Best Estimates for Reserves

Glen Barnett and Ben Zehnwirth

17 May, 1998

Abstract

In recent years a number of authors (Brosius, 1992; Mack, 1993, 1994; and Murphy, 1994) have shown that link ratio techniques for loss reserving can be regarded as weighted regressions of a certain kind. We extend these regression models to handle different exposure bases and modelling of trends in the incremental data, and develop a variety of diagnostic tools for testing the assumptions these techniques carry with them.

The new 'extended link ratio family' (ELRF) of regression models is used to test the assumptions made by the standard link ratio techniques and compare their predictive power with modelling (trends in) the incremental data. Not only does the ELRF of regression models indicate that for most, if not all, cumulative arrays the assumptions made by the standard link ratio techniques are not satisfied by the data, but that modelling the trends in the (log) incremental data has more predictive power.

The ELRF modelling structure creates a bridge to a statistical (probabilistic) modelling framework where the assumptions are more in keeping with what we see in actual data. There is a paradigm shift from the standard link ratio techniques to the statistical modelling framework; and the ELRF can be regarded as the bridge from the 'old' paradigm to the 'new'.

There are three (critical) stages involved in arriving at a reserve figure, namely, extraction of information from the data in terms of trends and stability thereof, and distributions about trends; formulation of assumptions about the future leading to forecasting of distributions of paid losses; and correlation between lines and security level sought.

Finally, other benefits of the new statistical paradigm are discussed, including segmentation, credibility and reserves or distributions for different layers.

1 Introduction and Summary

A model that is used to forecast reserves cannot include every variable that contributes to the variation of the final reserve amount. The exact future payment (being a random variable) is unknown and unknowable. Consequently a probabilistic model for future reserves is required. If the resulting predictive distribution of reserves is to be of any use, or have any meaning, the assumptions contained in that probabilistic model must be satisfied by the data. An appropriate probabilistic model will enable the calculation of the distribution of the reserve that reflects both the process variability producing the future payments and the parameter estimation error (parameter uncertainty).

The regression models based on link ratios developed by Brosius (1992), Murphy (1994) and Mack (1993, 1994) are described in Section 2 and extended to include trends in the incremental data, and different exposure bases. We refer to that family of models as the extended link ratio family (ELRF). The ELRF provides both diagnostic and formal tests of the standard link ratio techniques. It also facilitates the comparison of the relative predictive power of link ratios vis-a-vis modelling the trends in the (log) incremental data.

Very often, for real data, even the best model within the ELRF is not appropriate, because the data doesn’t satisfy the assumptions of that model. The common causes of this failure to satisfy assumptions motivate the development of the statistical modelling framework discussed in Section 3. The rich family of statistical models in the framework contains assumptions more in keeping with reality.

In Section 3, a statistical modelling framework, based on the analysis of the log incremental data, is described where each model in the framework has four components of interest. The first three components are trends in each of the directions: development period, accident period and payment/calendar period. The fourth component is the distribution of the data about the trends. Each model fits a distribution to each cell in the loss development array and relates cell distributions by trend parameters. This rich family of models we call the Probabilistic Trend Family (PTF). We describe how to identify the optimal model in the statistical modelling framework via a step by step model identification procedure and illustrate that in the presence of an unstable payment/calendar year trend, formulating assumptions about the future may not be straightforward. The statistical modelling framework allows separation of parameter uncertainty and process variability.

It also allows us to:

1. check that all the assumptions contained in the model are satisfied by the data,
2. calculate distributions of reserve forecasts, including the total reserve,
3. calculate distributions of, and correlations between future payment streams,
4. price future underwriting years including aggregate deductibles and excess layers,
5. easily update models and track forecasts as new data arrive.

The final part of the paper discusses how the combination of information extracted from the data and business knowledge allow the actuary to formulate appropriate assumptions for the future in terms of predicting distributions of loss reserves. Correlations between different lines and a prescribed security level are important inputs into a final reserve figure. Finally, other benefits of the statistical paradigm are alluded to, including segmentation, credibility and pricing different layers.

2 Extended Link Ratio Family

2.1 Introduction

Brosius (1992) points out that the use of regression in loss reserving is not new, dating back to at least the 1950's, and says that using link ratio techniques corresponds to fitting a regression line without an intercept term. Mack (1993) derives standard errors of development factors and forecasts (including the total) for the chain ladder regression ratios. He mentions the connection to weighted least squares regression through the origin, and he presents diagnostics that indicate that an intercept term may be warranted on the data he analyses.

Working directly in a regression framework, Murphy (1994) derives results for models without an intercept, such as the chain ladder ratios, and also models with an intercept.

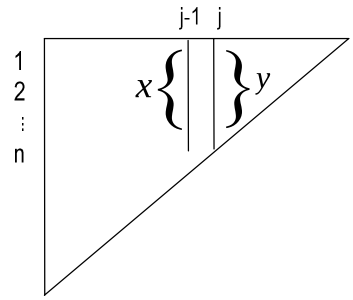
Under the assumption of (heteroscedastic) normality, we derive results for a more general family of models that also include accident year trends for each development year. This extended family we call the Extended Link Ratio Family (ELRF). We discuss calculations and diagnostics for fitting and choosing between models, and checking assumptions. Standard errors of forecasts for both cumulatives and incrementals are also derived.

In the current section we analyse a number of real loss development arrays. Diagnostics including graphs of the data and formal statistical testing both indicate that models based on link ratios suffer several common deficiencies and frequently even the optimal model in the ELRF is inappropriate. Moreover, models based on the log incremental data have more predictive power than the optimal model in the ELRF.

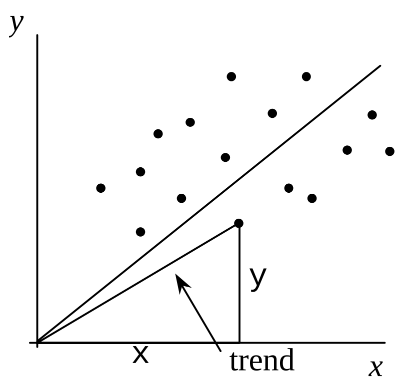
The standard link ratio models carry assumptions not usually satisfied by the data. This can lead to false indications and low predictive power, so that the standard errors of forecasts are meaningless. Hence, we relegate the calculation of standard errors to the Appendix.

2.2 Calculating Ratios using Regressions

Suppose $x(i)$; $i = 1,2,...,n$ represent the *cumulative* at development period $j-1$ for accident periods $i = 1,2,...,n$ and $y(i)$ are the corresponding cumulative values at development period *j*.



A graph of *y* versus *x* may appear as follows.



A link *ratio* $y(i)/x(i)$ is the slope of a line passing through the origin and the point $(x(i), y(i))$. So, each ratio is a trend.

Accordingly, a link ratio (trend) average method is based on the regression

$$y(i) = bx(i) + \varepsilon(i), \qquad \text{Var}[\varepsilon(i)] = \sigma^2 x(i)^\delta \qquad (2.1)$$

The parameter b represents the slope of the 'best' line through the origin and the data points $(x(i), y(i)); i = 1,2,\ldots,n$.

The variance of $y(i)$ about the line depends on $x(i)$, via the function $x^\delta(i)$, where δ is a "weighting" parameter.

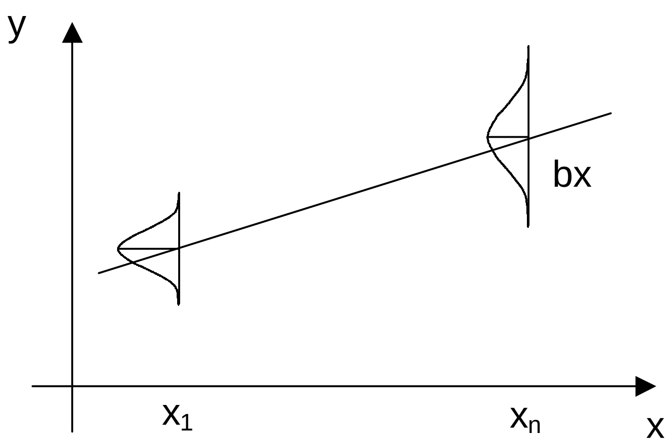


Figure 2.1   Chain Ladder Ratios Regression

In the above figure Var[$\varepsilon(i)$] = $\sigma^2 x(i)^\delta$, where δ = 1. Interestingly, the assumption that conditional on $x(i)$ the "average" value of $y(i)$ is $bx(i)$ is rarely true for real loss development arrays.

Consider the *following* cases:

Case (i): δ=1

The weighted least squares estimator of b is

$$\hat{b} = \frac{\sum x(i) \cdot y(i)/x(i)}{\sum x(i)} . \tag{2.2}$$

This is the weighted average by volume, i.e. the chain ladder average method, or chain ladder ratio.

Case (ii): δ=2

The weighted least squares estimator of b is

$$\hat{b} = \frac{1}{n} \sum y(i)/x(i) . \tag{2.3}$$

This is the simple arithmetic average of the ratios.

Case (iii): δ=0

This yields a weighted average weighted by volume squared.

So, by varying the *parameter* δ we obtain different link-ratio methods (averages).

One of the advantages of estimating link-ratios using regressions is that both standard errors of the average method selection and standard errors of the forecasts can be obtained. Another more important advantage is that the assumptions made by the method can be tested.

One important assumption is that $\varepsilon(i)/\sigma x(i)^{\delta/2}$, i = 1, 2, ..., n are normally distributed. Otherwise, the weighted least squares estimator of b is not necessarily efficient, and the reserve forecasts consequently may be biased for the mean and will have a large variance. The normality assumption can be tested by examining the three diagnostic displays: normal probability plot, Box-plot and histogram of the weighted standardised residuals. The Shapiro-Wiks test based on the normality plot is a formal test.

The link ratio method also makes other assumptions that should always be tested.

Another basic assumption is that

$$E\big(y(i)\big|x(i)\big) = bx(i)\,. \tag{2.4}$$

That is, in order to obtain the mean cumulative at development period j, take the cumulative at the previous development period j-1and multiply it by the ratio. A quick diagnostic check of this assumption is given by the graph of *y(i)* versus *x(i)*. Very often a (non-zero) intercept is also required. See Figure 2.4.

Equation (2.4) can be re-cast

$$E\big(y(i) - x(i)\big|x(i)\big) = (b-1)x(i)\,. \tag{2.5}$$

That is, the mean incremental at development period j equals the cumulative at development period j-1 multiplied by the link ratio b minus 1. What are the diagnostic tests for this assumption?

If the assumption (2.4) is valid, then the weighted standardised residuals versus fitted values should appear random. Instead, what you will usually see is a downward trend depicted in the Figure 2.2 below, representing the chain ladder ratios residuals for the Mack (1994) data. (See Example 1 below).

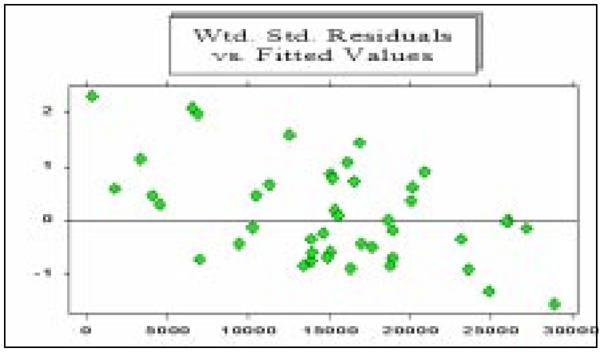


Figure 2.2

This indicates that large values are over fitted and small values are under fitted so that $\mathrm{E}(y \mid x) = bx$ is *not* true.

Comparison of graphs of weighted standardised residuals with graphs of the data will indicate that accident periods that have 'high' cumulatives are over fitted and those with 'low' cumulatives are under fitted. Here are the two displays for the Mack (1994) data. Note that as a result of the equivalence of equations (2.4) and (2.5), the residuals of the cumulative data are also the residuals of the incremental data.

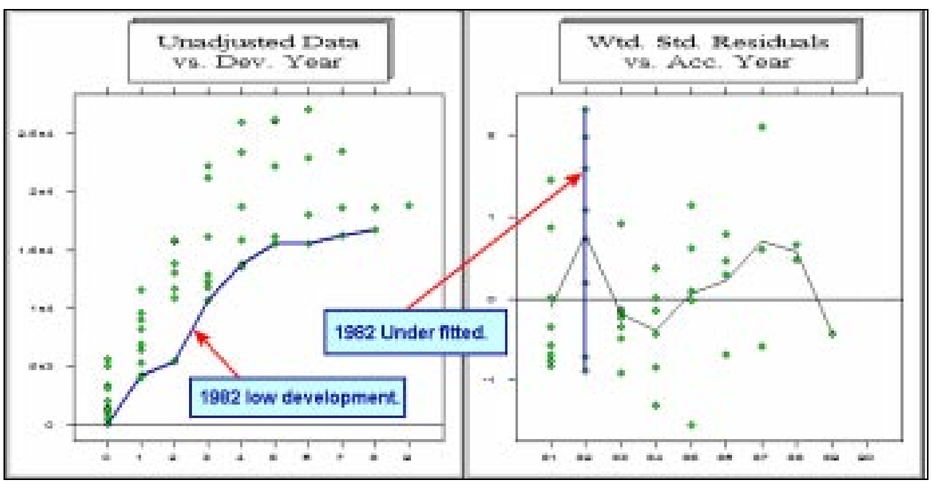


Figure 2.3a

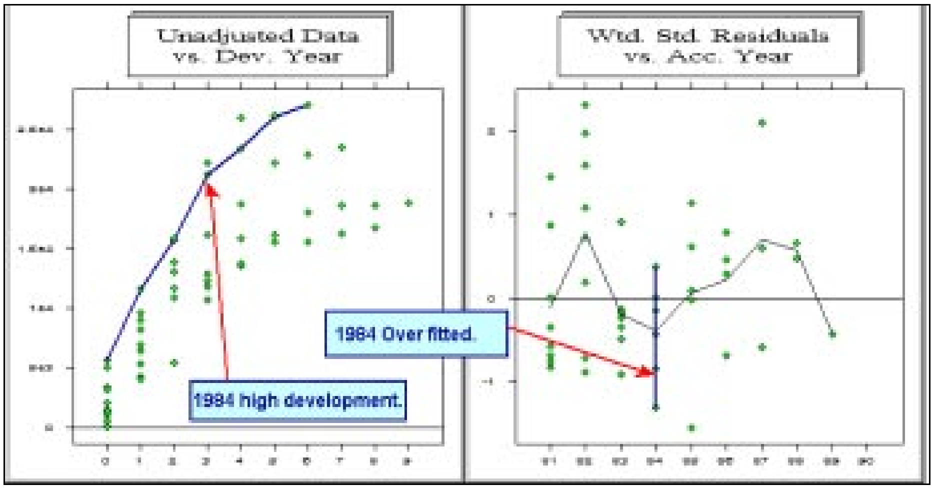


Figure 2.3b

If you think of the way the incrementals are generated and the fact that there are usually payment period effects, the cumulative at development period $j-1$ rarely is a good predictor of the next incremental (after adjusting for trends).

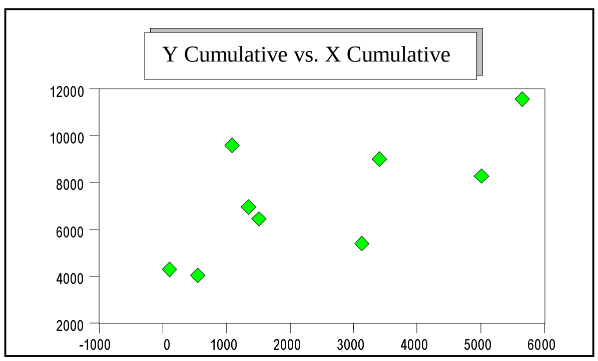


Figure 2.4. Cumulative Development Period 1 versus Cumulative Development Period 0

Murphy (1995) suggested to extend the regression model (2.1) to include the possibility of an intercept.

$$y(i) = a + bx(i) + \varepsilon(i) \tag{2.6}$$

such that $\mathrm{Var}[\varepsilon(i)] = \sigma^2 x(i)^\delta$.

If the intercept "*a*" is significant and we do not include it in the regression model then the estimate of the link ratio *b* (slope) is biased. Note that in the above graph (Figure 2.4) of cumulative at development period 1 versus cumulative at development period 0, the intercept appears significant. Indeed, it is significant between every pair of contiguous development periods. (See the data of Example 1 below).

We can rewrite (2.6) thus:

$$y(i) - x(i) = a + (b-1)\, x(i) + \varepsilon(i) \qquad (2.7)$$

So, here $y(i) - x(i)$ is the incremental at development period j.

Consider the following two situations:

1. $b > 1$ and $a = 0$

Here to forecast the mean incremental at development period j we take the cumulative x at development period $j-1$ and multiply it by $(b-1)$.

2. $b = 1$ and $a \neq 0$

This means that $x(i)$ has no predictive power in forecasting $y(i) - x(i)$. The estimate of a is a weighted average of the incrementals in development period j. So, we would forecast the 'next' accident periods incremental by averaging the incrementals down a development period. Accordingly, the standard link ratio technique is abandoned in favour of averaging incrementals for each development period down the accident periods.

