

# Pricing and Reserving Multiple Excess Layers

Ben Zehnirth

Glen Barnett and Michael Kunkler

## Abstract

Patterns and changing trends among several excess-type layers on the same business tend to be closely related. The changes in trend often occur at the same time in each layer. Therefore a good model for the changing trends in the ground-up data will often be a good model for the various layers. As might be expected, after adjusting for the trends, the corresponding values in each layer are still generally correlated. These relationships among layers should be taken into account when reserving or pricing excess layers.

We describe a framework within which such relationships can be explicitly incorporated and taken account of when reserving or pricing several excess layers. This framework allows for reserving and pricing for multiple layers with changes in trends in the same periods, correlations between layers, and possibly equal trends or trend changes across layers.

## 1. Introduction

Changes in trends in loss development arrays (generally triangles) occur in three directions – along the accident periods, the development periods and the payment periods.

When dealing with loss arrays, it is important to look at the incremental data. That way, changes in trends in the payment period direction, which affect the incremental losses directly, can be observed and modeled. Identifying changes in this direction can have a very large effect on the forecast of outstanding losses. Additionally, unidentified changes in this direction interfere with our ability to correctly identify trends and changes in trends in the other directions.

Trends in loss development arrays occur in multiplicative fashion. For example, inflation and superimposed inflation occur as multiplicative trends in the payment period direction. The tail in the development direction is usually exponential. These effects encourage us to model the logarithms in the data, so changes in trend become additive. Additionally, the data generally have larger standard deviation when the mean is larger, and often exhibit skewness. Both of these features are usually absent after taking logarithms. Hence, after adjusting for inflation (as appropriate) and normalizing for exposure, we take logarithms before modeling the trends in the data.

It is often necessary to reserve or price several excess-type layers. Naturally these are intimately related to the original (ground up) loss array, and to each other. Indeed, the changes in trends for the data for each layer will occur in the same places, though the changes will often be different in size, and the variability about the trends will be different. If we model the trends using regression models, the models are the same for each layer, but the identified parameters may be different. The layers, even after adjusting for trends, will often be correlated with each other.

It is possible to take advantage of the fact that the trends move together, leading to faster modeling, since identified trend change in one array (often the original loss array is best) apply to all of the associated layers. Some estimated trends may not be credible for an individual sublayer taken by itself, but when all the information is taken together, and after appropriate model reduction, the estimate can become more precise. Moreover, knowledge of the correlation between layers can be important when forecasting.

## 2 The Model

### 2.1 Model for a single layer

Let us take the array to be a full triangle, covering  $s$  periods (years). The formulas given here are easy to modify for triangles missing the early payment years, or the later accident or development years, but the discussion is clearer if we present the simplest array.

Let  $y_{w,d}$  be the lognormalized incremental payment in accident period  $w$ , development period  $d$ .

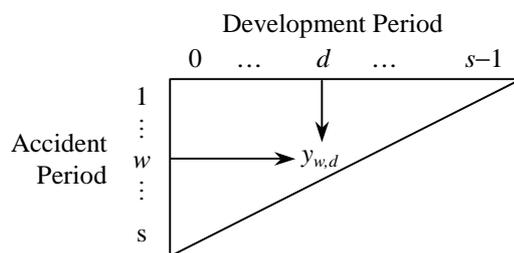


Figure 1. Lognormalized incremental data array of dimension  $s$ .

Incremental data exhibits trends in three directions - accident year, payment year and development year. Development year trends are normally thought of as the "run-off" pattern. Payment year trends represent the superimposed inflation (if the data are already adjusted for economic inflation). Accident years can move up and down in level. Trends in these directions are related - a trend in the payment year direction will also project onto the accident year and development year directions, and vice versa.

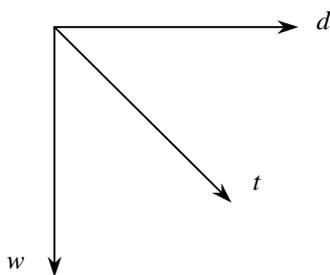


Figure 2. Trend directions

The payment year variable  $t$  can be expressed as  $t = w + d$ . That is, the payment year is just the sum of the accident and development years. This relationship between the three directions implies that there are only two independent directions.

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 normalizing for exposures) shows little change over many years, requiring only a few parameters. The parameters in the development year direction represent the trend in run-off from one development year to the next. This trend is often linear (on the log scale) across many of the later development years, requiring few parameters (often only one) 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. We see that very often only a few parameters are required to describe the trends in the data. 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.

Consider a single accident year (of log-data). We represent the expected level in the first development year by a parameter ( $\alpha$ ). We can model the trends across the development years by allowing for a (possible) parameter to represent the expected change (or trend) between each pair of development years ( $\gamma$  represents the average change between one development period and the next). We model the variation of the data about this process with a zero-mean normally distributed random error. That is:

$$y_d = \alpha + \sum_{j=1}^d \gamma_j + \varepsilon_d,$$

where  $\varepsilon_d$  is normally distributed with mean zero and variance  $\sigma^2$  – i.e.  $\varepsilon_d \sim N(0, \sigma^2)$ .

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

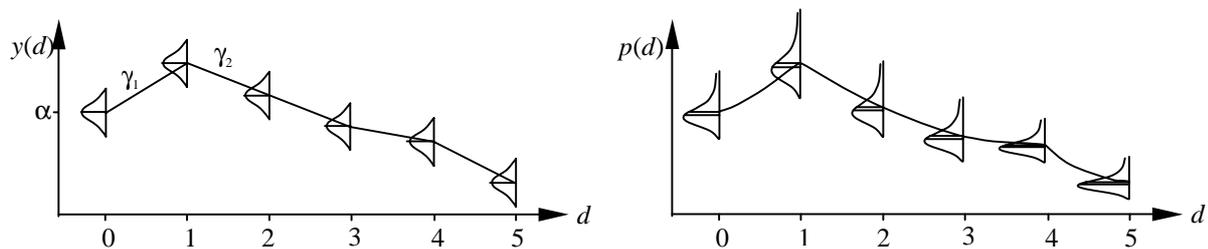


Figure 3. Probabilistic model for trends along a development year on the log scale and the original scale.

If there is no superimposed inflation or accident year changes, the above type of model may be suitable for a real run-off triangle. However, usually there are changing trends in these directions.

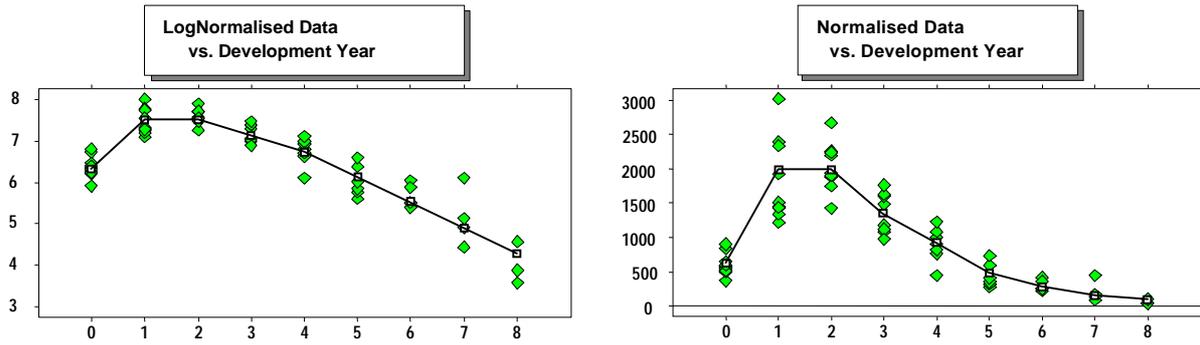


Figure 4. Plots of real data (log scale and original scale), normalized for exposures and adjusted for economic inflation against delay with the means from a fitted model. This data has no accident year trends or superimposed inflation.

