Section A
INTRODUCTION

Volume 2 of the Claims Reserving Manual was first published in 1989 by the Institute of Actuaries. At that time, Volume 2 consisted of five papers covering more advanced reserving methods, loosely described as “statistical” methods. The criteria for inclusion in Volume 2 were that the methods should be “statistical”, had been used by a practitioner and had been found to be of value.

The initial edition of Volume 2 did not include any commentary on, or summaries of, the five original papers, nor did it attempt to present them in any sort of context. The present edition includes a précis of each of the original papers, so that the reader can see the contents of each paper at a glance. In addition, two new papers have been added to the Manual. Further papers will be added in the future, as appropriate.

Since the Claims Reserving Manual was first published, a considerable number of actuarial papers on reserving have been published in a variety of journals, and some papers have been offered to the Faculty and Institute of Actuaries for inclusion in the Claims Reserving Manual. Clearly, not all the papers published or submitted since the initial edition of Volume 2 can be reproduced or referred to in the Claims Reserving Manual.

However, so that the reader is aware of some of this further work, précis of other selected papers have also been added to Volume 2. The criteria for inclusion are that the paper either puts forward a new approach to a claims reserving model, or gives some useful refinement of, or variation on, an existing model. The intention is that these summaries will be added to over time.

Whilst all the Volume 2 papers include an example where appropriate, illustrating the use of the models, it was also felt that, with the widespread use of personal computers, it would be useful to issue a disk with an illustration on a spreadsheet of the application of the models. This should further aid the reader’s understanding of the model, and assist any readers who want to try out the models in practice.

This revision of the Claims Reserving Manual therefore includes a disk and additional description of two of the Volume 2 papers. Further computerised illustrations are planned for other models where it is felt that this would be useful.

<>

Section B
DESCRIPTION OF STOCHASTIC MODELS

[B1]
WHAT IS A STOCHASTIC MODEL?

Section 2B of the Supplementary Introduction to Volume 1 gives a general description of reserving methodology. In that description, the process of arriving at an estimate of future payments is described as one of constructing a model, fitting it to some set of past observations, and using it to infer results about the future — in this case, the future events we are interested in are the payment of claims. Several distinctions are made between different types of model, including those between deterministic and stochastic models.

Deterministic reserving models are, broadly, those which only make assumptions about the expected value of future payments. Stochastic models also model the variation of those future payments. By making assumptions about the random component of a model, stochastic models allow the validity of the assumptions to be tested statistically, and produce estimates not only of the expected value of the future payments, but also of the variation about that expected value.

All the methods in Volume 2 could be described as stochastic to a greater or lesser extent. One can distinguish between them a little, since the methods described in Sections D1, D4, D5, D6 and D7 all allow the user to make estimates of the variation about the expected future payments. The methods described in sections D2 and D3, however, simply involve the fitting of curves to sets of data. The curves are then used to predict future payments, but do not allow the modeller to make estimates of the variation of these payments.

A further distinction can be made between those models based on individual claims, and those which project grouped claims data. This distinction is most commonly found amongst stochastic methods, although the only methods presently in Volume 2 which model individual claims information are those explained in Sections D4 and D7.

<>

[B2]
WHAT ARE THE ADVANTAGES / DISADVANTAGES OF A STOCHASTIC MODEL?

This section briefly highlights some of the perceived advantages and disadvantages of stochastic models, to give the reader some idea of their strengths and weaknesses.

Section 2B of the Supplementary Introduction to Volume 1 observed that deterministic models may often be applied without a clear recognition of the assumptions one is making. One of the main benefits of a stochastic model is that it is totally explicit about the assumptions being made. Further, it allows these assumptions to be tested by a variety of techniques. Because it models the random variation of future payments, estimates may be made of the likely variability of the estimated future payments.

This allows one to monitor whether the predictions of a model are within the bounds one would expect. For example, a deterministic model simply makes a point estimate of the expected future payments in a given period. The one sure thing one can say about these expected payments, is that the actual payments will be different from expected. Deterministic models do not give you any idea as to whether this difference is significant. Stochastic models enable the modeller to produce a band within which the modeller expects payments to fall with a certain level of confidence, and can be used as an indication as to whether the assumptions of the model hold good.

The strengths of stochastic models can also be their weaknesses.

A stochastic reserving method models an immensely complex series of events with a few parameters. Hence, as with any model, stochastic or otherwise, it is open to the criticism that its assumptions are far too simple and hence unrealistic. Because stochastic models are quite clear and rigid, there is very little scope for incorporating judgement, or extraneous factors into the model.

Finally, stochastic models can be computationally quite complex to perform, and may require a more in-depth statistical and computational ability than some of the more simple deterministic models. This in turn can mean that the results are more difficult to communicate than some of the more simple deterministic models.

<>

[B3]
WHAT MAKES A GOOD STOCHASTIC MODEL?

To appreciate what makes a good stochastic model, it is necessary to understand why one constructs a model in the first place.

Take as an example a set of data on (say) motor claims. This may consist of tens of thousands of claim payments, extending over a number of years. If we simply record each of those amounts individually on reams of paper, the human mind simply cannot grasp the essential characteristics of the data, or discern any pattern, let alone use the data to make sensible predictions.

To understand the data in any meaningful way, therefore, requires the formulation of a pattern that in some way represents the data. In this way, the important characteristics of the data can be represented by a limited number of terms that can be relatively easily understood.

Further, when considering any set of data over time, there will be some systematic influences affecting the claims experience, such as the inflation in the cost of car repairs in our example. There may also be some random influences, such as the variation in the frequency of cars having accidents. To understand the data effectively, one needs to differentiate between systematic influences and random variation.

It is this need to reduce complexity and to separate systematic influences from random variations that leads to a stochastic model. A stochastic model allows the modeller to replace the individual data values by a summary that both describes the essential characteristics of the data by a limited number of parameters, and distinguishes between the systematic and random influences underlying the data.

The parameters of a model are chosen to "fit" the data as closely as possible. The fit can be made better and better by having more and more parameters. However, this then becomes self-defeating, as a model with hundreds of parameters provides no real reduction in complexity from the raw data, and allows the user only a limited ability to grasp the key characteristics of the data.

An essential requirement of a good model, therefore, is that it has enough parameters to describe the characteristics of the data, but not so many that its descriptive power becomes limited. Additionally, as described in the first paper in Volume 2, as you increase the number of parameters of the model, you decrease its predictive power. That is, the model begins to adhere more and more closely to the raw data. Small changes in those data can then lead to large changes in the parameters of the model, making any predictions produced by the model unstable.
A good stochastic model should also enable one to appreciate the systematic influences underlying the data, together with the random influences. Some data points may be subject to considerable random variation, so the model should ensure that it is not

unduly affected by such isolated values. A good stochastic model should therefore be capable of testing the underlying assumptions. By applying such tests, the modeller will gain a greater understanding of the characteristics of the data and, hence, have better control over the projected values.

The above points are, of necessity, fairly general in nature, as any sort of modelling is as much an art as a science. To this end, it is worth observing what it takes to be a good modeller.

The first and most important requirement is to appreciate that all models are "wrong" to some extent. They are not "reality"; they are just a simplified representation of reality, enabling the user to make practical projections of the data. As a consequence, there is no one "right" model, and many different models may be more or less equally applicable.

So, the second requirement is that a good modeller should consider many different models, trying to recognise all those that might be useful, rather than whether they are "right" or "wrong".

A final requirement of the modeller is that they should check the fit of a model. The object of this exercise is to understand the past data, and to infer useful results about the development of those data. This cannot be done rigorously if the modeller does not understand where the model fits or deviates from the data.

<>

Section C
PRÉCIS OF PAPERS IN SECTION D

This section provides a précis of each paper included in Section D of Volume 2. The intention is to give a brief summary of the paper, a description of the reserving model on which the paper is based, and a few observations about the applicability of the model. The précis also deal with what data are required and what level of statistical and computational ability is needed, and offer some thoughts on the strengths and weaknesses of the model.

The numerical heading given to each paper refers to the relevant sub-section within Section D of Volume 2, where the full text of that paper is to be found.

<>

[C1]
THE CHAIN LADDER TECHNIQUE — A STOCHASTIC MODEL
Contributed by B Zehnwirth
(9 pages, see [D1])

Summary

The chain ladder technique is one of the oldest actuarial techniques to be applied widely for estimating loss reserves. It appears intuitively natural and was for some time widely regarded as being based on a non-stochastic model: that is, a model which is deterministic and accordingly does not include a random component.

The paper demonstrates the intimate connection between the chain ladder technique and a two-way analysis of variance model applied to the logarithms of the incremental paid losses. Recognition of this connection reveals the merits and defects of the chain ladder technique more clearly.

Description of the model

The basic model is as follows:

$$\text{Log}(P_{ij}) = Y_{ij} = a_i + b_j + e_{ij} \quad (e_{ij} \text{ are independent identically distributed normal error terms})$$

where Pij are the incremental payments for accident year i, development period j. This model implies that each incremental paid loss, P_{ij}, has a lognormal distribution. The model is fitted by least squares regression or by the application of an algorithm (known as "Expectation-Maximisation", or E–M) for the corresponding two-way analysis of variance.

General comments

The basic statistical chain ladder is generally considered to be over-parameterised, and can be criticised for not including any calendar year effects as part of the model. It is, however, a powerful diagnostic tool for exploring payment/calendar year trends.

It can also form a basis for more sophisticated models, which are not so heavily parameterised, and can include calendar year effects and incorporate additional information into the reserving process.

Some of these extensions to the basic statistical chain-ladder are described in the paper by S Christofides in Section D5, which is also summarised in section C5 of this Volume. Further extensions to the basic model are also described in another paper by B Zehnwirth, which is summarised in section E of this Volume. <>

[C2]
EXPONENTIAL RUN-OFF
Contributed by B Ajne
(11 pages, see [D2])

Summary

The paper describes a model of the exponential run-off of the incremental payments after the first few development years, based on observations for personal injuries in motor insurance. A brief example is provided, as well as possible adjustments for the effect of inflation.

Description of the model

The basic model is as follows:

$$C_{ij} = q_i.C_{ij-1},\ j = \alpha + 1,\ \ldots\ ,\ A - 1$$
$$C_{ij} = 0,\ j \geq A$$

where C_{ij} are the incremental payments for accident year i, development year j.

The q_i are estimated using an algorithm to maximise a likelihood function. The likelihood function is found assuming that:

$$P(X_i = j) = \beta_i .\ q_i^{j-a}$$

where X_i is the number of years between occurrence and settlement for claims occurred in year i ($X_i > \alpha$).

Each amount of claim payment is assumed to be independent of all others.

The reserves are calculated for each year of origin by multiplying the claims paid to date by a ratio based on q_i .

General comments

The concept of exponential run-off is particularly useful for long-tail lines of business. The method is fairly simple mathematically, and the only data required are incremental payments. Provided an equation "solver" is available, it can be programmed and used very easily in any spreadsheet.

The assumptions made by the model are very strong, and it is doubtful whether assumption (2.2) in the paper can ever be properly met in practice. Since this is central to the exponential run-off assumed by the model, it casts some doubt on the validity of the estimates, although the results of the model may still be useful.

The author suggests that an examination of the residuals would be “useful”. In fact, this may more properly be described as “essential”.

<>

[C3.a]
THE CURVE FITTING METHOD
Contributed by S Benjamin and L M Eagles
(9 pages, see [D3.a])

Summary

The paper describes the use of curve fitting to the progression of paid and incurred loss ratios. A Craighead curve (otherwise known as a Weibull distribution) is suggested with up to 3 parameters. A least squares method is proposed for the curve fitting, with graphical examples. The use of curve-fitting is compared with other methods.

Description of the model

The progression of loss ratios is considered by dividing the cumulative claims to date by the estimated ultimate premiums for each year of origin. For each year a Craighead curve y(t) is then fitted to the loss-ratios at time t, as follows:

$$y(t) = A(1 - e^{-(\frac{t}{b})^c})$$

where A is the estimated ultimate ratio, and b and c are parameters. b and c are fitted to all the years of origin, and A varies for each year. For data consisting of a mixture of short tail and long tail business, a double Craighead curve is proposed.

The fitting method is to minimise by iterations $\sum w(t).(y(t) - y_{obs}(t))^2$, where w(t) is the weighting and $y_{obs}(t)$ the observed loss ratio. The use of w(t) allows outliers to be excluded, or the curve to be forced through the most recent data point. Two methods of minimisation by iterations are mentioned, although they are not spelt out in any detail.

General comments

The model was originally intended to be applied to London Market business, but can be used for any type of business, provided that the run-off follows a Craighead curve.

There is no particular reason why the progression of the loss-ratios beyond the data should follow any particular type of curve, so the use of the model to extend the curve beyond the observed data should be treated with some caution.

The data required are paid and incurred claims, together with premiums or some appropriate measure of exposure.

Ideally, the simple visual examination of estimated relative to observed data suggested in the paper should be supplemented by a more formal statistical check of the goodness of fit of the model. The χ^2 test, which can be performed quite easily, is suitable for this purpose.

The paper only requires a few mathematical skills, although implementing the iterative techniques requires a certain level of statistical and computational ability.
The model is non-linear with 3 parameters, so it cannot easily be fitted into a formal spreadsheet. However, the existence of equation "solvers" in many spreadsheets may provide a pragmatic solution to the problem of fitting the curve.

<>

[C3.b]
THE REGRESSION METHOD
Contributed by S Benjamin and L M Eagles
(8 pages, see [D3.b])

Summary

The paper describes and illustrates a method of refining the ultimate loss ratios found by some other method (for example the curve fitting method). A suggestion is given as to how, using graphical means, one can assess likely upper and lower bounds for the estimates of ultimate loss-ratios.

Description of the model

Ultimate loss ratios need to be estimated prior to applying this method. For each year of origin and development year, IBNR loss ratios are determined by:

IBNR loss-ratio(development year t) = Ultimate loss ratio − Incurred loss-ratio(development year t)

For a given development year, a regression line is estimated, based on all the years of account, as:

IBNR loss-ratio(development year t) = a × Incurred loss ratio (development year t) + b

for some fixed a and b.

Reserves are then calculated from this formula.

The regression line can actually be reformulated in terms of credibility:

$$\text{Future claims} = Z \times \frac{a}{Z} \times \text{claims to date} + (1-Z) \times \frac{b}{1-Z} \times \text{premiums}$$

Giving no credibility to the premiums, by regressing with b set equal to zero, is equivalent to using a traditional chain-ladder method.

General comments

The method can be used for any type of business, provided that the ultimate loss ratios are already estimated. It is very easy to implement in a spreadsheet. As the method is based on regression, standard errors of the estimates of the parameters can easily be determined by statistical techniques, as well as by the graphical method suggested.

The paper is easily understandable, but the reader has to be familiar with the principle of regression. The user of the method is given many suggestions as to how the method can be presented simply to, for example, an underwriter.

<>

[C4]
REID'S METHOD
Contributed by D H Reid
(20 pages, see [D4])

Summary

This paper describes a class of models, set out in a series of papers written by the author. It considers the case where a relatively complete set of information on individual claims is available, and where past years' claim patterns may be expected to give insight into the more recent years.

This approach was first described by the author in a paper in the Journal of the Institute of Actuaries, "Claims Reserves in General Insurance", Volume 105, Part III (1978). Subsequent papers in this series are set out at the end of this précis.

The method provides the means by which to establish a probability distribution of claim reserves. Emphasis is given to the process of fitting and re-fitting models as necessary, prior to the extrapolation process. The model is very flexible and allows for the tendency of larger claims to take longer to settle, the proportion of nil claims to vary from one origin year to another, the rate of claim settlement to vary both across and within origin years, and for the effect of inflation on claim costs.

Description of the model

The basic model for the claims arising in a particular origin year consists of a number of components:

1. An underlying bivariate distribution of the cost of positive claims by claim settlement amount and development time.

2. A comparable univariate distribution by development time for nil claims.

3. The proportion of all claims represented by nil settlements.

4. A functional transformation of the settlement time axis from fixed calendar period time to real settlement time (i.e. operational time) represented by the underlying distributions (components 1 and 2).

5. A series of claim cost scale parameters intended to represent cost levels for fixed intervals of operational time, relative to the underlying bivariate distribution.

6. A separate treatment of the largest group of claims by size.

The model assumes that the ordering of claim settlements is not affected by the rate of settlement, and that this ordering is represented by the underlying distributions of components 1 and 2. Recent years' data are fitted to these underlying distributions and components 3–6 are estimated. In conjunction with appropriate assumptions, the fitted parameters are used to extrapolate the incomplete portion of recent years' settlements, from which reserves and reserving distributions are derived.

General comments

The methodology is intended for situations where a detailed analysis of claims behaviour can be obtained. It is likely to be of most relevance for Direct business, where data on amounts and numbers of claims are available by claim size. The method is quite complex, and requires a considerable amount of effort to implement in its fullest form.

The method is very flexible, and can be adapted to embrace more (or less) elaborate models of claim development. It can also be used to help develop sub-models relating claim cost movements to extraneous variables, such as inflation.

A considerable amount of statistical knowledge is required. Some steps in the process require the user to be able to use numerical techniques, for example finding parameters that maximise a likelihood function, without setting out explicitly how this may be achieved.

The original 1978 paper introduced the idea of Operational Time to the context of claim reserving. Although the detailed modelling of the underlying bivariate distribution has now been much simplified in the light of experience, the remainder of the original approach remains valid. The papers in the series, all by D H Reid, are set out below:

1. Claim reserves in general insurance, *Journal of the Institute of Actuaries*, 105, pp 211–296, 1978.

2. Reserves for outstanding claims in non-life insurance, *Transactions of the International Congress of Actuaries*, Zurich and Lausanne, 2, pp 229–241, 1980.

3. A method of estimating outstanding claims in motor insurance with applications to experience rating, *Cahiers du CERO*, Bruxelles, 23, pp 275–289, 1981.

4. Discussion of methods of claim reserving in non-life insurance, *Insurance: Mathematics and Economics* 5, pp 45–56, North Holland, Amsterdam, 1986.

5. Operational time and a fundamental problem of insurance in a data-rich environment, *Applied Stochastic Models and Data Analysis*, 1995, pp 257–269.

<>

[C5]
REGRESSION MODELS BASED ON LOG-INCREMENTAL PAYMENTS
Contributed by S Christofides
(54 pages, see [D5])

Summary

The paper describes a statistical reserving model, based on the logs of the incremental payments. It shows, by a simple example, how such models can be fitted and results derived using a spreadsheet. A more realistic example is then considered, and refinements of the model are described. Because it is a statistical model, standard errors (a measure of the variability of the estimate) for the future incremental payments can be calculated and statistical techniques used to test the fit of the model.

Description of the model

The basic model is as follows:

$$\mathrm{Log}(P_{ij}) = Y_{ij} = a_i + b_j + e_{ij} \quad (e_{ij} \text{ are independent identically distributed normal error terms})$$

where P_{ij} are the incremental payments for accident year i, development period j.

The a_i and b_j are fitted by regression, which can be done automatically in most spreadsheets. The future payments and standard errors are then calculated using matrix manipulation.

Refinements to the basic model are illustrated, including fitting a curve for the development parameters, and adjusting for claims volume and inflation. Models based on curves for the development factors can be useful for estimating tails, as they can be used to project beyond the existing data set.

General comments

The method is of general use and is not restricted to any particular class of business. The only data required are incremental payments. The basic method can be easily programmed in any spreadsheet, although the matrix manipulation necessary to calculate the standard errors may be somewhat time-consuming. Once the basic model has been set up in a spreadsheet, however, the model can be fitted and future payments predicted with very little time or effort for any data set of the same size.

The method does not work for negative incremental payments. There is also a limit to the number of future payments (n) that can be predicted in a spreadsheet, to the largest nxn matrix that a given spreadsheet package can manipulate.
The paper requires a basic level of statistical knowledge. Familiarity with matrix manipulation and regression in a spreadsheet would be helpful, although the worked example sets out all the steps clearly enough for this not to be a necessity.<>

[C6]
MEASURING THE VARIABILITY OF CHAIN LADDER RESERVE ESTIMATES
Contributed by T Mack
(65 pages, see [D6])

Summary

The author has written a series of papers on the subject of the variability of chain-ladder estimates, most notably the CAS prize-winning paper "Measuring The Variability Of Chain Ladder Reserve Estimates". The paper in Section D6 is a reproduction of this paper with some modifications and additions.

The paper derives a formula for the standard error of chain-ladder reserve estimates without assuming any specific claim amount distribution function. For ease of reference, these techniques are described as the "Distribution-free approach".

Description of the model

The foundation of the Distribution-free approach is the observation of three main assumptions which are shown to underlie traditional chain-ladder techniques. These are:

(i) $E(C_{i,k+1}|C_{i1}, ..., C_{ik}) = C_{ik}f_k$, $1 \le i \le I$, $1 \le k \le I - 1$,

(ii) $\{C_{i1}, ..., C_{iI}\}$, $\{C_{j1}, ..., C_{jI}\}$, $i \neq j$, are independent,

(iii) $Var(C_{i,k+1}|C_{i1}, ..., C_{ik}) = C_{ik}\sigma_k^2$, $1 \le i \le I$, $1 \le k \le I - 1$.

Where Cik denotes the accumulated total claims amount of accident year i up to development year k, f_k is the development factor from k to k+1, and σ_k are parameters.

The first two assumptions seem intuitively sensible, although these can be demonstrated to be the implicit assumptions of the formal chain-ladder model. The third assumption is deduced from the fact that the estimator of f_k, is the C_{ik}-weighted mean of the individual development factors.

An important corollary of assumption (i) is that the development factors are not correlated. That is, if we have a particularly high development factor in one period, there is no tendency for the subsequent factor to be particularly low (or high).

The main results of the paper are as follows. The estimate of the standard error of the reserve estimate for accident year i, $\hat{R}_i$, is:

$$\widehat{\mathrm{SE}}\left(\hat{R}_i\right)^2 = \hat{C}_{iI}^2 \sum_{k=I+1-i}^{I-1} \frac{\hat{\sigma}_k^2}{\hat{f}_k^2}\left(\frac{1}{\hat{C}_{ik}} + \frac{1}{\sum_{j=1}^{I-k} C_{jk}}\right)$$

The estimate of the standard error of the reserve estimate for all accident years combined, $\hat{R}$, is:

$$\widehat{\mathrm{SE}}\left(\hat{R}\right)^2 = \sum_{i=2}^{I}\left\{\widehat{\mathrm{SE}}(\hat{R}_i)^2 + \hat{C}_{iI}\left(\sum_{j=i+1}^{I}\hat{C}_{jI}\right)\sum_{k=I+1-i}^{I-1}\frac{2\hat{\sigma}_k^2 / \hat{f}_k^2}{\sum_{n=1}^{I-k} C_{nk}}\right\}$$

A hat indicates an estimator of the particular figure. The derivations of the estimators of C_{ik}, f_k and σ_k are straightforward, and are set out in the paper.

General comments

Although the above formulae look quite daunting, they consist of nothing more than basic arithmetic — addition, multiplication and so on — and are in fact quite easy to implement in a spreadsheet. Once the formulae have been set up, a new set of data can be brought into a spreadsheet. Activating a “calc” to the spreadsheet will then yield the estimates of the standard errors of the reserves for each accident year, and the reserve as a whole, for the new set of data. This is probably one of the easiest ways of obtaining estimates of reserve variability.

There are many potential drawbacks to simple chain-ladder reserve estimates, which are discussed in Volume 1. The approach in this paper does, however, have the significant benefit of making clear the assumptions one is making. Also, because it is a statistical model, it provides a series of diagnostic tools to test whether these assumptions are valid, as well as giving estimates of reserve variability. The use of these diagnostic tools in discussed further in Section F of Volume 2.

To understand fully the proofs in the paper requires a considerable amount of statistical knowledge. However, the general reasoning involved and the final formulae for the standard errors of the reserve estimates are quite simple, and within the reach of most people with a basic grasp of statistics.

<>

[C7]
PROBABILITY DISTRIBUTION OF OUTSTANDING LIABILITY FROM INDIVIDUAL PAYMENTS DATA
Contributed by T S Wright
(20 pages, see [D7]

Summary

The paper describes an approach to estimating future claims using data on individual claim payments, rather than the more usual aggregate data. The approach provides an estimate of the whole probability distribution of the outstanding liability, rather than just the first two moments. This additional information may be used to assess safety loadings of reserve estimates, allowing for the skewness of the distribution of the outstanding liability.

Description of the approach

The approach may be summarised as follows:

(i) Estimate the distribution functions, $F_i(x)$, for the size of payments made in development period i.

(ii) Use a weighted combination of the $F_i(x)$ to estimate the distribution of future payments, F(x).

(iii) Fit a curve to F(x) and discretise the fitted curve so it can be used in a compounding algorithm in step (v).

(iv) Construct a probability distribution for the number of future payments.

(v) Calculate the compound distribution of the amount of future payments based on the estimated probability distribution functions in (iii) and (iv). This is done using Panjer's recursive method.

General comments

The approach relies on the availability of individual claim size information, and is capable of implementation in a spreadsheet. To do so, one needs to be able to fit curves to distributions. The curve-fitting and calculating of the compound distribution would probably be quite time-consuming to implement. The approach is probably of most use for situations where one is not considering a very large number of claims.

The paper does make a few sweeping assumptions, which are not fully spelt out. It is intended, however, to illustrate a pragmatic approach to the use of individual claim size information. The paper illustrates the calculation of a safety loading using the Proportional Hazards criterion , suggested by Wang, which may not be widely known. The use of Panjer's recursive method may also be new to many readers.

The paper requires a moderate level of statistical and computational ability.

<>

Section D
PAPERS OF MORE ADVANCED METHODS

This Section includes the full text of seven papers, covering more advanced reserving methods than those dealt with in Volume 1. In each case, the paper is based on a formal statistical concept, and has been found to be of value when dealing with practical reserving issues. A précis of each paper is also given in Section C, for those who do not wish to read each paper in full.

The papers included are as follows:

D1. The Chain Ladder Technique — A Stochastic Model *by B Zehnwirth*

D2. Exponential Run-Off *by B Ajne*

D3. a. A Curve Fitting Method *by S Benjamin and L M Eagles*
  b. A Regression Method *by S Benjamin and L M Eagles*

D4. Reid's Method *by D H Reid*

D5. Regression Models Based on Log-Incremental Payments *by S Christofides*

D6. Measuring the Variability of Chain Ladder Reserve Estimates *by T Mack*

D7. Probability Distribution of Outstanding Liability from Individual Payments Data *by T S Wright*

<>

[D1]

THE CHAIN LADDER TECHNIQUE — A STOCHASTIC MODEL

Contributed by B Zehnwirth

1. Introduction

The chain ladder technique (equivalently, age-to-age development factors) is one of the oldest actuarial techniques to be applied widely for estimating loss reserves.

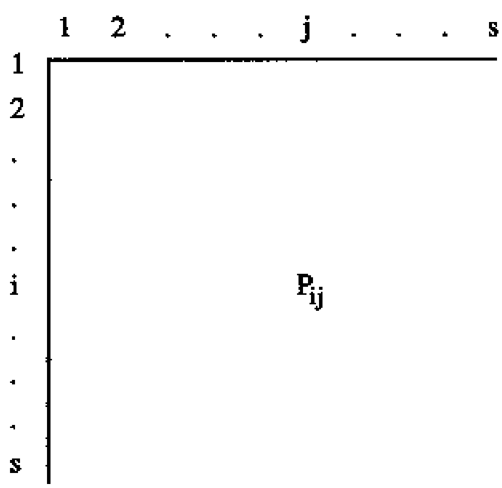
The technique appears intuitively natural and only until more recently was always regarded as being based on a non-stochastic model: that is, a model which is deterministic and accordingly does not include a random component.

The principal objective of this article is to demonstrate the intimate connection between the chain ladder technique and a two-way analysis of variance model applied to the logarithms of the incremental paid losses. Recognition of this connection reveals the merits and defects of the chain ladder technique more clearly.

2. Chain ladder technique

We first review the chain ladder technique in order to indicate two underlying model assumptions. The second model assumption is often not recognised by many users of the technique.

Let P_{ij} represent the incremental paid loss made in development year j, in respect of accident year i. The batch of data P_{ij}, i=1,...,s; j=1,...,s–i+1 is represented as a matrix thus:



Accident years (rows) range from 1 to s and development years (columns) also range from 1 to s.

We denote the cumulative paid loss in development year j, in respect of accident year I by C_{ij}. It is given by:

$$C_{ij} = \sum_{h=1}^{j} P_{ij} \ .$$

A matrix of development factors based on the { C_{ij} } array is constructed by computing the development factor D_{ij} as

$$D_{ij} = \frac{C_{ij}}{C_{i\,j1}} \qquad i = 1,...,s;$$

$$j = 2,...,s–i+1$$

The first basic assumption made is

Assumption 1: Each accident year has the same age-to-age development factors. Equivalently, for each j=2,...,s

$$D_{ij} = D_j \qquad \text{for all } i = 1,2,...s.$$

Under Assumption 1, the most popular estimator of the development factor D_j is the weighted average

$$\hat{D}_j = \frac{\sum_{i=1}^{sj+1} C_{ij}}{\sum_{i=1}^{sj+1} C_{i\,j1}}$$

$$= \frac{\sum_{i=1}^{sj+1} C_{i\,j1} * D_{ij}}{\sum_{i=1}^{sj+1} C_{i\,j1}}$$

The development factor D_{ij} is weighted by the corresponding "volume" measure $C_{i\,j–1}$.

Some users of the chain ladder technique do not use the weighted average estimator of D_j. This is an estimation issue that we address subsequently in this chapter. The fact remains that Assumption 1 is a model assumption associated with the chain ladder technique.

Projections of the quantities C_{ij}; i=2,...,s; j=s–i+2,...s are computed thus:

$$\hat{C}_{ij} = C_{i\,si+1} \prod_{k=s1+2}^{s} \hat{D}_k$$

This technique of projection is explicitly based on the fact that a second model assumption is valid. It is assumed that each accident year has **necessarily** a different level estimated by that year's individual experience. The quantity $C_{i\,s-i+1}$ represents an estimate of the level of accident year i.

Assumption 2: Each accident year has a parameter representing its level. The level parameter for accident year I is estimated by $C_{i\,s-i+1}$.

The last accident year s is represented by the single observation C_{i1}. Were we to assume that accident years are **completely** homogeneous, we should estimate the level of accident years by

$$\sum_{i=1}^{s} C_i 1/s \ ,$$

(or a better estimator of the mean level at development year 1).

Complete homogeneity means that the observations C_{i1}, C_{21},...,C_{s1} are generated by the same mechanism. The chain ladder technique explicitly assumes that the mechanisms generating the incremental paid losses C_{i1}, C_{21},...,C_{s1} are so unrelated that pooling of the information does not afford any increased efficiency. I would find it very difficult to believe that this assumption is **ever** true. In any case, why not find out first what the data indicate?

3. Statistical models related to chain ladder technique

Based on the two assumptions discussed in the preceding section, the following autoregressive model discussed in the paper by Kamreiter and Straub (1973) suggests itself.

$$C_{ij} = D_j\, C_{i\,j-1} + \delta_{ij} \ ; \ i = 1,...,s$$

where the random variables D_j, δ_{ij} and $C_{i\,j-1}$ are independent and satisfy

$$E[\delta_{ij}] = 0, \ E[D_j] = d_j.$$

The quantities { D_j } represent the development factors and are the same for each accident year. Note that it is implicitly assumed that the observations C_{i1}, C_{21},..., C_{s1} are **not** related (at all). Moreover, the additive error term δ_{ij} is questionable — the error term should be multiplicative (see Section 4).

We only remark that the above mentioned model satisfies Assumptions 1 and 2 of the preceding section and devote the remainder of this chapter to a second stochastic model, discussed by Kremer (1982).

The basic model is defined by the multiplicative representation,

$$P_{ij} = a_i' \,.\, b_j' \, e_{ij}' \tag{2.1}$$

where a_i' is the parameter representing the effect of accident year I;
b_j' is the parameter representing the effect of development year j;

and e_{ij}' is a random error term.

By taking logarithms of both sides of equation (2.1), the model may be re-formulated as a two-way analysis of variance model, viz.,

$$Y_{ij} = \log P_{ij} = \mu + a_i + b_j + e_{ij} \tag{2.2}$$

where the parameter μ represents the overall mean effect (on a logarithmic scale), the parameter a_i represents the residual effect due to accident year i and the parameter b_j represents the residual effect due to development year j. It is also assumed that

$$\sum_{i=1}^{s} a_i = \sum_{i=1}^{s} b_j = 0 \tag{2.3}$$

and that $\{ e_{ij} \}$ represent zero mean uncorrelated errors with $Var[e_{ij}]=\sigma^2$.

This model implies that each incremental paid loss P_{ij} has a lognormal distribution.

In the two-way analysis of variance model (2.2), accident year is regarded as a factor at s levels and development year is regarded as a factor at s levels. It is also assumed that the P_{ij}'s are independent random variables having a lognormal distribution with

$$\text{mean} = \exp(\mu + a_i + b_j + 0.5\sigma^2) \tag{2.4}$$

and

$$\text{variance} = \text{mean}^2 * (\exp(\sigma^2)-1) \tag{2.5}$$

Accident year effects and development year effects are assumed to be additive with no interaction. In other words, the effect of an accident year is the same for each development year and vice versa.

We now turn to the estimation of the parameters μ, $\{ a_i \}$ and $\{ b_j \}$.

Model (2.2) is essentially a regression model where the design matrix involves indicator variables. However, the design based on (2.2) alone is singular. In view of constraint (2,3), the actual number of free parameters is 2s–1, yet model (2.2) has 2s+1 parameters. By setting $a_1=b_1=0$, say, the resulting design is non-singular and estimates of parameters can be obtained using a statistical regression package.

Kremer (1982) presents three recursive equations for estimating the parameters μ, $\{ a_i \}$ and $\{ b_j \}$. These equations are essentially solutions to the normal equations of the model described by expression (2.2) and constraint (2.3). If there are no missing data values in the matrix, estimates of the parameters can be obtained using standard methods. When there are too many missing values, and standard methods cannot be used, the following technique, called the E–M algorithm has a fair amount of intuitive appeal.

For a complete matrix the estimates of the parameters are well known:

$$\hat{\mu} = \overline{Y}_{..} = \sum_{i=1}^{s}\sum_{j=1}^{s} Y_{ij} / s^2, \tag{2.6}$$

$$\hat{a}_i = \overline{Y}_{i.} \;\; \overline{Y}_{..}\,, \tag{2.7}$$

and

$$\hat{b}_j = \overline{Y}_{.j} \;\; \overline{Y}_{..}\,, \tag{2.8}$$

where

$$\overline{Y}_{i.} = \sum_{j=1}^{s} Y_{ij} / s \tag{2.9}$$

$$\overline{Y}_{.j} = \sum_{j=1}^{s} Y_{ij} / s \tag{2.10}$$

The E–M algorithm is a recursive technique for finding maximum likelihood estimates in the case of incomplete data. The estimates given by (2.6) to 2.8) are maximum likelihood but are based on a complete matrix. The E in the term "E–M algorithm" stands for Expectation and the M for Maximisation (of the likelihood).

Step 0: Complete the matrix by starting with some initial expected values. For instance, you may enter into the (empty) cell (i,j) the value $y_{i,s-i+1}$.

Step 1: Compute the maximum likelihood estimates for the completed matrix using equations (2.6) to (2.8).

Step 2: Use the estimates û, { $\hat{a}_i$ } and { $\hat{b}_j$ } obtained in Step 1 to compute new expected values û,+ $\hat{a}_i$ + $\hat{b}_j$ for the empty cells (lower triangle).

Now return to Step 1 and continue the recursions until a certain prescribed tolerance is reached, e.g. relative change in all estimates is less than 10^{-3}.

The final estimates û, { $\hat{a}_i$ } and { $\hat{b}_j$ } represent the maximum likelihood estimates. The variance σ^2 is estimated by the Mean Square Error

$$\hat{\sigma^2} = \sum_{i=1}^{s} \sum_{j=1}^{s-i+1} (y_{ij} - \hat{\mu} - \hat{a}_i - \hat{b}_j)^2 / (n - 2s - 1)$$

where n=total number of observations in the upper triangle, viz., s(s+1)/2.

Forecasts of P_{ij} for i=2,..., s and j=s–i+2,..., s are given by

$$\hat{P}_{ij} = \exp(\hat{\mu} + \hat{a}_i + \hat{b}_j + 0.5\hat{\sigma^2}).$$

Note that the two-way analysis of variance model can be applied and estimated for **any shape** array of the **incremental paid losses**. This means that a **formal** chain ladder technique can be applied to any shape array provided n>2s–1.

4. The importance of the log transform — removal of heterogeneity

Loss reservers often describe their data as being heterogeneous. For a long tail line of business, payments are necessarily made over time. Indeed, the main cause of heterogeneity is time itself! Time, almost always, almost everywhere, subjects incremental paid losses (and severities) to one type of heterogeneity we already know about: the variability in incremental paid losses (and in severities) increases as mean level increases.

Let's illustrate this well supported phenomenon with an example. If in 1965 average severity was 1,000 and standard deviation of severity 200, and if in 1988 average severity is 30,000, then the standard deviation of severity in 1988 is probably around 6,000. However, the standard deviation of the logarithms of severities has remained stable between 1965 and 1988. The logarithmic transformation stabilises the variance since it has a standard deviation that is proportional to the mean.

Based on the foregoing discussion, the model

$$P_{ij} = \mu + a_i + b_j + e_{ij}$$

in place of model (2.1) of Section 3, cannot be correct because the variance of the error term e_{ij} will necessarily depend on μ, a_i and b_j.

The foregoing discussion, moreover, also indicates that the geometric mean of development factors is a more efficient estimate of the mean development factor than an arithmetic average.

5. Estimation of development factors

Development factors are typically based on the cumulative paid losses and are ratios of numbers. It is not possible to determine, by eye, if two computed development factors are different in the sense that they are generated by a different process. For example, suppose the incremental paid losses for the first two development years, for two contiguous accident years, are generated by 100 tosses of a symmetric coin. The following scenario may be observed.

		Development Year	
		0	1
Accident	1	41	63
Year	2	59	38

The two computed development factors are 2.537 and 1.644. These, however, are generated by the same process.

Moreover, there is a substantial loss of information when data are cumulated. For instance, a constant incremental paid loss of 100 at every development year has development factors based on cumulative data that asymptote to one, and indeed, even if the incremental paid losses increase according to a polynomial trend, the development factors (based on the cumulative data) asymptote to one. Furthermore, any trends in the payment year direction are different to identify and estimate if the data are cumulated in the development direction.

6. Parameters

Consider the following quadratic trend model representing annual sales of a product,

$$y_t = 2 + 3t^2 + n_t$$

where $t = 1,2,...$ denotes year, y_t sales in year t, and the error terms $\{ n_t \}$ are zero mean and independent from a Normal distribution with variance σ^2.

Suppose we generate the values $y_1, y_2,..., y_7$ (seven year sales figures) and ask a colleague to forecast y_8. We know the complete specification of the model generating the sales including σ^2.

The colleague estimates the following models:

1. *Linear trend*

$$y_t = a_0 + a_1 t + n_t.$$

The regression output indicates that R^2=68% but the residuals appear to have a systematic pattern.

2. *Quadratic trend*

$$y_t = a_0 + a_1 t + a_2 t^2 + n_t$$

For this model R^2=76% and the residuals appear to be in good shape.

The colleague observes that as the number of parameters increases, the quality of fit is improved as measured by R^2. Accordingly, the next model suggests itself.

3. *A polynomial of degree six*

$$y_t = a_0 + a_1 t + \ldots a_6 t^6$$

Here R^2=100% and the fitted curve presents residuals that are all zero.

The colleague presents his forecast as

$$y_8 = a_0 + 8a_1 + \ldots + 8^6 a_6$$

When the colleague presents his solution, we mention to him that the data presented to him had an error, that is, the datum y_4 had been incorrectly generated.

The colleague has now to revise his forecast in the light of this information — the revised forecast is likely to bear no resemblance to the first forecast especially if σ^2 is large!!

The moral of this tale is that the polynomial model used by the colleague produces forecasts that are extremely sensitive to the random component in the data. The forecasts are subject to large uncertainties and accordingly are **not** useful. This is a feature possessed by any model that has many parameters — overparametrisation results in instability. The chain ladder model or technique has many parameters. An array comprising s accident years and s development years involves 2s–1 parameters. In particular, there is an accident year parameter for accident year s where there is only one observation — similarly for development year s.

Every model contains a priori information — the chain ladder **model** contains very little a priori information. The chain ladder model does not contain any information in respect of:

(i) trends and/or patterns in development factors;
(ii) trends across accident years;
(iii) trends across payment years.

Typically in Statistics, a two-way analysis of variance model is applied to a rectangular array involving two factors, each at a number of levels. A factor is a qualitative variable. We normally do not relate the different levels of a factor. For example, when analysing the effects of different soil types and fertilisers on yield of barley, we do not assume some kind of trend or systematic pattern across the fertilisers! It is absurd to treat accident years and development years as factors at different levels, the way we treat different soil types and different fertilisers.

The example involving the sixth degree polynomial gives us some insight as to when the chain ladder technique may work (provided the parameters are estimated efficiently). The chain ladder technique works when the mechanisms generating the paid losses are completely deterministic, that is, $\sigma^2 = 0$, or σ^2 is very close to 0 and development factors are homogeneous. Unfortunately, the real world is not like that.

References

Kramreiter, H. and Straub, E. (1973), On the calculation of IBNR reserves II, *Swiss Actuarial Journal*, 73, pp 177–190.

Kremer, E. (1982), IBNR — claims and the two-way model of ANOVA, *Scandinavian Actuarial Journal*, 1, pp 47–55.

<>

[D2]

EXPONENTIAL RUN-OFF

Contributed by B Ajne

1. Introduction

This method has been used to assess reserves for personal injury liabilities within direct motor third party insurance. It is a relatively straightforward method which simply models claims run-off by an exponential distribution. It is based on the observation that the claims in each development year for a particular year of business often show an exponentially decreasing shape apart perhaps from the first two years of development.

Thus, if the first few development years (often the first two years) are ignored, an exponential model can be applied. Care must obviously be taken that the model fits the data accurately and an examination of the residuals would perhaps be useful. The method has the advantage that prediction is possible for later development years than any in the triangle, unlike the chain ladder method.

A separate model is applied to each year of business written, but the results are inspected for trends and possible pooling of years of incurrence for which there is insufficient data for estimation.

The method is described first without taking account of inflation; inflation is dealt with in section 6.

2. The general case

In this section the model is discussed in detail, without inflation adjustments.

Define C_{ij} =amount paid in development year j in respect of claims incurred in year i.

It is assumed that all payments are made before year A.

i.e. $C_{ij} = 0$ for $j \geq A$ (2.1)

Also, after year α ($\alpha < A$) the claim payments are modelled by

$$C_{i\,j+1} = q_i\, C_{ij} \qquad (2.2)$$

for some fixed q_i.

This is equivalent to an exponential tail, since, under the exponential model,

$$C_{ij} = \lambda_i \ e^{\lambda_i j} \qquad \text{for some } \lambda \tag{2.3}$$

$$\text{then } C_{i\,j+1} = \lambda_i \ e^{\lambda_i (j+1)} = e^{\lambda_i} \ C_{ij} \tag{2.4}$$

and (2.4) is equivalent to (2.2) with a parameter transformation.

The following is a non-rigorous motivation of the likelihood which is used to estimate q_i. Considering one particular incurrence year i, the suffix on q_i is dropped, and it is assumed that there have been T years of run-off ($T > \alpha$).

Let X = development year by the end of which a claim is paid ($X > \alpha$).

$$\text{Then set } P(X = j) = \beta q^{j-\alpha} \qquad j = \alpha,...,T \tag{2.5}$$

Summing over j gives

$$\beta = \frac{1}{\left(\frac{1\ q^{T\ \alpha+1}}{1\ q}\right)} = \frac{1\ q}{1\ q^{T\ \alpha+1}} \tag{2.6}$$

The likelihood function is

$$L(q) = \prod_{j=\alpha}^{T} (\beta q^{j\ \alpha})^{N_{ij}} \tag{2.7}$$

where N_{ij} = number of claims in development year j in respect of claims incurred in year i

It can be shown that, if each pound of claim is independent of the rest, (2.7) can be replaced by

$$L(q) = \prod_{j=\alpha}^{T} (\beta q^{j\ \alpha})^{C_{ij}} \tag{2.8}$$

and (2.8) is used in all cases, even when the above assumption does not hold.

Taking logs of (2.8),

$$\log L(q) = \sum_{j=\alpha}^{T} C_{ij}(\log \beta + (j\ \ \alpha) \log q) \tag{2.9}$$

(2.9) can be maximised and the maximum likelihood estimate of q found. This is done for each row and thus a set of q_i estimates found.
The mathematical maximisation is contained in the appendix: this shows the uniqueness and existence of the maximum. However, it is more

straightforward to maximise (2.9) numerically using a simple search algorithm such as interval bisection. This is illustrated by the example in section 5.

3. Reserves

Let R_{ij} be the claims reserve at the end of the development year j (where $j > \alpha$) assuming no inflation.

So

$$R_{ij} = \sum_{k=\alpha} C_{ik} \tag{3.1}$$

Now

$$R_{ij} = \sum_{k=\alpha} C_{ik} \left[\frac{\sum_{k=j+1}^{\infty} C_{ik}}{\sum_{k=\alpha}^{j} C_{ik}} \right] \tag{3.2}$$

and, according to the model in section 2,

$$\frac{\sum_{k=j+1}^{\infty} C_{ik}}{\sum_{k=\alpha}^{j} C_{ik}} = \frac{\sum_{k=j+1}^{A} \gamma_i q_i^{k\alpha}}{\sum_{k=\alpha}^{j} \gamma_i q_i^{k\alpha}} \quad \text{where } \gamma_i = C_{i\alpha}$$

$$= \frac{\sum_{k=j+1}^{A} q_i^{k\alpha}}{\sum_{k=\alpha}^{j} q_i^{k\alpha}}$$

$$= \frac{q_i^{j\alpha+1}(1\ q_i^{Aj})(1\ q_i)^{1}}{(1\ q_i^{j\alpha+1})(1\ q_i)^{1}}$$

$$= \frac{q_i^{j\alpha}\ \ q_i^{A\alpha}}{q_i^{1}\ \ q_i^{j\alpha}}$$

$$= S_j(q_i), \text{ say} \tag{3.3}$$

Now, since we are using maximum likelihood estimation, the maximum likelihood estimate of the reserve is

$$\hat{R}_{ij} = \sum_{k=\alpha}^{j} C_{ik} S_j(\hat{q}_i) \tag{3.4}$$

The reserves in the example (section 5) have been calculated using (3.4) and the estimate of q_i from section 2.

4. The model in practice

The model which has been used in practice can be summarised in the following table.

For year of incurrence i,

development year	data	model
0	C_{i0}	C_{i0}
1	C_{i1}	C_{i1}
2	C_{i2}	γ_i
3	C_{i3}	$\gamma_i q_i$
4	C_{i4}	$\gamma_i q_i^2$
.	.	.
.	.	.
.	.	.
A−1	$C_{i\ A-1}$	$\gamma_i q_i^{A-3}$
A	C_{iA}	$\gamma_i q_i^{A-2}$
A+1	$C_{i\ A+1}$	0
.	.	.
.	.	.
.	.	.

where $\gamma_i = C_{i2}$.

It can be seen that this case has $\alpha = 2$. From the data, q_i is usually estimated to be around 0.9 for a succession of origin years i, and A is about 19.

Thus, in practice, the first two years are not modelled: the forecasting is applied only to run-off years of delay 3 or more. This means that the two most recent accident years have no forecast values of ultimate claims.

5. Example

The method is illustrated in this section by applying it to some actual data. In the example, the theoretical derivation of q_i is used (as set out in the appendix). As stated earlier, it is much easier to use a simple search method, but the theoretical approach is used in order to illustrate the method.

Year of origin 1974

$$T = 10,\ \frac{(T\ \ 2)}{2} = 4 \qquad \Sigma(j - 2)\ C_{ij} = 30{,}483 \qquad \Sigma C_{ij} = 10{,}335$$

From equation (A.9)

$$\overline{X} = 2.9495 < \frac{T\ \ 2}{2}$$

Hence a solution of f(q) = 0 is needed where

$$f(q) = \frac{9}{1\ \ q^9}\ \frac{1}{1\ \ q} - (8\ \ \overline{X}) \qquad \text{(equation (A10))}$$

As a first approximation, using equation (A11)

$$q = 1\left(\ \frac{T\ \ 2}{2}\ \ \overline{X}\ \right)\frac{12}{T(T\ \ 2)} = 0.842$$

q	f(q)
0.842	5.102 – 5.05 > 0
0.860	4.976 – 5.05 < 0
0.850	5.046 – 5.05 ≈ 0

$\therefore\ \hat{q} \approx 0.85$

Year of origin 1975

$$T = 9,\ \frac{(T\ \ 2)}{2} = 3.5 \qquad \Sigma(j - 2)\ C_{ij} = 24{,}090 \qquad \Sigma C_{ij} = 8{,}354$$

$$\overline{X} = 2.8836 < \frac{T\ \ 2}{2}$$

$$f(q) = \frac{8}{1\ \ q^8}\ \frac{1}{1\ \ q}\ \ (7\ \ \overline{X})$$

As a first approximation, q = 0.883

q	$f(q)$
0.883	$4.143 - 4.1164 > 0$
0.890	$4.103 - 4.1164 < 0$
0.885	$4.131 - 4.1164 > 0$

$\therefore \hat{q} \approx 0.89$

and so on.

Continuing the process for years of origin 1976 to 1981 gives the following table:

Year of origin	$\hat{q}$
1974	0.85
1975	0.89
1976	0.84
1977	0.78
1978	0.74
1979	0.69
1980	0.78
1981	0.79

These values of $\hat{q}$ can be substituted into the formula in section 3 to calculate the reserves.

The simpler search method can be illustrated by considering, for example, year of origin 1974. The values of q and I = logL(q) (which has to be maximised) in the relevant range are

q	I
0.82	–21875.7
0.83	–21854.2
0.84	–21841.7
0.85	–21837.9
0.86	–21842.7
0.87	–21855.8
0.88	–21877.2

Thus the maximum likelihood estimate of q is

$$\hat{q} \approx 0.85 \qquad \text{(as before)}$$

For the most recent years of origin there is very little data to use in the estimation procedure, and an IBNR computation is needed. For these years a "smoothed" common q value may be chosen which is a conservative estimate (e.g. 0.85 or 0.90) in the sense that it over-reserves: it is preferable that the predicted claims should be greater than the actual claims.

6. Adjustment for future inflation

Future inflation can be taken into account by modifying the claims reserve at the end of year j, R_{ij}.

Suppose future inflation with inflation factor r per year is to be taken into account.

This implies that R_{ij} has to be increased by a factor

$$\frac{\sum_{k=j+1}^{A} \gamma_i\, q_i^{k\alpha}\, r^{kj\text{‰}}}{\sum_{k=j+1}^{A} \gamma_i\, q_i^{k\alpha}}$$

$$= \frac{q_i^{j\alpha+1}\, r^{\text{‰}}(1\ \ (q_i\, r)^{Aj})\,(1\ \ q_i\, r)^{1}}{q_i^{j\alpha+1}(1\ \ q_i^{Aj})\,(1\ \ q_i)^{1}}$$

$$= \frac{r^{\text{‰}}[1\ \ (q_i\, r)^{Aj}](1\ \ q_i)}{(1\ \ q_i^{Aj})(1\ \ q_i\, r)} \qquad \text{if } q_i\, r \neq 1$$

or

$$= \frac{r^{\text{‰}}(A\ \ j)(1\ \ q_i)}{1\ \ q_i^{Aj}} \qquad \text{if } q_i\, r = 1 \qquad (6.1)$$

This factor is called $I_j(q_i, r)$.

If future inflation is to be taken into account, but its influence limited to n years ahead, then the factor by which R_{ij} has to be increased is instead (for $j \le A - n$)

$$\frac{\sum_{k=j+1}^{j+n1} \gamma_i q_i^{k\alpha} r^{kj\text{‰}} + \sum_{k=j+n}^{A} \gamma_i q_i^{k\alpha} r^{n\text{‰}}}{\sum_{k=j+1}^{A} \gamma_i q_i^{k\alpha}}$$

$$= \quad \frac{r^{\text{‰}}\ q_i^{j\alpha+1}\ (1 \quad (q_i\ r)^{n1})\ (1 \quad q_i\ r)^{1}\ +\ r^{n\text{‰}}\ q_i^{j+n\alpha}\ (1 \quad q_i^{Ajn+1})\ (1 \quad q_i)^{1}}{q_i^{j\alpha+1}\ (1 \quad q_i^{Aj})\ (1 \quad q_i)^{1}}$$

$$= \quad \frac{r^{\text{‰}}(1\ (q_i\ r)^{n1})(1\ q_i) + r^{n\text{‰}} q_i^{n1} (1\ q_i^{Ajn+1})(1\ q_i\ r)}{(1\ q_i^{Aj})\ (1\ q_i\ r)} \qquad \text{if } q_i\ r \neq 1$$

or

$$= \quad \frac{r^{\text{‰}}(n\ 1)\ (1\ q_i) + r^{\text{‰}}\ (1\ q_i^{Ajn+1})}{(1\ q_i^{Aj})} \qquad \text{if } q_i\ r = 1$$

(6.2)

This factor is called $I_j^{(n)}(q_i, r)$.

Summarising, it can be seen that if future inflation is taken into account then the reserve must be

$$R_{ij}\ I_j\,(\hat{q}_i, r) \qquad \text{or} \qquad R_{ij}\ I_j^{(n)}\ (\hat{q}_i, r)$$

depending on how many years' inflation are taken into account.

