression**

| Develop-ment year | Ave PPCI 1983-85 | log(PPCI) | Basis function | Parameter estimate | Parameter variance |
|---|---|---|---|---|---|
| 0 | 3,365 | 8.121 | 1 | 7.661 | 0.0010 |
| 1 | 6,914 | 8.841 | j-1 | -0.669 | 0.0005 |
| 2 | 8,691 | 9.070 | log(j) | 2.541 | 0.0005 |
| 3 | 9,287 | 9.136 | I(j=1) | 0.460 | 0.0010 |
| 4 | 7,511 | 8.924 | | | |
| 5 | 5,882 | 8.680 | | | |
| 6 | 7,488 | 8.921 | | | |
| 7 | 3,823 | 8.249 | | | |
| 8 | 2,707 | 7.903 | | | |
| 9 | 1,879 | 7.539 | | | |
| 10 | 1,214 | 7.102 | | | |
| 11 | 829 | 6.720 | | | |
| 12 | 381 | 5.943 | | | |

## 4.6 Parameter variance

Selecting appropriate parameter variances is a more difficult task. The parameter variances correspond to matrix $\Theta(0|0)$. From a Bayesian perspective these represent prior beliefs about the variance of the parameter estimates. Some ways of selecting these include:

- Putting high variances on these parameters; this corresponds to an uninformative prior which is appropriate in Bayesian problems if the user has no prior knowledge of what the parameters should be;
- Small variances on the parameters: in general in Bayesian problems this corresponds to having strong beliefs about what the parameters are. A large amount of evidence to the contrary within the data would then be required to materially alter the parameter estimates.

Within the context of an adaptive filter, these two cases have additional implications to those outlined above. High variances will effectively mean that a different model is fitted independently to each epoch of the filter. It therefore means that the number of parameters in the model is large since the parameters for each epoch are essentially independent of each other. The model is not parsimonious and is prone to over-fitting. Conversely using very low parameter variances leads to the filter being unlikely to change from one epoch to the next. Essentially this means that a static model is fitted to the data. The model is parsimonious, but loses its adaptive nature. Thus, it is not particularly useful as a reserving robot.

Not surprisingly, a compromise between the two extremes is required. Low parameter variances mean that there are strong relationships between the parameter estimates in one epoch and those in the next, meaning that the effective number of independent parameters is considerable less than the product of the number of parameters in each epoch and the number of epochs. Therefore, the parameter variances should be low enough to keep the effective number of parameters low so as to prevent the model over-fitting, yet high enough that the filter's adaptive ability is maintained.

At present this choice comes down to judgement. Typically values of the order of $10^{-3}$ to $10^{-5}$ are required. Some experimentation may be necessary to strike the necessary balance.

Further there is an additional reason not to use over-large parameter variances for the GLM filter. As described in Taylor (2008), the GLM filter is a second order approximation to the Bayesian updating of a GLM. It relies on a number of approximations based on Taylor series. High variances may lead to the higher order terms of these series becoming too large, and is likely to result in numerical instability of the filter.

It is customary to assume that the prior covariances between parameters are zero. It is also customary to assume that $R(s)$ (see equation (2.1)) is constant for all $s$ and is equal to $\Theta(0|0)$.

The variances assumed here are given above in Table 4.1. The level parameter (basis function of 1) has a variance of 0.001 and thus, a standard deviation of 0.03. This means that the level of the PPCI curve has about a two thirds chance of not shifting by more than 3% from one accident period to the next.

## 4.7 Applying the filter

Once the various inputs (process error and initial parameter estimates and variances) have been selected, the filter may be run. In this case the filter successively models accident years 1980 through to 1995.

## 4.8 Model diagnostics

It is important to ensure the model is fitting the data adequately. Some tests which may be used are given below.

### 4.8.1 Comparing the actual and fitted data

One useful diagnostic is to plot the actual and fitted log(PPCI)s by successive accident years. Such plots are given for three accident years (1982, 1984 and 1990) in Figure 4.3, Figure 4.4 and Figure 4.5 below. Also included in these graphs is the fitted curve from the previous accident period.

At the start of accident year 1982, the fit is poor. This was discussed in Section 4.5. However the fit to the latter development years is reasonable. The fit to accident year 1981 is very similar to that for 1982 suggesting comparable experience between the two accident years.

**Figure 4.3 Actual and fitted data - 1982**

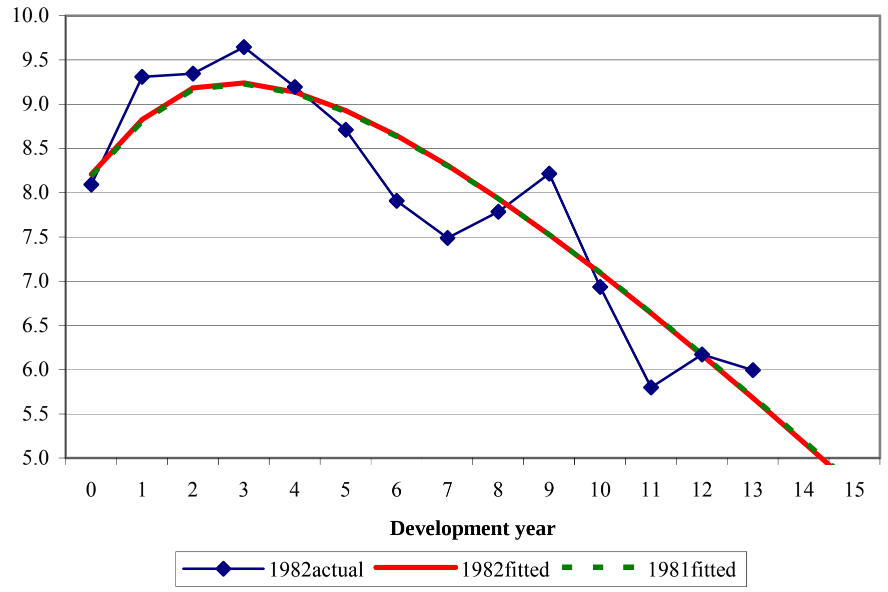
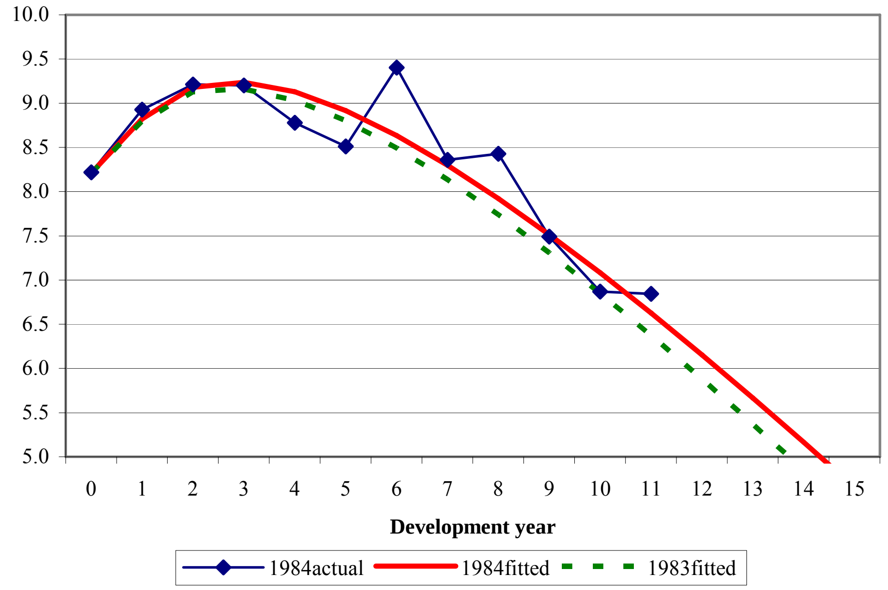



**Figure 4.4 Actual and fitted data - 1984**



The model fits accident year 1984 well. There has been some upward movement in the fitted values between 1983 and 1984.