To model the trends in the three directions, we require three sets of parameters. The level of accident period  $w$  in the first development period will be represented by  $\alpha_w$ . The *change* in level between development period  $d-1$  and  $d$  will be represented by  $\gamma_d$ . The change in level between payment period  $t-1$  and  $t$  will be represented by  $\iota_t$ . The basic model is

$$y_{w,d} = \alpha_w + \sum_{j=1}^d \gamma_j + \sum_{h=2}^{w+d} \iota_h + \varepsilon_{w,d} \quad \varepsilon_{w,d} \sim N(0, \sigma^2) \quad w = 1, \dots, s, \quad d = 0, \dots, s-1$$

This model may be fitted as an ordinary regression model. If there is indication of changing variance against development period, we may model the variance process and use generalized least squares regression to fit the above model. For example, sometimes the variance is larger in the later development periods.

Note that for any given triangle not all of the parameters are actually used – most of them are set to zero (omitted from the regression) – most arrays require very few parameters, and a model that included all of them would be highly overparameterized.

### Implementation as a regression model

The observations are stacked up in some order. For example, the data may be sorted by payment period, and within that by development, so that a new period of data arrives it is simply stacked under the existing data. However, any ordering will do. Let's call the row at which observation at  $(w, d)$  appears row  $i$ .

The parameter vector consists of the parameters for the levels for the accident years, the development year trend parameters and the payment year trend parameters. A predictor variable is constructed corresponding to each

parameter. Each predictor has a '1' if the corresponding parameter is in the model for that observation, and '0' if it is not.

Shown below is the parameter vector for the above model, and below that is shown the row of the set of predictors corresponding to the observation at  $(w, d)$ :

$$\beta = (\alpha_1, \dots, \alpha_{w-1}, \alpha_w, \alpha_{w+1}, \dots, \alpha_s, \gamma_1, \dots, \gamma_d, \gamma_{d+1}, \dots, \gamma_{s-1}, \iota_2, \dots, \iota_{w+d}, \iota_{w+d}, \dots, \iota_s)'$$

$$x_i = (0, \dots, 0, 1, 0, \dots, 0, 1, \dots, 1, 0, \dots, 0, 1, \dots, 1, 0, \dots, 0)$$

### Example

The purpose of this example is to illustrate the relationship between the layout of the triangle, the parameters and the predictor variables. We have the following (small) triangle of incremental paid losses, adjusted for inflation and normalized for exposures.

		$\gamma_1 \downarrow$	$\gamma_1 + \gamma_2 \downarrow$
		$\iota_2 \swarrow$	$\iota_2 + \iota_3 \swarrow$
$\alpha_1 \rightarrow$	96	83	42
$\alpha_2 \rightarrow$	118	74	
$\alpha_3 \rightarrow$	109		

If we were to begin with all possible parameters in the model, the data and predictor (independent) variables could be set up as follows:

P	Y	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\gamma_1$	$\gamma_2$	$\iota_2$	$\iota_3$
96	4.56435	1	0	0	0	0	0	0
83	4.41884	1	0	0	1	0	1	0
42	3.73767	1	0	0	1	1	1	1
118	4.77069	0	1	0	0	0	1	0
74	4.30406	0	1	0	1	0	1	1
109	4.69135	0	0	1	0	0	1	1

Because the level is captured for each year by the  $\alpha$  parameters, it is necessary to omit the constant term when doing the regression.

A typical triangle will only need a few parameters in each direction. If a model is missing a parameter it should have, a pattern will appear in the residuals against the direction the parameter relates to. For example, if only a single trend is fitted in the development year direction, but the run-off pattern in the data has a change in trends, then that will be clear in the graph of standardized residuals against development years. On the other hand, if an

unnecessary parameter is included, the difference between the unnecessary parameter and the one before will not be significantly different from zero.

## **2.2 Model for multiple layers**

When dealing with multiple layers, a good model is chosen for the data from the ground up, or the biggest layer available, and then applied to all the other layers. After checking that the model fits for all the layers, correlations between layers can be estimated by calculating the correlation between the residuals for each layer.

From the correlations and the variances for each layer, the variance-covariance matrix between layers can be estimated. Generalized least squares (GLS) can then be used to estimate the combined model to all layers at once. Some parameters may be set to be equal between adjacent layers.

Some trends in a model estimated to a layer on its own, particularly to a narrow excess layer, may have large standard error, making the estimate not significantly different from zero. However, it will not generally be appropriate to set the trend to zero (especially if it represents a rate of superimposed inflation that might reasonably be expected to continue into the future). We simply don't have a credible estimate. However, once a final model has been estimated to all the layers and appropriate reduction of the model undertaken (for example, some parameters set to be equal to each other across layers), often the worrisome trend estimate can have a much smaller standard error.

Many packages will perform GLS. See the Appendices for more details. The generalized least squares model can also be implemented in a standard regression package by transforming the predictors. This is outlined in Appendix 4.

## **2.3 Forecasting and reserving**

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 are, and 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.

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.

#### Forecasting a single layer model

On the log scale, forecasts operate just as with any normal regression model. Using standard regression results, forecasts and standard errors on the log scale are:  $\hat{y}_i = x_i \hat{\beta}$ ,  $\text{Var}(\hat{y}_i - y_i) = \sigma^2 (x_i' (X'X)^{-1} x_i + 1)$  (call this  $v_i$ ), and  $\text{Cov}(\hat{y}_i, \hat{y}_j) = \sigma^2 x_i' (X'X)^{-1} x_j$  ( $= c_{ij}$ , say). The  $X$  here is the matrix of predictors for the single layer model, of course.

On the original (dollar) scale: if there are no exposures, and no adjustments for future inflation, the formulas are

$\hat{P}_i = e^{\hat{y}_i + \frac{1}{2}v_i}$ ,  $\text{Var}(\hat{P}_i) = \hat{P}_i^2 (e^{v_i} - 1)$ , and  $\text{Cov}(\hat{P}_i, \hat{P}_j) = \hat{P}_i \hat{P}_j (e^{c_{ij}} - 1)$ . These results follow directly from the formulas for the mean and variance of a lognormal distribution. If there are exposures, the fitted values ( $\hat{P}_i$ ) will be multiplied by the appropriate exposure for their year, and that carries through the variances and covariances by adjusting the  $\hat{P}$  terms by the same factor; inflation factors, multiplied out to the appropriate payment year inflate the values in analogous fashion.

Forecasts of sums (e.g. accident period totals, payment period totals, overall total), their standard deviations, and covariance between totals can be obtained using the well known formulas for expectation and variances of sums of random variables.