Appendix

In section 2, the log likelihood is derived in equation (2.9) as

$$\log L(q) = \sum_{j=\alpha}^{T} C_{ij}\,(\log \beta + (j\ \ \alpha) \log q)$$

This expression has to be differentiated with respect to q. First of all, note that

$$\frac{d\beta}{dq} = \frac{1}{(1\ q^{T\alpha+1})} + \frac{(1\ q)\ q^{T-\alpha}\ (T\ \ \alpha + 1)}{(1\ q^{T\alpha+1})^2} \qquad \text{(A.1)}$$

and so

$$\frac{d}{dq}\log L \;=\; \sum_{j=\alpha}^{T} C_{ij}\Bigg[\;\frac{(1\ q)\,(T\ \ \alpha+1)\,q^{T\alpha}}{(1\ q^{T\alpha+1})^{2}}\,.\,\frac{(1\ q^{T\alpha+1})}{(1\ q)}$$

$$\frac{1}{(1\ q^{T\alpha+1})}\,.\,\frac{(1\ q^{T\alpha+1})}{(1\ q)}+\frac{(j\ \ \alpha)}{q}\;\Bigg]$$

$$=\;\sum_{j=\alpha}^{T} C_{ij}\Bigg[\;\frac{(T\ \ \alpha+1)q^{T\alpha}}{(1\ q^{T\alpha+1})}\;\;\frac{1}{(1\ q)}+\frac{(j\ \ \alpha)}{q}\;\Bigg]$$

(A.2)

Put $\overline{X} \;=\; \frac{1}{\sum_{j=\alpha}^{T} C_{ij}}\sum_{j=\alpha}^{T}(j\ \ \alpha)\,C_{ij}$ and note that

$$\frac{q^{T\alpha}}{(1\ q^{T\alpha+1})}=\frac{1}{q}\left(\;1+\frac{1}{(1\ q^{T\alpha+1})}\;\right). \tag{A.3}$$

Substituting into (A.2) it can be seen that

$$\frac{d}{dq}\log L \;=\; \sum_{j=\alpha}^{T} C_{ij}\left\{\;\frac{(T\ \ \alpha+1)}{q}\left(\;\frac{1}{(1\ q^{T\alpha+1})}\;1\;\right)\frac{1}{1\ q}+\frac{\overline{X}}{q}\;\right\}$$

$$=\left(\;\sum_{j=\alpha}^{T} C_{ij}\right)\left(\;\frac{1}{q}\left[\;\overline{X}\;\;\frac{q}{1\ q}\;\;(T\ \ \alpha+1)+\frac{T\ \ \alpha+1}{1\ q^{T\alpha+1}}\;\right]\;\right)$$

$$=\left(\;\sum_{j=\alpha}^{T} C_{ij}\right)\frac{1}{q}\left(\;\overline{X}\;\left[\;\frac{1}{1\ q}+(T\ \ \alpha)\;\;\frac{T\ \ \alpha+1}{1\ q^{T\alpha+1}}\;\right]\;\right)$$

(A.4)

A solution of $\frac{d}{dq}\log L = 0$ is needed, so consider

$$f(q) = \overline{X}\left[\frac{1}{1\ q}+(T\ \ \alpha)\;\;\frac{T\ \ \alpha+1}{1\ q^{T\alpha+1}}\right] \tag{A.5}$$

and note that

$$f(q) = \frac{T\ \alpha + 1}{1\ q^{T\ \alpha + 1}}\ \frac{1}{1\ q} - (T\ \alpha\ \overline{X}) \qquad \text{(A.6)}$$

Hence $f(0_+) = \overline{X} > 0$.

It is easy to show that f is a decreasing function of q for $0 < q < 1$ (just differentiate and show that the derivative is always < 0). So if it can be shown that $f(1_-) < 0$ it will have been proved that there is an unique solution of $f(q) = 0$ in $0 < q < 1$ and that this is the maximum likelihood estimator of q.

To calculate $f(1_-)$, first consider $q = 1 - \varepsilon$. Then

$$f(q) = \frac{T\ \alpha + 1}{1\ (1\ \varepsilon)^{T\alpha+1}}\ \frac{1}{\varepsilon}\ (T\ \alpha\ \overline{X})$$

$$= \frac{T\alpha + 1}{1[1(T\alpha + 1)\varepsilon + ‰(T\alpha + 1)(T\alpha)\varepsilon^2 \tfrac{1}{6}(T\alpha + 1)(T\alpha)(T\alpha 1)\varepsilon^3 + O(\varepsilon^4)]}$$

$$\frac{1}{\varepsilon}\ (T\ \ \alpha\ \ \overline{X})$$

$$= \frac{1}{\varepsilon[(1\ \ ‰(T\ \ \alpha)\ +\ \tfrac{1}{6}(t\ \ \alpha)(T\ \ \alpha\ \ 1)\varepsilon^2\ +\ O(\varepsilon^3)]}\ \frac{1}{\varepsilon}\ \left(T\ \ \alpha\ \ \overline{X}\right)$$

$$= \frac{1}{\varepsilon}\ \left(\frac{T\ \ \alpha}{2}\right)\ \frac{(T\ \ \alpha)(T\ \ \alpha\ \ 1)\varepsilon}{6} + \left(\frac{T\ \ \alpha}{2}\right)^2 \varepsilon\ +\ O(\varepsilon^2)\ \ \frac{1}{\varepsilon}\ \ (T\ \ \alpha\ \ \overline{X})$$

$$= \frac{T\ \ \alpha}{2}\ \left(\frac{T\ \ \alpha}{2}\right)\left(\frac{T\ \ \alpha\ \ 1}{3}\ \frac{T\ \alpha}{2}\right)\varepsilon\ +\ O(\varepsilon^2)\ \ (T\ \ \alpha\ \ \overline{X}) \qquad \text{(A.7)}$$

$$\to \overline{X}\ \frac{T\ \alpha}{2}\ \ \text{as } \varepsilon \to 0$$

So $f(1_-) < 0$ if $\overline{X} < \dfrac{T\ \alpha}{2}$.

A value of q is needed such that f(q) = 0. If $\overline{X} < \frac{T \;\; \alpha}{2}$, and if q = 1 – ε then from the above equation (A.7) it can be seen that a first order approximation for ε is given by

$$\varepsilon = \left(\frac{T \;\; \alpha}{2} \;\; \overline{X}\right) . \frac{12}{(T \;\; \alpha)(T \;\; \alpha + 2)} \qquad \text{(A.8)}$$

For the model in practice (section 4),

$$\overline{X} = \frac{1}{\sum_{j=2}^{T} C_{ij}} \sum_{j=2}^{T} (j \;\; 2)\, C_{ij} \qquad \text{(A.9)}$$

Then if $\overline{X} < \frac{T \;\; \alpha}{2}$, q is a solution of

$$f(q) = \frac{T \;\; 1}{1 \;\; q^{T \; 1}} \;\; \frac{1}{1 \;\; q} \;\; (T \;\; 2 \;\; \overline{X}) \qquad \text{(A.10)}$$

and a first order approximation for ε is given by

$$\varepsilon = \left(\frac{T \;\; 2}{2} \;\; \overline{X}\right) . \frac{12}{T(T \;\; 2)} \qquad \text{(A.11)}$$

A more accurate maximum likelihood estimate of q can be found by a numerical search method around this first order approximation. This is illustrated in the example in section 5.

The approximate maximum likelihood estimate of q can be found from equation (A.8) using

$$\hat{q} = 1 - \varepsilon.$$

<>

[D3]

A CURVE FITTING METHOD AND A REGRESSION METHOD

Contributed by S Benjamin and L M Eagles

Introduction

This method models the run-off triangle row-by-row and then ties the rows together. Each row, or year of account, is modelled by a Weibull distribution function. This model was suggested by D H Craighead, and so the Weibull distribution function has become to be known as the Craighead curve when it is used in this context. It is not a linear model and the three parameters have to be estimated using an iterative search method. Once this has been done, the ultimate loss ratio for each year of account can be estimated.

The second part of the method relates the known loss ratios (paid or incurred) to the predicted ultimate loss ratio. Taking a particular development year, each row of the triangle has a known loss ratio for that development year and a predicted ultimate loss ratio from the first part of the method. There is thus a set of pairs of known and dependent variables: one pair for each year of account. A line of best fit is found, using standard regression methods. From this regression line, another estimate of the ultimate loss ratio for each year of account can be read off. This new estimate has the advantage that it takes into account the information from all years of account, rather than just one particular year of account. The regression line can also be used to produce a confidence interval for the estimated ultimate loss ratio, and to estimate loss ratios for future development years (ie the lower triangle).

<>

[D3.a]

THE CURVE FITTING METHODContributed by S Benjamin and L M Eagles

1. Introduction

1.1 In the London Market details of numbers of claims are generally not available or not relevant. Data is usually available for each "year of account" or year of origin, i.e. for all risks written in a particular accounting year which is usually a calendar year. The items normally available are:

(i) Premiums paid to date
(ii) Claims paid to date
(iii) Claims outstanding, typically the case estimates as notified by the brokers to the companies for outstanding claims.

For each year of account separately, a past history of premium payments, claim payments and claims outstanding will be available. This information, split by year of development, may not be complete (e.g. no information may be available on the claims paid by the end of the first or second year of development). Sometimes quarterly development data is available. Data is normally available subdivided by currency and possibly by line of business.

1.2 The curve fitting method works by estimating the Ultimate Loss Ratio ("ULR") for each year of account, from which the necessary reserve is easily derived. Years of account do not need to be homogeneous.

1.3 The method will provide a reserve for each year of account for which sufficient past development data is available. As will be seen, even where little historical development for any one year is available, the information from adjacent years can be used to help.

1.4 The method lends itself well to interactive graphical illustration and is therefore easy to follow by actuaries and non-actuaries.

2. Method

2.1 Run-off triangles are drawn up for as many years of account as possible showing the development year by year (or quarter by quarter) of premiums and claims.

2.2 An estimate of the ultimate premiums receivable is made for each year of account. If it is necessary to calculate the estimate then development factors are normally applied which are calculated from the data without smoothing. Other methods, such as the Regression Method described below, could be used in appropriate circumstances. Often the underwriters' estimates are used since the underwriters have a better feel for the way in which, in practice, policies are being signed.

2.3 The estimates of ultimate premiums are divided into the relevant claims to give a run-off triangle of loss ratios.

2.4 The loss development patterns analysed may be either

(a) paid loss ratios
(b) incurred loss ratios (where incurred claims are paid claims plus claims outstanding), depending on which is believed to be more useful, or
(c) both.

2.5 For each year of accunt for which there is sufficient development a curve is fitted to the loss ratio development. The chosen curve y(t) is described below; it tends to a finite limit (the estimated ULR) as the development time, t, becomes very large.

3. Choice of curve

3.1 The curves used to fit the loss development pattern for a particular year of account are chosen:

(a) to fit the past history of claims payments as best as possible,
and
(b) to allow for additional future claim development.

3.2 Empirical considerations suggest that if a smooth curve is sought to fit the shape of the loss ratio at development time t, plotted against t, that curve would have a negative exponential shape. The actual formula is:

$$y(t) = A(1 - e^{-(t/b)^c})$$

The parameters of such a curve have specific meanings:

A: As t becomes very large y(t) tends to a value of A, i.e. A is the estimated ULR.

b: At t=b $y(t)=A(1 - e^{-1})$ i.e. b is the time taken to reach a loss ratio of about 63% of the ultimate loss ratio. b is measured in the same units of time as t; for example, if t is in years, then so is b.

c: c defines the steepness of the curve.

This curve has been called a "Craighead Curve". It was originally suggested in a paper by D H Craighead (1979).

Two graphs showing how this curve varies if the b and c parameters are varied are set out at the end of Section 4.

3.3 Other curves can be used; for example a double Craighead curve (if it is thought that the data consists of two parts, one short-tail and one long-tail), of the form:

$$y(t) = A_1\ (1\ e^{(t/b_1)^{c_1}}) + A_2\ (1\ e^{(t/b_2)^{c_2}})$$

can be fitted. This is useful when the data are given quarterly and hence the number of data points is larger than the number of parameters.

3.4 It is not always necessary, or advisable, to fit the curve to each year of account separately. The important more recent years of account cannot be fitted anyway. Within the same class of business it is useful to fit the same b and c (i.e. shape) to all years of account and let A (the ULR level) vary. Looking at all the curves thus generated on the same screen (graph) is helpful. Scaling them all to the same nominal ULR of 100% shows by eye whether acceptable homogeneity exists or whether, typically, one year may be an outlier and deserves to be fitted separately by allowing b or c or both to vary also. Knowledge of a structural break in the type or mix of business can be brought into the judgement.

3.5 It can also be helpful to fit A, b_1, c_1 to the paid loss ratios and A, b_2, c_2 (i.e. a common ULR, A) to the incurred loss ratios, and to view the graphical results on the same screen.

3.6 Certain classes of business have typical values of b and c. These may be imposed on data which is otherwise unhelpful.

4. Method of curve fitting

4.1 Any appropriate method of curve fitting could be used. Typically, however, a method which chose curve parameters in such a way as to minimise the sum of weighted least squared deviations of the curve from the data has been used. For example, if an individual year is being fitted in isolation, the quantity to be minimised is D where

$$D = \Sigma w(t)\ (y(t) - yobs(t))^2$$

where yobs(t) = observed loss ratio at time t
w(t) = weights assigned to the loss ratio at time t

and y(t) is the curve being fitted.

4.2 When using the Craighead curve, it is possible to determine algebraically the value of A needed to minimise the measure of deviation, D, provided b and c are known (or fixed).

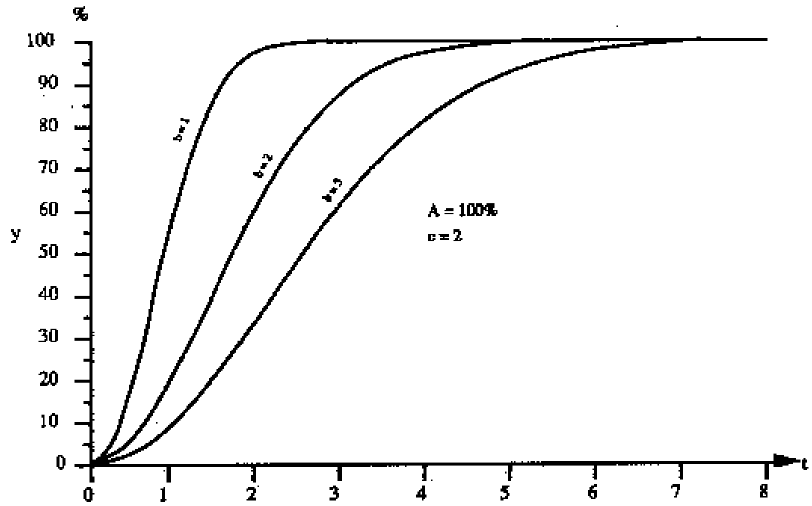
4.3 When b and/or c are not known, or a more complicated curve is being fitted, it is normally necessary to determine the curve parameters using an iterative minimisation technique, such as steepest descent, or the Davidon, Fletcher and Powell algorithm. These iterative techniques normally (but not always) converge quite quickly. The parameters they converge to will define a "local" minimum of D, which may not always be the "global" minimum.

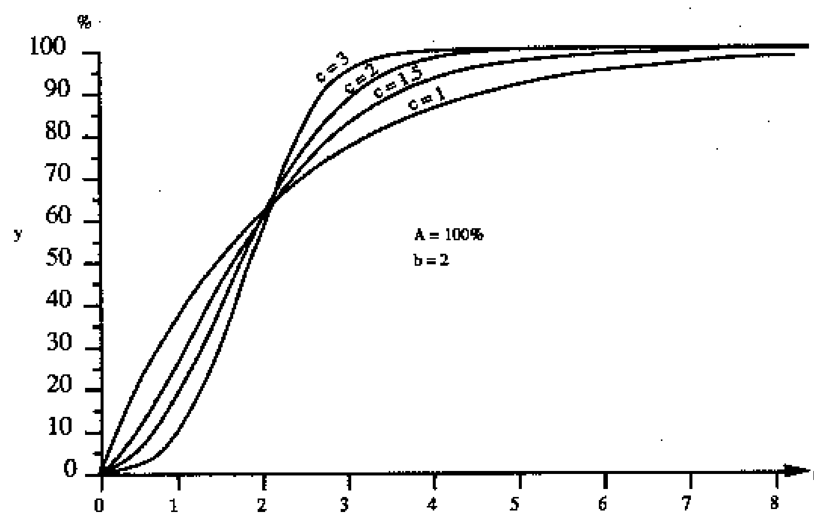
4.4 Any weights could be attached to the individual data points fitted by the above approach (thus outliers can be excluded by giving such points zero weight). Typically, however, the following could be used:

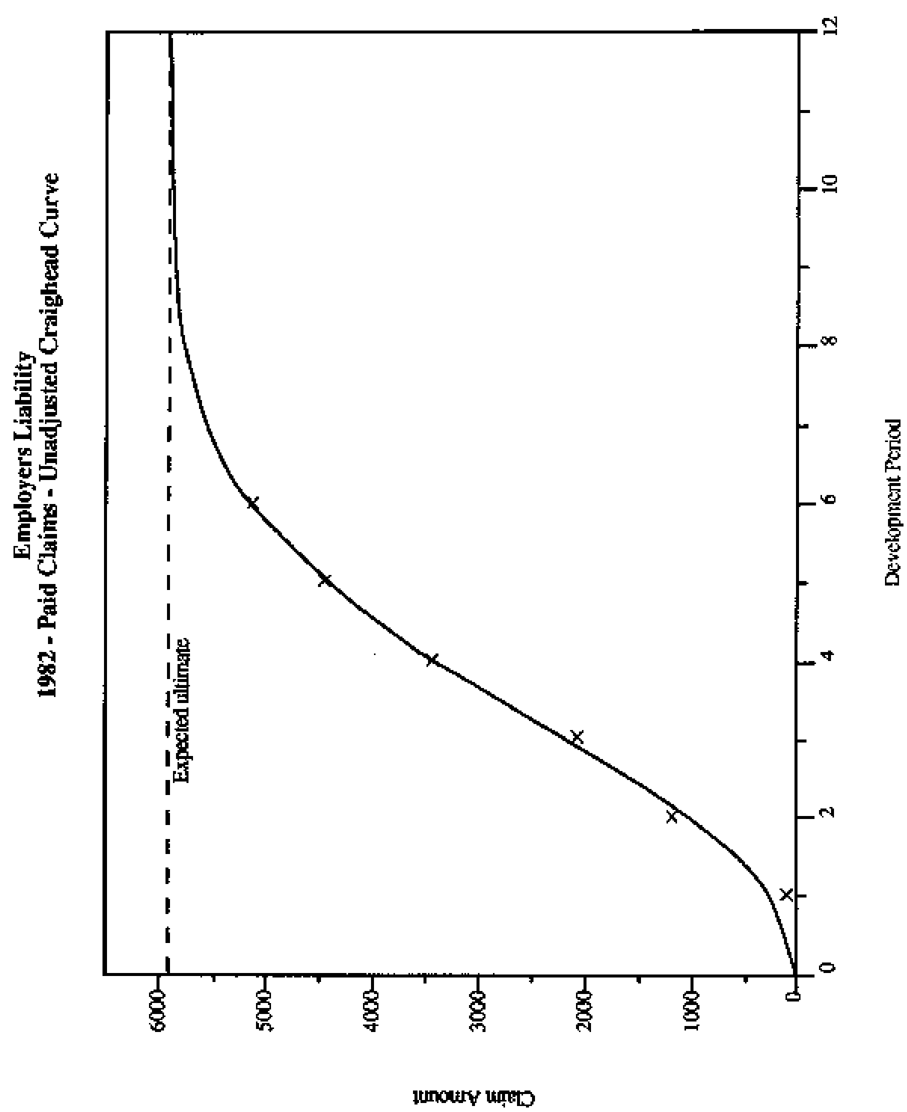
(a) equal weights of 1 to each non-zero data point,
or
(b) $w(t)=t$ thus giving more weight to the more developed data,
or
(c) A weight of 1 for the most recent data point for the given year, 0.9 for the next most recent point, 0.9^2 for the next most recent, etc.

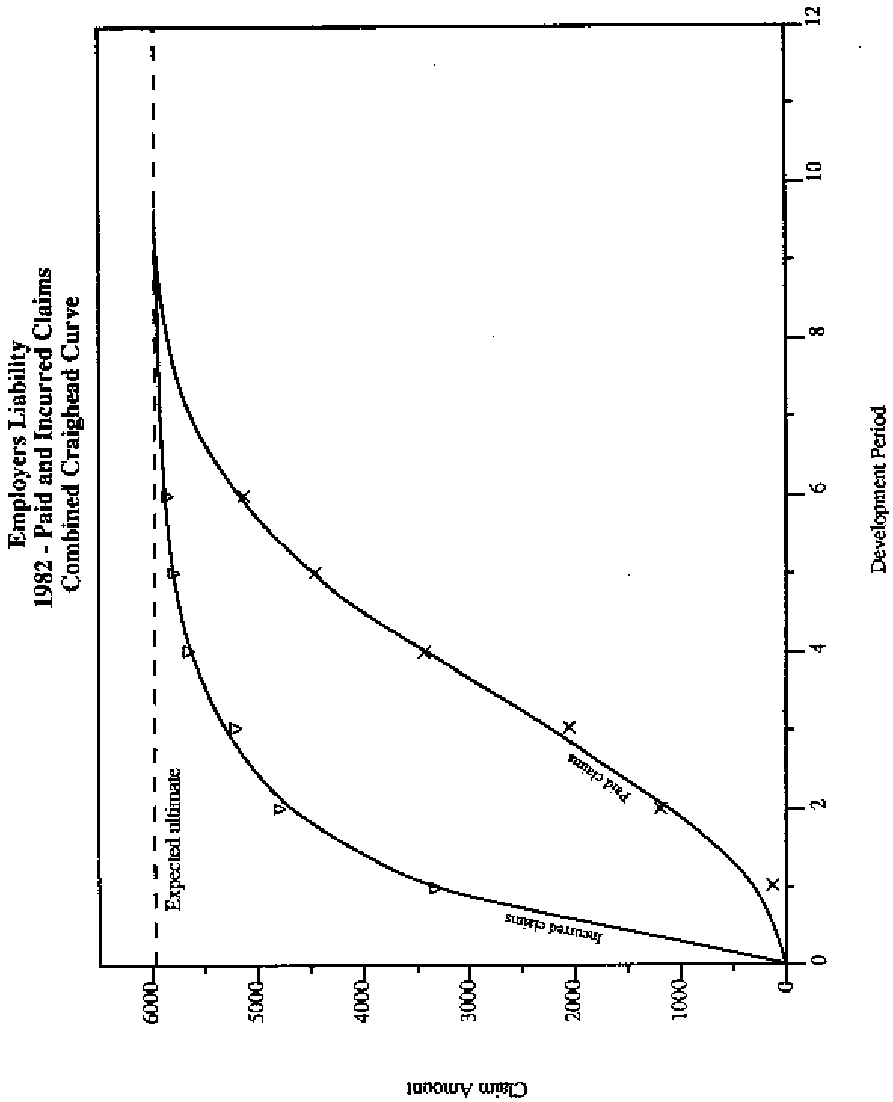
4.5 A special case, which can be considered as an "adjusted Craighead curve", is where the curve is forced through the most recent data point (e.g. by giving this point a very large weight, or to fit A, b, c, and to adjust A, keeping b and c fixed to make the curve pass through latest point). Curves fitted in this manner ensure that further development is non-negative.

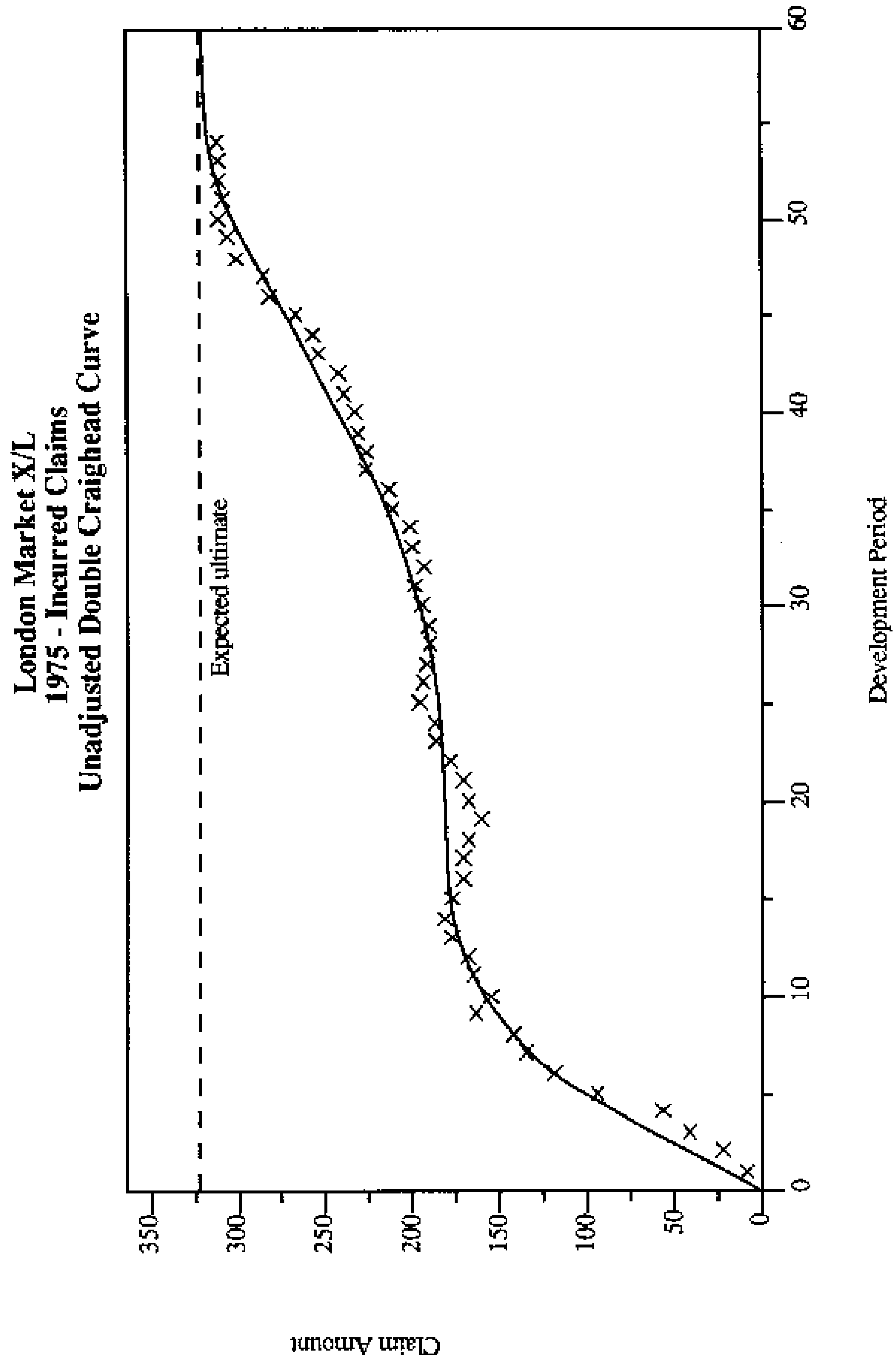
4.6 Typically, the loss ratios and the curves fitted to these loss ratios would be plotted graphically. A qualitative goodness of fit can then be determined, by visual examination. Signs of heterogeneity or structural breaks between different years of account can be seen visually. Curve fits can then be refined, if necessary, by the actuary if he or she believes that different curves would fit the data better.


%
100
90
80
70
60
50
40
30
20
10
0
y
b=1
b=2
b=3
A = 100%
c = 2
0
1
2
3
4
5
6
7
8
t


%
100
90
80
70
60
50
40
30
20
10
0
y
c=3
c=2
c=1.5
c=1
A = 100%
b = 2
0
1
2
3
4
5
6
7
8
t


Employers Liability
1982 - Paid Claims - Unadjusted Craighead Curve
Expected ultimate
Claim Amount
Development Period
0
1000
2000
3000
4000
5000
6000
0
2
4
6
8
10
12


Employers Liability
1982 - Paid and Incurred Claims
Combined Craighead Curve
Expected ultimate
Incurred claims
Paid claims
Claim Amount
Development Period
0
1000
2000
3000
4000
5000
6000
0
2
4
6
8
10
12


London Market X/L
1975 - Incurred Claims
Unadjusted Double Craighead Curve
Expected ultimate
Claim Amount
Development Period
0
50
100
150
200
250
300
350
0
10
20
30
40
50
60

5. Comparison of the curve fitting method with other methods

5.1 The curve fitting method can model claims development (or even premium development) directly, rather than modelling loss ratios. Graphically, however, it is easier to visualise and present loss ratios.

5.2 The link ratio method may technically be reformulated in a manner closely analogous to the curve fitting method. If a single curve is chosen for every year of account so that each successive link ratio matches those derived from a chain ladder or other link ratio method and if, when fitting this curve to the data no weight is given to any data point for a given year of account other than the most recent then the curve fitting approach will produce the same estimated reserves as the link ratio method.

5.3 Three potential weaknesses inherent in the link ratio method are:

(i) over parameterisation of the data (analogous to the curve chosen being fitted with too many parameters).

(ii) undue weight given to the most recent data point in each account year, and

(iii) inability to reflect structural breaks in the underlying business run-off patterns.

It can be seen from 5.2 that the first two weaknesses can be partially overcome using a curve fitting method which chooses curves according to fewer parameters than those implied by a link ratio method, and by giving some weight in the curve fitting to data points other than the most recent. The curve fitting method can also accommodate structural breaks (as mentioned in 3.6).

5.4 The curve fitting method as we have applied it uses premiums as a measure of exposure (by analysing loss ratios). In principle the curve fitting method could instead use other exposure measures, such as number of policies, if they were appropriate to the type of business being analysed.

<>

[D3.b]
THE REGRESSION METHODContributed by S Benjamin and L M Eagles

1. Introduction

1.1 The regression method is used to refine estimated ULR's (Ultimate Loss Ratios), particularly those derived from the curve fitting method.

1.2 The method can provide explicit confidence limits for the ULR thus estimated. These confidence limits are not those in the strict statistical sense but do give a practical range in which the ULR can be expected to lie.

1.3 This method also lends itself well to graphical illustration and is therefore easy to follow by actuaries and non-actuaries.

1.4 Further details of the method, and of its potential application to a minimum reserving basis used by Lloyd's were set out in 1986 in a paper by S Benjamin and L M Eagles.

2. Method

2.1 ULRs for each year of account are estimated using an alternative reserving method. It is not necessary to have estimates for every single year of account, although it is usually helpful to have them for as many years as possible.

2.2 IBNR loss ratios are calculated as the difference between the ULRs and the incurred loss ratios.

2.3 For each development year in turn, a regression line is fitted to the set of points obtained from plotting the IBNR loss ratios against the incurred loss ratios at the development year under consideration; each year of account being represented by a separate point.

2.4 At one extreme, the points may all lie on a straight line so that the relationship may be deduced with little uncertainty. At the other extreme, the points may appear to be scattered at random which suggests that there is little relationship between losses at the development year under consideration and the eventual level of future losses. In fitting this regression line the fitting can be found as a "confidence interval"; this is a region about the line where future results are likely to fall.

2.5 The regression line can be expressed in terms of its "Slope" and "Constant" and leads to an expression of the form:

Estimated IBNR loss ratio = Slope × (incurred Loss Ratio at development year under consideration) + Constant %

2.6 This expression can thus be used to estimate the IBNR loss ratio and hence the ULR for the year of account whose latest development year corresponds to the development year under consideration. The regression lines arising from other years of development can similarly be applied to estimate the ULRs for the other years of account.

2.7 Where very few years of account have advanced to a particular development year, a curve fitted to the underlying loss ratios can be used to provide estimates not only of IBNR loss ratios but also of the expected incurred loss ratios to each development year. These expected loss ratios for any development year can then be plotted against estimated IBNR loss ratios in exactly the same way as actual loss ratios. A line of best fit and a confidence interval can then be derived. For some recent years of account this approach may also be adopted where there is a large amount of fluctuation in the data for the early development years generally.

2.8 It is possible to estimate ULRs (or to estimate IBNR and/or outstandings) by regressing one of these items against either paid, incurred or outstanding loss ratios. It is also possible to use more complicated regression, e.g. log-linear regressions rather than linear regressions. The mechanics of the method remain unchanged.

2.9 The decision as to which regression to make is made after looking at the plots of residuals in each particular case.

3. Comments

3.1 A qualitative degree of confidence in the reserves established by this reserving method can be gained by visual examination of the plots mentioned in 2.3.

3.2 Plots consisting of points scattered apparently at random (such as described in 2.4) do occur, particularly for more recent years of account. In such cases it may be appropriate to take the regression line as the Estimated IBNR loss ratio = Constant (where the constant is chosen as the average IBNR loss ratio, or from other considerations).

In many other cases confidence intervals, again particularly for more recent years of account, can be quite large. It is believed that this usually reflects inherent difficulties facing any loss reserving method, rather than weaknesses of this particular method.

3.3 The regression line stated in 2.5 may be algebraically reformulated in terms of credibility theory, as:

Future Claims = $z \times k \times$ claims to date + $(1 - z) \times 1\times$ premiums
(z is the degree of credibility given to the claims data).

As the development year (and the credibility given to claims data) increases we would therefore expect the constant and the slope of the regression line both to tend to zero.

The actual regression lines adopted may be adjusted if desired to exhibit this behaviour. For years of account when the regression line constant has become zero the method becomes analogous to the link ratio method.

3.4 In practice, for discussion with underwriters, it is more effective to plot separately for each year of development the ULR as ordinate and the paid (or incurred) loss ratio as abscissa, and to fit the line of regression.

3.5 A simple measure of the inherent variability can then be demonstrated visually by:

(a) drawing a line parallel to the regression line and passing through the data point furthest from it, followed by

(b) drawing a further line parallel to it at the same distance on the other side.

Thus, a path symmetrical about the line of regression is formed which just encloses all the data points. The width of the path is an intuitively appealing display and measure of the inherent variability. The width of the path obtained by regressing ULRs against paid loss ratios can be compared with the width obtained by using incurred loss ratios.

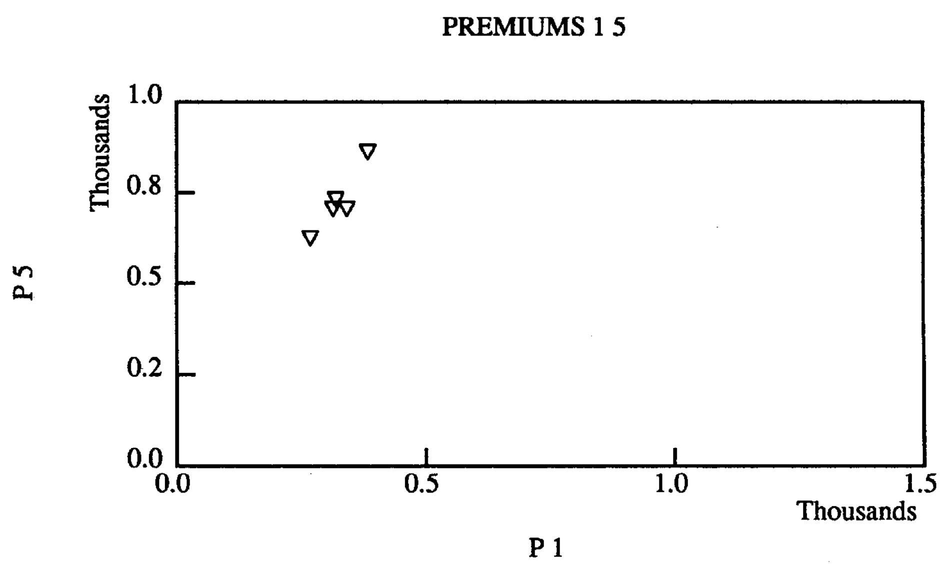
3.6 Outliers can be identified visually and lead to useful discussion.

3.7 At the first level of presentation, if curve fitting is to be avoided, a sufficient number of ULRs must be available. Hence either sufficient history must be available or the underwriter's own estimates of ULRs on partly developed years of account must be used.

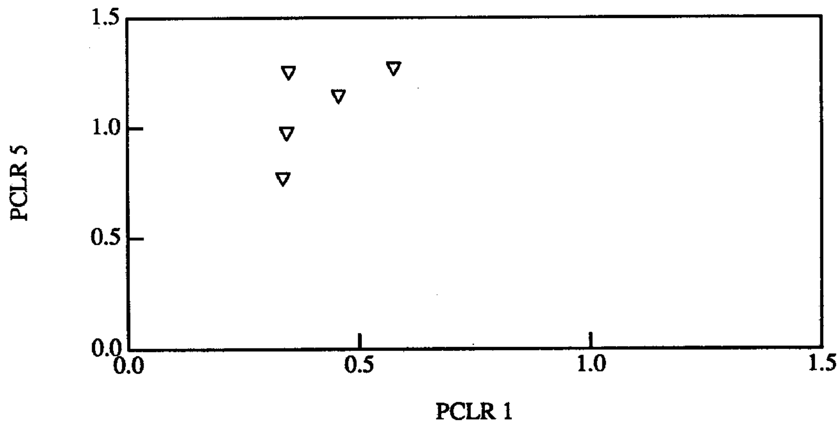
References

(1) 1979: Craighead, D.H. "Some aspects of the London reinsurance market in worldwide short-term business", *Journal of the Institute of Actuaries*, Vol. 106 Part III p 286.

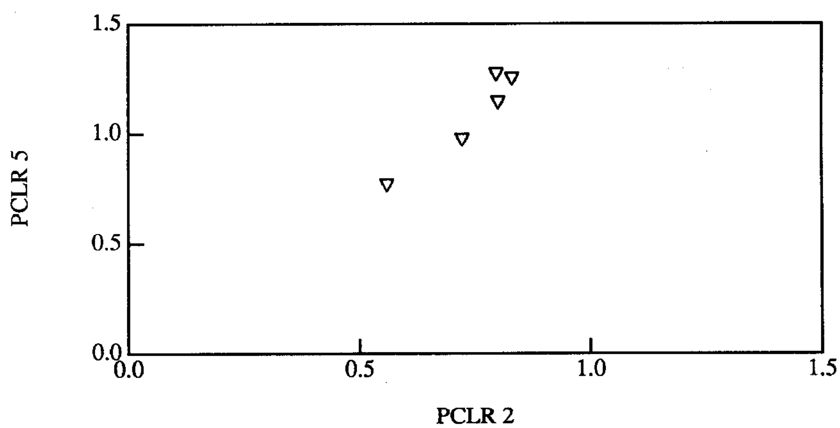
(2) 1986: Benjamin, S. and Eagles, L.M. "Reserves in Lloyd's and the London Market", *Journal of the Institute of Actuaries*, Vol. 113 Part II p 197.


PREMIUMS 1 5
Thousands
1.0
0.8
0.5
0.2
0.0
P 5
0.0
0.5
1.0
1.5
Thousands
P 1

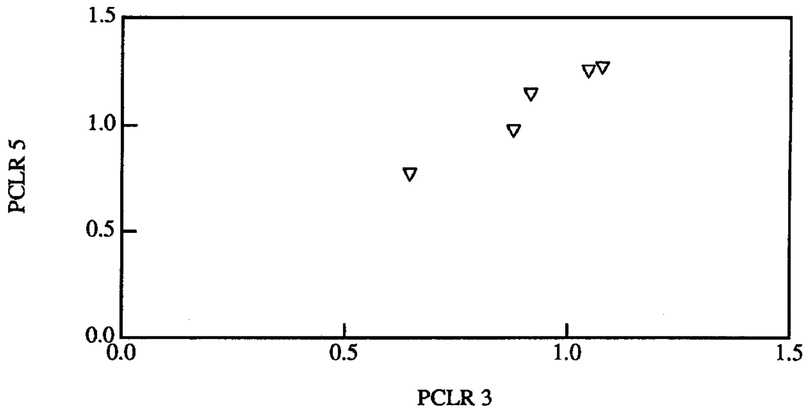
PAID CLAIMS LOSS RATIO 1 5



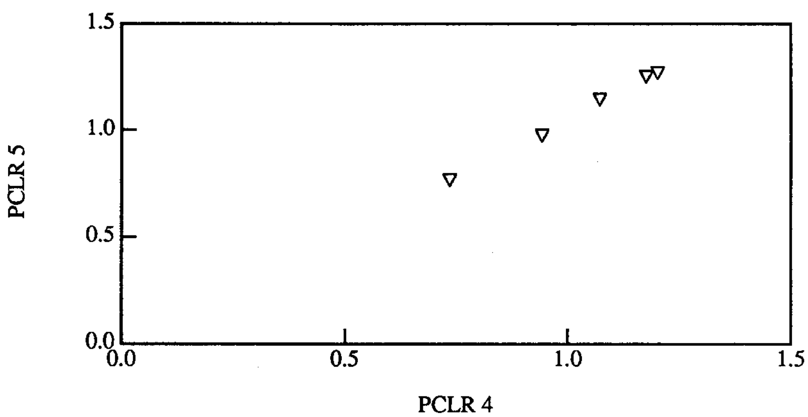
PAID CLAIMS LOSS RATIO 2 5



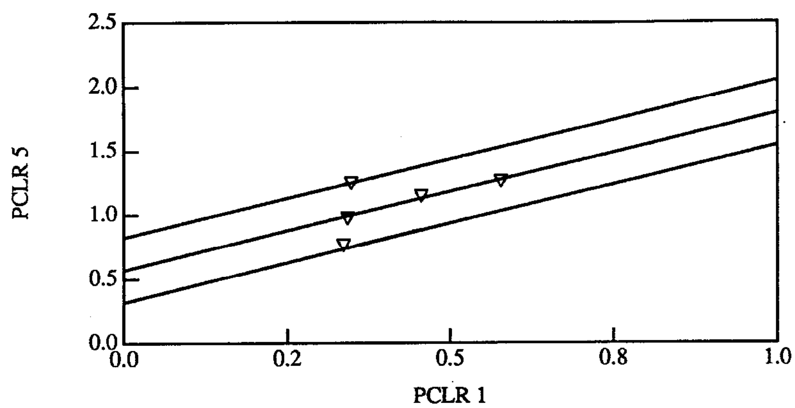
PAID CLAIMS LOSS RATIO 3 5



PAID CLAIMS LOSS RATIO 4 5

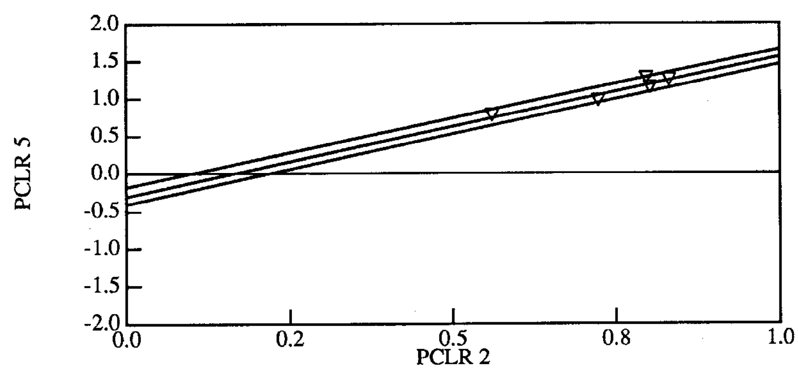


GRAPH PATH 1 5



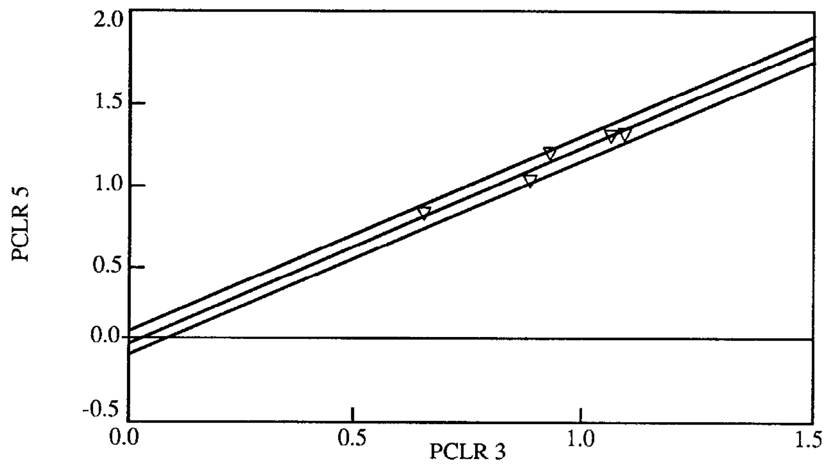
LINE 1 5: PCLR 5 = 1.224 * PCLR1 + 0.575
UPPER LINE 1 5: = LINE 1 5 + 0.250
LOWER LINE 1 5: = LINE 1 5 - 0.250

GRAPH PATH 2 5



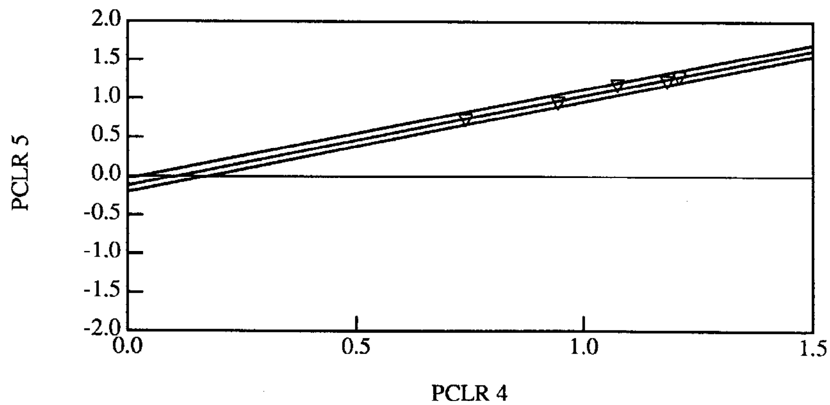
LINE 2 5: PCLR 5 = 1.827 * PCLR2 - 0.277
UPPER LINE 2 5: = LINE 2 5 + 0.087
LOWER LINE 2 5: = LINE 2 5 - 0.087

GRAPH PATH 3 5



LINE 3 5: PCLR 5 = 1.197 * PCLR3 - 0.013
UPPER LINE 3 5: = LINE 3 5 + 0.069
LOWER LINE 3 5: = LINE 3 5 - 0.069

GRAPH PATH 4 5



LINE 4 5: PCLR 5 = 1.096 * PCLR34 - 0.046
UPPER LINE 4 5: = LINE 4 5 + 0.022
LOWER LINE 4 5: = LINE 4 5 - 0.022

[D4]
REID'S METHOD

Contributed by D H Reid

1. Introduction

The method used by D H Reid is essentially non-parametric in nature. A central feature is the use of an empirical estimate of the distribution of claims by size and delay-time, based on one chosen year of origin, the "base year". This estimated distribution is assumed to underlie all the other years of business. The number of claims for each year is assumed to be known or estimated. Instead of estimating a rate of development for the other years of business, each is compared with the base year. For example, if the proportions of the total number of claims from year of origin 2, developed to the end of the first, second, third etc year of their run-off are calculated, then the corresponding points of time for the same proportions in the base year are labelled r_{21}, r_{22}, r_{23}, etc.

The way in which corresponding points of time in the run-offs, and the corresponding sizes of claims are developed and used are described in the paper and shown graphically in Figure 5.

The complications arise mostly from the special treatment required for the end-periods, and for large claims.

Various complicated expressions are evaluated using numerical techniques which are typical of modern computer-based calculations; in principle, they do not affect the method, although they are part of the work of implementation.

1.1 This is a description of a reserving method first proposed by D H Reid (1978) and subsequently developed in a series of papers (Reference 1 to 3). It is a very powerful method of most relevance in direct business where data is available subdivided by claim size.

1.2 The following aspects of the claims reserving situation provide motivation for the particular approach taken:

1.2.1 In most reserving contexts for the claims arising from a particular period a correlation exists between their cost and the period of time elapsing between origin and settlement. Rates of settlement are at least partially within the control of claim officials and are not necessarily constant from year to year. Prima facie, adequate understanding of the developing experience of a claim portfolio upon which reserves can be constructed can thus be gained only by monitoring **both** development time and cost variables jointly. Thus both of these variables and in particular **cost** must be treated in any thorough reserving process.

1.2.2 The standard actuarial approach to graduation by means of a standard table would appear to provide a framework suitable for generalisation in this context. It turns out, however, that, because of the subjectivity involved in what constitutes a claim at the practical level — leading to varying proportions of "nil" claims — the relevant features of the process require careful specification.

1.2.3 The actual cost of claims is determined partly by factors extrinsic to the company, and partly by company policy. It is important that historic cost movements should be visible in a form which enables the effect of these sources to be viewed, before extrapolation to the future, and that this view should as far as possible be free of corruption through point 1.2.1 above.

1.2.4 It goes almost without saying that the general insurance market is highly competitive. Particularly in the context of a small market share there is a considerable premium on estimating claims experience levels as precisely as possible, and to this end maximum benefit from available data is needed. The objective should be to create claim estimation methods which make explicit use of data and use visible valuation bases and are thus directly under management's control. Present day computing power is such as to render computational labour transparent to the user.

1.2.5 Although the method is relatively complex, it is readily capable of being implemented on a PC.

1.2.6 The underlying model represents a deliberate simplification of the claim process (although more elaborate than any other currently in use). The extent to which it is refined in application will depend on the context — in most applications the model as proposed will be adequate. For some, for example where radically different types of payment are embraced within a claim, and where factors affecting each may be different, a version of the model which reflects this may be necessary (see e.g. Section 4 of Reference 2), or a "DP" solution may be adequate which subdivides the experience into two or more parts each of which can be valued separately. Nevertheless, it should be said that experience over 10 years has shown the basic model to be applicable in most cases without further elaboration.

1.2.7 Any valuation method — from case estimates onwards — involves the use (explicitly or implicitly) of a conceptual model, and it is critical to the success of the method that the degree of adequacy of the model should be visible — and where inadequate the model should be capable of modification, or, in the last resort, rejection.

1.2.8 Whilst the method has wide application — and can easily cope with, e.g., the effect of Excess of Loss reinsurance cover on claims experience, it is not readily applicable to "pathological" types of claim, such as certain long term industrial diseases — at least until considerably extended data bases are available.

1.2.9 It is important to mention that in the original paper much of the presentation was concerned with modelling the two way distribution of the base year in considerable detail. The author has now indicated that a relatively simple approach to the distribution based on linear interpolation will suffice for most practical purposes. This results in a much less labour-intensive approach than formerly.

2. The structure of the method is quite involved and the flowchart (fig. 1) may help clarify the interdependence of the various stages. The heart of the method is a bivariate distribution intended to express the relationship between ultimate cost and time of settlement of claims originating in one year.

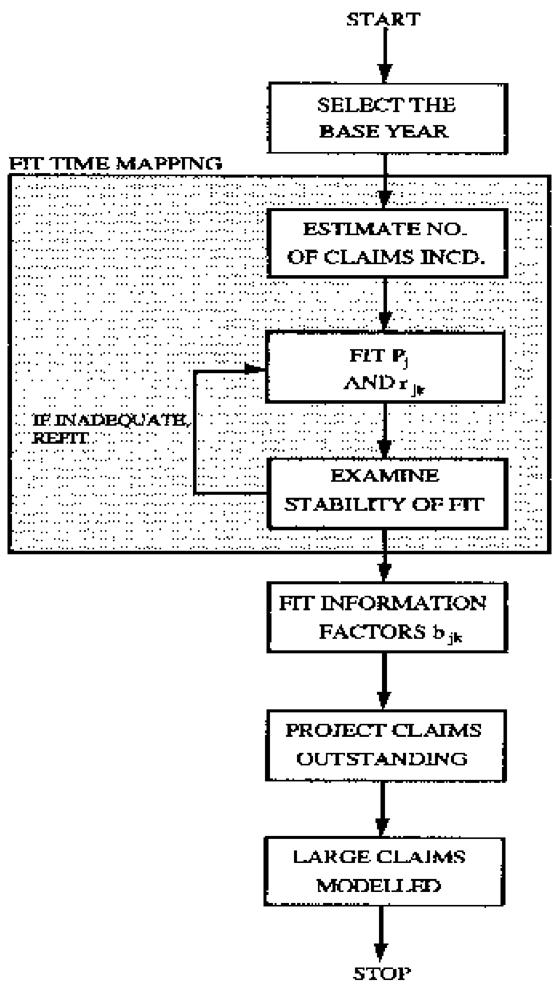


Figure 1

2.1 The variable x represents size of claim (aggregation of individual payments) The variable t represents time of development (time elapsing from the beginning of the origin year to the date of settlement).

Then $M(x,t)$ represents the probability that a claim exceeds amount x and is settled at a time of development greater than t.

2.1.1 Origin date can refer either to date of originating event or date of notification to the company: in the second case estimation for known outstandings is produced; in the first some method of forecasting IBNR numbers is needed — then the method can produce total reserves (known cases and IBNR).

2.1.2 Consistency is required in the definition of a claim and its "time of settlement". The latter could for example relate to "time of first closure" — leaving liability in respect of reopenings to be determined separately. The important thing is consistency — other arrangements are possible.

2.2 The underlying method is as follows for a particular year of origin of claim:

- a proportion p of the claims are zero (i.e. settled at no cost);
- of the zero claims, $M^z(t)$ is the proportion whose time to settlement exceeds t;
- of the non-zero claims $M^{nz}(x,t)$ is the proportion whose cost exceeds x and time to settlement exceeds t.

2.3 The distributions $M^z(t)$, $M^{nz}(x,t)$ are empirically determined from the experience of a well (i.e. nearly completed) developed year of origin (the "base year"). For later years of origin this distribution is assumed to apply, although the model allows for different rates of settlement and for claim inflation by fitting mappings from the actual time and monetary amounts of later years to the operational time and monetary amounts of the base year. It also allows for varying proportions of zero claims.

2.4 The function $M^{nz}(x,t)$ is truncated so that large claims and claims settled at very late durations are treated differently in this analysis.

2.4.1 The definitions of cut-off points for large and late claims should be chosen to fit convenience.

2.4.2 Late claims are included in the last time period analysed (the "end group"). For most direct lines of business the end group should form a very small proportion of all the claims and no significant distortion need be introduced by this approach.