If b=1 then the graph of $y(i) - x(i)$ against $x(i)$ should be flat, as depicted below in Figure 2.5, which represents the incrementals versus previous cumulatives (development period 0) for the Mack (1994) data. It is clear that the correlation is zero. This is also true for every pair of contiguous development periods.

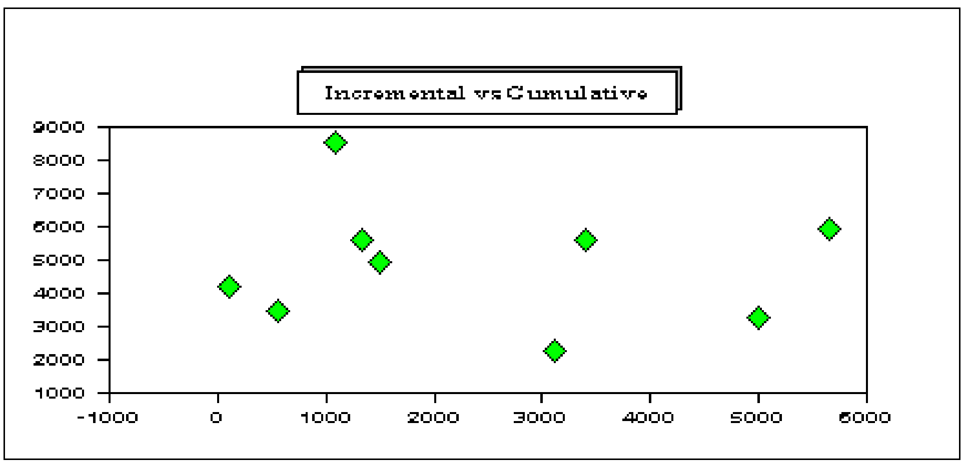


Figure 2.5 Incrementals Development Period 1 versus Cumulative Development Period 0

In conclusion, if the incrementals $y(i) - x(i)$ in development period j, say, appear random it is very likely that the graph of $y(i) - x(i)$ versus $x(i)$ is also random. That is, there is zero correlation between the incrementals and the previous cumulatives.

Now, if the incrementals possess a trend down the accident periods, the estimate of the parameter b in equation (2.7) will be significant and so the link ratio (b) plus the intercept (a) will have some predictive power. We should, however, incorporate an accident period trend parameter for the incremental data, namely,

$$y(i) - x(i) = a_0 + a_1 i + (b-1)\, x(i) + \varepsilon(i), \qquad \mathrm{Var}[\varepsilon(i)] = \sigma^2 x(i)^\delta. \tag{2.8}$$

For most real cumulative loss development arrays that possess a constant trend down the development period the trend parameter (a_1) will be more significant than the ratio minus 1 (b–1). Indeed, very often b–1 will be insignificant. That is, the trend will have more predictive power than the ratio, and the residual predictive power of the ratio after including the trend will be insignificant.

We use the following naming convention for the *three* parameters:

a_0 - Intercept
a_1 - Trend
b - Ratio (Slope)

Here are some models included in the ELRF described by equation (2.8).

- Chain Ladder Link Ratios

  Here Intercept = Trend = 0 and δ = 1.

- Cape Cod – Intercept Only

  Here it is assumed that the Ratio = 1 and the Trend = 0. The Cape Cod estimates a weighted (depending on δ) average of the incrementals in each development period. It also forecasts a weighted average down the accident periods for each development period.

  The model can be written

$$y(i) - x(i) = a_0 + \varepsilon(i), \quad \mathrm{Var}[\varepsilon(i)] = \sigma^2 x(i)^\delta.$$

- Trend with Ratio = 1

  The model estimates a weighted (depending on δ) trend (parameters a_0 and a_1) down the accident periods for each development period. It also forecasts a weighted trend down the accident periods for each development period.

Example 1: The Mack data

The data for the first example is from Mack (1994). The data are incurred losses for automatic facultative business in general liability, taken from the Historical Loss Development Study, 1991, published by the Reinsurance Association of America.

	0	1	2	3	4	5	6	7	8	9
1981	5012	8269	10907	11805	13539	16181	18009	18608	18662	18834
1982	106	4285	5396	10666	13782	15599	15496	16169	16704	
1983	3410	8992	13873	16141	18735	22214	22863	23466		
1984	5655	11555	15766	21266	23425	26083	27067			
1985	1092	9565	15836	22169	25955	26180				
1986	1513	6445	11702	12935	15852					
1987	557	4020	10946	12314						
1988	1351	6947	13112							
1989	3133	5395								
1990	2063									

Table 2.1. Incurred loss array for the Mack data. Rows are accident years and columns are development years. Note that 1982 accident year values are low.

We first fit the chain ladder ratios regression model. That is, we fit equation (2.1) with δ=1 for every pair of contiguous development periods. The standardised residuals are displayed in Figure 2.6. Note that the equivalence of equations (2.5) and (2.6) means that the residuals of the cumulative data are identical to the residuals of the incremental data.

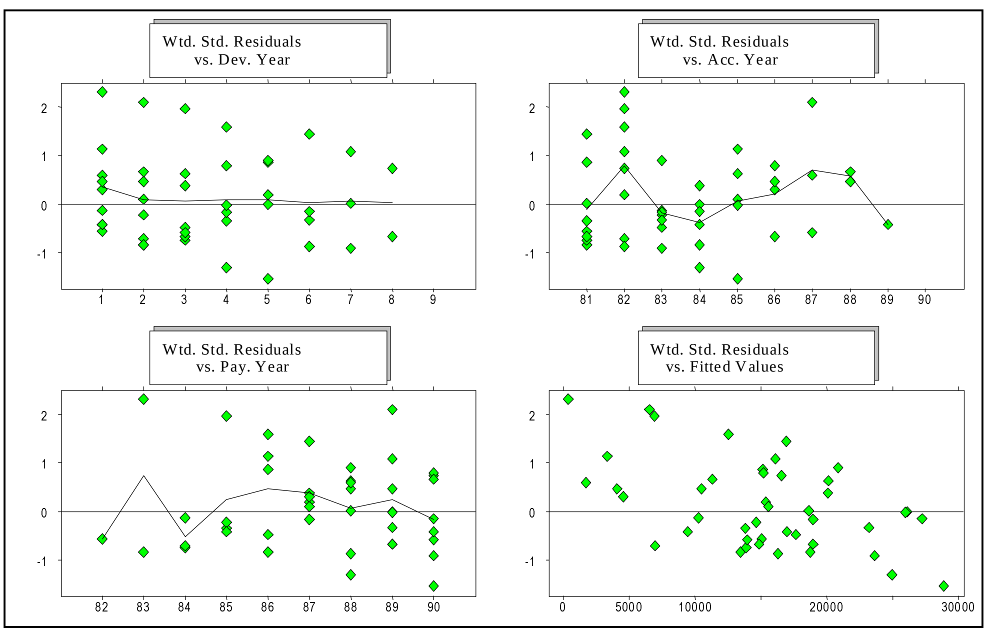


Figure 2.6 Residual plot for the chain ladder ratios model.

We have already observed the downward trend in the fitted values (Figure 2.2) and that the high cumulatives are overfitted whereas the low cumulatives are underfitted. This is mainly due to the fact that intercepts are required.

So, we now fit models (2.6) with intercepts except for the last two pairs of contiguous development periods, as there is insufficient data here. See Table 2.2 for the regression output. Note that none of the slope (ratio) parameters are significantly different from 1 and if both parameters are insignificant, the slope (ratio) is less significant. This means that the previous cumulative is not really of much help in predicting the next incremental incurred loss.

Intercept and Ratio Regression Table $\delta = 1$

Develop.	Intercept			Slope			
Period	Estimate	Std.Error	p value	Estimate	Slope - 1	Std. Error	p value
00-01	4,329	516	0.000	1.21445	0.21445	0.42131	0.626
01-02	4,160	2,531	0.151	1.06962	0.06962	0.35842	0.852
02-03	4,236	2,815	0.193	0.91968	-0.08032	0.24743	0.759
03-04	2,189	1,133	0.126	1.03341	0.03341	0.07443	0.677
04-05	3,562	2,031	0.178	0.92675	-0.07325	0.11023	0.554
05-06	589	2,510	0.836	1.0125	0.0125	0.12833	0.931
06-07	792	149	0.118	0.9911	-0.0089	0.00803	0.467
07-08	-	-	-	1.01694	0.01694	0.01506	0.463
08-09	-	-	-	1.00922	0.00922	-	-

(AIC=760.8)

Table 2.2. Fit of the model with intercept and ratio, with δ at 1.
(There is no intercept fitted for the last two years).

The model is overparameterized, so we eliminate the least significant parameter in each regression. We find that in each case the intercept is the parameter retained: that is, for every pair of contiguous development periods the model reduces to Cape Cod, that is,

$$y(i) - x(i) = a_0 + \varepsilon(i).$$

The residual plots for the reduced model (Cape Cod) are given in Figure 2.7.

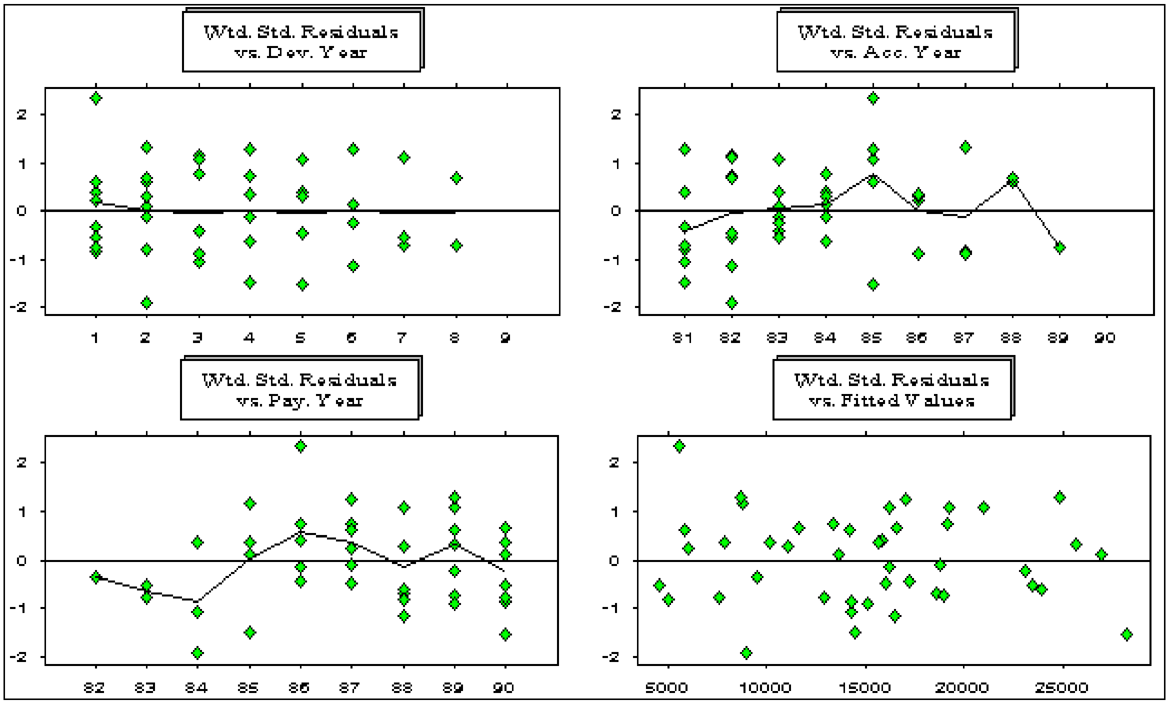


Figure 2.7. Residual plot for δ = 1, model with intercepts and with slopes set to 1. The line joins mean residuals.

Note that residuals versus fitted values are 'straight' now and that we do not have the high low effect in residuals versus accident periods. Since residuals versus accident years do not exhibit a trend, if we were to include a trend, that is, estimate

$$y(i) - x(i) = a_0 + a_1 i + \varepsilon \,,$$

we would find that the estimate of a_1 is insignificant.

We now present forecasts and coefficients of variation based on the Cape Cod (intercept only) with δ=1 model and compare this with the forecasts and coefficients of variation for the chain ladder ratios.

	Cape Cod			**Chain Ladder**		
Accident Year	Mean Forecast	Standard Error	Coeff. of Variation	Mean Forecast	Standard Error	Coeff. of Variation
1981	0	0	-	0	0	-
1982	172	41	0.244186	155	148	0.954839
1983	483	465	0.899142	616	586	0.951299
1984	1,113	498	0.385531	1633	702	0.429884
1985	1,941	1,218	0.51217	2779	1404	0.505218
1986	4,200	1,555	0.408791	3671	1976	0.538273
1987	6,878	1,677	0.271393	5455	2190	0.401467
1988	10,252	3,247	0.308234	10934	5351	0.489391
1989	14,874	3,657	0.25381	10668	6335	0.593832
1990	19,336	4,532	0.215021	16360	24606	1.504034
Total	**59,248**	**8,494**	**0.110347**	**52272**	**26883**	**0.514291**

Table 2.4. Comparison of Cape Cod coefficients of variation with those for the Chain Ladder Ratios.

Note that for the Cape Cod model the standard errors are generally decreasing as a percentage of the accident year forecast totals as we proceed down to the later years. This is because the model relates the numbers in the triangle to a certain degree- it assumes that the incremental values in the same development period are random from the same distribution. This does not happen with the chain ladder ratios, because the model does <u>not</u> relate the incrementals in the triangle in any meaningful way. For example, how are the values in the development period 0 related? Consequently, the coefficients of variation are substantially higher for the chain ladder ratios model, and moreover violate the fundamental statistical principle of insurance - risk reduction by pooling. It does not make sense that the coefficient of variation for 1990 is 150%, but for the previous year, 1989, it is 59%, when 1990 has only one more incremental value to forecast than 1989.

For the Mack data, the model with intercepts is reasonable, as there is no accident year trend in the incrementals. For data where a constant trend (on a dollar scale) does exist, then the trend will be significant, but very often the ratio -1 will be insignificant.

2.3 Summary

We have so far considered two modelling cases: incrementals for a particular development period have a zero trend, and incrementals have a constant trend (after possibly adjusting the data by accident year exposures). In both these real data cases link ratios are often insignificant and therefore also lack predictive power. The case encountered most often in practice, however, involves a trend change along the payment/calendar periods (diagonals). This means that as you look down each development period, the change in trend will occur in different accident periods. Consequently, none of the above models in the ELRF can capture these trends.

The weighted standardised residuals depicted in Figure 2.8 and Figure 2.9 are those of the chain ladder ratios and model (2.8) respectively, applied to project ABC (Worker's Compensation Portfolio) discussed in Section 3. Note that the chain ladder ratios indicate a payment year trend change and model (2.8) that fits a constant trend down the accident years for each development year, indicates that the trend before payment year 1984 is lower than the trend after 1984. This project (ABC) is analysed in more detail in Section 3.

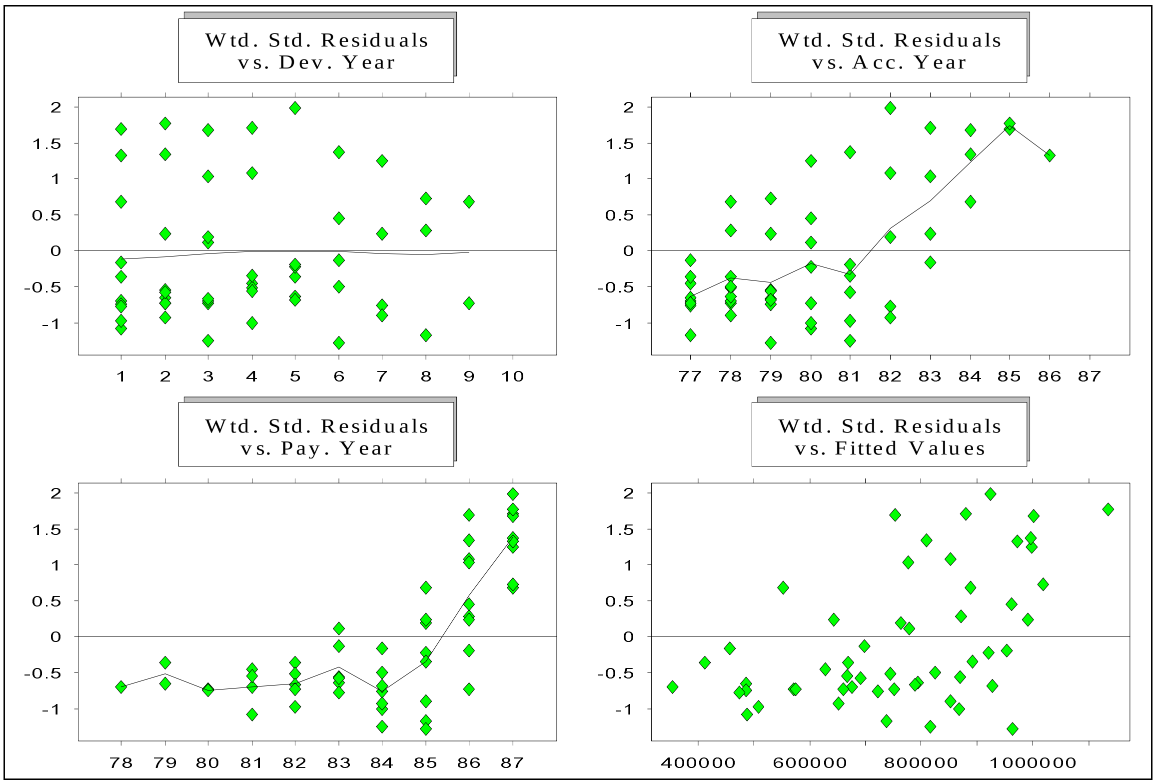


Figure 2.8. Residual plot for chain ladder ratios. The line joins the means of the residuals.

