Volume LXXXVII, Part 2 **No. 167**

PROCEEDINGS
November 12, 13, 14, 15, 2000

BEST ESTIMATES FOR RESERVES

GLEN BARNETT AND BEN ZEHNWIRTH

Abstract

Link ratio techniques can be regarded as weighted regressions. We extend these regression models to handle different exposure bases and modeling of trends in the incremental data, and we develop a variety of diagnostic tools for testing the assumptions of these models.

This "extended link ratio family" (ELRF) of regression models is used to test the assumptions made by standard link ratio techniques, and compare their predictive power with modeling trends in the incremental data. Most loss arrays don't satisfy the assumptions of standard link ratio techniques. The ELRF modeling structure creates a bridge to a statistical modeling framework where the assumptions are more consistent with actual data. There is a paradigm shift from standard link ratio techniques to the statistical modeling framework—the ELRF models form a bridge from the "old" paradigm to the "new."

There are three critical stages involved in arriving at a reserve figure: extracting information from the data *in terms of trends and stability, and distributions about these trends; formulating assumptions about the future leading to forecasting of distributions of paid losses; and consideration of the correlations between lines and their effect on the desired security level.*

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 much use, or to 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 [2], Murphy [8], and Mack [6], [7] are described in Section 2, and are extended to include trends in both 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* modeling 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 modeling framework discussed in Section 3. The rich family of statistical models in that framework contains assumptions more in keeping with reality.

This statistical modeling framework is based on the analysis of the logarithms of the incremental data. 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, while 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 modeling 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. Because it is statistical, the modeling framework allows separation of parameter uncertainty and process variability.

It also allows us to:

- check that all the assumptions contained in the model are satisfied by the data;
- calculate distributions of reserve forecasts, including the total reserve;
- calculate distributions of, and correlations between, future payment streams;
- price future underwriting years, including aggregate deductibles and excess layers;
- 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 allows 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 the pricing of different layers.

2. EXTENDED LINK RATIO FAMILY

2.1. Introduction

Brosius [2] points out that the use of regression in loss reserving is not new, dating back to at least the 1950s, and says that using link ratio techniques corresponds to fitting a regression line without an intercept term. Mack [6] 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 [8] derives results for models without an intercept (such as the chain ladder ratios), as well as models with an intercept.

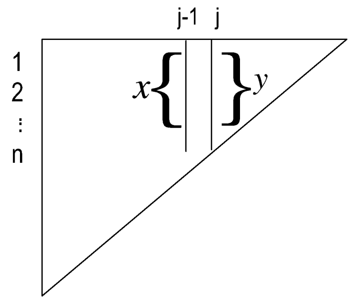
Under the assumption of heteroscedastic (i.e., with non-constant variance) normality, we derive results for a more general family of models (ELRF) that also include accident-year trends for each development year. We discuss calculations and diagnostics for fitting and choosing between models and checking assumptions. Standard errors of forecasts for both cumulatives and incrementals are also derived.

In the current section, we analyze a number of real loss development arrays. Diagnostics, including graphs of the data and formal statistical testing, 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

FIGURE 1

CUMULATIVE LOSS ARRAY



predictive power, so that the standard errors of forecasts become meaningless. Hence, we relegate the calculation of standard errors to the appendices to this paper.

2.2. *Calculating Ratios Using Regressions*

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

A graph of y versus x may appear as in Figure 2.

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

$$y(i) = bx(i) + \varepsilon(i), \tag{2.1.a}$$

where

$$\mathrm{Var}[\varepsilon(i)] = \sigma^2 x(i)^\delta. \tag{2.1.b}$$

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 $x(i)^\delta$, where δ is a "weighting" parameter. The term

FIGURE 2

CUMULATIVE LOSSES VERSUS PREVIOUS DEVELOPMENT YEAR

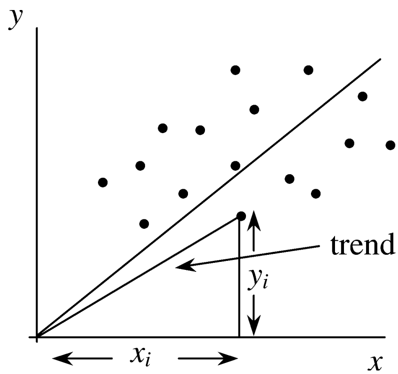
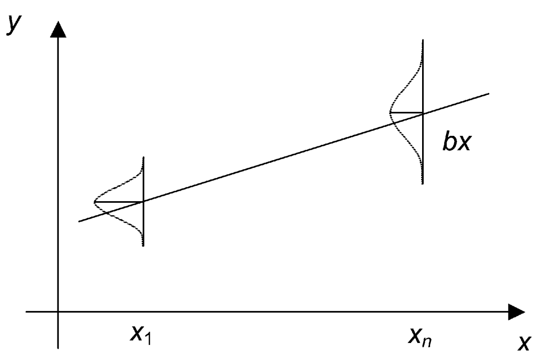


FIGURE 3

CHAIN LADDER RATIO REGRESSION



σ^2 represents an underlying level of variance (or base variance) common to the whole development period.

In Figure 3, $\text{Var}[\varepsilon(i)] = \sigma^2 x(i)^\delta$, where $\delta = 1$. It turns out that 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 1 $\delta = 1$. 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 (by volume) average ratio (i.e., the chain ladder average method, or chain ladder ratio).

CASE 2 $\delta = 2$. 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.

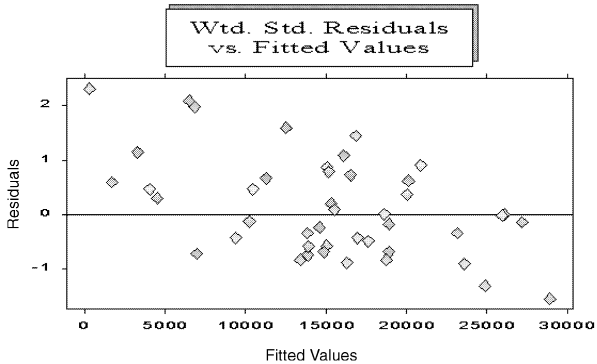
CASE 3 $\delta = 0$. This yields a weighted average (weighted by volume squared) corresponding to ordinary least squares regression through the origin.

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 the standard errors of the parameters in the average method selection and the standard errors of the forecasts can be obtained. A more important advantage is that the assumptions made by the method can be tested.

One important assumption is that the standardized errors, $\varepsilon(i)/\sigma x(i)^{\delta/2}$, $i = 1,2,\ldots,n$, are normally distributed with mean 0 and standard deviation 1. Otherwise, the weighted least squares estimator of b is not necessarily efficient; and the reserve forecasts consequently may be poor estimates of the mean—they will have a large variance. The normality assumption can be checked by examining a number of diagnostic displays, including the normal probability plot, box-plot, and histogram of the weighted standardized residuals. The Shapiro–Francia test [10], based on the normality plot, is a formal test for normality of the residuals.

FIGURE 4



The link ratio method also carries with it other assumptions that should always be tested.

Another basic assumption is that

$$E(y(i) \mid x(i)) = bx(i). \tag{2.4}$$

That is, in order to obtain the mean cumulative at development period j, take the cumulative at the previous development period, $j-1$, and 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 this shows that a (non-zero) intercept is also required. (See Figure 6.)

Equation 2.4 can be re-cast

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

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 underlying Equation 2.4 is valid, then the weighted standardized residuals versus fitted values should appear random. Instead, what you will usually see is a downward trend like that depicted in Figure 4, representing the residuals from the chain ladder ratios model for the Mack [7] data. (See Example 1 below.)

This indicates that large values are overpredicted and small values are underpredicted, so that $E(y \mid x) = bx$ is *not* true.

Comparison of graphs of weighted standardized residuals with graphs of the data will indicate that accident periods that have "high" cumulatives are overfitted and those with "low" cumulatives are underfitted. Figure 5 shows the two displays for the Mack [7] 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.

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 accident period trends).