2.4.3 Large claims are modelled separately. The assumption here is that the number of large claims is binomially distributed and that the amount can be modelled by a Pareto curve. The most recent years will not (yet) supply much data for large claims and older years, including years earlier than the base year, can be used. The large claims modelled in this section may also include claims included in the main analysis so as to provide more data; obviously in any final calculations care should be taken to avoid double counting on any claim.

2.4.3.1 The proportion of large claims is settled from the base year data by judgement. This same proportion is then used to determine the actual large claims in each year of origin j.

2.4.3.2 Data from a number of years of origin are deflated back to the base year by a number of assumed large claims inflation factors. When a Pareto curve provides a good fit, this combination of inflation factor and curve is adopted. Projections of future sizes of large claims can then be carried out on a future large claims inflation factor, which may be based upon the fitting above.

2.4.3.3 The chosen proportion of large claims is used on the future probability of a large claim arising. This probability is applied to the total number of claims in each year of origin j to find the expected number of large claims for year j. This number is then combined with the size as found above.

2.5 A year which is well developed is chosen as the base year. Experience has shown that the actual choice of base year does not have a major effect on the final results, as the fitting of mappings between the base and later years will correct for any unusual effects. In any case of doubt, alternative base years can often be selected for comparison.

3. Fitting the time mappings

3.1 *"Operational Time"*

A critical feature of the model (necessary because of 1.2.1 above) is the manner in which allowance is made for varying rates of claim settlement on the claims arising in each origin year. This is achieved by the introduction of an "operational time" mapping in each origin year subsequent to the base year. The idea is that the proportions of claims settled at each point of development of a given origin year equate to those for the base year at the operational time value specific to that point of development.

The operational time scale for each origin year is determined entirely on the number of claims settled, and not the cost of claims.

Thus (refer to figure 2) r_{21} represents the value of operational time for origin year 2 corresponding to the stage of development (by number of claims and not amounts) which that origin year has reached at development time 1 year.

More generally r_{jk} is the operational time at which the base year has reached the same stage of development as origin year j at its kth year of development.

3.2 *Total number of claims by origin year*

If the origin year is defined as year of notification, there is no difficulty in ascertaining the number of claims. If however the origin year is defined as year of event, some means is required at this stage for estimating the eventual number of IBNR claims. Usually a simple procedure based on reporting patterns will suffice.

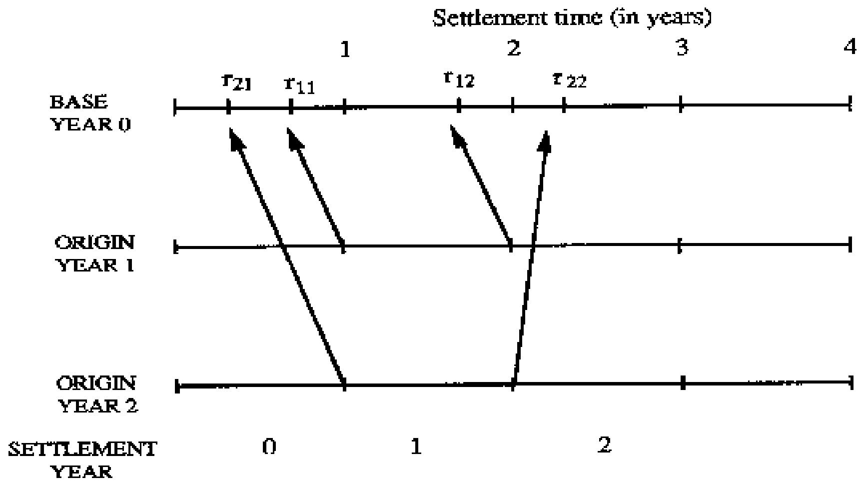


Figure 2

3.3 *Fitting r_{jk} and p_j*

The data will then provide us with:

Q^{nz}_{jk} = the number of origin year j claims settled at positive cost in calendar time (j+k, j+k+1).

Q^{z}_{jk} = the number of origin year j claims settled at zero cost in calendar time (j+k, j+k+1).

Q_j = the number of origin year j claims still unsettled at calendar time s (where s is the final calendar year of development currently available).

Diagrammatically:

non-zero origin year j	Q_{j1}	Q_{j2}		$Q_{j\ldots}$	Q_j
zero	Q_j	Q_j		Q_j	
settlement time t = 0	1	2 . .	s–j–1	s–j	

Figure 3

We would expect these relative ratios to be:

$(1-p_j)$* $[M^{nz}(0,0)-M^{nz}(0,r_{j1})]$	$(1-p_j)$* $[M^{nz}(0,r_{j1})-M^{nz}(0,r_{j2})]$	. . .		$p_j\, M^z(r_{j\,s-j})$
p_j* $[M^z(0)-M^z(r_{j1})]$	p_j* $[M^z(r_{j1})-M^z(r_{j2})]$	. . .		$+ (1-p_j)M^{nz}(r_{j\,s-j})$

Figure 4

We can construct a log-likelihood function for this fit:

$$L = \sum_{j=1}^{s-1} \left\{ \sum_{k=0}^{s-1-j} [N^z_{jk} + N^{nz}_{jk}] + N_j \right\}$$

where

$$N^z_{jk} = Q^z_{jk} \ln\{M^z[r_{jk}] - M^z[r_{j\,k+1}]\} + Q^z_{jk} \ln p_j$$

$$N^{nz}_{jk} = Q^{nz}_{jk} \ln\{M^{nz}[r_{jk}, 0] - M^{nz}[r_{j\,k+1}, 0]\} + Q^{nz}_{jk} \ln (1 - p_j)$$

$$N_j = Q_j \ln\{p_j M^z[r_{j\,s-j}] + (1 - p_j)\, M^{nz}[r_{j\,s-j}, 0]\}$$

Estimates of r_{jk} and p_j are found by maximising the log-likelihood function. This is achieved by standard computational methods.

3.4 *Validation*

At this stage the actual and predicted results can be compared, and if the base year is inappropriate for the data this should become evident.

3.5 *Alternatives*

In the event that a significant diversion should be found at this stage between actual and predicted results, various alternative models are available — for example it may be appropriate in some contexts to sever completely the connection between zero and non-zero claims and proceed accordingly, allowing different r_{jk} for zero and non-zero claims.

4. Fitting the monetary mapping

The development of numbers of claims in each origin year has been compared to the base year, and it is now necessary to compare each origin year to the base year on the basis of the cost of claims. This will produce a set of inflation factors.

4.1 b_{jk} represents the factor by which the cost of claims originating in year j and settled in development year k exceeded that of those base year claims which were settled between r_{jk} and $r_{j\,k+1}$, i.e. the equivalent time in the base year.

4.2 Again it is possible to fit b_{jk} for all the j,k available using a log-likelihood method.

To bring size of claim into the analysis we group by size into bands (x_i, x_{i+1}). If we then let Q_{ijk} represent those claims included in Q^{nz}_{jk} falling between x_i and x_{i+1}, where the x_i are dividing points between the bands, we can express the distribution of Q_{ijk} as:

$$M^*_{ijk} - M^*_{i+1jk}$$

where

$$M^*_{ijk} = M^{nz}\left\{ \frac{x_i}{b_{jk}}, r_{jk} \right\} M^{nz}\left\{ \frac{x_i}{b_{jk}}, r_{j\,k+1} \right\}$$

The log-likelihood function is proportional to:

$$\Sigma Q_{ijk} \ \log\ (M^*_{ijk} - M^*_{i+1\,jk})$$

and the values of b_{jk} which maximise this can be found by numerical techniques.

4.3 Again, at this stage we can examine the b_{jk} to determine whether they accord with intuitive understanding. We can also compare the fitted and actual distributions of claims by size. In the event that what appears to be a poor fit is obtained three possibilities need to be distinguished.

- The sample of claims involved may be so small that appreciable random fluctuation is anticipated.

- The choice of bands of claim size on which the fit has been carried out may be inappropriate relative to the weight of the distribution concerned.

- It may be that a substantive change in the underlying distribution has taken place which does not permit of representation by the underlying surface mapped by the successive transformations involved.

In the first of these cases typically recourse would be had to comparative values from earlier years, or alternatively the period concerned would be grouped with one or more surrounding periods in order to provide a more statistically viable sample.

In the second case, an alternative choice of fitting points can be examined, and it would be the intention to develop an "expert" system to provide this facility automatically.

In the third case a decision must be made as to whether the feature concerned is one which can be expected to persist in future and should therefore be carried forward or may safely be disregarded from the point of view of future projections.

4.4 It is now possible to examine the effect of inflation on past settlements. Before doing so, however, it is necessary to allow for changes in rate of settlement, which may distort the observed values of b_{jk}, (see section 1.2.1). This is done by deriving adjusted inflation factors B_{jk}, which relate to fixed periods of operational time rather than calendar time.

i.e. whereas b_{jk} relates to the period between **settlement** times k and k+1, corresponding to operational times r_{jk} and $r_{j\ k+1}$, B_{jk} relates to the period between **operational** times k and k+1.

Comparisons of B_{jk} for successive origin years should then be free of the distorting effect of changes in rates of settlement and should reflect the "true" effect of inflation.

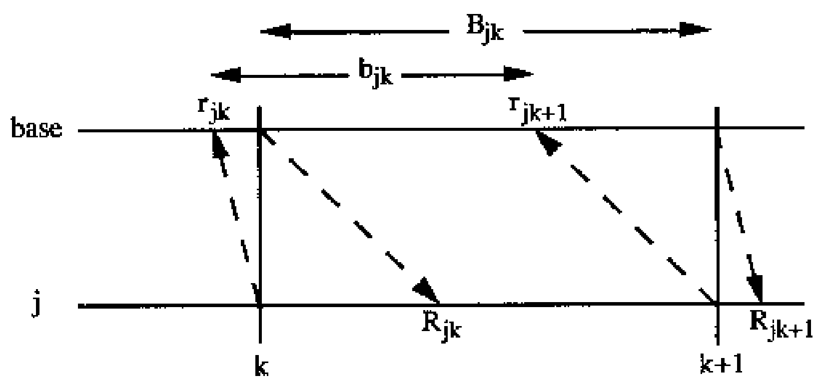
Note — for notational convenience the fixed periods of operational time to which the B_{jk} relate are termed "groups".

4.4.1 The B_{jk} are obtained as weighted averages of those b_{jk} which lie wholly or partly between operational time k and k+1. The weights used are the contributions to the mean non-zero claim cost from the component parts of the base year distribution.

This derivation assumes that each b_{jk} applies uniformly over the period to which it relates. Other assumptions, and other methods of combining the component b_{jk} could be used. It would also be possible to calculate the B_{jk} directly, using b_{jk} derived from them to fit the data.

4.4.2 Since the B_{jk} correspond to periods of operational time it is desirable to be able to relate them to calendar times so that secular changes in claim cost can be properly measured. This is done via R_{jk}, which is defined as the time to settlement in origin year j of the beginning of group k; i.e. B_{jk} relates to the period of settlement R_{jk} to R_{jk+1}. The R_{jk} are obtained from the r_{jk} by linear interpolation, though other means for obtaining them could be used. Figure 5 demonstrates the relationship between the b_{jk} and B_{jk} and between the r_{jk} and R_{jk}.

4.4.3 Because groups and calendar years of development rarely correspond exactly, the situation often arises, for a particular origin year, where past b_{jk} provide information for only part of a group, the remaining part being outstanding. Such outstanding parts of groups are known as fringe



groups. Usually each origin year will have one fringe group (never more than one).

The treatment of these groups has to be considered carefully. If only a small part of the group is outstanding then it may be appropriate to apply the B_{jk} obtained from the settled part to the outstanding part. Conversely, if only a small part is settled then it would be more appropriate to use preceding origin years' B_{jk} for the same group, along with projected inflation, as a guide to the B_{jk} for the outstanding part.

Figure 5

5. Estimating Reserves

5.1 The process of estimating reserves arises as a natural consequence of the underlying conceptual framework — a bivariate claims distribution (by size and time) is modified to reflect the observed experience of individual origin years. The required modifications are produced by the B_{jk}.

In the reserving context, what is required is that part of the modified bivariate distribution which is outstanding at the time of inspection, i.e. that part which lies after $r_{j(s-j+1)}$.

5.2 Estimation of reserves thus boils down to estimation of the outstanding B_{jk}. This may normally be done by assuming a constant rate of future claims inflation (f, say) and applying the formula $B_{j+1\,k}=(1+f)B_{jk}$ for outstanding groups but other adjustments may also be made to allow for, e.g., anticipated changes in rate of settlement or inflation rates which vary by origin year or settlement year.

5.2.1 The expected settlement amount for each outstanding group is then calculated as the product of

(i) B_{jk}
(ii) mean non-zero claim cost for the group
(iii) expected number of non-zero claims in the group.

To obtain (iii), the expected number of non-zero claims in each group, the estimated total number of outstanding non-zero claims (as obtained from the fitting procedure described in 3.3) is spread over the group pro rata to the proportion of all non-zero claims attributable to each group.

It is then a simple matter to accumulate the total settlement amount outstanding.

5.2.2 It should be borne in mind that, as mentioned in 2.1, the claim costs reserved represent aggregations of individual payments, i.e. the ultimate cost of outstanding claims. To obtain the outstanding monetary amounts any payments made on account on these claims should be deducted from the reserve calculated as described above.

5.2.3 Fringe groups, described in 4.4.3, require special consideration. The appropriate B_{jk} should be obtained as discussed in 4.4.3 and the mean non-zero claim cost and expected number of non-zero claims are derived from the distribution for only the outstanding part of the group.

5.2.4 Large claims also require special consideration. When assigning the outstanding non-zero claims to groups, the proportion of large claims may be reduced in line with the number of such claims already settled for each origin year, in order to allow for the time development of these claims. The appropriate B_{jk} may be obtained from an assumed constant inflation rate, though it should be borne in mind that for these claims origin year is the important determining factor for cost level.

6. Miscellaneous

6.1 The advantages of the method cover two main headings:

(i) *Analysis*

By removing the effect changes in the rate of settlement have on observed claim cost the method allows a proper analysis of the underlying movements in claim cost and, unlike other standard methods (e.g. chain ladder, separation method) is free from the distorting influence of such changes. Inspection of the fits, both by size and time, may also indicate whether observed changes can be attributed to secular movement or relate to underlying changes in the nature of the business, in which case appropriate steps may be taken to amend any assumptions for the future.

(ii) *Control*

By suitable adjustment of the parameters affecting the reserves (e.g. rate of inflation, individual B_{jk}, proportion of large claims) senior management can ensure that the reserves reflect their judgements as to general economic conditions and the nature of the business (possibly as indicated by the method's own analysis). The flexibility of the method allows such adjustments to be precise and specific.

6.2 The method can also be extended to experience rating of larger commercial risk, though a number of constraints may be required because of greater variability, e.g. by reference to an extended model using B_{jk} from a larger portfolio. This is of particular significance to those contexts where rating is based on "burning cost". (For a fuller discussion see Reference 3.)

6.3 As pointed out in the original paper, the method gives rise to the possibility of estimating confidence intervals for outstanding claims. Beyond this it becomes a practical possibility to examine the "strength" of reserves in terms of the trade off of variability against mean cost at a given reserve level.

7. Example

Example Accounting Date: 31.12.87

MODEL (Section 2)

The Base Year for the model is 1982 with truncation points of 6 for operational time t and £80,000 for claim size x (see paragraph 2.3–2.5).

The model for Nil claims, $M^z(t)$, is:
(x 10^{-5})

t:	0.00	0.5	1	2	3	4	5	6
$M^z(t)$:	100,000	84,835	37,299	1,288	46	13	3	1

The model for Non-nil claims, $M^{nz}(x,t)$, is:
(x 10^{-5})

t: x	0	0.5	1	2	3	4	5	6
£0	100,000	71,556	28,296	2,278	598	243	112	70
£25	95,126	68,026	26,889	2,147	568	234	106	68
£100	56,228	43,280	19,881	1,852	513	213	99	65
£200	35,885	28,612	14,386	1,552	454	201	97	63
£500	15,727	13,189	7,577	1,136	384	180	93	59
£1,000	7,336	6,529	4,158	859	334	160	84	53
£1,500	4,512	4,115	2,798	728	304	150	80	4
£2,000	3,030	2,792	1,989	615	275	139	76	49
£3,000	1,529	1,425	1,126	462	234	122	68	44
£4,000	918	876	745	376	211	112	68	44
£5,000	595	570	511	310	184	104	65	42
£6,500	414	407	372	255	160	99	59	40
£8,000	285	279	258	190	133	87	55	40
£10,000	203	203	186	148	112	76	49	38
£15,000	104	104	97	85	74	51	34	28
£20,000	68	68	66	63	57	47	34	28
£25,000	55	55	55	51	49	40	28	25
£30,000	40	40	40	38	36	27	21	19
£40,000	28	28	28	28	27	21	15	13
£50,000	17	17	17	17	17	15	11	9
£65,000	17	17	17	17	17	15	11	9
£80,000	13	13	13	13	13	13	9	8
£100,000	6	6	6	6	6	6	4	2

The above matrices are based on 23,679 Nil claims and 52,643 Non-nil claims forming the actual data for the Base Year as seen at 31/12/87. The values of $M^z(t)$ and $M^{nz}(x,t)$ at intermediate values of x and t are derived from linear interpolation of ln(M) against t and/or ln(x). For this purpose, the device is used of re-assigning the lowest value of x to 1, in order to avoid singularities when taking logs.

For large claims (see paragraph 2.4.3) the model consists of a truncated pareto distribution over the range £80,000 to £420,000 with parameter 1.29. The proportion of large claims is 0.00012. This is slightly different from the proportion shown in the bivariate model above because it is derived from the inspection of a number of years' data as described in paragraph 2.4.3, as are also the truncated limit £420,000 and the pareto parameter 1.29.

FITTING THE "TIME" MAPPINGS (Section 3)

The actual number of claims settled and outstanding for later years is:

ORIGIN YEAR		YEAR OF DEVELOPMENT 0	1	2	3	4		OUTSTANDING
1983	Nil	15,204	7,609	289	13	6	}	
	Non-nil	39,052	14,006	964	192	76	}	92
1984	Nil	16,381	9,569	457	35		}	
	Non-nil	43,971	16,561	1,299	280		}	217
1985	Nil	18,132	11,025	786			}	
	Non-nil	51,167	18,919	1,712			}	725
1986	Nil	18,511	12,885				}	
	Non-nil	58,257	21,218				}	4,104
1987	Nil	19,080					}	
	Non-nil	62,522					}	46,531

Fitting the p_j and r_{jk} as described in paragraph 3.3 gives the following parameter values:

ORIGIN YEAR	p_j	r_{j1}	r_{j2}	r_{j3}	r_{j4}	r_{j5}
1983	0.7015	1.014	1.982	2.918	3.720	4.486
1984	0.7019	0.981	1.900	2.743	3.569	
1985	0.7073	0.974	1.826	2.638		
1986	0.7195	0.958	1.780			
1987	0.7361	0.906				

which generate fitted data as follows:

ORIGIN YEAR		YEAR OF DEVELOPMENT 0	1	2	3	4	OUTSTANDING
1983	Nil	14,902	7,917	302	10	3	2
	Non-nil	39,515	13,558	931	193	79	91
1984	Nil	16,282	9,702	449	23		6
	Non-nil	44,051	16,430	1,302	302		223
1985	Nil	18,314	10,980	647			46
	Non-nil	50,959	18,964	1,856			704
1986	Nil	19,358	12,016				871
	Non-nil	57,423	22,027				3,280
1987	Nil	19,091					14,717
	Non-nil	62,548					31,777

FITTING THE MONETARY MAPPING (Section 4)

The actual numbers of claims settled in size bands are shown in detail in the Appendix. The numbers shown represent the numbers of claims settled at a cost greater than the corresponding value of x (claim size). In this way, the number of claims settled greater than £0 can be seen to correspond to the number of non-nil claims settled in fitting the time mapping above. Numbers of claims settled within size bands, Q_{ijk}, can easily be obtained by differencing.

Fitting b_{jk} as described in paragraph 4.2 gives the following results:

ORIGIN YEAR	b_{j0}	b_{j1}	b_{j2}	b_{j3}	b_{j4}
1983	1.138	1.091	1.235	1.103	1.113
1984	1.216	1.190	1.216	1.406	
1985	1.327	1.355	1.458		
1986	1.372	1.471			
1987	1.416				

which generate the modelled numbers of claims by size shown in the Appendix.

As described in paragraph 4.4, the fitted b_{jk} are transformed to B_{jk}, which are free of the distorting effect of changes in settlement rates evidence from the r_{jk}. The resultant B_{jk}, using the method described in paragraph 4.4.1, are:

ORIGIN YEAR	B_{j0}	B_{j1}	B_{j2}	B_{j3}	B_{j4}
1983	1.138	1.093	1.228	1.105	1.113
1984	1.216	1.191	1.251	1.406	
1985	1.328	1.363	1.458		
1986	1.376	1.471			
1987	1.416				

Note that each of the B_{jk} in the bottom diagonal corresponds to a fringe group, as discussed in paragraph 4.4.3. The effect of the special considerations applied to fringe groups will be apparent in the section on estimating reserves.

The R_{jk}, discussed in paragraph 4.4.2, define the periods of calendar time covered by groups in the fitted data. Derived from the r_{jk}, they have the following values.

ORIGIN YEAR	R_{j0}	R_{j1}	R_{j2}	R_{j3}	R_{j4}
1983	0.986	2.019	3.103	4.366	5.000
1984	1.020	2.119	3.311	4.000	
1985	1.031	2.214	3.000		
1986	1.051	2.000			
1987	1.000				

ESTIMATING RESERVES (Section 5)

In this example the outstanding B_{jk} have been estimated by assuming a constant rate of 7.5% p.a. for future claims inflation but the following two cases of special treatment should be noted:

(a) *Fringe Groups (see paragraph 4.4.3)*

For 1985–87 the outstanding part of the fringe group has been assigned the same value of B_{jk} as was fitted to the settled part. For 1983–84, however, the settled part has been deemed to be too small to give a reliable indication of the B_{jk} for the whole group and the B_{jk} for the outstanding part has been projected at the assumed rate of inflation from the B_{jk} for the previous year.

(b) *End and Large Claims Groups (see paragraph 2.4)*

The B_{jk} for these groups are projected at the assumed rate of inflation from the Base Year, for which the B_{jk} are automatically equal to 1. A similar consideration also applies to Group 5. The resultant B_{jk} are:

ORIGIN YEAR	GROUP 0	1	2	3	4	5	END	LARGE CLAIMS
1983					1.075	1.075	1.075	1.075
1984				1.188	1.156	1.156	1.156	1.156
1985			1.458	1.277	1.242	1.242	1.242	1.242
1986		1.471	1.567	1.373	1.335	1.335	1.335	1.336
1987	1.416	1.581	1.685	1.476	1.436	1.436	1.436	1.436

The corresponding mean claim costs, derived from the model after application of the estimated outstanding B_{jk} given above, are:

ORIGIN YEAR	GROUP 0	1	2	3	4	5	END	LARGE CLAIMS
1983					6,229	7,257	15,768	165,539
1984				5,396	6,160	7,802	16,951	177,954
1985			2,653	5,107	6,622	8,387	18,222	191,300
1986		1,174	2,287	5,490	7,119	9,016	19,589	205,648
1987	448	796	2,459	5,902	7,653	9,692	21,058	221,071

Note the effect of the positive correlation between time to settlement and claim size on the outstanding fringe groups' mean claim costs, which are greater than those for the corresponding complete groups. The number of outstanding non-zero claims for each year, as estimated in "Fitting the Time Mapping" above, is assigned to individual outstanding groups pro rata to the corresponding probabilities from the model. This gives the following expected numbers of non-zero claims in each group:

ORIGIN YEAR	GROUP 0	1	2	3	4	5	END	LARGE CLAIMS
1983					28	21	34	7
1984				68	75	23	37	9
1985			260	248	89	28	44	10
1986		1,379	1,373	290	104	33	51	12
1987	5,091	24,562	1,585	335	120	38	59	14

Note that for 1983 the probability of being assigned to the Large Claims group has been reduced slightly because of the known settlement of one large claim. The resultant outstanding settlement amounts (product of numbers and averages) are, in £000:

ORIGIN YEAR	GROUP 0	1	2	3	4	5	END	LARGE CLAIMS
1983					177	155	530	1,171
1984				365	460	183	624	1,567
1985			691	1,269	590	234	799	2,007
1986		1,619	3,140	1,594	741	294	1,004	2,522
1987	2,281	19,560	3,898	1,979	920	365	1,246	3,130
Total	2,281	21,179	7,729	5,208	2,887	1,231	5,154	12,451

Aggregating these amounts for each origin year and subtracting payments made on account (see paragraph 5.2.2) gives the following table of claims reserves, in £000:

ORIGIN PERIOD	OUTSTANDING SETTLEMENT COSTS	AMOUNT PAID ON ACCOUNT	RESERVE
1983	2,033	789	1,244
1984	3,199	1,075	2,124
1985	5,590	1,525	4,065
1986	10,915	3,836	7,079
1987	33,380	10,435	22,945
Total	58,120	19,377	38,743

References

REID, D.H.:-

(1) 1978: Claim Reserves in General Insurance (with discussion), *Journal of the Institute of Actuaries*, Vol. 105 Part III pp 211–296, discussion p 297.
(2) 1980: Reserves for outstanding claims in non-life insurance. *Transactions of the International Congress of Actuaries*, Zurich and Lausanne, 2, pp 229–241.
(3) 1981: *Cahiers du CERO*, Vol. 23, 3–4.
(4) 1986: *Insurance: Mathematics and Economics* 5 pp 45–56.

<>

PAPERS OF MORE ADVANCED METHODS

MBERS OF CLAIMS BY SIZE (ACTUAL DATA)

	YEAR OF ORIGIN											
	1983	1984	1985	1986	1987							
CLAIM SIZE	YEAR OF DEVELOPMENT					YEAR OF DEVELOPMENT				YEAR OF DEVELOPM		
	0	1	2	3	4	0	1	2	3	0	1	
£0	39,052	14,006	964	192	76	43,971	16,561	1,299	280	51,167	18,919	1,
£25	37,285	13,407	920	181	69	42,140	15,859	1,236	270	49,496	18,284	1
£100	20,991	10,066	792	168	62	24,616	12,086	1,071	245	30,599	14,356	1
£200	12,506	7,201	658	141	54	14,593	8,841	909	217	18,295	10,732	1
£500	5,097	3,735	442	113	46	6,114	4,625	600	170	7,696	5,730	
£1,000	2,138	1,948	326	96	42	28	211	326	119	64	3,039	
£1,500	1,235	1,248	261	83	38	1,456	1,546	334	120	1,916	1,966	
£2,000	802	840	212	74	36	920	1,286	273	105	1,264	1,373	
£3,000	340	414	138	62	32	417	541	187	83	613	760	
£4,000	177	236	96	42	28	211	326	119	64	316	388	
£5,000	88	140	67	34	24	123	204	84	48	206	231	
£6,500	41	79	42	26	20	68	97	56	41	122	118	
£8,000	29	43	31	22	16	37	60	42	31	61	69	
£10,000	14	19	22	18	13	21	34	22	17	33	35	
£15,000	5	6	11	9	9	4	9	7	8	9	12	
£20,000	2	1	7	3	5	0	5	6	3	3	3	
£25,000	1	0	5	3	3	0	3	4	2	2	1	
£30,000	1	0	5	2	3	0	2	3	0	1	1	
£40,000	0	0	2	0	2	0	1	2	0	0	1	
£50,000	0	0	1	0	2	0	1	1	0	0	0	
£65,000	0	0	0	0	2	0	0	1	0	0	0	
£80,000	0	0	0	0	1	0	0	1	0	0	0	
£100,000	0	0	0	0	1	0	0	0	0	0	0	
£150,000	0	0	0	0	0	0	0	0	0	0	0	
£200,000	0	0	0	0	0	0	0	0	0	0	0	

APPENDIX — ACTUAL & MODELLED NUMBERS OF CLAIMS BY SIZE (MODELLED DATA)

	YEAR OF ORIGIN									
	1983	1984	1985	1986	1987					
CLAIM SIZE	YEAR OF DEVELOPMENT					YEAR OF DEVELOPMENT				YEAR OF DEVEL
	0	1	2	3	4	0	1	2	3	0	1
£0	39,052	14,006	964	192	76	43,971	16,561	1,299	280	51,167	18,919
£25	37,239	13,336	910	181	73	41,972	15,793	1,228	265	48,906	18,080
£100	21,106	9,913	786	162	66	24,369	11,857	1,045	240	29,505	13,918
£200	13,014	7,228	666	138	59	15,260	8,786	871	212	18,962	10,642
£500	5,158	3,723	465	109	49	6,142	4,572	590	164	7,832	5,699
£1,000	2,103	1,948	329	91	42	2,540	2,408	405	133	3,321	3,067
£1,500	1,159	1,244	264	80	39	1,408	1,549	317	116	1,863	2009
£2,000	723	846	222	71	35	888	1,071	260	104	1,196	1,427
£3,000	306	425	155	58	29	391	550	175	84	555	766
£4,000	144	241	115	50	25	189	315	126	70	279	448
£5,000	72	140	88	42	22	96	192	95	61	146	283
£6,500	34	77	64	32	20	43	99	66	48	63	144
£8,000	20	47	51	24	17	25	63	51	39	36	91
£10,000	12	26	30	18	13	15	35	30	29	21	53
£15,000	5	8	12	11	9	7	12	11	16	9	18
£20,000	2	3	5	6	6	3	5	5	10	5	8
£25,000	1	2	3	4	5	1	2	3	6	2	3
£30,000	0	1	1	4	4	0	2	1	4	1	2
£40,000	0	0	1	3	2	0	1	1	4	0	1
£50,000	0	0	1	1	2	0	0	1	3	0	0
£65,000	0	0	0	1	1	0	0	0	1	0	0
£80,000	0	0	0	0	1	0	0	0	1	0	0
£100,000	0	0	0	0	1	0	0	0	0	0	0
£150,000	0	0	0	0	0	0	0	0	0	0	0
£200.000	0	0	0	0	0	0	0	0	0	0	0

[D5]

REGRESSION MODELS BASED ON LOG-INCREMENTAL PAYMENTS

Contributed by S Christofides

The first article in Volume 2 of this Manual by B Zehnwirth has shown the close connection between the intuitive Chain Ladder technique and the more formal two way analysis of variance model based on the log-incremental payments.

Models initiated by this more formal definition of the basic chain ladder have recently started to gain acceptance in loss reserving work and a number of papers on the subject have now been published. These models differ from the traditional techniques by a more formal definition of both the model assumptions and the parameter estimation and testing. With the formal models statistical estimates of reserves, that is both mean estimates and the associated model standard errors, can be calculated. The basic chain ladder is deterministic and produces point estimates of reserves.

The purpose of this paper is to serve as a basic introduction to these methods for the practitioner. To facilitate this a PC spreadsheet package is used to show how run-off models of the log-incremental payments can be identified and fitted in practice using multiple regression.

The approach adopted considers the basic chain ladder technique first and shows how the intuitive chain ladder model can be made more formal. The parameters of this model are then estimated and the implied underlying payment pattern compared with the chain ladder derived pattern. Both models are used to fill in the square and the results compared. In the case of the formal model it is also shown how the regression results are used to derive estimates of the individual future payments and their standard errors and how accident year and overall standard errors can be calculated.

The simple example makes it easier to follow the calculations and is intended to allow the reader to focus on the more interesting modelling aspects of the later sections.

A more realistic example is then analysed. The data is first viewed graphically to identify an appropriate run-off model to fit. The identified model is fitted and tested. The model is then redefined with fewer parameters and refitted. The results, both future payments and their standard errors, from these models are calculated and compared.

The data is then adjusted for inflation and for claim volume and a series of models are identified and tested. Three of these are used to obtain estimates which are then compared.

A degree of theory is assumed. The model parameters are estimated using multiple regression and matrix operations are used to calculate the variance-covariance matrices. All the computations and graphs are done in a PC spreadsheet package, Supercalc 5 in this case although Lotus 123 could have been used equally effectively.

The wide availability, ease of use and power of these packages makes these methods accessible to all. Alternatively any programming language with matrix manipulation capabilities, such as APL or SAS, could be used for this work. Programs have also been written in GLIM (see A Renshaw, 2).

A. Introduction

Almost all actuarial methods for estimating claims reserves have an underlying statistical model. Obtaining estimates of the parameters is not always carried out in a formal statistical framework and this can lead to estimates which are not statistically optimal. These traditional methods generally produce only point estimates.

The models, such as the basic chain ladder, are often overparameterised and adhere too closely to the actual observed data. This process can lead to a high degree of instability in values predicted from the model as the close adherence to the observed values results in parameter estimates which are very sensitive to small changes in the observed values. A small change in an observed value, particularly in the south-west or north-east regions of the data triangle, can result in a large change in the predicted values. In practice attempts may be made to achieve some stability in the results by using benchmark patterns, by selection of development factors and a number of other such techniques.

Formal statistical models are used extensively in data analysis elsewhere to obtain a better understanding of the data, for smoothing and for prediction. Explicit assumptions are made and the parameters estimated via rigorous mathematics. Various tests can then be applied to test the goodness of fit of the model and, once a satisfactory fit has been obtained, the results can be used for prediction purposes.

This process allows us to focus on the model being fitted and should also highlight any inadequacies in the model. The estimates of the parameters, on the basis of the model, can be made statistically optimal. Peculiarities in the data may be identified and often investigation of these can yield useful additional information to the modeller.

All modelling, whether based on the traditional actuarial techniques such as the chain ladder or on more formal statistical models, requires a fair amount of skill and experience on the part of the modeller. All these models are attempting to describe the very complex claims process in relatively simple terms and often with very little data. The advantage of the more formal approach is that the appropriateness of the model can be tested and its shortcomings, if any, identified before any results are obtained.

B. The basic chain ladder technique and the underlying stochastic model

The following simple example considers a 4 by 4 triangle of cumulative payments:

CUMULATIVE PAID CLAIMS

DEVELOPMENT YEAR

ACC YR	0	1	2	3
0	11073	17500	19339	20105
1	14799	24156	26500	
2	15636	26159		
3	16913			

The usual (weighted) basic chain ladder development factors are (see Vol 1 Section E8):

0 to 1	1 to 2	2 to 3
1.633781	1.100418	1.039609

where 1.633781 = (17500 + 24156 + 26159)/(11073 + 14799 + 15636) etc.

Using these factors the square can be completed in the usual way:

CUMULATIVE PAID CLAIMS

DEVELOPMENT YEAR

ACC YR	0	1	2	3
0	11073	17500	19339	20105
1	14799	24156	26500	27550
2	15636	26159	28786	29926
3	16913	27632	30407	31611

The actual and fitted portions of the square have been separated for illustration. It is assumed in this example that there are no payments beyond the 3rd development period so that the first (zero th) accident year is complete.

The chain ladder produces successive cumulative losses from which the future incremental payments can be derived by subtraction. It is therefore possible to split

the overall chain ladder derived reserve estimate for an accident year into its incremental or payment year values.
The underlying model is better illustrated by these incremental payments which are shown in the table below.

INCREMENTAL PAID CLAIMS

DEVELOPMENT YEAR

ACC YR	0	1	2	3	0/S
0	11073	6427	1839	766	
1	14799	9357	2344	1050	1050
2	15636	10523	2627	1140	3767
3	16913	10719	2775	1204	14698
				Total	19515

The accident year projected future payments and the overall estimate are shown in the last column. The chain ladder estimate of future payments to development period 3 for all accident years is 19515.

Dividing each of these incremental amounts by the final, or ultimate, accident year value gives the following:

PERCENTAGE PAID CLAIMS

DEVELOPMENT YEAR

ACC YR	Ultimate	0	1	2	3
0	20105	55.08	31.97	9.15	3.81
1	27550	53.72	33.96	8.51	3.81
2	29926	52.25	35.16	8.78	3.81
3	31611	53.50	33.91	8.78	3.81

The basic chain ladder has produced the following underlying incremental payment pattern:

Development year	0	1	2	3
Incremental paid %	53.50	33.91	8.78	3.81

Note that this underlying pattern can be calculated directly from the development factors.

The basic chain ladder assumptions can be restated as follows:

a: Each accident year has its own unique level.

b: Development factors for each period are independent of accident year or, equivalently, there is a constant payment pattern.

These assumptions can now be used to define the model more formally.

Let:

A_i be the ultimate (cumulative) payments for the i-th accident year.

B_j be the percentage of ultimate claims paid during the j-th development period.

P_{ij} be the incremental paid claims for accident year i paid during development period j

The chain ladder model can thus be described by the following equations

$P_{ij} = A_i \times B_j$ for i,j from 0 to 3

and the condition

$\Sigma B_j = 1$ where j is summed from 0 to 3

The next section considers how these equations may be solved and estimates of the parameters obtained.

C. Estimating the parameters of the formal chain ladder model

As the main set of relations involves products the usual approach is first to make these linear by taking logarithms and then use multiple regression to obtain estimates of the parameters in log-space. It will eventually be necessary to reverse this transformation to get back to the original data space.

Dealing with the main set of equations is relatively easy. Taking logarithms (natural logarithms will be assumed throughout and denoted by ln) gives

$\ln (P_{ij}) = \ln A_i + \ln B_j$

Unfortunately taking logarithms of the second condition does not produce a linear equation as

$\ln(\Sigma B_j) \neq \Sigma (\ln B_j)$

It is possible to obtain estimates of these parameters using iterative procedures but this is not pursued here. It is more convenient to drop the condition and concentrate initially on obtaining the parameter estimates from the remaining, now linear, set of equations.
Dropping the condition gives rise to a singularity and so it is necessary to introduce a new condition in order to obtain the parameter estimates. This does not affect the eventual results but it does change the interpretation of the parameters.

For ease of reference the parameters are now redefined ($\ln A_i = a_i$ etc) and an error term introduced.

$$\ln(P_{ij}) = Y_{ij} = a_i + b_j + e_{ij}$$

where e_{ij} is some error term.

As indicated above without some restriction these equations are singular. Note for example that a_3 appears only in one equation which involes b_0 and an error term. An infinite number of combinations of a_3 and b_0 are possible as long as they sum to the same view.

For convenience in this example b_0 is set to zero. Another approach is to set both a_0 and b_0 equal to zero and introduce a constant, k, into the model. The chain ladder assumes each accident year has a unique level so the model to be fitted below will follow the former description. The alternative definition is considered later in Section H and the advantages of this choice outlined.

The predictions obtained by either approach will be the same so the restriction can be chosen at the convenience of the modeller.

The model to be fitted is described by:

$$\ln(P_{ij}) = Y_{ij} = a_i + b_j + e_{ij}$$

where i and j go from 0 to 3 and $b_0 = 0$

The model has seven parameters to be estimated, the same number as the basic chain ladder model.

The following table is in the form most convenient for the regression facility of any of the popular spreadsheet packages.

		Y-variate		← Design Matrix **X** →						
i	j	P_{ij}	Y_{ij}	a_0	a_1	a_2	a_3	b_1	b_2	b_3
0	0	11073	9.31226	1	0	0	0	0	0	0
0	1	6427	8.76826	1	0	0	0	1	0	0
0	2	1839	7.51698	1	0	0	0	0	1	0
0	3	766	6.64118	1	0	0	0	0	0	1
1	0	14799	9.60231	0	1	0	0	0	0	0
1	1	9357	9.14388	0	1	0	0	1	0	0
1	2	2344	7.75961	0	1	0	0	0	1	0
2	0	15636	9.65733	0	0	1	0	0	0	0
2	1	10523	9.26132	0	0	1	0	1	0	0
3	0	16913	9.73584	0	0	0	1	0	0	0

Each row corresponds to a data value and its representation by the model parameters. The last but one row, for example, describes the accident year 2, development year 1, value in log-space as the sum of the a_2 and b_1 parameters. The coefficients of the other parameters are zero for this data value.

The resulting matrix of parameter coefficients, made up of ones and zeros in this case, will be referred to as the model design matrix **X**. It is governed by the model chosen.

Within the class of log-linear models changing the model just involves changing the design matrix.

The regression takes the $\ln(P_{ij})$ or Y_{ij} values as the dependent variable and each of the columns of the matrix **X** as the independent variables.

The spreadsheet regression command, which requires a columm for the dependent values and a range for the independent values (i.e. the design matrix) is then used to carry out the regression and output the result. It is necessary to specify that the fit is without a constant and to define a results or output range. This is quite straightforward in practice and the results are produced almost instantly.

The spreadsheet output in this case will be:

Regression Output:

Constant	0
Std Err of Y Est	.0524
R Squared(Adj,Raw)	.9976 .9992
No. of Observations	10
Degrees of Freedom	3

Coefficient(s)	9.288	9.591	9.692	9.736	−.4661	−1.801	−2.647
Std Err of Coef.	.0400	.0400	.0428	.0524	.04277	.05015	.06591

A brief description of this fairly standard spreadsheet regression output will be found in Appendix 2.

The results can also be obtained by matrix manipulation. An indication of how this can be done is given in Section D.

The coefficients are the parameter estimates and are in the same order as the columns of the design matrix.

So the model estimate for a_0 is 9.288, for a_1 it is 9.591 and so on until b_3 which is estimated as –2.647.

The payment pattern can be derived from this output. This is done by exponentiating the development year parameters b_j's, remembering to bring in the b_0 which was set to 0, and scaling so that the exponentiated values add up to the required 100%.

A formal proof of this is beyond the scope of this paper and the interested reader is referred to Verrall's paper (5) "Chain Ladder and Maximum Likelihood". The table below, and the comparison with the basic chain ladder result, may be sufficient to satisfy the majority of practitioners.

The following table shows these basic calculations

Parameter	b_0	b_1	b_2	b_3	
Coefficient	0	–.4662	–1.8015	–2.6472	sum
exp (b_j)	1	0.6274	0.1651	0.0709	1.8634
Payment %	53.67	33.67	8.86	3.80	100

This is very close to the basic chain ladder derived pattern.

BCL Payment %	53.50	33.91	8.78	3.81

The slight differences arise from the way the parameter estimates are derived. The same underlying model is assumed in both cases. Unfortunately however a fair amount of further manipulation is necessary to obtain estimates of ultimate values for each accident year. These cannot be derived simply from the accident year regression coefficients.

In order to progress further it is now necessary to go back and consider what assumptions were made by the spreadsheet in deriving the parameter estimates. This requires a more detailed consideration of the formal model and in particular the structure of the assumed error term.
These aspects are considered in the following section.

D. Fitting assumptions and error terms

The spreadsheet regression is fitted by least squares. That is by minimizing the sum of the squares of the error terms e_{ij}.

It is usual and convenient to assume that the error values e_{ij} are identically and independently distributed with a normal distribution whose mean is zero and variance some fixed σ^2.

i.e. e_{ij} = IID $N(0, \sigma^2)$

In matrix form it can be shown that, under these assumptions, the parameter estimates are given by

$$(\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}^T\mathbf{Y}$$

where $\mathbf{X}$ is the design matrix and $\mathbf{X}^T$ its transpose and $\mathbf{Y}$ is the data vector. The standard errors can also be calculated in matrix form.

These assumptions can be tested by analysis of the residual (error) terms, by plots and other diagnostic tests. Residual plots are shown and discussed later.

Recalling that the original payments were transformed by taking logarithms the error normality assumption in log-space implies that the data in the original space are log-normally distributed.

The IID assumption estimates are also the maximum likelihood estimates in this case and it can be shown that the parameter estimates so obtained are unbiased. Since maximum likelihood estimates are invariant under transformation Verrall (5) shows in “Chain Ladder and Maximum Likelihood” how maximum likelihood estimates of development factors can be obtained by direct substitution.

As the log-normal distribution is skewed with a tail to the right some extreme high values are to be expected. This is sometimes a feature of incremental claims payment triangles. The cause is usually a large claim payment in later development periods, the settlement perhaps of a particularly large claim, when the overall level of payments is low.

These assumptions are not claimed to be theoretically justified for log-incremental claims payments. They have an intuitive appeal and are chosen primarily for convenience. Alternative assumptions, which may well be more generally applicable to claim payments, can be made and results obtained. These tend to require more complex computations or iterative procedures which generally necessitate the use of specially written software.

Further comments on the error terms are to be found in the final section of this paper which also includes some suggestions for dealing with negative incremental payments.

E. Predicting future payments and their standard errors

In order to derive estimates of the model parameters it was convenient to take logarithms and work in log-space. To obtain results in the original space it is necessary to reverse this transformation.

Obtaining the parameter estimates in log-space is relatively straightforward. To revert back to the original space is not so simple and it is necessary to use the relationships between the parameters of the log-normal distribution and the underlying normal distribution.

Again for simplicity the easiest approach is adopted here. This approach is also used by Zehnwirth and by Renshaw and again the justification can be found in their papers. These estimates, in the original space, are not necessarily unbiased especially where a small number of data points are being fitted. Verrall (6) shows how it is possible to obtain unbiased estimates but the calculations are more complicated.

The estimates to be used here are given by the following

The future values $\hat{P}_{ij}$'s are calculated from the estimates obtained in the log-space $\hat{Y}_{ij}$ as follows

a) $\hat{P}_{ij} = \exp(\hat{Y}_{ij} + 0.5\ \mathrm{var}(\hat{Y}_{ij}))$

Their standard errors are given by

b) $\mathrm{s.e.}(\hat{P}_{ij}) = \hat{P}_{ij}\ \mathrm{sqrt}(\exp(\mathrm{var}(\hat{Y}_{ij}))^{-1})$

So the first step is to derive the predicted values and their standard errors in log-space.

The predicted values in log-space are obtained from the estimates of the parameters produced by the regression.

For example the first future value to be predicted is for accident year 1 development year 3 and this is given by

$$\begin{aligned}\hat{Y}_{13} &= a_1 + b_3 \\ &= 9.591 - 2.647 \\ &= 6.944\end{aligned}$$

To obtain the variance of this, and the other estimates, it is necessary to calculate the variance-covariance matrix.
This matrix is given by

$$\sigma^2\ \mathbf{X}_f\ (\mathbf{X}^T\mathbf{X})^{-1}\ \mathbf{X}_f^T$$

where σ^2 is the model variance (scalar) and depends on the data

$\mathbf{X}_f$ is the design matrix of the future values and

$\mathbf{X}_f^T$ is its transpose and

$(\mathbf{X}^T\mathbf{X})^{-1}$ is the model information matrix

with $\mathbf{X}$ the design matrix and $\mathbf{X}^T$ its transpose.

In a spreadsheet a small macro can be written to carry out this calculation. The results of each stage of this calculation for the simple example above are to be found in Appendix 1.

Note that changing data values in the original triangle only affects the scalar factor σ^2 and so the lengthy matrix calculation only need be done once for a given size model.

The usual practice therefore is to calculate the matrix product

$$\mathbf{X}_f\,(\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}_f^T$$

and multiply by the specific data σ^2 as necessary.

A library of these matrices could be built up for the models to be used, to cater for different sizes of triangles for instance, and stored for future use.

The design matrix of future values $\mathbf{X}_f$, following the same format as the original design matrix, is as follows:

←Future Design Matrix **X** →

i	j	a_0	a_1	a_2	a_3	b_1	b_2	b_3
1	3	0	1	0	0	0	0	1
2	2	0	0	1	0	0	1	0
2	3	0	0	1	0	0	0	1
3	1	0	0	0	1	1	0	0
3	2	0	0	0	1	0	1	0
3	3	0	0	0	1	0	0	1

The matrix

$$\mathbf{X}_f\,(\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}_f^T$$

in this case is (see Appendix 1)

1.66667	.00000	1.33333	.00000	.00000	1.33333
.00000	1.25000	.75000	.00000	.75000	.25000
1.33333	.75000	1.91667	.00000	.25000	1.41667
.00000	.00000	.00000	1.66667	1.33333	1.33333
.00000	.75000	.25000	1.33333	1.91667	1.41667
1.33333	.25000	1.41667	1.33333	1.41667	2.58333

The variance-covariance matrix of future values is calculated from the above by just multiplying through by the model σ^2 which in this case is

$$.0524^2 = .002744$$

The variance-covariance matrix is then

.00457	.00000	.00366	.00000	.00000	.00366
.00000	.00343	.00206	.00000	.00206	.00069
.00366	.00206	.00526	.00000	.00069	.00389
.00000	.00000	.00000	.00457	.00366	.00366
.00000	.00206	.00069	.00366	.00526	.00389
.00366	.00069	.00389	.00366	.00389	.00709

Note that these matrices are square and symmetric with each side equal to the number of future values to be projected. The diagonal elements contain the variances of each of these values and are in the same order as the future design matrix elements.

To obtain the variances to be used for projecting future values we will follow common practice and add the model variance (σ^2) to the variances calculated above. These two sources of error are the estimation and statistical errors. These variances recognise that the parameter coefficients are estimates (and subject to error) as well as the inherent noise in the process or data. We do not attempt to correct or estimate any specification or selection errors which may well be equally significant contributors to a total overall error term. Our final example gives some indications of how projected values can be affected by the choice of model parameters. For a more detailed explanation of these types of error the reader is referred to the paper by Taylor (3).

The variances for the future values in log-space are the sum of the variance-covariance matrix values obtained above and the model variance σ^2.

So the variance for the first projected value which was estimated above, Y_{13}, is

$$1.66667 \times 0.05238^2 + 0.05238^2 = .007317$$

The following table shows the various values and their variances and standard errors

i	j	$\hat{Y}_{ij}$	Var($\hat{Y}_{ij}$)	$\hat{P}_{ij}$	var($\hat{P}_{ij}$)	se($\hat{P}_{ij}$)
1	3	6.94395	.007317	1041	7953	89
2	2	7.89094	.006174	2681	44520	211
2	3	7.04521	.008003	1152	10662	103
3	1	9.26969	.007317	10650	833010	913
3	2	7.93438	.008003	2803	63122	251
3	3	7.08865	.009832	1204	14328	120

We note here that the sum of the variances is 973595 which is a value that will be used later.

F. Accident year and overall standard errors

Calculating the variances or standard errors across accident years and in total requires one further step involving the covariances. The information needed is in the last matrix above together with the values calculated for $\hat{P}_{ij}$ ’s and their variances.

The variance of the sum of two values A and B is given by

$$\text{Var}(A+B) = \text{Var}(A) + \text{Var}(B) + 2\text{Cov}(A,B)$$

and this extends to sums of more than two values by including all pairs of covariances. Note that Cov(A,B) = Cov(B,A).

A justification is given in Renshaw’s paper that in the case of log-linear models the covariances can be calculated in the original space by the following convenient formula

$$\text{Cov}(\hat{P}_{ij}\ ,\ \hat{P}_{kl}) = E(\hat{Y}_{ij})\ E(\hat{Y}_{kl})\ (\exp(\text{Cov}(\hat{Y}_{ij}\ ,\ \hat{Y}_{kl}\ ) - 1)$$

In practice this can be set up fairly easily in the spreadsheet once the individual values have been estimated and the variance-covariance matrix computed. It does nevertheless involve a fair amount of computation. To illustrate the calculation consider the standard error for the second accident year.

Two values are involved $\hat{P}_{22}$ and $\hat{P}_{23}$, which were estimated as 2681 and 1152.

Their standard errors obtained above were 211 and 103 respectively. The covariance, in log-space, for these estimates can be found in the variance-covariance matrix and is 0.00206. So the covariance in the original space is

$$
\begin{aligned}
\mathrm{Cov}(\hat{p}_{22},\ \hat{p}_{23}) &= 2681 \times 1152\ (\exp(.00206) - 1) \\
&= 6363
\end{aligned}
$$

The required variance of the sum is then given by

$$\mathrm{Var}(\hat{p}_{22} + \hat{p}_{23}) = 211^2 + 103^2 + 2 \times 6363 = 67868$$

So the estimated standard error of the total assumed outstanding claims for this year is 261 or just under 7% of the estimated value of 3833 (2681 + 1152).

This process can be applied to obtain the standard errors for any combination of values, for instance for each accident year or each payment year and more interestingly for the overall total reserve estimate.

The total reserve estimate is the sum of all the projected values and so its variance calculation will include all possible combinations of covariances (of pairs) of values involved in the calculation. This, surprisingly, makes the spreadsheet calculation easier as there is no need to exclude or select any values. One simply sums a range.

The calculations are as in the previous example and can be tabulated fairly easily to produce the following matrix of covariances.

(i,j)	(1,3)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
(1,3)	—	0	4394	0	0	4593
(2,2)	0	—	6363	0	15481	2216
(2,3)	4394	6363	—	0	2216	5403
(3,1)	0	0	0	—	109410	47006
(3,2)	0	15481	2216	109410	—	13145
(3,3)	4593	2216	5403	47006	13145	—
					Total =	420452

Note that the diagonal elements are left blank as the values here should be the variances which were estimated previously. The matrix is symmetric, as is to be expected, and so summing the range produces the sum of covariances of all possible pairs of values. This sum of all pairs of covariances is 420452.

The sum of the variances of the projected values obtained earlier was 973595 and so the overall variance, which is the sum of these two values, is 1394047.

The overall standard error, which is the square root of this value, is therefore estimated as 1181 or just 6% of the overall reserve estimate of 19531. The overall error is relatively small in this simple example. In practice, with real data involving more accident and development years, the percentage errors tend to be higher. The table below summarizes the results.

Project values and their standard errors:

		Development Period				
Acc Yr		0	1	2	3	Tot Acc Yr
1	Amount				1041	1041
	s error				89	89
2	Amount			268	1152	3833
	s error			1	103	261
				211		
3	Amount		10650	280	1204	14657
	s error		913	3	120	1118
				251		
				Overall Total		19531
				Standard Error		1181

The chain ladder overall estimate was 19515. The individual values obtained by the two methods are also close but the chain ladder estimates are point estimates whereas the regression based estimates are statistical estimates with both a mean and a standard error estimate.