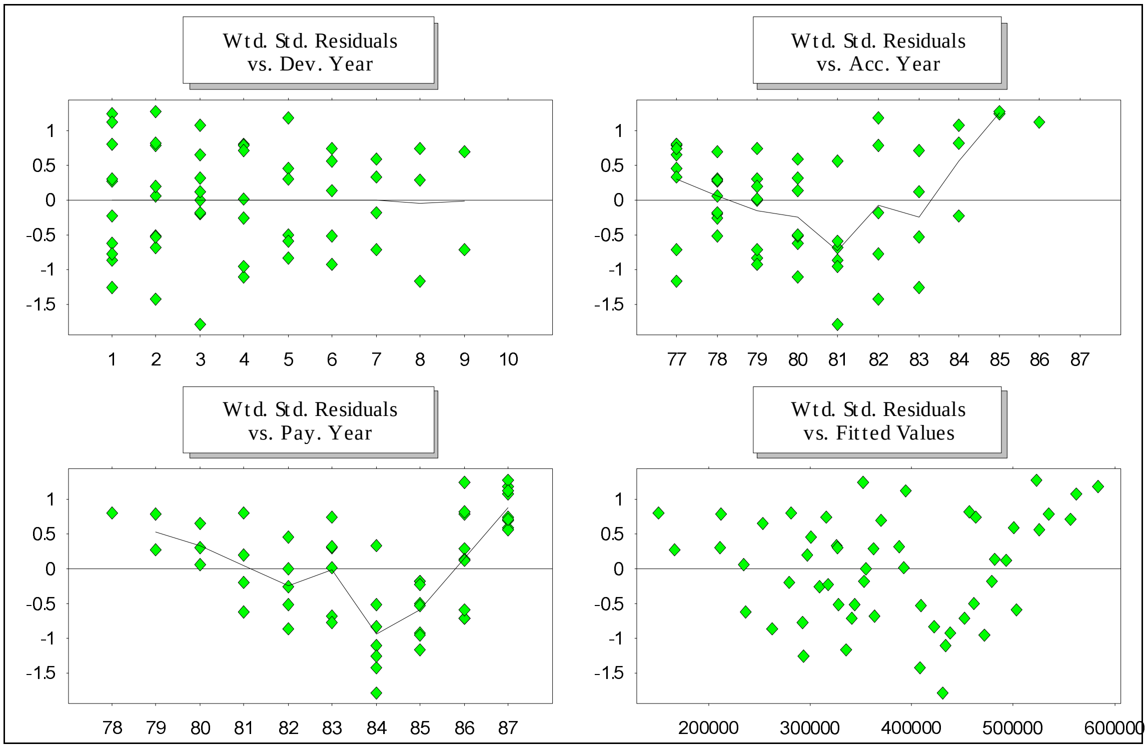
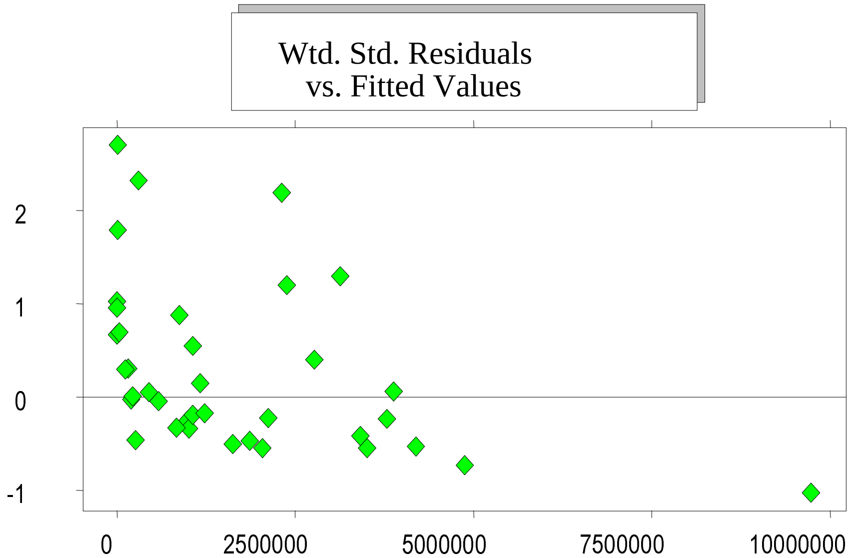


Figure 2.9 Residual plots for trends plus ratio model

The types of models described by equation (2.8) can be used to diagnostically identify payment period trend changes, but cannot estimate these trend changes or forecast with them. These models in the ELRF form a bridge to models that also include payment period trend parameters, that is, statistical models in the PTF.

It is important to note that ELRF models also make the implicit assumption that the weighted standardised errors come from a normal distribution. If the assumption is true, the estimates of the regression parameters are optimal. If the assumption is not true, the estimates may be very poor. This normality assumption is rarely true for loss reserving data. In fact, the weighted standardised residuals are generally skewed to the right, suggesting that the analysis should be conducted on the logarithmic scale. The graph below illustrates the skewness of a set of weighted standardised residuals based on chain ladder ratios for Project Pan6 analysed in detail in Section 3. The positive weighted standardised residuals are further from zero than the negative ones. If the normality assumption were correct, the plot would look roughly symmetric about the zero line.



In summary, using the regression methodology of ELRF, you will discover that for most real loss development arrays of any data type, standard development factor (link-ratio) techniques are inappropriate. Analysing the incrementals on the logarithmic scale with the inclusion of payment period trend parameters has more predictive power.

Finally, but importantly, the estimate of a mean forecast of outstanding (reserve) and corresponding standard deviation based on a model are meaningless unless the assumptions made by the model are supported by the data.

3 Statistical Modelling Framework

3.1 Introduction

Clearly we require a model that is able to deal with changing trends; trends in the data on the original (dollar) scale are hard to deal with, since trends on that scale are not generally linear, but move in percentage terms – for example, 5% superimposed (social) inflation in early years, and 3% in later years. It is the logarithms of the incremental data that show linear trends. Consequently we introduce a modelling framework for the logarithms of the incremental data that allows for changes in trends. The models of this type provide a high degree of insight into the loss development processes. Moreover, they facilitate the extraction of maximum information from the loss development array.

The details of the modelling framework and its inherent benefits are described in Zehnwirth (1994). However, given that there is a paradigm shift from the standard link ratio methodology to the statistical modelling framework, we review the salient features of the statistical modelling framework.

3.2 Trend Properties of Loss Development Arrays

Since a model is suppose to capture the trends in the data, it behoves us to discuss the geometry of trends in the three directions, viz., *development year*, *accident year* and *payment/calendar year*.

Development years are denoted by j; j = 0, 1, 2, ..., s-1; accident years by i; i = 1, 2, ..., s; and payment years by t; t = 1, 2, ..., s.

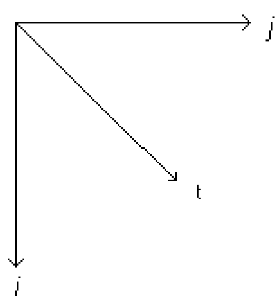


Figure 3.1

The payment year variable t can be expressed as $t = i + j$. This relationship between the three directions implies that there are only two 'independent' directions.

The two directions, development year and accident year, are orthogonal, equivalently, they have zero correlation. That is, trends in either direction are not projected onto the other. The payment year direction t however, is not orthogonal to either the development or accident year directions. That is, a trend in the payment year direction is also projected onto the development year and accident year directions. Similarly, accident year trends are projected onto payment year trends.

The main idea is to have the possibility of parameters in each of the three directions – development years, accident years and payment years. The parameters in the accident year direction determine the level from year to year; often the level (after adjusting for exposures) shows little change over many years, requiring only a few parameters. The parameters in the development year direction represent the trend from one development year to the next. This trend is often linear (on the log scale) across many of the later development years, often requiring only one parameter to describe the tail of the data. The parameters in the payment year direction describe the trend from payment year to payment year. If the original data are inflation adjusted before being transformed to the log scale, the payment year parameters represent superimposed (social) inflation, which may be stable for many years or may not be stable. This is determined in the analysis. Consequently, the (optimal) identified model for a particular loss development array is likely to be parsimonious. This allows us to have a clearer picture of what is happening in the incremental loss process.

The mathematical formulation of the models in the statistical modelling framework is given by equation (3.6) below. We now illustrate the geometry of trends with a simulation example.

Example 2 - Simulated Data:

To illustrate the trend properties of a loss development array, let us examine a situation where we know the trends, because we have selected them. Consider a set of data where the underlying paid loss (at this point without any payment year trends or even randomness – just the underlying development) is of the form

$$y(i, j) = \ln(p_{ij}) = 11.51293 - 0.2\,j$$

On a log scale this is a line with a slope of -0.2. The accident years are completely homogeneous. Let's add some payment/calendar year trends. A trend of 0.1 from 1978 to 1982, 0.3 from 1982 to 1983 and 0.15 from 1983 to 1991.

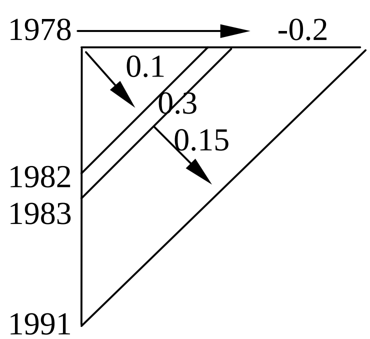


Figure 3.1. Diagram of the trends on the log scale in the data array.

Patterns of change like this are quite common in real data. Trends in the payment/calendar year direction project onto the other two directions. The resultant trends for the first six accident years are shown below.

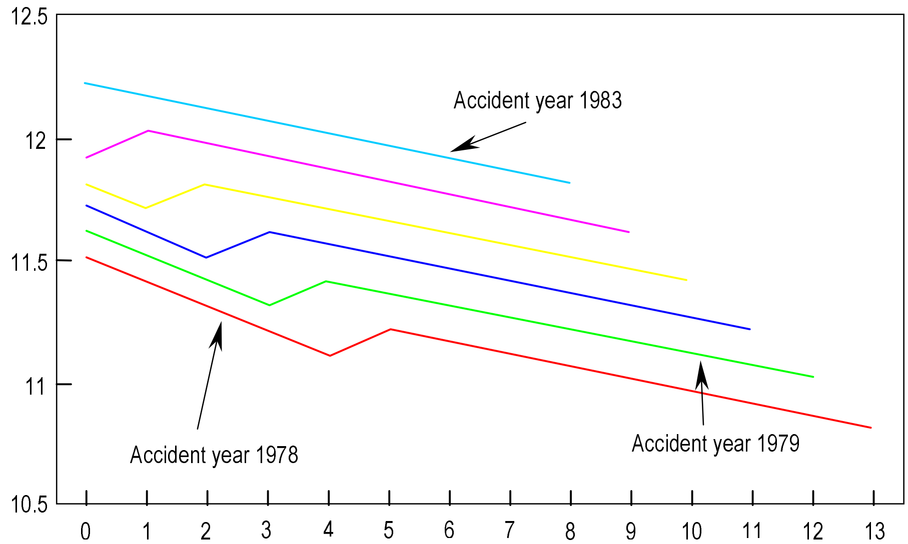


Figure 3.3. Plot of the log(paid) data against delay for the first six accident years.

Note that each line in the graph is the resultant development year trend for a single accident year. As you go down the accident years (1978 to 1983) the 30% trend always kicks in, one development period earlier. The payment year trends also project onto the accident years, which is why the early years are at the bottom and the later years are at the top. Note how the "kink" moves back as we go up to the more recent accident years. The resultant development year trends are different for each accident year now. We can't model even this simple situation with link ratios, or any model in ELRF.

Of course, real data is never so smooth. On the same log scale, we add some noise – random numbers with mean zero and standard deviation 0.1.

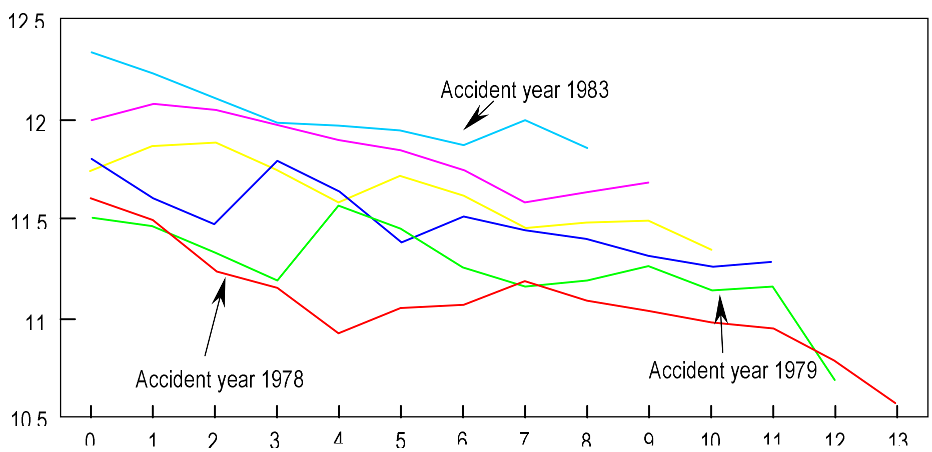


Figure 3.4. Trend plus randomness for the first six accident years.

Now the underlying changes in trends are not at all clear for two reasons. The payment year trends project onto development years and the data always exhibits randomness that tends to obscure the underlying trend changes. It has many of the properties we observe in real data – and yet it is plain that even with the extensions, the regression models in ELRF from Section 2 are inadequate for this data. We instead model the trends (in the three directions) and the variability. We measure these things on the log scale. In this Section, let $y(i,j)$ be the natural log of the incremental payment data in accident year i and development year j. This is different from our use of $y(i,j)$ in Section 2, but we do it for consistency with the literature appropriate to the models in each Section. We will analyze this data shortly.

Consider a single accident year. We represent the expected level in the first development year by a parameter (α). We can model the trends across the development years by allowing for a (possible) parameter to represent the expected change (trend) between each pair of development years. We model the variation of the data about this process with a zero-mean normally distributed random error. That is:

$$y(j) = \alpha + \sum_{k=1}^{j} \gamma_k + \varepsilon_j \qquad (3.1)$$

This probabilistic model is depicted below (for the first six development years).

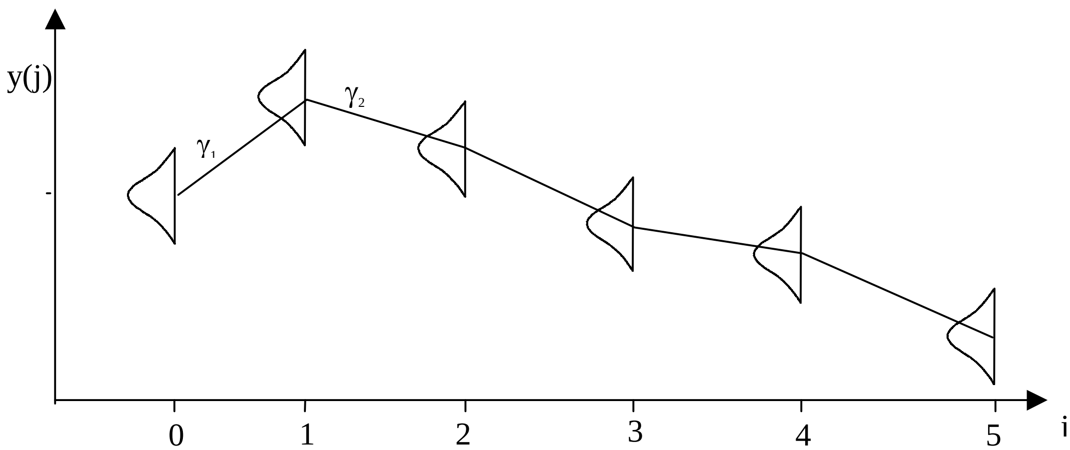


Figure 3.5. Probabilistic model for trends along a development year on the log scale.

For this probabilistic model, α is not the value of y observed at delay 0. It is the mean of $y(0)$. Indeed, $y(0)$ has a normal distribution with mean α and variance σ^2. Similarly, γ_j is not the observed trend between development year j-1 and j, but rather it is the mean trend between those development years – $E[y(j)–y(j–1)] = \gamma_j$.

The parameters of the probabilistic model represent means of random variables. Indeed, the model (on a log scale) comprises a normal distribution for each development year where the means of the normal distributions are related by the parameter α and the trend parameters γ_1, γ_2, ..., .

Based on the model in equation 3.1, the random variable $p(j)$ has a lognormal distribution with

$$\text{Median} \qquad = \exp\left[\alpha + \sum_{j=1}^{d} \gamma_j\right] \tag{3.2}$$

$$\text{Mean} \qquad = \text{median} \times \exp[0.5\sigma^2] \tag{3.3}$$

and

$$\text{Standard Deviation} \qquad = \text{mean} \times \sqrt{\exp[\sigma^2] - 1} \tag{3.4}$$

The probabilistic model for $p(j)$ comprises a lognormal distribution for each development year where the medians of the lognormal distributions are related by equation 3.2 and the means are related by equation (3.3). So, in fitting or estimating the model we are essentially fitting a lognormal distribution to each development year. The trend (on a log scale) comprising the straight line segments is only one component of the model. A principal component comprises the distributions about the trends.