**Figure 4.5 Actual and modelled data - 1990**

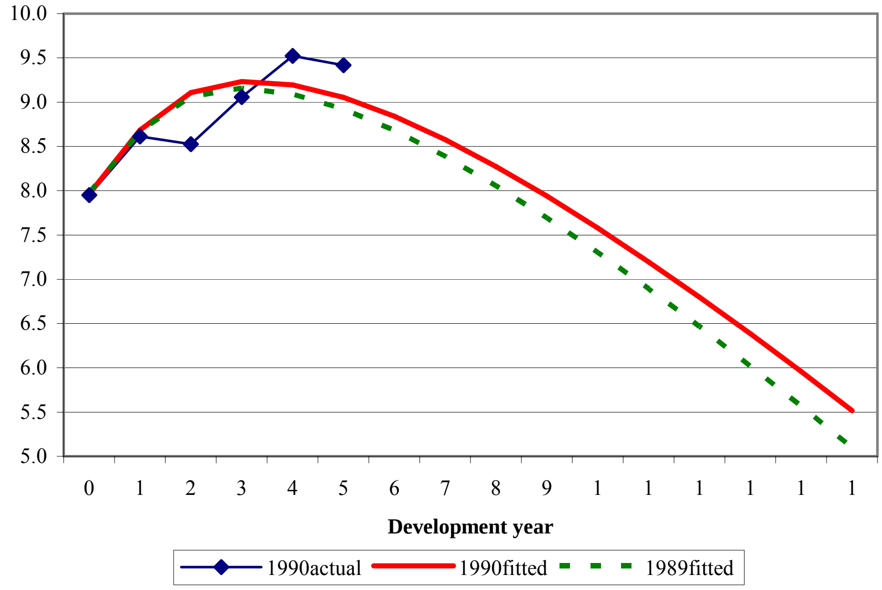



Figure 4.5 provides a good demonstration of the filter in action. The 1990 data suggest slightly higher payments and a slightly longer mean term to payment than the curve fitted to accident year 1989. Consequently the curve moves upwards. However, it does not fully move to reflect the 1990 experience due to the relatively low parameter variances imposed (Section 4.6).

### 4.8.2 Actual and expected triangle

The triangular arrangement of the ratio of actual to expected PPCI values is also useful to identify regions of poor fit. Figure 4.6 presents these ratios for this model. This triangle is colour-coded with blue meaning that actual PPCI values are greater than fitted and vice versa for pink.

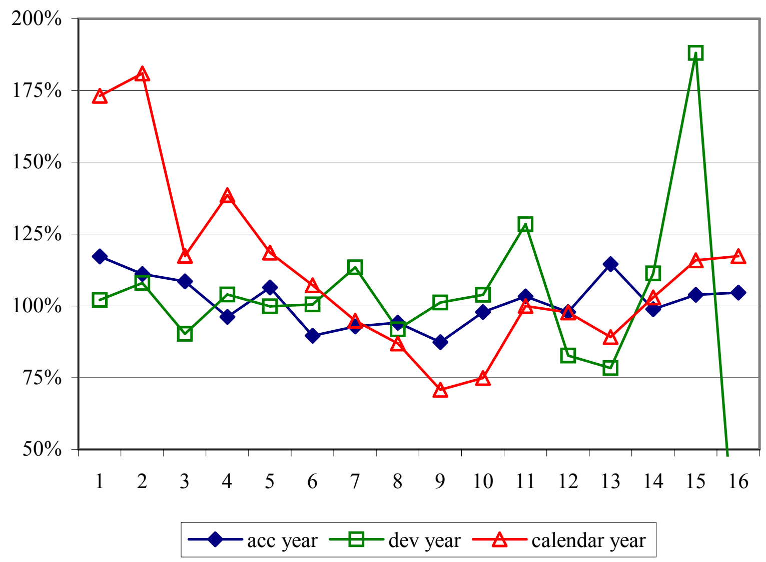
Some problems are identified in this triangle. Firstly, the fit is poor in the early to mid development years for the first three accident years. However, that is expected due to the initialisation of the parameters. For mid-range accident years (1985-1990) there do appear to be some problems in development years 0 to 4 where the model tends to over-estimate the PPCI.

It is useful looking at this information summarised in the three directions of the triangle – by rows (accident year), by columns (development year) and by diagonal (calendar or payment year). Figure 4.7 presents this information. If the model fits well, then the lines on this graph should be randomly located around 100%. It is seen that the ratios appear satisfactory by accident and development period, but show a definite pattern by calendar year. The diagnostic is picking up the problems identified in the paragraph above.

**Figure 4.6 Actual/Expected ratios**

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1980 | 173% | 196% | 119% | 130% | 120% | 99% | 86% | 45% | 72% | 41% | 118% | 92% | 109% | 189% | 290% | 0% |
| 1981 | 158% | 131% | 145% | 125% | 91% | 89% | 79% | 81% | 69% | 104% | 236% | 37% | 9% | 16% | 95% | |
| 1982 | 89% | 162% | 118% | 150% | 106% | 80% | 48% | 44% | 86% | 199% | 85% | 43% | 100% | 137% | | |
| 1983 | 102% | 118% | 93% | 102% | 106% | 67% | 87% | 98% | 91% | 41% | 131% | 122% | 107% | | | |
| 1984 | 101% | 111% | 103% | 97% | 70% | 67% | 216% | 106% | 166% | 98% | 81% | 124% | | | | |
| 1985 | 75% | 82% | 79% | 82% | 79% | 113% | 106% | 96% | 50% | 175% | 119% | | | | | |
| 1986 | 71% | 98% | 71% | 77% | 98% | 94% | 136% | 120% | 122% | 53% | | | | | | |
| 1987 | 74% | 70% | 68% | 81% | 83% | 115% | 114% | 152% | 144% | | | | | | | |
| 1988 | 71% | 104% | 49% | 95% | 71% | 96% | 149% | 77% | | | | | | | | |
| 1989 | 92% | 69% | 71% | 73% | 131% | 142% | 103% | | | | | | | | | |
| 1990 | 98% | 93% | 56% | 84% | 139% | 144% | | | | | | | | | | |
| 1991 | 104% | 77% | 106% | 90% | 109% | | | | | | | | | | | |
| 1992 | 99% | 99% | 78% | 159% | | | | | | | | | | | | |
| 1993 | 120% | 93% | 96% | | | | | | | | | | | | | |
| 1994 | 100% | 106% | | | | | | | | | | | | | | |
| 1995 | 105% | | | | | | | | | | | | | | | |

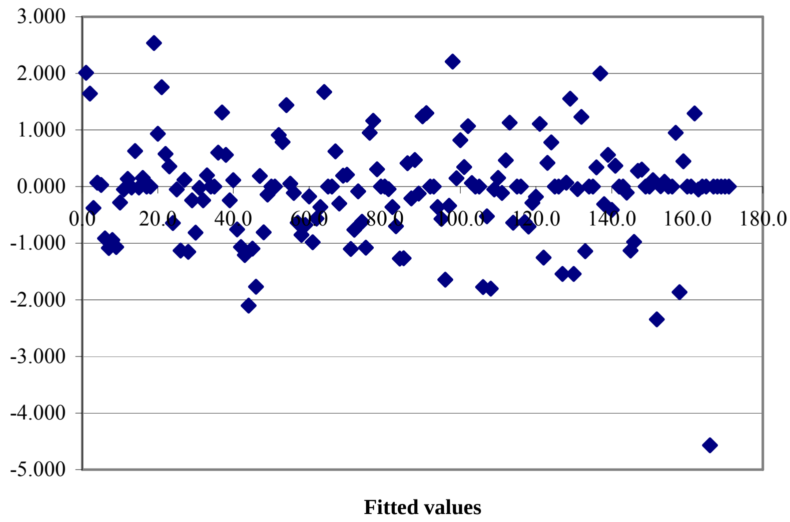
**Figure 4.7 Actual/expected graph**




### 4.8.3 Other residual plots

Usual residual plots like scatterplots, histograms and quantile-quantile plots may also be used. For the GLM filter it is best to use the deviance residuals since these are approximately normal. An example of these plots is given in Figure 4.8, a scatterplot of the deviance residuals against the fitted values. These residuals should form a cloud around zero with no changes in spread by fitted value, as seen in Figure 4.8. However it should be noted that the residuals from a filter are, by their nature, not independent so a greater tolerance is needed in interpreting these residuals. Some evidence of this may be seen in Figure 4.6 and Figure 4.7.

**Figure 4.8 Deviance residuals by fitted values**




## 4.9 Risk margins

Prudential standards in Australia require the specification of risk margins as well as the central estimate of liability. These are defined as the maximum of the 75-percentile (as the excess over the mean) and half the coefficient of variation.

Any reserving robot must be able to calculate risk margins and indeed it is relatively straightforward to do so. Since the adaptive filter is a stochastic method, it is possible to use bootstrapping to obtain a distribution of the reserves from which the risk margins may be calculated.

The bootstrapping process is described in McGuire and Taylor (2007) and is not repeated here. However, it is worth noting that the usual bootstrapping methods for outstanding claims liability estimation (see, e.g., Taylor 2000, Pinheiro et al 2003) may not be used for the adaptive filter since these methods require independent residuals. If that procedure were applied to the GLM filter, then the predictive variance would be under-estimated. Stoffer and Wall (1991) discuss an appropriate procedure for bootstrapping the Kalman filter and its dependent residuals. This approach is appropriate for the GLM filter, and the modification required to adapt it to the GLM filter is outlined in McGuire and Taylor (2007).