All the usual information that can be produced from the traditional chain ladder can be derived from the regression chain ladder including estimates of development factors. The stochastic approach as shown above can produce additional information, based on the model assumptions, such as standard errors of parameters and reserve estimates, that the traditional approach does not. The statistical estimates obtained by the regression approach also facilitate stability comparisons across companies and classes.

This completes our consideration of the regression chain ladder. The technique does not require that we have a complete triangle of data and can work with almost any shape data as long as there are sufficient points from which to obtain estimates of the parameters.

In the next section a log-linear regression model is fitted which is motivated by the run-off shape of the data. This model has fewer parameters as the development parameters are subject to some curve fitting. This is used to project values outside the original triangle shape, that is a tail is projected. The computation approach is identical to the above. The only differences are that there are now more data points to be fitted and the design, and future design matrices are different.

G. Identifying and fitting a regression model

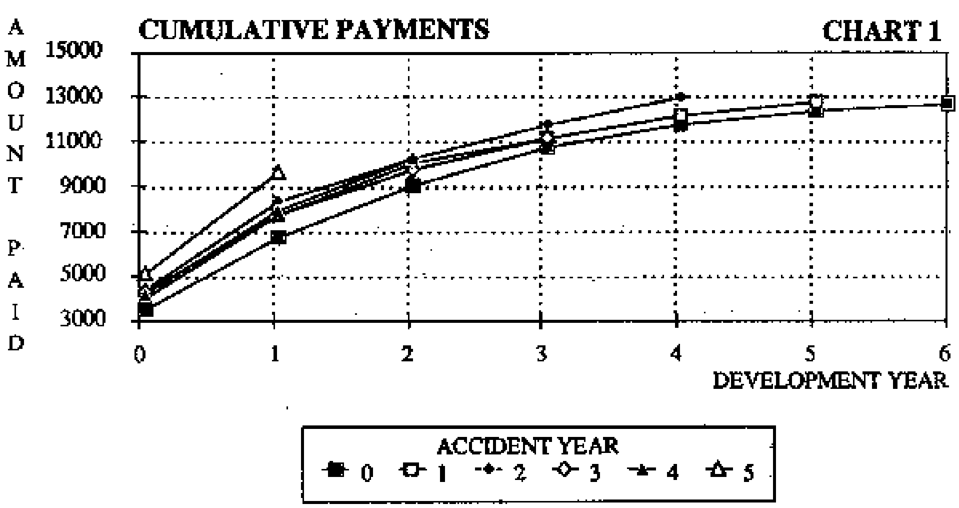
1. Preliminary analysis: Identifying the model.

We will now consider a new data set and attempt to identify and fit an appropriate log-linear model to this data.

The first stage is a visual examination of the data. As a spreadsheet is being used it is very easy to plot the values and look at the resulting line charts rather than attempt to visualize these by looking at the data triangles.

The cumulative claims payments, which are from a UK Motor Non-Comprehensive account, are as follows:

	Development Year						
Acc Yr	0	1	2	3	4	5	6
0	3511	6726	8992	10704	11763	12350	12690
1	4001	7703	9981	11161	12117	12746	
2	4355	8287	10233	11755	12993		
3	4295	7750	9773	11093			
4	4150	7897	10217				
5	5102	9650					
6	6283						



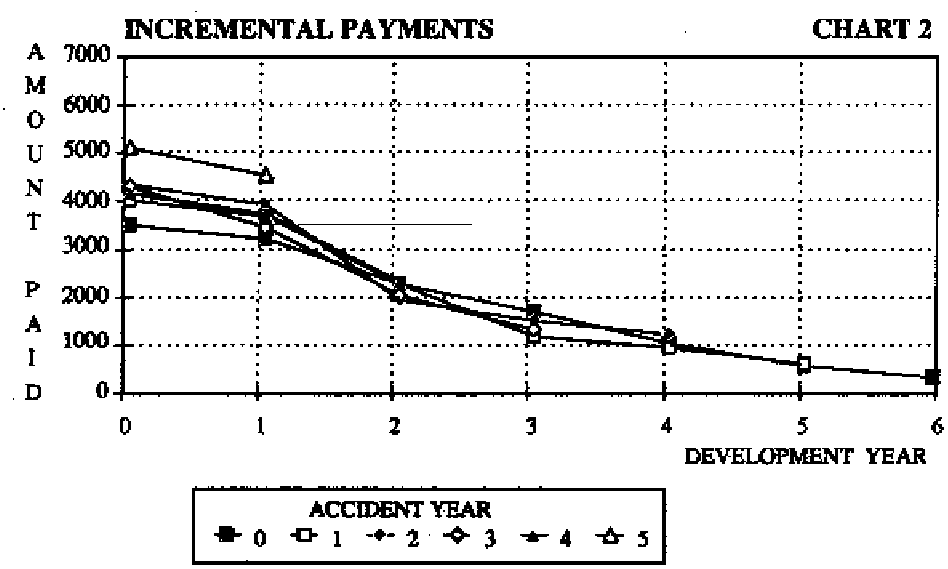
The graph below shows these figures as line charts.
This is a useful presentation but it is hard to identify from this alone an appropriate model to use. Part of the problem arises from the fact that cumulative payments are clearly not independent. The incremental payments are expected to eventually decline but it is not easy to see any pattern or trend from this cumulative plot alone.

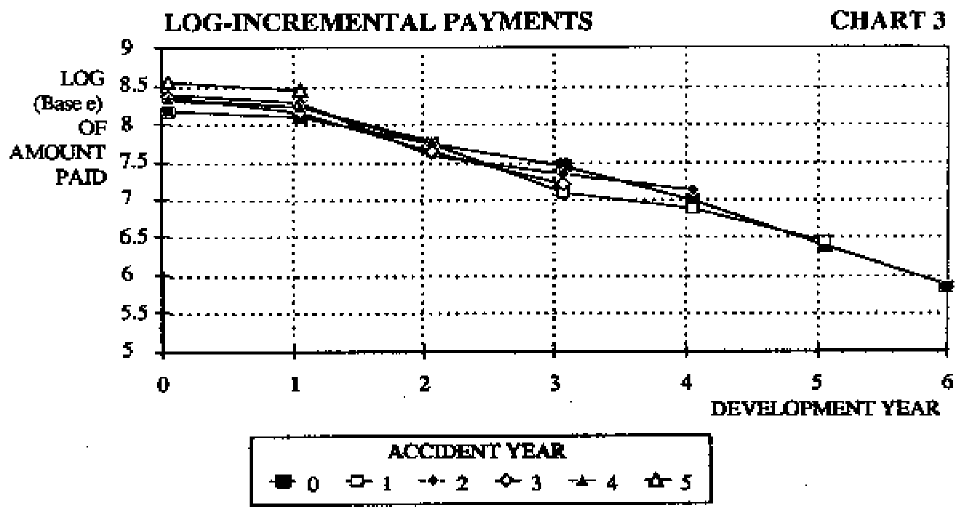
For these reasons the incremental data are now considered.

The incremental payments are

	Development Year						
Acc Yr	0	1	2	3	4	5	6
0	3511	3215	2266	1712	1059	587	340
1	4001	3702	2278	1180	956	629	
2	4355	3932	1946	1522	1238		
3	4295	3455	2023	1320			
4	4150	3747	2320				
5	5102	4548					
6	6283						

Even before these values are plotted a more promising trend can be detected across the development direction. Plotting these values we have:





Finally taking logarithms (base e) of these values and plotting as before produces the following line chart:

Looking at the accident year lines the first four or five look fairly bunched together and the last two (the last one is only a single point) appear to be at a higher level. From development year one the lines look reasonably straight and to have the same slope. These observations indicate that incremental payments from development year 1 on are decaying exponentially, as their logarithms appear to lie approximately on a straight line.

The first model to be fitted is based on these observations and will assume that each accident year has its own parameter or level. Development year zero will be assumed to have its own parameter and in line with the observation above the development parameters from d_1 on will be assumed to be linearly related or to lie along a straight line with some slope to be determined.

This is a start to the modelling process for this data set. The model is not expected to be the final or best for the data but is being used to illustrate various aspects of the modelling process. Note in particular that the plotted log-incremental data has been used to identify an appropriate model to start the process.

The techniques here can be applied in exactly the same way to more complex situations. As an example a different decay rate can be assumed for each accident year if the plot indicates that there is support for such a hypothesis. The model will then be very similar to the one described by Ajne in the second article of this volume. The only difference, apart from the decay rates, is that he fits the first two development periods before curve fitting whereas the example here curve fits from development one as this appears to be supported by the data.

The use of spreadsheets with their comprehensive graphics capabilities enables the modeller to carry out the initial stages of the data analysis phase very quickly as the above charts illustrate. Graphical presentation can also enhance reserving reports to management who may be less actuarially inclined than the writers of such reports.

H. Defining the model

The first model as identified above will now be defined more formally. There is a unique level for each accident year and a unique value for the zero development period. The parameters for development periods 1 to 6 are assumed to follow some linear relationship (straight line) with the same slope or parameter s.

Using the terminology developed earlier we have

$$Y_{ij} = a_i + d_j + e_{ij} \qquad \text{for i, j from 0 to 6}$$

where $d_0 = d$, $d_j = s \times j$ for $j > 0$

and e_{ij} is the error term assumed iid normal with zero mean.

Following the previous example, the spreadsheet table and design matrix are as shown below.

Table 1: Regression Table for the Full Parameter Model

				←		design matrix		→				
i	j	P_{ij}	Y_{ij}	a_0	a_1	a_2	a_3	a_4	a_5	a_6	d	s
0	0	351	8.164	1	0	0	0	0	0	0	1	0
0	1	1	8.076	1	0	0	0	0	0	0	0	1
0	2	321	7.726	1	0	0	0	0	0	0	0	2
0	3	5	7.445	1	0	0	0	0	0	0	0	3
0	4	226	6.965	1	0	0	0	0	0	0	0	4
0	5	6	6.375	1	0	0	0	0	0	0	0	5
0	6	171	5.829	1	0	0	0	0	0	0	0	6
1	0	2	8.294	0	1	0	0	0	0	0	1	0
1	1	105	8.217	0	1	0	0	0	0	0	0	1
1	2	9	7.731	0	1	0	0	0	0	0	0	2
1	3	587	7.073	0	1	0	0	0	0	0	0	3
1	4	340	6.863	0	1	0	0	0	0	0	0	4
1	5	400	6.444	0	1	0	0	0	0	0	0	5
2	0	1	8.379	0	0	1	0	0	0	0	1	0
2	1	370	8.227	0	0	1	0	0	0	0	0	1
2	2	2	7.574	0	0	1	0	0	0	0	0	2
2	3	227	7.328	0	0	1	0	0	0	0	0	3
2	4	8	7.121	0	0	1	0	0	0	0	0	4
3	0	118	8.365	0	0	0	1	0	0	0	1	0
3	1	0	8.148	0	0	0	1	0	0	0	0	1
3	2	956	7.612	0	0	0	1	0	0	0	0	2
3	3	629	7.185	0	0	0	1	0	0	0	0	3
4	0	435	8.331	0	0	0	0	1	0	0	1	0
4	1	5	8.229	0	0	0	0	1	0	0	0	1
4	2	393	7.749	0	0	0	0	1	0	0	0	2
5	0	2	8.537	0	0	0	0	0	1	0	1	0
5	1	194	8.422	0	0	0	0	0	1	0	0	1
6	0	6	8.746	0	0	0	0	0	0	1	1	0
		152										
		2										
		123										
		8										
		429										
		5										
		345										
		5										
		202										
		3										
		132										
		0										
		415										
		0										

374
7
232
0
510
2
454
8
623
8

The regression output for this model is given below. For ease of reference two extra lines have been inserted in this output. Firstly the parameter labels are shown above the parameter coefficient estimates and secondly the T-Ratios are shown.

Regression output:

Constant		0
Std Err of Y Est		.1139
R squared(Adj,Raw)	.9762	.9832
No. of Observations		28
Degrees of Freedom		19

	a_0	a_1	a_2	a_3	a_4	a_5	a_6	d	s
Coefficient(s)	8.57	8.574	8.665	8.554	8.637	8.846	9.042	–.296	–.435
Std Err of Coef.	3	.072	.069	.070	.076	.091	.134	.070	.018
T-ratios	.076	119.9	124.9	121.8	113.8	97.6	67.6	–4.2	–23.5
	113.								
	3								

The development parameters, d and s are significantly different from zero as their T-Ratios (parameter estimate divided by its standard error estimate) are –4.2 and –23.5 respectively which are well outside the usual 95% confidence interval (critical) range of –2 to 2.

The accident year parameters are also all significantly different from zero, as they surely have to be with this model's assumptions (all accident year levels are significantly above zero), but they do look close to one another. In order to test whether these are distinct it is necessary to redefine the model by dropping the a_0 parameter and replacing it with a constant. The only change to the design matrix is that the first column is now made up of ones.

The regression output of the redefined model is almost identical:

Regression Output:

Constant		0
Std Err of Y Est		.1139
R Squared (Adj,Raw)	.9762	.9832
No. Of Observations		28

Degrees of Freedom 19

	k	a_1	a_2	a_3	a_4	a_5	a_6	d	s
Coefficient(s)	8.573	.001	.092	−.019	.064	.273	.469	−.296	−.435
Std Err of Coef.	.076	.064	.069	.075	.084	.098	.132	.070	.018
T-ratios	113.3	.0	1.3	−.2	.8	2.8	3.6	−4.2	−23.5

The output clearly shows a much better definition of the same model as it identifies that the accident years 1,2,3 and 4 parameters are not significantly different from zero or, in comparison to the previous definition, significantly different from the zero'th accident year parameter which has now become the constant level value k. Based on this definition the model parameters for accident years 0,1,2,3 and 4 can be set to zero and be effectively estimated by a new common value k. This new constant of the reduced parameter model should now be an average value for the five accident years whose individual parameters have been dropped from the model.

A theoretically more appealing approach for inducing a partition in the accident year parameters, based on the multicomparison t-criterion test, can be found in Renshaw (2).

Setting a_0 to a_4 to zero reduces the model parameters to just the five parameters k, a_5, a_6, d and s which we expect to be significant.
The design matrix is now simpler as can be seen from Table 2 below.

Table 2: Regression Table for the Reduced Parameter Model

				←design matrix →				
i	j	P_{ij}	Y_{ij}	k	a_5	a_6	d	s
0	0	3511	8.164	1	0	0	1	0
0	1	3215	8.076	1	0	0	0	1
0	2	2266	7.726	1	0	0	0	2
0	3	1712	7.445	1	0	0	0	3
0	4	1059	6.965	1	0	0	0	4
0	5	587	6.375	1	0	0	0	5
0	6	340	5.829	1	0	0	0	6
1	0	4001	8.294	1	0	0	1	0
1	1	3702	8.217	1	0	0	0	1
1	2	2278	7.731	1	0	0	0	2
1	3	1180	7.073	1	0	0	0	3
1	4	956	6.863	1	0	0	0	4
1	5	629	6.444	1	0	0	0	5
2	0	4355	8.379	1	0	0	1	0
2	1	3932	8.277	1	0	0	0	1
2	2	1946	7.574	1	0	0	0	2
2	3	1522	7.328	1	0	0	0	3

2	4	1238	7.121	1	0	0	0	4
3	0	4295	8.365	1	0	0	1	0
3	1	3455	8.148	1	0	0	0	1
3	2	2023	7.612	1	0	0	0	2
3	3	1320	7.185	1	0	0	0	3
4	0	4150	8.331	1	0	0	1	0
4	1	3747	8.229	1	0	0	0	1
4	2	2320	7.749	1	0	0	0	2
5	0	5102	8.537	1	1	0	1	0
5	1	4548	8.422	1	1	0	0	1
6	0	6283	8.746	1	0	1	1	0

The regression output for this reduced parameter model is

Regression Output:

Constant		0
Std Err of Y Est		.1119
R Squared(Ajd,Raw)	.9770	.9804
No. of Observations		28
Degrees of Freedom		23

	k	a_5	a_6	d	s
Coefficient(s)	8.608	.244	.441	–.303	–.440
Std Err of Coef.	.052	.085	.122	.068	.017
T-ratio	167.1	2.9	3.6	–4.5	–26.4

As expected the constant has now changed as it is an average value for the first five accident years. The other parameters have also changed slightly.

All the parameters are now significantly different from zero, with t-ratios exceeding absolute 2, as expected. The quality of fit is still good and the number of parameters has been reduced from nine to five. The model looks reasonable enough to warrant further investigation.

The next section considers some basic testing using residual analysis plots of the first (all parameter) model and this reduced parameter model.

Projections from both these models will be calculated and compared after this analysis.

I. Testing the models by residual analysis plots

The parameter estimates from the regressions can now be used to calculate the model estimates, in log-space, which can then be compared with the observed values in log-space. It is usual to use standardized residuals, defined as the difference between observed and fitted values divided by the model standard error, and considering these in graphical form. Under the IID assumptions used to derive the model estimates these residuals should exhibit a fair degree of randomness.

Testing now turns to the analysis of these standardized residuals. In practice these are plotted against development, accident and payment year and also against the fitted values. Working in a spreadsheet makes this process very easy as each chart can be defined as an X-Y chart with Y the standardized residuals and X the other variable in turn.

Table 3 below shows the actual values, their logarithms and the model fitted values in log-space for the full parameter model as defined in Table 1. The residuals are just the differences between the observed and fitted values in log-space and the standardized residuals are the residuals divided by the model standard error, which was .1139 for this model.

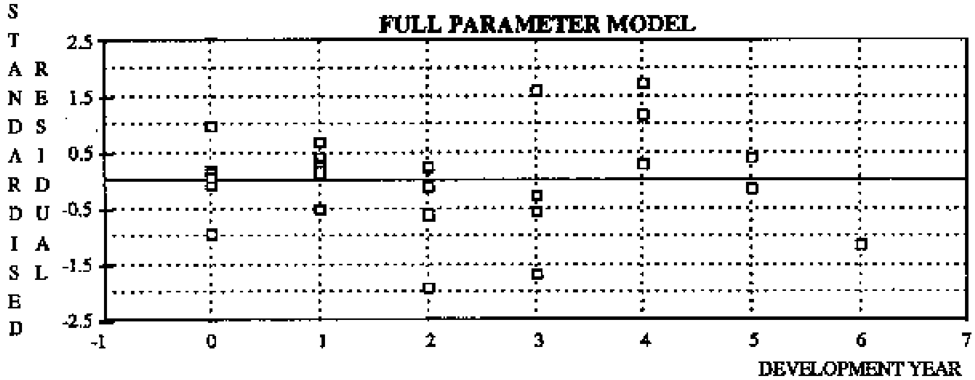
Table 3: Residuals Table for the Full Parameter Model.

Acc i	Dev j	Pay $i+j$	P_{ij}	Y_{ij}	$\hat{Y}_{ij}$	Resid	Stand Resid
0	0	0	3511	8.164	8.277	−.113	−.991
0	1	1	3215	8.076	8.138	−.062	−.547
0	2	2	2266	7.726	7.703	.023	.201
0	3	3	1712	7.445	7.268	.177	1.557
0	4	4	1059	6.965	6.833	.132	1.159
0	5	5	587	6.375	6.398	−.023	−.202
0	6	6	340	5.829	5.963	−.134	−1.177
1	0	1	4001	8.294	8.278	.017	.147
1	1	2	3702	8.217	8.139	.078	.683
1	2	3	2278	7.731	7.704	.027	.239
1	3	4	1180	7.073	7.269	−.196	−1.717
1	4	5	956	6.863	6.834	.029	.253
1	5	6	629	6.444	6.399	.045	.396
2	0	2	4355	8.379	8.369	.010	.091
2	1	3	3932	8.277	8.230	.047	.412
2	2	4	1946	7.574	7.795	−.221	−1.943
2	3	5	1522	7.328	7.360	−.032	−.283
2	4	6	1238	7.121	6.925	.196	1.722
3	0	3	4295	8.365	8.258	.107	.942
3	1	4	3455	8.148	8.119	.028	.249
3	2	5	2023	7.612	7.684	−.072	−.631
3	3	6	1320	7.185	7.249	−.064	−.560
4	0	4	4150	8.331	8.340	−.010	−.084
4	1	5	3747	8.229	8.202	.027	.237
4	2	6	2320	7.749	7.767	−.017	−.153
5	0	5	5102	8.537	8.549	−.012	−.104
5	1	6	4548	8.422	8.411	.012	.104
6	0	6	6283	8.746	8.746	.000	.000

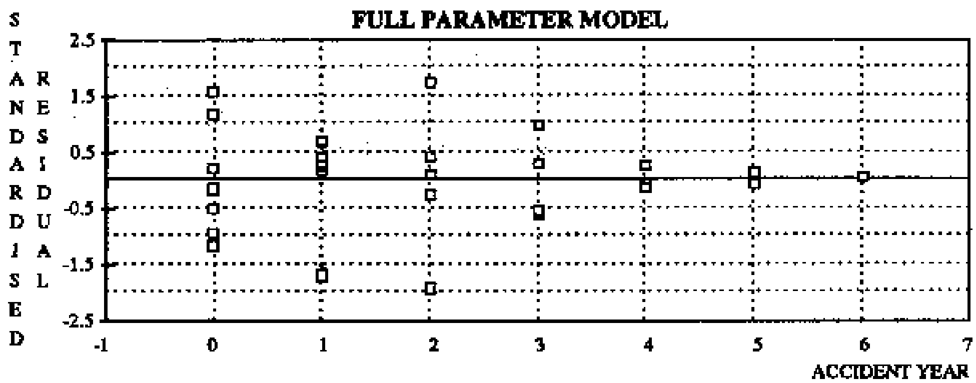
To produce the residual plots in X-Y chart form the standardized residuals column is defined as the Y-variate and the first three columns in turn as the X-variate for the accident year, development year and payment year plots. For the final plot the fitted values column is picked instead.

The various residual plots from this model are shown below in Charts 4 to 7.

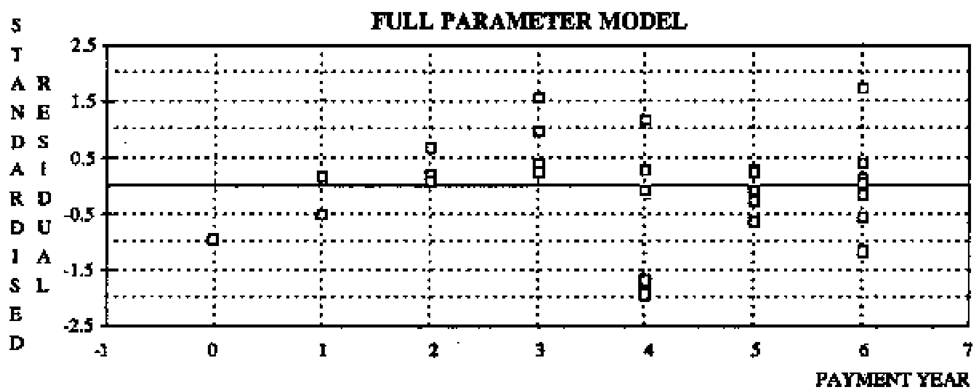
RESIDUAL ANALYSIS CHART 4

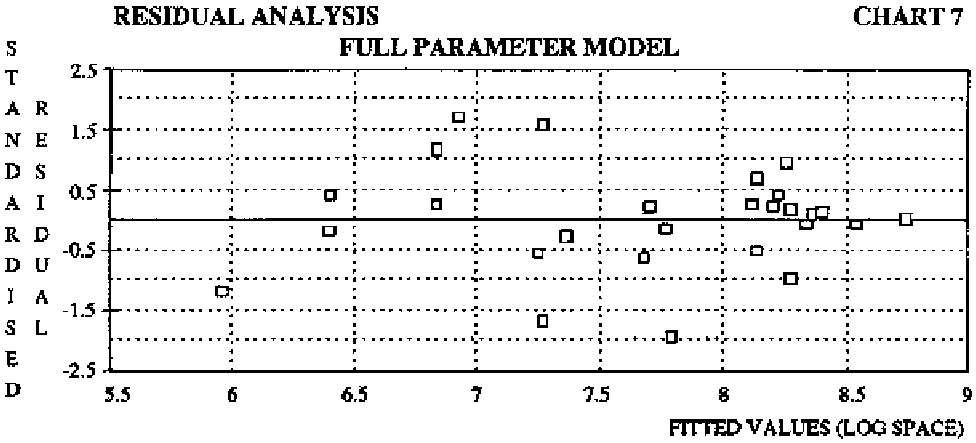


RESIDUAL ANALYSIS CHART 5



RESIDUAL ANALYSIS CHART 6

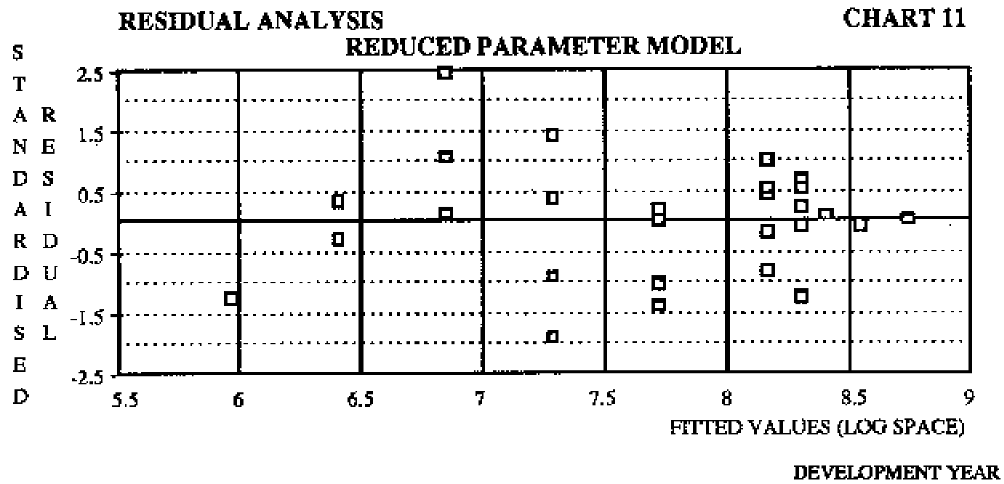


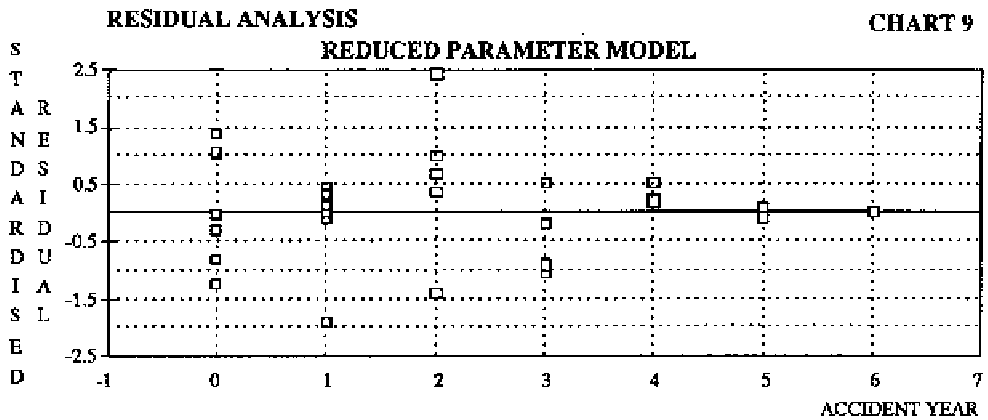


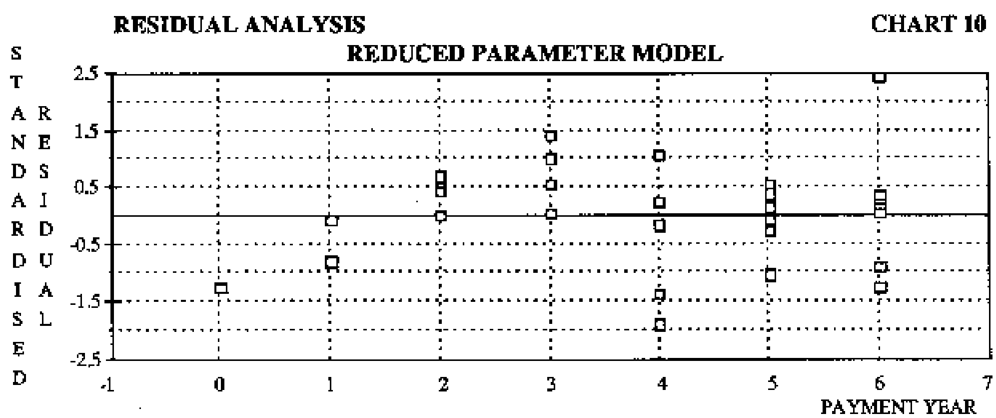
The residuals for the Reduced Parameter Model, which is defined in Table 2 (common level value for the first five accident years), are shown in Table 4 below.

Table 4: Residuals Table for the Reduced Parameter Model.

Acc i	Dev j	Pay i+j	P_{ij}	Y_{ij}	$\hat{Y}_{ij}$	Resid	Stand Resid
0	0	0	3511	8.164	8.304	–.141	–1.259
0	1	1	3215	8.076	8.168	–.093	–.828
0	2	2	2266	7.726	7.729	–.003	–.025
0	3	3	1712	7.445	7.289	.156	1.398
0	4	4	1059	6.965	6.849	.116	1.035
0	5	5	587	6.375	6.410	–.035	–.309
0	6	6	340	5.829	5.970	–.141	–1.260
1	0	1	4001	8.294	8.304	–.010	–.091
1	1	2	3702	8.217	8.168	.048	.432
1	2	3	2278	7.731	7.729	.002	.022
1	3	4	1180	7.073	7.289	–.216	–1.927
1	4	5	956	6.863	6.849	.013	.121
1	5	6	629	6.444	6.410	.035	.309
2	0	2	4355	8.379	8.304	.075	.667
2	1	3	3932	8.277	8.168	.109	.971
2	2	4	1946	7.574	7.729	–.155	–1.386
2	3	5	1522	7.328	7.289	.039	.347
2	4	6	1238	7.121	6.849	.272	2.431
3	0	3	4295	8.365	8.304	.061	.543
3	1	4	3455	8.148	8.168	–.021	–.185
3	2	5	2023	7.612	7.729	–.116	–1.039
3	3	6	1320	7.185	7.289	–.104	–.925
4	0	4	4150	8.331	8.304	.026	.236
4	1	5	3747	8.229	8.168	.060	.540
4	2	6	2320	7.749	7.729	.021	.185
5	0	5	5102	8.537	8.548	–.011	–.095
5	1	6	4548	8.422	8.412	.011	.095
6	0	6	6283	8.746	8.746	.000	.000







The various residual plots from this model are shown below in Charts 8 to 11.

This reduced parameter model has a standardized residual for accident year 2, development period 4, of 2.431 as the maximum (absolute) standardized residual value. The full parameter model had a lowest standardized residual of –1.943 (i=2, j=2). The second model has a slightly smaller standard error of .1119 compared to the .1139 of the full parameter model. There is however little difference overall between these models detectable from the above tables. Both seem to fit the data fairly well.

The next stage is to consider these residuals in graphic form to examine whether any unmodelled trends are detectable.

In all these residual plots, according to the model error assumptions, we expect a set of fairly random points bounded in about 95% of cases within the −2 to 2 range.

As Table 3 and Table 4 above indicate, all the standardized residuals for the full parameter model are just in this range (Table 3) with just one value outside the range in the case of the reduced parameter model (Table 4). Values outside this range will sometimes occur and often identify outliers that may warrant further investigation.

The development year plots (Charts 4 and 8) will generally be the most interesting and particularly where, as in these cases, it has been assumed that there is some relationship connecting the development parameters. A particular feature worth looking out for in these plots is any tendency for the residuals to spread or fan out with development. This is not too noticeable in these examples. Note however that the residuals for development periods 4 to 6 in both cases do not appear very random. There are however only a few values involved and these may well be impacted by the outlier identified earlier (i=2, j=4). We have used a very simple shape to describe the run-off from development period 1 and these residual plots are quite reasonable in the circumstances.

The accident year residual plots are shown in Chart 5 for the full parameter model and in Chart 9 for the reduced model. Considering the former first, as each accident year has its own parameter in this model, the plot should be boringly predictable with the residuals balanced about the zero horizontal. Chart 5 shows this quite clearly.

The reduced model accident year residuals, Chart 9, look very similar although here the first five accident years have effectively been fitted by a single parameter. The only visible differences are the accident 4 residuals which are all greater than zero. In a fuller analysis this parameter should be added back to the reduced model and tested for significance. It is possible that it may become more significant if measured against the average for accident years 0 to 3 although this turns out not to be the case in this instance.

In both cases the accident year residuals appear to get closer to the zero horizontal line, with increasing accident year, resulting in the left half of both charts diverging from this line. This is due, at least in part, to over-parameterisation. In the extreme right, for example, as only one point is fitted and with its own parameter a perfect fit is obtained and the residual has to be zero. For accident year 5 two points are fitted and so the accident year parameter is again effective in ensuring a close fit. The values in these late accident years are also relatively large, as they are from earlier development periods when payments tend to be higher, and they may be relatively more stable. This is considered later.

The payment year residuals (Charts 6 and 10) can be interesting but more difficult to interpret. Inflationary forces are expected to operate along this direction but as accident year levels have been assumed independent this may mask any such influences. The plots for both models look very similar, which is not very surprising, as neither model considers this direction in its definition. Both these charts appear to show a definite non-random shape for the early payment years and this would warrant further investigation. Changes in claims inflation rates during the period concerned, which are not incorporated in the model, may well be the cause. This is not pursued here. The regression analysis at least identifies areas that would warrant further investigation in practice.

It was indicated earlier that higher values, generally in earlier development periods, may be relatively more stable than later, generally lower, values. This can be tested by plotting residuals against fitted values as is shown in Charts 7 and 11. In both these charts the last few residuals on the extreme right look close to the horizontal zero line but these points are the same points identified earlier as the last two or three accident year values. The residuals show a tendency to increase (in absolute terms) as values decrease. This effect, generally known as heteroscedasticity, is also detectable from the development year plots as incremental payments eventually decrease with development. No attempt is made here to overcome any heteroscedasticity.

The error term normality assumption can also be tested more formally within the spreadsheet if required. It is possible for instance to use the Data Distribution command to calculate and tabulate a frequency distribution of the residuals and compare values in this table with preset values calculated from the standard normal distribution.

The residual analysis indicates that these models have some weakness along the payment year direction and there are sufficient reasons to doubt some of the model assumptions. A full analysis would follow these up. In particular some inflation adjustment should be made to the data and the modelling process repeated to see whether this adjustment removes the non-random look of these residuals along the payment year direction. However for the time being it will be assumed that both these models are satisfactory and the regression results will be used in the next section to project the future payments and their standard errors from these two models.

A later section will consider a model with inflation and claim volume adjustment to see if a better model can be found.

J. Using the models to project future payments and standard errors

When the basic chain ladder model with independent development parameters is fitted it is not possible to extend the projections beyond the latest development contained in the triangle without resorting to some form of external curve fitting of development factors such as the Sherman inverse power curve for example.

In these examples as a curve (straight line) has been fitted to the development parameters it is possible to extend the model projections to development periods beyond those contained in the data triangle.

The model has a natural stop as the payments are decaying exponentially and so become small relatively quickly. So we could simply sum the implied geometric series or take the values to some development period beyond which we would expect no more payments in practice.

In what follows it is assumed that there are no payments beyond development 12, as this is sufficient for purposes of illustration and cuts down the values to be projected. In practice this will need to be decided on the merits of each case and knowledge of the likely run-off period of the particular class being investigated.

The data triangle contained 28 values and our completed rectangle has a total of 91 data points (7×13). There are therefore 63 individual payments and their standard errors to calculate.

The design and future design matrices are first produced and these are used to produce the variance-covariance matrix of the future values. This is now a 63 × 63 matrix and should be within the capability of a reasonable spreadsheet. Both Lotus 123 Version 2.2 and SuperCalc5 Version 5.0 can handle square matrices of around 89 × 89.

For producing the future values and their associated (individual) standard errors only the diagonal elements of this matrix are needed. The calculations from here are fairly simple and are shown in the Tables 5 and 7 below. These tables are set in the way one would normally produce them in a spreadsheet. The values are arranged by accident year first, as this is how the future design matrix was set out. The accident year order was adopted here as this order facilitates the computation of the accident year standard errors.

The second table in each set (Tables 6 and 8) show the projected values and standard errors in a more traditional format and also include accident year and overall totals for both values and standard errors. These calculations are also set out in the spreadsheet as explained in Section F. In view of the size of the matrices involved they have not been shown here.

The various matrix products needed to calculate the variance-covariance matrix (as set out in Appendix 1 for the earlier chain ladder example) took under two minutes on a 12MHz PC fitted with a maths co-processor.

Table 5: Projection for the Full Parameter Model: Part a

i	j	$\hat{Y}_{ij}$	var $\hat{Y}_{ij}$	$\hat{P}_{ij}$	se $\hat{P}_{ij}$	% error
0	7	5.528	.0195	254	36	14.0%
0	8	5.093	.0223	165	25	15.0%
0	9	4.658	.0258	107	17	16.2%
0	10	4.223	.0301	69	12	17.5%
0	11	3.788	.0350	45	8	18.9%
0	12	3.353	.0405	29	6	20.3%
1	6	5.964	.0185	393	54	13.7%
1	7	5.529	.0210	255	37	14.6%
1	8	5.094	.0241	165	26	15.6%
1	9	4.659	.0280	107	18	16.8%
1	10	4.224	.0325	69	13	18.2%
1	11	3.789	.0377	45	9	19.6%
1	12	3.354	.0436	29	6	21.1%
2	5	6.490	.0179	664	89	13.4%
2	6	6.055	.0199	431	61	14.2%
2	7	5.620	.0227	279	42	15.2%
2	8	5.185	.0261	181	29	16.3%
2	9	4.750	.0302	117	21	17.5%
2	10	4.315	.0350	76	14	18.9%
2	11	3.880	.0405	49	10	20.3%
2	12	3.445	.0467	32	7	21.9%
3	4	6.814	.0177	919	123	13.3%
3	5	6.379	.0193	595	83	14.0%
3	6	5.944	.0216	386	57	14.8%
3	7	5.509	.0246	250	39	15.8%
3	8	5.074	.0283	162	27	17.0%
3	9	4.639	.0327	105	19	18.2%
3	10	4.205	.0378	68	13	19.6%
3	11	3.770	.0435	44	9	21.1%
3	12	3.335	.0499	29	7	22.6%
4	3	7.332	.0181	1542	209	13.5%
4	4	6.897	.0193	999	139	14.0%
4	5	6.462	.0212	647	95	14.6%
4	6	6.027	.0237	419	65	15.5%
4	7	5.592	.0269	272	45	16.5%
4	8	5.157	.0308	176	31	17.7%
4	9	4.722	.0354	114	22	19.0%
4	10	4.287	.0407	74	15	20.4%
4	11	3.852	.0466	48	11	21.8%
4	12	3.417	.0532	31	7	23.4%
5	2	7.976	.0202	2939	420	14.3%
5	3	7.541	.0208	1903	276	14.5%
5	4	7.106	.0220	1232	184	14.9%
5	5	6.671	.0240	798	124	15.6%
5	6	6.236	.0266	518	85	16.4%

5	7	5.801	.0298	336	58	17.4%
5	8	5.366	.0338	218	40	18.5%
5	9	4.931	.0385	141	28	19.8%
5	10	4.496	.0438	92	19	21.2%
5	11	4.061	.0498	59	13	22.6%
5	12	3.626	.0565	39	9	24.1%

Table 5: Projection for the Full Parameter Model: Part b

i	j	$\hat{Y}_{ij}$	var $\hat{Y}_{ij}$	$\hat{P}_{ij}$	se $\hat{P}_{ij}$	% error
6	1	8.607	.0296	5550	962	17.3%
6	2	8.172	.0290	3592	616	17.1%
6	3	7.737	.0290	2325	399	17.2%
6	4	7.302	.0298	1506	262	17.4%
6	5	6.867	.0313	975	174	17.8%
6	6	6.432	.0334	632	116	18.4%
6	7	5.997	.0362	410	79	19.2%
6	8	5.562	.0397	266	53	20.1%
6	9	5.127	.0439	172	36	21.2%
6	10	4.692	.0487	112	25	22.4%
6	11	4.257	.0543	73	17	23.6%
6	12	3.822	.0605	47	12	25.0%

TOTAL = 34377

Table 6: Projected values and Standard Errors.

Full Parameter Model.

		Development Year												
Yr		1	2	3	4	5	6	7	8	9	10	11	12	Total
0	£							254	165	107	69	45	29	669
	se							36	25	17	12	8	6	79
1	£						393	255	165	107	69	45	29	1063
	se						54	37	26	18	13	9	6	119
2	£					664	431	279	181	117	76	49	32	1830
	se					89	61	42	29	21	14	10	7	196
3	£				919	595	386	250	162	105	68	44	29	2559
	se				123	83	57	39	27	19	13	9	7	265
4	£			1542	999	647	419	272	176	114	74	48	31	4324
	se			209	139	95	65	45	31	22	15	11	7	443
5	£		2939	1903	1232	798	518	336	218	141	92	59	39	8274
	se		420	276	184	124	85	58	40	28	19	13	9	890
6	£	5550	3592	2325	1506	975	632	410	266	172	112	73	47	15659
	se	962	616	399	262	174	116	79	53	36	25	17	12	2158

Overall Total 34377

Standard Error	2742
Percent. Error	7.98

Table 7: Projection for the Reduced Parameter Model: Part a

i	j	$\hat{Y}_{ij}$	var $\hat{Y}_{ij}$	$\hat{P}_{ij}$	se $\hat{P}_{ij}$	% error
0	7	5.530	.0182	255	35	13.6%
0	8	5.091	.0209	164	24	14.5%
0	9	4.651	.0241	106	17	15.6%
0	10	4.211	.0279	68	11	16.8%
0	11	3.772	.0322	44	8	18.1%
0	12	3.332	.0371	29	6	19.4%
1	6	5.970	.0161	395	50	12.8%
1	7	5.530	.0182	255	35	13.6%
1	8	5.091	.0209	164	24	14.5%
1	9	4.651	.0241	106	17	15.6%
1	10	4.211	.0279	68	11	16.8%
1	11	3.772	.0322	44	8	18.1%
1	12	3.332	.0371	29	6	19.4%
2	5	6.410	.0146	612	74	12.1%
2	6	5.970	.0161	395	50	12.8%
2	7	5.530	.0182	255	35	13.6%
2	8	5.091	.0209	164	24	14.5%
2	9	4.651	.0241	106	17	15.6%
2	10	4.211	.0279	68	11	16.8%
2	11	3.772	.0322	44	8	18.1%
2	12	3.332	.0371	29	6	19.4%
3	4	6.849	.0136	950	111	11.7%
3	5	6.410	.0146	612	74	12.1%
3	6	5.970	.0161	395	50	12.8%
3	7	5.530	.0182	255	35	13.6%
3	8	5.091	.0209	164	24	14.5%
3	9	4.651	.0241	106	17	15.6%
3	10	4.211	.0279	68	11	16.8%
3	11	3.772	.0322	44	8	18.1%
3	12	3.332	.0371	29	6	19.4%
4	3	7.289	.0132	1474	170	11.5%
4	4	6.849	.0136	950	111	11.7%
4	5	6.410	.0146	612	74	12.1%
4	6	6.970	.0161	395	50	12.8%
4	7	5.530	.0182	255	35	13.6%
4	8	5.091	.0209	164	24	14.5%
4	9	4.651	.0241	106	17	15.6%
4	10	4.211	.0279	68	11	16.8%
4	11	3.772	.0322	44	8	18.1%
4	12	3.332	.0371	29	6	19.4%
5	2	7.972	.0195	2927	411	14.0%
5	3	7.532	.0199	1886	267	14.2%

5	4	7.093	.0208	1216	176	14.5%
5	5	6.653	.0224	784	118	15.0%
5	6	6.213	.0244	506	79	15.7%
5	7	5.774	.0270	326	54	16.6%
5	8	5.334	.0302	210	37	17.5%
5	9	4.894	.0340	136	25	18.6%
5	10	4.455	.0382	88	17	19.7%
5	11	4.015	.0431	57	12	21.0%
5	12	3.575	.0485	37	8	22.3%

Table 7: Projection for the Reduced Parameter Model: Part b

i	j	$\hat{Y}_{ij}$	var $\hat{Y}_{ij}$	$\hat{P}_{ij}$	se $\hat{P}_{ij}$	% error
6	1	8.609	.0285	5562	946	17.0%
6	2	8.170	.0279	3582	603	16.8%
6	3	7.730	.0279	2308	388	16.8%
6	4	7.290	.0284	1487	252	17.0%
6	5	6.851	.0295	959	166	17.3%
6	6	6.411	.0311	618	110	17.8%
6	7	5.971	.0333	399	73	18.4%
6	8	5.532	.0361	257	49	19.2%
6	9	5.092	.0394	166	33	20.0%
6	10	4.652	.0432	107	23	21.0%
6	11	4.213	.0476	69	15	22.1%
6	12	3.773	.0526	45	10	23.2%

TOTAL = 33847

Table 8: Projected values and Standard Errors

Reduced Parameter Model.

		Development Year												
Yr		1	2	3	4	5	6	7	8	9	10	11	12	Total
0	£							255	164	106	68	44	29	666
	se							35	24	17	11	8	6	75
1	£						395	255	164	106	68	44	29	1060
	se						50	35	24	17	11	8	6	106
2	£					612	395	255	164	106	68	44	29	1672
	se					74	50	35	24	17	11	8	6	146
3	£				950	612	395	255	164	106	68	44	29	2622
	se				111	74	50	35	24	17	11	8	6	200
4	£			1474	950	612	395	255	164	106	68	44	29	4096
	se			170	111	74	50	35	24	17	11	8	6	275
5	£		2927	1886	1216	784	506	326	210	136	88	57	37	8173
	se		411	267	176	118	79	54	37	25	17	12	8	851
6	£	5562	3582	2308	1487	959	618	399	257	166	107	69	45	15558
	se	946	603	388	252	166	110	73	49	33	23	15	10	2101
												Overall Total		33847

Standard Error	2545
Percent. Error	7.52

K. Overall standard error and accident year standard errors

The calculations necessary to produce the accident year and overall standard errors shown in Tables 6 and 8 above are a repeat of those shown in Section G. The only complication is that in the above cases there are more values to project (63 rather than 6) so there is a lot more to calculate.

The results are very close. The full model produces estimated future payments of 34377 with a standard error of 2742 or 7.98%. The reduced parameter model produces estimated future payments of 33847 with a standard error of 2545 or 7.52%. The two estimated values are not significantly different but the second model has a proportionately smaller standard error. This is purely due to the smaller number of parameters used in defining this model. The second model may therefore be considered to have the slight advantage over the first.

The closeness of these results is not particularly surprising as the two models are very similar. Most of the future payments relate to the last two accident years and here both models have assumed these years to have independent levels (just like the chain ladder model) and so any smoothing from the reduced parameter model affects only the earliest accident years where the projected future payment values are not so large.

In fact assumptions about the most recent accident years are crucial to any reserve analysis. The base data used in this example is unadjusted for inflation and claim volume and the levels for the various accident years are not normally expected to be as close as those of the first five accident years above.

The next section will consider modelling the inflation and volume adjusted data.

L. Adjusting for inflation and claim volumes

It is possible to reduce the model parameters further by using an inflation index to bring all payments to current value and a claims volume adjustment or weight for each accident year so as to normalize these payments.

The claim volume values to be used in this example are based on the number of claims reported by the end of the first development period. They are scaled for convenience.

Accident Year	0	1	2	3	4	5	6
Claim Volume	1.43	1.45	1.52	1.35	1.29	1.47	1.91

An earnings index for the relavant period will be used in this case to bring payment values to payment year 6 (the latest payment year) values. In practice case is needed to ensure that the index used is the most appropriate index for the class of claims under investigation.

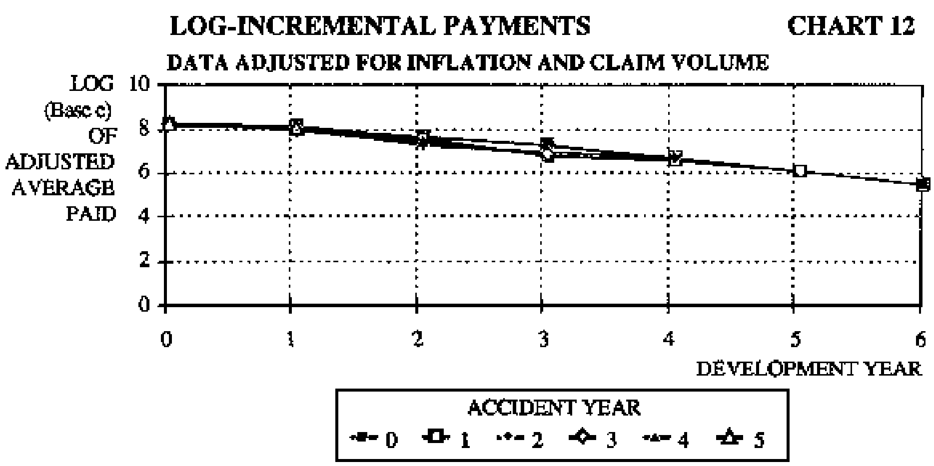
Payment year	0	1	2	3	4	5	6
Index	1.55	1.41	1.30	1.23	1.13	1.05	1

The inflation adjusted, volume normalized incremental payments (shown in integer format but calculated and used to many decimal places) are now as follows:

	Development Year						
Acc Yr	0	1	2	3	4	5	6
0	3806	3170	2060	1473	837	431	238
1	3891	3319	1932	920	692	434	
2	3725	3182	1447	1051	814		
3	3913	2892	1573	978			
4	3635	3050	1798				
5	3644	3094					
6	3290						

Even before any further analysis is carried out it is clear from this triangle that there is a fair amount of consistency and stability in the adjusted data.

Plotting the log-incremental adjusted data, as can be seen from Chart 12 below,



appears to confirm this observation. The various lines, each representing an accident year, look closely grouped together for at least the first couple of development periods.
The chart indicates that accident year effects may have been reduced or eliminated and the first test will be to confirm whether this is the case. As the shape of these lines is as before the same assumptions will be made in modelling the shape.

The design matrix is initially exactly as in the previous example which assumed accident years 1 to 6 as independent variates and had an independent first development level (d) and then a linear trend with common slope s.

The regression output using the adjusted values and including the extra two lines as before is:

Regression Output: Full Parameter Model.

Constant		0
Std Err of Y Est		.1153
R Squared(Adj,Raw)	0.9788	.9851
No. of Observations		28
Degrees of Freedom		19

	k	a_1	a_2	a_3	a_4	a_5	a_6	d	s
Coefficient(s)	8.627	−.087	−.114	−.175	−.120	−.110	−.237	−.292	−.505
Std Err of Coef.	.077	.065	.070	.076	.085	.099	.133	.071	.019
T-Ratios	112.6	−1.3	−1.6	−2.3	−1.4	−1.1	−1.8	−4.1	−27.0

Accident year 3 turns out to be the only one whose parameter has a T-ratio whose absolute value exceeds 2 and may be considered significant.

So the next stage is to eliminate all the accident years with T-Ratios less than absolute 2 and refit. There are now four parameters namely

k a_3 d and s

The regression output of this model is:

Regression Output:

Constant		0
Std Err of Y Est		.1157
R Squared(Adj,Raw)	.9787	.9810
No. Of Observations	28	
Degrees of Freedom	24	

	k	a_3	d	s
Coefficient(s)	8.523	−.088	−.296	−.493
Std Err of Coef.	.054	.063	.068	.017
T-Ratios	157.2	−1.4	−4.3	−28.7

The parameters of this model can still be reduced as the accident year three parameter is now not significant. What has happened is that it is now being measured against the "average" of all the other accident year levels rather than just the first accident year level and this has been sufficient to make this last accident year parameter close enough to the average value. Care needs to be taken to ensure that none of the other parameters have become significant in the new model.

So this remaining accident year parameter will be dropped, leaving only three parameters, one for the common level k, and the two shape parameters d and s.

The regression output of this three parameter model is:

Regression Output:

Constant		0
Std Err of Y Est		.1179
R Squared(Adj,Raw)	.9779	.9795
No. Of Observations		28
Degrees of Freedom		25

	k	d	s
Coefficient(s)	8.501	–.286	–.489
Std Err of Coef.	.053	.069	.017
T-Ratios	161.3	–4.1	–28.3

This is an interesting stage. There are now only three parameters and all are significant. The model has a high R-squared value and appears to describe the data reasonably well. It is now tempting to use this model to project future payments.

The process is as before with the minor irritation of scaling the estimated values for claim volumes and using some future inflation index to take the projected payments to final values. The inflation rate to be used here is 7.5% p.a. which is chosen as it is close to the average annual historic rate implied by the index used to adjust the historic payments and will facilitate the comparison of the results. In practice a more appropriate prospective rate or rates will normally be utilized and a number of these used to obtain estimates.

Table 10 below shows the results derived from the full parameter model and inflation at 7.5% p.a.

Table 10: Projected values and Standard Errors.

Full Parameter Model with inflation at 7.5%

		Development Year												
Yr		1	2	3	4	5	6	7	8	9	10	11	12	Total
0	£							253	164	107	70	45	29	669
	se							36	25	18	12	9	6	80
1	£						389	253	164	107	70	45	29	1058
	se						54	37	26	18	13	9	6	120
2	£					658	427	278	181	117	76	50	32	1820
	se					89	61	43	30	21	15	10	7	198
3	£				911	592	384	250	162	106	69	45	29	2547
	se				123	84	58	40	28	19	14	10	7	267
4	£			1524	990	643	418	271	177	115	75	49	32	4292
	se			209	140	95	65	45	32	22	15	11	7	445
5	£		2910	1889	1226	797	518	336	219	142	93	60	39	8229
	se		421	277	185	126	86	59	41	29	20	14	10	896
6	£	5544	3596	2334	1515	984	639	415	270	176	114	74	48	15709
	se	972	624	406	267	177	119	81	55	38	26	18	12	2191

Overall Total	34324
Standard Error	2779
Percent. Error	8.10

The results are very close to those obtained earlier (Table 6) from the almost identical model without explicit inflation assumptions.