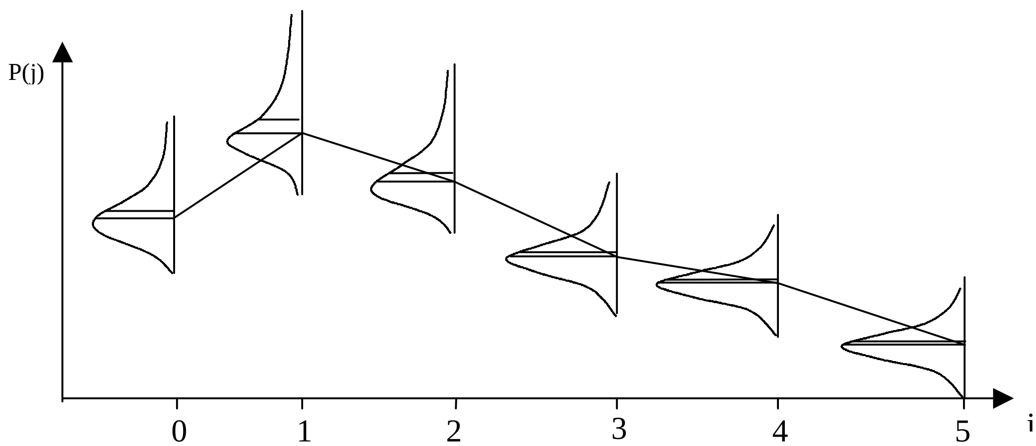


Figure 3.6. Model for trends along a development year (dollar scale). Means and medians of the distributions are marked.

Note equation (3.2) that exponentiating the mean on the log scale gives the median on the dollar scale (which is why the line above joins the medians). We will normally use the mean as our forecast, rather than the median, but the uncertainty (measured by the standard deviation) of the lognormal distribution is just as important a component of the forecast.

If we compute expected values of the logs of the development factors on the *incremental* data with this model, we obtain $E[ln(p(j)/p(j-1))] = E[(\gamma_j+\varepsilon_j-\varepsilon_{j-1})] = \gamma_j$. That is, trend parameters also underpin this new model, but in a way that will allow it to appropriately model the trends in the incremental data, in the three directions.

The model described so far only covers a single accident year. We have not yet accounted for the payment year and accident year trends. Let the mean of the (random) inflation between payment year *t* and *t+1* be represented by ι_t (*iota-t*).

Hence the *family* of models can be written:

$$y(i,j) = \alpha_i + \sum_{k=1}^{j} \gamma_k + \sum_{t=1}^{i+j} \iota_t + \varepsilon_{i,j} \tag{3.6}$$

We call this family of models the probabilistic trend family (PTF). Note that the mean trend between cells (*i*,*j*-*1*) and (*i*,*j*) is $\gamma_j + \iota_{i+j}$, and the mean trend between cells (*i*-*1*,*j*) and (*i*,*j*) is $\alpha_{i+1} - \alpha_i + \iota_{i+j}$.

A member of this *family* of models relates the lognormal distributions of the cells in the triangle. On a log scale the distribution for each cell is normal where the means of the normal distributions are related by the "trends" described by the member.

If the error terms $\varepsilon_{i,j}$ coming from a normal distribution with mean zero do not have a constant variance, then the changing variance also has to be modelled. Note that there are numerous models in PTF, even if we do not include the varying (stochastic) parameter models discussed in Section 3.3. The actuary has to identify the most appropriate model for the loss development array being analysed. The assumptions made by the 'optimal' model must be satisfied by the data. In doing so, one extracts information in terms of trends, stability thereof and the distributions of the data about the trends.

Example 2 continued - Estimation:

Let's now try to identify the model that created the data. We begin by fitting a model with all the development year trends equal (one γ) and all payment year trends equal (one ι and with no accident year trends (one α). That is, with $\gamma_k = \gamma$, $\iota_t = \iota$, and $\alpha_i = \alpha$, for all parameters. The parameter estimates are given in Table 3.2.

Parameter	Estimate	Std. Error	t-ratio
α	11.4256	0.0302	378.57
γ	-0.2062	0.0037	-55.08
ι	0.1563	0.0037	41.74

s = 0.1129 R^2 = 97.0%

Table 3.2. Parameter estimates for the model with constant trends.

The estimate of ι (iota) 0.1563 is the weighted average of the three trends 0.1, 0.3 and 0.15.

Removing constant trends makes any changes in trend more obvious. The residuals are shown in Figure 3.7.

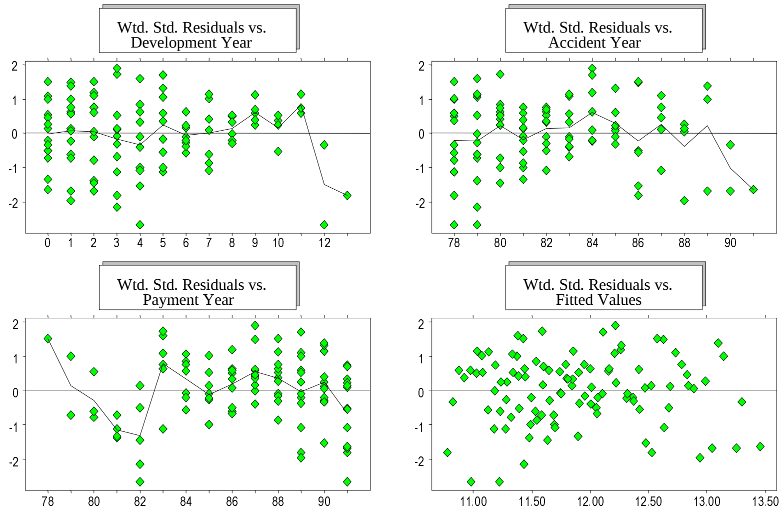


Figure 3.7. Plots of standardized residuals against the three directions, and against fitted values for the single payment year trend model. The lines join mean residuals.

The residuals need to be interpreted as the data adjusted for what has been fitted. Accordingly, the residuals versus payment years represent the data minus the fitted 0.1563.

Immediately the changes in trends in the payment year direction become obvious. We can see that the trend in the early years is substantially less than the estimated average of 0.1563, that the trend from 1982 to 1983 is much larger than it, and after that, the trend is pretty close to the fitted trend, as 0.15 - 0.1563 is approximately zero. This suggests that we should introduce another ι (iota) between 1982-1983 and another ι between 1983-1984 (that will continue to 1991).

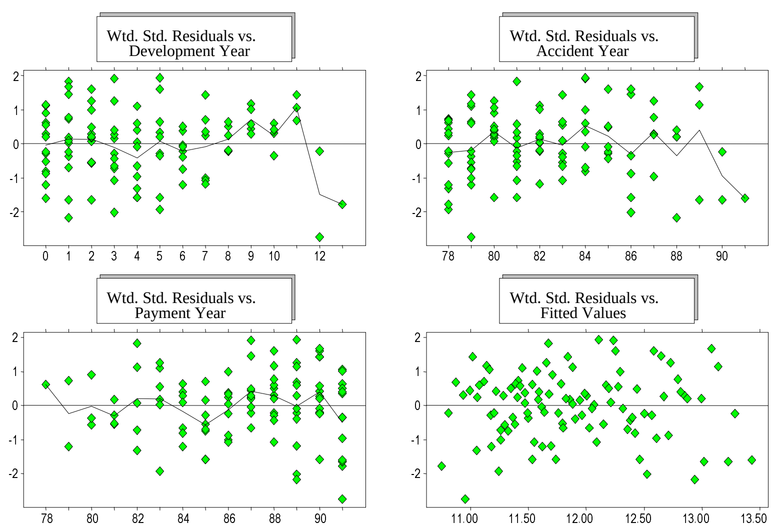


Figure 3.8. Plots of standardized residuals against the three directions, and against fitted values for the model with three payment year trends.

The residuals of the model with three payment year trends are given in Figure 3.8. This model seems to have captured the trends.

Parameter		Estimate	Std. Error	t-ratio
α		11.5321	0.0612	188.34
γ		-0.2062	0.0033	-61.91
ι	78-82	0.0873	0.0209	4.18
	82-83	0.3927	0.0442	8.90
	83-91	0.1446	0.0046	31.72

s = 0.1005 R^2 = 97.7%

Table 3.3. Parameter estimates for the model with three payment year trends.

Note that the estimates of the trend parameters 0.1, 0.3, 0.15 are not equal to the true values, indeed 0.3927 (±0.0442) is a bit off the mark (but not significantly). That is because in the payment years 1982 and 1983 there aren't many data points. Given the trend of 0.15 is in the data since 1983, we would expect stability of forecasts, and trend parameter estimates as we remove years.

The forecasts are stable – if we remove the most recent data, the forecasts of this model don’t change much relative to the standard error in the forecast, as we can see in Table 3.4.

Yrs in Estim.	N	γ (83-91)	std. err.	ι (83-91)	std. err.	Mean Fcst	std. error. Fcst
78-91	105	-0.2062	0.0033	0.1446	0.0046	23,426,542	927,810
78-90	91	-0.2075	0.0036	0.1527	0.0051	25,333,522	1,191,129
78-89	78	-0.2086	0.0042	0.1512	0.0064	24,850,972	1,526,246
78-88	66	-0.2119	0.0045	0.1575	0.0075	26,296,366	1,997,089
78-87	55	-0.2131	0.0055	0.1563	0.0103	25,894,931	2,868,948

Table 3.4. Forecasts and standard errors, and trend estimates (and their standard errors), for the selected model as the later payment years are removed.

Note that the estimate of γ (= -0.2) is pretty stable, as we remove the latest years.

The display below, Figure 3.9, gives the prediction errors (on a log scale) for the four payment years 1988 - 1991 based on the model estimated at year end 1987.

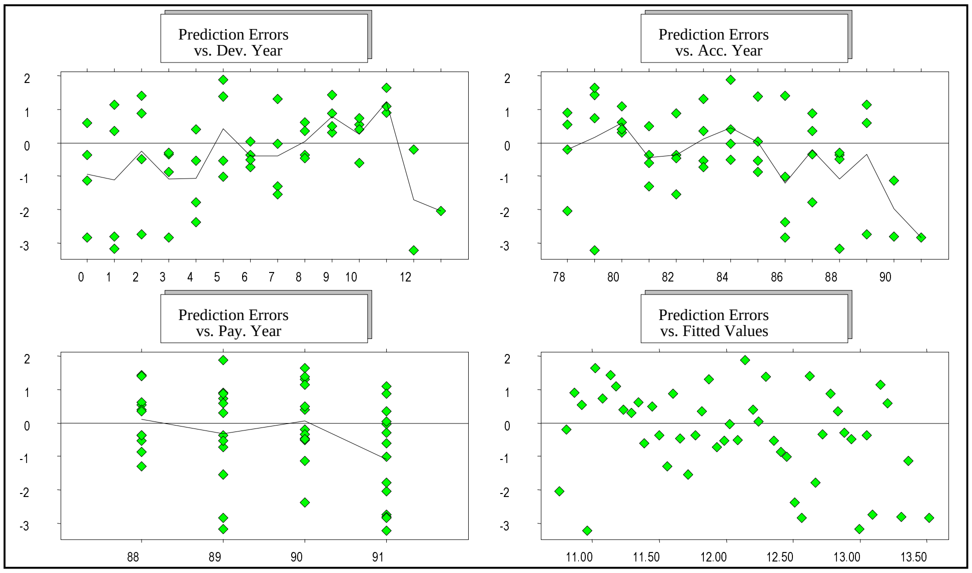


Figure 3.9 Prediction errors for 1988-1991, based on model estimated at year end 1987.

So the estimated model at the end of payment year 1987 slightly over-predicts the payment periods 1988-1991. That is because the trend estimate (since 1983) is now 15.63% ± 1.03%, in place of 14.46% ± 0.46% when we use all the years in the estimation. Hence the forecast of \$25.89M (±\$2.87M) is 'higher' than \$23.4M (±\$0.93M). When you test for a trend change between 1987 and 1988 it is not significant (as we would expect). Note that removal of payment years (validation analysis) is part of the model identification procedure and extraction of information process.

Payment Years in Estimation	Estimate of gamma	Estimate of iota (83-91)	Forecast ± SE \$M
1978-91	-20.62 ± 0.33	14.46 ± 0.46	23 ± 0.9
1978-90	-20.75 ± 0.36	15.27 ± 0.51	25 ± 1.2
1978-89	-20.86 ± 0.42	15.15 ± 0.64	25 ± 1.5
1978-88	-21.19 ± 0.45	15.75 ± 0.75	26 ± 2.0
1978-87	-21.31 ± 0.55	15.63 ± 1.03	26 ± 2.9

Example 3 - Real data with major payment year trend instability

We now analyze a real data set.

	0	1	2	3	4	5	6	7	8	9	10
1977	153,638	188,412	134,534	87,456	60,348	42,404	31,238	21,252	16,622	14,440	12,200
1978	178,536	226,412	158,894	104,686	71,448	47,990	35,576	24,818	22,662	18,000	
1979	210,172	259,168	188,388	123,074	83,380	56,086	38,496	33,768	27,400		
1980	211,448	253,482	183,370	131,040	78,994	60,232	45,568	38,000			
1981	219,810	266,304	194,650	120,098	87,582	62,750	51,000				
1982	205,654	252,746	177,506	129,522	96,786	82,400					
1983	197,716	255,408	194,648	142,328	105,600						
1984	239,784	329,242	264,802	190,400							
1985	326,304	471,744	375,400								
1986	420,778	590,400									
1987	496,200										

Acci. Yr	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987
Exposure	2.2	2.4	2.2	2.0	1.9	1.6	1.6	1.8	2.2	2.5	2.6

Table 3.5. Incremental paid losses and exposures for ABC.

This loss development array has a major trend change between payment years 1984 and 1985, even though the data and link ratios are relatively smooth. Indeed, it needs to be understood that in general trend instability has nothing to do with volatility or smoothness of the data and link ratios. Formulation of the assumptions about the future trend will depend on the explanation for the trend change (when there is one).

The individual link ratios for the cumulated data are very stable, as can be seen in Figure 3.10 below. It is very dangerous to try to make judgements about the suitability of development factor techniques from the individual link ratios on the cumulated data.

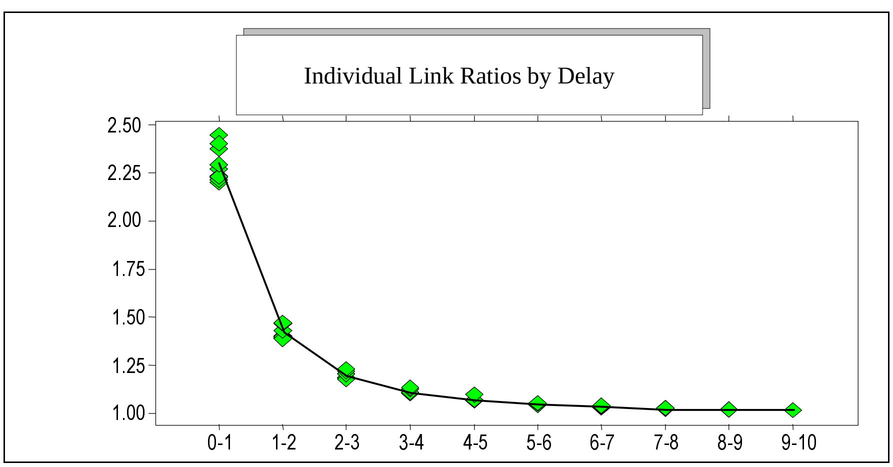


Figure 3.10. Plot of individual link ratios by delay. The line joins Chain Ladder ratios.

We first conduct some *diagnostic* PTF analysis and then show how the ELRF modelling structure also indicates payment year trend change and moreover that any method based on link ratios is quite meaningless. Figure 3.11 below shows the standardised residuals of the statistical chain ladder in PTF, i.e. the statistical chain ladder fits all the gamma parameters and all the alpha parameters (no iotas). So, the residuals are the data adjusted for the (average) trend between every pair of contiguous development periods and every pair of contiguous accident periods. This is why the residuals versus development years and residuals versus accident years are centred on zero! We use this model only as a diagnostic tool to determine (speedily) whether there are payment year trend changes which can be attributed solely to the payment years.

Contrast the smoothness of the above ratios with the plot of the residuals from this model.