The output from the bootstrap is a large number of simulations, say 1000, each one containing the cell by cell projection of outstanding claims liabilities. These may be summed to get the distribution of the total outstanding by accident year and overall. Table 4.2 presents the results for the PPCI model based on the inputs discussed above.

The overall coefficient of variation is quite high. This is a reflection of the fact that the PPCI model is a poor model for these data (as evidenced by the higher coefficients of variation associated with this model relative to the PPCF in Taylor, 2000) which is perhaps not surprising given that these are motor bodily injury data and therefore more amenable to modelling by a lump sum payment model like the PPCF model.

Note that the coefficients of variation quoted in Taylor (2000) using the Kalman filter with the PPCI model are not directly comparable with the coefficients of variation here. The former coefficients of variation are based on the bootstrap for independent residuals which, as indicated above, will under-estimate variability in an adaptive filter.

**Table 4.2 Results from bootstrapping**

| Accident year | Liability estimate $'(000) | Standard Deviation $'(000) | Coefficient of variation % | 75-percentile (% of mean) % |
|---|---|---|---|---|
| 1980 | 135 | 69 | 51 | 128 |
| 1981 | 244 | 128 | 52 | 140 |
| 1982 | 388 | 253 | 65 | 124 |
| 1983 | 498 | 317 | 64 | 123 |
| 1984 | 1,166 | 842 | 72 | 116 |
| 1985 | 1,912 | 1,390 | 73 | 121 |
| 1986 | 2,947 | 1,640 | 56 | 140 |
| 1987 | 5,285 | 2,837 | 54 | 130 |
| 1988 | 6,858 | 3,743 | 55 | 116 |
| 1989 | 12,149 | 5,490 | 45 | 120 |
| 1990 | 20,205 | 8,388 | 42 | 118 |
| 1991 | 28,910 | 11,683 | 40 | 115 |
| 1992 | 44,442 | 14,203 | 32 | 118 |
| 1993 | 52,551 | 15,142 | 29 | 114 |
| 1994 | 61,467 | 16,905 | 28 | 114 |
| 1995 | 68,180 | 17,576 | 26 | 111 |
| Total | 307,337 | 91,171 | 30 | 113 |

## 4.10 Multiple models and model blending

Since different models perform best for different types of data, it is common practice to apply several models and blend the results of these to form the overall liability estimate at a valuation date. A possibility is the blending process developed in Taylor (1985). The actual process used here is described in McGuire and Taylor (2007) and is a modification from the process described in Taylor (2000), which in turn is a slight variation of the original 1985 version. The blending procedure takes account of the variances of each of the model's results, the smoothness of the liability estimate relative to case estimates and the smoothness of the weights themselves in the production of model blending weights.

For the data above it would be expected that the PPCF model would perform better than the PPCI model and that the PCE (Projected Case Estimates) model would perform better for earlier accident periods than both payments based models. Taylor (2000) uses the Kalman filter for all three models and finds exactly that, as indicated by the coefficients of variation associated with each accident year. The blending process therefore downweights the PPCI relative to the PPCF and downweights both payments based models relative to the PCE in earlier accident periods and vice versa in more recent accident years.

### 4.11 Running the robot

As seen above, there is a substantial time cost in setting up the robot. However, once the adaptive model has been set up and tested at one valuation period, the workload in future quarters should be greatly reduced. Since the model is adaptive, it will automatically adapt to new experience as that arises in future valuation periods. Often all that will be required in future periods is to prepare the data, run the model (the inputs remain the same except for any economic inflation adjustments that may be needed, or changes to numbers incurred) and examine the model diagnostics to ensure the model fit is still adequate. Thus, a large part of the model fitting process happens automatically in subsequent periods.

Of course, some changes will be too great or too sudden for the filter to handle. Examples might be major changes due to legislation or the changing of benefits etc. Such cases should be detected through poor model diagnostics and would require intervention, possibly in the form of a full remodelling of the data as discussed above. For a company with many lines of business, all reserved using the filter robot, this might mean that at any given valuation, 80%-90% of the lines of business require no intervention while the remainder would need moderate to major remodelling. Moderate intervention might involve changing some of the inputs such as the parameter variances or data error. Major intervention may be changing the basis functions, or a change in data modelled or anything else that would require full remodelling of the experience.

# 5. The robot in action

The next example of the reserving robot in action has been taken from McGuire and Taylor (2007). It works through the PPCI, PPCF and PCE models for a long-tailed data set (described in Section 3). The three models are bootstrapped and the results blended to produce final estimates of outstanding claims liability and associated risk margin.

## 5.1 Payments per claim incurred model

Let

$$PPCI_{ij}^{L} = C_{ij}^{L} / U_{i}^{L} \tag{5.1}$$

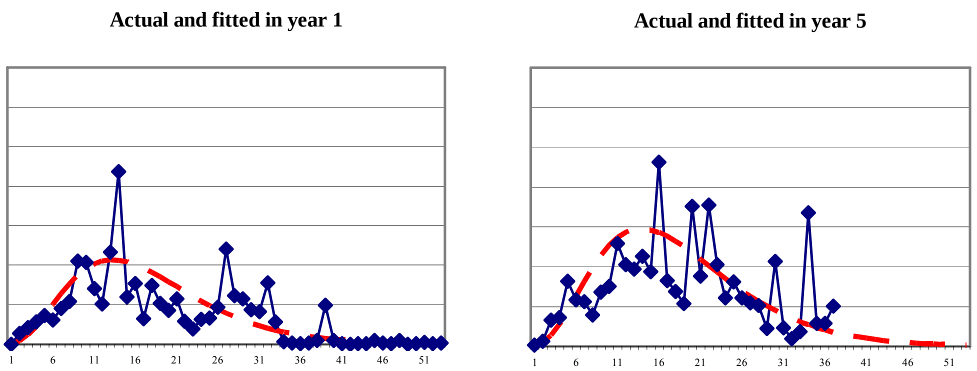
For this data set, the Hoerl curve is used as a starting point. Further, each jurisdictional group is permitted to have its own shape. The applied model has the form:

$$\begin{aligned} & PPCI_{ij}^{L} \sim Gamma(\mu_{ij}^{L}, v_j) \\ & \mu_{ij}^{L} = \exp\{\beta_{i0} + \beta_{i1}\log j + \beta_{i2} j + \beta_{i3}\min(j,16) \\ & \qquad + \mathrm{I}(L=1)\{\beta_{i4} + \beta_{i5}\log j + \beta_{i6} j + \beta_{i7}\min(j,16)\} \end{aligned} \tag{5.2}$$

where $v_j$ = coefficient of variation in development quarter $j$ and may be estimated based on the data and actuarial judgement as discussed in Section 4.3.

Examples of the actual data and fitted model are shown below in Figure 5.1, which plots PPCIs for the one of the jurisdictional groups (labelled as group 0). Note that both graphs are presented on the same vertical scale. Thus, movement in average claim payments from accident year 1 to 5 is apparent in these plots.

**Figure 5.1 Fitted PPCI model**



## 5.2 Payments per claims finalised model

The payments per claims finalised model (PPCF) actually consists of two sub-models:

- Average payments per claim finalised;
- The probability a claim finalises in a particular quarter.

**Model of Average payments per claim finalised**

Let

$$PPCF_{ij}^{L} = C_{ij}^{L} / F_{ij}^{L} \qquad (5.3)$$

Average payments per finalised claim tend to increase with age of claim since the more complicated and serious claims are more likely to take longer to settle. For the data in this paper, the following model is used:

$$PPCF_{ij}^{L} \sim Gamma(\mu_{ij}^{L}, v_j)$$

$$\mu_{ij}^{L} = \exp\left\{\beta_{i0} + \mathrm{I}(L=0)\left[\beta_{i1}\left(\frac{j-1}{4}\right)^{-0.4} + \beta_{i2}\min\left(6, j-1\right)\right]\right. \qquad (5.4)$$
$$\left. + \mathrm{I}(L=1)\left[\beta_{i3} + \beta_{i4}\left(\frac{j-1}{4}\right)^{-0.75} + \beta_{i5}\min\left(6, j-1\right)\right]\right\}$$

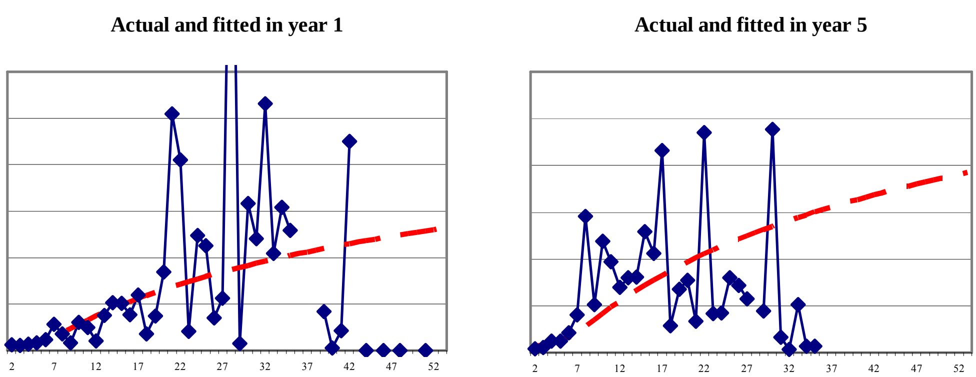
where

$\mu_{ij}^{L}$ = the mean $PPCF_{ij}^{L}$ in cell (*i,j*);

$v_j$ = coefficient of variation in development quarter *j*; and

modelling begins in the second development quarter (i.e. *j*=2).