Murphy [8] suggested an extension of the regression model represented by Equation 2.1 to include the possibility of an intercept:

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

such that

$$\mathrm{Var}[\varepsilon(i)] = \sigma^2 x(i)^\delta. \tag{2.6b}$$

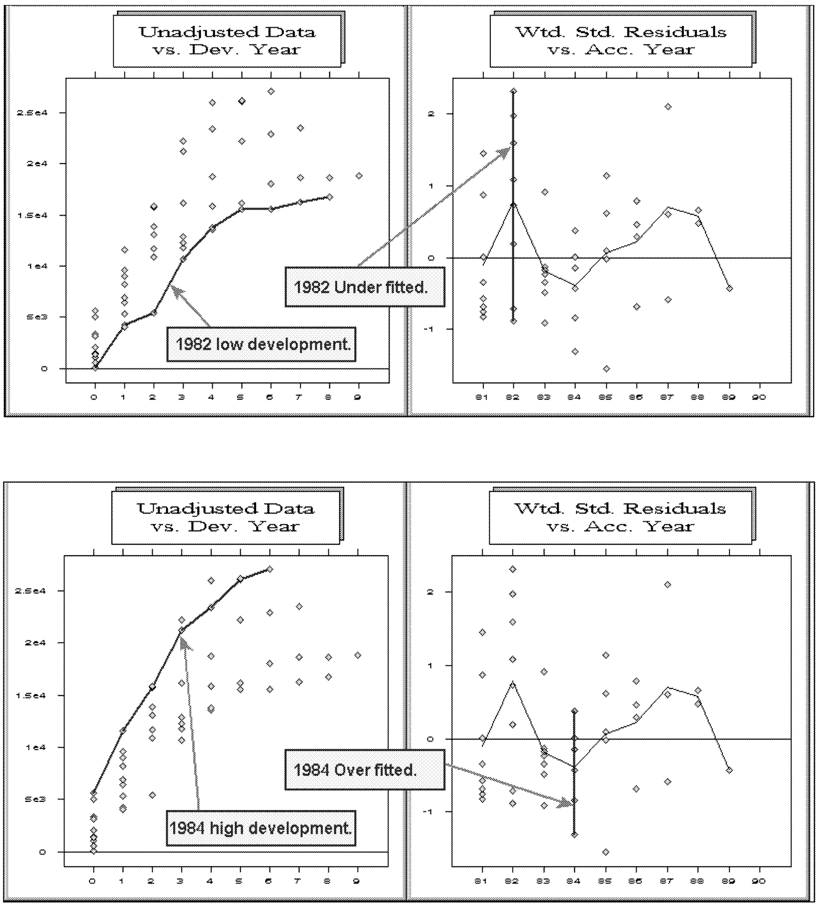
If the intercept a is non-zero and we do not include it in the regression model, then the estimate of the link ratio b (slope) is biased. Note that in the graph in Figure 6 of cumulative values at development period 1 versus cumulative values at development period 0, the intercept appears to be different from zero (the origin sits well below the graph). Indeed, it is significant between every pair of contiguous development periods. See the data of Example 1 below. We can rewrite Equation 2.6 thus:

$$y(i) - x(i) = a + (b-1)x(i) + \varepsilon(i). \tag{2.7}$$

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

FIGURE 5

RAW DATA AND RESIDUALS FROM CHAIN LADDER MODEL



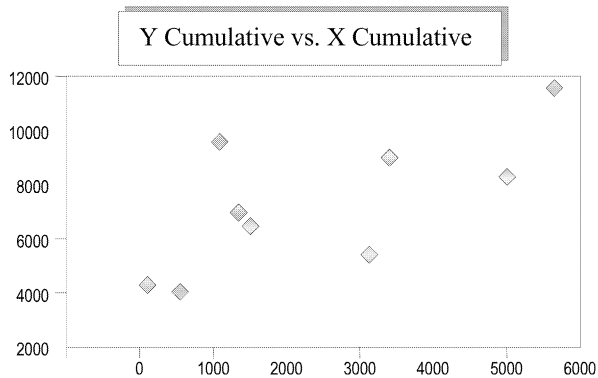
Consider the following two situations:

- $b > 1$ 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 $(b-1)$.

FIGURE 6

CUMULATIVE IN DEVELOPMENT PERIOD 1 VERSUS CUMULATIVE IN DEVELOPMENT PERIOD 0



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

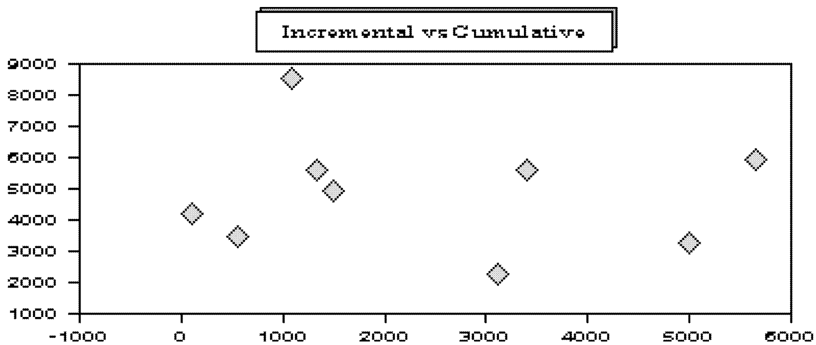
  In this case, $x(i)$ has no predictive power in forecasting $y(i) - x(i)$. The estimate of a is a weighted average of the incrementals in development period j. We would therefore forecast the next accident period's incremental by averaging the incrementals down a development period. Accordingly, the standard link ratio approach is abandoned in favor 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 in Figure 7, which represents the incrementals versus previous cumulatives (development period 0) for the Mack [7] data. It is clear that the correlation is essentially zero. This is also true for every pair of contiguous development periods.

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

FIGURE 7

INCREMENTALS IN DEVELOPMENT PERIOD 1 VERSUS CUMULATIVE IN DEVELOPMENT PERIOD 0



correlation between the incrementals and the previous cumulatives.

Now, if the incrementals possess a trend down the accident periods, it is likely that the cumulatives in the previous development period also trend down the accident periods. In this case, the estimate of the parameter b in Equation 2.7 will be significant; and so the link ratio b, together with the intercept a, will have some predictive power. In this circumstance, we should incorporate an accident period trend parameter for the incremental data; that is,

$$y(i) - x(i) = a_0 + a_1 i + (b-1)x(i) + \varepsilon(i), \qquad (2.8a)$$

where

$$\mathrm{Var}[\varepsilon(i)] = \sigma^2 x(i)^{\delta}. \qquad (2.8b)$$

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 (i.e., $b-1$). Indeed, more often than not, $b-1$ will be insignificant, if the trend parameter a_1 is included in the equation. That is, more often than not, 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:

$$a_0 = \text{intercept};$$
$$a_1 = \text{trend};$$
$$b = \text{ratio (slope)}.$$

Here are some models included in the ELRF described by Equation 2.8.

- Chain Ladder Link Ratios

  In this model, $a_0 = a_1 = 0$ and $\delta = 1$.

- Cape Cod—intercept only

  Here it is assumed that $b = 1$ and $a_1 = 0$. The Cape Cod estimates a weighted average (with weights depending on δ) of the incrementals in each development period. The forecasts are also based on a weighted average down the accident periods for each development period.

  The model can be written as:

$$y(i) - x(i) = a_0 + \varepsilon(i), \tag{2.9a}$$

  where

$$\mathrm{Var}[\varepsilon(i)] = \sigma^2 x(i)^\delta. \tag{2.9b}$$

- Trend with $b = 1$

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

2.3. *Example 1: The Mack Data*

The data for the first example is from Mack [7] (see Table 2.1). The data are incurred losses for automatic facultative business

TABLE 2.1

INCURRED LOSS ARRAY FOR THE MACK DATA†

Accident Year	Development Year 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									

†Note that 1982 accident year values are low.

in general liability, taken from the Reinsurance Association of America's Historical Loss Development Study [9].

We first fit the chain ladder ratios regression model; that is, we fit Equation 2.1 with $\delta = 1$ for every pair of contiguous development periods. The standardized residuals are displayed in Figure 8. Note that the equivalence of Equations 2.5 and 2.6 means that the residuals of the model for the cumulative data are identical to the residuals for the model of the incremental data.

We have already observed the downward trend in the fitted values (Figure 4), and that the high cumulatives are overpredicted whereas the low cumulatives are underpredicted. This is mainly due to the fact that intercepts are required.

So we now fit a model of the type given in Equation 2.6 to each year (i.e., with intercepts, except for the last two pairs of contiguous development periods, as there is insufficient data there). (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

TABLE 2.2

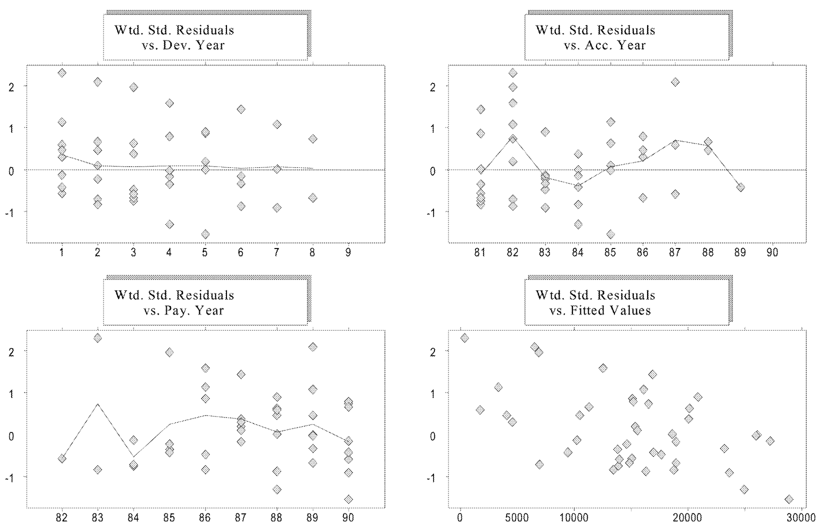
FIT OF THE MODEL WITH INTERCEPT† AND RATIO, WITH $\delta = 1$

Develop. Period	Intercept Estimate	Intercept Std. Error	Intercept p value	Slope Estimate	Slope−1 Estimate	Slope 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	1.01250	0.01250	0.12833	0.931
06–07	792	149	0.118	0.99110	−0.00890	0.00803	0.467
07–08	—	—	—	1.01694	0.01694	0.01506	0.463
08–09	—	—	—	1.00922	0.00922	—	—

† Due to lack of observations in the tail, there is no intercept fitted for the last two years.