Increasing the inflation rate to 8.5% p.a. increases the overall estimate to 35210 with a standard error of 2858. So the one percentage change in the assumed future inflation rate impacts the estimated future payments by 2.6%.

Turning now to the reduced parameter model, that is the three parameter model with no accident year effects apart from the common level we obtain the following results assuming future inflation at 7.5% p.a.

Table 11: Projected values and Standard Errors.

Reduced Parameter Model with inflation at 7.5%.

Development Year

Yr		1	2	3	4	5	6	7	8	9	10	11	12	Total
0	£							249	165	109	72	47	31	673
	se							36	25	18	13	9	6	79
1	£						412	272	179	118	78	52	34	1145
	se						55	39	27	19	14	10	7	120
2	£					703	464	306	202	134	88	58	39	1994
	se					90	62	44	31	22	16	11	8	184
3	£				1018	671	443	292	193	127	84	56	37	2921
	se				1125	86	59	42	29	21	15	11	7	235
4	£			1585	1045	690	455	300	198	131	87	57	38	4586
	se			192	129	88	61	43	30	21	15	11	8	323
5	£		2945	1942	1280	845	557	368	243	160	106	70	46	8563
	se		358	235	158	108	75	52	37	26	19	13	9	541
6	£	6241	4114	2712	1788	1180	778	514	339	224	148	98	65	18201
	se	777	500	329	220	151	105	73	52	37	26	19	13	1090

Overall Total	38083
Standard Error	1725
Percent. Error	4.53

The results now look, and are, different. The overall estimate is significantly up on the previous estimates and the standard error is much reduced. The reduction in the overall standard error is due to the smaller number of parameters left in the reduced model and reflects the increased degree of smoothing that this parameter reduction has produced.

The increase in the overall projection, at just under 11%, is however too high to be explained by the derived standard errors. The main contributor can be clearly identified from the tables as the last accident year. This is not too surprising with hindsight. There is only a single data point from which to project. If it is assumed, as in the first case, that each year has an independent level then this point alone determines the level of the last accident year. The accident year residual plot for the latter model (Chart 14) shows the standardized residual for accident year 6 at around −1. Although this will not generally be considered statistically significant its impact, in a reserving context, has become significant.

Assuming a common level (in the adjusted figures) substantially reduces the influence of this last data point on its accident year estimate of future payments. As the adjusted triangle figures show, the one and only value for this last accident year is substantially below the corresponding values of the prior years. Using the same average value for all accident years gives the last accident year an average value which is now just under 16% higher than the value estimated from its own single data point.

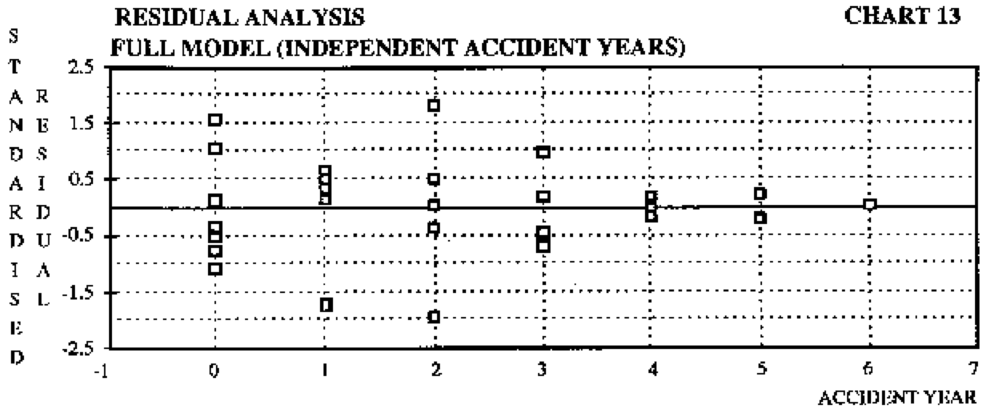
Putting the last accident year back into the model will produce results which will broadly match the full model overall estimate but with a reduced standard error. These are shown below.

Table 12: Projected values and Standard Errors.

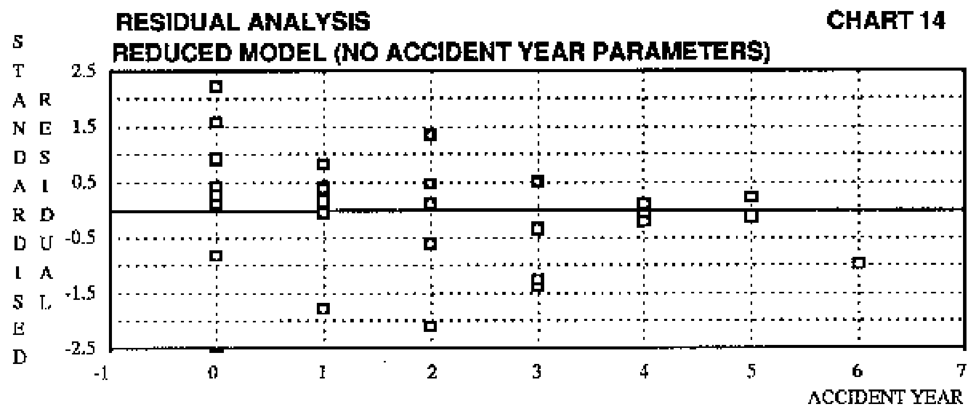
Reduced Parameter Model with Acc Yr 6, inflation at 7.5%.

		Development Year												
Yr		1	2	3	4	5	6	7	8	9	10	11	12	Total
0	£							249	165	109	72	47	31	673
	se							35	25	18	13	9	6	79
1	£						412	272	179	118	78	52	34	1145
	se						55	39	27	19	14	10	7	120
2	£					703	464	306	202	134	88	58	39	1994
	se					90	62	44	31	22	15	11	8	183
3	£				1017	671	443	292	193	127	84	56	37	2921
	se				125	85	59	42	29	21	15	11	7	234
4	£			1585	1045	689	455	300	198	131	87	57	38	4585
	se			192	128	88	61	43	30	21	15	11	8	322
5	£		2945	1942	1280	845	557	368	243	160	106	70	46	8562
	se		357	235	157	108	75	52	37	26	19	13	9	540
6	£	5494	3621	2387	1574	1038	685	452	299	197	130	86	57	16021
	se	981	639	421	280	188	127	87	60	41	28	20	14	2258

Overall Total	35902
Standard Error	2609
Percent. Error	7.27



(NOTE HERE THE RESIDUAL FOR ACC YR 6)



(IT RESULTS IN A 7.3% INCREASE IN THE OVERALL ESTIMATE)

Both these models are reasonable. They fit the data well and the standard errors are quite small. The results are quite different and these differences are clearly not explained by the standard errors, and are primarily due to the choice of parameters. As we know little about the underlying account it will be very difficult to choose between these models. In practice additional information, and informed views, will need to be sought to assist in this choice. This can then be used directly in deciding which parameters are to be left in the model.

A theoretically more appealing approach is to use some form of external or prior distribution and estimate in a Bayesian framework. This is explained in more detail by Verrall (6). It is possible to carry out the necessary calculations in the spreadsheet but more computation is necessary. The Bayesian approach combines formal statistical theory and informed prior estimates (knowledge and expertise!) and would appear to represent almost an ideal combination of theory and practice for reserving work. In practice more work is necessary in order to understand how sensitive the results are to these prior estimates, especially as these are made in log-space which, while convenient, are nevertheless somewhat alien from the immediate everyday experience of practitioners.

M. Final comments

This section will briefly consider some other aspects of these models which were deliberately avoided in earlier sections as the main emphasis has been a practical rather than a theoretical one.

a. Standard errors of reserve estimates

In practice, and as an approximation, as long as a sufficiently large number of future values are being projected it may be assumed that the distribution of the overall estimate obtained is normal with mean and standard error as calculated above.

That is we can use normal probability tables to establish approximate confidence intervals around the model reserve estimate. In the last example shown in Table 12 above for instance and under the conditions of the model, we have (approximately) a 95% probability that the required reserve will be less than 40194 ($35902 + 1.645 \times 2609$). Recall however that the error estimate may be incomplete and future inflation is assumed fixed reducing the possible error further.

In practice the specific variability of a particular class reserve estimate may be less important to management than the variability of the overall company claims reserve Balance Sheet figure.

The individual class standard errors may be used to obtain estimates of this overall variability. For example if mutual independence of reserve estimates by class is assumed the overall variance may be obtained as the sum of the individual variances. Under these circumstances the percentage error in the overall reserves can drop to low figures.

Much work remains to be done in this area. At least these methods provide a start point to such considerations.

There will clearly be other factors, not incorporated into the model, that in practice will add to the error terms. There was no attempt to explicitly adjust for inflation in the first examples although the models incorporated an implicit assumption which is then implicitly projected into the future.

In the later examples values were adjusted for past inflation, using an index that may or may not have been the most appropriate, and projected values calculated using an assumed future rate of inflation, or more correctly claims cost escalation. The examples assumed a future rate which was based on the average past inflation used in adjusting the data.

Relatively small changes in these assumed future rates can lead to relatively large changes to the overall projected values. These models can be used to produce a series of results, with varying future claims escalation assumptions, from which it may be possible to derive a measure of the additional variability that may arise from this source.

These models do not attempt to allow for changes in the speed of settlement of claims. Payment developments may appear stable due to a combination of accelerating costs counteracted by a slowdown in settlements. Clearly under such circumstances estimates from a regression model on log-incremental payments, or a chain ladder projection based on cumulative payments, are likely to produce estimates which may be seriously biased.

Finally there will generally be a lot more information available to management than that used in fitting any statistical model. It is just possible that a combination of statistical derived estimates with informed estimates based on specific and detailed knowledge of the particular business, its environment and claims, may produce final estimates that have reduced variability. This will be however difficult to prove.

b. Negative values in incremental data sets

One particular problem with log-linear models is the occasional negative value in the original space.

Negative values occur in practice especially in net of reinsurance incremental payments and in classes of business subject to large subrogation or salvage recoveries. Various alternative approaches are available to the modeller to deal with negative values in practice. One approach, adopted in a commercial package (ICRFS), is to add a sufficiently large constant to all the incremental values, so that they all become positive, before the logarithmic transformation and an adjustment made in the projected values.

An alternative approach, that may be acceptable in practice, is to shift payments from one period to an adjacent one so as to eliminate a negative value. This may be justified if it is known, or suspected, that the negative value is the result of some serial correlation, for example when preceded by a relatively large value. Another possibility, which may be tried where the negative value is small is to ignore the value totally or to set it to some small positive value such as 1 ($\log_e 1=0$).

No particular approach is recommended here as ideal for dealing with negative values. In practice the reason for such negative values has to be investigated and this process often helps identify an appropriate approach to deal with the problem. Clearly one should not ignore a genuine feature of the data for the sake of convenience.

c. Parsimony

The chain ladder model is sometimes considered overparameterised as it involves a parameter for each accident year and each development period. Too many parameters can lead to model instability. Increasing the number of explanatory variables improves the quality of the fitted data but such slavish adherence to the data often results in unstable projections. At the extreme one can always obtain a perfect fit by including enough parameters in the model. Such a model fails to achieve any smoothing of the data and will be very poor for prediction purposes. Parsimonious models, that is with fewer parameters are to be preferred for this reason. This is explained in more detail in the first article in this Volume of the Manual.

d. Serial Correlation and Heteroscedasticity

The triangular shaped incremental payments data tend to decrease as the development years increase and there is usually some serial correlation present in these payments for a particular accident year. Such correlation may occur when a low payment period, due to administrative problems for example, is followed by a catching up high payment period or vice versa. On net paid claims data this may happen when a gross payment is made in one period with the incoming associated reinsurance processed during the following period.

The decline in values in the development direction tends to result in the residuals increasing with development period. This characteristic is an example of heteroscedasticity. In effect the IID assumption implies that the error terms in the original space are subject to the same percentage variation irrespective of their absolute values. Experience with payments triangles indicates that as payments diminish in the tail the percentage variation of these payments tends to be much higher than that seen in the first few development periods when a greater volume of payments is usually being made. This may be more pronounced in net rather than gross payments.

Methods to overcome this are being developed. One approach followed by Zehnwirth in the ICRFS package (Interactive Claims Reserving and Forecasting System) is to use weights. Alternative error assumptions, which may well turn out to be more appropriate, are being investigated by others. The main disadvantage of these approaches is the difficulty of obtaining the parameter estimates compared to the comparatively easy spreadsheet regression approach.

e. The Hoerl run-off curves

A particularly useful family of curves for run-off patterns is the Gamma family defined by

$$P_{ij} = K_j\,(1 + j)^b \exp(aj)$$

Each curve has a level parameter K_j and two shape parameters b and a the latter being an exponential. They have the immediate advantage of becoming linear in log-space and can be fitted simply by multiple regression using the techniques of this article. These curves form the start point in the ICRFS package.

As the example above illustrated these curves do not always produce good fits for all development periods. They can be particularly poor in fitting the first few development periods which clearly have a significant influence on the reserves projected for the most recent accident years where a substantial amount of the overall reserve is generally to be found.

It is possible to use the simple techniques outlined in this article to fit “mixed” models where some shape is fitted for later development periods and independent parameters fitted for the earlier periods. The example above fitted an independent first development parameter and an exponential decay curve thereafter. Any shape that can be expressed linearly (in log-space) can be tried even if in practice restrictions in “allowable” shapes will inevitably be necessary to keep any package to reasonable size.

f. *Conclusion*

Regression techniques are now beginning to dominate developments in claims reserving methodology. The formal approach adopted, whether utilizing maximum likelihood and IID normal errors or any other error model, at least enables the modeller to test the reasonableness of the assumptions. The model testing phase itself can often reveal interesting aspects of the data which may not be immediately obvious from looking at the cumulative payments.

These models can be very useful for inter-company comparisons and for comparing the stability of run-off triangles. Some results along these lines are to be found in Section E of the Claims Run-Off Patterns Working Party report presented to the 1989 GISG (General Insurance Study Group) Conference in Brighton.

This article is intended to give a practical introduction to these techniques and does not claim any original theoretical developments. The writer is particularly grateful to Arthur Renshaw and Richard Verrall of City University for their invaluable and patient explanations on this subject. The hope is that other practitioners can now begin to benefit by experimenting with these techniques.

References

(1) 1982: Kremer, E. "IBNR claims and the Two-Way Model of ANOVA", *Scandinavian Actuarial Journal*, 1, 47-55.

(2) 1989: Renshaw, A.E. "Chain Ladder and Interactive Modelling", *Journal of the Institute of Actuaries*, Vol. 116, 559-587

(3) 1988: Taylor, G.C. "Regression Models in Claims Analysis (II)". *William M.M. Campbell, Cook, Knight*, Sydney, Australia

(4) 1988: Verrall, R. "Bayes Linear Models and the Claims Run-Off Triangle". *Actuarial Research Paper No. 7*, The City University, London

(5) 1989: Verrall, R. "The Chain Ladder and Maximum Likelihood", *Actuarial Research Paper No. 12*, City University, London.

(6) 1989: Verrall, R. "On the Unbiased Estimation of Reserves from Loglinear Models", *Actuarial Research Paper No. 13*, City University, London.

(7) 1985: Zehnwirth, B. "Interactive Claims Reserving Forecasting System (ICRFS)" *Benhar Nominees Pty Ltd.* Tarramurra NSW Australia.

Appendix 1

Matrix calculations for the formal chain ladder example

Design matrix **X**

1	0	0	0	0	0	0
1	0	0	0	1	0	0
1	0	0	0	0	1	0
1	0	0	0	0	0	1
0	1	0	0	0	0	0
0	1	0	0	1	0	0
0	1	0	0	0	1	0
0	0	1	0	0	0	0
0	0	1	0	1	0	0
0	0	0	1	0	0	0

Design matrix **X** transposed $\mathbf{X}^T$

1	1	1	1	0	0	0	0	0
0	0	0	0	1	1	0	0	0
0	0	0	0	0	0	1	1	0
0	0	0	0	0	0	0	0	1
0	1	0	0	0	1	0	1	0
0	0	0	0	0	0	0	0	0
0	0	1	1	0	0	0	0	0

Product of $\mathbf{X}^T\mathbf{X}$

4	0	0	0	1	1	1
0	3	0	0	1	1	0
0	0	2	0	1	0	0
0	0	0	1	0	0	0
1	1	1	0	3	0	0
1	1	0	0	0	2	0
1	0	0	0	0	0	1

Inverse of $\mathbf{X}^T\mathbf{X}$ i.e. $(\mathbf{X}^T\mathbf{X})^{-1}$

.58333	.25000	.16667	.00000	−.33333	−.41667	−.58333
.25000	.58333	.16667	.00000	−.33333	−.41667	−.25000
.16667	.16667	.66667	.00000	−.33333	−.16667	−.16667
.00000	.00000	.00000	1.00000	.00000	.00000	.00000
.33333	−.33333	−.33333	.00000	.66667	.33333	.33333
.41667	−.41667	−.16667	.00000	.33333	.91667	.41667
.58333	−.25000	−.16667	.00000	.33333	.41667	1.58333

Future design $\mathbf{X}_f$

0	1	0	0	0	0	1
0	0	1	0	0	1	0
0	0	1	0	0	0	1
0	0	0	1	1	0	0
0	0	0	1	0	1	0
0	0	0	1	0	0	1

Transpose of Future Design Matrix $\mathbf{X}_f^T$

0	0	0	0	0	0
1	0	0	0	0	0
0	1	1	0	0	0
0	0	0	1	1	1
0	0	0	1	0	0
0	1	0	0	1	0
1	0	1	0	0	1

Product of Future Design $\mathbf{X}_f$ and Inverse of $\mathbf{X}^T\mathbf{X}$

i.e $\mathbf{X}_f\,(\mathbf{X}^T\mathbf{X})^{-1}$

.33333	.33333	.00000	.00000	.00000	.00000	1.33333
.25000	−.25000	.50000	.00000	.00000	.75000	.25000
.41667	−.08333	.50000	.00000	.00000	.25000	1.41667
.33333	−.33333	−.33333	1.00000	.66667	.33333	.33333
.41667	−.41667	−.16667	1.00000	.33333	.91667	.41667
.58333	−.25000	−.16667	1.00000	.33333	.41667	1.58333

Final product ($\mathbf{X}_f\ (\mathbf{X}^T\mathbf{X})^{-1}$) and $\mathbf{X}_f^T$

i.e. $\mathbf{X}_f\ (\mathbf{X}^T\mathbf{X})^{-1}\ \mathbf{X}_f^T$

1.66667	.00000	1.33333	.00000	.00000	1.33333
.00000	1.25000	.75000	.00000	.75000	.25000
1.33333	.75000	1.91667	.00000	.25000	1.41667
.00000	.00000	.00000	1.66667	1.33333	1.33333
.00000	.75000	.25000	1.33333	1.91667	1.41667
1.33333	.25000	1.41667	1.33333	1.41667	2.58333

And finally the data specific Var-Cov matrix is derived from the above values by multiplying by σ^2.

So the first entry is $1.66667 \times .0524^2 = .00457$ etc.

The Variance-Covariance matrix in this case is then

i.e. $\sigma^2\ \mathbf{X}_f\ (\mathbf{X}^T\mathbf{X})^{-1}\ \mathbf{X}_f^T$

.00457	.00000	.00366	.00000	.00000	.00366
.00000	.00343	.00206	.00000	.00206	.00069
.00366	.00206	.00526	.00000	.00069	.00389
.00000	.00000	.00000	.00457	.00366	.00366
.00000	.00206	.00069	.00366	.00526	.00389
.00366	.00069	.00389	.00366	.00389	.00709

Appendix 2

Spreadsheet Regression Output tables

The raw spreadsheet regression output table for the first example (4×4 chain ladder) was

Regression Output:

Constant	0
Std Err of Y Est	.05238
R Squared(Ajd,Raw) .99758	.99919
No. of Observations	10
Degrees of Freedom	3

Coefficient(s)	9.2884	9.5911	9.6924	9.7358	−.4662	−1.801	−2.647
Std Err of Coef.	.0400	.0400	.0428	.0524	.0428	.0502	.0660

This is very typical of all the spreadsheet regression output.

A brief description of this output is given below:

a) Constant (=0)

The spreadsheet regression command usually has an option of either fitting through the origin or calculating a constant. In the case above the model was fitted through the origin so the constant calculated is zero. In the models described in the article a parameter is used in place of this constant as this makes the analysis more convenient. The calculated values will be the same but in the latter case the regression shows the standard error associated with this constant.

b) Std Err of Y Est (0.0524)

This is the estimated standard error of the residuals. It is the square root of the estimated model variance σ^2.

It is in other words the estimate of the standard deviation of the assumed underlying normal error term.

This value plays a very significant role in the estimates of future values and their standard errors.

c) R Squared (Adj, Raw) (0.9976 0.9992)

This is a statistic ranging from 0 to 1 which indicates how much variation in the data is explained by the model. The closer to 1, the more variation explained by the model. The difference in the two values is from a correction for the degrees of freedom.

In crude terms it indicates that the model explains 99.76% of the values, in the log-space.

d) No of Observations (10)

The 4 by 4 triangle contained ten values all of which were used in the fitting process.

e) Degrees of Freedom (3)

The model assumed 7 independent parameters (including the constant) and used 10 observations to estimate these. The difference, (10-7), is the number of degrees of freedom.

Note that in this case there are a lot of parameters in relation to the number of data values in the triangle. This tends to produce a high quality of fit, i.e. a high R^2 but forced adherence to the actual data by incorporating many parameters in the model can lead to a model with poor predictive qualities.

f) Coefficient(s) (9.288 9.591 etc.)

These values are the estimates of the model parameter values. They appear in the order defined by the Design Matrix one for each independent variable.

Least squares are being used to calculate these values and the solution is given by

$(\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}^T\,\mathbf{Y}$ where $\mathbf{Y}$ is the vector of data values.

g) Std. Err of Coef. (0.0400 0.0400 etc...)

These are the estimated standard errors of the coefficient estimates. They are the square roots of the diagonal elements of the variance-covariance matrix of the coefficients

$\sigma^2\,(\mathbf{X}^T\mathbf{X})^{-1}$

Changing values in the data triangle does not affect the design matrix $\mathbf{X}$ and only changes the scalar element or σ^2.

So different data sets result in standard errors of the model coefficients which differ only by a constant factor which is equal to the ratio of the data specific model standard errors or σ's.

<>

[D6]

MEASURING THE VARIABILITY OF CHAIN LADDER RESERVE ESTIMATES

Contributed by T Mack

Abstract

The variability of chain ladder reserve estimates is quantified without assuming any specific claims amount distribution function. This is done by establishing a formula for the so-called standard error which is an estimate for the standard deviation of the outstanding claims reserve. The information necessary for this purpose is extracted only from the usual chain ladder formulae. With the standard error as a tool it is shown how a confidence interval for the outstanding claims reserve and for the ultimate claims amount can be constructed. Moreover, the analysis of the information extracted and of its implications shows when it may be appropriate to apply the chain ladder method and when it may not be.

Note

The original version of this paper was submitted to the prize paper competition "Variability of Loss Reserves" held by the Casualty Actuarial Society and was awarded a joint second prize. The present text differs from that paper in a few changes to the text and a changed and more thorough test procedure in Appendix H. This paper is included in the Claims Reserving Manual with the specific permission of the Casualty Actuarial Society, which otherwise retains ownership and all rights to continue to publish and disseminate this paper anywhere.

Contents

1. Introduction and Overview

The chain ladder method is probably the most popular method for estimating outstanding claims reserves. The main reason for this is its simplicity and the fact that it is distribution-free, that is, it seems to be based on almost no assumptions. In this paper, it will be seen that this impression is wrong and that the chain ladder algorithm has far-reaching implications. These implications also allow it to measure the variability of chain ladder reserve estimates. With the help of this measure it is possible to construct a confidence interval for the estimated ultimate claims amount and for the estimated reserves.

Such a confidence interval is of great interest for the practitioner because the estimated ultimate claims amount can never be an exact forecast of the true ultimate claims amount and therefore a confidence interval is of much greater information value. A confidence interval also allows one to consider business strategy in conjunction with the claims reserving process, using specific confidence probabilities. Moreover, there are many other claims reserving procedures and the results of all these procedures can vary widely. With the help of a confidence interval it can be seen whether the difference between the results of the chain ladder method and any other method is significant or not.

The structure of the paper is as follows. In section 2 a first basic assumption underlying the chain ladder method is derived from the formula used to estimate the ultimate claims amount. In section 3, the comparison of the age-to-age factor formula used by the chain ladder method with other possibilities leads to a second underlying assumption regarding the variance of the claims amounts. Using both of these derived assumptions and a third assumption on the independence of the accident years, it is possible to calculate the so-called standard error of the estimated ultimate claims amount. This is done in section 4, where it is also shown that this standard error is the appropriate measure of variability for the construction of a confidence interval. Section 5 illustrates how any given run-off triangle can be checked using some plots to ascertain whether the assumptions mentioned can be considered to be met. If these plots show that the assumptions do not seem to be met, the chain ladder method should not be applied without adaptation. In section 6 the formulae and two statistical tests (set out in Appendices G and H) are applied to a numerical example. For the sake of comparison, the reserves and standard errors according to a well-known claims reserving software package are also quoted. Complete and detailed proofs of all results and formulae are given in the Appendices A–F.

The proofs are quite long and take up about one fifth of the paper. However, the resulting formula for the standard error is very simple and can be applied directly after reading the basic notations in the first two paragraphs of section 2. In the numerical example, too, the formula for the standard error could be applied immediately to the run-off triangle. Instead, an analysis of whether the chain ladder assumptions are met in this particular case is made first. Because this analysis comprises many tables and plots, the example takes up another two fifths of the paper (including the tests in Appendices G and H).

2. Notation and First Analysis of the Chain Ladder Method

Let C_{ik} denote the accumulated total claims amount of accident year i, $1 \le i \le I$, either paid or incurred up to development year k, $1 \le k \le I$. The values of C_{ik} for $i + k \le I + 1$ are known to us (run-off triangle) and we want to estimate the values of C_{ik} for $i + k > I + 1$, in particular the ultimate claims amount C_{iI} of each accident year $i = 2, ..., I$. Then

$$R_i = C_{iI} - C_{i,I+1-i}$$

is the outstanding claims reserve of accident year i as $C_{i,I+1-i}$ has already been paid or incurred up to now.

The chain ladder method consists of estimating the ultimate claims amounts C_{iI} by

$$(1) \quad \mathbf{C_{iI}} = C_{i,I+1-i} \cdot \mathbf{f_{I+1-i}} \cdot \ldots \cdot \mathbf{f_{I-1}}, \quad 2 \le i \le I$$

where

$$(2) \quad \mathbf{f_k} = \sum_{j=1}^{I-k} C_{j,k+1} \Big/ \sum_{j=1}^{I-k} C_{jk}, \quad 1 \le k \le I-1$$

are the so-called age-to-age factors.

This manner of projecting the known claims amount $C_{i,I+1-i}$ to the ultimate claims amount $\mathbf{C_{iI}}$ uses for all accident years $i \ge I + 1 - k$ the same factor $\mathbf{f_k}$ for the increase of the claims amount from development year k to development year k+1, although the observed individual development factors $C_{i,k+1} / C_{ik}$ of the accident years $i \le I - k$ are usually different from one another and from $\mathbf{f_k}$. This means that each increase from C_{ik} to $C_{i,k+1}$ is considered a random disturbance of an expected increase from C_{ik} to $C_{ik} f_k$ where f_k is an unknown 'true' factor of increase which is the same for all accident years and which is estimated from the available data by $\mathbf{f_k}$.

Consequently, if we imagine to be at the end of development year k we have to consider $C_{i,k+1}, \ldots, C_{iI}$ as random variables whereas the realizations of $C_{i1}, \ldots, C_{ik}$ are known to us and are therefore no longer random variables but scalars. This means that for the purposes of analysis every C_{ik} can be a random variable or a scalar, depending on the development year at the end of which we imagine to be but independently of whether C_{ik} belongs to the known part $i + k \le I + 1$ of the run-off triangle or not. When taking expected values or variances we therefore must always also state the development year at the end of which we imagine to be. This will be done by explicitly indicating those variables C_{ik} whose values are assumed to be known. If nothing is indicated all C_{ik} are assumed to be unknown.

What we said above regarding the increase from C_{ik} to $C_{i,k+1}$ can now be formulated in stochastic terms as follows. The chain ladder method assumes the existence of accident-year-independent factors $f_1, \ldots, f_{I-1}$ such that, given the development $C_{i1}, \ldots, C_{ik}$, the realization of $C_{i,k+1}$ is 'close' to $C_{ik} f_k$, the latter being the expected value of $C_{i,k+1}$ in its mathematical meaning, that is

(3) $E(C_{i,k+1} | C_{i1}, \ldots, C_{ik}) = C_{ik} f_k, \quad 1 \le i \le I, \; 1 \le k \le I-1$

Here to the right of the '$|$' those C_{ik} are listed which are assumed to be known. Mathematically speaking, (3) is a conditional expected value which is just the exact mathematical formulation of the fact that we already know $C_{i1}, \ldots, C_{ik}$, but do not know $C_{i,k+1}$. The same notation is also used for variances since they are specific expectations. The reader who is not familiar with conditional expectations should not refrain from further reading because this terminology is easily understandable and the usual rules for the calculation with expected values also apply to conditional expected values. Any special rule will be indicated wherever it is used.

We want to point out again that the equations (3) constitute an assumption which is not imposed by us but rather implicitly underlies the chain ladder method. This is based on two aspects of the basic chain ladder equation (1). One is the fact that (1) uses the same age-to-age factor $\mathbf{f}_k$ for different accident years i = I + 1 − k, ..., I. Therefore equations (3) also postulate age-to-age parameters f_k which are the same for all accident years. The other is the fact that (1) uses only the most recent observed value $C_{i,I+1-i}$ as basis for the projection to ultimate ignoring on the one hand all amounts $C_{i1}, \ldots, C_{i,I-i}$ observed earlier and on the other hand the fact that $C_{i,I+1-i}$ could substantially deviate from its expected value.

Note that it would easily be possible to also project to ultimate the amounts $C_{i1}, \ldots, C_{i,I-i}$ of the earlier development years with the help of the age-to-age factors $\mathbf{f}_1, \ldots, \mathbf{f}_{I-1}$ and to combine all these projected amounts together with $C_{i,I+1-i}\mathbf{f}_{I+1-i} \cdot \ldots \cdot \mathbf{f}_{I-1}$ into a common estimator for C_{iI}. Moreover, it would also easily be possible to use the values $C_{j,I+1-i}$ of the earlier accident years j < i as additional estimators for $E(C_{i,I+1-i})$ by translating them into accident year i with the help of a measure of volume for each accident year.

These possibilities are all ignored by the chain ladder method which uses $C_{i,I+1-i}$ as the only basis for the projection to ultimate. This means that the chain ladder method implicitly must use an assumption which states that the information contained in $C_{i,I+1-i}$ cannot be augmented by additionally using $C_{i1}, \ldots, C_{i,I-i}$ or $C_{1,I+1-i}, \ldots, C_{i-1,I+1-i}$. This is very well reflected by the equations (3) which state that, given $C_{i1}, \ldots, C_{ik}$, the expected value of $C_{i,k+1}$ only depends on C_{ik}.

Having now formulated this first assumption underlying the chain ladder method we want to emphasize that this is a rather strong assumption which has important consequences and which cannot be taken as met for every run-off triangle. Thus the widespread impression that the chain ladder method would work with almost no assumptions is not justified. In section 5 we will elaborate on the linearity constraint contained in assumption (3). But here we want to point out another consequence of formula (3). We can rewrite (3) in the form

$$E(C_{i,k+1} / C_{ik} \mid C_{i1}, \ldots, C_{ik}) = f_k$$

because C_{ik} is a scalar under the condition that we know $C_{i1}, \ldots, C_{ik}$. This form of (3) shows that the expected value of the individual development factor $C_{i,k+1} / C_{ik}$ equals f_k irrespective of the prior development $C_{i1}, \ldots, C_{ik}$ and especially of the foregoing development factor $C_{ik} / C_{i,k-1}$.

As is shown in Appendix G, this implies that subsequent development factors $C_{ik} / C_{i,k-1}$ and $C_{i,k+1} / C_{ik}$ are uncorrelated. This means that after a rather high value of $C_{ik} / C_{i,k-1}$ the expected size of the next development factor $C_{i,k+1} / C_{ik}$ is the same as after a rather low value of $C_{ik} / C_{i,k-1}$.

We therefore should not apply the chain ladder method to a business where we usually observe a rather small increase $C_{i,k+1} / C_{ik}$ if $C_{ik} / C_{i,k-1}$ is higher than in most other accident years, and vice versa. Appendix G also contains a test procedure to check this for a given run-off triangle.

3. Analysis of the Age-to-Age Factor Formula: the Key to Measuring the Variability

Because of the randomness of all realizations C_{ik} we can not infer the true values of the increase factors $f_1, ..., f_{I-1}$ from the data. They only can be estimated and the chain ladder method calculates estimators $\mathbf{f_1}, ..., \mathbf{f_{I-1}}$ according to formula (2). Among the properties which a good estimator should have, one prominent property is that the estimator should be unbiased, that is its expected value $E(\mathbf{f_k})$ (under the assumption that the whole run-off triangle is not yet known) is equal to the true value f_k, in other words $E(\mathbf{f_k}) = f_k$. Indeed, this is the case here as is shown in Appendix A under the additional assumption that

(4) the variables $\{C_{i1}, ..., C_{iI}\}$ and $\{C_{j1}, ..., C_{jI}\}$ of different accident years $i \neq j$ are independent

Because the chain ladder method neither in (1) nor in (2) takes into account any dependency between the accident years we can conclude that the independence of the accident years is also an implicit assumption of the chain ladder method. We will therefore assume (4) for all further calculations. Assumption (4), too, cannot be taken as being met for every run-off triangle because certain calendar year effects (such as a major change in claims handling or in case reserving or greater changes in the inflation rate) can affect several accident years in the same way and can thus distort the independence. How such a situation can be recognized is shown in Appendix H.

A closer look at formula (2) reveals that

$$\mathbf{f_k} = \frac{\sum_{j=1}^{I-k} C_{j,k+1}}{\sum_{j=1}^{I-k} C_{jk}} = \sum_{j=1}^{I-k} \frac{C_{jk}}{\sum_{j=1}^{I-k} C_{jk}} \cdot \frac{C_{j,k+1}}{C_{jk}}$$

is a weighted average of the observed individual development factors $C_{j,k+1}/C_{jk}$, for $1 \leq j \leq I-k$, where the weights are proportional to C_{jk}. Like $\mathbf{f_k}$ every individual development factor $C_{j,k+1}/C_{jk}$, $1 \leq j \leq I-k$, is also an unbiased estimator of f_k because

$$\begin{aligned}
E(C_{j,k+1}/C_{jk}) &= E(E(C_{j,k+1}/C_{jk} \mid C_{j1}, ..., C_{jk})) && \text{(a)} \\
&= E(E(C_{j,k+1} \mid C_{j1}, ..., C_{jk})/C_{jk}) && \text{(b)} \\
&= E(C_{jk} f_k / C_{jk}) && \text{(c)} \\
&= E(f_k) \\
&= f_k && \text{(d)}
\end{aligned}$$

Here equality (a) holds due to the iterative rule $E(X) = E(E(X|Y))$ for expectations, (b) holds because, given C_{j1} to C_{jk}, C_{jk} is a scalar, (c) holds due to assumption (3) and

(d) holds because f_k is a scalar. (When applying expectations iteratively, e.g. E(E(X|Y)), one first takes the conditional expectation E(X|Y) assuming Y being known and then averages over all possible realizations of Y.)

Therefore the question arises as to why the chain ladder method uses just $\mathbf{f_k}$ as estimator for f_k and not the simple average

$$\frac{1}{I-k}\sum_{j=1}^{I-k} C_{j,k+1}/C_{jk}$$

of the observed development factors which also would be an unbiased estimator as is the case with any weighted average

$$g_k = \sum_{j=1}^{I-k} w_{jk} C_{j,k+1}/C_{jk} \qquad \text{with } \sum_{j=1}^{I-k} w_{jk} = 1$$

of the observed development factors. (Here, w_{jk} must be a scalar if $C_{j1}, \ldots, C_{jk}$ are known.)

Here we recall one of the principles of the theory of point estimation which states that among several unbiased estimators preference should be given to the one with the smallest variance, a principle which is easy to understand. We therefore should choose the weights w_{jk} in such a way that the variance of g_k is minimal. In Appendix B it is shown that this is the case if and only if (for fixed k and all j)

w_{jk} is inversely proportional to $\mathrm{Var}(C_{j,k+1}/C_{jk} \mid C_{j1}, \ldots, C_{jk})$

The fact that the chain ladder estimator $\mathbf{f_k}$ uses weights which are proportional to C_{jk} therefore means that C_{jk} is assumed to be inversely proportional to $\mathrm{Var}(C_{j,k+1}/C_{jk} \mid C_{j1}, \ldots, C_{jk})$, or stated the other way around, that

$$\mathrm{Var}(C_{j,k+1}/C_{jk} \mid C_{j1}, \ldots, C_{jk}) = \alpha_k^2 / C_{jk}$$

with a proportionality constant α_k^2 which may depend on k but not on j and which must be non-negative because variances are always non-negative.

Since here C_{jk} is a scalar and because generally Var(X/c) = $\mathrm{Var}(X)/c^2$ for any scalar c, we can state the above proportionality condition also in the form

(5) $\mathrm{Var}(C_{j,k+1} \mid C_{j1}, \ldots, C_{jk}) = C_{jk}\alpha_k^2, \; 1 \le j \le I, \; 1 \le k \le I-1$

with unknown proportionality constants α_k^2, $1 \le k \le I-1$.

As with assumptions (3) and (4), assumption (5) also has to be considered a basic condition implicitly underlying the chain ladder method. Again, condition (5) cannot a priori be assumed to be met for every run-off triangle. In section 5 we will show how to check a given triangle to see whether (5) can be considered met or not. But before doing so we turn to the most important consequence of (5): together with (3) and (4) it enables us to quantify the uncertainty in the estimation of C_{iI} by $\mathbf{C_{iI}}$.

4. Quantifying the Variability of the Ultimate Claims Amount

The aim of the chain ladder method and of every claims reserving method is the estimation of the ultimate claims amount C_{iI} for the accident years i = 2, ..., I. The chain ladder method does this by formula (1), that is

$$\mathbf{C_{iI}} = C_{i,I+1-i} \cdot \mathbf{f_{I+1-i}} \cdot \ldots \cdot \mathbf{f_{I-1}}$$

This formula yields only a point estimate for C_{iI} which will normally turn out to be more or less wrong, that is there is only a very small probability for C_{iI} being equal to $\mathbf{C_{iI}}$. This probability is even zero if C_{iI} is considered to be a continuous variable. We therefore want to know in addition if the estimator $\mathbf{C_{iI}}$ is at least on average equal to the mean of C_{iI} and how large on average the error is. Precisely speaking we first would like to have the expected values $E(\mathbf{C_{iI}})$ and $E(C_{iI})$, $2 \le i \le I$, being equal. In Appendix C it is shown that this is indeed the case as a consequence of assumptions (3) and (4).

The second thing we want to know is the average distance between the forecast $\mathbf{C_{iI}}$ and the future realization C_{iI}. In Mathematical Statistics it is common to measure such distances by the square of the ordinary Euclidean distance ('quadratic loss function'). This means that one is interested in the size of the so-called mean squared error

$$\text{mse}(\mathbf{C_{iI}}) = E((C_{iI} - \mathbf{C_{iI}})^2 \mid D)$$

where $D = \{C_{ik} \mid i + k \le I + 1\}$ is the set of all data observed so far. It is important to realize that we have to calculate the mean squared error on the condition of knowing all data observed so far because we want to know the error due to <u>future</u> randomness only. If we calculated the unconditional error $E(C_{iI} - \mathbf{C_{iI}})^2$, which due to the iterative rule for expectations is equal to the mean value $E(E((C_{iI} - \mathbf{C_{iI}})^2 \mid D))$ of the conditional mse over all possible data sets D, we also would include all deviations from the data observed so far which obviously makes no sense if we want to establish a confidence interval for C_{iI} on the basis of the given particular run-off triangle D.

The mean squared error is exactly the same concept which also underlies the notion of the variance

$$\text{Var}(X) = E(X - E(X))^2$$

of any random variable X. Var(X) measures the average distance of X from its mean value E(X).

Due to the general rule $E(X-c)^2 = \mathrm{Var}(X) + (E(X)-c)^2$ for any scalar c we have

$$\mathrm{mse}(\mathbf{C_{iI}}) = \mathrm{Var}(C_{iI} \mid D) + (E(C_{iI} \mid D) - \mathbf{C_{iI}})^2$$

because $\mathbf{C_{iI}}$ is a scalar under the condition that all data D are known. This equation shows that the mse is the sum of the pure future random error $\mathrm{Var}(C_{iI} \mid D)$ and of the estimation error which is measured by the squared deviation of the estimate $\mathbf{C_{iI}}$ from its target $E(C_{iI} \mid D)$. On the other hand, the mse does not take into account any future changes in the underlying model, that is future deviations from the assumptions (3), (4) and (5), an extreme example of which was the emergence of asbestos. Modelling such deviations is beyond the scope of this paper.

As is to be expected and can be seen in Appendix D, $\mathrm{mse}(\mathbf{C_{iI}})$ depends on the unknown model parameters f_k and $\alpha_k{}^2$. We therefore must develop an estimator for $\mathrm{mse}(\mathbf{C_{iI}})$ which can be calculated from the known data D only. The square root of such an estimator is usually called '<u>standard error</u>' because it is an estimate of the standard deviation of C_{iI} in cases in which we have to estimate the mean value, too. The standard error $\mathrm{s.e.}(\mathbf{C_{iI}})$ of $\mathbf{C_{iI}}$ is at the same time the standard error $\mathrm{s.e.}(\mathbf{R_i})$ of the reserve estimate

$$\mathbf{R_i} = \mathbf{C_{iI}} - C_{i,I+1-i}$$

of the outstanding claims reserve

$$R_i = C_{iI} - C_{i,I+1-i}$$

because

$$\mathrm{mse}(\mathbf{R_i}) = E((\mathbf{R_i} - R_i)^2 \mid D) = E((\mathbf{C_{iI}} - C_{iI})^2 \mid D) = \mathrm{mse}(\mathbf{C_{iI}})$$

and because the equality of the mean squared errors also implies the equality of the standard errors. This means that

(6) $\quad \mathrm{s.e.}(\mathbf{R_i}) = \mathrm{s.e.}(\mathbf{C_{iI}})$

The derivation of a formula for the standard error $\mathrm{s.e.}(\mathbf{C_{iI}})$ of $\mathbf{C_{iI}}$ turns out to be the most difficult part of this paper; it is done in Appendix D. Fortunately, the resulting formula is simple

(7) $$(\mathrm{s.e.}(\mathbf{C_{iI}}))^2 = \mathbf{C_{iI}}^2 \sum_{k=I+1-i}^{I-1} \frac{\boldsymbol{\alpha}_\mathbf{k}{}^2}{\mathbf{f_k}{}^2} \cdot \left(\frac{1}{\mathbf{C_{ik}}} + \frac{1}{\sum_{j=1}^{I-k} C_{jk}} \right)$$

where

$$(8) \qquad \widehat{\alpha}_k^{\,2} = \frac{1}{I-k-1} \sum_{j=1}^{I-k} C_{jk} \left(\frac{C_{j,k+1}}{C_{jk}} - \widehat{f}_k \right)^2, \; 1 \le k \le I-2$$

is an unbiased estimator of $\alpha_k^{\,2}$ (the unbiasedness being shown in Appendix E) and

$$\widehat{C}_{ik} = C_{i,I+1-i} \cdot \widehat{f}_{I+1-i} \cdot \ldots \cdot \widehat{f}_{k-1}, \; k > I+1-i$$

are the amounts which are automatically obtained if the run-off triangle is completed step by step according to the chain ladder method. In (7), for notational convenience we have also set

$$\widehat{C}_{i,I+1-i} = C_{i,I+1-i}$$

Formula (8) does not yield an estimator for α_{I-1} because it is not possible to estimate the two parameters f_{I-1} and α_{I-1} from the single observation $C_{1,I} / C_{1,I-1}$ between development years I – 1 and I. If $\widehat{f}_{I-1} = 1$ and if the claims development is believed to be finished after I – 1 years we can put $\widehat{\alpha}_{I-1} = 0$. If not, we extrapolate the usually decreasing series $\widehat{\alpha}_1, \widehat{\alpha}_2, \ldots, \widehat{\alpha}_{I-3}, \widehat{\alpha}_{I-2}$ by one additional member, for instance by means of loglinear regression (see the example in section 6) or more simply by requiring that

$$\widehat{\alpha}_{I-3} / \widehat{\alpha}_{I-2} = \widehat{\alpha}_{I-2} / \widehat{\alpha}_{I-1}$$

holds at least as long as $\widehat{\alpha}_{I-3} > \widehat{\alpha}_{I-2}$.

This last possibility leads to

$$(9) \qquad \widehat{\alpha}^{\,2}_{I-1} = \min(\widehat{\alpha}^{\,4}_{I-2} / \widehat{\alpha}^{\,2}_{I-3}, \min(\widehat{\alpha}^{\,2}_{I-3}, \widehat{\alpha}^{\,2}_{I-2}))$$

We now want to establish a confidence interval for our target variables C_{iI} and R_i. Because of the equation

$$C_{iI} = C_{i,I+1-i} + R_i$$

the ultimate claims amount C_{iI} consists of a known part $C_{i,I+1-i}$ and an unknown part R_i. This means that the probability distribution function of C_{iI} (given the observations D which include $C_{i,I+1-i}$) is completely determined by that of R_i. We therefore need to establish a confidence interval for R_i only and can then simply shift it to a confidence interval for C_{iI}.

For this purpose we need to know the distribution function of R_i. Up to now we only have estimates $\mathbf{R_i}$ and $\mathrm{s.e.}(\mathbf{R_i})$ for the mean and the standard deviation of this distribution. If the volume of the outstanding claims is large enough we can, due to the central limit theorem, assume that this distribution function is a Normal distribution with an expected value equal to the point estimate given by $\mathbf{R_i}$ and a standard deviation equal to the standard error $\mathrm{s.e.}(\mathbf{R_i})$. A symmetric 95%-confidence interval for R_i is then given by

$$(\mathbf{R_i} - 2\cdot \mathrm{s.e.}(\mathbf{R_i}), \mathbf{R_i} + 2\cdot \mathrm{s.e.}(\mathbf{R_i}))$$

But the symmetric Normal distribution may not be a good approximation to the true distribution of R_i if this latter distribution is rather skewed. This will especially be the case if $\mathrm{s.e.}(\mathbf{R_i})$ is greater than 50 % of $\mathbf{R_i}$. This can also be seen at the above Normal distribution confidence interval whose lower limit then becomes negative even if a negative reserve is not possible.

In this case it is recommended to use an approach based on the Lognormal distribution. For this purpose we approximate the unknown distribution of R_i by a Lognormal distribution with parameters μ_i and σ_i^2 such that mean values as well as variances of both distributions are equal, so that

$$\exp(\mu_i + \sigma_i^2/2) = \mathbf{R_i}$$

$$\exp(2\mu_i + \sigma_i^2)(\exp(\sigma_i^2) - 1) = (\mathrm{s.e.}(\mathbf{R_i}))^2$$

This leads to

$$(10) \quad \sigma_i^2 = \ln(1 + (\mathrm{s.e.}(\mathbf{R_i}))^2 / \mathbf{R_i}^2)$$

$$\mu_i = \ln(\mathbf{R_i}) - \sigma_i^2/2$$

Now, if we want to estimate the 90th percentile of R_i, for example, we proceed as follows. First we take the 90th percentile of the Standard Normal distribution which is 1.28. Then $\exp(\mu_i + 1.28\sigma_i)$ with μ_i and σ_i^2 according to (10) is the 90th percentile of the Lognormal distribution and therefore also approximately of the distribution of R_i.

For instance, if $\mathrm{s.e.}(\mathbf{R_i})/\mathbf{R_i} = 1$ then $\sigma_i^2 = \ln(2)$ and the 90th percentile is $\exp(\mu_i + 1.28\sigma_i) = \mathbf{R_i}\exp(1.28\sigma_i - \sigma_i^2/2) = \mathbf{R_i}\exp(.719) = 2.05\cdot\mathbf{R_i}$. If we had assumed that R_i has approximately a Normal distribution, we would have obtained in this case $\mathbf{R_i} + 1.28\cdot \mathrm{s.e.}(\mathbf{R_i}) = 2.28\cdot\mathbf{R_i}$ as 90th percentile.

This may come as a surprise since we might have expected that the 90th percentile of a Lognormal distribution always must be higher than that of a Normal distribution with same mean and variance. But there is no general rule, it depends on the percentile chosen and on the size of the ratio $\text{s.e.}(\mathbf{R_i})/\mathbf{R_i}$. The Lognormal approximation only prevents a negative lower confidence limit. In order to set a specific lower confidence limit we choose a suitable percentile, for instance 10%, and proceed analogously as with the 90% before. The question of which confidence probability to choose has to be decided from a business policy point of view. The value of 80% = 90% – 10% taken here must be regarded merely as an example.

We have now shown how to establish confidence limits for every R_i and therefore also for every $C_{iI} = C_{i,I+1-i} + R_i$. We may also be interested in having confidence limits for the overall reserve

$$R = R_2 + \ldots + R_I$$

and the question is whether, in order to estimate the variance of R, we can simply add the squares $(\text{s.e.}(\mathbf{R_i}))^2$ of the individual standard errors as would be the case with standard deviations of independent variables. But unfortunately, whereas the R_i's themselves are independent, the estimators $\mathbf{R_i}$ are not because they are all influenced by the same age-to-age factors $\mathbf{f_k}$, that is the $\mathbf{R_i}$'s are positively correlated. In Appendix F it is shown that the square of the standard error of the overall reserve estimator

$$\mathbf{R} = \mathbf{R_2} + \ldots + \mathbf{R_I}$$

is given by

$$(11)\quad (\text{s.e.}(\mathbf{R}))^2 = \sum_{i=2}^{I} \left\{ (\text{s.e.}(\mathbf{R_i}))^2 + \mathbf{C_{iI}} \left(\sum_{j=i+1}^{I} \mathbf{C_{jI}} \right) \sum_{k=I+1-i}^{I-1} \frac{2\boldsymbol{\alpha}_\mathbf{k}^2 / \mathbf{f_k}^2}{\sum_{n=1}^{I-k} C_{nk}} \right\}$$

Formula (11) can be used to establish a confidence interval for the overall reserve amount R in quite the same way as it was done before for R_i. Before giving a full example of the calculation of the standard error, we will deal in the next section with the problem of how to decide for a given run-off triangle whether the chain ladder assumptions (3) and (5) are met or not.

5. Checking the Chain Ladder Assumptions Against the Data

As has been pointed out, the three basic implicit chain ladder assumptions

(3) $E(C_{i,k+1} | C_{i1}, ..., C_{ik}) = C_{ik} f_k$
(4) Independence of accident years
(5) $Var(C_{i,k+1} | C_{i1}, ..., C_{ik}) = C_{ik} \alpha_k^2$

are not met in every case. In this section we will indicate how these assumptions can be checked for a given run-off triangle. We have already mentioned in section 3 that Appendix H develops a test for calendar year influences which may violate (4). We can therefore concentrate in the following on assumptions (3) and (5).

First, we look at the equations (3) for an arbitrary but fixed k and for i = 1, ..., I. There, the values of C_{ik}, $1 \le i \le I$, are to be considered as given non-random values and equations (3) can be interpreted as an ordinary regression model of the type

$$Y_i = c + x_i b + \varepsilon_i, \quad 1 \le i \le I$$

where c and b are the regression coefficients and ε_i the error term with $E(\varepsilon_i) = 0$, that is $E(Y_i) = c + x_i b$. In our special case, we have c = 0, b = f_k and we have observations of the dependent variable $Y_i = C_{i,k+1}$ at the points $x_i = C_{ik}$ for i = 1, ..., I − k. Therefore, we can estimate the regression coefficient b = f_k by the usual least squares method

$$\sum_{i=1}^{I-k} (C_{i,k+1} - C_{ik} f_k)^2 = \text{minimum}$$

If the derivative of the left hand side with respect to f_k is set to 0 we obtain for the minimizing parameter f_k the solution

$$(12) \quad f_{k0} = \sum_{i=1}^{I-k} C_{ik} C_{i,k+1} \Big/ \sum_{i=1}^{I-k} C_{ik}^2$$

This is not the same estimator for f_k as according to the chain ladder formula (2). We therefore have used an additional index '0' at this new estimator for f_k. We can rewrite f_{k0} as

$$f_{k0} = \sum_{i=1}^{I-k} \frac{C_{ik}^2}{\sum_{i=1}^{I-k} C_{ik}^2} \cdot \frac{C_{i,k+1}}{C_{ik}}$$

which shows that f_{k0} is the $C_{ik}{}^{2}$-weighted average of the individual development factors $C_{i,k+1} / C_{ik}$, whereas the chain ladder estimator $\mathbf{f_k}$ is the C_{ik}-weighted average. In section 3 we saw that these weights are inversely proportional to the underlying variances $\text{Var}(C_{i,k+1} / C_{ik} \mid C_{i1}, \ldots, C_{ik})$.

Correspondingly, the estimator f_{k0} assumes

$$\text{Var}(C_{i,k+1} / C_{ik} \mid C_{i1}, \ldots, C_{ik}) \text{ being proportional to } 1 / C_{ik}{}^{2}$$

or equivalently

$$\text{Var}(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) \text{ being proportional to } 1$$

which means that $\text{Var}(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik})$ is the same for all observations $i = 1, \ldots, I - k$. This is not in agreement with the chain ladder assumption (5).