**Figure 5.2 Fitted PPCF model**



Examples of the actual data and fitted model are shown in Figure 5.2. Again both graphs are presented on the same vertical scale.

**Model of probability of finalisation**

Let

$$PRF_{ij}^{L} = F_{ij}^{L} \Big/ \left( O_{ij}^{L} + R_{ij}^{L} / 3 \right) \qquad (5.5)$$

The denominator in (5.5) is an exposure measure of the number of claims that may be finalised in cell (*i,j*). The probability of finalisation for the data in this paper is quite a volatile quantity; therefore simple models have been used.

$$PRF_{ij}^{L} \sim \text{Over-dispersed Poisson}(\mu_{ij}^{L}, v_j)$$

$$\mu_{ij}^{L} = \exp\Big\{ \beta_{i0} + \mathrm{I}(L=0)\left[ \beta_{i1} \min\left(j-1,15\right) + \beta_{i2} \max\left(1, j-13\right)^{-0.25} \right] + \mathrm{I}(L=1)\left[ \beta_{i3} + \beta_{i4} \min\left(j-1,15\right) + \beta_{i5} \max\left(1, j-13\right)^{-0.25} \right] \Big\} \qquad (5.6)$$

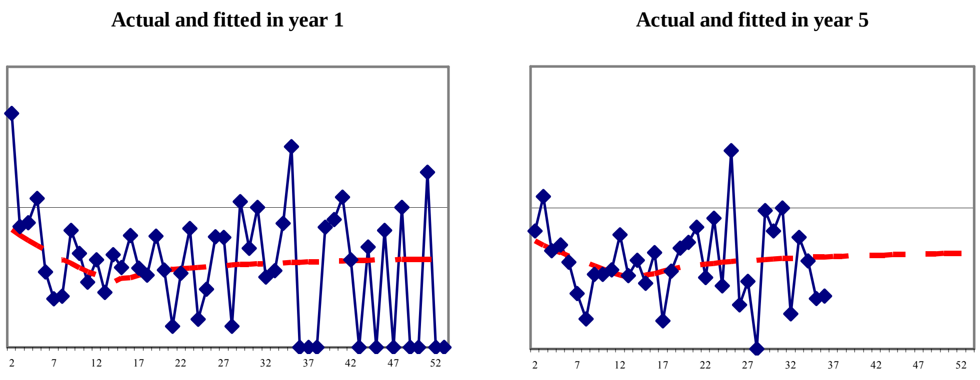
where

$\mu_{ij}^{L}$ = the mean $PRF_{ij}^{L}$ in cell (*i,j*); and

$v_j$ = ratio of the variance to the mean, which may exceed 1 (standard Poisson) through the use of an over-dispersed Poisson.

Examples of the actual data and fitted model are shown in Figure 5.3. Again both graphs are presented on the same vertical scale.

**Figure 5.3 Modelled finalisation probabilities**




## 5.3 Projected case estimates model

The projected case estimates model examines the further development required by case estimates at a given point in time to be sufficient to settle claims in full. In any given development period and accident period, the hindsight estimate of case estimates at the end of the quarter (the sum of payments in that quarter and closing case estimates) may be compared against opening case estimates. If the opening case estimates were sufficient, then the ratio would be one; a value greater than one indicates that the opening estimates are now considered insufficient, while a value less than one indicates they are now considered excessive.

Therefore, there are two quantities to be modelled – case estimates and payments. Both may be expressed as factors relative to the opening case estimates.

**Payment factors**

Payment factors are defined, for $j$=2, 3, …, as:

$$PF_{ij}^{L} = C_{ij}^{L} \big/ E_{i,j-1}^{L} \qquad (5.7)$$

The $PF_{ij}^{L}$ are modelled as follows:

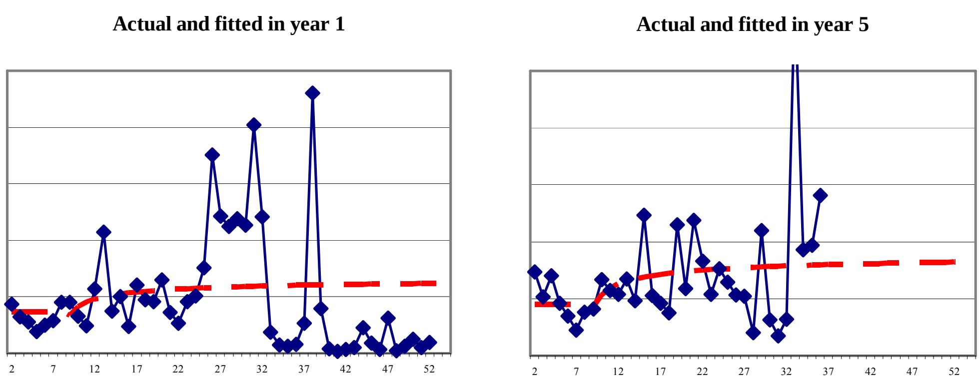
$$\begin{aligned} &PF_{ij}^{L} \sim Gamma(\mu_{ij}^{L}, v_j) \\ &\mu_{ij}^{L} = \exp\Big\{\beta_{i0} + \mathrm{I}(L=0)\Big[\beta_{i1}\max\big(1, j-9\big)^{-0.5} + \beta_{i2}\mathrm{I}(j \le 9)\Big] \\ &\qquad + \mathrm{I}(L=1)\Big[\beta_{i3} + \beta_{i4}\big(j-1\big)^{-0.5} + \beta_{i5}\max\big(0, 7-j\big)^{2}\Big]\Big\} \end{aligned} \qquad (5.8)$$

where

$\mu_{ij}^{L}$ = the mean $PF_{ij}^{L}$ in cell ($i,j$); and

$v_j$ = coefficient of variation of the payment factors in development quarter $j$.

**Figure 5.4 Modelled payment factors**



Examples of the actual data and fitted model are shown in Figure 5.4.

**Case estimate development factors**

Traditionally, the case estimate development factors model is based on the development of the hindsight case estimates. Thus, the modelled quantities are defined, for $j$=2, 3, … as:

$$CEDF_{ij}^{L} = \left(E_{ij}^{L} + C_{ij}^{L}\right) \Big/ E_{i,j-1}^{L} \qquad (5.9)$$

However, comparison with (5.7) indicates that, based on this definition, the $\{CEDF\}$ and $\{PF\}$ would not be independent. Although this is inconsequential for the traditional deterministic application of the PCE model, it does matter for the stochastic application. Therefore, in this paper, the following definition of a case estimate development factor is used:

$$CEDF_{ij}^{L} = E_{ij}^{L} \Big/ E_{i,j-1}^{L} \qquad (5.10)$$

To preserve the relationship with amount paid in a cell ($i$,$j$), the payment factors are included within the model for the $\{CEDF\}$. Thus, the model is:

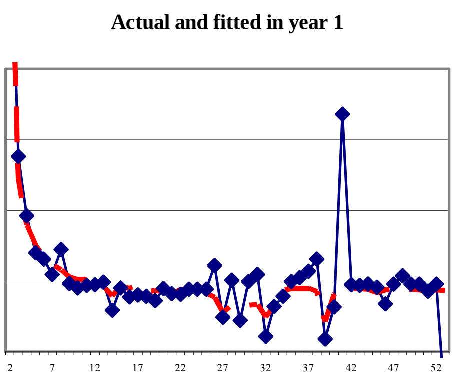
$$\begin{aligned} &CEDF_{ij}^{L} \sim Gamma(\mu_{ij}^{L}, v_j) \\ &\mu_{ij}^{L} = \exp\Big\{\beta_{i0} + \beta_{i1}PF_{ij}^{L} + \mathrm{I}(L=0)\left[\beta_{i1}j^{-1} + \beta_{i2}\mathrm{I}(j=1)\right] \\ &\qquad + \mathrm{I}(L=1)\left[\beta_{i3} + \beta_{i4}j^{-1} + \beta_{i5}\mathrm{I}(j=1)\right]\Big\} \end{aligned} \qquad (5.11)$$

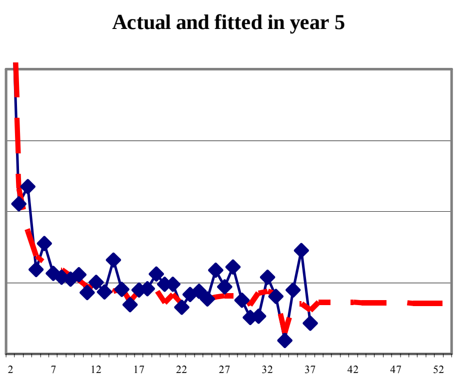
where

$\mu_{ij}^{L}$ = the mean $CEDF_{ij}^{L}$ in cell ($i$,$j$); and

$v_j$ = coefficient of variation in development quarter $j$.