FIGURE 8

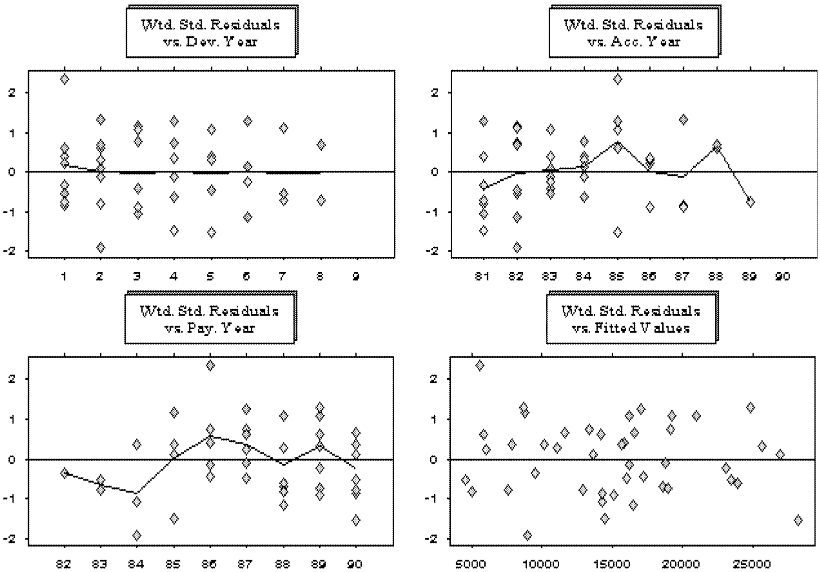
RESIDUAL PLOT FOR THE CHAIN LADDER RATIOS MODEL††



†† Note that the lines join the means at each period.

FIGURE 9

RESIDUAL PLOT FOR MODEL WITH INTERCEPTS FITTED, ALL SLOPES SET TO 1 AND $\delta = 1$



significant. This means that the previous cumulative is not really of much help in predicting the next incremental incurred loss.

The model is overparameterized (i.e., has many unnecessary parameters), 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. This results in the model:

$$y(i) - x(i) = a_0 + \varepsilon(i) \qquad (2.10)$$

The residual plots for the reduced model (Cape Cod) are given in Figure 9.

Note that residuals versus fitted values are “straight” now and that we do not have the high-low effect in the plot of residu-

TABLE 2.3

COMPARISON OF CAPE COD COEFFICIENTS OF VARIATION WITH THOSE FOR THE CHAIN LADDER

	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	1,633	702	0.429884
1985	1,941	1,218	0.512170	2,779	1,404	0.505218
1986	4,200	1,555	0.408791	3,671	1,976	0.538273
1987	6,878	1,677	0.271393	5,455	2,190	0.401467
1988	10,252	3,247	0.308234	10,934	5,351	0.489391
1989	14,874	3,657	0.253810	10,668	6,335	0.593832
1990	19,336	4,532	0.215021	16,360	24,606	1.504034
Total	59,248	8,494	0.110347	52,272	26,883	0.514291

als versus accident period. The plot of residuals versus accident year does not exhibit a trend; if we were to include a trend, by estimating

$$y(i) - x(i) = a_0 + a_1 i + \varepsilon, \qquad (2.11)$$

we would find that the estimate of a_1 would not be significantly different from zero.

We now present forecasts and coefficients of variation (mean divided by standard deviation) of forecasts based on the Cape Cod (intercept-only) model with $\delta = 1$, and compare this with the forecasts and coefficients of variation for the chain ladder ratios (see Table 2.3).

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 randomly drawn from the same distribution. This does not happen with the chain ladder ratios, because the model does *not* 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. Moreover, the coefficient of variation for 1990 is 150%, but for the previous year it is 59%. Note that 1990 has one more incremental value to forecast than 1989; if anything, a good model will on average have smaller coefficients of variation for totals of years with more observations. Since the 1990 accident-year total pools one more uncertain value than 1989 and the remaining values (conditionally on the first) could be expected to have similar coefficients of variation to the corresponding values from 1989, this appears to violate the fundamental statistical principle of insurance—risk reduction by pooling.

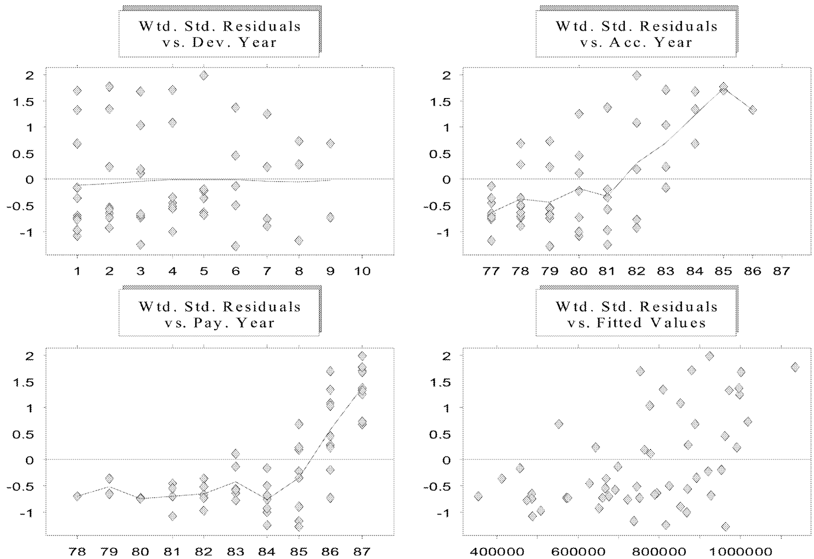
For the Mack data, the model with intercepts is reasonable, as there is no accident-year trend in the incrementals. For data where a constant trend (on a dollar scale) does exist, then the trend will be significant, but very often the ratio will not be significantly different from one.

2.4. Summary

We have so far considered two cases that can occur in real data: incrementals for a particular development period have no trend, and incrementals have trend in the accident period direction (after possibly adjusting the data by accident period exposures). In the first case, link ratios are often insignificant and so lack predictive power. In the second case, when incrementals versus accident periods for a particular development period have a constant trend, it is likely that the cumulatives in the preceding development period also exhibit a trend so that the ratio has some predictive power (equivalently, the ratio is significantly different from one). However, more often than not, the accident period trend has more predictive power than the ratio; and, once

FIGURE 10

RESIDUAL PLOTS FOR THE CHAIN LADDER RATIOS MODEL

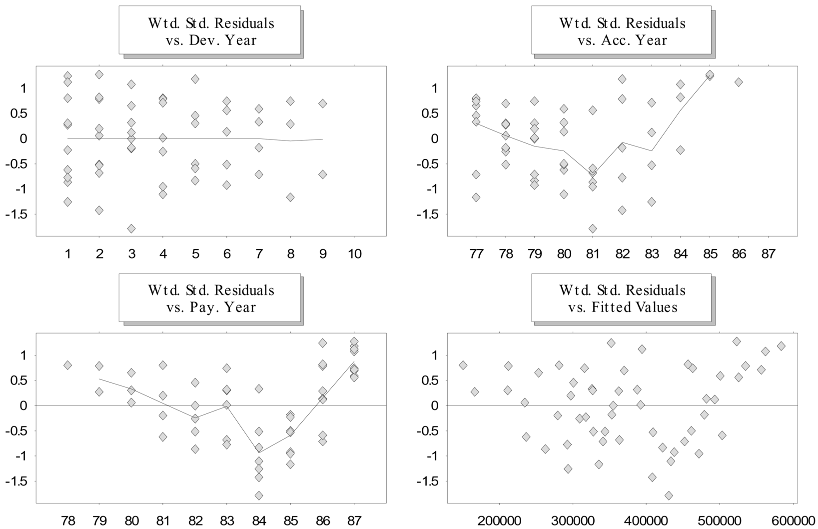


it is included in the model, the term $(b - 1)$ is often insignificant (i.e., the ratio does not have any residual predictive power). The situation encountered most often in practice, however, involves a trend change along the payment/calendar periods (diagonals). This means that as you look down each 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 standardized residual plots depicted in Figures 10 and 11 are those of the chain ladder ratios and Equation 2.8, respectively, applied to project ABC (Workers Compensation Portfolio) discussed in Section 3. Note that the chain ladder ratios indicate a payment-year trend change, and the model in Equation 2.8, which fits a constant trend down the accident years for each development year, indicates that the trend before pay-

FIGURE 11

RESIDUAL PLOTS FOR TREND PLUS RATIO MODEL



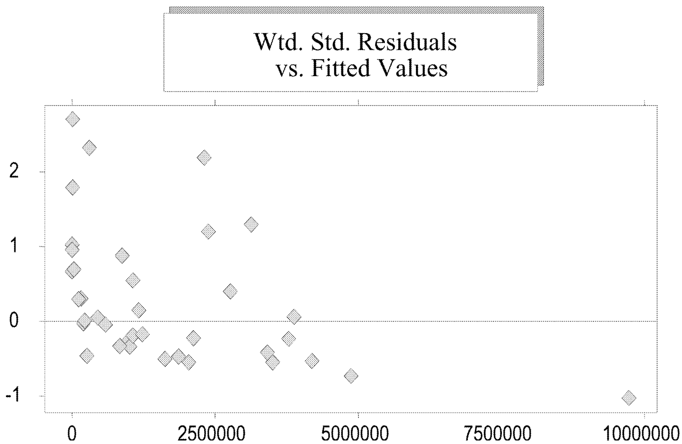
ment year 1984 is lower than the trend after 1984. This project (ABC) is analyzed in more detail in Section 3.