Here we remember that indeed the least squares method implicitly assumes equal variances $\text{Var}(Y_i) = \text{Var}(\varepsilon_i) = \sigma^2$ for all i. If this assumption is not met, that is if the variances $\text{Var}(Y_i) = \text{Var}(\varepsilon_i)$ depend on i, one should use a weighted least squares approach which consists of minimizing the weighted sum of squares

$$\sum_{i=1}^{I} w_i (Y_i - c - x_i b)^2$$

where the weights w_i are in inverse proportion to $\text{Var}(Y_i)$.

Therefore, in order to be in agreement with the chain ladder variance assumption (5), we should use regression weights w_i which are proportional to $1 / C_{ik}$ (more precisely to $1 / (C_{ik} \alpha_k{}^{2})$, but $\alpha_k{}^{2}$ can be amalgamated with the proportionality constant because k is fixed).

Then minimizing

$$\sum_{i=1}^{I-k} (C_{i,k+1} - C_{ik} f_k)^2 / C_{ik}$$

with respect to f_k yields indeed

$$f_{k1} = \sum_{i=1}^{I-k} C_{i,k+1} / \sum_{i=1}^{I=k} C_{ik}$$

which is identical to the usual chain ladder age-to-age factor $\mathbf{f_k}$.

It is tempting to try another set of weights, namely $1/C_{ik}^{\ 2}$ because then the weighted sum of squares becomes

$$\sum_{i=1}^{I-k} (C_{i,k+1} - C_{ik} f_k)^2 / C_{ik}^{\ 2} = \sum_{i=1}^{I-k} \left(\frac{C_{i,k+1}}{C_{ik}} - f_k \right)^2$$

Here the minimizing procedure yields

$$(13) \quad f_{k2} = \frac{1}{I-k} \sum_{i=1}^{I-k} \frac{C_{i,k+1}}{C_{ik}}$$

which is the ordinary unweighted average of the development factors. The variance assumption corresponding to the weights used is

$$\mathrm{Var}(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) \text{ being proportional to } C_{ik}^{\ 2}$$

or equivalently

$$\mathrm{Var}(C_{i,k+1} / C_{ik} \mid C_{i1}, \ldots, C_{ik}) \text{ being proportional to } 1$$

The benefit of transforming the estimation of the age-to-age factors into the regression framework is the fact that the usual regression analysis instruments are now available to check the underlying assumptions, especially the linearity and the variance assumption. This check is usually done by carefully inspecting plots of the data and of the residuals, as described below.

First, we plot $C_{i,k+1}$ against C_{ik}, i = 1, ..., I – k, in order to see if we really have an approximately linear relationship around a straight line through the origin with slope $\mathbf{f_k} = f_{k1}$. Second, if linearity seems acceptable, we plot the weighted residuals

$$(C_{i,k+1} - C_{ik}\mathbf{f_k}) / \sqrt{C_{ik}}, \quad 1 \le i \le I - k,$$

(whose squares have been minimized) against C_{ik} in order to see if the employed variance assumption really leads to a plot in which the residuals do not show any specific trend but appear purely random. It is recommended to compare all three residual plots (for i = 1, ..., I – k)

Plot 0: $C_{i,k+1} - C_{ik} f_{k0}$ against C_{ik}
Plot 1: $(C_{i,k+1} - C_{ik} f_{k1}) / \sqrt{C_{ik}}$ against C_{ik}
Plot 2: $(C_{i,k+1} - C_{ik} f_{k2}) / C_{ik}$ against C_{ik}

and to find out which one shows the most random behaviour. All this should be done for every development year k for which we have sufficient data points, say at least 6, that is for $k \leq I - 6$.

Some experience with least squares residual plots is useful, especially because in our case we have only very few data points. Consequently, it is not always easy to decide whether a pattern in the residuals is systematic or random. However, if Plot 1 exhibits a non-random pattern, and either Plot 0 or Plot 2 does not, and if this holds true for several values of k, we should seriously consider replacing the chain ladder age-to-age factors $f_{k1} = \mathbf{f_k}$ with f_{k0} or f_{k2} respectively.

The following numerical example will clarify the situation a bit more.

6. Numerical Example

The data for the following example are taken from the 'Historical Loss Development Study', 1991 Edition, published by the Reinsurance Association of America (RAA). There, we find on page 96 the following run-off triangle of Automatic Facultative business in General Liability (excluding Asbestos & Environmental):

	C_{i1}	C_{i2}	C_{i3}	C_{i4}	C_{i5}	C_{i6}	C_{i7}	C_{i8}	C_{i9}	C_{i10}
i=1	5012	8269	10907	11805	13539	16181	18009	18608	18662	18834
i=2	106	4285	5396	10666	13782	15599	15496	16169	16704	
i=3	3410	8992	13873	16141	18735	22214	22863	23466		
i=4	5655	11555	15766	21266	23425	26083	27067			
i=5	1092	9565	15836	22169	25955	26180				
i=6	1513	6445	11702	12935	15852					
i=7	557	4020	10946	12314						
i=8	1351	6947	13112							
i=9	3133	5395								
i=10	2063									

The above figures are cumulative incurred case losses in $1000. We have taken the accident years from 1981 (i=1) to 1990 (i=10) which is enough for the sake of example but does not mean that we believe to have reached the ultimate claims amount after 10 years of development.

We first calculate the age-to-age factors $\mathbf{f_k} = f_{k1}$ according to formula (2). The result is shown in the following table together with the alternative factors f_{k0} according to (12) and f_{k2} according to (13)

	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9
f_{k0}	2.217	1.569	1.261	1.162	1.100	1.041	1.032	1.016	1.009
f_{k1}	2.999	1.624	1.271	1.172	1.113	1.042	1.033	1.017	1.009
f_{k2}	8.206	1.696	1.315	1.183	1.127	1.043	1.034	1.018	1.009

If one has the run-off triangle on a personal computer it is very easy to produce the plots recommended in section 5 because most spreadsheet programs have the facility of plotting X−Y graphs. For every k = 1, ..., 8 we make a plot of the amounts $C_{i,k+1}$ (y-axis) of development year k+1 against the amounts C_{ik} (x-axis) of development year k for i = 1, ..., 10 − k, and draw a straight line through the origin with slope f_{k1}.

The plots for k = 1 to 8 are shown in the upper graphs of Figures 1 to 8, respectively. (All figures are to be found at the end of the paper after the appendices.) The number above each point mark indicates the corresponding accident year. (Note that the point mark at the upper or right hand border line of each graph does not belong to the plotted points $(C_{ik}, C_{i,k+1})$, it has only been used to draw the regression line.) In the lower graph of each of the Figures 1 to 8 the corresponding weighted residuals $(C_{i,k+1} - C_{ik}) / \sqrt{C_{ik}}$ are plotted against C_{ik} for i = 1, ... ,10 – k.

The two plots for k = 1 (Figure 1) clearly show that the regression line does not capture the direction of the data points very well. The line should preferably have a positive intercept on the y-axis and a flatter slope. However, even then we would have a high dispersion. Using the line through the origin we will probably underestimate any future C_{i2} if C_{i1} is less than 2000 and will overestimate it if C_{i1} is more than 4000. Fortunately, in the one relevant case i = 10 we have C_{i1} = 2063 which means that the resulting forecast $C_{10,2} = C_{10,1}\mathbf{f_2}$ = 2063·2.999 = 6187 is within the bulk of the data points plotted. In any case, Figure 1 shows that any forecast of $C_{10,2}$ is associated with a high uncertainty of about ±3000 or almost ±50% of an average-sized C_{i2} which is subsequently even larger when extrapolating to ultimate. If in a future accident year we have a value C_{i1} outside the interval (2000, 4000) it is reasonable to introduce an additional parameter by fitting a regression line with positive intercept to the data and using it for the projection to C_{i2}. Such a procedure of employing an additional parameter is acceptable between the first two development years in which we have the highest number of data points of all years.

The two plots for k = 2 (Figure 2) are more satisfactory. The data show a clear trend along the regression line and quite random residuals. The same holds for the two plots for k = 4 (Figure 4). In addition, for both k = 2 and k = 4 a weighted linear regression including a parameter for intercept would yield a value of the intercept which is not significantly different from zero. The plots for k = 3 (Figure 3) seem to show a curvature to the left but because of the few data points we can hope that this is incidental. Moreover, the plots for k = 5 have a certain curvature to the right such that we can hope that the two curvatures offset each other. The plots for k = 6, 7 and 8 are quite satisfactory. The trends in the residuals for k = 7 and 8 have no significance in view of the very few data points.

We need not look at the regression lines with slopes f_{k0} or f_{k2} as these slopes are very close to $\mathbf{f_k}$ (except for k=1). But we should look at the corresponding plots of weighted residuals in order to see whether they appear more satisfactory than the previous ones. (Note that due to the different weights the residuals will be different even if the slopes are equal.) The residual plots for f_{k0} and k = 1 to 4 are shown in Figures 9 and 10. Those for f_{k2} and k = 1 to 4 are shown in Figures 11 and 12. In the residual plot for $f_{1,0}$ (Figure 9, upper graph) the point furthest to the left is not an outlier as it is in the plots for $f_{1,1} = \mathbf{f_1}$ (Figure 1, lower graph) and $f_{1,2}$ (Figure 11, upper graph).

But with all three residual plots for k=1 the main problem is the missing intercept of the regression line which leads to a decreasing trend in the residuals. Therefore the improvement of the outlier is of secondary importance. For k = 2 the three residuals plots do not show any major differences between each other. The same holds for k = 3 and 4. The residual plots for k = 5 to 8 are not important because of the small number of data points. Altogether, we decide to keep the usual chain ladder method, that is the age-to-age factors $\mathbf{f_k} = f_{k,1}$, because the alternatives $f_{k,0}$ or $f_{k,2}$ do not lead to a clear improvement.

Next, we can carry through the tests for calendar year influences (see Appendix H) and for correlations between subsequent development factors (see Appendix G). For our example neither test leads to a rejection of the underlying assumption as is shown in the appendices mentioned.

Having now finished all preliminary analyses we calculate the estimated ultimate claims amounts $\mathbf{C_{iI}}$ according to formula (1), the reserves $\mathbf{R_i} = \mathbf{C_{iI}} - C_{i,I+1-i}$ and its standard errors (7). For the standard errors we need the estimated values of $\boldsymbol{\alpha}_k^2$ which according to formula (8) are given by

k	1	2	3	4	5	6	7	8	9
$\boldsymbol{\alpha}_k^2$	27883	1109	691	61.2	119	40.8	1.34	7.88	

A plot of $\ln(\boldsymbol{\alpha}_k^2)$ against k is given in Figure 13 and shows that there indeed seems to be a linear relationship which can be used to extrapolate $\ln(\boldsymbol{\alpha}_9^2)$. This yields $\boldsymbol{\alpha}_9^2$ = exp(–.44) = .64. But we use formula (9) which is more easily programmable and in the present case is a bit more on the safe side: it leads to $\boldsymbol{\alpha}_9^2$ = 1.34. Using formula (11) for s.e.(**R**) as well we finally obtain

	$\mathbf{C_{i,10}}$	$\mathbf{R_i}$	$\mathrm{s.e.}(\mathbf{C_{i,10}}) = \mathrm{s.e.}(\mathbf{R_i})$	$\mathrm{s.e.}(\mathbf{R_i})/\mathbf{R_i}$
i=2	16858	154	206	134%
i=3	24083	617	623	101%
i=4	28703	1636	747	46%
i=5	28927	2747	1469	53%
i=6	19501	3649	2002	55%
i=7	17749	5435	2209	41%
i=8	24019	10907	5358	49%
i=9	16045	10650	6333	59%
i=10	18402	16339	24566	150%
Overall		52135	26909	52%

(The numbers in the 'Overall'-row are **R**, s.e.(**R**) and s.e.(**R**)/**R**.) For i = 2, 3 and 10 the percentage standard error (last column) is more than 100% of the estimated reserve $\mathbf{R_i}$. For i = 2 and 3 this is due to the small amount of the corresponding reserve and is not important because the absolute amounts of the standard errors are rather small. But the standard error of 150% for the most recent accident year i = 10 might lead to some concern in practice. The main reason for this high standard error is the high uncertainty of forecasting next year's value $C_{10,2}$ as was seen when examining the plot of C_{i2} against C_{i1}. Thus, one year later we will very likely be able to give a much more precise forecast of $C_{10,10}$.

Because all standard errors are close to or above 50% we use the Lognormal distribution in all years for the calculation of confidence intervals. We first calculate the upper 90%-confidence limit (or with any other chosen percentage) for the overall outstanding claims reserve R. Denoting by μ and σ^2 the parameters of the Lognormal distribution approximating the distribution of R and using s.e.($\mathbf{R}$)/$\mathbf{R}$ = .52 we have σ^2 = .236 (cf. (10)) and, in the same way as in section 4, the 90th percentile is $\exp(\mu + 1.28\sigma) = \mathbf{R}\cdot\exp(1.28\sigma - \sigma^2/2) = 1.655\cdot\mathbf{R} = 86298$.

Now we allocate this overall amount to the accident years i = 2,..., 10 in such a way that we reach the same level of confidence for every accident year. Each level of confidence corresponds to a certain percentile t of the Standard Normal distribution and — according to section 4 — the corresponding percentile of the distribution of R_i is $\mathbf{R_i}\exp(t\sigma_i - \sigma_i^2/2)$ with $\sigma_i^2 = \ln(1 + (\mathrm{s.e.}(\mathbf{R_i}))^2/\mathbf{R_i}^2)$. We therefore only have to choose t in such a way that

$$\sum_{i=2}^{I}\mathbf{R_i}.\exp(t\sigma_i - \sigma_i^2/2) = 86298$$

This can easily be solved with the help of spreadsheet software (for example. by trial and error, or by using a “Solver”) and yields t = 1.13208 which corresponds to the 87th percentile per accident year and leads to the following distribution of the overall amount 86298:

	$\mathbf{R_i}$	$\mathrm{s.e.}(\mathbf{R_i})/\mathbf{R_i}$	σ_i^2	upper confidence limit $\mathbf{R_i}\exp(t\sigma_i - \sigma_i^2/2)$
i=2	154	1.34	1.028	290
i=3	617	1.01	.703	1122
i=4	1636	.46	.189	2436
i=5	2747	.53	.252	4274
i=6	3649	.55	.263	5718
i=7	5435	.41	.153	7839
i=8	10907	.49	.216	16571
i=9	10650	.59	.303	17066
i=10	16339	1.50	1.182	30981
Total	52135			86298

In order to arrive at the lower confidence limits we proceed completely analogously. The 10th percentile, for instance, of the total outstanding claims amount is $\mathbf{R}\cdot\exp(-1.28\sigma - \sigma^2/2) = .477\cdot\mathbf{R} = 24871$. The distribution of this amount over the individual accident years is made as before and leads to a value of t = –.8211 which corresponds to the 21st percentile. This means that a 87% – 21% = 66% confidence interval for each accident year leads to a 90% – 10% = 80% confidence interval for the overall reserve amount. In the following table, the confidence intervals thus obtained for R_i are already shifted (by adding $C_{i,I+1-i}$) to confidence intervals for the ultimate claims amounts C_{iI} (for instance, the upper limit 16994 for i=2 has been obtained by adding $C_{2,9}$ = 16704 and 290 from the preceding table):

	$\mathbf{C_{i,10}}$	confidence intervals for 80% prob. overall	empirical limits
i=2	16858	(16744, 16994)	(16858, 16858)
i=3	24083	(23684, 24588)	(23751, 24466)
i=4	28703	(28108, 29503)	(28118, 29446)
i=5	28927	(27784, 30454)	(27017, 31699)
i=6	19501	(17952, 21570)	(16501, 22939)
i=7	17749	(15966, 20153)	(14119, 23025)
i=8	24019	(19795, 29683)	(16272, 48462)
i=9	16045	(11221, 22461)	(8431, 54294)
i=10	18402	(5769, 33044)	(5319, 839271)

The column "empirical limits" contains the minimum and maximum size of the ultimate claims amount resulting if, in formula (1), each age-to-age factor $\mathbf{f_k}$ is replaced with the minimum (or maximum) individual development factor observed so far. These factors are defined by

$$f_{k,min} = \min \{C_{i,k+1} / C_{ik} \mid 1 \le i \le I - k\}$$
$$f_{k,max} = \max \{C_{i,k+1} / C_{i,k} \mid 1 \le i \le I - k\}$$

and can be taken from the table of all development factors which can be found in Appendices G and H. They are

	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9
$f_{k,min}$	1.650	1.259	1.082	1.102	1.009	0.993	1.026	1.003	1.009
$f_{k,max}$	40.425	2.723	1.977	1.292	1.195	1.113	1.043	1.033	1.009

In comparison with the confidence intervals, these empirical limits are narrower in the earlier accident years $i \le 4$ and wider in the more recent accident years $i \ge 5$. This was to be expected because the small number of development factors observed between the late development years only leads to a rather small variation between the minimum and maximum factors. Therefore these empirical limits correspond to a confidence probability which is rather small in the early accident years and becomes larger and larger towards the recent accident years. Thus, this empirical approach to establishing confidence limits does not seem to be reasonable.

If we used the Normal distribution instead of the Lognormal we would obtain a 90th percentile of $\mathbf{R} + 1.28\cdot\mathbf{R}\cdot(\text{s.e.}(\mathbf{R})/\mathbf{R}) = 1.661\cdot\mathbf{R}$ (which is almost the same as the $1.655\cdot\mathbf{R}$ with the Lognormal) and a 10th percentile of $\mathbf{R} - 1.28\cdot\mathbf{R}\cdot(\text{s.e.}(\mathbf{R})/\mathbf{R}) = .34\cdot\mathbf{R}$

(which is lower than the .477·**R** with the Lognormal). Also, the allocation to the accident years would be different.

Finally, we compare the standard errors obtained to the output of the claims reserving software package ICRFS by Ben Zehnwirth.

This package is a modelling framework in which the user can specify his own model within a large class of models. But it also contains some predefined models, inter alia also a 'chain ladder model'. But this is not the usual chain ladder method, instead, it is a log-linearized approximation of it. This is very similar to the model described in the paper, Regression Model Based on Log-Incremental Payments by S.Christofides, see Section D5, Volume 2 of the Claims Reserving Manual.

The slight difference in the results is due to a different estimator for the variance, σ^2. Therefore, the estimates of the outstanding claims amounts differ from those obtained here with the usual chain ladder method. Moreover, it works with the logarithms of the incremental amounts $C_{i,k+1} - C_{ik}$ and one must therefore eliminate the negative increment $C_{2,7} - C_{2,6}$. In addition, $C_{2,1}$ was identified as an outlier and was eliminated. Then the ICRFS results were quite similar to the chain ladder results as can be seen in the following table

	est. outst. claims amount $\mathbf{R_i}$		standard error	
	chain ladder	ICRFS	chain ladder	ICRFS
i=2	154	387	206	528
i=3	617	674	623	624
i=4	1636	1993	747	1435
i=5	2747	2602	1469	1688
i=6	3649	4097	2002	2476
i=7	5435	5188	2209	3156
i=8	10907	12174	5358	7685
i=9	10650	15343	6333	11158
i=10	16339	27575	24566	28333
Overall	52135	70032	26909	33637

Even though the reserves $\mathbf{R_i}$ for i=9 and i=10 as well as the overall reserve **R** differ considerably they are all within one standard error and therefore not significantly different. But it should be remarked that this manner of using ICRFS is not intended by Zehnwirth because any initial model should be further adjusted according to the indications and plots given by the program. In this particular case there were strong indications for developing the model further but then one would have to give up the 'chain ladder model'.

7. Final Remarks

This paper develops a complete methodology of how to attack the claims reserving task in a statistically sound manner on the basis of the well-known and simple chain ladder method. However, the well-known weak points of the chain ladder method should not be concealed. These are the fact that the estimators of the last two or three factors f_I, f_{I-1}, f_{I-2} rely on very few observations and the fact that the known claims amount C_{I1} of the last accident year (sometimes $C_{I-1,2}$, too) forms a very uncertain basis for the projection to ultimate.

This is most clearly seen if C_{I1} happens to be 0: Then we have $\mathbf{C_{iI}} = 0$, $\mathbf{R_I} = 0$ and $\text{s.e.}(\mathbf{R_I}) = 0$ which obviously makes no sense. (Note that this weakness can often be overcome by translating and mixing the amounts C_{i1} of earlier accident years i < I into accident year I with the help of a measure of volume for each accident year.)

Thus, even if the statistical instruments developed do not reject the applicability of the chain ladder method, the result must be judged by an actuary and/or underwriter who knows the business under consideration. Even then, unexpected future changes can make all estimations obsolete. But for the many normal cases it is good to have a sound and simple method. Simple methods have the disadvantage of not capturing all aspects of reality but have the advantage that the user is in a position to know exactly how the method works and where its weaknesses are. Moreover, a simple method can be explained to non-actuaries in more detail. These are important advantages of simple models over more sophisticated ones.

Appendix A: Unbiasedness of Age-to-Age Factors

Proposition: Under the assumptions

(3) There are unknown constants $f_1, \ldots, f_{I-1}$ with

$$E(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) = C_{ik} f_k, \qquad 1 \le i \le I,\ 1 \le k \le I-1$$

(4) The variables $\{C_{i1}, \ldots, C_{iI}\}$ and $\{C_{j1}, \ldots, C_{jI}\}$ of different accident years $i \ne j$ are independent

the age-to-age factors $\mathbf{f_1}, \ldots, \mathbf{f_{I-1}}$ defined by

(2) $$f_k = \sum_{j=1}^{I-k} C_{j,k+1} / \sum_{j=1}^{I-k} C_{jk}, \qquad 1 \le k \le I-1$$

are unbiased, that is we have $E(\mathbf{f_k}) = f_k, \qquad 1 \le k \le I-1$

Proof: Because of the iterative rule for expectations we have

(A1) $$E(\mathbf{f_k}) = E(E(\mathbf{f_k} \mid B_k))$$

for any set B_k of variables C_{ij} assumed to be known. We take

$$B_k = \{C_{ij} \mid i + j \le I + 1, j \le k\}, \qquad 1 \le k \le I\text{-}1$$

According to the definition (2) of $\mathbf{f_k}$ and because C_{jk}, $1 \le j \le I-k$, is contained in B_k and therefore has to be treated as scalar, we have

(A2) $$E(\mathbf{f_k} \mid B_k) = \sum_{j=1}^{I-k} E(C_{j,k+1} \mid B_k) / \sum_{j=1}^{I-k} C_{jk}$$

Because of the independence assumption (4) conditions relating to accident years other than that of $C_{j,k+1}$ can be omitted, that is we get

(A3) $$E(C_{j,k+1} \mid B_k) = E(C_{j,k+1} \mid C_{j1}, \ldots, C_{jk}) = C_{jk} f_k$$

using assumption (3) as well.

Inserting (A3) into (A2) yields

$$\text{(A4)}\quad E(\mathbf{f_k} \mid B_k) = \sum_{j=1}^{I-k} C_{jk} f_k \Big/ \sum_{j=1}^{I-k} C_{jk} = f_k$$

Finally, (A1) and (A4) yield $E(\mathbf{f_k}) = E(f_k) = f_k$ because f_k is a scalar.

Appendix B: Minimizing the Variance of Independent Estimators

Proposition: Let $T_1, \ldots, T_I$ be independent unbiased estimators of a parameter t, that is with

$$E(T_i) = t, \qquad 1 \le i \le I$$

then the variance of a linear combination

$$T = \sum_{i=1}^{I} w_i T_i$$

under the constraint

$$\text{(B1)} \quad \sum_{i=1}^{I} w_i = 1$$

(which guarantees E(T) = t) is minimal iff the coefficients w_i are inversely proportional to $\mathrm{Var}(T_i)$, that is iff

$$w_i = c / \mathrm{Var}(T_i), \qquad 1 \le i \le I$$

Proof: We have to minimize

$$\mathrm{Var}(T) = \sum_{i=1}^{I} w_i^2 \mathrm{Var}(T_i)$$

(due to the independence of $T_1, \ldots, T_I$) with respect to w_i under the constraint (B1).

A necessary condition for an extremum is that the derivatives of the Lagrangian are zero, that is

$$\text{(B2)} \quad \frac{\partial}{\partial w_i}\left(\sum_{i=1}^{I} w_i^2 \, \mathrm{Var}(T_i) + \lambda \left(1 - \sum_{i=1}^{I} w_i \right) \right) = 0, \qquad 1 \le i \le I$$

with a constant multiplier λ whose value can be determined by additionally using (B1).

(B2) yields

$$2w_i \mathrm{Var}(T_i) - \lambda = 0$$

or

$$w_i = \lambda / (2 \cdot \mathrm{Var}(T_i))$$

These weights w_i indeed lead to a minimum as can be seen by calculating the extremal value of Var(T) and applying Schwarz's inequality.

Corollary: In the chain ladder case we have estimators $T_i = C_{i,k+1} / C_{ik}$, $1 \le i \le I - k$, for f_k where the variables of the set

$$A_k = \bigcup_{i=1}^{I-k} \{C_{i1}, \ldots, C_{ik}\}$$

of the corresponding accident years i = 1, ..., I – k up to development year k are considered to be given. We therefore want to minimize the conditional variance

$$\mathrm{Var}\left(\sum_{i=1}^{I-k} w_i T_i \,\middle|\, A_k \right)$$

From the above proof it is clear that the minimizing weights should be inversely proportional to $\mathrm{Var}(T_i | A_k)$. Because of the independence (4) of the accident years, conditions relating to accident years other than that of $T_i = C_{i,k+1} / C_{ik}$ can be omitted. We therefore have

$$\mathrm{Var}(T_i | A_k) = \mathrm{Var}(C_{i,k+1} / C_{ik} | C_{i1}, \ldots, C_{ik})$$

and arrive at the result that the minimizing weights should be inversely proportional to

$$\mathrm{Var}(C_{i,k+1} / C_{ik} | C_{i1}, \ldots, C_{ik})$$

Appendix C: Unbiasedness of the Estimated Ultimate Claims Amount

Proposition: Under the assumptions

(3) There are unknown constants $f_1, \ldots, f_{I-1}$ with

$$E(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) = C_{ik} f_k, \qquad 1 \le i \le I,\ 1 \le k \le I\text{-}1$$

(4) The variables $\{C_{i1}, \ldots, C_{iI}\}$ and $\{C_{j1}, \ldots, C_{jI}\}$ of different accident years $i \neq j$ are independent

the expected values of the estimator

(1) $\mathbf{C_{iI}} = C_{i,I+1-i}\mathbf{f_{i+1-i}} \cdot \ldots \cdot \mathbf{f_{I-1}}$

for the ultimate claims amount and of the true ultimate claims amount C_{iI} are equal, that is we have $E(\mathbf{C_{iI}}) = E(C_{iI})$, $2 \le i \le I$.

Proof: We first show that the age-to-age factors $\mathbf{f_k}$ are uncorrelated. With the same set

$$B_k = \{ C_{ij} \mid i + j \le I + 1, j \le k \}, \qquad 1 \le k \le I - 1$$

of variables assumed to be known as in Appendix A we have for $j < k$

$$\begin{aligned} E(\mathbf{f_j}\mathbf{f_k}) &= E(E(\mathbf{f_j}\mathbf{f_k} \mid B_k)) && \text{(a)} \\ &= E(\mathbf{f_j} E(\mathbf{f_k} \mid B_k)) && \text{(b)} \\ &= E(\mathbf{f_j} f_k) && \text{(c)} \\ &= E(\mathbf{f_j}) f_k && \text{(d)} \\ &= f_j f_k && \text{(e)} \end{aligned}$$

Here (a) holds because of the iterative rule for expectations, (b) holds because $\mathbf{f_j}$ is a scalar for B_k given and for $j < k$, (c) holds due to (A4), (d) holds because f_k is a scalar and (e) was shown in Appendix A.

This result can easily be extended to arbitrary products of different $\mathbf{f_k}$'s, that is we have

(C1) $E(\mathbf{f_{I+1-i}} \cdot \ldots \cdot \mathbf{f_{I-1}}) = f_{i+1-i} \cdot \ldots \cdot f_{I-1}$

This yields

$$
\begin{aligned}
E(\mathbf{C_{iI}}) &= E(E(\mathbf{C_{iI}} \mid C_{i1}, \ldots, C_{i,I+1-i})) && \text{(a)}\\
&= E(E(C_{i,I+1-i}\mathbf{f_{I+1-i}} \cdot\ldots\cdot \mathbf{f_{I-1}} \mid C_{i1}, \ldots, C_{i,I+1-i})) && \text{(b)}\\
&= E(C_{i,I+1-i}E(\mathbf{f_{I+1-i}} \cdot\ldots\cdot \mathbf{f_{I-1}} \mid C_{i1}, \ldots, C_{i,I+1-i})) && \text{(c)}\\
&= E(C_{i,I+1-i}E(\mathbf{f_{I+1-i}} \cdot\ldots\cdot \mathbf{f_{I-1}})) && \text{(d)}\\
&= E(C_{i,I+1-i}) \cdot E(\mathbf{f_{I+1-i}} \cdot\ldots\cdot \mathbf{f_{I-1}}) && \text{(e)}\\
&= E(C_{i,I+1-i}) \cdot f_{I+1-i} \cdot\ldots\cdot f_{I-1} && \text{(f)}
\end{aligned}
$$

Here (a) holds because of the iterative rule for expectations, (b) holds because of the definition (1) of $\mathbf{C_{iI}}$, (c) holds because $C_{i,I+1-i}$ is a scalar under the stated condition, (d) holds because conditions which are independent from the conditioned variable $\mathbf{f_{I+1-i}} \cdot\ldots\cdot \mathbf{f_{I-1}}$ can be omitted (observe assumption (4) and the fact that $\mathbf{f_{I+1-i}}, \ldots, \mathbf{f_{I-1}}$ only depend on variables of accident years ≤ i), (e) holds because $E(\mathbf{f_{I+1-i}}, \ldots, \mathbf{f_{I-1}})$ is a scalar and (f) holds because of (C1).

Finally, repeated application of the iterative rule for expectations and of assumption (3) yields for the expected value of the true reserve C_{iI}

$$
\begin{aligned}
E(C_{iI}) &= E(E(C_{iI} \mid C_{i1}, \ldots, C_{i,I-1}))\\
&= E(C_{i,I-1}f_{I-1})\\
&= E(C_{i,I-1})f_{I-1}\\
&= E(E(C_{i,I-1} \mid C_{i1}, \ldots, C_{I-2}))f_{I-1}\\
&= E(C_{i,I-2}f_{I-2})f_{I-1}\\
&= E(C_{i,I-2})f_{I-2}f_{I-1}\\
&= \text{and so on}\\
&= E(C_{i,I+1-i})f_{I+1-i} \cdot\ldots\cdot f_{I-1}\\
&= E(\mathbf{C_{iI}})
\end{aligned}
$$

Appendix D: Calculation of the Standard Error of $\mathbf{C_{iI}}$

Proposition: Under the assumptions

(3) There are unknown constants $f_1, \ldots, f_{I-1}$ with

$$E(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) = C_{ik} f_k, \qquad 1 \le i \le I,\ 1 \le k \le I-1$$

(4) The variables $\{C_{i1}, \ldots, C_{iI}\}$ and $\{C_{j1}, \ldots, C_{jI}\}$ of different accident years $i \neq j$ are independent

(5) There are unknown constants $\alpha_1, \ldots, \alpha_{I-1}$ with

$$\mathrm{Var}(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) = C_{ik} \alpha_k^2, \qquad 1 \le i \le I,\ 1 \le k \le I-1$$

the standard error $\mathrm{s.e.}(\mathbf{C_{iI}})$ of the estimated ultimate claims amount $\mathbf{C_{iI}} = C_{i,I+1-i} \mathbf{f_{I+1-i}} \cdot \ldots \cdot \mathbf{f_{I-1}}$ is given by the formula

$$(\mathrm{s.e.}(\mathbf{C_{iI}}))^2 = \mathbf{C_{iI}}^2 \sum_{k=I+1-i}^{I-1} \frac{\boldsymbol{\alpha}_k^2}{\mathbf{f}_k^2} \left(\frac{1}{\mathbf{C_{ik}}} + \frac{1}{\sum_{j=1}^{I-k} C_{jk}} \right)$$

where $\mathbf{C_{ik}} = C_{i,I+1-i} \mathbf{f_{I+1-i}} \ldots \mathbf{f_{k-1}}$, $k > I + 1 - i$, are the estimated values of the future $C_{i,k}$ and $\mathbf{C_{i,I+1-i}} = C_{i,I+1-i}$.

Proof: As stated in section 4, the standard error is the square root of an estimator of $\mathrm{mse}(\mathbf{C_{iI}})$ and we have also seen that

(D1) $\mathrm{mse}(\mathbf{C_{iI}}) = \mathrm{Var}(C_{iI} \mid D) + (E(C_{iI} \mid D) - \mathbf{C_{iI}})^2$

In the following, we use the abbreviations

$$E_i(X) = E(X \mid C_{i1}, \ldots, C_{i,I+1-i})$$
$$\mathrm{Var}_i(X) = \mathrm{Var}(X \mid C_{i1}, \ldots, C_{i,I+1-i})$$

Because of the independence of the accident years we can omit in (D1) that part of the condition $D = \{C_{ik} \mid i + k \le I + 1\}$ which is independent from C_{iI}, that is we can write

(D2) $\mathrm{mse}(\mathbf{C_{iI}}) = \mathrm{Var}_i(C_{iI}) + (E_i(C_{iI}) - \mathbf{C_{iI}})^2$

We first consider $\mathrm{Var}_i(C_{iI})$. Because of the general rule Var(X) = $E(X^2) - (E(X))^2$ we have

$$\text{(D3)}\quad \mathrm{Var}_i(C_{iI}) = E_i(C_{iI}^2) - (E_i(C_{iI}))^2$$

For the calculation of $E_i(C_{iI})$ we use the fact that for $k \geq I + 1 - i$

$$\begin{aligned}\text{(D4)}\quad E_i(C_{i,k+1}) &= E_i(E(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik})) \\ &= E_i(C_{ik} f_k) \\ &= E_i(C_{ik}) f_k\end{aligned}$$

Here, we have used the iterative rule for expectations in its general form $E(X|Z) = (E(X|Y)|Z)$ for $\{Y\} \supset \{Z\}$ (mostly $\{Z\}$ is the empty set). By successively applying (D4) we obtain for $k \geq I + 1 - i$

$$\begin{aligned}\text{(D5)}\quad E_i(C_{i,k+1}) &= E_i(C_{i,I+1-i}) f_{I+1-i} \cdot \ldots \cdot f_k \\ &= C_{i,I+1-i} f_{I+1-i} \cdot \ldots \cdot f_k\end{aligned}$$

because $C_{i,I+1-i}$ is a scalar under the condition 'i'.

For the calculation of the first term $E_i(C_{iI}^2)$ of (D3) we use the fact that for $k \geq I + 1 - i$

$$\begin{aligned}\text{(D6)}\quad E_i(C_{i,k+1}^2) &= E_i(E(C_{i,k+1}^2) \mid C_{i1}, \ldots, C_{ik}) && \text{(a)} \\ &= E_i(\mathrm{Var}(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) + (E(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}))^2) && \text{(b)} \\ &= E_i(C_{ik}\alpha_k^2 + (C_{ik} f_k)^2) && \text{(c)} \\ &= E_i(C_{ik})\alpha_k^2 + E_i(C_{ik}^2) f_k^2\end{aligned}$$

Here, (a) holds due to the iterative rule for expectations, (b) due to the rule $E(X^2) = \mathrm{Var}(X) + (E(X))^2$ and (c) holds due to (3) and (5). Now, we apply (D6) and (D5) successively to get

$$\begin{aligned}\text{(D7)}\quad E_i(C_{iI}^2) &= E_i(C_{i,I-1})\alpha_{I-1}^2 + E_i(C_{i,I-1}^2) f_{I-1}^2 && \text{(D6)} \\ &= C_{i,I+1-i} f_{I+1-i} \cdots f_{I-2} \alpha_{I-1}^2 + && \text{(D5)} \\ &\quad + E_i(C_{i,I-2})\alpha_{I-2}^2 f_{I-1}^2 + && \text{(D6)} \\ &\quad + E_i(C_{i,I-2}^2) f_{I-2}^2 f_{I-1}^2\end{aligned}$$

$$
\begin{aligned}
&= C_{i,I+1-i} f_{I+1-i} \cdots f_{I-2} \alpha_{I-1}{}^2 + \\
&\quad + C_{i,I+1-i} f_{I+1-i} \cdots f_{I-3} \alpha_{I-2}{}^2 f_{I-1}{}^2 + \qquad \text{(D5)} \\
&\quad + E_i(C_{i,I-3}) \alpha_{I-3}{}^2 f_{I-2}{}^2 f_{I-1}{}^2 + \qquad \text{(D6)} \\
&\quad + E_i(C_{i,I-3}{}^2) f_{I-3}{}^2 f_{I-2}{}^2 f_{I-1}{}^2
\end{aligned}
$$

= and so on

$$
= C_{i,I+1-i} \sum_{k=I+1-i}^{I-1} f_{I+1-i} \cdots f_{k-1} \alpha_k{}^2 f_{k+1}{}^2 \cdots f_{I-1}{}^2
$$

$$
+ C_{i,I+1-i}{}^2 f_{I+1-i}{}^2 \cdot \ldots \cdot f_{I-1}{}^2
$$

where in the last step we have used $E_i(C_{i,I+1-i}) = C_{i,I+1-i}$ and $E_i(C_{i,I+1-i}{}^2) = C_{i,I+1-i}{}^2$ because under the condition 'i' $C_{i,I+1-i}$ is a scalar. Due to (D5) we have

(D8) $(E_i(C_{iI}))^2 = C_{i,I+1-i}{}^2 f_{I+1-i}{}^2 \cdot \ldots \cdot f_{I-1}{}^2$

Inserting (D7) and (D8) into (D3) yields

(D9) $\mathrm{Var}_i(C_{iI}) = C_{i,I+1-i} \sum_{k=I+1-i}^{I-1} f_{I+1-i} \cdots f_{k-1} \alpha_k{}^2 f_{k+1}{}^2 \cdots f_{I-1}{}^2$

We estimate this first summand of $\mathrm{mse}(\mathbf{C_{iI}})$ by replacing the unknown parameters $f_k, \alpha_k{}^2$ with their unbiased estimators $\mathbf{f_k}$ and $\boldsymbol{\alpha}_\mathbf{k}{}^2$, that is by

(D10) $$C_{i,I+1-i} \sum_{k=I+1-i}^{I-1} \mathbf{f_{I+1-i}} \cdots \mathbf{f_{k-1}} \cdot \boldsymbol{\alpha}_\mathbf{k}{}^2 \cdot \mathbf{f_{k+1}}{}^2 \cdots \mathbf{f_{I-1}}{}^2$$

$$
= C_{i,I+1-i}{}^2 \mathbf{f_{I+1-i}}{}^2 \cdots \mathbf{f_{I-1}}{}^2 \sum_{k=I+1-i}^{I-1} \frac{\boldsymbol{\alpha}_\mathbf{k}{}^2 / \mathbf{f_k}{}^2}{C_{i,I+1-i} \mathbf{f_{I+1-i}} \cdots \mathbf{f_{k-1}}}
$$

$$
= \mathbf{C_{iI}}{}^2 \sum_{k=I+1-i}^{I-1} \frac{\boldsymbol{\alpha}_\mathbf{k}{}^2 / \mathbf{f_k}{}^2}{\mathbf{C_{ik}}}
$$

where we have used the notation $\mathbf{C_{ik}}$ introduced in the proposition for the estimated amounts of the future $C_{i,k}$, $k > I + 1 - i$, including $\mathbf{C_{i,I+1-i}} = C_{i,I+1-i}$.

We now turn to the second summand of the expression (D2) for $\text{mse}(\mathbf{C_{iI}})$. Because of (D5) we have

$$E_i(C_{iI}) = C_{i,I+1-i} f_{I+1-i} \cdot \ldots \cdot f_{I-1}$$

and therefore

$$\text{(D11)} \quad (E_i(C_{iI}) - \mathbf{C_{iI}})^2 = C_{i,I+1-i}{}^2 (f_{I+1-i} \cdot \ldots \cdot f_{I-1} - \mathbf{f_{I+1-i}} \cdot \ldots \cdot \mathbf{f_{I-1}})^2$$

This expression cannot simply be estimated by replacing f_k with $\mathbf{f_k}$ because this would yield 0 which is not a good estimator because $\mathbf{f_{I+1-i}} \cdot \ldots \cdot \mathbf{f_{I-1}}$ generally will be different from $f_{I+1-i} \cdot \ldots \cdot f_{I-1}$ and therefore the squared difference will be positive. We therefore must take a different approach. We use the algebraic identity

$$F = f_{I+1-i} \cdot \ldots \cdot f_{I-1} - \mathbf{f_{I+1-i}} \cdot \ldots \cdot \mathbf{f_{I-1}}$$

$$= S_{I+1-i} + \ldots + S_{I-1}$$

with

$$S_k = \mathbf{f_{I+1-i}} \cdot \ldots \cdot \mathbf{f_{k-1}} f_k f_{k+1} \cdot \ldots \cdot f_{I-1} - \\ - \mathbf{f_{I+1-i}} \cdot \ldots \cdot \mathbf{f_{k-1}} \mathbf{f_k} f_{k+1} \cdot \ldots \cdot f_{I-1}$$

$$= \mathbf{f_{I+1-i}} \cdot \ldots \cdot \mathbf{f_{k-1}} (f_k - \mathbf{f_k}) f_{k+1} \cdot \ldots \cdot f_{I-1}$$

This yields

$$F^2 = (S_{I+1-i} + \ldots + S_{I-1})^2$$

$$= \sum_{k=I+1-i}^{I-1} S_k{}^2 + 2 \sum_{j<k} S_j S_k$$

where in the last summation j and k run from I + 1 − i to I − 1. Now we replace $S_k{}^2$ with $E(S_k{}^2 \mid B_k)$ and $S_j S_k$, $j < k$, with $E(S_j S_k \mid B_k)$. This means that we approximate $S_k{}^2$ and $S_j S_k$ by varying and averaging as little data as possible so that as many values C_{ik} as possible from data observed are kept fixed. Due to (A4) we have $E(f_k - \mathbf{f_k} \mid B_k) = 0$ and therefore $E(S_j S_k \mid B_k) = 0$ for $j < k$ because all $\mathbf{f_r}$, $r < k$, are scalars under B_k.

Because of

$$\text{(D12)}\quad E((f_k - \mathbf{f_k})^2 \mid B_k) = \mathrm{Var}(\mathbf{f_k} \mid B_k)$$

$$= \sum_{j=1}^{I-k} \mathrm{Var}(C_{j,k+1} \mid B_k) \Big/ \left(\sum_{j=1}^{I-k} C_{jk} \right)^2$$

$$= \sum_{j=1}^{I-k} \mathrm{Var}(C_{j,k+1} \mid C_{j1}, \ldots, C_{jk}) \Big/ \left(\sum_{j=1}^{I-k} C_{jk} \right)^2$$

$$= \sum_{j=1}^{I-k} C_{jk}\alpha_k{}^2 \Big/ \left(\sum_{j=1}^{I-k} C_{jk} \right)^2$$

$$= \alpha_k{}^2 \Big/ \sum_{j=1}^{I-k} C_{jk}$$

we obtain

$$E(S_k{}^2 \mid B_k) = \mathbf{f}^2_{\mathbf{I+1-i}} \cdots \mathbf{f}^2_{\mathbf{k-1}} \alpha_k^2 f^2_{k+1} \cdots f^2_{I-1} \Big/ \sum_{j=1}^{I-k} C_{jk}$$

Taken together, we have replaced $F^2 = (\Sigma S_k)^2$ with $\Sigma_k E(S_k{}^2 \mid B_k)$ and because all terms of this sum are positive we can replace all unknown parameters f_k, $\alpha_k{}^2$ with their unbiased estimators $\mathbf{f_k}, \boldsymbol{\alpha}_\mathbf{k}{}^2$. Altogether, we estimate
$F^2 = (f_{I+1-i} \cdot \ldots \cdot f_{I-1} - \mathbf{f_{I+1-i}} \cdot \ldots \cdot \mathbf{f_{I-1}})^2$ by

$$\sum_{k=I+1-i}^{I-1} \left(\mathbf{f}^2_{\mathbf{I+1-i}} \cdots \mathbf{f}^2_{\mathbf{k-1}} \cdot \boldsymbol{\alpha}^2_\mathbf{k} \cdot \mathbf{f}^2_{\mathbf{k+1}} \cdots \mathbf{f}^2_{\mathbf{I-1}} \Big/ \sum_{j=1}^{I-k} C_{jk} \right) = \mathbf{f}^2_{\mathbf{I+1-i}} \cdot \ldots \cdot \mathbf{f}^2_{\mathbf{I-1}} \sum_{k=I+1-i}^{I-1} \frac{\boldsymbol{\alpha}_\mathbf{k}{}^2 / \mathbf{f_k}{}^2}{\sum_{j=1}^{I-k} C_{jk}}$$

Using (D11), this means that we estimate $(E_i(C_{iI}) - \mathbf{C_{iI}})^2$ by

(D13) $$C_{i,I+1-i}^2 \mathbf{f}_{\mathbf{I+1-i}}^2 \cdot \ldots \cdot \mathbf{f}_{\mathbf{I-1}}^2 \sum_{k=I+1-i}^{I-1} \frac{\boldsymbol{\alpha}_{\mathbf{k}}^2 / \mathbf{f}_{\mathbf{k}}^2}{\sum_{j=1}^{I-k} C_{jk}}$$

$$= \mathbf{C}_{\mathbf{iI}}^2 \sum_{k=I+1-i}^{I-1} \frac{\boldsymbol{\alpha}_{\mathbf{k}}^2 / \mathbf{f}_{\mathbf{k}}^2}{\sum_{j=1}^{I-k} C_{jk}}$$

From (D2), (D10) and (D13) we finally obtain the estimator $(\text{s.e.}(\mathbf{C_{iI}}))^2$ for $\text{mse}(\mathbf{C_{iI}})$ as stated in the proposition.

Appendix E: Unbiasedness of the Estimator α_k^2

Proposition: Under the assumptions

(3) There are unknown constants $f_1, ..., f_{I-1}$ with

$$E(C_{i,k+1} \mid C_{i1}, ..., C_{ik}) = C_{ik} f_k, \qquad 1 \le i \le I,\ 1 \le k \le I-1$$

(4) The variables $\{C_{i1}, ..., C_{iI}\}$ and $\{C_{j1}, ..., C_{jI}\}$ of different accident years $i \ne j$ are independent.

(5) There are unknown constants $\alpha_1, ..., \alpha_{I-1}$ with

$$Var(C_{i,k+1} \mid C_{i1}, ..., C_{ik}) = C_{ik} \alpha_k^2, \qquad 1 \le i \le I,\ 1 \le k \le I-1$$

the estimators

$$\boldsymbol{\alpha}_{\mathbf{k}}^2 = \frac{1}{I-k-1} \sum_{j=1}^{I-k} C_{jk} \left(\frac{C_{j,k+1}}{C_{jk}} - \mathbf{f_k} \right)^2, \qquad 1 \le k \le I-2$$

of α_k^2 are unbiased, that is we have

$$E(\boldsymbol{\alpha}_{\mathbf{k}}^2) = \alpha_k^2, \qquad 1 \le k \le I-2$$

Proof: In this proof all summations are over the index j from j = 1 to j = I – k. The definition of $\boldsymbol{\alpha}_{\mathbf{k}}^2$ can be rewritten as

$$\text{(E1)} \quad (I-k-1)\boldsymbol{\alpha}_{\mathbf{k}}^2 = \Sigma(C_{j,k+1}^2 / C_{jk} - 2 \cdot C_{j,k+1}\mathbf{f_k} + C_{jk}\mathbf{f_k}^2)$$
$$= \Sigma(C_{j,k+1}^2 / C_{jk}) - \Sigma(C_{jk}\mathbf{f_k}^2)$$

using $\Sigma C_{j,k+1} = \mathbf{f_k} \Sigma C_{jk}$ according to the definition of $\mathbf{f_k}$. Using again the set

$$B_k = \{ C_{ij} \mid i + j \le I + 1, j \le k \}$$

of variables C_{ij} assumed to be known, (E1) yields

$$\text{(E2)} \quad E((I-k-1)\boldsymbol{\alpha}_{\mathbf{k}}^2 \mid B_k) = \Sigma E(C_{j,k+1}^2 \mid B_k) / C_{jk} - \Sigma C_{jk} E(\mathbf{f_k}^2 \mid B_k)$$

because C_{jk} is a scalar under the condition of B_k being known.

Due to the independence (4) of the accident years, conditions which are independent from the conditioned variable can be omitted in $E(C_{j,k+1}{}^2 \mid B_k)$, that is

$$\text{(E3)}\quad E(C_{j,k+1}{}^2 \mid B_k) = E(C_{j,k+1}{}^2 \mid C_{j1}, \ldots, C_{jk})$$

$$= \mathrm{Var}(C_{j,k+1} \mid C_{j1}, \ldots, C_{jk}) + (E(C_{j,k+1} \mid C_{j1}, \ldots, C_{jk}))^2$$

$$= C_{jk}\alpha_k{}^2 + (C_{jk} f_k)^2$$

where the rule $E(X^2) = \mathrm{Var}(X) + (E(X))^2$ and the assumptions (5) and (3) have also been used.

From (D12) and (A4) we gather

$$\text{(E4)}\quad E(\mathbf{f_k}{}^2 \mid B_k) = \mathrm{Var}(\mathbf{f_k}{}^2 \mid B_k) + (E(\mathbf{f_k} \mid B_k))^2$$

$$= \alpha_k{}^2 / \Sigma C_{jk} + f_k{}^2$$

Inserting (E3) and (E4) into (E2) we obtain

$$E((I-k-1)\boldsymbol{\alpha}_{\mathbf{k}}{}^2 \mid B_k) = \sum_{j=1}^{I-k}(\alpha_k{}^2 + C_{jk} f_k{}^2) - \sum_{j=1}^{I-k}\left(C_{jk}\alpha_k{}^2 / \sum_{j=1}^{I-k} C_{jk} + C_{jk} f_k{}^2\right)$$

$$= (I-k)\alpha_k{}^2 - \alpha_k{}^2$$

$$= (I-k-1)\alpha_k{}^2$$

From this we immediately obtain $E(\boldsymbol{\alpha}_{\mathbf{k}}{}^2 \mid B_k) = \alpha_k{}^2$

Finally, the iterative rule for expectations yields

$$E(\boldsymbol{\alpha}_{\mathbf{k}}{}^2) = E(E(\boldsymbol{\alpha}_{\mathbf{k}}{}^2 \mid B_k)) = E(\alpha_k{}^2) = \alpha_k{}^2$$

Appendix F: The Standard Error of the Overall Reserve Estimate

Proposition: Under the assumptions

(3) There are unknown constants $f_1, \ldots, f_{I-1}$ with

$$E(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) = C_{ik} f_k, \qquad 1 \le i \le I,\ 1 \le k \le I-1$$

(4) The variables $\{C_{i1}, \ldots, C_{iI}\}$ and $\{C_{j1}, \ldots, C_{jI}\}$ of different accident years $i \ne j$ are independent.

(5) There are unknown constants $\alpha_1, \ldots, \alpha_{I-1}$ with

$$\mathrm{Var}(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) = C_{ik} \alpha_k^{\,2}, \qquad 1 \le i \le I,\ 1 \le k \le I-1$$

the standard error s.e.($\mathbf{R}$) of the overall reserve estimate

$$\mathbf{R} = \mathbf{R_2} + \ldots + \mathbf{R_I}$$

is given by

$$(\mathrm{s.e.}(\mathbf{R}))^2 = \sum_{i=2}^{I} \left\{ (\mathrm{s.e.}(\mathbf{R_i}))^2 + \mathbf{C_{iI}} \left(\sum_{j=i+1}^{I} \mathbf{C_{jI}} \right) \sum_{k=I+1-i}^{I-1} \frac{2\boldsymbol{\alpha}_\mathbf{k}^{\,2} / \mathbf{f_k}^{\,2}}{\sum_{n=1}^{I-k} C_{nk}} \right\}$$

Proof: This proof is analogous to that in Appendix D. The comments will therefore be brief. We first must determine the mean squared error mse($\mathbf{R}$) of $\mathbf{R}$. Using again D=$\{C_{ik} \mid i + k \le I + 1\}$ we have

$$\text{(F1)}\quad \mathrm{mse}\left(\sum_{i=2}^{I} \mathbf{R_i} \right) = E\left(\left(\sum_{i=2}^{I} \mathbf{R_i} - \sum_{i=2}^{I} R_i \right)^2 \middle| D \right)$$

$$= E\left(\left(\sum_{i=2}^{I} \mathbf{C_{iI}} - \sum_{i=2}^{I} C_{iI} \right)^2 \middle| D \right)$$

$$= \mathrm{Var}\left(\sum_{i=2}^{I} C_{iI} \,\middle|\, D \right) + \left(E\left(\sum_{i=2}^{I} C_{iI} \,\middle|\, D \right) - \sum_{i=2}^{I} \mathbf{C_{iI}} \right)^2$$

The independence of the accident years yields

$$\text{(F2)}\quad \operatorname{Var}\left(\sum_{i=2}^{I} C_{iI} \,\middle|\, D\right) = \sum_{i=2}^{I} \operatorname{Var}(C_{iI} \mid C_{i1}, \ldots, C_{i,I+1-i})$$

whose summands have been calculated in Appendix D, see (D9). Furthermore

$$\text{(F3)}\quad \left(E\left(\sum_{i=2}^{I} C_{iI} \,\middle|\, D\right) - \sum_{i=2}^{I} \mathbf{C_{iI}}\right)^2 = \left(\sum_{i=2}^{I} (E(C_{iI} \mid D) - \mathbf{C_{iI}})\right)^2$$

$$= \sum_{2 \le i,j \le I} (E(C_{iI} \mid D) - \mathbf{C_{iI}}) \cdot (E(C_{jI} \mid D) - \mathbf{C_{jI}})$$

$$= \sum_{2 \le i,j \le I} C_{i,I+1-i} C_{j,I+1-j} F_i F_j$$

$$= \sum_{i=2}^{I} (C_{i,I+1-i} F_i)^2 + 2 \sum_{i<j} C_{i,I+1-i} C_{j,I+1-j} F_i F_j$$

with (as for (D11))

$$F_i = f_{I+1-i} \cdots f_{I-1} - \mathbf{f_{I+1-i}} \cdots \mathbf{f_{I-1}}$$

which is identical to F of Appendix D but here we have to carry the index i, too. In Appendix D we have shown (cf. (D2) and (D11)) that

$$\operatorname{mse}(\mathbf{R_i}) = \operatorname{Var}(C_{iI} \mid C_{i1}, \ldots, C_{i,I+1-i}) + (C_{i,I+1-i} F_i)^2$$

Comparing this with (F1), (F2) and (F3) we see that

$$\text{(F4)}\quad \operatorname{mse}\left(\sum_{i=2}^{I} \mathbf{R_i}\right) = \sum_{i=2}^{I} \operatorname{mse}(\mathbf{R_i}) + \sum_{2 \le i<j \le I} 2 \cdot C_{i,I+1-i} C_{j,I+1-j} F_i F_j$$

We therefore need only develop an estimator for $F_i F_j$. A procedure completely analogous to that for F^2 in the proof of Appendix D yields for $F_i F_j,\ i < j,$ the estimator

$$\sum_{k=I+1-i}^{I-1} \mathbf{f_{I+1-j}} \ldots \mathbf{f_{I-i}} \mathbf{f}^2_{\mathbf{I+1-i}} \ldots \mathbf{f}^2_{\mathbf{k-1}} \boldsymbol{\alpha}^2_{\mathbf{k}} \mathbf{f}^2_{\mathbf{k+1}} \ldots \mathbf{f}^2_{\mathbf{I-1}} \Big/ \sum_{n=1}^{I-k} C_{nk}$$

which immediately leads to the result stated in the proposition.