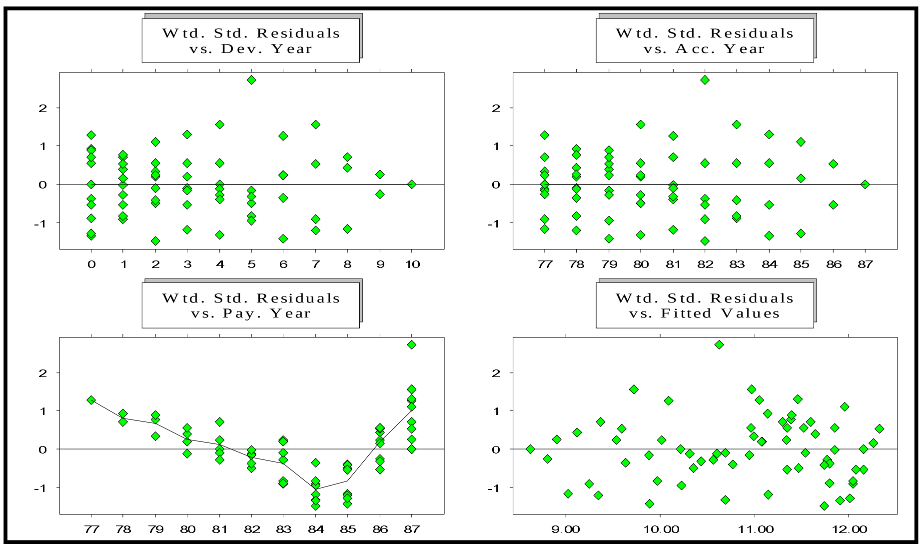


Figure 3.11 Standardised residuals of the statistical chain ladder model.

We can now see dramatic changes in the payment year direction. It might be very dangerous to use forecasts from any model assuming no changes in payment year trend, such as a model from the ELRF – it would correspond to forecasting along the zero line in Figure 2.8. (the residuals of the standard chain ladder ratios). There is a difference between Figures 3.10 and 2.8. The statistical chain ladder shows the payment year trends after adjusting for the trends in the other two directions. The chain ladder ratios (Figure 2.8) do not do that. But the change in trend is clear in either graph. In the current statistical modelling framework, we are able to model this change; we have a lot more control over how we incorporate the trend changes into our model and hence into the forecasts. Even the best ELRF model here hardly uses ratios and is deficient because it gives us no control in the payment year direction. It turns out that the trend before 1984 is approximately 10% whereas the trend past 1984 is approximately 20%. So which trend should we assume for the future? This depends on the explanation for the change. If the trend instability is due to new legislation that applies retrospectively (to all accident periods) then one would revert to the 10%. If there is no explanation for the trend change, except that the payments have increased, then calling the future in terms of trends is more difficult.

Example 4 - Volatile data with stable trends

We now consider an array where the paid losses are very volatile, but the trends are stable. Recall that trend stability/instability is not dependent on the volatility of the data nor of the link ratios. Since the random component is an integral part of the model, this model captures the behaviour of this volatile data very well. We call this array PAN6.

	Development Year					
Acci. Yr	0	1	2	3	4	5
1986	194324	571621	327880	249194	524483	1724274
1987	1469	57393	485791	169614	121410	599021
1988	1860	161538	408008	314614	6744000	****
1989	23512	185604	260725	1134272	851099	2174200
1990	1044	70096	93600	1283752	1595466	913215
1991	****	3730	869959	187019	2764795	****
1992	****	443205	180064	683407	878117	
1993	****	12808	433511	118017		
1994	1431	77765	151161			
1995	51539	****				
1996	****					

Table 3.6. Paid loss array for the PAN6 data for Example 4.

A good model can be identified quickly for the logarithms of these data; it has no payment year trends, and only two different development year trends; between development years 0-1, and for all later years. The residual plot is given in Figure 3.12.

However, note that the spread of the first two development years is wider than for the later years and the spread for 'small' fitted values is larger than the spread for 'large' fitted values. If we estimate the standard deviations in the two sections, we find that they are 3.0177 and 0.8015 respectively. This requires a weighted regression; development years 0 and 1 are given weight $(0.8015/3.0177)^2$, and the other years (2+) have weight 1. The weighted standardized residual plots now look fine; see Figure 3.13. A check of the plot of residuals against normal scores (not presented here) indicates that the assumption of normality of the logarithms of the data is very reasonable; the squared correlation is greater than 0.99.

The normal distributions for this model have relatively large variances. The estimate of σ^2 for development periods 0-1 is 2.923 and for development periods 2+ is 0.80346. Note that if a normal distribution has a variance σ^2, then the corresponding lognormal distribution has a coefficient variation of $\sqrt{\exp(\sigma^2)-1} > \sigma$.

This model also has forecasts that are stable as we remove the most recent data, as we see in Table 3.7. This is a very important attribute of this identified model that captures the information in the data – if the trends in the data are stable, then so are the forecasts based on the estimated model. In this case we were able to remove almost half the (most recent) data. The standard errors of the forecasts are large because the lognormal distributions are skewed - insurance is about measuring variance (not just mean).

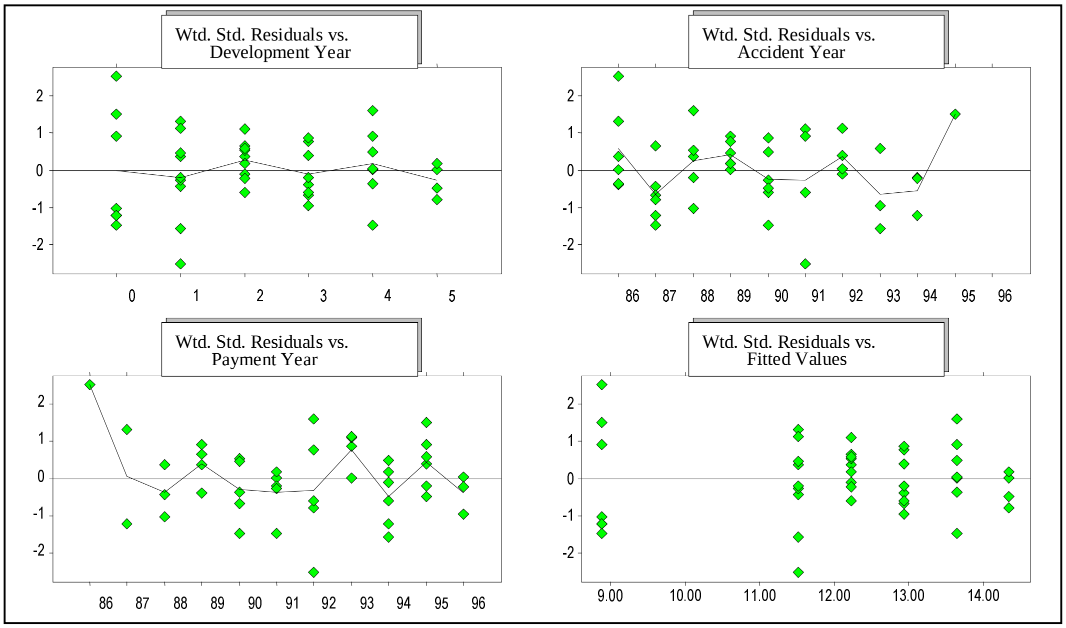


Figure 3.12. Plot of standardized residuals for the model with two gamma parameters and one alpha parameter.

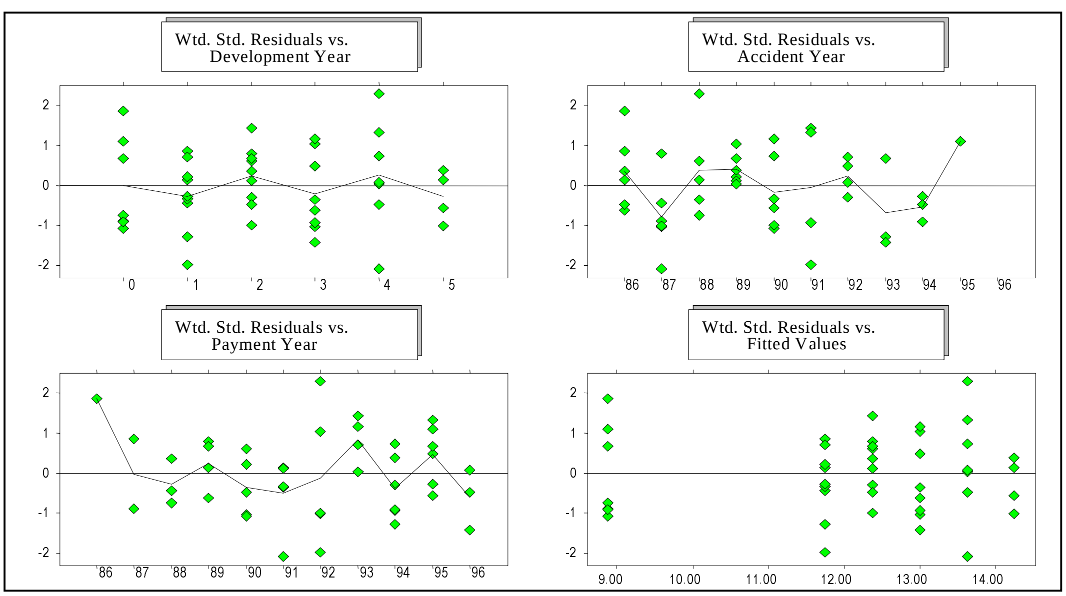


Figure 3.13. Plot of weighted standardized residuals after the weighted regression.

While the variability of the data and hence the standard errors of the forecasts are large, the message from the data has been consistent over many years. We are predicting the distribution of the data in each cell, not merely their mean and standard deviation, so a large standard deviation does not imply a bad model. Indeed, the model is very good. It captures the variances, indeed the distributions, in each cell;

Years in Estimation	N	Trend (dev period 1+)	standard error	Mean Fcst	Standard Error
86-96	44	0.6250	0.1432	20,352,011	9,136,870
86-95	41	0.6102	0.1479	21,410,781	9,839,127
86-94	35	0.6149	0.1681	21,037,520	10,654,173
86-93	29	0.5024	0.1977	19,755,944	11,647,274
86-92	25	0.5631	0.2143	18,567,664	11,529,359

Table 3.7. Forecasts and standard errors, and the final trend estimates (and their standard errors), for the final model as the later payment years are removed.

The high standard errors of forecasts are due to large process variability. As we remove recent years (diagonals) from the estimation we note the stability of forecasts (outstanding). This is further evidence of a stable trend in the data.

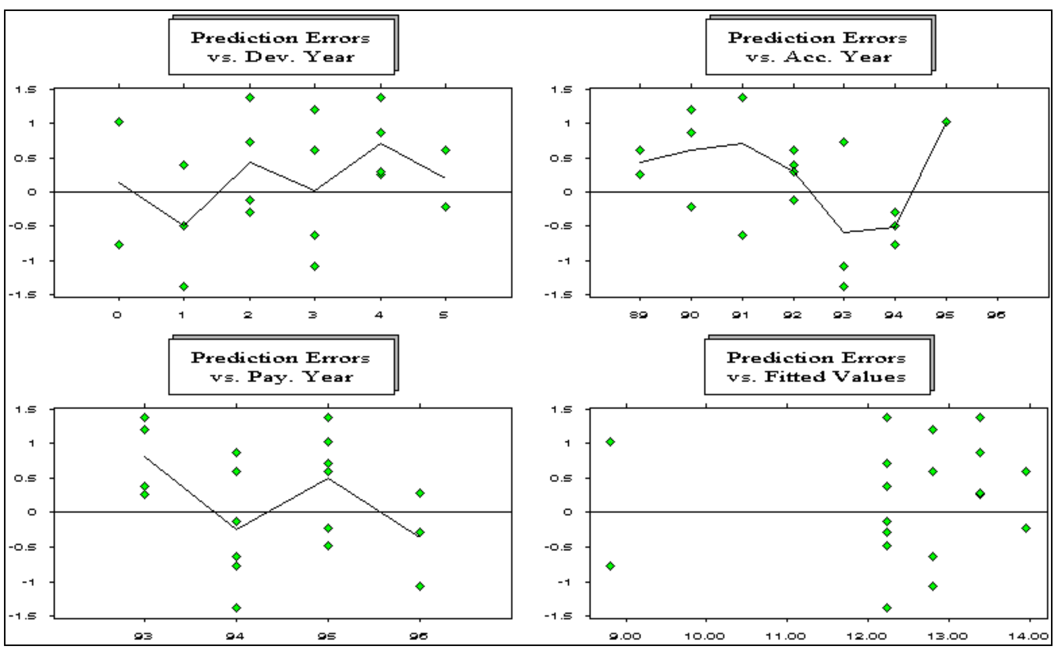


Figure 3.14 - Prediction errors for years 1993-1996

Note that at end of year 1992, the estimated model would have predicted the normal distributions for the log(payments) in years 1993-1996 and would have produced statistically the same forecast outstanding.

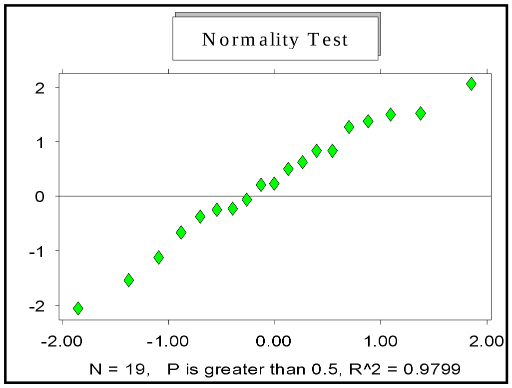


Figure 3.15. Normality plot of prediction errors for 1993-96 based on model estimated at year end 1992.

We now turn to ELRF analysis. Since the data are extremely skewed (lognormal with large coefficient of variation), the residuals of the chain ladder (regression) ratios in ELRF are extremely skewed to the right. See Figure 3.16 below. The plot of residuals against fitted values shows a downward trend indicating that we overpredict the large values and underpredict the small ones. The residuals also show strong indications of non-normality. Moreover, all ratios have no predictive power (provided there is an intercept). In any event, residuals are skewed (not normal), so even the best model in ELRF, the Cape Cod ($y - x = a_0 + \varepsilon$), is not a good one.

Recall that if model assumptions are not satisfied by the data, then any forecast calculations are quite meaningless.

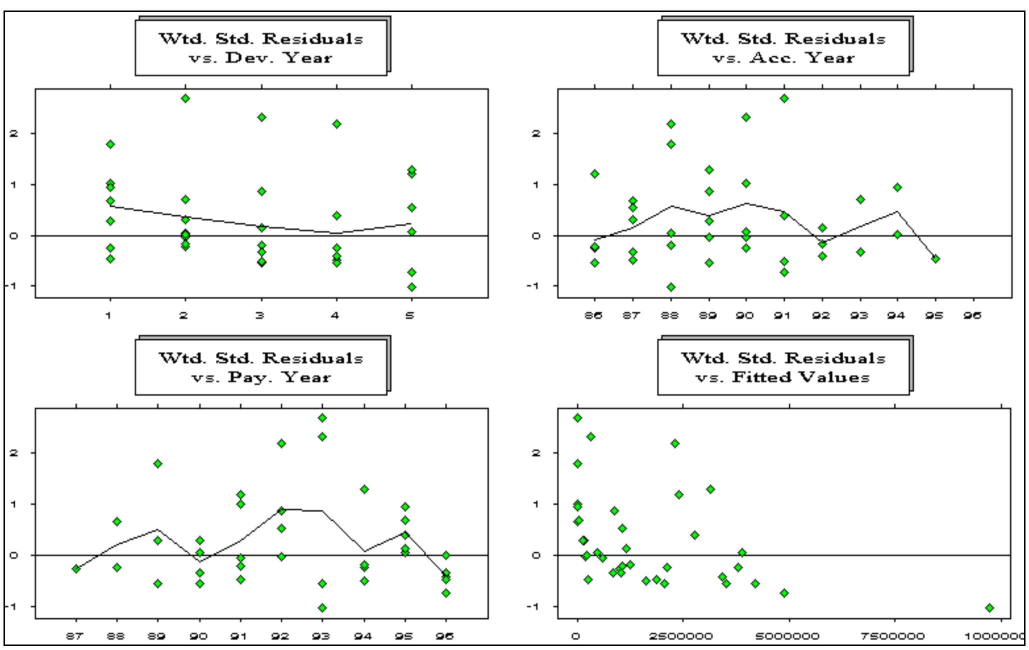


Figure 3.16. Residuals of chain ladder ratios regression model

Example 5 - Simulated array based on a (simple) model with only two parameters.

This array termed SDF contains a simulated data set where the incremental paid losses have accident years that are completely homogeneous. The actual model driving the data has one alpha (α) = 10, one gamma (γ) = -0.3 and σ^2 = 0.4. That is,

$$y(i, j) = 10 - 0.3d + \varepsilon(i, j)$$

where the $\varepsilon(i,j)$ are i.i.d. from N(0, 0.4).

The simulated data is presented in Table 3.8.

Acci.	Development Year								
Year	0	1	2	3	4	5	6	7	8
1978	24,307	44,260	3,900	13,393	17,731	5,802	1,975	6,294	4,758
1979	19,122	7,003	8,147	2,872	5,639	2,967	3,455	2,364	4,005
1980	18,082	27,708	27,901	5,699	9,297	2,899	7,461	6,885	1,854
1981	80,451	11,411	68,627	6,703	6,430	2,693	5,560	847	4,706
1982	49,099	7,144	11,979	3,481	2,279	5,975	4,472	2,066	509
1983	33,475	54,717	8,774	4,859	5,808	20,750	1,903	2,449	937
1984	23,070	49,554	5,659	9,909	9,123	8,805	7,008	8,634	3,053
1985	14,324	8,352	7,955	8,092	6,044	8,542	7,700	2,849	1,130
1986	58,785	16,833	6,068	5,227	3,276	16,592	2,407	908	2,508
1987	9,017	7,999	10,796	12,737	3,880	6,536	6,779	2,336	
1988	12,205	24,980	13,835	9,881	4,978	3,380	7,105		
1989	17,883	5,194	11,429	2,769	9,540	5,107			
1990	25,584	14,468	12,543	5,774	10,414				
1991	49,089	22,514	24,075	9,591					
1992	24,064	49,272	3,231						
1993	17,858	19,689							
1994	24,869								

Acci.	Development Year							
Year	9	10	11	12	13	14	15	16
1978	1,529	2,383	161	760	445	640	501	255
1979	2,029	1,370	718	432	1,282	456	232	
1980	2,681	792	1,338	820	702	179		
1981	1,245	1,289	717	754	436			
1982	2,345	755	1,702	596				
1983	796	1,446	1,271					
1984	493	2,108						
1985	1,754							

Table 3.8. Incremental paid loss data for simulated example SDF.