The models subsumed by Equation 2.8 can be used to diagnostically identify payment period trend changes but do not identify these trend changes or forecast with them. The models in the ELRF form a bridge to models that also include payment period trend parameters; that is, statistical models in the Probabilistic Trend Family (PTF) of the next section.

It is important to note that ELRF models also make the implicit assumption that the weighted standardized 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 standardized residuals are generally skewed to the right, suggesting that the analysis should be conducted on the

FIGURE 12

RESIDUALS VERSUS FITTED VALUES FOR THE CHAIN LADDER RATIOS



logarithmic scale. The graph in Figure 12 illustrates the skewness of a set of weighted standardized residuals based on chain ladder ratios for Project PAN6 (analyzed in detail in Example 4 of Section 3). The positive-weighted standardized 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 ELRF regression methodology you will discover that, for any type of real loss development array, the standard development factor (link ratio) techniques are frequently inappropriate. Analyzing 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 may be quite meaningless, unless the assumptions made by the model are supported by the data.

3. STATISTICAL MODELING 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 instead 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 modeling 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 a great deal of easily communicated information from the loss development array.

The details of the modeling framework and its inherent benefits are described in Zehnwirth [12]. However, given that there is a paradigm shift from the standard link ratio methodology to the statistical modeling framework, we review the salient features of the statistical modeling framework.

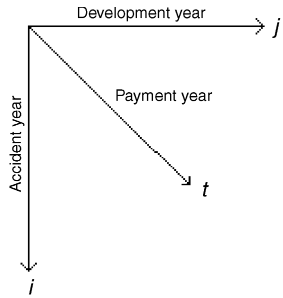
3.2. Trend Properties of Loss Development Arrays

Since a model is supposed to capture the trends in the data, it behooves 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, $j = 0,1,2,\ldots,s-1$; accident years by i, $i = 1,2,\ldots,s$; and payment years by t, $t = 1,2,\ldots,s$. See Figure 13.

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.

FIGURE 13



The two directions, development-year and accident-year, are orthogonal. That is, trends in either direction are not projected onto the other. The payment-year direction t 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 (by a price or wage index) 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 at all. 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.

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. The mathematical formulation of the models in the statistical modeling framework is given by Equation 3.5. We now illustrate the geometry of trends with a simulation example.

3.3. 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

$$y(i,j) = \ln(p_{ij}) = 11.51293 - 0.2j. \qquad (3.1)$$

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. Note that a linear trend of 0.1 per year on the log scale is about a 10% *per annum* increase on the original scale.

The trends are depicted in Figure 14. Patterns of change like this are quite common in real data. Note that trends in the payment/calendar year direction project onto the other two directions, as they must. The resultant trends for the first six accident years are shown in Figure 15.

Note that each line in the graph is the resultant development-year trend for a single accident year. As you go down the acci-

FIGURE 14

DIAGRAM OF THE TRENDS ON THE LOG SCALE IN THE DATA ARRAY

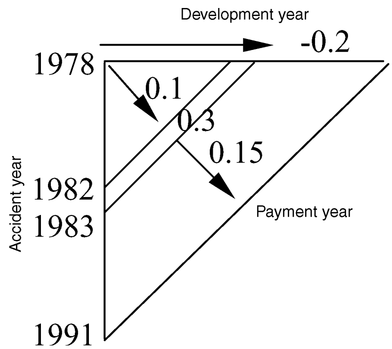


FIGURE 15

PLOT OF THE LOG(PAID) DATA AGAINST DELAY FOR THE FIRST SIX ACCIDENT YEARS

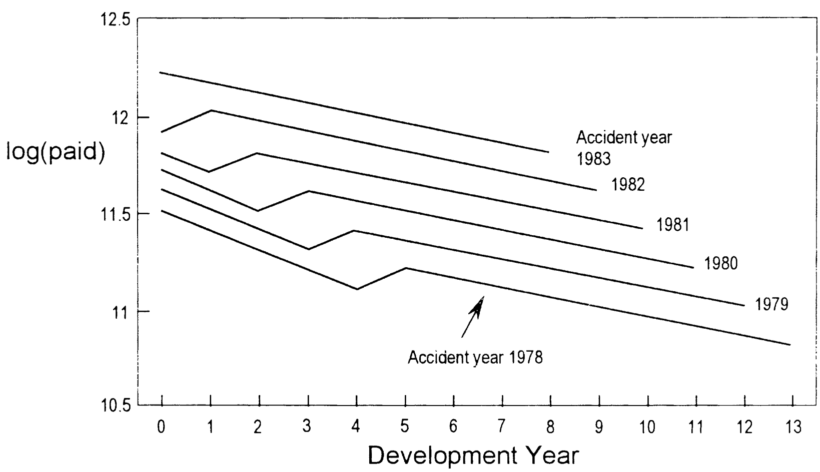
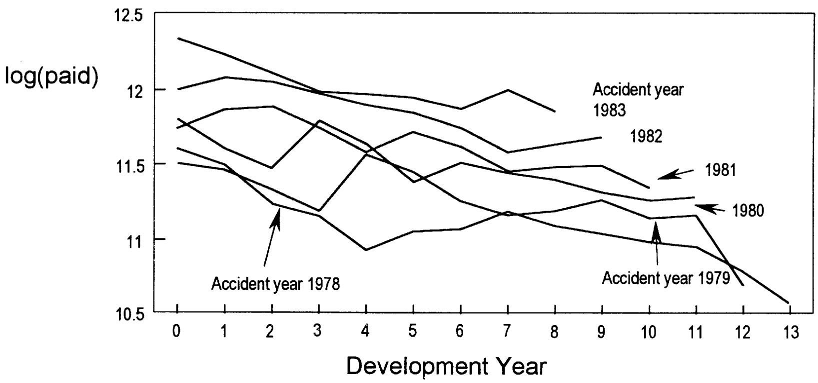


FIGURE 16

TREND PLUS RANDOMNESS FOR THE FIRST SIX ACCIDENT YEARS



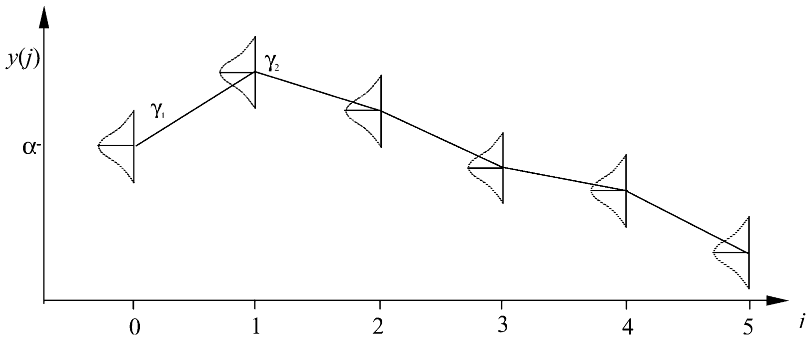
dent years (1978 to 1983), the 30% trend always kicks in one development-period earlier. The payment-year trends also project onto the accident years, which is why the early years are at the bottom and the later years are at the top. Note how the "kink" moves back as we go up to the more recent accident 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 other ELRF model.

Of course, real data are never so smooth. On the same log scale, we add some noise—random numbers with mean 0 and standard deviation 0.1, as shown in Figure 16.

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 presented there, the regression models in ELRF from Section 2 are inadequate for this data. We instead look to a model

FIGURE 17

PROBABILISTIC MODEL FOR DEVELOPMENT-YEAR TRENDS (LOG SCALE)



that incorporates the trends in the three directions and the variability about those trends, measured on a log scale.

Consider a single accident year (dropping the i subscript for the moment). We represent the expected level in the first development year by the 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, represented as:

$$y(j) = \alpha + \sum_{k=1}^{j} \gamma_k + \varepsilon_j. \tag{3.2}$$

This probabilistic model is depicted in Figure 17 (for the first six development years).

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, and so on.

Based on the model in Equation 3.2, the random variable $p(j)$ has a lognormal distribution with

$$\text{median} = \exp\left[\alpha + \sum_{k=1}^{j} \gamma_k\right], \tag{3.3}$$

$$\text{mean} = \text{median} \times \exp[\tfrac{1}{2}\sigma^2], \tag{3.4}$$

and

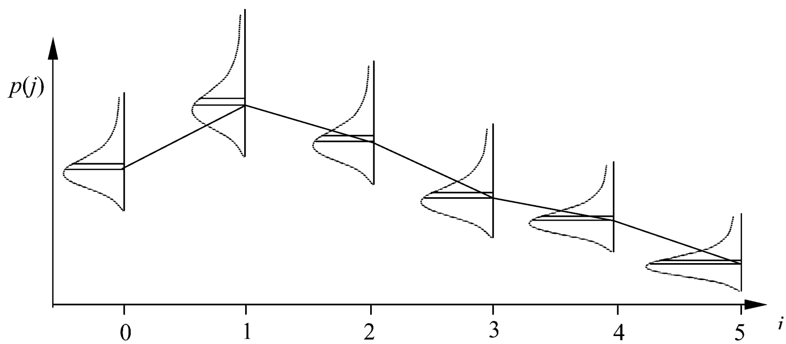
$$\text{standard deviation} = \text{mean} \times \sqrt{\exp[\sigma^2] - 1}. \tag{3.5}$$

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 3.3 and the means are related by Equation 3.4. So, in 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.