#### Forecasting multilayer models

Once the parameter estimates are available, forecasts on the log scale are produced in the obvious fashion, as they would be for a single layer model, but with the parameter estimates from the combined model. Variances and covariances are obtained as with any GLS model. See Appendix 3. From the variance-covariance matrix of forecasts, individual variances and covariances may be obtained directly, and the variance of a single element, or covariances between elements obtained in the usual way, resulting in formulas similar to those given for a single layer.

Forecasts on the original (dollar) scale:

Calculations proceed from the log-scale forecasts in the same fashion as given for the single layer.

## 2.4 Predictive Aggregate Loss Distributions and Value-at-Risk

The distribution of sums of payments, for example, accident year outstanding payments, is the distribution of a sum of correlated lognormal variables, under the model. Generally when forecasting accident or payment year totals, or the overall outstandings, more information than just mean and standard deviation is required. It is not possible to compute the distribution of the various totals analytically. Nor is it reasonable to assume that the total for any year is well approximated by a normal distribution. Given the small number of values in the sum, the typical degrees of skewness in the forecast distributions, and the correlations between forecasts, it is rare that the sample size will be sufficient to apply the central limit theorem to the distribution of the total. 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 process can be repeated over and over to obtain the entire distribution of any desired total to whatever level of accuracy is required.

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 Zehnwrith (1998) for more details.

With knowledge of the entire predictive distribution of an aggregate forecast, it is a relatively simple matter to perform Value-at-Risk calculations. For any given quantile, the Value at Risk is simply that quantile minus the provision. Alternatively, the maximum dollar loss at the  $100(1-\alpha)\%$  confidence level is the V-a-R for the  $(1-\alpha)$  quantile.

If a reinsurer writes more than one layer on a single piece of long-tail business and aims for a  $100(1-\alpha)\%$  security level – selecting the  $(1-\alpha)$  quantile – on all the layers combined, then the risk margin per layer decreases the more layers the company writes. This is always true, even though there is some dependence (and so correlation) between the various layers.

Consider a company that writes  $n$  layers. Suppose that the standard error of loss reserve  $L(j)$  of layer  $j$  is  $s.e.(j)$ , and the correlation between layers  $i$  and  $j$  is  $r_{ij}$ . Under the assumption that the identified common trend changes are the only ones that will occur in the future, the standard error for the combined layers  $L(1) + \dots + L(n)$  is

$$s.e.(Total) = [\sum_j s.e.^2(j) + 2\sum_{i<j} r_{ij} s.e. (i) s.e. (j)]^{0.5}$$

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

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

Of course, in general, we might expect layers to have similar future payment year trend changes, as they generally do in the past. In that situation, it would be useful to model the combined data from all of the layers in question. Of course, where unobserved trend changes in layer  $i$  and  $j$  are related, then one could use  $s.e.(i) + s.e.(j)$  as the upper bound of the standard error of  $L(i) + L(j)$ .

## 2.5 Pricing a future year

### Pricing an existing layer

If the layer in question is one observed in the past, or it can be constructed from information to hand, after making some appropriate assumption about the corresponding exposure and for the  $\alpha$  parameter for the new year (the two assumptions are related, as the decision boils down to deciding about the change in the difference between the log-exposure and the new level parameter from the last year), forecasts for the new accident year can be obtained exactly as outlined in section 2.3. The total and standard deviation for that year could be used to calculate the price for the corresponding layer, if a premium based on the mean, or mean +  $k$  standard deviations is desired.

As before, we can also simulate from the distribution of the total for the future accident year, allowing pricing to be based on any percentile of the predictive distribution of the aggregate, and can be useful for planning/pricing excess-type reinsurance on the aggregate for that line of business.

### Estimating mean and standard deviation for an unobserved layer

If there is only one limit that is varying (say, all layers being considered have a lower limit of zero, or no upper limit), then we can simply examine the relation between the means of the price (and the standard deviation) and the limit that varies, and fit that relationship with a smooth curve. For example, if we look at it graphically:

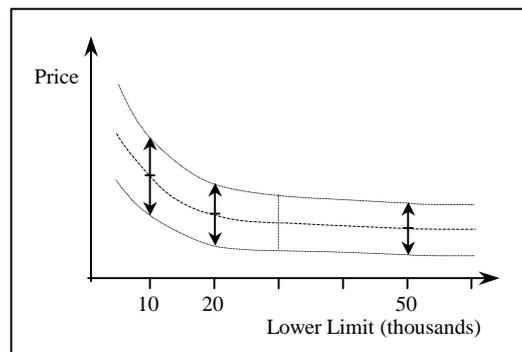


Figure 5. Plot of mean price  $\pm$  one standard deviation, with smooth curves and interpolated values at an upper limit of 30,000.

If both limits are varying, then we would need to examine the mean vs limit (and standard deviation vs limit) relationships in two dimensions, which may require modeling of data for many pairs of limits to get enough information for the analysis.

### 3 Examples

Example 1: XS10 and P40X10

These are two reinsurance layers on the same business. Looking at a plot of the changing trends in the data, we see:

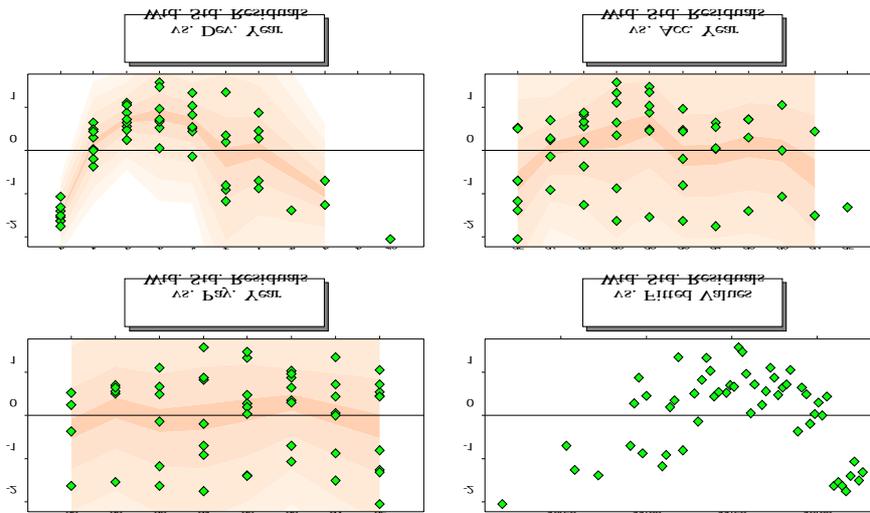


Figure 6: XS10's changing trends. The average trend in the development year and payment year directions has been removed.

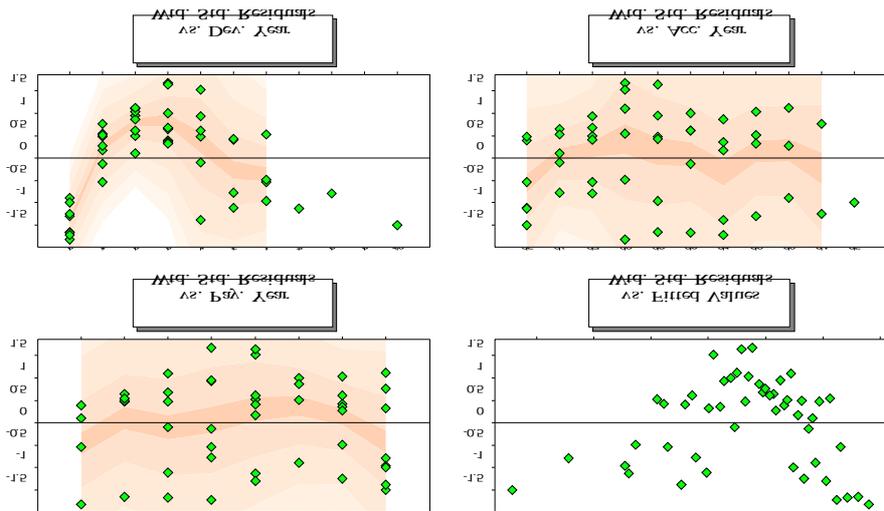


Figure 7: P40X10's changing trends. The average trend in the development year and payment year directions has been removed. Note how the trend changes occur in the same years in both arrays.