**Figure 5.5 Modelled case estimate development factors**



Examples of the actual data and fitted model are shown in Figure 5.5.

## 5.4 Results

The models described above were applied to the data using the GLM filter based robot described in Section 4.

Table 5.1 displays the results for each of the three models for jurisdiction 0, together with the bootstrapped estimates of coefficient of variation. Note that the results have been scaled for confidentiality reasons. The PPCF and PCE models perform best as measured by coefficient of variation. For these data, the three models give broadly consistent results for all years but the most recent.

**Table 5.1 Results for jurisdiction 0**

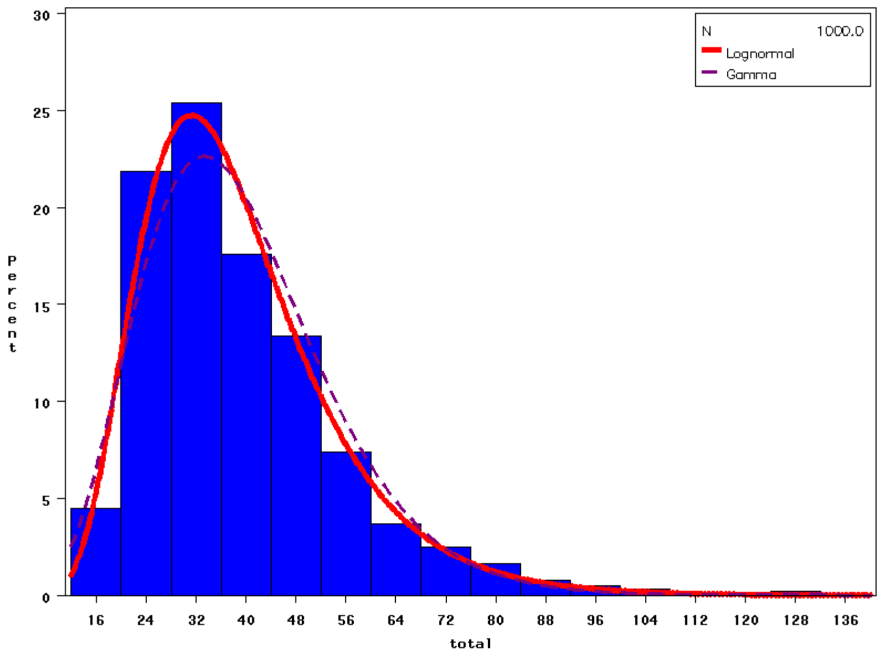
| Accident | PPCI | | PPCF | | PCE | |
|---|---|---|---|---|---|---|
| year | Mean | CV | Mean | CV | Mean | CV |
| 1 | 8 | 229% | 132 | 55% | 22 | 105% |
| 2 | 20 | 216% | 242 | 47% | 56 | 108% |
| 3 | 58 | 166% | 165 | 58% | 23 | 98% |
| 4 | 110 | 135% | 268 | 47% | 70 | 90% |
| 5 | 242 | 100% | 861 | 30% | 317 | 62% |
| 6 | 292 | 71% | 1,216 | 27% | 671 | 64% |
| 7 | 680 | 59% | 1,257 | 27% | 799 | 44% |
| 8 | 819 | 53% | 1,672 | 27% | 1,319 | 40% |
| 9 | 2,262 | 49% | 3,366 | 25% | 2,040 | 32% |
| 10 | 3,546 | 49% | 3,510 | 22% | 2,368 | 31% |
| 11 | 6,363 | 48% | 6,041 | 21% | 5,480 | 31% |
| 12 | 7,151 | 46% | 6,742 | 20% | 6,700 | 31% |
| 13 | 8,461 | 44% | 8,664 | 21% | 7,234 | 33% |
| 14 | 8,904 | 42% | 9,015 | 21% | 3,749 | 98% |
| Total ex 14 | 30,011 | | 34,136 | | 27,099 | |
| Total | 48,589 | 42% | 41,721 | 18% | 29,366 | 22% |

Table 5.2 displays the results for the other jurisdictional grouping of data. It is seen that the PPCF and PCE models perform best for these data. The PPCI model with universally high coefficients of variation performs poorly. These data show changes in the rates of finalisation in recent years, which changes the profile of average payments paid per development period, which in turn, impacts on the reliability of the PPCI model. Further, the results of the PPCF and PCE models are quite similar (ignoring the most recent year – PCE models generally perform poorly on recent accident periods) while the PPCI model gives substantially higher results.

**Table 5.2 Results for jurisdiction 1**

| Accident | PPCI | | PPCF | | PCE | |
|---|---|---|---|---|---|---|
| year | Mean | CV | Mean | CV | Mean | CV |
| 1 | 45 | 224% | 6 | 232% | 0 | 0% |
| 2 | 118 | 210% | 146 | 58% | 215 | 148% |
| 3 | 159 | 187% | 36 | 116% | 0 | 0% |
| 4 | 140 | 205% | 236 | 45% | 72 | 112% |
| 5 | 236 | 159% | 226 | 48% | 52 | 73% |
| 6 | 329 | 113% | 365 | 40% | 417 | 60% |
| 7 | 440 | 98% | 686 | 34% | 565 | 54% |
| 8 | 635 | 94% | 780 | 34% | 519 | 53% |
| 9 | 1,824 | 90% | 1,430 | 27% | 1,303 | 48% |
| 10 | 5,654 | 83% | 6,470 | 24% | 6,225 | 55% |
| 11 | 8,246 | 84% | 6,219 | 21% | 6,923 | 59% |
| 12 | 9,023 | 83% | 6,580 | 20% | 5,536 | 65% |
| 13 | 9,210 | 79% | 7,716 | 21% | 5,713 | 74% |
| 14 | 12,530 | 77% | 10,826 | 21% | 1,827 | 128% |
| Total ex 14 | 36,058 | | 30,895 | | 27,539 | |
| Total | 38,915 | 76% | 43,151 | 19% | 30,849 | 45% |

**Figure 5.6 Distribution of PPCI results**




Histograms of the bootstrapped distributions for jurisdiction 0 for each of the models are presented in Figure 5.6, Figure 5.7 and Figure 5.8.