Note from Equation 3.3 that exponentiating the mean on the log scale gives the median on the dollar scale. This is why the line in Figure 17, after exponentiation in Figure 18, joins the medians (the lower of the two lines on each density) not the means. 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.

FIGURE 18

MODEL FOR TRENDS ALONG A DEVELOPMENT YEAR (DOLLAR SCALE)



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:

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

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 $\alpha_{i+1} - \alpha_i + \iota_{i+j}$.

TABLE 3.1

PARAMETER ESTIMATES FOR THE MODEL WITH CONSTANT TRENDS

Parameter	Estimate	Standard Error	t-ratio
α	11.4256	0.0302	378.57
γ	−0.2062	0.0037	−55.08
ι	0.1563	0.0037	41.74
$s = 0.1129$		$R^2 = 97.0\%$	

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 particular model.

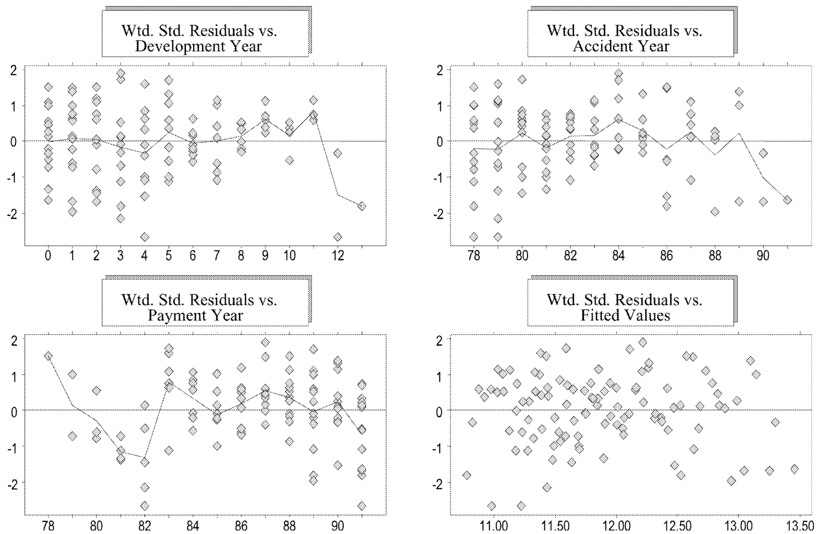
If the error terms $\varepsilon_{i,j}$ (each coming from a normal distribution with mean 0) do not have a constant variance, then the changes in variance must also be modeled. Note that there are numerous models in the PTF, even if we do not include the varying (stochastic) parameter models discussed in Section 3.7. The actuary has to identify the most appropriate model for the loss development array being analyzed. The assumptions made by the "optimal" model must be satisfied by the data. In doing so, one extracts information in terms of trends, stability of trends, and the distributions of the data about the trends.

3.3. 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 to each other (one γ), all payment-year trends equal to each other (one ι), and 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 3.1.

FIGURE 19

PLOTS OF STANDARDIZED RESIDUALS VERSUS THE THREE DIRECTIONS AND FITTED VALUES FOR THE SINGLE PAYMENT-YEAR TREND MODEL



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

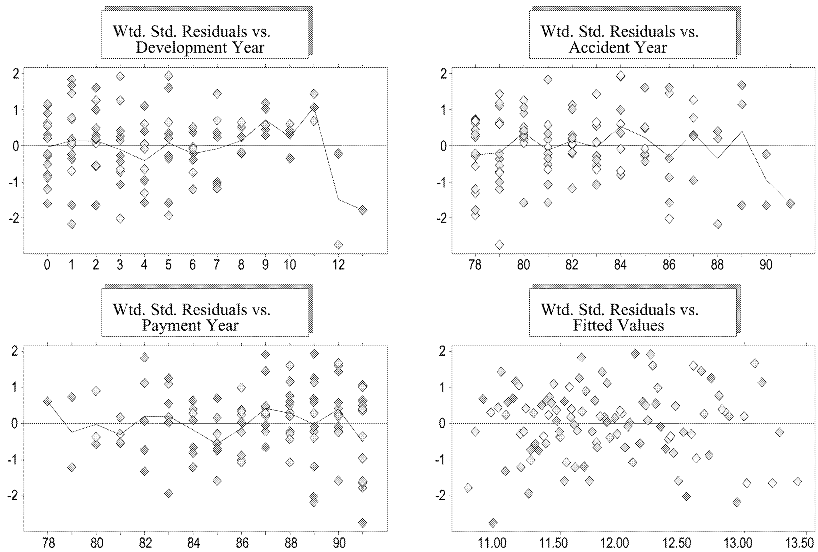
Removing constant trends makes any changes in trend more obvious; the residuals are shown in Figure 19.

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 value of 0.1563 per year.

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

FIGURE 20

PLOTS OF STANDARDIZED RESIDUALS VERSUS THE THREE DIRECTIONS AND FITTED VALUES FOR THE THREE-PAYMENT-YEAR-TRENDS MODEL



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) parameter between 1982–1983, and a further ι parameter between 1983–1984 (that will continue to 1991).

The residuals of the model with three payment-year trends are given in Figure 20—this model seems to have captured the trends. The parameter estimates are given in Table 3.2.

Note that the estimates of the trend parameters 0.1, 0.3, 0.15 are not equal to the true values; indeed, 0.3927 (standard error 0.0442) is a bit off the mark (which is about two standard errors). The estimate is far from 0.3 because in the payment years 1982 and 1983, there aren't many data points. Given that the

TABLE 3.2

PARAMETER ESTIMATES FOR THE THREE-PAYMENT-YEAR-TRENDS MODEL

Parameter	Estimate	Standard 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
ι: 82–83	0.3927	0.0442	8.90
ι: 83–91	0.1446	0.0046	31.72
$s = 0.1005$		$R^2 = 97.7\%$	

TABLE 3.3

FORECASTS, STANDARD ERRORS, TREND ESTIMATES AND THEIR STANDARD ERRORS AS THE LATER PAYMENT YEARS ARE REMOVED

Years in Estimation	N	γ (83–91)	Standard Error (γ)	ι (83–91)	Standard Error (ι)	Mean Forecast	Standard Error of Forecast
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

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 3.3.

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

FIGURE 21

PREDICTION ERRORS FOR 1988–1991, FOR MODEL ESTIMATED IN 1987

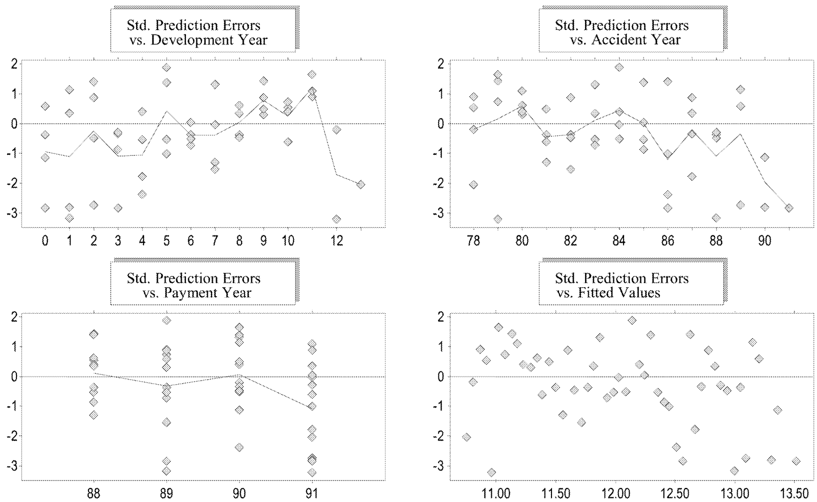


Figure 21 gives the prediction errors (on a log scale) for the four payment years 1988–1991, based on the model estimated at year end 1987.

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 (see Table 3.4) is now $15.63\% \pm 1.03\%$ (where we are writing mean $\pm$ standard deviation as shorthand), in place of $14.46\% \pm 0.46\%$ when we use all the years in the estimation. Hence the forecast of \$26M ($\pm$\$2.9M) is "higher" than \$23M ($\pm$\$0.9M). When you test for a trend change between 1987 and 1988, it is not significant (as we would expect). Note that removal of recent payment years to check the model's ability to predict them (validation analysis) is part of the model identification procedure and extraction of information process.

TABLE 3.4

VALIDATION RESULTS—PARAMETER ESTIMATES AND FORECASTS AS PAYMENT YEARS ARE REMOVED FROM THE SELECTED MODEL

Payment Years in Estimation	Estimate of gamma	Estimate of iota (83–91)	Forecast ± Standard Error ($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

3.4. Example 3—Real Data With Major Payment-Year Trend Instability

We now analyze a real data array as presented in Table 3.5.

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. When there is a trend change, formulation of the assumptions about the future trend will depend on the explanation for that trend change.

The individual link ratios for the cumulated data are very stable, as can be seen in Figure 22. 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.