Initial model selection

We will find a suitable model for one array (XS10) and use it as the starting point for a combined analysis. The model has three trends in the development year direction (0-1, 1-3, and after 3). It has two accident year levels - one before 1990, and one from 1990. There is superimposed inflation in the middle payment years (1991-93), but it is zero everywhere else. There is heteroscedasticity in the data - the variance after delay 4 is larger, so correspondingly less weight is given to observations after that development year - about 0.219 of the weight to the earlier observations.

XS10 Alpha Parameter Estimates							40X10 Alpha Parameter Estimates						
Acci Yr	Alpha	S.E.	T-Ratio	Diff.	S.E.	T-Ratio	Acci Yr	Alpha	S.E.	T-Ratio	Diff.	S.E.	T-Ratio
1985-89	10.322	0.2248	45.92				1985-89	10.4695	0.2453	42.67			
1990-95	8.8355	0.2513	35.16	-1.4865	0.2474	-6.01	1990-95	8.9959	0.2853	31.53	-1.4736	0.297	-4.96

Gamma Parameter Estimates							Gamma Parameter Estimates						
Dev. Yr	Gamma	S.E.	T-Ratio	Diff.	S.E.	T-Ratio	Dev. Yr	Gamma	S.E.	T-Ratio	Diff.	S.E.	T-Ratio
1	2.4952	0.2119	11.77				1	2.391	0.2299	10.4			
2-3	0.0000	0.0000	-	-2.4952	0.2119	-11.77	2-3	0.0000	0.0000	-	-2.391	0.2299	-10.4
4-10	-1.0688	0.0904	-11.82	-1.0688	0.0904	-11.82	4-10	-1.142	0.1058	-10.79	-1.142	0.1058	-10.79

Iota Parameter Estimates							Iota Parameter Estimates						
Pay. Yr	Iota	S.E.	T-Ratio	Diff.	S.E.	T-Ratio	Pay. Yr	Iota	S.E.	T-Ratio	Diff.	S.E.	T-Ratio
1989-90	0.0000	0.0000	-				1989-90	0.0000	0.0000	-			
1991-93	0.4919	0.0893	5.51	0.4919	0.0893	5.51	1991-93	0.3994	0.1057	3.78	0.3994	0.1057	3.78
1994-95	0.0000	0.0000	-	-0.4919	0.0893	-5.51	1994-95	0.0000	0.0000	-	-0.3994	0.1057	-3.78

Table 1. Parameter estimates for the same model estimated on the two layers.

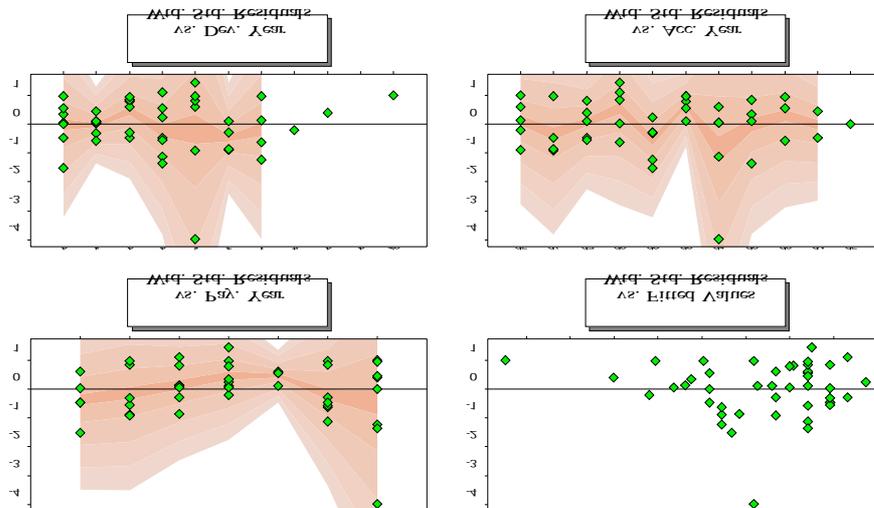


Figure 8. Residual plots from the fitted model to the XS10 data.

Estimation of the correlation between arrays:

We will estimate a single correlation between the two arrays. An initial examination of correlation against delay – after combining some years to get adequate sample sizes – gives the impression of fairly constant correlation except that it looks low in delays 6-10, taken together. After removing a single, somewhat discrepant, observation in the tail of the XS10 project (which is also the  $x$ -component of point A below), the correlation does not appear to change with delay. If the observation is retained, there is a minor indication that there may be a decrease (the confidence intervals around the estimated correlations are quite wide). We proceed with the more parsimonious assumption – there is no need to introduce an extra level of complication based on an impression caused by a single point.

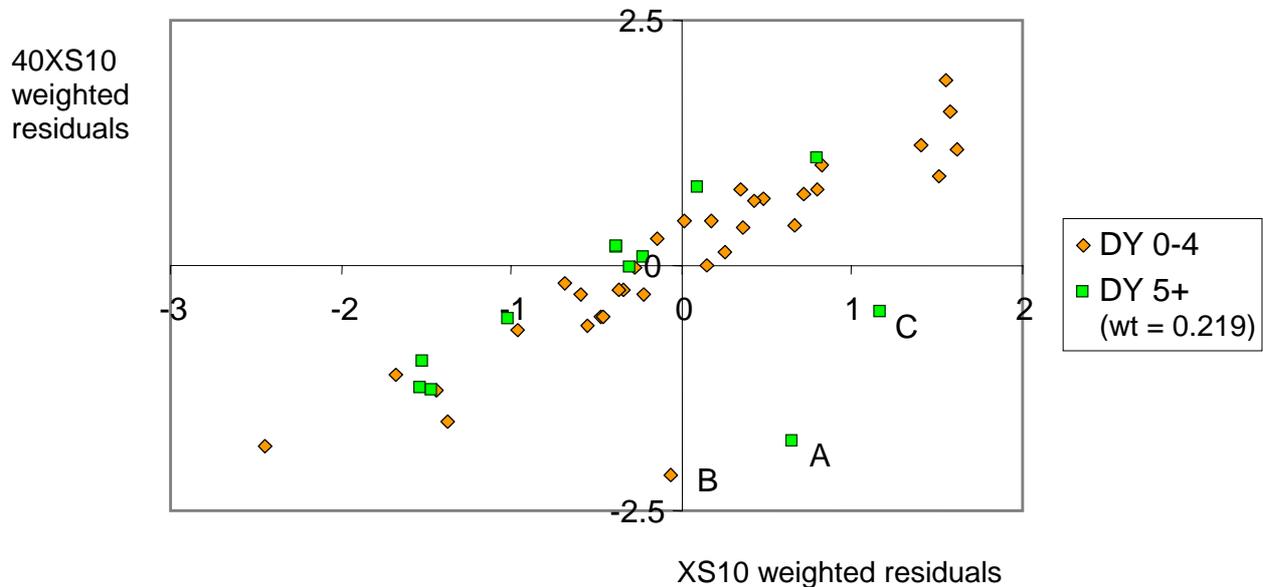


Figure 9. Plot of corresponding weighted residuals for the two layers. Points A, B and C possibly don't fit the main relationship between layers.