**Figure 5.7 Distribution of PPCF results**

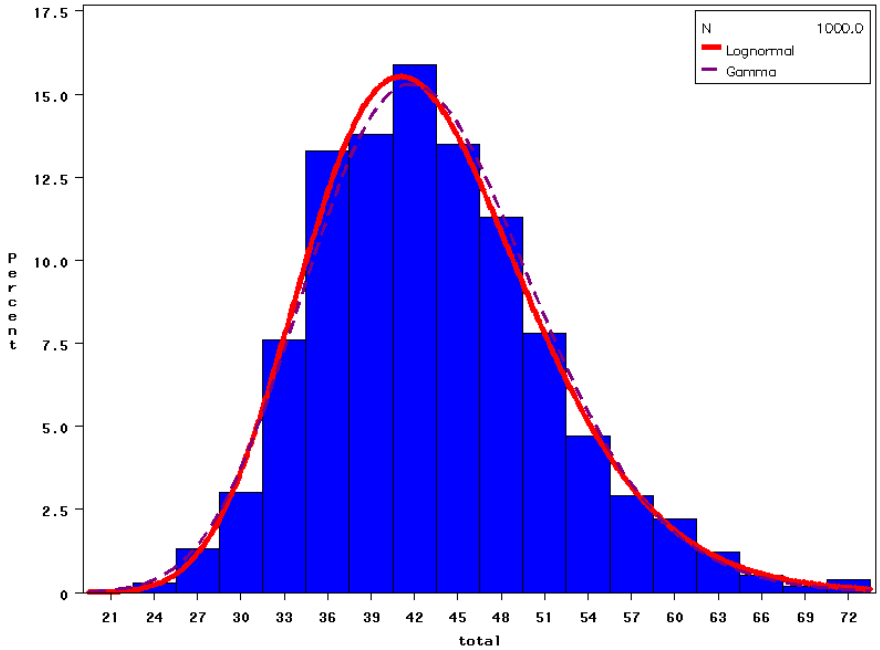
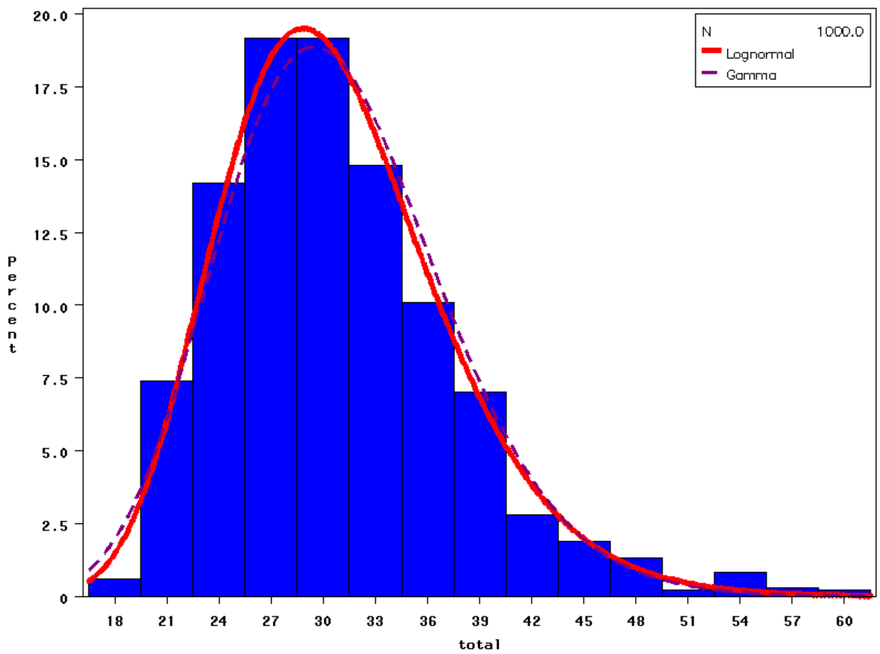
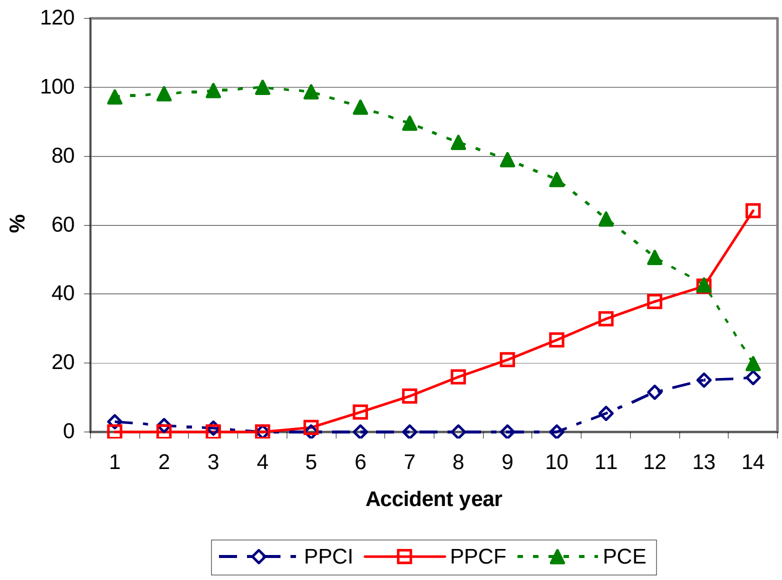



**Figure 5.8 Distribution of PCE results**



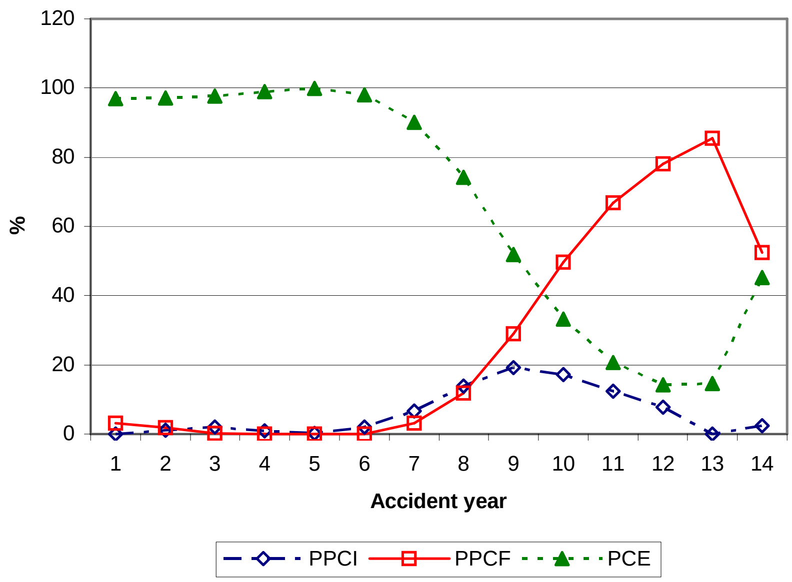

The blending algorithm discussed in McGuire and Taylor (2007) is applied. Through some inputted constants, this takes account of the predictive variances of the results, the smoothness of the weights and the smoothness of the progression of the blended estimates to the case estimates in determining final weights. The reader is referred to McGuire and Taylor (2007) for details of the input selections made in the production of the weights. The model blending weights are shown in Figure 5.9 and Figure 5.10 for each of the two jurisdictional groupings.

**Figure 5.9 Selected weights for jurisdiction 0**



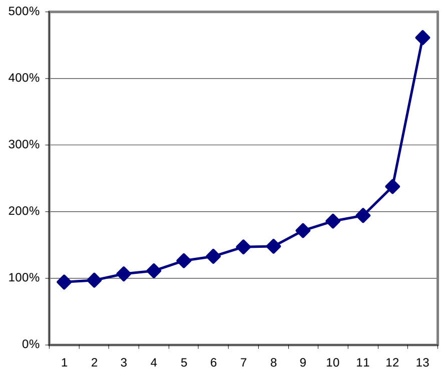

A final point to note is the relatively high weight assigned to the PCE model for accident year 14, particularly for jurisdictional grouping 1. Although the PCE results for year 14 have a high coefficient of variation (refer to Table 5.1 and Table 5.2), the quantum of liability is small relative to the other models, meaning in turn, that the standard deviation in absolute terms is low. The practitioner may wish to judgementally increase the standard deviation to put it on more similar terms with the other models; however, this has not been implemented here.

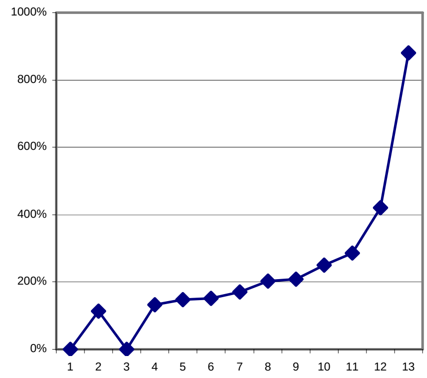
**Figure 5.10 Selected weights for jurisdication 1**



The ratio of the blended liability (that results from the weights above) to case estimates is shown in Figure 5.11 in which the horizontal axis relates to accident years. Given the immaturity of the case estimates in year 14 (the most recent), this comparison has been omitted. It is seen that the progression is satisfactory. Ratios of zero appear for Years 1 and 3 in jurisdiction 1, but in fact case estimates are zero in these years.

**Figure 5.11 Ratio of blended results to case estimates for L=0 (left graph) and L=1 (right graph)**




The final results are given in Table 5.3. It is observed that the coefficients of variation are generally lower than the individual coefficients of variation from each component model, particularly for the coefficient of variation for the total liability.

**Table 5.3 Blended results**

| Accident year | L=0 | | L=1 | |
|---|---|---|---|---|
| | Mean | CV | Mean | CV |
| 1 | 22 | 104% | 0 | 20051% |
| 2 | 56 | 107% | 213 | 145% |
| 3 | 24 | 96% | 3 | 1413% |
| 4 | 70 | 90% | 73 | 110% |
| 5 | 324 | 60% | 52 | 72% |
| 6 | 702 | 58% | 415 | 59% |
| 7 | 847 | 38% | 561 | 49% |
| 8 | 1,375 | 32% | 567 | 39% |
| 9 | 2,317 | 24% | 1,447 | 32% |
| 10 | 2,672 | 21% | 6,273 | 26% |
| 11 | 5,712 | 20% | 6,649 | 22% |
| 12 | 6,771 | 18% | 6,655 | 20% |
| 13 | 8,035 | 17% | 7,425 | 20% |
| 14 | 7,963 | 20% | 6,811 | 23% |
| Total | 36,891 | 12.7% | 37,144 | 16.6% |

# 6. Discussion

This paper discusses the creation of a reserving robot for automatically carrying out much of the analysis involved in outstanding claims estimation.