We first conduct some diagnostic PTF analysis, then show how the ELRF modeling structure also indicates payment-year trend change, indeed that any method based on link ratios is quite meaningless for this data. Consequently, there is little to be gained by forecasting any of the ELRF models. Figure 23 shows

TABLE 3.5

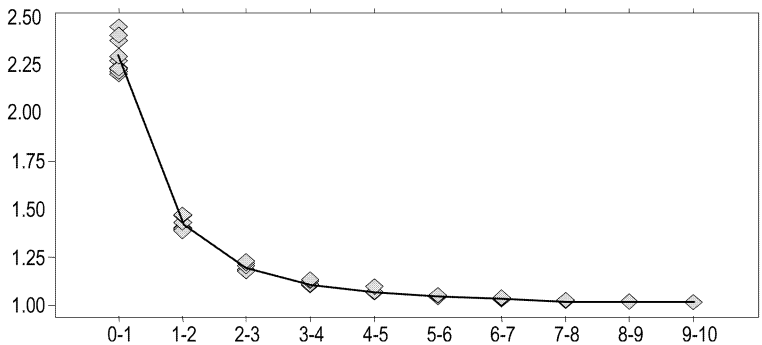
INCREMENTAL PAID LOSSES AND EXPOSURES FOR ABC

	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										

Accident Year	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

FIGURE 22

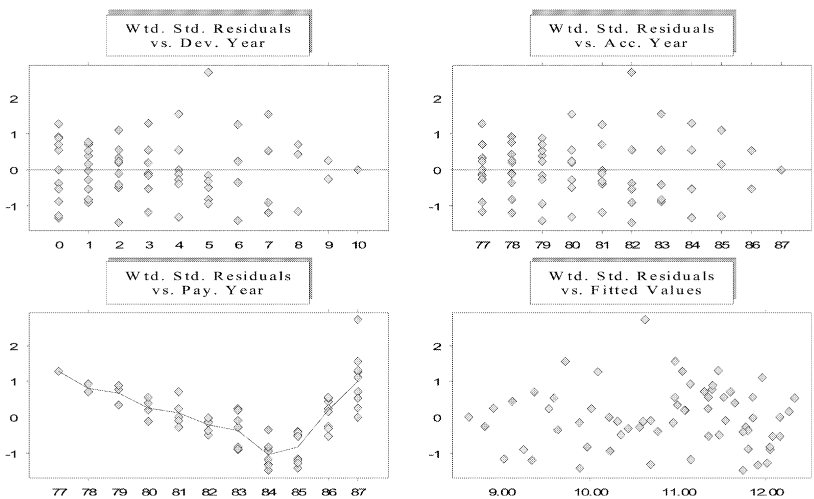
PLOT OF INDIVIDUAL LINK RATIOS BY DELAY†



† The line joins Chain Ladder ratios.

FIGURE 23

STANDARDIZED RESIDUALS OF THE STATISTICAL CHAIN LADDER MODEL



the standardized residuals of the statistical chain ladder in PTF (i.e., the statistical chain ladder fits all the gamma parameters and all the alpha parameters with no iotas). The residuals are just the data (less the average level) adjusted for the (average) trend between every pair of contiguous development periods and every pair of contiguous accident periods. This is why the plots of standardized residuals versus development years and accident years are centered on zero! We use this model only as a diagnostic tool to determine quickly whether there are payment-year trend changes that can be attributed solely to the payment years.

Contrast the smoothness of the ratios above with the plot of the residuals from this 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. There is a difference between Figures 23 and 10—the statistical chain ladder shows the payment-year trends after adjusting for the trends in the other two directions, while the chain ladder ratios (Figure 10) do not do that. But the change in trend is clear in either graph. In the current statistical modeling 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 for this data 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 after 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%—as a change to the level of payments will be a single jump in level (possibly taking several years to be completely manifested). If there is no explanation for the trend change, except that the payments have

TABLE 3.6

PAID LOSS ARRAY FOR THE PAN6 DATA FOR EXAMPLE 4

Accident Year	Development Year 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	****					

increased, then calling the future in terms of trends is more difficult.

3.5. Example 4—Volatile Data With Stable Trends

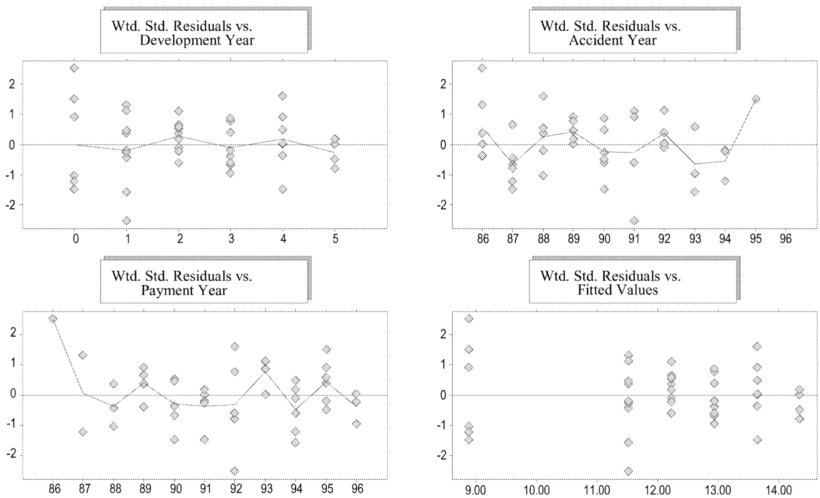
We now consider an array where the paid losses are very volatile, but the trends are stable (see Table 3.6). Recall that trend stability/instability is dependent on neither the volatility of the data nor the volatility of the link ratios. Since the random component is an integral part of the model, this model captures the behavior of this volatile data very well. We call this array PAN6.

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 24).

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

FIGURE 24

PLOT OF STANDARDIZED RESIDUALS FOR THE MODEL WITH TWO GAMMA PARAMETERS AND ONE ALPHA PARAMETER

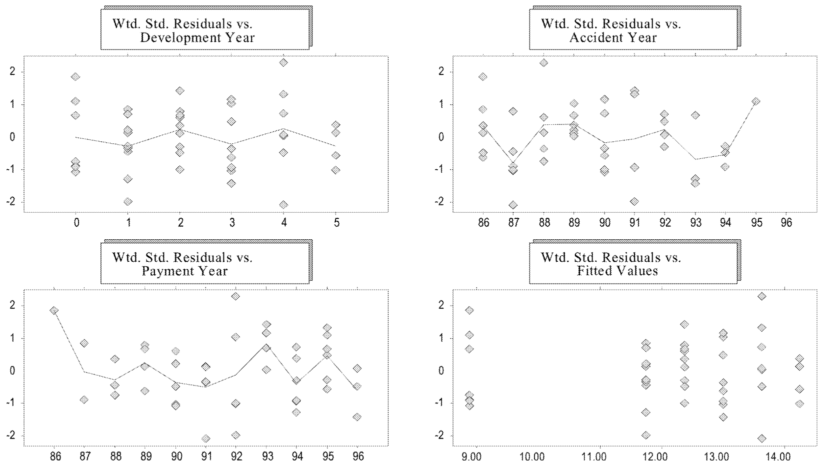


that they are 3.0177 and 0.8015, respectively. This requires a weighted regression; development years 0 and 1 are given weight $(0.8015/3.0177)^2$, and the other years (2+) have a weight of 1. The weighted, standardized residual plots now look fine (see Figure 25). A check of the plot of residuals against 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 $\sqrt{\exp(\sigma^2) - 1} > \sigma$.

FIGURE 25

PLOT OF WEIGHTED STANDARDIZED RESIDUALS AFTER THE WEIGHTED REGRESSION



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 the 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 means.

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.

TABLE 3.7

FORECASTS, FORECAST STANDARD ERRORS, FINAL TREND ESTIMATES AND THEIR STANDARD ERRORS FOR THE FINAL MODEL AS THE LATER PAYMENT YEARS ARE REMOVED

Years in Estimation	N	Trend (dev. period 1+)	Standard Error	Mean Forecast	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

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

Note that at the end of year 1992, the estimated model would have predicted the normal distributions for the log(payments) in years 1993–1996 (see Figure 26), and would have produced statistically the same forecast of outstanding claims. Figure 27 indicates that the assumption of normality is reasonable.

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 28). 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 the 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.