Appendix G: Testing for Correlations between Subsequent Development Factors

In this appendix we first prove that the basic assumption (3) of the chain ladder method implies that subsequent development factors $C_{ik} / C_{i,k-1}$ and $C_{i,k+1} / C_{ik}$ are not correlated. Then we show how we can test if this uncorrelatedness is met for a given run-off triangle. Finally, we apply this test procedure to the numerical example of section 6.

Proposition: Under the assumption

(3) There are unknown constants $f_1, \ldots, f_{I-1}$ with

$$E(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) = C_{ik} f_k, \quad 1 \le i \le I, \ 1 \le k \le I-1$$

subsequent development factors $C_{ik} / C_{i,k-1}$ and $C_{i,k+1} / C_{ik}$ are uncorrelated, that is we have (for $1 \le i \le I$, $2 \le k \le I-1$)

$$E\left(\frac{C_{ik}}{C_{i,k-1}} \cdot \frac{C_{i,k+1}}{C_{ik}}\right) = E\left(\frac{C_{ik}}{C_{i,k-1}}\right) E\left(\frac{C_{i,k+1}}{C_{ik}}\right)$$

<u>Proof</u>: For $j \le k$ we have

$$\begin{aligned}
\text{(G1)}\quad E(C_{i,k+1} / C_{ij}) &= E(E(C_{i,k+1} / C_{ij} \mid C_{i1}, \ldots, C_{ik})) && \text{(a)}\\
&= E(E(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) / C_{ij}) && \text{(b)}\\
&= E(C_{ik} f_k / C_{ij}) && \text{(c)}\\
&= E(C_{ik} / C_{ij}) f_k && \text{(d)}
\end{aligned}$$

Here equation (a) holds due to the iterative rule E(X) = E(E(X|Y)) for expectations, (b) holds because, given $C_{i1}, \ldots, C_{ik}$, C_{ij} is a scalar for $j \le k$, (c) holds due to (3) and (d) holds because f_k is a scalar.

From (G1) we obtain through the special case $j = k$

$$\text{(G2)}\quad E(C_{i,k+1} / C_{ik}) = E(C_{ik} / C_{ik}) f_k = f_k$$

and through $j = k - 1$

$$\begin{aligned}
\text{(G3)}\quad E\left(\frac{C_{ik}}{C_{i,k-1}} \cdot \frac{C_{i,k+1}}{C_{ik}}\right) &= E\left(\frac{C_{i,k+1}}{C_{i,k-1}}\right) && \text{(G1)}\\
&= E\left(\frac{C_{ik}}{C_{i,k-1}}\right) f_k
\end{aligned}$$

Inserting (G2) into (G3) completes the proof.

Designing the test procedure

The usual test for uncorrelatedness requires that we have identically distributed pairs of observations which come from a Normal distribution. Both conditions are usually not fulfilled for adjacent columns of development factors. (Note that due to (G2) the development factors $C_{i,k+1} / C_{ik}$, $1 \le i \le I - k$, have the same expectation but assumption (5) implies that they have different variances.) We therefore use the test with Spearman's rank correlation coefficient because this test is distribution-free and because by using ranks the differences in the variances of $C_{i,k+1} / C_{ik}$, $1 \le i \le I - k$, become less important. Even if these differences are negligible the test will only be of an approximate nature because, strictly speaking, it is a test for independence rather than for uncorrelatedness. But we will take this into account when fixing the critical value of the test statistic.

For the application of Spearman's test we consider a fixed development year k and rank the development factors $C_{i,k+1} / C_{ik}$ observed so far according to their size starting with the smallest one on rank one and so on. Let r_{ik}, $1 \le i \le I - k$, denote the rank of $C_{i,k+1} / C_{ik}$ obtained in this way, $1 \le r_{ik} \le I - k$. Then we do the same with the preceding development factors $C_{ik} / C_{i,k-1}$, $1 \le i \le I - k$, leaving out $C_{I+1-k,k} / C_{I+1-k,k-1}$ for which the subsequent development factor has not yet been observed. Let s_{ik}, $1 \le i \le I - k$, be the ranks obtained in this way, $1 \le s_{ik} \le I - k$. Now, Spearman's rank correlation coefficient T_k is defined to be

$$\text{(G4)} \quad T_k = 1 - 6\sum_{i=1}^{I-k}(r_{ik} - s_{ik})^2 / ((I - k)^3 - I + k)$$

It can be seen that

$$-1 \le T_k \le +1$$

and, under the null-hypothesis,

$$E(T_k) = 0$$
$$Var(T_k) = 1/(I - k - 1)$$

A value of T_k close to 0 indicates that the development factors between development years k – 1 and k and those between years k and k + 1 are not correlated. Any other value of T_k indicates that the factors are (positively or negatively) correlated.

For a formal test we do not want to consider every pair of columns of adjacent development years separately in order to avoid an accumulation of the error probabilities. We therefore consider the triangle as a whole. This also is preferable from a practical point of view because it is more important to know whether correlations globally prevail than to find a small part of the triangle with correlations. We therefore combine all values T_2, T_3, ..., T_{I-2} obtained in the same way like T_k. (There is no T_1 because there are no development factors before development year k=1 and similarly there is also no T_I; even T_{I-1} is not included because there is only one rank and therefore no randomness.)

According to Appendix B we should not form an unweighted average of $T_2, \ldots, T_{I-2}$ but rather use weights which are inversely proportional to $\mathrm{Var}(T_k) = 1/(I-k-1)$. This leads to weights which are just equal to one less than the number of pairs (r_{ik}, s_{ik}) taken into account by T_k which seems very reasonable.

We thus calculate

$$\text{(G5)} \qquad T = \sum_{k=2}^{I-2} (I-k-1) T_k \Big/ \sum_{k=2}^{I-2} (I-k-1)$$

$$= \sum_{k=2}^{I-2} \frac{I-k-1}{(I-2)(I-3)/2} T_k$$

$$E(T) = \sum_{k=2}^{I-2} E(T_k) = 0$$

$$\text{(G6)} \quad \mathrm{Var}(T) = \sum_{k=2}^{I-2} (I-k-1)^2 \, \mathrm{Var}(T_k) \Big/ \left(\sum_{k=2}^{I-2} (I-k-1) \right)^2$$

$$= \sum_{k=2}^{I-2} (I-k-1) \Big/ \left(\sum_{k=2}^{I-2} (I-k-1) \right)^2$$

$$= \frac{1}{(I-2)(I-3)/2}$$

where for the calculation of Var(T) we used the fact that under the null-hypothesis subsequent development factors and therefore also different T_k's are uncorrelated.

Because the distribution of a single T_k with $I - k \geq 10$ is Normal in good approximation and because T is the aggregation of several uncorrelated T_k's (which all are symmetrically distributed around their mean 0) we can assume that T has approximately a Normal distribution and use this to design a significance test. Usually, when applying a significance test one rejects the null-hypothesis if it is very unlikely to hold, e.g. if the value of the test statistic is outside its 95% confidence interval. But in our case we propose to use only a 50% confidence interval because the test is only of an approximate nature and because we want to detect correlations already in a substantial part of the run-off triangle. Therefore, as the probability for a

Standard Normal variate lying in the interval (−.67, .67) is 50% we do not reject the null-hypothesis of having uncorrelated development factors if

$$-\frac{0.67}{\sqrt{((I-2)(I-3)/2)}} \leq T \leq +\frac{0.67}{\sqrt{((I-2)(I-3)/2)}}$$

If T is outside this interval we should be reluctant with the application of the chain ladder method and analyze the correlations in more detail. In such a case, an autoregressive model of an order > 1 is probably more appropriate, for example by replacing the fundamental chain ladder assumption (3) with

$$E(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) = C_{ik} f_k + C_{i,k-1} g_k$$

Application to the example of section 6:

We start with the table of all development factors:

	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9
i=1	1.6	1.32	1.08	1.15	1.20	1.11	1.033	1.00	1.01
i=2	40.4	1.26	1.98	1.29	1.13	0.99	1.043	1.03	
i=3	2.6	1.54	1.16	1.16	1.19	1.03	1.026		
i=4	2.0	1.36	1.35	1.10	1.11	1.04			
i=5	8.8	1.66	1.40	1.17	1.01				
I=6	4.3	1.82	1.11	1.23					
I=7	7.2	2.72	1.12						
I=8	5.1	1.89							
I=9	1.7								

As described above we first rank column F_1 according to the size of the factors, then leave out the last element and rank the column again. Then we do the same with columns F_2 to F_8. This yields the following table:

r_{i1}	s_{i2}	r_{i2}	s_{i3}	r_{i3}	s_{i4}	r_{i4}	s_{i5}	r_{i5}	s_{i6}	r_{i6}	s_{i7}	r_{i7}	s_{i8}	r_{i8}
1	1	2	2	1	1	2	2	5	4	4	3	2	1	1
9	8	1	1	7	6	6	5	3	2	1	1	3	2	2
4	3	4	4	4	3	3	3	4	3	2	2	1		
3	2	3	3	5	4	1	1	2	1	3				
8	7	5	5	6	5	4	4	1						
5	4	6	6	2	2	5								
7	6	8	7	3										
6	5	7												
2														

We now add the squared differences between adjacent rank columns of equal length, that is we add $(s_{ik} - r_{ik})^2$ over i for every k, $2 \le k \le 8$. This yields 68, 74, 20, 24, 6, 6 and 0. (Remember that we have to leave out k = 1 because there is no s_{i1}, and k = 9 because there is only one pair of ranks and therefore no randomness.) From these figures we obtain Spearman's rank correlation coefficients T_k according to formula (G4):

k	2	3	4	5	6	7	8
T_k	4/21	−9/28	3/7	−1/5	2/5	−1/2	1
I − k −1	7	6	5	4	3	2	1

The (I − k − 1)-weighted average of the T_k's is T = .070 (see formula (G5)). Because of Var(T) = 1/28 (see (G6)) the 50% confidence limits for T are ± .67/√28 = ±.127. Thus, T is within its 50%-interval and the hypothesis of having uncorrelated development factors is not rejected.

Appendix H: Testing for Calendar Year Effects

One of the three basic assumptions underlying the chain ladder method was seen to be assumption (4) of the independence of the accident years. The main reason why this independence can be violated in practice is the fact that we can have certain calendar year effects such as major changes in claims handling or in case reserving or external influences such as substantial changes in court decisions or inflation. Note that a constant rate of inflation which has not been removed from the data is extrapolated into the future by the chain ladder method. In the following, we first generally describe a procedure to test for such calendar year influences and then apply it to our example.

Designing the test procedure:

A calendar year influence affects one of the diagonals

$$D_j = \{C_{j1}, C_{j-1,2}, \ldots, C_{2,j-1}, C_{1j}\}, \qquad 1 \le j \le I$$

and therefore also influences the adjacent development factors

$$A_j = \{C_{j2} / C_{j1}, C_{j-1,3} / C_{j-1,2}, \ldots, C_{1,j+1} / C_{1j}\}$$

and

$$A_{j-1} = \{C_{j-1,2} / C_{j-1,1}, C_{j-2,3} / C_{j-2,2}, \ldots, C_{1j} / C_{1,j-1}\}$$

where the elements of D_j form either the denominator or the numerator. Thus, if due to a calendar year influence the elements of D_j are larger (smaller) than usual, then the elements of A_{j-1} are also larger (smaller) than usual and the elements of A_j are smaller (larger) than usual.

Therefore, in order to check for such calendar year influences we only have to subdivide all development factors into 'smaller' and 'larger' ones and then to examine whether there are diagonals where the small development factors or the large ones clearly prevail. For this purpose, we order for every k, $1 \le k \le I - 1$, the elements of the set

$$F_k = \{C_{i,k+1} / C_{ik} \mid 1 \le i \le I - k\}$$

that is of the column of all development factors observed between development years k and k + 1, according to their size and subdivide them into one part LF_k of larger factors being greater than the median of F_k and into a second part SF_k of smaller factors below the median of F_k. (The median of a set of real numbers is defined to be a number which divides the set into two parts with the same number of elements.) If the number $I - k$ of elements of F_k is odd there is one element of F_k which is equal to the median and therefore assigned to neither of the sets LF_k and SF_k; this element is eliminated from all further considerations.

Having done this procedure for each set F_k, $1 \le k \le I - 1$, every development factor observed is

- either eliminated (like e.g. the only element of F_{I-1})
- or assigned to the set L = $LF_1 + ... + LF_{I-2}$ of larger factors
- or assigned to the set S = $SF_1 + ... + SF_{I-2}$ of smaller factors

In this way, every development factor which is not eliminated has a 50% chance of belonging to either L or S.

Now we count for every diagonal A_j, $1 \le j \le I - 1$, of development factors the number L_j of large factors, that is elements of L, and the number S_j of small factors, that is elements of S. Intuitively, if there is no specific change from calendar year j to calendar year j + 1, A_j should have about the same number of small factors as of large factors, that is L_j and S_j should be of approximately the same size apart from pure random fluctuations. But if L_j is significantly larger or smaller than S_j or, equivalently, if

$$Z_j = \min(L_j, S_j)$$

that is the smaller of the two figures, is significantly smaller than $(L_j + S_j)/2$, then there is some reason for a specific calendar year influence.

In order to design a formal test we need the probability distribution of Z_j under the null-hypothesis that each development factor has a 50 % probability of belonging to either L or S. This distribution can easily be established. We give an example for the case where $L_j + S_j = 5$, that is where the set A_j contains 5 development factors without counting any eliminated factor. Then the number L_j has a Binomial distribution with n = 5 and p = .5, that is

$$\text{prob}(L_j = m) = \binom{n}{m}\frac{1}{2^n} = \binom{5}{m}\frac{1}{2^5}, \quad m = 0, 1, ..., 5$$

Therefore

$$\text{prob}(S_j = 5) = \text{prob}(L_j = 0) = 1/32$$
$$\text{prob}(S_j = 4) = \text{prob}(L_j = 1) = 5/32$$
$$\text{prob}(S_j = 3) = \text{prob}(L_j = 2) = 10/32$$
$$\text{prob}(S_j = 2) = \text{prob}(L_j = 3) = 10/32$$
$$\text{prob}(S_j = 1) = \text{prob}(L_j = 4) = 5/32$$
$$\text{prob}(S_j = 0) = \text{prob}(L_j = 5) = 1/32$$

This yields

$$\text{prob}(Z_j = 0) = \text{prob}(L_j = 0) + \text{prob}(S_j = 0) = 2/32$$
$$\text{prob}(Z_j = 1) = \text{prob}(L_j = 1) + \text{prob}(S_j = 1) = 10/32$$
$$\text{prob}(Z_j = 2) = \text{prob}(L_j = 2) + \text{prob}(S_j = 2) = 20/32$$

In this way we obtain very easily the following table for the cumulative probability distribution function of Z_j:

n	$\text{prob}(Z_j \le 0)$	$\text{prob}(Z_j \le 1)$	$\text{prob}(Z_j \le 2)$ ...
≤4	>10%	>10%	>10%
5	6.25%	>10%	>10%
6	3.1%	>10%	>10%
7	1.6%	>10%	>10%
8	0.8%	7.0%	>10%
9	0.4%	3.9%	>10%
10	0.2%	2.1%	>10%
11	0.1%	1.2%	6.5%
...			

Now, we use this table in the following way: Any realization $Z_j = z_j$ with a cumulative probability $\text{prob}(Z_j \le z_j) \le 10\ \%$ indicates that the corresponding set $A_j = \{C_{j2} / C_{j1}, C_{j-1,3} / C_{j-1,2}, \ldots\}$ contains either significantly many "larger" or significantly many "smaller" development factors. Then, the factors of the predominant type (either the larger or the smaller factors of A_j) are assumed to be influenced by a specific calendar year effect and are viewed to be outliers. Therefore, it seems to be advisable to reduce their weight when calculating the age-to-age factors $\mathbf{f_k}$.

Specifically, it is proposed to reduce the weight of each of these outlying development factors to 50 % of its original weight, that is to calculate

$$\mathbf{f_k} = \sum_{i=1}^{I-k} w_{ik} C_{i,k+1} \Big/ \sum_{i=1}^{I-k} w_{ik} C_{ik}$$

with $w_{ik} = .5$ if $C_{i,k+1} / C_{ik}$ belongs to the factors of the predominant type (either larger or smaller) of its corresponding set A_{i+k-1} and if A_{i+k-1} shows a significant calendar year effect, that is if $\text{prob}(Z_{i+k-1} \le z_{i+k-1}) \le 10\ \%$. In all other cases we put $w_{ik} = 1$ as usual. Strictly speaking, the formulae for $\boldsymbol{\alpha}_{\mathbf{k}}^2$ and for the standard error must be changed analogously, if some $w_{ik} < 1$ are used.

As with every test procedure which is applied several times there is an accumulation of the error probabilities, that is the danger increases that we find a significant case which in reality is not extraordinary. But here, this will not cause any essential disadvantage as we only change weights and do not discard anything entirely.

Application to the example of section 6:

We start with the triangle of all development factors observed:

	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9
i=1	1.6	1.32	1.08	1.15	1.20	1.11	1.033	1.00	1.01
i=2	40.4	1.26	1.98	1.29	1.13	0.99	1.043	1.03	
i=3	2.6	1.54	1.16	1.16	1.19	1.03	1.026		
i=4	2.0	1.36	1.35	1.10	1.11	1.04			
i=5	8.8	1.66	1.40	1.17	1.01				
i=6	4.3	1.82	1.11	1.23					
i=7	7.2	2.72	1.12						
i=8	5.1	1.89							
i=9	1.7								

We have to subdivide each column F_k into the subset SF_k of 'smaller' factors below the median of F_k and into the subset LF_k of 'larger' factors above the median. This can be done very easily with the help of the rank columns r_{ik} established in Appendix G: The half of factors with small ranks belongs to SF_k, those with large ranks to LF_k and if the total number is odd we have to eliminate the mean rank. Replacing a small rank with 'S', a large rank with 'L' and a mean rank with '*' we obtain the following picture:

	j=1	j=2	j=3	j=4	j=5	j=6	j=7	j=8	j=9
j=1	S	S	S	S	L	L	*	S	*
j=2	L	S	L	L	*	S	L	L	
j=3	S	S	*	S	L	S	S		
j=4	S	S	L	S	S	L			
j=5	L	L	L	L	S				
j=6	*	L	S	L					
j=7	L	L	S						
j=8	L	L							
j=9	S								

We now count for every diagonal A_j, $2 \le j \le 9$, the number L_j of L's and the number S_j of S's. We have left out A_1 because it contains at most one element which is not eliminated, and therefore Z_1 is not a random variable but always = 0. With the notations s_j, l_j, z_j for the realizations of the random variables S_j, L_j, Z_j and with $n = s_j + l_j$ as above, we obtain the following table:

j	s_j	l_j	z_j	n	$\text{prob}(Z_j \le z_j)$
2	1	1	1	2	>10%
3	3	0	0	3	>10%
4	3	1	1	4	>10%
5	1	3	1	4	>10%
6	1	3	1	4	>10%
7	2	4	2	6	>10%
8	4	4	4	8	>10%
9	4	4	4	8	>10%

According to the probabilities $\text{prob}(Z_j \le z_j)$ there does not seem to be any calendar year effect. Therefore, there is no reason to change any weight in the calculation of the age-to-age factors.

As a final check for calendar year effects we can plot all standardized residuals

$$(C_{i,k+1} / C_{ik} - \mathbf{f_k})\sqrt{C_{ik}} / \boldsymbol{\alpha}_{\mathbf{k}}, \qquad 2 \le i + k \le I$$

against the calendar years j = i + k. For the data of our example, the resulting plot is shown in Figure 14. There does not seem to be any specific trend or irregularity in the pattern of these residuals. The fact that only positive residuals are absolutely larger than 1.6 hints at a positive skewness of the distribution of the development factors.

<>

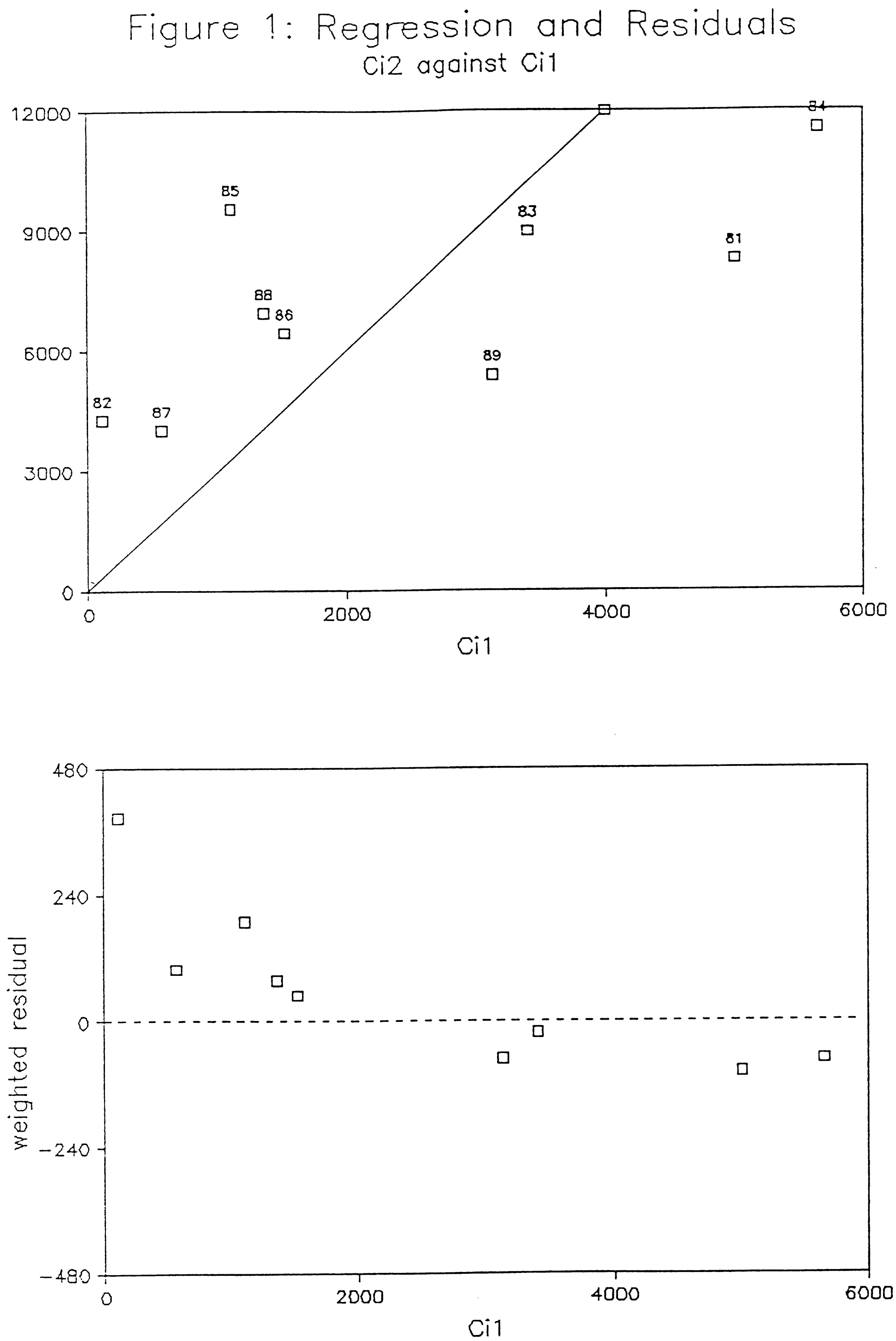
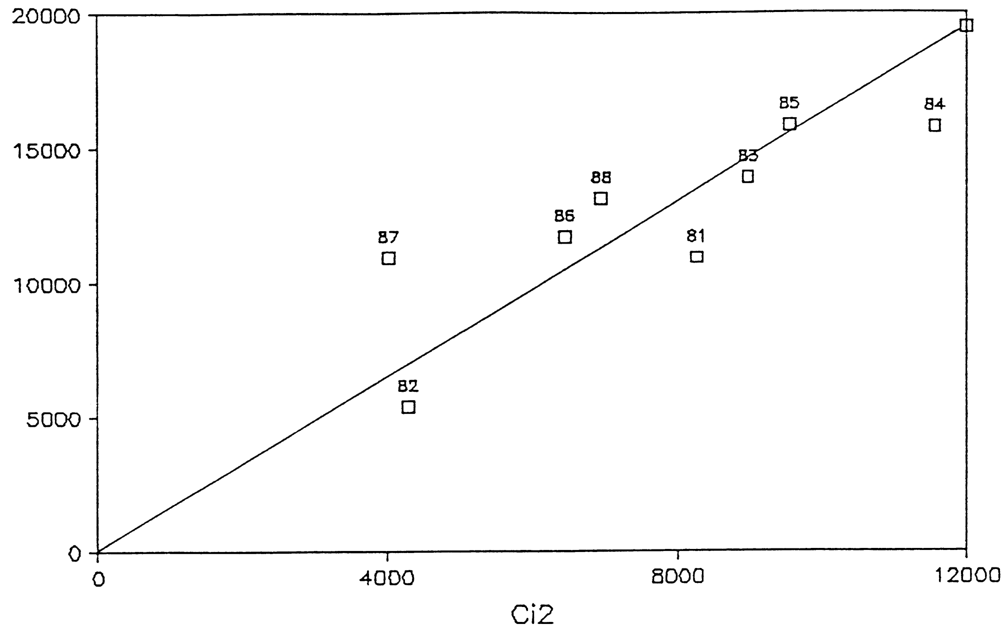
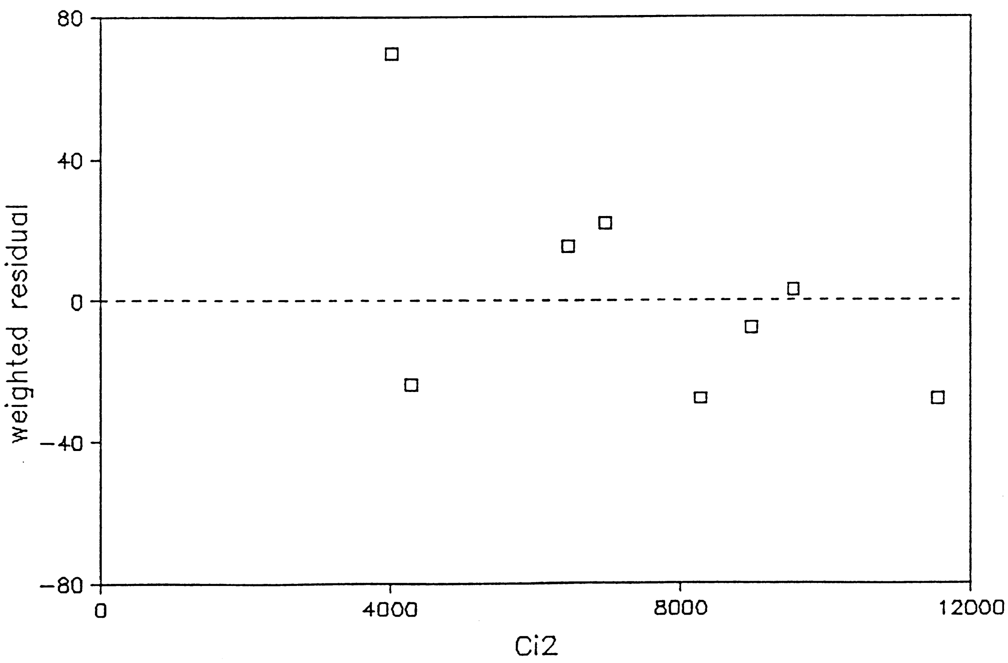


Figure 1: Regression and Residuals
Ci2 against Ci1
12000
9000
6000
3000
0
2000
4000
6000
Ci1
85
83
81
88
86
89
82
87
480
240
-240
-480
weighted residual

Figure 2: Regression and Residuals

Ci3 against Ci2





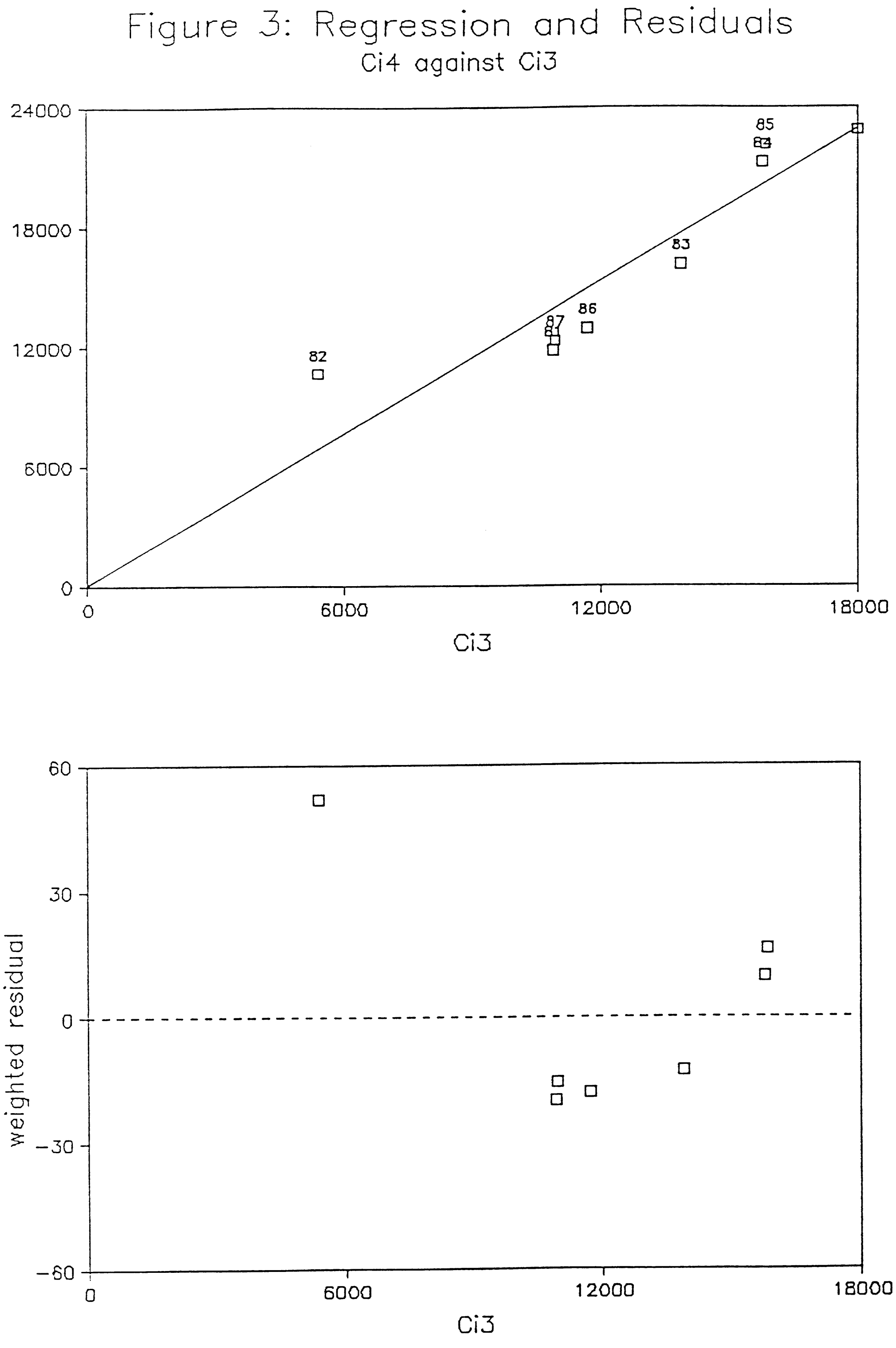


Figure 3: Regression and Residuals
Ci4 against Ci3
24000
18000
12000
6000
0
0
6000
12000
18000
Ci3
85
84
83
86
87
81
82
60
30
0
−30
−60
weighted residual
0
6000
12000
18000
Ci3

Figure 4: Regression and Residuals

Ci5 against Ci4

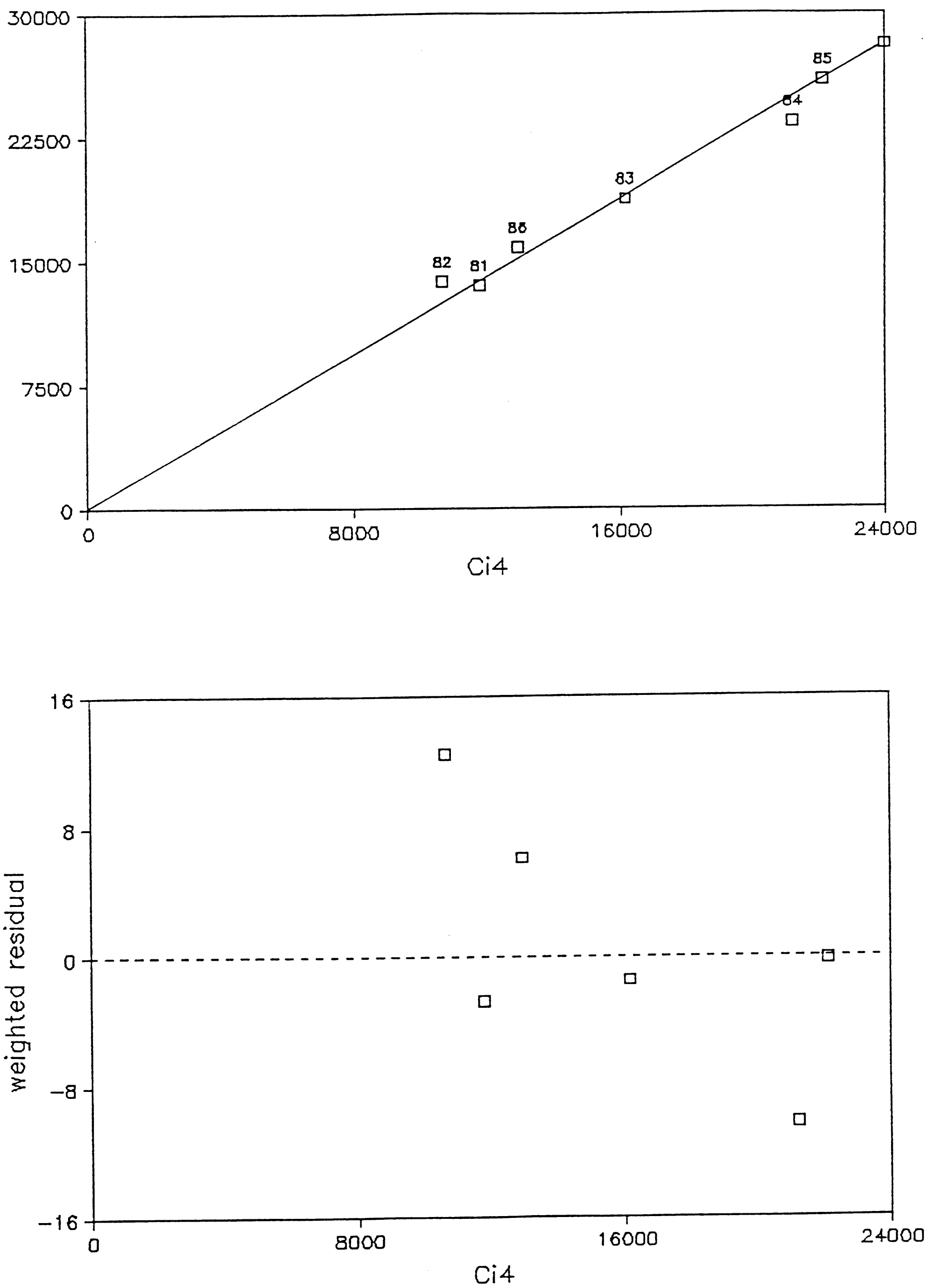


Figure 5: Regression and Residuals

Ci6 against Ci5

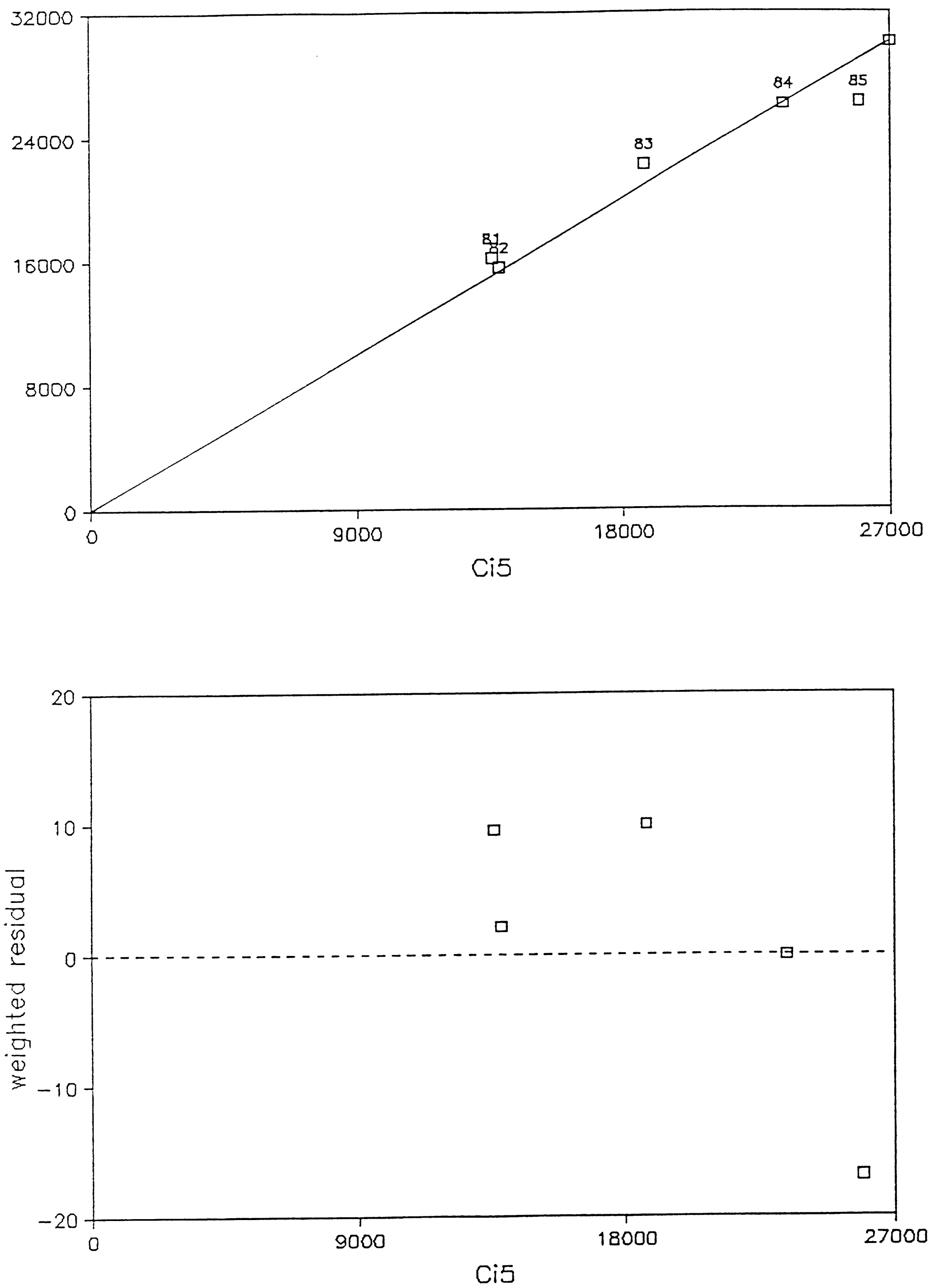
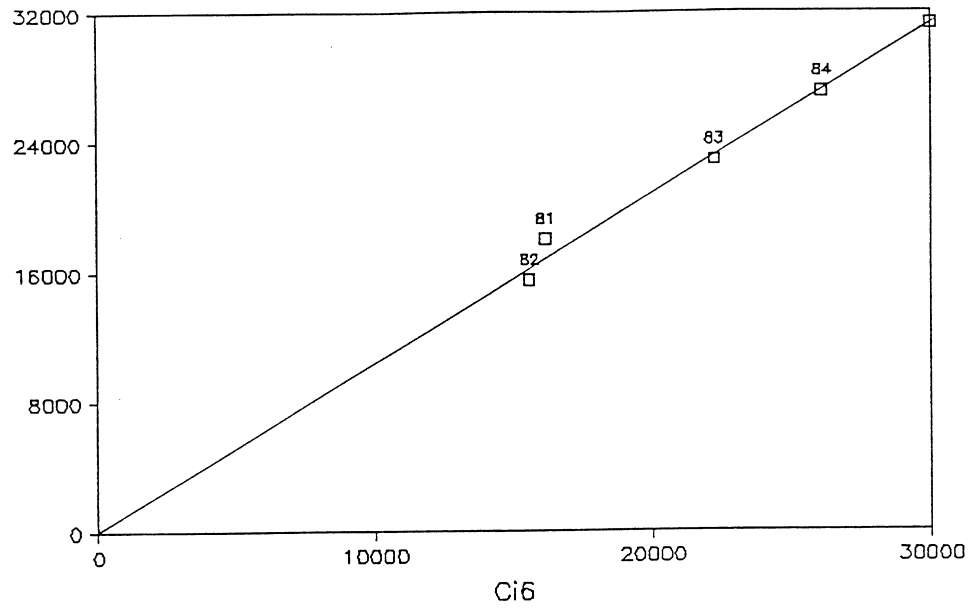


Figure 6: Regression and Residuals
Ci7 against Ci6



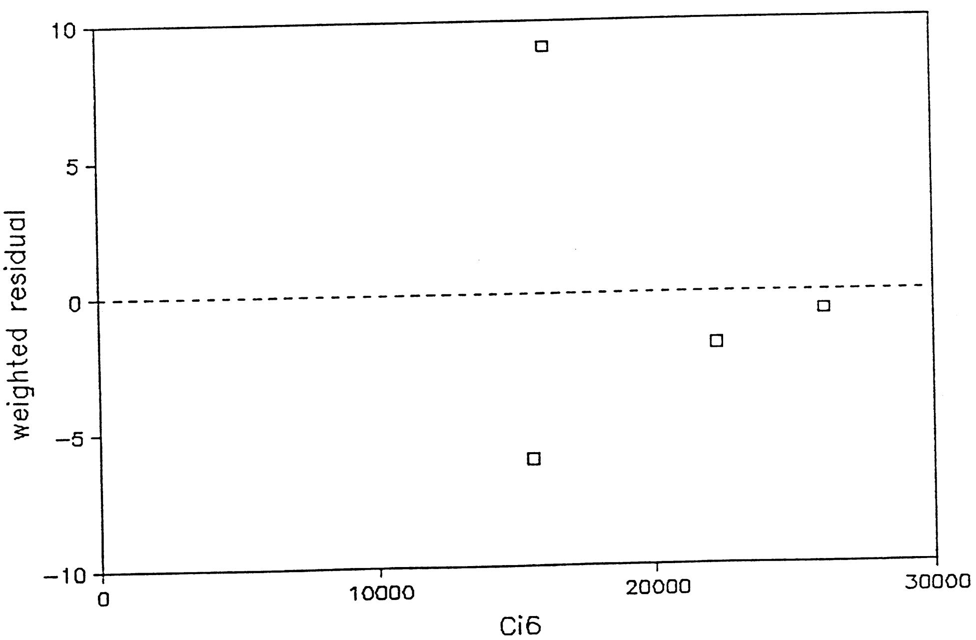
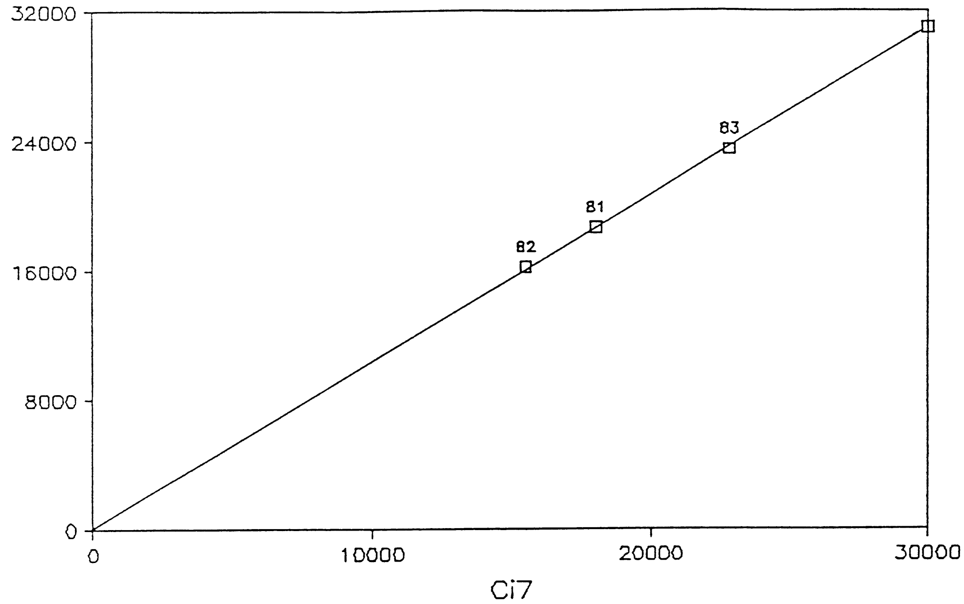


Figure 7: Regression and Residuals
Ci8 against Ci7



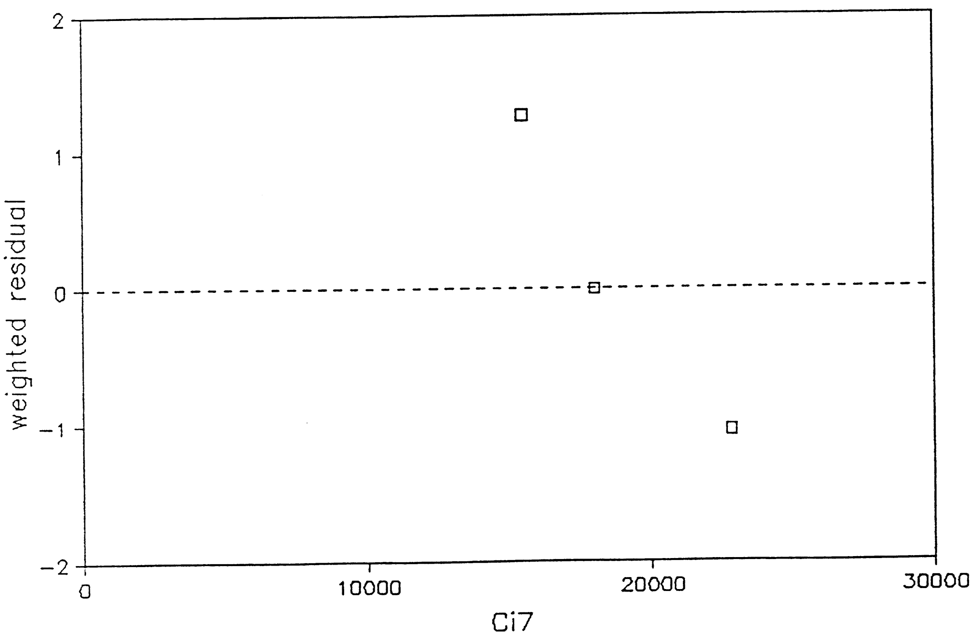
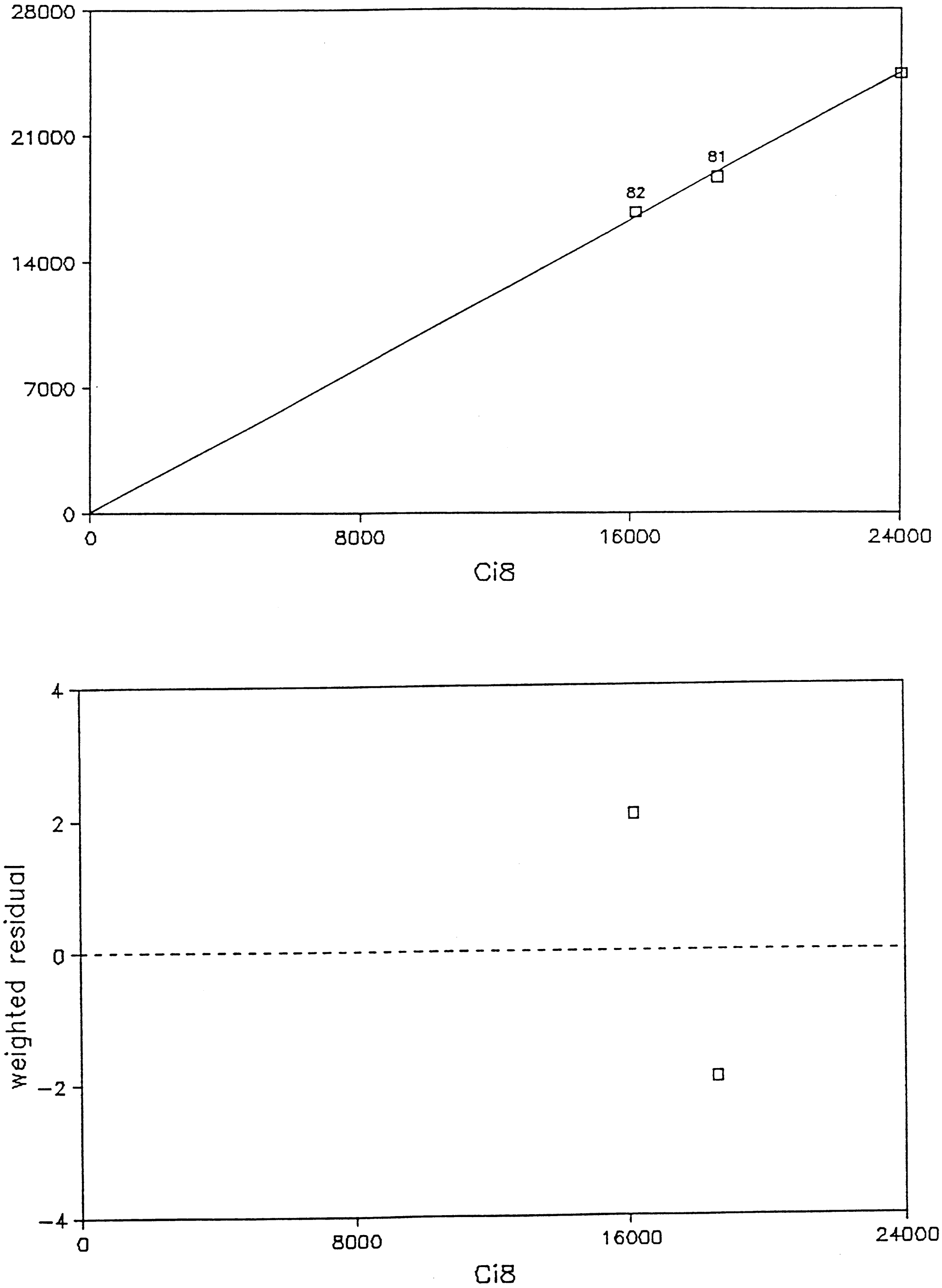