Recall the features that any such automatic process, or reserving robot, would need to have, as specified in the introduction:

- The model must be able to adapt to changing experience without human intervention;
- There should be an objective means of measuring the goodness of fit and performance of the model.

Adaptive filters such as the GLM filter satisfy the first criterion. This was illustrated in Sections 4 and 5. They satisfy the second criterion through the use of model diagnostic tests based on residuals and comparisons of actual and expected values as well as the use of the bootstrap to estimate the predictive distribution of the results.

Section 4 discusses the steps required in setting up the reserving robot while Section 5 gives an example of the full robot in action at a valuation date.

Reserving robots such as the one discussed here have to potential to offer considerable time and cost savings to companies with many lines of business. While there is a substantial cost of setting up the models in the first place, significant savings should be realised in subsequent valuations. This comes hand in hand with the implementation of a full stochastic model and the availability of risk margin estimates through bootstrapping.

Of course the robot cannot do the entire reserving job. For instance, the adaptive filters may be applied to gross data with subsequent allowance for recoveries by the actuary (though those could also be filtered). Further, the bootstrapping process as described here will not provide information on the diversification benefit that should be applied to the individual risk margins. However the synchronised bootstrap (a procedure for jointly bootstrapping many lines of business to preserve correlations; Taylor and McGuire, 2007) could be incorporated into the bootstrapping procedure and used for the estimation of the overall risk margin.

A number of items are still under investigation. Firstly, the parameter variance estimation is currently driven by judgement as is standard practice in Bayesian problems. The possibility of developing a more objective way for selection of these variances is being examined. Secondly, the model diagnostics need to be manually examined each quarter. It may be possible to develop some key indicators to give warning when possible evidence of model misfit has been found – a case of building primitive eyes for the robot.

It is likely that there is a wide range of potential application for a robotic reserver. It is also equally likely that there will be lines of business which are not suitable for an automated stochastic process. Some of these will be obvious – cases where there are little data and/or highly volatile data. Other cases may not be so clear-cut. Therefore, any reserving robot should be run in tandem with existing methodology for a number of valuations until confidence is gained that the robot is sufficient for the task.

**Acknowledgements**

We would like to thank Insurance Australia Group for their support of this work.

# 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.

England, P.D. and Verrall, R.J. (2002). Stochastic claims reserving in general insurance. **British Actuarial Journal**. 8, 443-518.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. **Journal of Basic Engineering**, 82, 340-345.

McCullagh, P. and Nelder, J. A. (1989). **Generalised linear models.** (3rd edition). Chapman and Hall, New York.

McGuire, G. and Taylor, G. (2007). Adaptive Reserving using Bayesian revision for the Exponential Dispersion Family. Taylor Fry working paper, available from www.taylorfry.com.au or by emailing grainne.mcguire@taylorfry.com.au.

Nelder, J.A. and Wedderburn, R. W. M. (1972). Generalized linear models. **Journal of the Royal Statistical Society, Series A.** 135, 370-384.

Pinheiro, P. J. R., Andrade e Silva, J. M. and de Lourdes Centeno, M. (2003). Bootstrap methodology in claim reserving. **Journal of Risk and Insurance**. 70(4), 701-714.

Stoffer, D.S. and Wall, K.D. (1991). Bootstrapping state=space models: Gaussian maximum likelihood estimation and the Kalman Filter. **Journal of the American Statistical Association**, 86 (416), 1024:1033.

Taylor, G. (1985). Combination of estimates of outstanding claims in non-life insurance. **Insurance: Mathematics and Economics**. 4, 81:91

Taylor, G. (2000). **Loss reserving: an actuarial perspective.** Kluwer Academic Publishers. London, New York, Dordrecht.

Taylor, G. and McGuire, G. (2007) A synchronised bootstrap to account for dependencies between lines of business in the estimation of loss reserve prediction error. **North American Actuarial Journal**, in press..

Taylor, G. (2008). Second order Bayesian revision of a generalised linear model. **Scandinavian Actuarial Journal**, in press.

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

# Appendix A Gamma distribution/log link filter

The system equation for the Analytical Gamma GLM filter is the same as that in (2.1). However, the observation equation differs due to the log link and is

$$Y(s) = \exp\{X(s)\beta(s \mid s-1)\} + v(s) \qquad (A.1)$$

Some other quantities of importance within the Gamma GLM filter are

$$\begin{aligned} 1/h^{-1}(s \mid s-j) &= \mathrm{E}\left[\exp\{-M(s \mid s-j)\beta(s-1 \mid s-1)\}\right] \\ \Gamma(s \mid s-j) &= \mathrm{Var}\left[\exp\{-M(s \mid s-j)\beta(s-1 \mid s-1)\}\right] \end{aligned} \qquad (A.2)$$

where $M(s|s\text{-}j)$ is defined in (A.5) and (A.14) below.

The Gamma GLM filter is applied to these quantities, from which the parameter means and variances are derived. The steps of the filter are:

**Step 1: parameter estimates for new epoch based with no new information**

$$\mathrm{E}\left[\beta(s) \mid Y(s-1)\right] = \Phi(s)\,\mathrm{E}\left[\beta(s-1) \mid Y(s-1)\right] \qquad (A.3)$$

$$\mathrm{Var}\left[\beta(s) \mid Y(s-1)\right] = \Phi(s)\,\mathrm{Var}\left[\beta(s-1) \mid Y(s-1)\right]\Phi(s) + R(s) \qquad (A.4)$$

where $\Phi(s+1)$ is the identity matrix in the example of filtering in this paper (where filtering is applied by row, each row having different parameters).

Let $h^{-1}(s|s\text{-}j)$ be the reciprocal of $1/h^{-1}(s|s\text{-}j)$. Define $M(s|s\text{-}1)$ and $Q(s|s\text{-}1)$ as the orthogonal and diagonal matrices respectively satisfying

$$M(s \mid s-1)^T Q(s \mid s-1) M(s \mid s-1) = Var\left[\beta(s) \mid Y(s-1)\right] \qquad (A.5)$$

Also define:

$$\frac{1}{h^{-1}(s \mid s-1)} = \left[I + 0.5\,Q(s \mid s-1)\right]\exp\left\{-\mathrm{E}\left[M(s \mid s-1)\beta(s) \mid Y(s-1)\right]\right\} \qquad (A.6)$$

$$\Gamma(s \mid s-1) = Q(s \mid s-1)\,\mathrm{diag}\exp\left\{-2\,\mathrm{E}\left[M(s \mid s-1)\beta(s) \mid Y(s-1)\right]\right\} \qquad (A.7)$$

**Step 2: taking the new data into account**

Define $N$, $W$ as

$$N(s\mid s-1)=[\Gamma(s\mid s-1)]^{-1}[1+\tfrac{1}{2}Q(s\mid s-1)] \\ \operatorname{diag}\exp\left\{-M(s\mid s-1)E\left[\beta(s)\mid Y(s-1)\right]\right\} \tag{A.8}$$

$$W(s\mid s-1)=N(s\mid s-1)\left[\frac{1}{h^{-1}(s\mid s-1)}\right] \tag{A.9}$$

where diag($v$) means the diagonal matrix corresponding to the vector $v$.