The first thing to note with this data is that once noise is added, it looks like incremental paid data for a real array, even though it was generated from a very simple model.

The relatively large $\sigma^2 = 0.4$ explains the high variability in the observed paid losses. The incremental data displayed in Table 3.8 appear volatile, but the values in the same development period are independent realizations from the same lognormal distribution.

For example, in development period zero, the simulated values 80,451 and 9,017 come from a lognormal distribution with mean 26,903 and standard deviation 18,867. Since a lognormal distribution is skewed to the right, realizations larger than the mean are typically 'far' away, whereas realizations less than the mean are bounded by zero and the mean, and so are 'closer' to the mean.

The apparent volatility in the data is not due to instability in trends - indeed the reality is quite the opposite - though volatile, the incremental paid losses have stable trends. Since we know the exact probability distributions driving the data we can compute the exact mean and exact standard deviation for each cell in the rectangle and also the exact means and standard deviations of sums.

The exact mean of the total outstanding is \$284,125 with an exact standard deviation of \$30,970. (So, the process variance is $30{,}970^2$). When we analyse the data in PTF we identify only two significant parameters $\hat{\alpha} = 9.9667 \pm 0.0847$ and $\hat{\gamma} = -0.2867 \pm 0.0126$. The estimate of σ^2 is 0.4085. Residuals of this estimated model are displayed in Figure 3.17 below.

The table below gives forecasts of total outstanding, including validation forecasts. Note stability as expected.

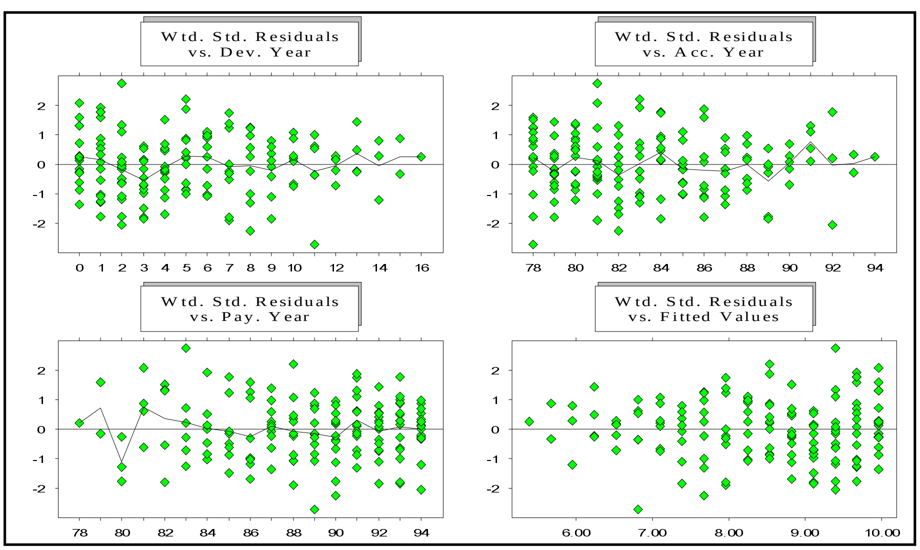


Figure 3.17 Residuals based on the estimated parameters of the true model.

Payment Years in Estimation	Estimate of Gamma	Mean Forecast ± SE
1978-94	-0.2867 ± 0.0126	299,660 ± 35,487
1978-93	-0.2858 ± 0.0146	303,980 ± 37,886
1978-92	-0.2865 ± 0.0166	302,601 ± 38,843
1978-91	-0.2926 ± 0.0195	304,711 ± 42,148
1978-90	-0.2940 ± 0.0228	296,650 ± 43,625
1978-89	-0.2861 ± 0.0271	313,604 ± 50,001

We now study the cumulative array.

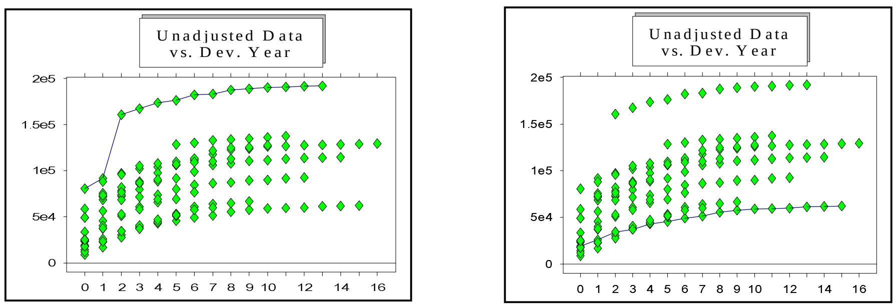


Figure 3.18 Accident year 1981 high development, 1979 low development.

Even though the incremental data was generated with accident years homogeneous, the cumulated data have each accident year at a completely different level; the plot against accident years jumps all over the place – the values along an accident year tend to be high or low. This is a common feature with cumulative arrays.

The cumulative values for 1979 lie entirely below those for 1982, (Figure 3.18) yet most of the incremental payments are 'close' together. One 'large' incremental value from the tail of the lognormal has a major impact on the cumulative data. The link ratio techniques assume that the next incremental payment will be high if the current cumulative is high, and this looks like what is going on with the cumulative data. So, the cumulatives deliver a false indication, even for data where there are no payment year trend changes.

Note that for 1979, cumulative paid at development year 5 is $45,750, whereas for 1981 it is $176,315. So, "current emergence is not a predictor of future emergence," a term used by Gary Venter.

The chain ladder ratios model gives a mean outstanding forecast of $254,130 and a standard error of the outstanding forecast of $59,419. The plot of residuals against fitted values makes it clear where the problem lies, as we see in Figure 3.19 below.

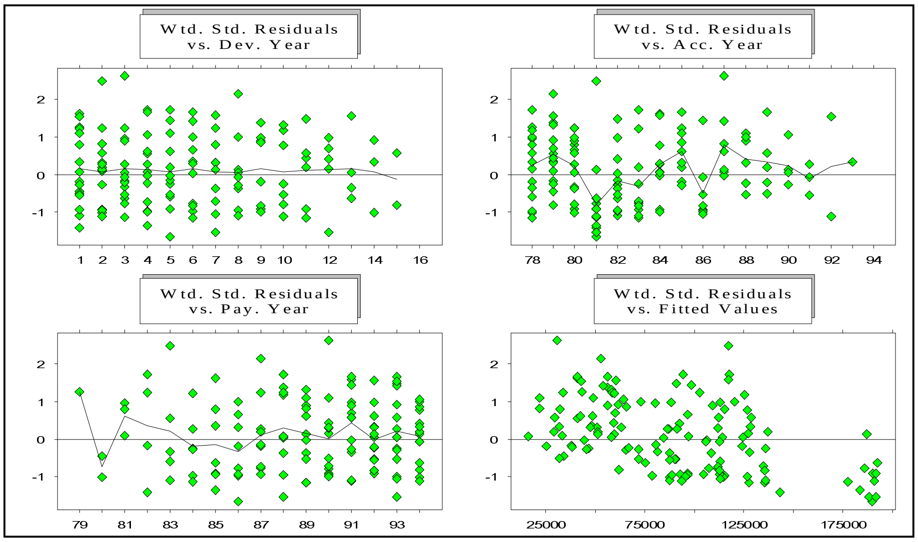


Figure 3.19 Plot of weighted standardized residuals for chain ladder ratios.

Again, we have a year with high cumulatives over fitted and year with low cumulatives under fitted. So, 1979 accident year is under fitted and 1981 accident year is over fitted.

Note how there is a distinct downward trend in the fitted values plot. It indicates that the model overpredicts the high cumulative values and underpredicts the low values – which it will do if the cumulatives don't really contain information on the subsequent incrementals. Normal scores plots show the non-normality. If we look at the plot of the incremental paid losses against the previous cumulative, we can see that models involving ratios will be inappropriate, since there is no relationship.

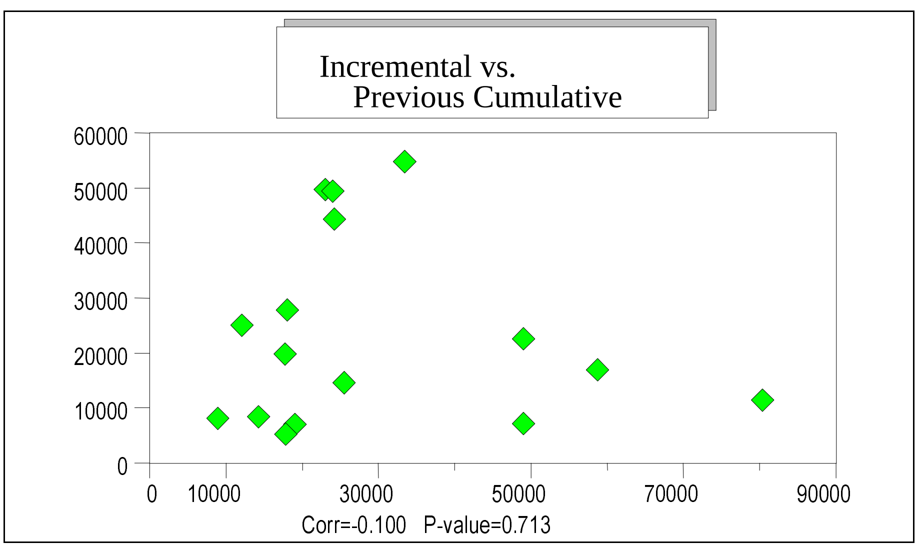


Figure 3.20. Plot of incremental payments against previous cumulative.

The best model in ELRF sets ratios to 1 and only uses intercepts. That is, it takes averages of incrementals in each development year. But due to non-normality, this is not good enough. At least ELRF analysis informs us that the incrementals in a development period are random from a distribution and these incrementals are not correlated to the previous cumulatives - the way the data were generated. It also tells us that the data are skewed and so we need to take a transformation. By way of summary, the ELRF analysis informs us that the data were created incrementally, accident years are homogeneous and we should be modelling the log incremental data. It is telling us the truth.

If you generate (simulate) data using ratios, ELRF will tell you that ratios have predictive power and that the data were generated cumulatively. But, and this is an extremely important qualification, for most real loss development arrays ELRF analysis will indicate that the data were generated incrementally, that ratios have no, or much less predictive power than trends in the log incrementals and that there may be payment/calendar year trend changes.

3.3 Varying (stochastic) parameters

In view of the trend relationships between the three directions, development year, accident year and payment year, a model with several parameters in the payment year and accident year directions will suffer from multi-collinearity problems. Zehnwirth (1994) in Section 7.2 discusses the importance of varying (stochastic) parameter models, especially the introduction of a varying alpha parameter (in place of adding parameters) to overcome multi-collinearity. This is akin to exponential smoothing in the accident year direction. This approach is necessary, very powerful and increases the stability of the model, especially if in the more recent accident years there are some slight changes in levels. The 'amount' of stochastic variations in alpha is determined by the SSPE statistic which is explained in Zehnwirth (1994).

3.4 Model Identification

The aim is to identify a parsimonious model in PTF that separates the (systematic) trends from the random fluctuations and moreover determines whether the trend in the payment/ calendar year direction is stable.

The model identification procedure is discussed in Section 10 of Zehnwirth (1994). We start off with a model that only has one parameter in each direction, model (sequentially) the trends in the development year direction followed by payment year or accident year directions, depending on which direction exhibits more dramatic trend changes. Heteroscedastic adjustments may also be necessary. Validation analysis is an integral component of model identification, extraction of information and testing for stability of trends.

3.5 Assumptions about the future

Stability and assumptions about the future are discussed in Section 9.6.2 and 10.2 of Zehnwirth (1994). If payment/calendar year trend has been stable in the more recent years, then the assumption about the future is relatively straightforward. For example, if the estimate in the last seven years of i, is $\hat{i} \pm se(\hat{i})$, then we assume for the future a mean trend of $\hat{i}$ with a standard deviation of trend of $se(\hat{i})$. We do not assume the trend in the future is constant. Our model includes the variability (uncertainty) in trend in the future in addition to the process variability (about the trend).

If on the other hand, payment/calendar year trend has been unstable as was illustrated with Project ABC, assumptions about the future will depend on the explanation for the instability - for Project ABC we revert to the 10% trend if the dramatic change is explained by new legislation. Zehnwirth (1994) also cites some other practical examples where special knowledge about the business is a contributing factor in formulating assumptions about the future, especially in the presence of trend instability. Importantly, however, that special knowledge is combined with the information that is extracted from past experience.

It is not possible to enumerate all possible cases, though several cases are discussed in Zehnwirth (1994). The more experience the actuary has with the new statistical paradigm, the better he/she is equipped to formulate assumptions about the future, in the presence of unstable trends. Bear in mind, of course, that quite often trends are stable. We only know after performing a PTF analysis.

3.6 How do we know that real data triangles are generated by the members of the rich PTF?

Let's conduct the following experiment. I give you 100 triangles, say, and for each triangle I tell you which model in the PTF is 'best'. So, corresponding to each triangle there is a corresponding 'best' fit model in PTF. You do your own testing and you agree. You can also conduct ELRF analysis on each triangle. Recall that a fitted (best) model relates the distributions of each cell in terms of trends on the log scale. Now, I tell you that some of the triangles are real company's data but some are not. That is, for some triangles the data represent a sample path from the so called "fitted" distributions. Which is real company's data and which is simulated data from the so called "fitted" model? As you cannot distinguish between real triangles and simulated triangles generated from models in the PTF, these kinds of models must be valid. That is, the rich family of models in PTF possess probabilistic mechanisms for generating real data. Of course, the models do not represent the complex underlying generating process that is driven by many variables. However, the variables that drive the data are implicitly included in the trends and the noise (σ^2). We do the same thing when we fit a loss distribution (e.g. Pareto) to a bunch of severities. The estimated Pareto did <u>not</u> create the severities. But, it has probabilistic mechanisms for creating the data as a sample.

4 The reserve figure

Loss reserves often constitute the largest single item in an insurer's balance sheet. An upward or downward 10% movement of loss reserves could change the whole financial picture of the company.

4.1 Prediction Intervals

We have argued for the use of probabilistic models, especially in assessing the variability or uncertainty inherent in loss reserves. The probability that the loss reserve, carried in the balance sheet, will be realized in the future, is necessarily zero, even if the loss reserve is the true mean!

Future (incremental) paid losses may be regarded as a sample path from the forecast (estimated) lognormal distributions. The estimated distributions include both process risk and parameter risk. Forecasting of distributions is discussed in Zehnwirth (1994).

The forecast distributions are accurate provided the assumptions made about the future will remain true. For example, if it is assumed that future payment/calendar year trend (inflation) has a mean of 10% and a standard deviation of 2%, and in two years time it turns out that inflation is 20%, then the forecast distributions are far from accurate.

Accordingly, any prediction interval computed from the forecast distributions is conditional on the assumptions about the future remaining true. The assumptions are in terms of mean trends, standard deviations of trends and distributions about the trends.

It is important to note that there is a difference between a fitted distribution and the corresponding predictive distribution. A predictive distribution necessarily incorporates parameter estimation error (parameter risk); a fitted distribution does not. Ignoring parameter risk can result in substantial underestimation of reserves and premiums. See the paper by Dickson, Tedesco and Zehnwirth (1998) for more details.

The distribution of sums, for example, accident year outstanding payments, is the distribution of a sum of lognormal variables that are correlated. The exact distribution of the sum can be obtained by generating (simulating) samples from the estimated multivariate lognormal distributions. The same could be done for payment year totals (important for obtaining the distributions of the future payment stream), or for the overall total. This information is relevant to Dynamic Financial Analysis. Distributions for future underwriting years can also be computed. This information is useful for pricing, including aggregate deductibles and excess layers.

Insurer's risk can be defined in many different ways. Most definitions are related to the standard deviation of the risk, in particular a multiple of the standard deviation.

If an insurer writes more than one long-tail line and aims for a 100 $(1-\alpha)$% security level on all the lines combined, then the risk margin per line decreases the more lines the company writes. This is always true, even if there exists some dependence (correlation) between the various lines.

Consider a company that writes n independent long tail lines. Suppose that the standard error of loss reserve $L(j)$ of line j is $s.e.(j)$. That is, $s.e.(j)$ is the standard error of the loss reserve variable $L(j)$. The standard error for the combined lines $L(1) +...+ L(n)$ is

$$s.e.(\text{Total}) = [s.e.^2(1)+...+ s.e.^2(n)]^{0.5}$$

If the risk margin for all lines combined is $k \times s.e.(\text{Total})$, where k is determined by the level of security required, then the risk margin for line j is

$$k \times s.e.(\text{Total}) \times s.e.(j)/ [s.e.(1)+...+ s.e.(n)] < k \times s.e.(j)\ (n).$$

The last inequality is true even when $s.e.(\text{Total})$ is not given by the above expression.