Taking into account the reduced weight to the observations after delay 4, the overall correlation between the residuals from the two arrays is 0.781. Without point A, the correlation is 0.853. Without the three marked points, the correlation is 0.957. We will proceed with correlation based on all of the data. The regression estimates are given in Table 4:

Combined model

		XS10			40X10-XS10		
	Years	Estimate	S.E.	t-ratio	Diff.	S.E.	t-ratio
Alpha	1985-89	10.3281	0.2100	49.19	0.1327	0.1481	0.90
	1990-95	8.7682	0.2332	37.61	0.1146	0.1742	0.66
Gamma	0:1	2.4750	0.1980	12.50	-0.1075	0.1389	-0.77
	3:10	-1.0811	0.0810	-13.34	0.0183	0.1012	0.18
Iota	1991-93	0.5250	0.0822	6.39	-0.0684	0.0652	-1.05

Table 4. Parameter estimates for the combined model

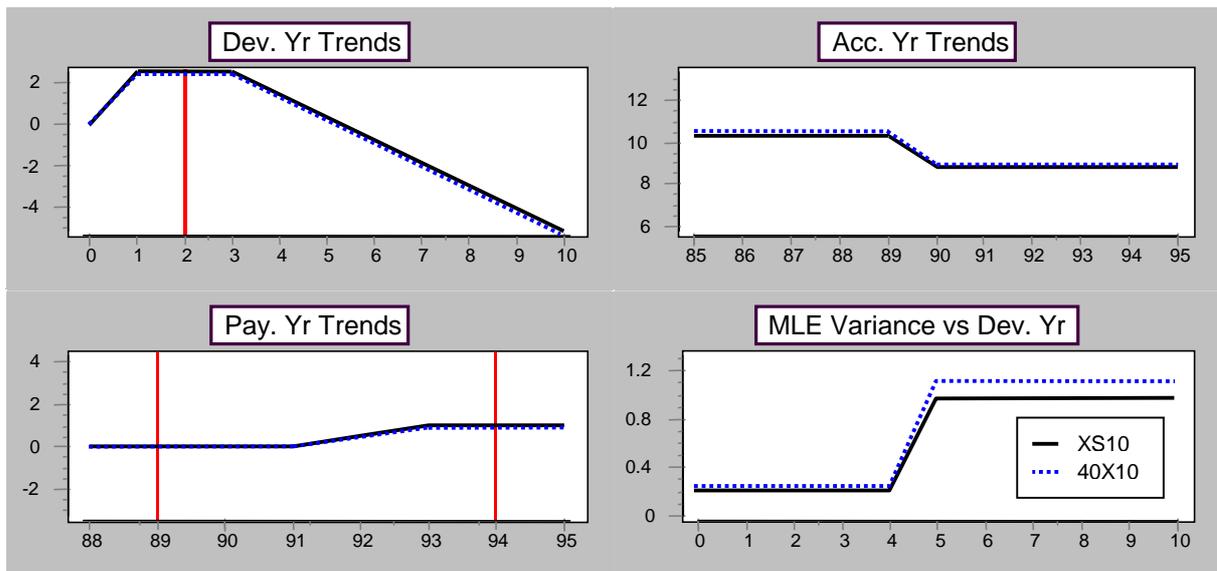


Figure 10. Fitted trends in the three directions and variances for the combined model. The fitted trends are very similar.

Clearly not all the differences are required. Proceeding to eliminate the least significant parameters and re-estimate at each stage, we obtain a final model as shown below. A mere six parameters describe the trends in three directions for both layers.

Final model:

		XS10			40X10-XS10		
	Years	Estimate	S.E.	t-ratio	Diff.	S.E.	t-ratio
Alpha	1985-89	10.3745	0.2006	51.71	0.0000	0.0000	-
	1990-95	8.8064	0.2239	39.33	0.0000	0.0000	-
Gamma	0:1	2.4386	0.1891	12.90	0.0000	0.0000	-
	3:10	-1.0841	0.0773	-14.02	0.0000	0.0000	-
Iota	1991-93	0.5212	0.0796	6.55	-0.0546	0.0242	-2.26

Table 3. Parameter estimates for the reduced model.

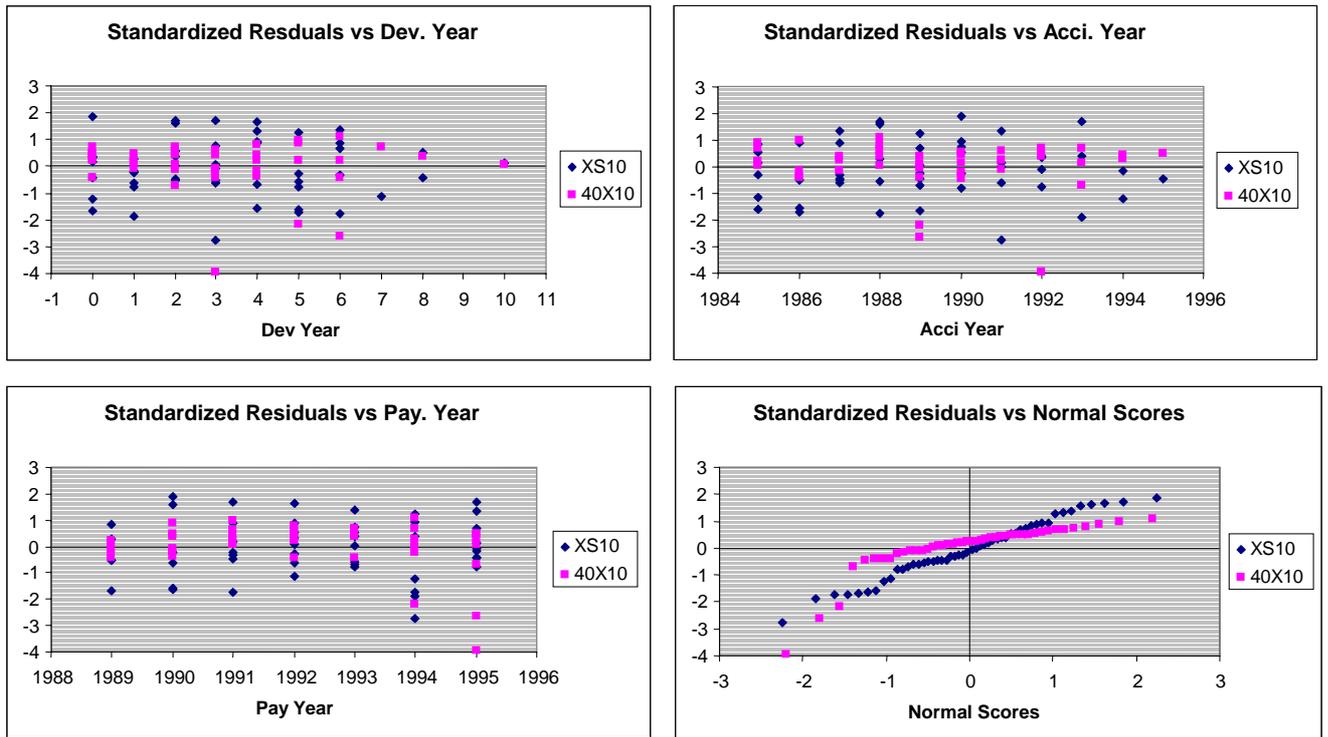


Figure 11. Residual plots against the three directions, and normal scores plots of the residuals for each layer. The plots look reasonable, except, perhaps, for a few possible outliers in the left tail of the 40X10 layer.