Then

$$\beta(s\mid s-1)=-M(s\mid s-1)^{T}\log\left(N(s\mid s-1)^{-1}W(s\mid s-1)\right) \tag{A.10}$$

Further define:

$$B(s\mid s-j)=-\operatorname{diag}\exp\left\{-M(s\mid s-j)\beta(s\mid s-1)\right\} \qquad j=0,1 \tag{A.11}$$

$$G(s)=-\operatorname{diag}\exp\left\{-X(s)\beta(s\mid s-1)\right\} \tag{A.12}$$

and $P(s)$ and $D(s)$ as the orthogonal and diagonal matrices respectively satisfying

$$P(s)^{T}D(s)P(s)=N(s\mid s-1)B(s\mid s-1) \\ +M(s\mid s-1)X(s)^{T}\operatorname{diag}\left[G(s)\Lambda(s)Y(s)\right]X(s)M(s\mid s-1)^{T} \tag{A.13}$$

Then

$$M(s\mid s)=P(s)M(s\mid s-1) \tag{A.14}$$

and

$$J(s)=P(s)^{T}B(s\mid s)P(s) \\ \times\begin{bmatrix} N(s\mid s-1)B(s\mid s-1)+M(s\mid s-1)X(s)^{T} \\ \operatorname{DIAG}\left\{G(s)\Lambda(s)y(s)\right\}X(s)M(s\mid s-1)^{T}\end{bmatrix}^{-1} \\ \times M(s\mid s-1)X(s)^{T}\Lambda(s) \tag{A.15}$$

Finally, measuring the extent to which the new data should be allowed to adjust the parameter estimates:

$$K(s) = B(s \mid s)^{-1} P(s) J(s) G(s) \qquad \textit{(A.16)}$$

**Step 1a: estimates of the observations without new data**

(Note this is defined here rather than in Step 1 due $\beta(s|s\text{-}1)$ forming part of the definition)

$$\hat{Y}(s \mid s-1) = \exp\{X(s)\beta(s \mid s-1)\} \qquad \textit{(A.17)}$$

**Step 3: updating the parameters to take account of the new data**

$$1/h^{-1}(s \mid s) = \operatorname{diag}\exp\left[-M(s \mid s)\beta(s \mid s-1)\right]\left[1 - K(s)\left(Y(s) - \hat{Y}(s \mid s-1)\right)\right] \qquad \textit{(A.18)}$$

where 1 represents a vector with all entries unity, rather than the unit matrix.

$$\Gamma(s \mid s) = -D(s)^{-1} B(s \mid s)^2 \qquad \textit{(A.19)}$$

$$E\left[\beta(s) \mid Y(s)\right] = M(s \mid s)^T \left\{-\log\frac{1}{h^{-1}(s \mid s)} + \frac{1}{2}\Gamma(s \mid s) H_{(-1)}(s \mid s)^2 1\right\} \qquad \textit{(A.20)}$$

$$Var\left[\beta(s) \mid Y(s)\right] = M(s \mid s)^T H_{(-1)}(s \mid s)\Gamma(s \mid s) H_{(-1)}(s \mid s) M(s \mid s) \qquad \textit{(A.21)}$$

In (A.21) and (A.20),

$$H_{(-1)}(s \mid s) = \operatorname{diag} h^{-1}(s \mid s) \qquad \textit{(A.22)}$$

where $h^{-1}(s|s)$ is the reciprocal of $1/h^{-1}(s|s)$.

The filter is initialised by setting initial values (i.e. prior values) for $\mathrm{E}\left[\beta(0) \mid Y(0)\right]$ and $\mathrm{Var}\left[\beta(0) \mid Y(0)\right]$. Further, input values are required for $\Lambda(s)$ - for the gamma distribution, the coefficients of variation of the data *Y(s)* are the appropriate choice. The $R(s)$ may be all be set to a constant matrix.

# Appendix B Motor bodily injury data

## B.1 Gross payments

These have been taken from Appendix 3, Table B.3.5 in Taylor (2000)

| Accident | Payments in $(31/12/95) in $(000) in development year | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| year | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 1980 | 4,677 | 9,989 | 8,746 | 10,228 | 8,572 | 5,786 | 3,855 | 1,445 | 1,612 | 626 | 1,172 | 589 | 438 | 473 | 446 | 0 |
| 1981 | 5,288 | 8,089 | 12,839 | 11,830 | 7,760 | 6,182 | 4,118 | 3,016 | 1,775 | 1,785 | 2,645 | 266 | 38 | 45 | 159 | |
| 1982 | 2,924 | 9,869 | 10,242 | 13,808 | 8,785 | 5,409 | 2,425 | 1,597 | 2,149 | 3,296 | 917 | 295 | 428 | 359 | | |
| 1983 | 3,600 | 7,514 | 8,247 | 9,327 | 8,584 | 4,245 | 4,096 | 3,216 | 2,014 | 592 | 1,188 | 691 | 367 | | | |
| 1984 | 3,642 | 7,394 | 9,838 | 9,734 | 6,377 | 4,884 | 11,920 | 4,189 | 4,492 | 1,760 | 944 | 922 | | | | |
| 1985 | 2,463 | 5,033 | 6,980 | 7,722 | 6,702 | 7,834 | 5,579 | 3,622 | 1,300 | 3,069 | 1,370 | | | | | |
| 1986 | 2,267 | 5,959 | 6,175 | 7,051 | 8,102 | 6,339 | 6,978 | 4,396 | 3,107 | 903 | | | | | | |
| 1987 | 2,009 | 3,701 | 5,297 | 6,886 | 6,496 | 7,550 | 5,855 | 5,751 | 3,871 | | | | | | | |
| 1988 | 1,860 | 5,282 | 3,650 | 7,528 | 5,156 | 5,766 | 6,862 | 2,573 | | | | | | | | |
| 1989 | 2,331 | 3,517 | 5,310 | 6,066 | 10,149 | 9,265 | 5,262 | | | | | | | | | |
| 1990 | 2,314 | 4,486 | 4,113 | 6,999 | 11,163 | 10,058 | | | | | | | | | | |
| 1991 | 2,607 | 3,952 | 8,228 | 7,905 | 9,307 | | | | | | | | | | | |
| 1992 | 2,595 | 5,404 | 6,578 | 15,546 | | | | | | | | | | | | |
| 1993 | 3,155 | 4,975 | 7,961 | | | | | | | | | | | | | |
| 1994 | 2,626 | 5,703 | | | | | | | | | | | | | | |
| 1995 | 2,827 | | | | | | | | | | | | | | | |

## B.2 Ultimate claim numbers

These have been taken from Table 2.4 in Taylor (2000)

| Accident year | Ultimate no. of claims |
|---|---|
| 1980 | 779 |
| 1981 | 930 |
| 1982 | 894 |
| 1983 | 964 |
| 1984 | 982 |
| 1985 | 938 |
| 1986 | 957 |
| 1987 | 855 |
| 1988 | 874 |
| 1989 | 873 |
| 1990 | 816 |
| 1991 | 870 |
| 1992 | 900 |
| 1993 | 886 |
| 1994 | 887 |
| 1995 | 903 |

## B.3 PPCI

| Accident year | PPCI in $(31/12/95) in development year 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1980 | 6,004 | 12,823 | 11,228 | 13,129 | 11,004 | 7,428 | 4,949 | 1,856 | 2,069 | 803 | 1,504 | 757 | 562 | 607 | 573 | 0 |
| 1981 | 5,686 | 8,698 | 13,805 | 12,720 | 8,344 | 6,648 | 4,428 | 3,243 | 1,908 | 1,919 | 2,844 | 286 | 41 | 48 | 171 | |
| 1982 | 3,271 | 11,039 | 11,457 | 15,446 | 9,826 | 6,050 | 2,712 | 1,786 | 2,403 | 3,687 | 1,026 | 331 | 478 | 401 | | |
| 1983 | 3,735 | 7,794 | 8,555 | 9,675 | 8,904 | 4,404 | 4,249 | 3,336 | 2,089 | 614 | 1,233 | 717 | 381 | | | |
| 1984 | 3,708 | 7,530 | 10,018 | 9,912 | 6,494 | 4,973 | 12,138 | 4,265 | 4,574 | 1,792 | 961 | 938 | | | | |
| 1985 | 2,625 | 5,366 | 7,442 | 8,233 | 7,145 | 8,352 | 5,948 | 3,861 | 1,386 | 3,271 | 1,461 | | | | | |
| 1986 | 2,369 | 6,227 | 6,452 | 7,368 | 8,466 | 6,624 | 7,291 | 4,594 | 3,246 | 944 | | | | | | |
| 1987 | 2,349 | 4,328 | 6,196 | 8,053 | 7,598 | 8,831 | 6,848 | 6,726 | 4,528 | | | | | | | |
| 1988 | 2,128 | 6,043 | 4,177 | 8,614 | 5,900 | 6,598 | 7,851 | 2,944 | | | | | | | | |
| 1989 | 2,670 | 4,028 | 6,082 | 6,948 | 11,626 | 10,613 | 6,027 | | | | | | | | | |
| 1990 | 2,836 | 5,498 | 5,040 | 8,578 | 13,680 | 12,325 | | | | | | | | | | |
| 1991 | 2,996 | 4,543 | 9,458 | 9,086 | 10,698 | | | | | | | | | | | |
| 1992 | 2,883 | 6,004 | 7,309 | 17,273 | | | | | | | | | | | | |
| 1993 | 3,561 | 5,615 | 8,985 | | | | | | | | | | | | | |
| 1994 | 2,961 | 6,430 | | | | | | | | | | | | | | |
| 1995 | 3,131 | | | | | | | | | | | | | | | |

Calculated from B.1 and B.2