If as a result of analyzing each line using the statistical modelling framework, we find that for some lines, trends change in the same years and the changes are of the same order of magnitudes, then the lines are not independent. (There may also be some correlations between the residuals, but that would be negligible).

In that situation, line i and j are correlated, say, then one should use $s.e.(i) + s.e.(j)$ as the upper bound of the standard error of $L(i)+ L(j)$. (Based on our experience, it is not often that different lines are correlated in terms of trends.)

Suppose we assume for the future payment/calendar years a mean trend of $\hat{\iota}$ with a standard deviation (standard error) $s.e.(\hat{\iota})$. Specifically we are saying that the trend ι, a random variable, has a normal distribution with mean $\hat{\iota}$ and standard deviation $s.e.(\hat{\iota})$. Recognition of the relationship between the lognormal and normal distributions tells us that the mean payment increases as $s.e.(\hat{\iota})$ increases (and $\hat{\iota}$ remains constant). The greater the uncertainty in a parameter (the mean remaining constant), the more money is paid out. The same argument applies to the other estimated parameters in the model. This is what is known as Jensen's inequality, explained in college finance texts. See for example Brealey & Myers (1991). It is dangerous to ignore this concept.

4.2 Risk Based Capital

There are a number of misconceptions regarding risk-based capital. It is important to note that:

- The uncertainty in loss reserves (for the future) should be based on a probabilistic model (for the future) that may bear no relationship to reserves carried by the company in the past;

- The uncertainty for each line for each company should be based on a probabilistic model, derived from the company's experience, that describes the particular line for that company. A model appropriate for one loss development array will not be appropriate for another.

- The company's experience may bear very little relationship to the industry as a whole.

The approach discussed here allows the actuary to determine the relationships within and between companies' experiences and their relationships to the industry in terms of simple well understood features of the data.

In establishing the loss reserve, recognition is often given to the time value of money by discounting. The absence of discounting implies that the (median) estimate contains an implicit risk margin. But this implicit margin may bear no relationship to the security margin sought. The risk should be computed before discounting (at a zero rate of return).

4.3 Booking of Reserve

There are no hard and fast rules here. However, three very important steps are critical.

<u>Step 1</u>

Extract information, in terms of trends, stability thereof and distributions about trends, for the loss development array, in particular the incremental paid losses. Information is extracted by identifying the 'best' model in PTF. Model identification and extraction of information necessarily involves validation analysis (removal of past recent payment/calendar years).

<u>Step 2</u>

Assumptions about the future are formulated. If payment/calendar year trend is stable this is straightforward. If more recently trends are unstable, then an attempt is made to determine the cause by analyzing other data types, and using any relevant business knowledge. A number of examples are given in Zehnwirth (1994), but it is impossible to give an exhaustive list as each case may be different.

<u>Step 3</u>

Using the distributions of reserves, security margin sought on combined lines, risk capital available to the company, a percentile can be selected. Incidentally, the more uncertain the trends are for the future, the higher security margin may be called for.

4.4 Other benefits of the statistical paradigm

Finally, the statistical modelling framework has other benefits, including:

- Credibility models

  If a particular trend parameter estimate for an individual company is non-credible, it can be formally "shrunk" towards an industry estimate.

- Segmentation and layers

  Very often the statistical model identified for all payment types applies to some of the segments. By the same model we mean the same parameter structure but the estimates are not identical. Indeed, the variance of the normal distribution for a segment is larger than for the whole. These ideas can also be applied to territories etc. and to layers.

References

Brealy, R.A. and Myers, S.C. (1991) *Principles of Corporate Finance.* McGraw Hill Inc., New York.

Brosius, E. (1992), Loss Development using Credibility. *Casualty Actuarial Society Part 7 Exam Study Kit*

Dickson, D., L Tedesco and B. Zehnwirth (1998), Predictive Aggregate Claims Distributions, *Journal of Risk and Insurance* (to appear).

Halliwell, L. (199?), Loss Prediction by Generalized Least Squares,

Mack T. (1993). Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates. *ASTIN Bulletin*, Vol 23 No. 2. 213-225.

Mack T. (1994). Which stochastic model is underlying the chain ladder method? *Insurance Mathematics and Economics*, Vol 15 No. 2/3. 133-138.

Murphy D. M. (1994). Unbiased Loss Development Factors. *Proceedings of the Casualty Actuarial Society*, Vol LXXXI No. 144-155, 154-222.

Shapiro S.S. and Francia R.S. (1972). Approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 67, 215-216.

Venter G. (1996). Testing assumptions of age-to-age Factors. Draft Report, Instrat/Sedgwick Re, New York.

Zehnwirth, B. (1994) Probabilistic Development Factor Models with Applications to Loss Reserve Variability, Prediction Intervals and Risk Based Capital. *Casualty Actuarial Society Forum*, Spring 1994, Vol.2.

Appendices

Appendix A.1: Calculations for Link Ratio models with intercepts and accident year trends.

Let there be n accident years, numbering the most recent accident year as 0, and the first as n-1, as in Murphy (1994). Let y_{ij} be the cumulative amount paid in accident year i, development year j, $i=0,...,n-1$, $j=0,...,n-1$, as in Figure 1. This simplifies many of the formulas. Let $x_{ij} = y_{i,j-1}$, so that y_{ij}/x_{ij} is the observed development factor from j-1 to j in accident year i.

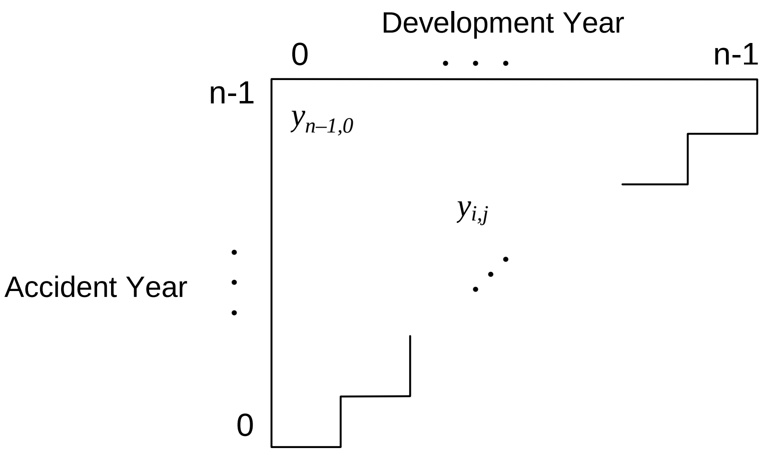


Figure A.1.1 Triangular loss development array of size n, with accident years labelled in reverse order.

The only difference a different array shape will make is to change the limits on summations.

Let $p_{ij} = \alpha_j + \lambda_j z_{ij} + (\beta_j - 1) x_{ij} + u_{ij}$, where p_{ij} is the incremental paid loss in accident year i, at development year j, (denote the cumulative by y_{ij}), z_{ij} is the count of accident years from the top, starting from 0; since we number from the bottom, in the current notation $z_{ij} = n-1-i$, and where x_{ij} is the cumulative paid in accident year i, up to development year j–1. Here, α_j is an intercept (level) term, λ_j represents accident year trend and β_j the dependence on the previous cumulative. We can also write the model as $y_{ij} = \alpha_j + \lambda_j z_{ij} + \beta_j x_{ij} + u_{ij}$, and we will proceed with this formulation of the model. As before, $\mathrm{Var}(u_{ij}) = \sigma_j^2 x_{ij}^{\delta}$.

The regressions are independent, so most of the calculations are straightforward.

Parameter estimates and standard errors

With 3 parameters in each regression, it will be easiest to use a standard regression routine. We now describe how to do a weighted regression with an unweighted routine.

Writing the j^{th} regression in matrix form (and dropping the j subscript), we have:
$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$, where $\boldsymbol{y} = (y_{n-1}, y_{n-2}, ..., y_j)'$, $\boldsymbol{\beta} = (\alpha, \lambda, \beta)'$, $\boldsymbol{u} = (u_{n-1}, u_{n-2}, ..., u_j)'$,

$$X = \begin{bmatrix} 1 & z_{n-1} & x_{n-1} \\ 1 & z_{n-2} & x_{n-2} \\ \vdots & \vdots & \vdots \\ 1 & z_j & x_j \end{bmatrix} \text{ and } \operatorname{Var}(\boldsymbol{u}) = \sigma^2 \boldsymbol{U} = \sigma^2 \begin{bmatrix} x_{n-1}^{\delta} & 0 & \cdots & 0 \\ 0 & x_{n-2}^{\delta} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & x_j^{\delta} \end{bmatrix}.$$

Let $\boldsymbol{y}^* = \boldsymbol{U}^{-1/2}\boldsymbol{y}$, $\boldsymbol{X}^* = \boldsymbol{U}^{-1/2}\boldsymbol{X}$, $\boldsymbol{e} = \boldsymbol{U}^{-1/2}\boldsymbol{u}$ Then we have $\boldsymbol{y}^* = \boldsymbol{X}^*\boldsymbol{\beta} + \boldsymbol{e}$, with the e_i's independent normal with variance σ^2. That is, $\boldsymbol{y}^* = (y_{n-1}x_{n-1}^{\delta/2}, y_{n-2}x_{n-2}^{\delta/2}, \ldots, y_j x_j^{\delta/2})'$ and similarly with each column of $\boldsymbol{X}$, including the column of 1's. The parameter estimates and variance-covariance matrix of this new regression are the same as that of the old regression.

Consequently, we will simply take $\hat{\alpha}, \hat{\lambda}, \hat{\beta}$ and their estimated variances and covariances as being available. Note that in the final development years it isn't possible to fit all three parameters; usually we would choose to fit either α and/or β as appropriate.

Residuals

Note that the residuals are the same whether we consider incremental or cumulative values – $\hat{u}_{ij} = y_{ij} - \hat{y}_{ij} = p_{ij} - \hat{p}_{ij}$. Again considering only the regression for development year j, $\operatorname{Var}(u_i) = \sigma^2[1-(x_i^*)'(X^{*\prime}X^*)^{-1}x_i^*]$, where x_i^* is the i^{th} row of $\boldsymbol{X}^*$.

Forecasts and standard errors

Forecasts: Obviously $\hat{y}_{i,i+k} = \hat{\alpha}_{i+k} + \hat{\lambda}_{i+k} z_{i,i+k} + \hat{\beta}_{i+k}\hat{y}_{i,i+k-1}$, where as usual $\hat{y}_{ii} = y_{ii}$.

Standard Errors: Conditioning on the data,

$$\begin{aligned} \operatorname{Var}(\hat{y}_{i,i+k} - y_{i,i+k}) &= \operatorname{Var}(\hat{y}_{i,i+k} - \mu_{i,i+k} + \mu_{i,i+k} - y_{i,i+k}) \\ &= \operatorname{Var}(\hat{y}_{i,i+k} - \mu_{i,i+k}) + \operatorname{Var}(y_{i,i+k} - \mu_{i,i+k}) \\ &= v_{i,i+k}^{p} + v_{i,i+k}^{e}, \text{ say} \end{aligned}$$

The first term is what Murphy calls the parameter variance, and the second the process variance. Now

$$
\begin{aligned}
v_{i,i+k}^{p} &= \mathrm{Var}(\hat{y}_{i,i+k}) \\
&= \mathrm{Var}(\hat{\alpha}_{i+k} + \hat{\lambda}_{i+k} z_{i,i+k} + \hat{\beta}_{i+k}\hat{y}_{i,i+k-1}) \\
&= \mathrm{Var}(\hat{\alpha}_{i+k}) + 2z_{i,i+k}\mathrm{Cov}(\hat{\alpha}_{i+k}, \hat{\lambda}_{i+k}) + z_{i,i+k}^{2}\mathrm{Var}(\hat{\lambda}_{i+k}) \\
&\quad + 2\hat{y}_{i,i+k-1}\mathrm{Cov}(\hat{\alpha}_{i+k}, \hat{\beta}_{i+k}) + 2z_{i,i+k}\hat{y}_{i,i+k-1}\mathrm{Cov}(\hat{\alpha}_{i+k}, \hat{\beta}_{i+k}) \\
&\quad + \mathrm{Var}(\hat{\beta}_{i+k}\hat{y}_{i,i+k-1}).
\end{aligned}
$$

As with Murphy (1994), we note that here

$$
\begin{aligned}
\mathrm{Var}(\hat{\beta}_{i+k}\hat{y}_{i,i+k-1}) = \hat{\beta}_{i+k}^{2}\mathrm{Var}(\hat{y}_{i,i+k-1}) + \hat{y}_{i,i+k-1}^{2}\mathrm{Var}(\hat{\beta}_{i+k}) \\
+ \mathrm{Var}(\hat{\beta}_{i+k})\mathrm{Var}(\hat{y}_{i,i+k-1}).
\end{aligned}
$$

Hence

$$
\begin{aligned}
v_{i,i+k}^{p} &= \mathrm{Var}(\hat{\alpha}_{i+k}) + 2z_{i,i+k}\mathrm{Cov}(\hat{\alpha}_{i+k}, \hat{\lambda}_{i+k}) + z_{i,i+k}^{2}\mathrm{Var}(\hat{\lambda}_{i+k}) \\
&\quad + 2\hat{y}_{i,i+k-1}\mathrm{Cov}(\hat{\alpha}_{i+k}, \hat{\beta}_{i+k}) + 2z_{i,i+k}\hat{y}_{i,i+k-1}\mathrm{Cov}(\hat{\alpha}_{i+k}, \hat{\beta}_{i+k}) \\
&\quad + \hat{y}_{i,i+k-1}^{2}\mathrm{Var}(\hat{\beta}_{i+k}) + [\hat{\beta}_{i+k}^{2} + \mathrm{Var}(\hat{\beta}_{i+k})]v_{i,i+k-1}^{p}.
\end{aligned}
$$

We estimate this by substituting the estimated variance of β_{i+k} in for the variance above. Note that v_{ii}^{p} is zero, since we are conditioning on the data.

$$
\begin{aligned}
v_{i,i+k}^{e} &= \mathrm{Var}(y_{i,i+k} - \mu_{i,i+k}) \\
&= \mathrm{Var}(\alpha_{i+k} + \lambda_{i+k} z_{i,i+k-1} + \beta_{i+k} y_{i,i+k-1} + u_{i,i+k}) \\
&= \beta_{i+k}^{2}\mathrm{Var}(y_{i,i+k-1}) + \mathrm{Var}(u_{i,i+k}) \\
&= \beta_{i+k}^{2}\mathrm{Var}(y_{i,i+k-1} - \mu_{i,i+k-1}) + \sigma_{i+k}^{2} x_{i,i+k}^{\delta} \\
&= \beta_{i+k}^{2} v_{i,i+k-1}^{e} + \sigma_{i+k}^{2} x_{i,i+k}^{\delta}
\end{aligned}
$$

which we estimate as:

$$
\begin{aligned}
\hat{v}_{i,i+k}^{e} &= \widehat{\mathrm{Var}}(y_{i,i+k}) \\
&= \hat{\beta}_{i+k}^{2}\widehat{\mathrm{Var}}(y_{i,i+k-1}) + \hat{\sigma}_{i+k}^{2}\hat{\mathrm{E}}(y_{i,i+k-1}^{\delta}) \\
&= \hat{\beta}_{i+k}^{2}\hat{v}_{i,i+k}^{e} + \hat{\sigma}_{i+k}^{2}\hat{f}_{i,i+k-1}^{\delta}.
\end{aligned}
$$

where $f_{i,j}^{\delta} = \mathrm{E}(y_{i,j}^{\delta})$. We estimate the process variance as:

$$\widehat{\mathrm{Var}}(y_{i,i+k}) = \hat{\beta}_{i+k}^{2}\ \widehat{\mathrm{Var}}(y_{i,i+k-1}) + \hat{\sigma}_{i+k}^{2}\hat{\mathrm{E}}(y_{i,i+k-1}^{\delta})$$

$$= \hat{\beta}_{i+k}^{2}\ \widehat{\mathrm{Var}}(y_{i,i+k-1}) + \hat{\sigma}_{i+k}^{2}\hat{f}_{i,i+k-1}^{\delta}, \quad \text{say},$$

where

$$\hat{f}_{ij}^{\delta} = \begin{cases} 1, & \delta = 0 \\ \hat{y}_{ij}, & \delta = 1 \\ \hat{y}_{ij}^{\ 2} + \widehat{\mathrm{Var}}(y_{ij}), & \delta = 2 \end{cases}$$

just as with Murphy (1994); with the normality assumption we can obtain estimates at other values of δ, but we'll omit details here.

Forecasts and standard errors on the paid scale can be obtained in similar fashion.