FIGURE 26

PREDICTION ERRORS FOR YEARS 1993–1996

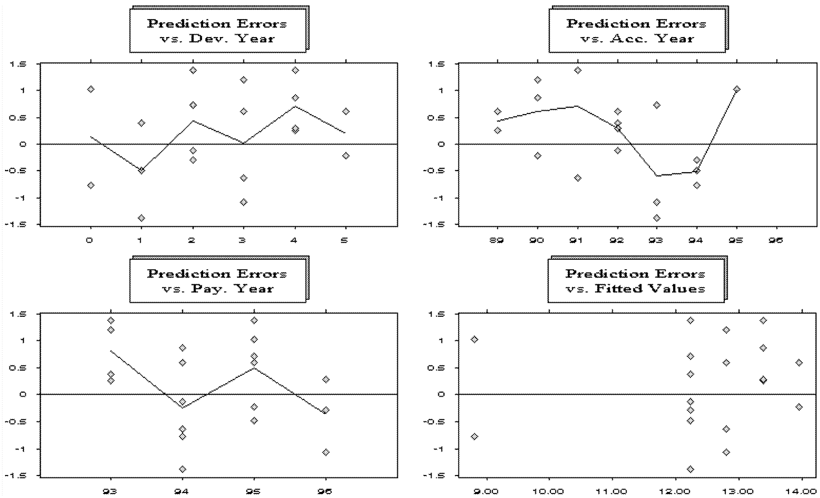


FIGURE 27

NORMALITY PLOT OF PREDICTION ERRORS FOR 1993–96 BASED ON MODEL ESTIMATED AT YEAR END 1992

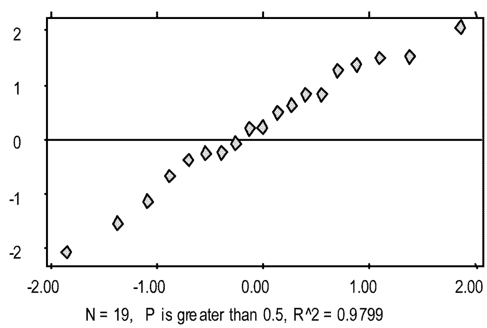
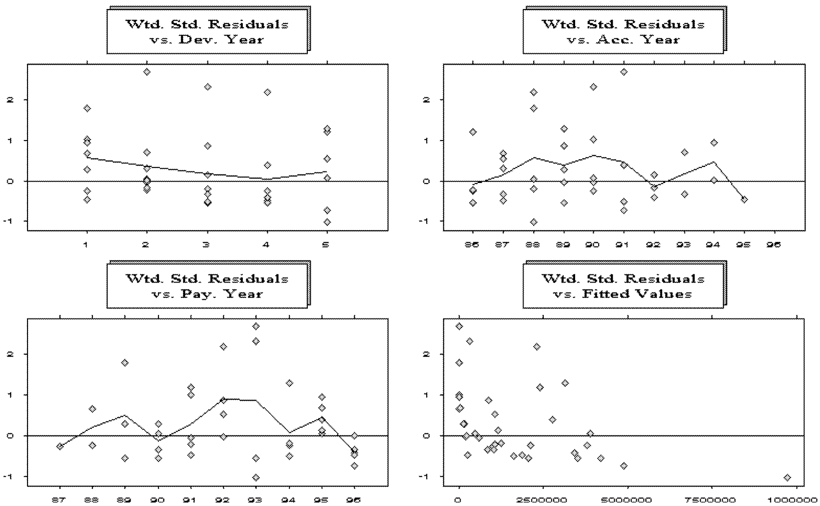


FIGURE 28

RESIDUALS OF CHAIN LADDER RATIOS REGRESSION MODEL



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

3.6. Example 5—Simulated Array Based on a Model With Only Two Parameters

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

$$y(i,j) = 10 - 0.3j + \varepsilon(i,j), \qquad (3.7)$$

where the $\varepsilon(i,j)$ are independent and identically distributed from a normal distribution with mean 0 and variance 0.4.

The simulated data is presented in Table 3.8.

TABLE 3.8

INCREMENTAL PAID LOSS DATA FOR SIMULATED EXAMPLE SDF

Accident	Development Year																
Year	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1978	24,307	44,260	3,900	13,393	17,731	5,802	1,975	6,294	4,758	1,529	2,383	161	760	445	640	501	255
1979	19,122	7,003	8,147	2,872	5,639	2,967	3,455	2,364	4,005	2,029	1,370	718	432	1,282	456	232	
1980	18,082	27,708	27,901	5,699	9,297	2,899	7,461	6,885	1,854	2,681	792	1,338	820	702	179		
1981	80,451	11,411	68,627	6,703	6,430	2,693	5,560	847	4,706	1,245	1,289	717	754	436			
1982	49,099	7,144	11,979	3,481	2,279	5,975	4,472	2,066	509	2,345	755	1,702	596				
1983	33,475	54,717	8,774	4,859	5,808	20,750	1,903	2,449	937	796	1,446	1,271					
1984	23,070	49,554	5,659	9,909	9,123	8,805	7,008	8,634	3,053	493	2,108						
1985	14,324	8,352	7,955	8,092	6,044	8,542	7,700	2,849	1,130	1,754							
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																

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 σ^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 further away from the mean compared to realizations less than the mean, which are bounded below by zero (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, and so the process variance is $30{,}970^2$. When we analyze the data in PTF, we identify only two significant parameters—$\hat{\alpha} = 9.9667$ (with standard error 0.0847) and $\hat{\gamma} = -0.2867$ (standard error 0.0126), and the estimate of σ^2 is 0.4085. Residuals from this estimated model are displayed in Figure 29.

Table 3.9 gives forecasts of total outstanding, including validation forecasts (note that the forecasts are stable, as expected).

We now study the cumulative array, displayed in Figure 30. Even though the incremental data was generated with homogeneous accident years, the cumulated data have each accident year

FIGURE 29

RESIDUALS BASED ON THE ESTIMATED PARAMETERS OF THE MODEL FROM WHICH THE DATA WERE GENERATED

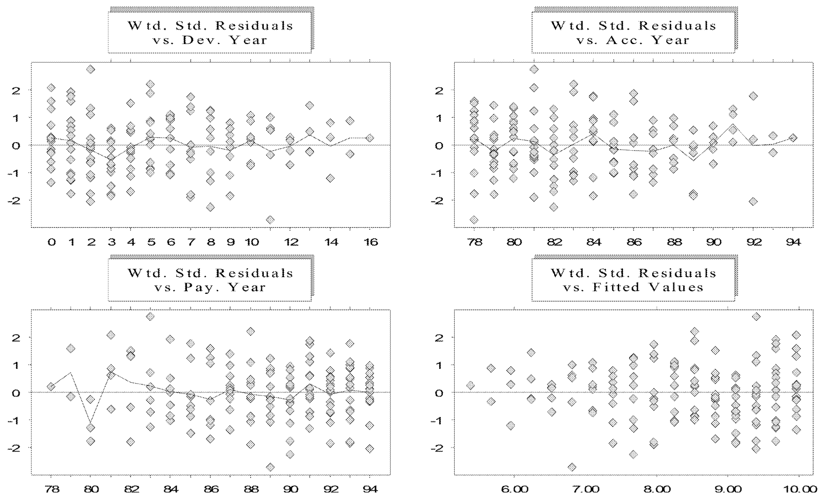


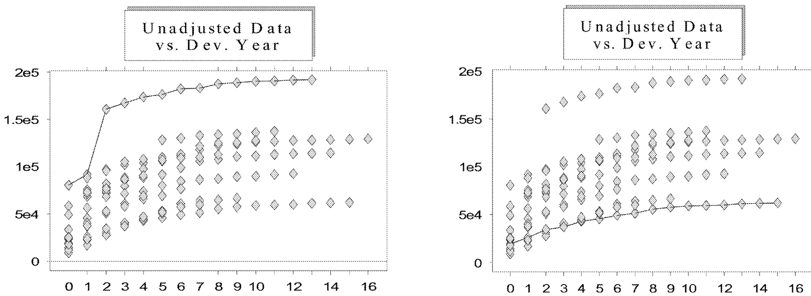
TABLE 3.9

ESTIMATES OF DEVELOPMENT-YEAR TRENDS AND TOTAL OUTSTANDING WITH STANDARD ERRORS AS YEARS ARE REMOVED FROM THE ANALYSIS

Payment Years in Estimation	Estimate of Gamma	Standard Error of Gamma	Mean Forecast	Standard Error of Forecast
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

FIGURE 30

CUMULATIVE DATA (SDF)†



†Accident year 1981 has high development; 1979 has low development.

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.

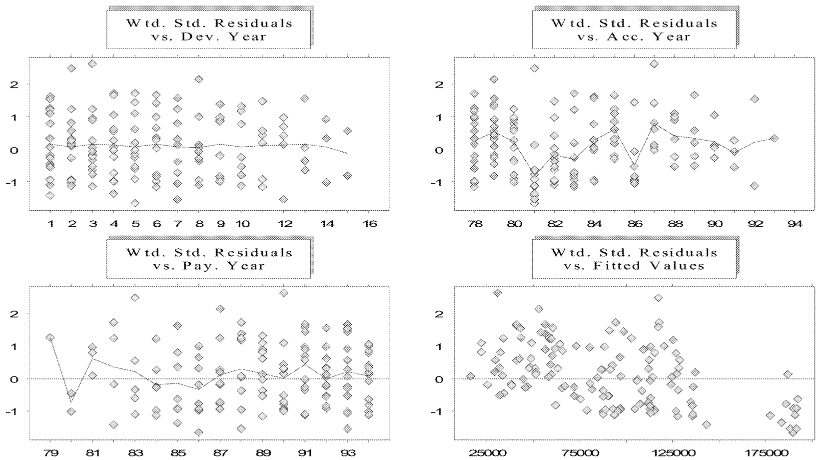
The cumulative values for 1979 lie entirely below those for 1982 (see Figure 30), yet most of the incremental payments are "close" together. One "large" incremental value from the tail of the lognormal has a major impact on the cumulative data. The link ratio techniques assume that the next incremental 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. For this array, we know that current emergence is not a predictor of future emergence.

The chain ladder ratios model gives a mean outstanding forecast of $254,130 and a standard error of the outstanding forecast

FIGURE 31

PLOT OF WEIGHTED STANDARDIZED RESIDUALS FOR CHAIN LADDER RATIOS FOR TRIANGLE SDF



of $59,419. The plot of residuals against fitted values makes it clear where the problem lies, as we see in Figure 31.