ALL2M and ALL1M

These are two reinsurance layers (limited to 2 million and 1 million) of a long-tail line of business. All2M is the base layer for this analysis. As we see from the figure below, the changes in trends in the two layers seem to occur in the same places.

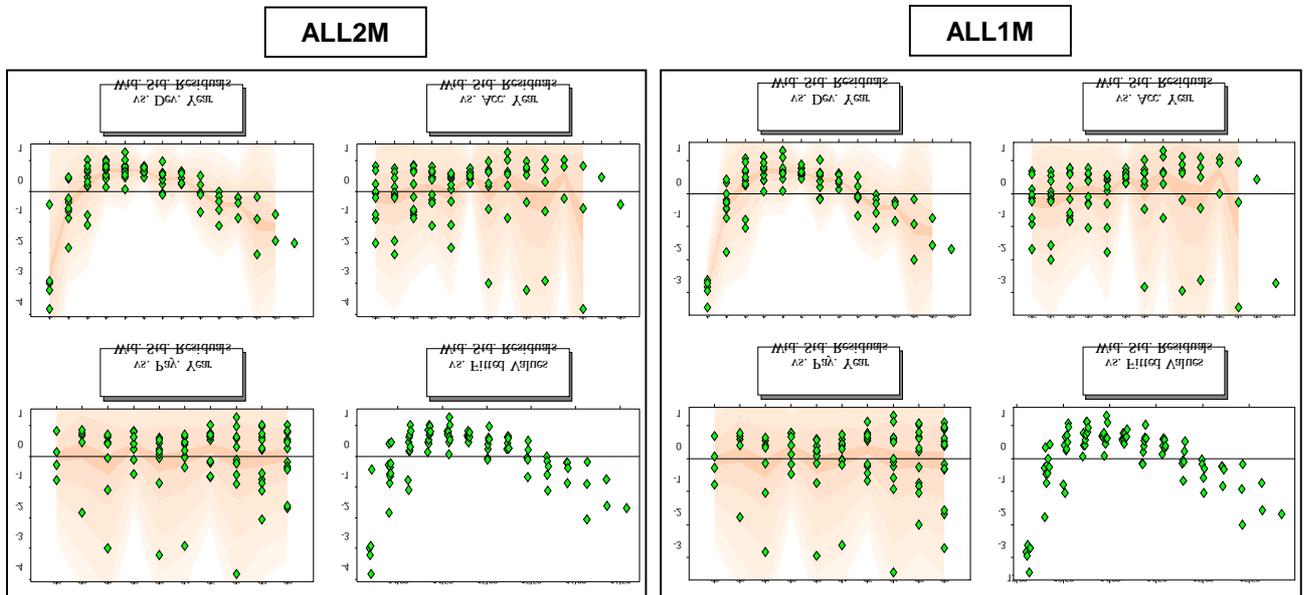


Figure 12. Plot of residuals against the three directions and fitted values for the two layers. The average trend in the payment year and development year directions has been removed so that the changes in trend can be seen.

Consequently, it makes sense to fit the same model (though with possibly different parameters) to both layers. Table 1 lists the parameters for the model identified for the base layer and fitted to both layers. Note that there is no superimposed inflation in any of the models in the analysis of these arrays –  $\iota$  (iota) is zero for all years.

Alpha Parameter Estimates

Acci. Year	All2M						All1M					
	Alpha	S.E.	t-ratio	Diff.	S.E.	t-ratio	Alpha	S.E.	t-ratio	Diff.	S.E.	t-ratio
1985-89	8.1322	0.3316	24.52				7.6229	0.3300	23.10			
1990-96	8.7119	0.2957	29.46	0.5797	0.1501	3.86	8.3914	0.2943	28.52	0.7686	0.1494	5.15
1997-98	10.1524	0.6908	14.70	1.4405	0.6243	2.31	9.6168	0.6874	13.99	1.2254	0.6212	1.97

Gamma Parameter Estimates

Dev. Year	All2M						All1M					
	Gamma	S.E.	t-ratio	Diff.	S.E.	t-ratio	Gamma	S.E.	t-ratio	Diff.	S.E.	t-ratio
0:1	4.4912	0.3570	12.58				4.4275	0.3552	12.46			
1:2	1.5704	0.2593	6.06	-2.9207	0.5214	-5.60	1.5157	0.2581	5.87	-2.9207	0.5214	-5.60
2:4	0.6755	0.1035	6.52	-0.8950	0.3280	-2.73	0.6374	0.1030	6.19	-0.8950	0.3280	-2.73
4:7	0.0000	0.0000	–	-0.6755	0.1035	-6.52	0.0000	0.0000	–	-0.6755	0.1035	-6.52
7:13	-0.1820	0.0535	-3.40	-0.1820	0.0535	-3.40	-0.1911	0.0532	-3.59	-0.1820	0.0535	-3.40

Table 4. Parameter estimates for the same model estimated on the two layers.

The next step is to estimate the correlation between the layers. The correlation is high – 0.99027 – so a stable regression algorithm becomes more important in this case. We then estimate the combined model in a single regression.

Combined model

Acci. Year	All2M			All1M – All2M		
	Alpha	S.E.	t-ratio	Diff. in Alpha	S.E.	t-ratio
1985-89	8.1322	0.3316	24.52	-0.5093	0.0462	-11.03
1990-96	8.7119	0.2957	29.46	-0.3204	0.0412	-7.78
1997-98	10.1524	0.6908	14.70	-0.5355	0.0962	-5.57

Dev. Year	All2M			All1M – All2M		
	Gamma	S.E.	t-ratio	Diff. in Gamma	S.E.	t-ratio
0:1	4.4912	0.3570	12.58	-0.0637	0.0497	-1.28
1:2	1.5704	0.2593	6.06	-0.0548	0.0361	-1.52
2:4	0.6755	0.1035	6.52	-0.0381	0.0144	-2.64
4:7	0.0000	0.0000	–	0.0000	0.0000	–
7:13	-0.1820	0.0535	-3.40	-0.0092	0.0074	-1.23

Table 5. Parameter estimates for the combined model – this is the same model as in Table 4; the correlation between layers doesn't affect the parameter estimates or standard errors.