Forecasts and Standard Errors of Development Year Totals

Forecasts:
Let D_j be the unknown future development year total forecast, so:

$$D_j = \sum_{i=0}^{j-1} y_{ij}\text{, and}$$

$$\hat{D}_j = \sum_{i=0}^{j-1} \hat{y}_{ij}$$

Standard Errors:

$$\mathrm{Var}(\hat{D}_j - D_j) = \mathrm{Var}(\sum_{i=0}^{j-1} \hat{y}_{ij} - y_{ij})$$

$$= \mathrm{Var}(\sum_{i=0}^{j-1} \hat{y}_{ij} - \mu_{ij}) + \mathrm{Var}(\sum_{i=0}^{j-1} y_{ij} - \mu_{ij})$$

$$= V_j^{p} + V_j^{e}.$$

$$
\begin{aligned}
V_j^p &= \mathrm{Var}(\sum_{i=0}^{j-1} \hat{y}_{ij}) \\
&= \mathrm{Var}(\sum_{i=0}^{j-1} \hat{\alpha}_j + \hat{\lambda}_j z_{ij} + \hat{\beta}_j \hat{y}_{i,j-1}) \\
&= \mathrm{Var}(n_j \hat{\alpha}_j + Z_j \hat{\lambda}_j + \hat{\beta}_j [\hat{D}_{j-1} + y_{j-1,j-1}])
\end{aligned}
$$

(where $Z_j = \sum_{i=0}^{j-1} z_{ij}$)

$$
\begin{aligned}
&= n_j^2 \mathrm{Var}(\hat{\alpha}_j) + 2n_j Z_j \mathrm{Cov}(\hat{\alpha}_j, \hat{\lambda}_j) + Z_j^2 \mathrm{Var}(\hat{\lambda}_j) \\
&\quad + 2n_j [\hat{D}_{j-1} + y_{j-1,j-1}] \mathrm{Cov}(\hat{\alpha}_j, \hat{\beta}_j) \\
&\quad + 2Z_j [\hat{D}_{j-1} + y_{j-1,j-1}] \mathrm{Cov}(\hat{\lambda}_j, \hat{\beta}_j) + \mathrm{Var}(\hat{\beta}_j [\hat{D}_{j-1} + y_{j-1,j-1}]) \\
&= n_j^2 \mathrm{Var}(\hat{\alpha}_j) + 2n_j Z_j \mathrm{Cov}(\hat{\alpha}_j, \hat{\lambda}_j) + Z_j^2 \mathrm{Var}(\hat{\lambda}_j) \\
&\quad + 2[\hat{D}_{j-1} + y_{j-1,j-1}][n_j \mathrm{Cov}(\hat{\alpha}_j, \hat{\beta}_j) + Z_j \mathrm{Cov}(\hat{\lambda}_j, \hat{\beta}_j)] \\
&\quad + [\hat{D}_{j-1} + y_{j-1,j-1}]^2 \mathrm{Var}(\hat{\beta}_j) + [\hat{\beta}_j^2 + \mathrm{Var}(\hat{\beta}_j)] \mathrm{Var}(\hat{D}_{j-1}) \\
&= n_j^2 \mathrm{Var}(\hat{\alpha}_j) + 2n_j Z_j \mathrm{Cov}(\hat{\alpha}_j, \hat{\lambda}_j) + Z_j^2 \mathrm{Var}(\hat{\lambda}_j) \\
&\quad + 2[\hat{D}_{j-1} + y_{j-1,j-1}][n_j \mathrm{Cov}(\hat{\alpha}_j, \hat{\beta}_j) + Z_j \mathrm{Cov}(\hat{\lambda}_j, \hat{\beta}_j)] \\
&\quad + [\hat{D}_{j-1} + y_{j-1,j-1}]^2 \mathrm{Var}(\hat{\beta}_j) + [\hat{\beta}_j^2 + \mathrm{Var}(\hat{\beta}_j)] V_{j-1}^p
\end{aligned}
$$

We estimate this by replacing the variance and covariance terms by their estimates. Also,

$$
\begin{aligned}
V_j^e &= \mathrm{Var}(\sum_{i=0}^{j-1} y_{ij}) \\
&= \mathrm{Var}(n_j \alpha_j + Z_j \lambda_j + \beta_j \sum_{i=0}^{j-1} y_{i,j-1} + \sum_{i=0}^{j-1} u_{i,j}) \\
&= \mathrm{Var}(\beta_j [D_{j-1} + y_{j-1,j-1}]) + \mathrm{Var}(\sum_{i=0}^{j-1} u_{i,j}) \\
&= \beta_j^2 \mathrm{Var}(D_{j-1}) + \sigma_j^2 \sum_{i=0}^{j-1} x_{ij}^{\delta} \\
&= \beta_j^2 V_{j-1}^e + \sigma_j^2 (y_{j-1,j-1}^{\delta} + \sum_{i=0}^{j-2} y_{i,j-1}^{\delta})\,.
\end{aligned}
$$

Due to independence across accident years, $\mathrm{E}(\sum_{i=0}^{j-2} y_{i,j-1}^{\delta}) = \sum_{i=0}^{j-2} \mathrm{E}(y_{i,j-1}^{\delta}) = \sum_{i=0}^{j-2} f_{i,j-1}^{\delta}$ so the process variance term is estimated by $\hat{V}_j^e = \hat{\beta}_j^2 \hat{V}_{j-1}^e + \hat{\sigma}_j^2 (y_{j-1,j-1}^{\delta} + \sum_{i=0}^{j-2} \hat{f}_{i,j-1}^{\delta})$. The estimated standard error of $\hat{D}_j$ is then $\sqrt{\hat{V}_j^p + \hat{V}_j^e}$.

Appendix A.2: Exposures and Forecasting

Let y_{ij}^o be the observed cumulative at accident year *i*, development year *j*, let y_{ij}^n be the corresponding normalized-for-exposures cumulative. Similarly, let p_{ij}^o and p_{ij}^n be the corresponding paid values.

Let c_i be the exposure for accident year *i*. Then $y_{ij}^n = y_{ij}^o / c_i$, and $p_{ij}^n = p_{ij}^o / c_i$. We fit the OLRT model to y^n, but we use it to forecast y^o.

Individual Forecasts and Standard Errors (all models).

$\hat{y}_{ij}^o = c_i \hat{y}_{ij}^n$, and similarly for *p*, $\hat{p}_{ij}^o = c_i \hat{p}_{ij}^n$, and
$\mathrm{Var}(\hat{y}_{ij}^o - y_{ij}^o) = c_i^2 \mathrm{Var}(\hat{y}_{ij}^n - y_{ij}^n)$ and $\mathrm{Var}(\hat{p}_{ij}^o - p_{ij}^o) = c_i^2 \mathrm{Var}(\hat{p}_{ij}^n - p_{ij}^n)$.

So individual observed forecasts and standard errors are just the corresponding normalized values multiplied by the exposure.

Cumulative Development Year Totals

$$\mathrm{Var}(\hat{D}_j^o - D_j^o) = V_j^{op} + V_j^{oe}$$

$$\begin{aligned}
V_j^{op} &= \mathrm{Var}(\textstyle\sum_{i=0}^{j-1} \hat{y}_{ij}^o) \\
&= \mathrm{Var}(\textstyle\sum_{i=0}^{j-1} c_i \hat{y}_{ij}^n) \\
&= \mathrm{Var}[\textstyle\sum_{i=0}^{j-1} c_i (\hat{\alpha}_j + \hat{\lambda}_j z_{ij} + \hat{\beta}_j \hat{y}_{i,j-1}^n)] \\
&= \mathrm{Var}[\hat{\alpha}_j (\textstyle\sum_{i=0}^{j-1} c_i) + \hat{\lambda}(\textstyle\sum_{i=0}^{j-1} c_i z_{ij}) + \hat{\beta}_j \textstyle\sum_{i=0}^{j-1} \hat{y}_{i,j-1}^o] \\
&= \mathrm{Var}[\hat{\alpha}_j C_j + \hat{\lambda}_j Z_j^* + \hat{\beta}_j \textstyle\sum_{i=0}^{j-1} \hat{y}_{i,j-1}^o]
\end{aligned}$$

$$
\begin{aligned}
&= C_j^2 \mathrm{Var}(\hat{\alpha}_j) + 2C_j Z_j^* \mathrm{Cov}(\hat{\alpha}_j, \hat{\lambda}_j) + (Z_j^*)^2 \mathrm{Var}(\hat{\lambda}_j) \\
&\quad + 2C_j (y_{j-1,j-1}^o + \hat{D}_{j-1}^o) \mathrm{Cov}(\hat{\alpha}_j, \hat{\beta}_j) \\
&\quad + 2Z_j^* (y_{j-1,j-1}^o + \hat{D}_{j-1}^o) \mathrm{Cov}(\hat{\lambda}_{\mathrm{j}}, \hat{\beta}_j) \\
&\quad + (y_{j-1,j-1}^o + \hat{D}_{j-1}^o)^2 \mathrm{Var}(\hat{\beta}_j) + [\hat{\beta}_j^2 + \mathrm{Var}(\hat{\beta}_j)] \mathrm{Var}(\hat{D}_{j-1}^o) \\
&= C_j^2 \mathrm{Var}(\hat{\alpha}_j) + 2C_j Z_j^* \mathrm{Cov}(\hat{\alpha}_j, \hat{\lambda}_j) + (Z_j^*)^2 \mathrm{Var}(\hat{\lambda}_j) \\
&\quad + 2(y_{j-1,j-1}^o + \hat{D}_{j-1}^o)[C_j \mathrm{Cov}(\hat{\alpha}_j, \hat{\beta}_j) + Z_j^* \mathrm{Cov}(\hat{\lambda}_{\mathrm{j}}, \hat{\beta}_j)] \\
&\quad + (y_{j-1,j-1}^o + \hat{D}_{j-1}^o)^2 \mathrm{Var}(\hat{\beta}_j) + [\hat{\beta}_j^2 + \mathrm{Var}(\hat{\beta}_j)] V_{j\text{-}1}^{op}
\end{aligned}
$$

$$
\begin{aligned}
V_j^{oe} &= \mathrm{Var}(\textstyle\sum_{i=0}^{j-1} y_{ij}^o) \\
&= \mathrm{Var}(\textstyle\sum_{i=0}^{j-1} c_i y_{ij}^n) \\
&= \mathrm{Var}[\textstyle\sum_{i=0}^{j-1} c_i (\alpha_j + \lambda_j z_{ij} + \beta_j y_{i,j-1}^n + u_{ij})] \\
&= \mathrm{Var}(\textstyle\sum_{i=0}^{j-1} c_i u_{ij}) + \hat{\beta}_j^2 \mathrm{Var}[\textstyle\sum_{i=0}^{j-1} y_{i,j-1}^o] \\
&= \textstyle\sum_{i=0}^{j-1} \mathrm{Var}(c_i u_{ij}) + \hat{\beta}_j^2 \mathrm{Var}[\textstyle\sum_{i=0}^{j-1} y_{i,j-1}^o] \\
&= \sigma_j^2 c_{j-1}^2 (y_{j-1,j-1}^n)^\delta + \textstyle\sum_{i=0}^{j-2} c_i^2 v_{i,j-1}^{ne} + \hat{\beta}_j^2 V_{j-1}^{oe} \\
&= \sigma_j^2 c_{j-1}^{2-\delta} (y_{j-1,j-1}^o)^\delta + \textstyle\sum_{i=0}^{j-2} v_{i,j-1}^{oe} + \hat{\beta}_j^2 V_{j-1}^{oe}
\end{aligned}
$$

Appendix A.3: Likelihood and Conditional regressions

Let $y(j)$ be the data in development year j. Let $\theta(j)$ be all the parameters for that development year. Let $\boldsymbol{y} = (y(0),...,y(n-1))$ and $\theta = (\theta(1), ..., \theta(n-1))$. Then straight-forward application of conditional probability and some simplification gives us:

$$
\begin{aligned}
L(\theta|\boldsymbol{y}) &\propto \mathrm{p}[\boldsymbol{y}\,|\theta] \\
&\propto \mathrm{p}[y(n-1)|\theta(n-1), \mathrm{y}(n-2), \mathrm{y}(n-3), \ldots, \mathrm{y}(0)] \\
&\qquad \cdot \mathrm{p}[y(n-2)|\theta(n-2), \mathrm{y}(n-3), \mathrm{y}(n-4), \ldots, \mathrm{y}(0)] \\
&\qquad\qquad \vdots \\
&\qquad \cdot \mathrm{p}[y(1)|\theta(1), y(0)] \\
&\qquad \cdot \mathrm{p}[y(0)] \\
&\propto \mathrm{p}[y(n-1)|\theta(n-1), \mathrm{y}(n-2)]\cdot \mathrm{p}[y(n-2)|\theta(n-2), \mathrm{y}(n-3)]\cdot\ldots\cdot \mathrm{p}[y(1)|\theta(1), y(0)]\cdot \mathrm{p}[y(0)]
\end{aligned}
$$

Since, for each regression, we are conditioning on the data from previous develop-ment years, the fact that the previous development data is stochastic and not fixed is not an issue – the conditional likelihoods still correspond to ordinary regressions.

The likelihood for y(0) doesn't contain any of the parameters. At any value for θ, then, the likelihood of y(0) is just a constant; consequently it cannot affect the location of the maximum of the likelihood, nor its curvature there. So the way that the forecasts depend on the parameters isn't affected by y(0) apart from the way it enters the regression for y(1).

The model for the data says that the values in future development years depend on the earlier development years. We've observed the whole of y(0), so we know exactly how it will impact the future runoff, because the model describes that. Of course, the model may be wrong (and we argue that it is), but given the model, the regressions may all be performed as ordinary regressions.

The forecasts are made conditionally on the data. We've argued above that even the stochastic nature of $y(0)$ can be ignored in the forecasting because the model fully describes its impact on the future observations. However, this is not an important point – if an argument were made that the stochastic nature of $y(0)$ should somehow affect the forecasts, it would not affect any of our arguments about the unsuitability of these models.

Appendix A.4: Design Matrices for the models described in Part 2.

Readers may wish to fit the regression models described in Part 2 of this paper. The models described there can be fitted to data in any of the more common statistical packages, of course, or in a spreadsheet such as Excel. Here we briefly describe what the various predictors look like. We begin by describing the full model (which is not used in practice – it's overparameterized), and then some of the more common simpler models.

Let us examine the expected values in each cell in the log(incremental) array under the general model; using the notation of Part 2:

	0	1	…	j	…	$n-1$
0	α_0	$\alpha_0+\gamma_1+\iota_1$	$\cdots$	$\alpha_0+\sum_{k=1}^{j}\gamma_k+\sum_{r=1}^{j}\iota_r$	$\cdots$	$\alpha_0+\sum_{k=1}^{j}\gamma_k+\sum_{r=1}^{j}\iota_r$
1	$\alpha_1+\iota_1$	$\alpha_1+\gamma_1+\sum_{r=1}^{2}\iota_t$	$\cdots$	$\alpha_1+\sum_{k=1}^{j}\gamma_k+\sum_{r=1}^{j+1}\iota_r$	$\cdots$	
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\iddots$	
i	$\alpha_i+\sum_{r=1}^{i}\iota_r$	$\alpha_1+\gamma_1+\sum_{r=1}^{i+1}\iota_r$		$\alpha_i+\sum_{k=1}^{j}\gamma_k+\sum_{r=1}^{i+j}\iota_r$		
$\vdots$	$\vdots$	$\vdots$	$\iddots$			
$n-1$	$\alpha_{n-1}+\sum_{r=1}^{t}\iota_r$					

Table A.4.1 Expected values of log(incremental) under the general model of Part 2.

Produce the vector of observations by stacking up the development years one on top of another: $\boldsymbol{y} = (y(0)', y(1)', \ldots, y(n-1)')'$. Similarly, produce a column in the X-matrix for each parameter, and the parameters become a column with rows in the same order as the corresponding columns of the X-matrix (design matrix). Note that α is an intercept parameter, so we don't add an intercept. That is, the regression is written $\boldsymbol{y} = X\boldsymbol{\beta}+\boldsymbol{\varepsilon}$. A good approach is to do all the α's, then all the γ's and then all the ι's.

For n=4, this corresponds to the X-matrix below (the zeroes have been suppressed to make the patterns more clear):

	α_0	α_1	α_2	α_3	γ_1	γ_2	γ_3	ι_1	ι_2	ι_3
y(0,0)	1									
y(1,0)	1				1			1		
y(2,0)	1				1	1		1	1	
y(3,0)	1				1	1	1	1	1	1
y(0,1)		1						1		
y(1,1)		1			1			1	1	
y(2,1)		1			1	1		1	1	1
y(0,2)			1					1	1	
y(1,2)			1		1			1	1	1
y(0,3)				1				1	1	1

Table A.4.2 Design matrix (X-matrix) for the full model for a triangle with 4 years' data. The zeroes have been suppressed.

In general, the (i,j) row for an array of size n would have a 1 for the column for α_j, it would have 1s for the columns for γ_k, where $k \leq j$, and it would have 1s for the columns for ι_r, where $r \leq i+j$, with zeroes everywhere else.

Setting some of the parameters to be equal is simply a matter of adding together columns from the full design matrix. For example, here is the design matrix for the array of size 4, with level of log payments for all years, with two development year trends (0-1, and all later years), and a single payment year trend – that is, all α's equal, $\gamma_2 = \gamma_3$, and all ι's equal.

	α	γ_1	γ_{2+}	ι
y(0,0)	1			
y(1,0)	1	1		1
y(2,0)	1	1	1	2
y(3,0)	1	1	2	3
y(0,1)				1
y(1,1)		1		2
y(2,1)		1	1	3
y(0,2)				2
y(1,2)		1		3
y(0,3)				2

Table A.4.3 Design matrix (X-matrix) for a simple model for a triangle with 4 years' data. The zeroes have been suppressed.