Figure 8: Regression and Residuals

Ci9 against Ci8



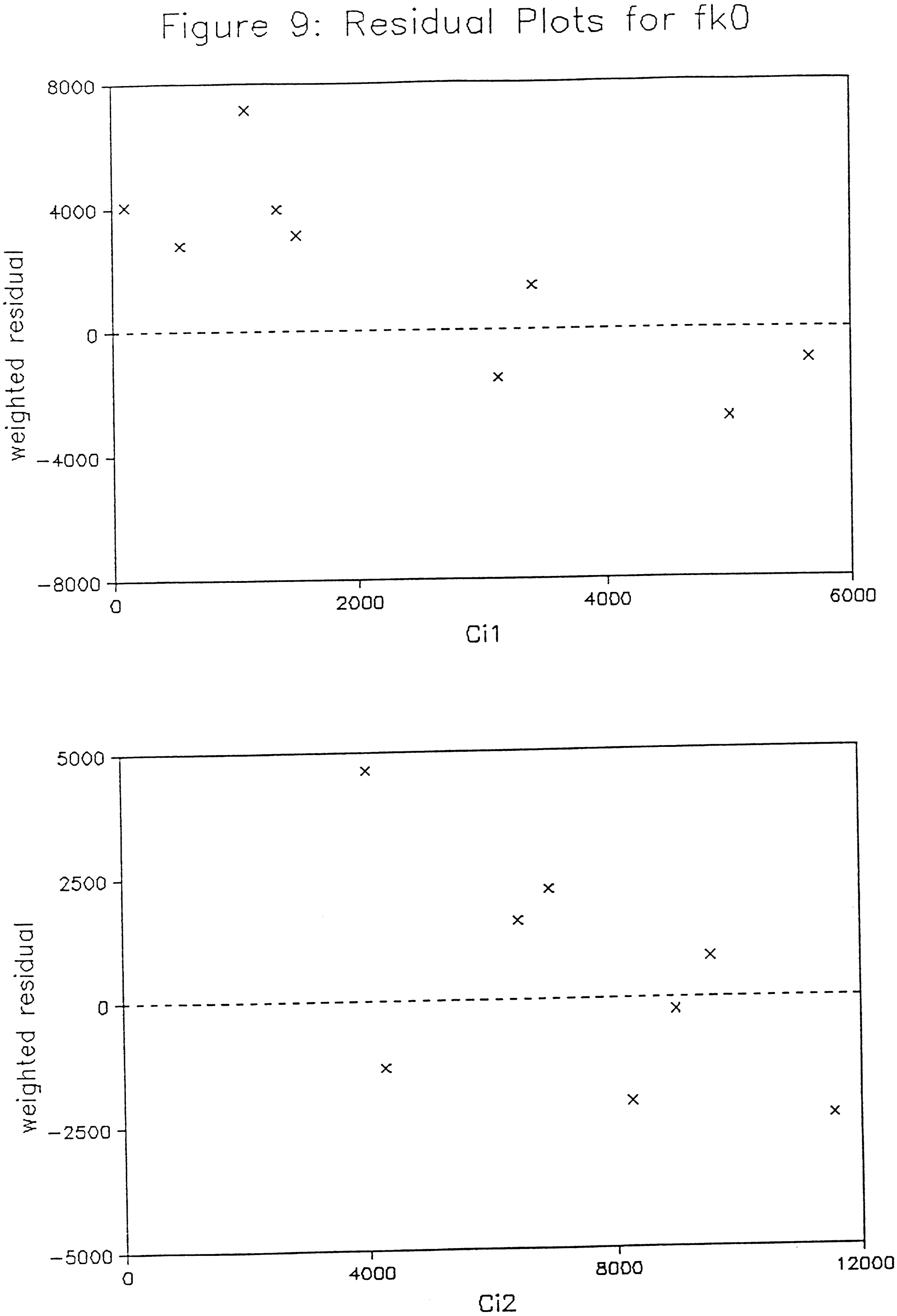


Figure 9: Residual Plots for fk0
8000
4000
0
-4000
-8000
weighted residual
0
2000
4000
6000
Ci1
5000
2500
0
-2500
-5000
weighted residual
0
4000
8000
12000
Ci2

Figure 10: Residual Plots for fk0

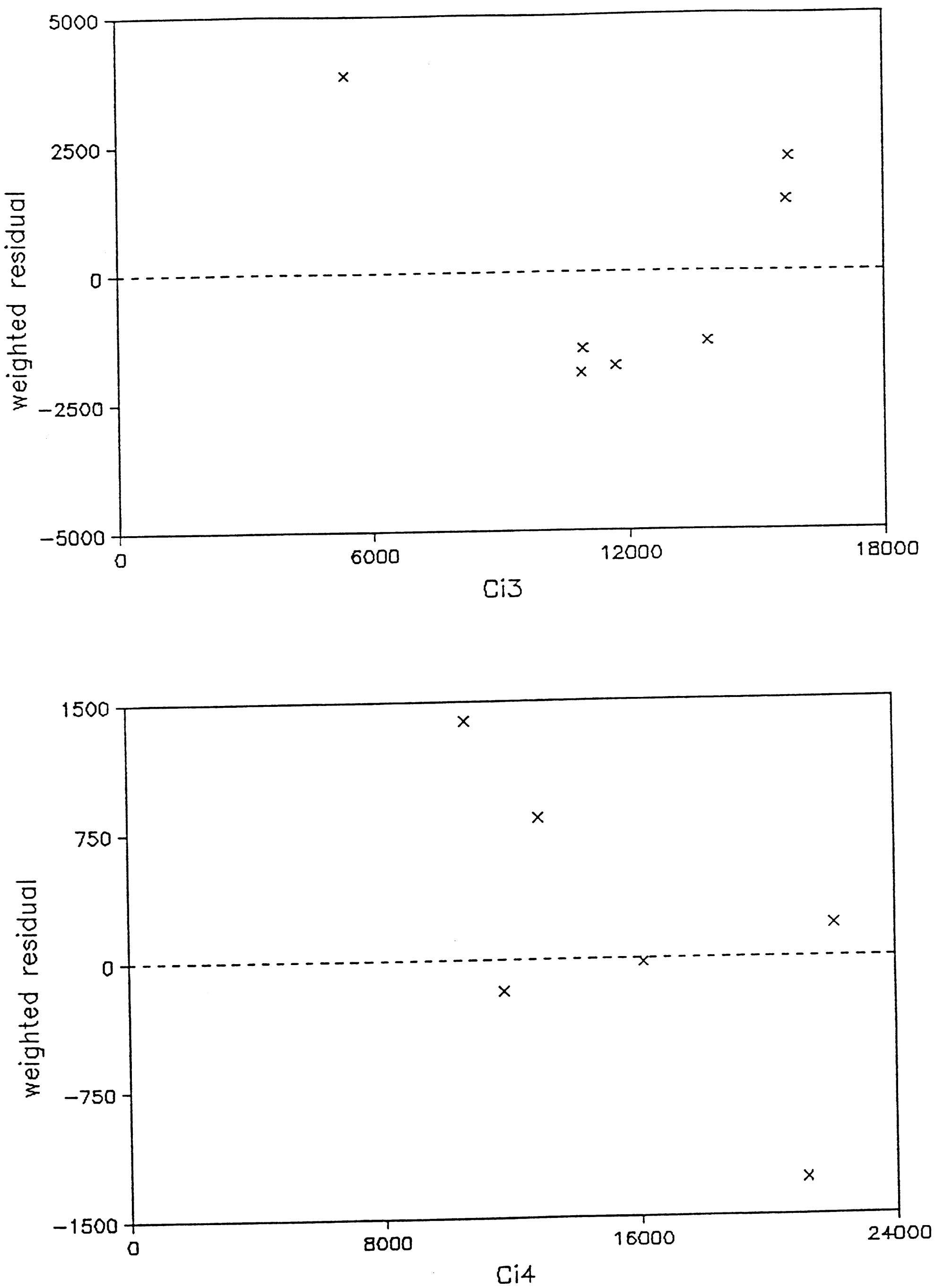


Figure 11: Residual Plots for fk2

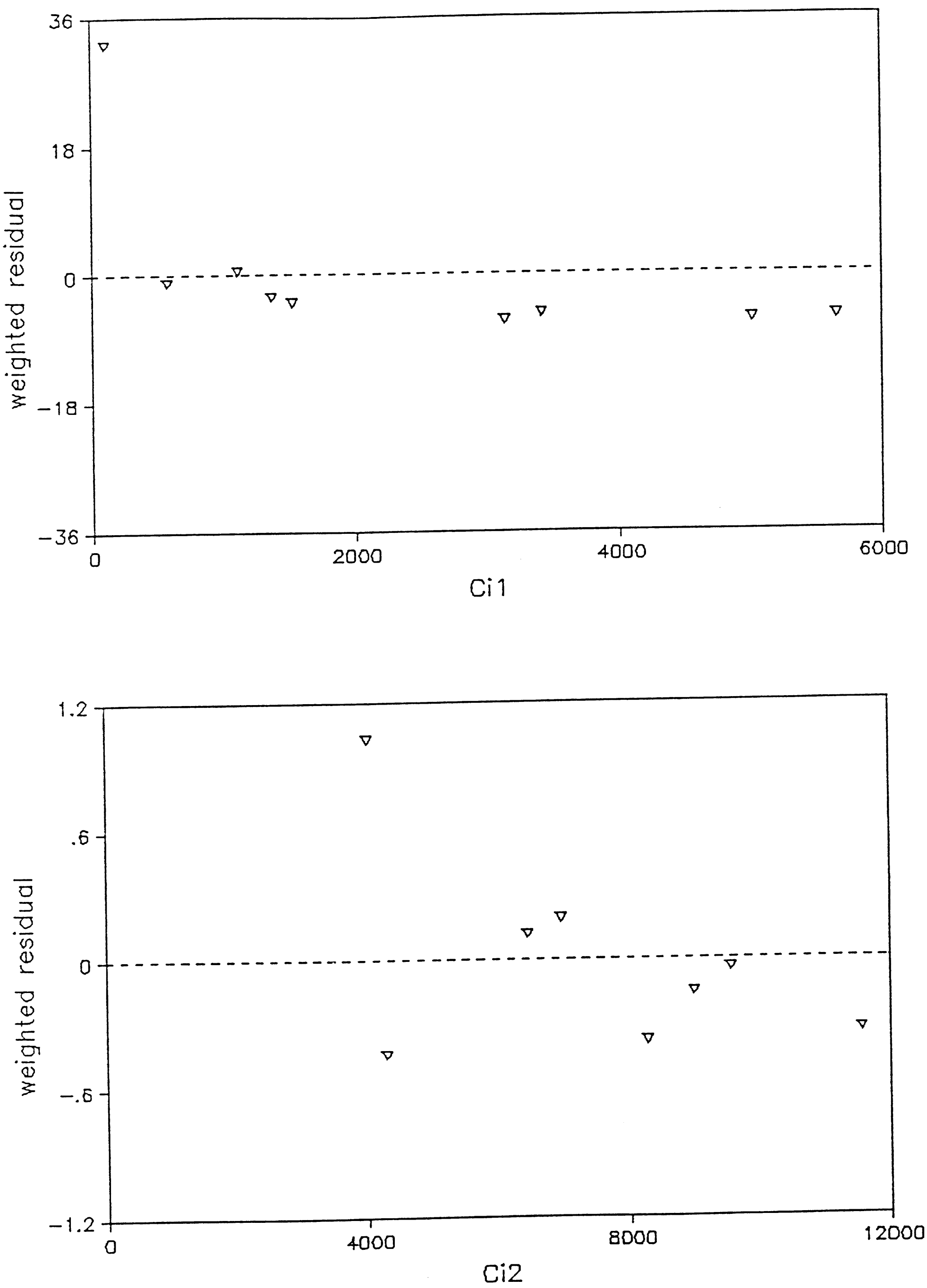
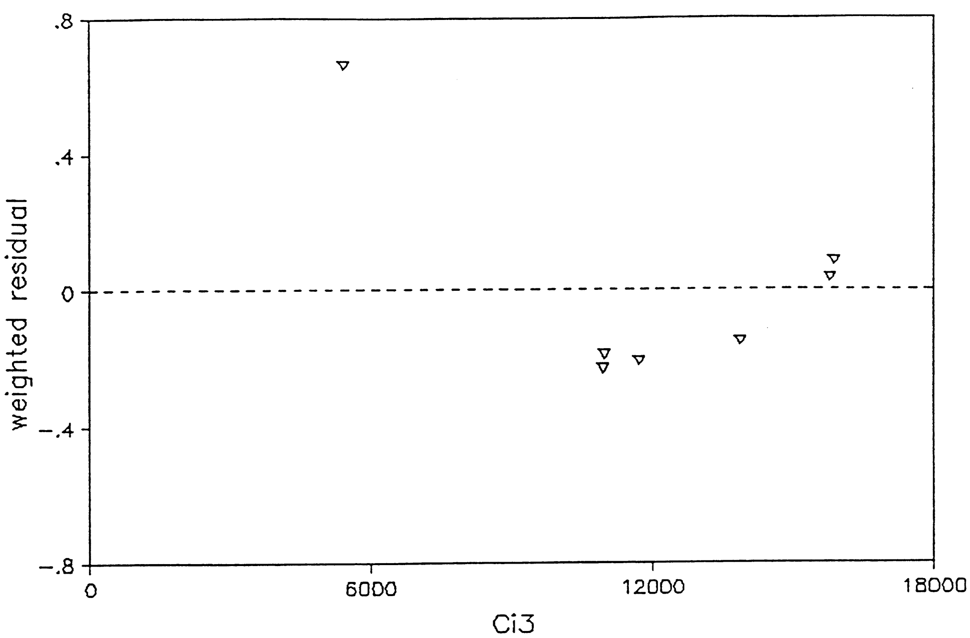


Figure 12: Residual Plots for fk2



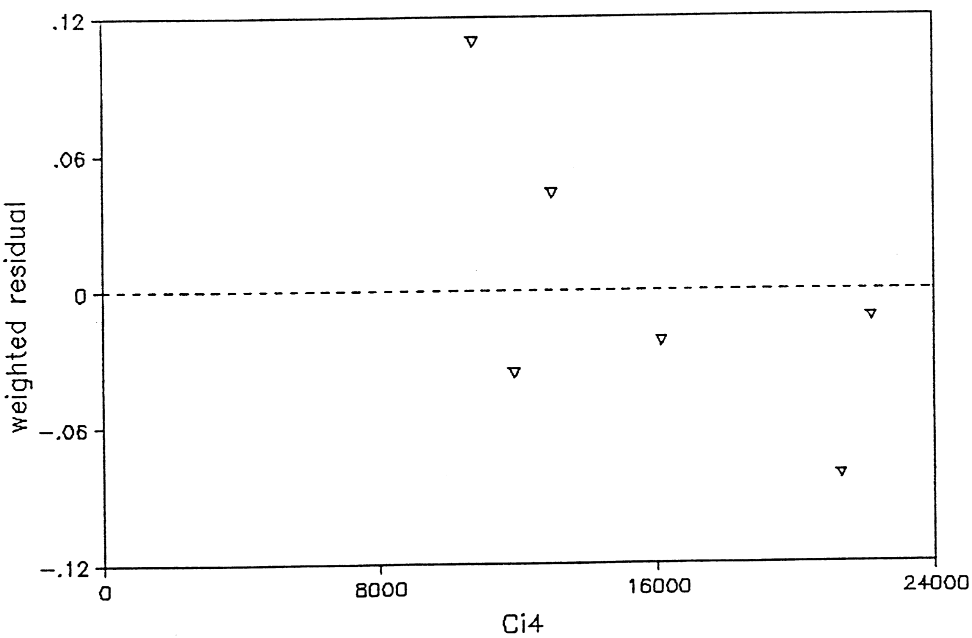


Figure 13: Plot of $\ln(\alpha_k^2)$ against k

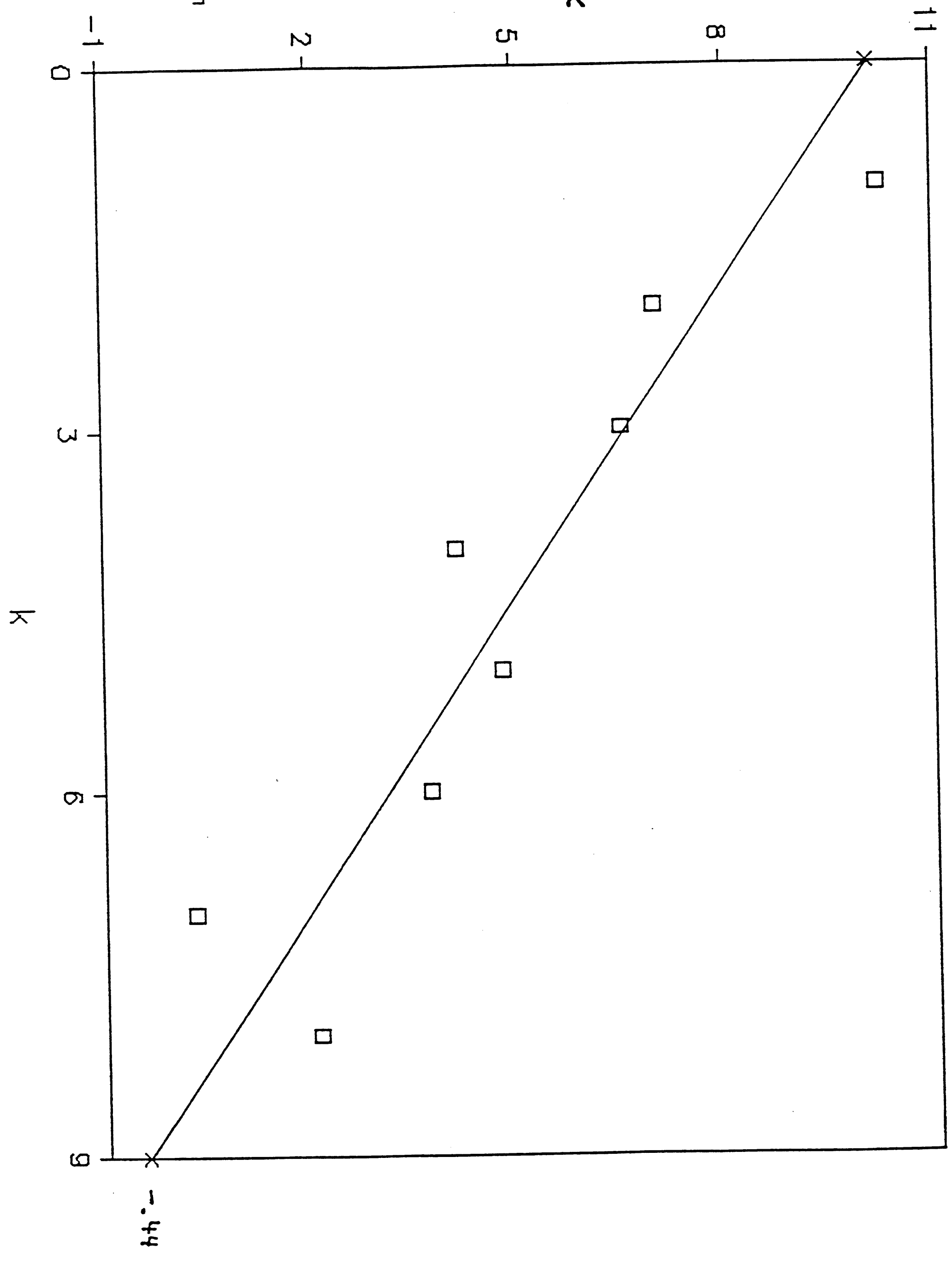
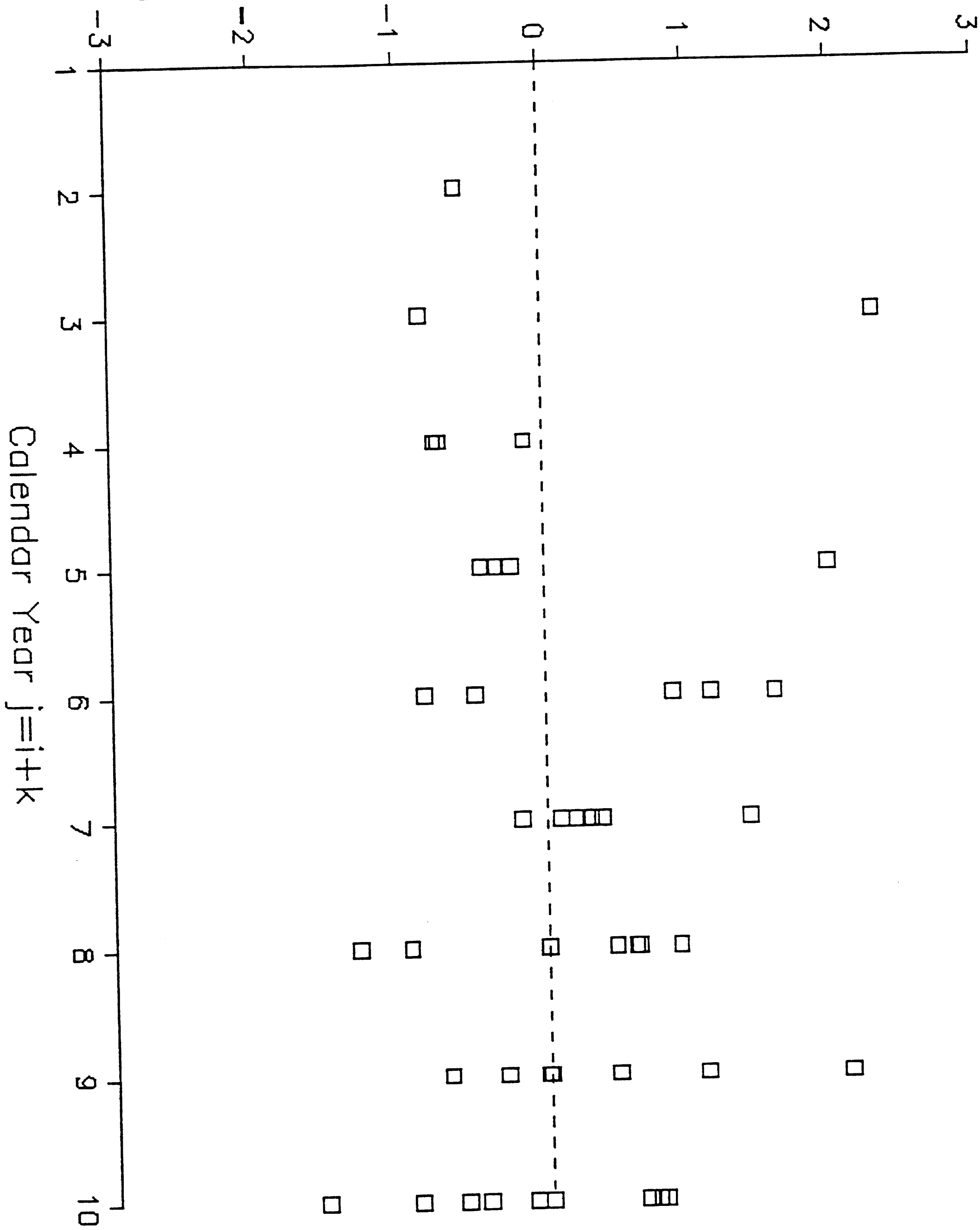


Figure 14: Std.Residuals vs. Calendar Year



Section E
PRÉCIS OF OTHER ACTUARIAL PAPERS

This section provides a series of précis of several other papers published since the first edition of the Claims Reserving Manual was produced. The intention is to give the reader an overview of the paper, together with a description of the reserving model on which the paper is based. A few observations are also made about the applicability of the model, what data are required, what level of statistical and computational ability is needed, plus some thoughts on the strengths and weaknesses of the model.

The first three of these papers provide variations on the theme of regression models based on log-incremental payments. The paper by R J Verrall and Z Li gives a suggestion to overcome the problem of negative incremental payments. The paper by R J Verrall uses the log-incremental regression model as a basis for allowing the practitioner to enter prior information, or to estimate the parameters dynamically. A Bayesian method is used and the data are analysed recursively, using the Kalman filter. Finally, the paper by B Zehnwirth sets out a framework based on the log-incremental payments. His models include systematic components by development year, accident year and calendar year, as well as a random component.

The next two papers make use of more detailed claims information than just aggregate claims payments. The paper by T S Wright sets out a comprehensive approach using Generalised Linear Models to fit Operational Time models. These allow estimates of reserves and different components of reserve variability to be produced. The paper by D H Reid extends the basic Operational Time concept to allow for sudden changes in the nature and mix of business.

The paper by D M Murphy examines the standard link-ratio methods from the point of view of classical regression theory, and considers the circumstances under which the standard link-ratio methods may be considered optimal.

The final, brief, paper by D Gogol describes an approach to estimating loss reserves, using recent loss experience and two probability distributions. The distributions are of the ultimate losses, based on prior experience and rate adequacy changes, and the ratio of the estimator based on recent experience to the true ultimate loss.

[E1]
NEGATIVE INCREMENTAL CLAIMS:
CHAIN LADDER AND LINEAR MODELS
By R J Verrall and Z Li
(13 pages)
***Journal of the Institute of Actuaries,* Vol. 120, p. 171 (1993)**

Summary

One of the problems of many models based on log-incremental payments is the inability to deal with negative incremental payments. One approach to this problem is to add a suitably large arbitrary constant to all the payments, and then subtract the constant after the forecasts are made.

The paper shows that the addition of such a constant (a threshold parameter) is equivalent to modelling the incremental payments by a three parameter log-normal distribution, for which the choice of constant can be performed by maximum likelihood estimation rather than arbitrarily.

Description of the model

The basic model, based on log-incremental payments (see for example the Christofides paper in Section D5 of Volume 2), is adjusted as follows:

$$\mathrm{Log}(P_{ij}+c) = Y_{ij} = a_i + b_j + e_{ij}$$ (e_{ij} are independent identically distributed normal random errors)

where P_{ij} are the incremental payments for accident year i, development period j, and c is the threshold parameter.

Standard procedures for producing maximum likelihood estimates yield a set of equations that can be solved iteratively for a_i, b_j and c. The technique could also be applied to other models for Y_{ij}. The implications for the standard errors of the estimated future payments are not discussed.

General comments

As for most models based on log-incremental payments, the technique is not restricted to any particular class of business, and the only data required are incremental payments.

The technique gives a theoretically sound solution to the problem of negative incremental payments, rather than relying on arbitrary adjustments to the data. The user should, however, examine the sensitivity of the results to the level of threshold parameter used.

The paper requires a basic level of statistical knowledge. The authors include a worked example, but the steps in the iterative process to calculate the parameters are not spelt out. Familiarity with matrix manipulation and regression in a spreadsheet is therefore essential.

<>

[E2]

A STATE SPACE REPRESENTATION OF THE CHAIN LADDER MODEL

By R J Verrall

(21 pages)

***Journal of the Institute of Actuaries,* Vol. 116, p. 589 (1989)**

Summary

The model treats the development triangle as a dynamic system, with development taking place over time, t, in the direction of the diagonal (calendar year). A recursive relationship between the parameters at time t and t+1 is developed, with the ability to enter prior information. The recursive estimation of the parameters is based on a process known as the Kalman Filter.

Description of the model

The basic model begins with the familiar:

$\text{Log}(X_{ij}) = Y_{ij} = \mu + a_i + b_j + e_{ij}$ (e_{ij} are independent identically distributed Normal random errors)

where X_{ij} are the incremental payments for accident year i, development period j.

The model then becomes quite unfamiliar as it defines:

The Observation equation, $\mathbf{Y_t} = F_t \,.\, \boldsymbol{\theta}_\mathbf{t} + \mathbf{e_t}$

The System equation, $\boldsymbol{\theta}_\mathbf{t+1} = G_t \,.\, \boldsymbol{\theta}_\mathbf{t} + H_t \,.\, {}_\mathbf{t} + \mathbf{w_t}$

The bold symbols denote vectors. For example, $\mathbf{Y_t}$ is a vector of the Y_{ij} at time t, and $\mathbf{e_t}$ is a vector of the e_{ij} at time t. $\boldsymbol{\theta}_\mathbf{t}$ is known as the State vector, which is a vector of the parameter estimates (that is estimates of a_i and b_j) at time t. $\mathbf{u_t}$ is a stochastic input vector assumed to be independent of $\boldsymbol{\theta}_\mathbf{t}$, and $\mathbf{w_t}$ is a disturbance vector. F_t, G_t and H_t are matrices. The Observation and System equations together comprise the "State Space representation" of this particular chain ladder model.

When $\mathbf{u_t} = \mathbf{w_t} = \mathbf{0}$, the System equation reduces to $\boldsymbol{\theta}_\mathbf{t+1} = G_t.\boldsymbol{\theta}_\mathbf{t}$, and G_t can be defined such that the parameters at times t and t+1 are equal. This equates to least squares estimation when the parameters are identical for each row and each column.

When $\mathbf{w_t} = \mathbf{0}$, and $\mathbf{u_t}$ has the prior distribution of the new parameters, Bayesian estimates are obtained with distinct parameters.

When $\mathbf{w_t} \neq \mathbf{0}$, the parameters at times t and t+1 are related but not necessarily the same. This is known as dynamic parameter estimation, which in a sense lies in between the two previous cases of identical and distinct parameter estimation.

The paper considers a specific case of the State Space system, where $\mathbf{e_t}$, $\mathbf{u_t}$, $\mathbf{w_t}$ and

$\boldsymbol{\theta}_{t|t-1}$ are independent and normally distributed with defined means and variances, for which the State vector, $\boldsymbol{\theta}_t$, can be calculated recursively by a series of matrix manipulations.

General comments

The paper illustrates by way of examples how prior information and dynamic estimation of parameters can enhance traditional chain ladder methods, and squeeze the maximum amount of information from the available data. The use of a Bayesian approach should lead to greater parameter and predictor stability than ordinary chain ladder models.

The standard of mathematics and computational ability is very high and may be beyond the scope of most people. Whilst numerical examples are given, the intermediate steps in arriving at the results are not, so it may be tricky to replicate the examples. Realistically, anyone wanting to use these methods may be best advised to do so using commercially available software packages, although it is important to understand the theory underlying the model when doing so.

One possible problem when using this type of model is that the assumptions and inputs can become somewhat divorced from reality, including as they do estimates of the variances of parameters of a model of the logs of incremental payments. These are not concepts that are readily translatable to ones knowledge of the payment of claims, and it is not always easy to understand the implications for the future payments of changes in these inputs to the model.

<>

[E3]
PROBABILISTIC DEVELOPMENT FACTOR MODELS WITH APPLICATION TO LOSS RESERVE VARIABILITY, PREDICTION INTERVALS AND RISK BASED CAPITAL
By B Zehnwirth
(159 pages)
***Casualty Actuarial Society Spring Forum,* Vol. 2, p. 447 (1994)**

Summary

The paper describes a statistical modelling framework. Each model contained in the framework has four components. The first three components are the trends in the development year, accident year and calendar year, and the fourth component is random fluctuation (or distribution of the deviations) about the trends. The emphasis of the paper is to focus on the calendar year direction.

The modelling framework is relatively simple, allowing the testing of assumptions (for example, looking at the stability of models) and the quantification of reserve variability.

Description of the model

The family of models includes:

$$\text{Log}(P_{ij}) = Y_{ij} = a_i + b_j + c_k + e_{ij} \quad (e_{ij} \text{ are independent identically distributed Normal random errors})$$

where P_{ij} are the incremental payments for accident year i, development period j, and $k = i + j$.

Models are fitted by using weighted least squares regression. As a result of multi-collinearity, principally due to the non-orthogonality of the calendar year direction with the other two directions, varying parameter models are necessary and are also included in the framework. Other Bayesian approaches are included, which are of particular use if estimates of certain parameters in a parsimonious model are subject to large uncertainties.

General comments

This family of models is an extension of the type of model described by S Christofides in Section D5 of Volume 2, and can be used for a variety of types of business or types of incremental data. Whilst the basic model can be easily programmed in a spreadsheet, the more complex variations are probably beyond the means of most programmers.

As for other models based on the logs of incremental payments, the models do not work for negative incremental payments, and there is a limit in a spreadsheet to the number of future payments that can be predicted.

Most of the paper requires a basic level of statistical knowledge, whilst some of the variations on the basic model require a more advanced level. Familiarity with matrix manipulation and regression in a spreadsheet is required to implement the models.

<>

[E4]
STOCHASTIC CLAIMS RESERVING WHEN PAST CLAIM NUMBERS ARE KNOWN
By T S Wright
(93 pages)
***Proceedings of the Casualty Actuarial Society,* Vol. 79, p. 255 (1992)**

Summary

The model attempts to represent the underlying claims settlement process.

The starting premise is that the cost of settling claims and the order in which they are settled are related — that is, typically, the longer the period to settlement, the greater the final settlement cost is likely to be. The method therefore develops a model of the claim settlement cost as a function of the relative proportion of claims settled (this time-frame is known as Operational Time).

The model is fitted using the theory of Generalised Linear Modelling ("GLMs"). Because it is a statistical model, standard errors (as a measure of the variability of the estimate) for the future incremental payments can be calculated, and statistical techniques used to test the fit of the model.

Description of the model

Operational Time (τ) is the number of claims closed to date, expressed as a proportion of the ultimate number of claims. The mean claim size, $m(\tau)$, can be modelled by a wide variety of different types of function of τ. For example:

$$m(\tau) = \exp(\beta_0 + \beta_1\tau + \ldots..\beta_n\tau^n)$$

Alternatives include polynomial functions of τ, or functions such as $\beta_n\tau^{-n}$, or some combination of these functions. The parameters of the model are fitted using GLMs, for example using the software package GLIM.

The modelling technique involves fitting a basic model that adheres closely to the data, then examining alternative models. The nature of the model means that familiar measures of goodness of fit, such as "sums of squares", are not appropriate. An alternative measure, deviance, is therefore considered, as well as other indications as to the goodness of fit of the model.

Certain restrictive assumptions are made at the initial fitting stage, which are then examined and may subsequently be relaxed.

General comments

The method is likely to be of most use where the greatest cause of uncertainty in predicting ultimate claims is due to individual claim costs — for example, classes involving bodily injury claims. It should also be of particular use when it is believed that settlement rates are changing, as the model may be able to capture these changes more effectively than traditional techniques.

It requires data on both the amount and number of claims settled.

The model is sensitive to the estimated future number of settled claims, and these estimates need careful scrutiny. Inflation is a parameter that may be modelled, and this is also an area where close scrutiny is required. The approach to comparing different functions for $m(\tau)$ is open to some criticism, as the comparison of non-nested models using deviances is not strictly valid — though the author recognises that this is a pragmatic approach.

A high degree of statistical knowledge is required to implement and understand the model, as well as considerable computer literacy. Knowledge of a GLM software package such as GLIM is essential.

<>

[E5]

OPERATIONAL TIME AND A FUNDAMENTAL PROBLEM OF INSURANCE IN A DATA-RICH ENVIRONMENT

By D H Reid

(13 pages)

***Applied Stochastic Models and Data Analysis,* Vol. 11, No. 3, Wiley (1995)**

Summary

This paper is the latest in a series developing a particular approach to claims reserving where relatively complete information on individual claims is available, and where past years' claims patterns are relevant — albeit with modifications — to the development of more recent years' experience.

Specifically, this paper addresses a problem which has arisen in recent years, where relatively rapid changes in size and factor mix of the claims portfolio are taking place. Most, if not all, previous claim reserving methodologies have implicitly assumed that factors change relatively slowly, to such an extent that the effect of this trend on claim development is not significant.

The present paper, by contrast, models the effect of factor trends explicitly, both on the level of claim cost itself and on the development patterns. By doing so, it proposes an approach which may then be applied directly to the development of premium rates, as well as reserves.

Description of the model

The model proposed is based upon that described in Section D4 of Volume 2, but develops that model to allow for the incorporation of a rating factor or set of classificatory factors into the analysis. This is done by first elaborating the structure of claim cost development for recent years as represented by the original model, and then introducing an approach which makes the resulting complex picture more readily comprehensible and, at the same time, statistically estimable.

General comments

Given that this methodology is intended for situations where significant resources are available for claims modelling, and where it is important to achieve as close an understanding of the claims development process as possible, the proposed methodology is relatively flexible and can be adapted to a wide range of situations.

<>

[E6]

UNBIASED LOSS DEVELOPMENT FACTORS

By D M Murphy

(60 pages)

***Casualty Actuarial Society Spring Forum,* Vol. 1, p. 183 (1994)**

Summary

Standard link ratio methods are examined from the viewpoint of classical regression theory. The circumstances under which the standard link ratio methods could be considered optimal are discussed. Formulae for variances of, and confidence intervals around, point estimates of ultimate loss and loss reserves are derived. A triangle of incurred losses is used to demonstrate the techniques.

A summary of a simulation study is presented which suggests that the performance of the link ratio method, using least squares linear estimates, may approach that of the Bornhuetter-Ferguson and Stanard-Bühlmann techniques in some situations.

Description of the model

The estimates of ultimate loss for n accident years are derived using recursion:

$$\hat{M}_1 = \hat{a}_1 + \hat{b}_1 x_{0,0}$$
$$\hat{M}_n = n\hat{a}_n + \hat{b}_n(\hat{M}_{n-1} + x_{n-1,n-1})$$

$x_{i,j}$ denotes the cumulative incurred loss from accident year i, development year j, and

$$\hat{M}_n = E(\sum_{i=0}^{n-1} x_{i,n} | x_{i,i})$$

The variance is given by the sum of the parameter risk and the process risk. Each are defined for n=1, and then recursively for n>1, as follows:

parameter risk

$$\mathrm{Var}(\hat{M}_1) = \frac{\sigma_1^2}{I_1} + \left(x_{0,0} - \overline{x}_0\right)^2 \mathrm{Var}(\hat{b}_1)$$
$$\mathrm{Var}(\hat{M}_n) = n^2 \frac{\sigma_n^2}{I_n} + \left(\hat{M}_{n-1} + x_{n-1,n-1} - n\overline{x}_{n-1}\right)^2 \mathrm{Var}(\hat{b}_n) + b_n^2 \mathrm{Var}(\hat{M}_{n-1}) + \mathrm{Var}(\hat{b}_n)\mathrm{Var}(\hat{M}_{n-1})$$

where:

$$\overline{x}_{n-1} = \frac{1}{I_n} \sum_{i=n}^{N} x_{i,n-1}$$

is the average "x value" and $I_n = N - n + 1$ (assuming a full column in the triangle) is the number of data points in the regression estimate of the n^{th} link ratio.

process risk

$$\mathrm{Var}(E_1) = \sigma_1^2$$

$$\mathrm{Var}(E_n) = n\sigma_n^2 + b_n^2 \mathrm{Var}(E_{n-1})$$

where E_i is an error term.

General comments

A modest level of mathematics is required to follow the paper. Proofs of the theory are relegated to a bulky appendix. The example provided helps the reader to follow the theory by showing practical application of the formulae. The calculation of the least squares estimates and their variances can readily be done in most spreadsheet packages.

The use of the confidence intervals depends on whether the assumptions made regarding the probability distribution of the error terms are appropriate. The paper does not address how these should be tested.

The Benjamin-Eagles paper in Section D3 of the Manual describes a method which is the same as the least squares linear method described in this paper, but without the mathematical rigour.

The Stanard-Bühlmann technique (also known as the "Cape Cod Method") is not explained. Reference would need to be made to the paper by J Stanard in the 1985 Proceedings of the CAS (Casualty Actuarial Society) for explanantion.

<>

[E7]
USING EXPECTED LOSS RATIOS IN RESERVING
By D F Gogol
(3 pages)
***Casualty Actuarial Society Fall Forum*, p. 241 (1995)**

Summary

The paper describes an approach to estimating loss reserves using the recent loss experience and two probability distributions. The first distribution is that of the ultimate losses for the recent period, based on prior experience and rate adequacy changes. The second distribution is that of the ratio of the estimator based on recent experience to the true ultimate loss.

Description of the model

The model is:

$$h(x|y) = g(y|x)f(x) \Big/ \int_0^\infty g(y|x)f(x)dx$$

where, for losses in respect of an exposure period E:

> f(x) is the probability density function of the distribution of ultimate losses for exposure period E, prior to considering the losses for exposure period E.
>
> g(y|x) is the probability density function of the distribution of y, the developed losses at the point of time under consideration, for exposure period E, given that the ultimate losses are x.
>
> h(x|y) is the probability density function of the distribution of the ultimate losses, given that the developed losses are y.

The functions f(x) and g(y|x) are estimated, and the mean of the distribution given by h(x|y) is the estimate of ultimate losses. For certain choices of f(x) and g(y|x), an explicit formula for the mean of h(x|y) is known, for example when f(x) and g(y|x) are both log-normal.

The paper compares the Bayesian estimate of the ultimate loss ratio with the actual developed loss ratio and the Bornhuetter-Ferguson estimate of the ultimate loss ratio.

General comments

The model is particularly useful for recent accident years and for lines of business with slow development. The model should be capable of fairly easy implementation in most spreadsheet packages.

A modest level of statistical knowledge is required. One approach to estimating the distributions given by f(x) and g(y|x) is to assume f(x) and g(y|x) are of a known type, such as log-normal, and estimate their means and variances to obtain the parameters of the distributions. To do this, a certain amount of judgment may be needed, as the estimates will usually have to be based on somewhat limited information. Thus, although the model provides a rigorous way of incorporating prior information, some of the information used in applying the model may be rather unreliable.

<>

Section F
COMPUTERISED ILLUSTRATION OF VOLUME 2 PAPERS

Accompanying this revision of the Claims Reserving Manual is a disk illustrating the application of two of the methods described in Volume 2. The methods included are those described in Sections E5 and E6 of Volume 2, by S Christofides and T Mack respectively.

Both spreadsheet programs on this disk are solely for illustration, and are intended to help the user understand better the mechanics of performing the methods described in the papers. They are designed to replicate exactly the calculations shown in those papers. This will allow the user to follow the intermediate steps, and assist in understanding how the methods can be applied in practice.

Note that the spreadsheets are simply a mechanical reproduction of the particular calculations illustrated in the two papers. As such, they have not been designed to be used as a generalised reserving tool on other data. Readers should not, therefore, attempt to substitute their own data into this software for practical reserving purposes.

<>

[F1]
COMPUTERISED ILLUSTRATION (1) — REGRESSION MODEL BASED ON LOG-INCREMENTAL PAYMENTS
by S Christofides

The first file on the disk distributed with the Claims Reserving Manual demonstrates the model described in the paper by S Christofides in Section D5 of Volume 2. The filename is **crmsc.xls**, and is written in Excel version 5.

The file illustrates step-by-step the “full parameter” example given in pages D5.16 to D5.33 of the paper. The paper sets out clearly all the steps involved. Further brief instructions are included on the disk as to the operation of the spreadsheet regression analysis and matrix manipulation, so no further instructions are felt necessary here.

The spreadsheet also includes graphs of the various Residual analyses.

<>

[F2]
COMPUTERISED ILLUSTRATION (2) — MEASURING THE VARIABILITY OF CHAIN LADDER RESERVE ESTIMATES
by T Mack

The Lotus version 3 file, **crmmack.wk3**, on the disk distributed with the Claims Reserving Manual, demonstrates the model described in the paper by T Mack in Section D6 of Volume 2.

The file illustrates the calculation of the standard errors of the reserve estimates, and the use of a variety of diagnostics to test the assumptions made when using the model.

To make the calculation of the standard errors easier to follow, the calculations from the example in Section D6, on pages D6.19 to D6.24, have been broken down into small sections, for ease of reference. This should assist the user in seeing how the techniques can be applied in practice, as the formulae for calculating the standard errors, whilst being quite simple, do look a bit daunting at first sight.

The examples of some of the diagnostic tests are also based on the examples included in section 5 and Appendix H of Mack's paper. The diagnostics involve checking the three assumptions made when using the model. For a summary of the assumptions made, see the précis of this paper given in Section C of Volume 2.

The checks of the three assumptions are briefly described and illustrated below.

Checking Assumption 1

One way of checking assumption 1 is simply to conduct a visual examination of the data, to see if there is a consistent linear relationship between cumulative claims from one period to the next. A further way of checking the assumption is to use regression diagnostics, as explained in the section on checking assumption 3.

The attached tables and graphs are reproduced from the spreadsheet, and illustrate a visual examination of the data and the standardised residuals used to check assumption 1. When checking the residuals, if the model holds good, one expects to see the residuals randomly scattered, without any systematic patterns or distortions.

The diagnostic checks shown here correspond to those in Figure 1 of the paper. They differ slightly in that they plot incremental payments on the Y-axis, and examine the standardised residuals. Both types of diagnostic check are equally valid, and are just two ways of looking at the same thing.

INSERT PAGE 20 OF THE PAPER

Checking Assumption 2

One possible distortion that may invalidate assumption 2 is the presence of calendar year influences in the data. If there are such calendar year influences (for example, increasing payments in just one calendar year due to a new type of tax), then consecutive sets of development factors will be larger/smaller than expected. It is possible, however, to construct a statistical test to see whether there are diagonals with a preponderance of "Large" or "Small" development factors.

For each development period, k, the development factors are ordered and described as "L" or "S", depending on whether they are larger or smaller than the median. Then, for each of the j different diagonals, the numbers of L or S factors are counted. The actual median development factor, if we are looking at an odd-numbered set of factors, is described as "M", and is excluded from the subsequent construction of the test statistics.

In the absence of any calendar year effects, the number of L's and S's should be about the same. Similarly, the minimum of the number of L's and S's, described as Z_j, should not be significantly different from the average number of L's plus S's. The paper shows how the distribution of Z_j can be calculated and used for a significance test. Where this test indicates the presence of a calendar year effect, it is suggested that the weights of the relevant outlying development factors are reduced.

Alternatively, one can construct a formula for the first two moments of Z_j, which are:

$$E\left(Z_j\right)=\frac{n}{2}-\binom{n-1}{m}\frac{n}{2^n}$$

$$\mathrm{Var}\left(Z_j\right)=\frac{n\times(n-1)}{4}-\binom{n-1}{m}\frac{n\times(n-1)}{2^n}+E\left(Z_j\right)-E\left(Z_j\right)^2$$

Looking at individual diagonals may be misleading, so one considers $Z = Z_2 + \ldots + Z_{I-1}$. The expected value and variance of Z are just the sums of the individual expectations and variances of the Z_j's respectively (under the initial assumptions, the Z_j's are uncorrelated). Assuming Z is Normal, it can be concluded that there is no significant calendar year affect, at a 95% confidence level, if the actual Z is within two standard errors of the expected Z.

Assumption 2 can also be checked by the use of Residual diagnostics, as described in the section on checking assumption 3. The following examples of some of these tests are from the **crmmack.wk3** spreadsheet.

The statistical test illustrates the alternative approach, based on the first two moments of Z_j. Whilst the examples illustrate the application of these tests, they also show the difficulties in applying statistical tests to the quite small volumes of data one is invariably considering with such reserving exercises.

In the diagnostics shown below, the triangle of "S" and "L" factors does not exhibit any statistically significant calendar year effect, although some columns do show a sharp change from "S" to "L", and vice versa, as one runs down the accident years.

INSERT PAGE 22 OF THE PAPER

INSERT PAGE 23 OF THE PAPER

Checking Assumption 3

The underlying assumptions described so far have been based on using the volume-weighted chain-ladder. The chosen estimate of the development factor is, as the name implies, a weighted average of the actual development factors, where the weights are the measures of claims volume — namely the cumulative claims to date at the relevant points. Other weights of the actual development factors could be used — indeed, any set of weights that sums to one is applicable. In each case, it can be shown that they produce an unbiased estimator of the development factor.

Assumption 3 is derived by T Mack by noting that, out of a collection of unbiased estimators, one prefers the estimator with the smallest variance. Hence the weights are chosen so that the variance is minimised — this can be shown to be the case if and only if the weights are inversely proportional to $\mathrm{Var}(C_{i,k+1}/C_{i,k}|C_{i1}, ..., C_{ik})$. So, as the volume-weighted chain-ladder uses weights that are proportional to C_{ik}, this corresponds to assuming that C_{ik} is inversely proportional to $\mathrm{Var}(C_{i,k+1}/C_{i,k}|C_{i1}, ..., C_{ik})$. Put the other way round, and noting that $\mathrm{Var}(X/a) = \mathrm{Var}(X)/a^2$, gives us our assumption 3, namely

$$\mathrm{Var}(C_{i,k+1}|C_{i1}, ..., C_{ik}) = C_{ik}\sigma_k^2.$$

Other weights could just as easily be used to arrive at the development factors. For example, the simple average of the development factors could have been used. Another alternative is the C_{ik}^2-weighted average. To distinguish the alternative versions of the estimator of the development factors, the volume-weighted estimator is denoted as f(k,1), the simple average as f(k,2) and the C_{ik}^2-weighted average as f(k,0). The results for the estimates of the variance of the reserves can be extended to encompass these different ways of arriving at the development factors.

If the assumption about the variance is reasonable, one can look at the residuals for different types of estimator of f_k, and see which, if any, shows the most random behaviour. A check of this assumption involves plotting the residuals for the three possible types of weight used for the different f_k, for all k. Examples of these plots are given for k=1 on the spreadsheet, and are reproduced below.

With the small number of data points typically present when making chain-ladder reserve estimates, it is hard to form any meaningful conclusions. Nevertheless, if the plots for f(k,1), corresponding to the volume-weighted chain-ladder used as the basis for assumption 3, look non-random, and one of the plots for the alternative weights does not, then one might question whether the variance assumption 3 is reasonable. One might then consider using alternative weights when making our estimates of the chain-ladder factors.

These diagnostic checks are illustrated on the attached extract from the **crmmack.wk3** spreadsheet. They correspond to the diagnostic checks described in section D5 of the paper, and set out the intermediate steps necessary to calculate the Residual diagnostics used to check assumption 3.

INSERT PAGE 25 OF THE PAPER

APPENDIX — ACTUAL & MODELLED NUMBERS OF CLAIMS BY SIZE (ACTUAL DATA)

CLAIM SIZE	YEAR OF ORIGIN														
	1983					1984				1985			1986		1987
	YEAR OF DEVELOPMENT					YEAR OF DEVELOPMENT				YEAR OF DEVELOPMENT			YEAR OF DEVELOPMENT		YEAR OF DEVELOP-MENT
	0	1	2	3	4	0	1	2	3	0	1	2	0	1	0
£0	39,052	14,006	964	192	76	43,971	16,561	1,299	280	51,167	18,919	1,712	58,257	21,218	62,522
£25	37,285	13,407	920	181	69	42,140	15,859	1,236	270	49,496	18,284	1,632	56,728	20,627	61,101
£100	20,991	10,066	792	168	62	24,616	12,086	1,071	245	30,599	14,356	1,437	36,973	16,883	42,290
£200	12,506	7,201	658	141	54	14,593	8,841	909	217	18,295	10,732	1,193	21,837	12,710	25,081
£500	5,097	3,735	442	113	46	6,114	4,625	600	170	7,696	5,730	828	8,878	6,714	9,772
£1,000	2,138	1,948	326	96	42	28	211	326	119	64	3,039	581	3,891	3,660	4,312
£1,500	1,235	1,248	261	83	38	1,456	1,546	334	120	1,916	1,966	457	2,298	2,422	2,520
£2,000	802	840	212	74	36	920	1,286	273	105	1,264	1,373	371	1,544	1,697	1,655
£3,000	340	414	138	62	32	417	541	187	83	613	760	247	867	949	908
£4,000	177	236	96	42	28	211	326	119	64	316	388	164	527	577	522
£5,000	88	140	67	34	24	123	204	84	48	206	231	122	336	356	344
£6,500	41	79	42	26	20	68	97	56	41	122	118	70	188	192	183
£8,000	29	43	31	22	16	37	60	42	31	61	69	49	109	106	95
£10,000	14	19	22	18	13	21	34	22	17	33	35	34	57	61	42
£15,000	5	6	11	9	9	4	9	7	8	9	12	10	20	14	15
£20,000	2	1	7	3	5	0	5	6	3	3	3	7	8	4	4
£25,000	1	0	5	3	3	0	3	4	2	2	1	3	4	2	3
£30,000	1	0	5	2	3	0	2	3	0	1	1	3	3	2	2
£40,000	0	0	2	0	2	0	1	2	0	0	1	2	1	1	0
£50,000	0	0	1	0	2	0	1	1	0	0	0	1	1	1	0
£65,000	0	0	0	0	2	0	0	1	0	0	0	1	1	1	0
£80,000	0	0	0	0	1	0	0	1	0	0	0	1	0	1	0
£100,000	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
£150,000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
£200,000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

APPENDIX — ACTUAL & MODELLED NUMBERS OF CLAIMS BY SIZE (MODELLED DATA)

CLAIM SIZE	YEAR OF ORIGIN														
	1983					1984				1985			1986		1987
	YEAR OF DEVELOPMENT					YEAR OF DEVELOPMENT				YEAR OF DEVELOPMENT			YEAR OF DEVELOPMENT		YEAR OF DEVELOP-MENT
	0	1	2	3	4	0	1	2	3	0	1	2	0	1	0
£0	39,052	14,006	964	192	76	43,971	16,561	1,299	280	51,167	18,919	1,712	58,257	21,218	62,522
£25	37,239	13,336	910	181	73	41,972	15,793	1,228	265	48,906	18,080	1,625	55,712	20,304	59,823
£100	21,106	9,913	786	162	66	24,369	11,857	1,045	240	29,505	13,918	1,399	34,067	15,844	36,915
£200	13,014	7,228	666	138	59	15,260	8,786	871	212	18,962	10,642	1,195	22,088	12,336	24,071
£500	5,158	3,723	465	109	49	6,142	4,572	590	164	7,832	5,699	814	9,202	6,710	10,068
£1,000	2,103	1,948	329	91	42	2,540	2,408	405	133	3,321	3,067	558	3,934	3,653	4,316
£1,500	1,159	1,244	264	80	39	1,408	1,549	317	116	1,863	2009	433	2,214	2,417	2,427
£2,000	723	846	222	71	35	888	1,071	260	104	1,196	1,427	357	1,423	1,717	1,558
£3,000	306	425	155	58	29	391	550	175	84	555	766	250	675	962	753
£4,000	144	241	115	50	25	189	315	126	70	279	448	179	342	566	384
£5,000	72	140	88	42	22	96	192	95	61	146	283	135	183	361	208
£6,500	34	77	64	32	20	43	99	66	48	63	144	93	80	194	92
£8,000	20	47	51	24	17	25	63	51	39	36	91	69	45	115	50
£10,000	12	26	30	18	13	15	35	30	29	21	53	49	26	70	28
£15,000	5	8	12	11	9	7	12	11	16	9	18	19	11	24	12
£20,000	2	3	5	6	6	3	5	5	10	5	8	8	6	11	7
£25,000	1	2	3	4	5	1	2	3	6	2	3	5	3	5	3
£30,000	0	1	1	4	4	0	2	1	4	1	2	3	1	2	1
£40,000	0	0	1	3	2	0	1	1	4	0	1	1	0	2	0
£50,000	0	0	1	1	2	0	0	1	3	0	0	1	0	1	0
£65,000	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0
£80,000	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
£100,000	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
£150,000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
£200,000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0