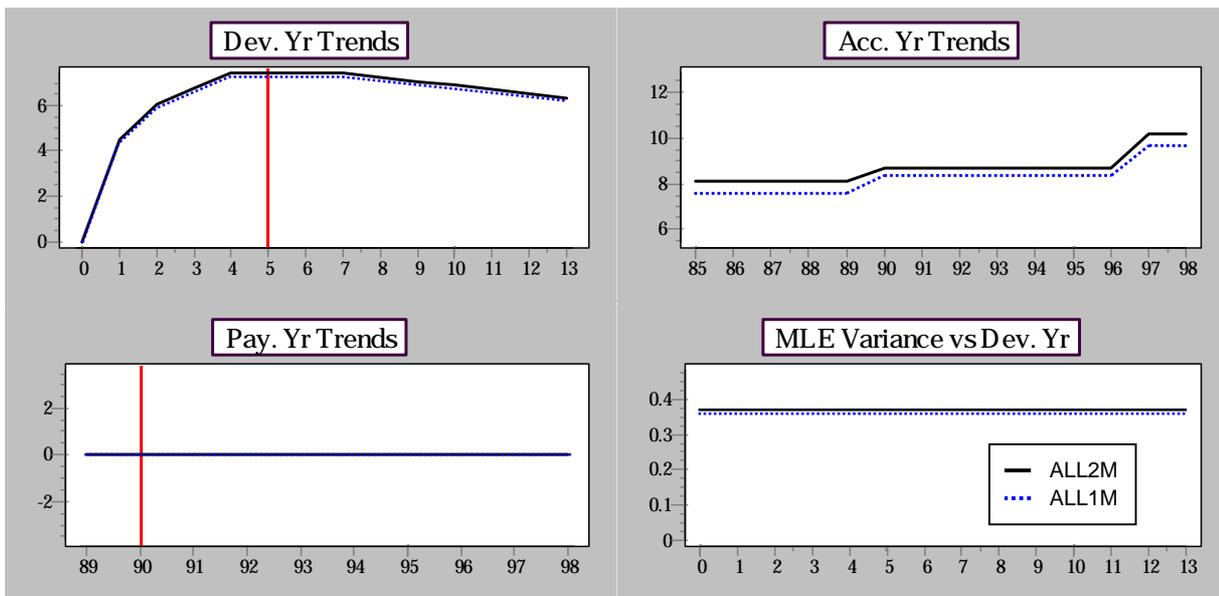


Figure 13. Fitted trends in the three directions and variances for the combined model. The fitted trends are again very similar.

Only one of the parameters for the difference between layers is significantly different from zero in the full model. As we eliminate non-significant parameters, the standard errors of the remaining parameters decrease. After eliminating the difference between layers for the first and last gamma parameters (so percentage changes between development years 0-1 and after development year 7 are treated as the same for the two layers), the remaining difference parameters are all significant. The final model is below.

Final Model

Acci. Year	All2M			All1M – All2M		
	Alpha	S.E.	t-ratio	Diff. in Alpha	S.E.	t-ratio
1985-89	8.1717	0.3316	24.65	-0.5617	0.0285	-19.73
1990-96	8.7440	0.2955	29.59	-0.3630	0.0231	-15.70
1997-98	10.2004	0.6920	14.74	-0.5993	0.0826	-7.25

Dev. Year	All2M			All1M – All2M		
	Gamma	S.E.	t-ratio	Diff. in Gamma	S.E.	t-ratio
0:1	4.4432	0.3561	12.48	0.0000	0.0000	–
1:2	1.5832	0.2599	6.09	-0.0716	0.0329	-2.18
2:4	0.6777	0.1038	6.53	-0.0410	0.0142	-2.89
4:7	0.0000	0.0000	–	0.0000	0.0000	–
7:13	-0.1889	0.0533	-3.54	0.0000	0.0000	–

Table 6. The final combined model for All1M and All2M.

It is also important to examine some diagnostics for the combined model. Residual plots against development years, accident years, and payment years, and normal scores plots are below.

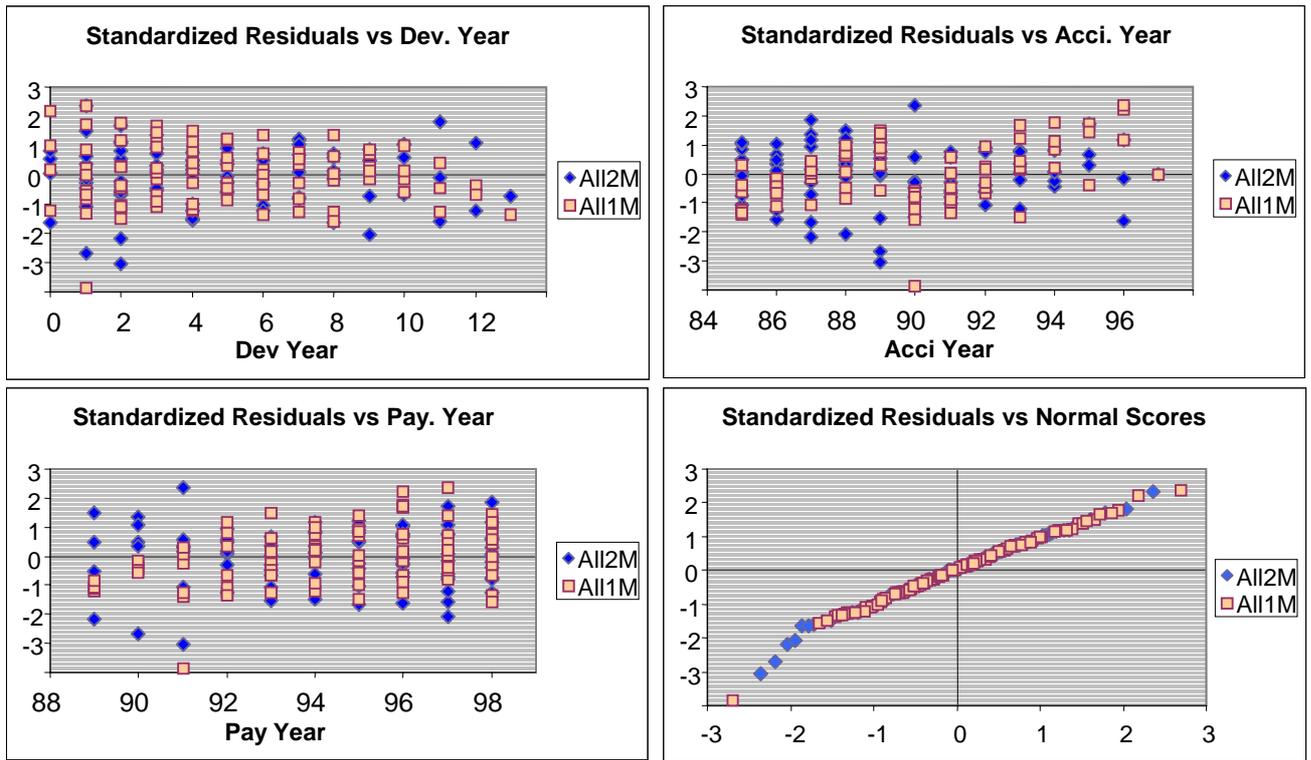


Figure 14. Residual plots against the three directions, and normal scores plots of the residuals for the two layers combined.

It appears that the changes between accident years 1989 and 1990 are not being picked up as well as before, even though there is a parameter there for each layer. Otherwise the diagnostics look reasonable.

## References

- Dickson, D., L Tedesco and B. Zehnwirth (1998), Predictive Aggregate Claims Distributions, *Journal of Risk and Insurance* (to appear).
- Searle, S.R. (1997) *Linear Models*. Wiley Classics. John Wiley and Sons, 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 1: Multilayer models

Let there be  $k$  data triangles (one for each layer).

Call the lognormalized data from the first array  $Y_1$ , where  $Y_1 = (y_{11}, y_{12}, \dots, y_{1n})'$ , with the observations stacked up in order. Similarly, call the log-normalized data for the other layers  $Y_2, \dots, Y_k$ . Note that the two subscripts here refer to the layer and the position in the vector, and not directly to their position in the triangle.

Let  $y_1 = (y_{11}, y_{21}, \dots, y_{k1})'$  be all of the first observations. Further, let  $y_2, y_3, \dots, y_n$  be the 2nd, 3rd, ...  $n$ th observations. Let  $V_t$  be the variance-covariance matrix of the error terms for the  $k$  observations making up  $y_t$ . Then the model becomes:

$$y_t = X_t \beta_t + e_t \quad \text{Var}(e_t) = V_t, \quad t = 1, 2, \dots, n$$

where  $X_t$  and  $\beta_t$  are chosen depending on how the model is parameterized. For example, each layer may have its own parameters as with the individual regressions, or you might model each layer in terms of its differences from the base layer, or in terms of differences from the previous layer. This last approach makes it easy to detect when parameters don't change across adjacent layers. For that parameterization,  $X_t$  and  $\beta_t$  are defined below.

Let  $Y = (y_1', y_2', \dots, y_n)'$  be all of the observations, where we have all the 'first' observations, then all the 'second' observations, and so on. Note that it is possible to stack the observations up in the other way, first array above second array, and so on –  $Y^* = (Y_1', Y_2', \dots, Y_k)'$ , and this might be easier if you are using a package – but the first formulation can make the computations simpler if you are implementing the algorithm yourself. If you change the formulation to that layer-by-layer approach, then the  $X$ 's and  $\beta$ 's are rearranged correspondingly. The same values are all present, they just move in the matrix. The first formulation will be retained for now as it makes it easier to discuss the relationships between layers, albeit one observation at a time.

The regression for all of the arrays (layers) is then:

$$Y = X\beta + e, \quad \text{Var}(e) = V,$$

where  $X = (X_1', X_2', \dots, X_n)'$ ,  $\beta = \beta_1 (= \beta_2 = \dots = \beta_n)$ , and, suppressing zero elements,

$$V = \begin{bmatrix} V_1 & & & & \\ & V_2 & & & \\ & & V_3 & & \\ & & & \ddots & \\ & & & & V_n \end{bmatrix}$$

because  $V$  is of a special form – block diagonal – the regression can be implemented quite efficiently. With at least some of the  $V_t$  equal (normally they are all equal), any of several appropriately chosen algorithms will be fast. Note that  $V$  is estimated from the variances to the layers and the correlation(s) between layers. If the correlations are very high, numerical stability will be a more important consideration than speed, and the generalized least squares algorithm must be selected for numerical stability. Some packages don't take great care to implement numerically stable algorithms, and they may fail if the correlations between layers are very high. However most dedicated statistical packages are stable.

## Appendix 2: The form of $X_t$ and $\beta_t$

Let us take the form where we deal with differences between layers directly. Let  $x_{it}$  be the design vector and  $\beta_{it}$  the parameter vector for  $y_{it}$ , and let  $x_{i-j,t}$  be the design vector and  $\beta_{i-j,t}$  the parameter vector for the difference between layers  $y_{it} - y_{jt}$ . Note that when the models are the same for each layer, as we will generally have,  $x_{1t} = x_{2t} \dots = x_{kt}$ , and the  $\beta$ 's are all of the same form. Similarly, if the difference between layers is not set to zero,  $x_{i-j,t}$  is also the same as  $x_{1,t}$  and the  $\beta$ 's for the differences (which are the differences in regression parameters for the layers) are also of the same form. Note that  $x_{1t}$  is a row vector and  $\beta_{1t}$  is a column vector. As before, blank elements indicate zeros.

$$X_t = \begin{bmatrix} x_{1t} & & & & & \\ x_{1t} & x_{2-1,t} & & & & \\ x_{1t} & x_{2-1,t} & x_{3-2,t} & & & \\ & \vdots & & \ddots & & \\ x_{1t} & x_{2-1,t} & x_{3-2,t} & \dots & x_{k-(k-1),t} & \end{bmatrix} \quad \beta_t = \begin{bmatrix} \beta_{1t} \\ \beta_{2-1,t} \\ \beta_{3-2,t} \\ \vdots \\ \beta_{k-(k-1),t} \end{bmatrix}$$

Even though  $X_t$  appears to be lower triangular above, it isn't, because the 'elements' are row vectors.

Note that when we multiply the  $i^{th}$  row of  $X_t$  by  $\beta_t$  that we obtain the mean of the base layer plus the difference between the second layer and the base layer, plus the difference between the next layer and that one, and so on. For example, for the third row (which represents the  $t^{th}$  observation in layer 3), we obtain

$$E(y_{3t}) = x_{1t}\beta_{1t} + x_{2-1,t}\beta_{2-1,t} + x_{3-2,t}\beta_{3-2,t} = E(y_{1t}) + E(y_{2t} - y_{1t}) + E(y_{3t} - y_{2t})$$

When we wish to set two parameters to be equal, we delete the columns of  $X$  and rows of  $\beta$  corresponding to the difference between them. With this parameterization, testing that a difference in parameters for two layers is equal to zero is simply a matter of testing that the parameter for the difference is zero.

### Appendix 3: Formulas for multilayer forecasts

The mean forecast is given by  $\hat{Y}_f = X_f \hat{\beta}$  where  $X_f$  is the matrix of predictors for those observations being forecast, and  $\hat{\beta}$  is the estimated. The variance-covariance matrix of forecasts is  $\text{Var}(\hat{Y}_f) = X_f \text{Var}(\hat{\beta}) X_f' = X_f (X'V^{-1}X)^{-1} X_f'$ .

The above formulas apply whichever parameterization of  $X$  and  $\beta$  is used.

Note that if the original model from which  $V$  was estimated has sufficient sample size to estimate all the parameters in the model, captures all the features of the data, and the only change to the model has been to eliminate some parameters which are essentially zero, it should not be necessary to re-estimate any of the layer variances or inter-layer covariances since the residual variances will be estimates of 1, and the residual covariances across layers will be estimates of 0. However, it is easy to incorporate updated variance estimates if required.

### Appendix 4: Implementation via regression for a two-layer model

Firstly the base layer is modeled, and then that model is fitted to the second layer and checked that it is a reasonable model for each. The correlation between the residuals for the layers, and the residual variances are calculated, from which an estimate of the covariance may be readily obtained.

Then the data from the two layers are put together into a single data set, but are transformed to independence and normalized to have equal variance. First the observations in each layer are divided by their standard deviation (giving  $y^*$ , say). The observations in the second layer are then further transformed to make them uncorrelated with the first layer:

$$y_{2t}^{**} = (y_{2t}^* - r y_{1t}^*) / (1 - r^2)$$

It doesn't really matter how you stack up the observations as long as you keep the relationships between the observations in each layer straight. It is easier to do it layer by layer when working this way. The  $x$ 's for each layer are transformed with that same transformation that was used for their corresponding observation. The parameter estimates produced from this transformed model are the same as those for the untransformed GLS model.