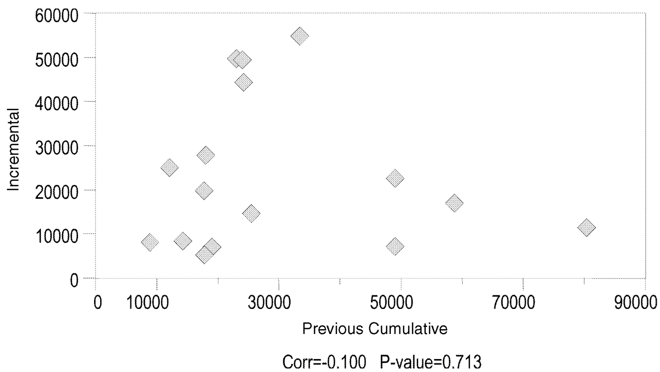
Again, we have a year with high cumulatives (1981) overpredicted, and a year with low cumulatives (1979) underpredicted.

Note how there is a distinct downward trend in the plot of fitted values. 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. Plots of residuals against normal scores show non-normality (not presented here). If we look at the plot of the incremental paid losses against the previous cumulative (see Figure 32), we can see that models involving ratios will be inappropriate since there is no relationship.

The best model in the ELRF sets all of the ratios to 1 and only uses intercepts (i.e., it takes averages of incrementals in

FIGURE 32

PLOT OF INCREMENTAL PAYMENTS AGAINST PREVIOUS CUMULATIVE



each development year). But, due to the non-normality, this is insufficient. At least the ELRF analysis informs us that the incrementals in a development period may be regarded as random values from the same distribution, and that these incrementals are not correlated with the previous cumulatives (just as the data were generated). It also tells us that the data are skewed, 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 modeling the logs of the incremental data—it is telling us the truth.

If you generate (simulate) data using link ratios, the ELRF will tell you that ratios have predictive power and that the data were generated cumulatively. Importantly however, for most real loss development arrays, ELRF analysis indicates that the data were generated incrementally and that ratios have much less predictive power than trends in the log incrementals. The ELRF analysis also shows when there may be payment/calendar year trend changes.

3.7. Varying (Stochastic) Parameters

In view of the trend relationships between the development-year, accident-year, and payment-year directions, a model with several parameters in the payment-year and accident-year directions may suffer from multi-collinearity problems. Zehnwirth [12, 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 variation in α is determined by the SSPE statistic, which is explained in Zehnwirth [12].

3.8. Model Identification

It is important 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 [12]. We start off with a model that only has one parameter in each direction, model (sequentially) the trends in the development-year direction, and follow that by looking at the trends in the payment-year or accident-year directions (depending on which direction exhibits more dramatic trend changes). Adjustments for different variances may also be necessary. Validation analysis is an integral component of model identification, extraction of information, and testing for stability of trends.

3.9. Assumptions About the Future

Stability and assumptions about the future are discussed in Section 9.6.2 and 10.2 of Zehnwirth [12]. 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 ι is $\hat{\iota}$ with standard error $s.e.(\hat{\iota})$, then we assume for the future a mean trend of $\hat{\iota}$, with a standard deviation of trend of $s.e.(\hat{\iota})$; 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 [12] 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 feasible cases, though several cases are discussed in Zehnwirth [12]. 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; but we only know this after performing an analysis like that described for the PTF.

3.10. How Do We Know That Real Data Triangles Can Be Generated By the Members of the PTF?

Let's conduct the following experiment. Begin with 100 data arrays (triangles) and, for each triangle, the "best" model in the PTF is given. Recall that a fitted (best) model relates the distributions of each cell in terms of trends on the log scale. These models are tested to ensure they are good models. ELRF analysis is then conducted on each triangle. Now assume that some of the triangles are real data from some companies, but some are not.

That is, for some triangles the data represent a sample path from the so-called "fitted" distributions. Which are the real data and which are simulated data from the model?

Both the real and simulated data display similar features to each other, whether we analyze them using PTF models or ELRF models. For example, both usually indicate that once you fit an intercept (and accident-year trend, if present), ratios are not needed. Both tend to display trends in the payment years. Since 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 underlying complex generation 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 set of severities. The estimated Pareto did *not* create the severities, but it has probabilistic mechanisms for creating the data as a sample.

Suppose we now simulate (cumulative) loss development arrays from ratio models. The ELRF methodology applied to the simulated arrays will inform us that ratios have predictive power (indeed more predictive power than using the trends in the incrementals). If we now analyze the corresponding incremental arrays using PTF models, we notice a phenomenon that is very rarely observed for real data. After removing trends in the development-year direction, there are distinct patterns in the accident-year direction—the levels jump up and down dramatically, much more so than is typical for real data. This is because if an incremental value in development period zero is low, then the subsequent incrementals for the same accident period remain relatively low, as all accident years are multiplied by the same ratios. Similarly, if the initial incremental value is high, the subsequent incrementals from the model are high. This pattern occurs even when you include considerable randomness.

Finally, suppose an actuary is presented with many development arrays, some real, and the others simulated using models based on ratio-techniques. By applying the ELRF and PTF methodologies, the actuary is very frequently able to distinguish the real arrays from the ones simulated from the ratio models.

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 effectively 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, which include both process risk and parameter risk. Forecasting of distributions is discussed in Zehnwirth [12].

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.

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 [4] for more details).

Under the model, the distribution of a sum of payments (e.g., accident-year outstanding payments) is the distribution of a sum of correlated lognormal variables. The lognormal variables are correlated because of the correlations between the estimated parameters describing their mean level—indeed many forecasts will even share some parameters. The 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.

An insurer's risk can be defined in many different ways. One common definition is related to the standard deviation of the risk, in particular a multiple of the standard deviation. If the reserve is based on a given percentile of the distribution of the total outstanding, the size of the loading as a multiple of the standard deviation will be dependent on the skewness of the distribution.

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, no matter which allocation principle is used. This is always true, even if there is some dependence (and so correlation) between the various lines. In the following example, the standard deviation principle is used.

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

$$s.e.(\text{Total}) = [s.e.^2(1)+\cdots+s.e.^2(n)]^{0.5}. \tag{4.1}$$

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 an allocation of the risk margin for line j is

$$k \times s.e.(\text{Total}) \times s.e.(j)/[s.e.(1)+\cdots+s.e.(n)] < k \times s.e.(j). \tag{4.2}$$

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

If as a result of analyzing each line using the statistical modeling framework, we find that, for some lines, trends change in the same years and the changes are of similar size, then the lines are not independent. (There may also be correlations between the residuals, but that is generally only important when forecasting sums of lines.)

In that situation, if line i and j are correlated, say, then one could 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 the case that different lines are much 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 known as Jensen's inequality (this concept is in many college finance texts—see for example Brearley & Myers [1]). 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 the following:

- 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 usually 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 the Reserve

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

Step 1

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

Step 2

Assumptions about the future are formulated. If payment/calendar year trend is stable, this is straightforward. If trends are unstable in more recent years, 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 [12], but it is impossible to give an exhaustive list as each case may be different.

Step 3

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

4.4. Other Benefits of the Statistical Paradigm

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

- Credibility models

  If a particular trend parameter estimate for an individual company is not fully credible, it can be formally shrunk towards an industry estimate.

- Segmentation and layers

Very often the statistical model identified for a combined array of all payment types applies to some of the segments (by the same model, we mean the same parameter structure, not that the estimates are identical). Indeed, the variance of the normal distribution for a segment is larger than for the whole—on the original scale, the coefficient of variation is larger for the components. These ideas can also be applied to territories, etc., and to layers.

REFERENCES

[1] Brearly, Richard A. and Stewart C. Myers, *Principles of Corporate Finance*, New York, NY: McGraw Hill Inc., 1991.

[2] Brosius, Eric, "Loss Development Using Credibility," *Casualty Actuarial Society Part 7 Exam Study Kit*, 1992.

[3] Cook, R. Dennis and Sanford Weisberg, *Residuals and Influence in Regression*, New York, NY: Chapman and Hall, 1982.

[4] Dickson, David, Leanna Tedesco and Ben Zehnwirth, "Predictive Aggregate Claims Distributions," *Journal of Risk and Insurance* 65, 1998, pp. 689–709.

[5] Halliwell, Leigh, "Loss Prediction by Generalized Least Squares," *PCAS* LXXXIII, 1996, pp. 436–489.

[6] Mack, Thomas, "Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates," *ASTIN Bulletin*, 23, 2, 1993, pp. 213–225.

[7] Mack, Thomas, "Which Stochastic Model is Underlying the Chain Ladder Method?" *Insurance Mathematics and Economics*, 15, 2/3, 1994, pp. 133–138.

[8] Murphy, Daniel M., "Unbiased Loss Development Factors," *PCAS* LXXXI, 144–155, 1994, pp. 154–222.

[9] Reinsurance Association of America, *Historical Loss Development Study, 1991 Edition*, Washington, D.C.: R.A.A., 1991.

[10] Shapiro, Samuel S. and R. S. Francia, "Approximate Analysis of Variance Test for Normality," *Journal of the American Statistical Association* 67, 1972, pp. 215–216.

[11] Venter, Gary G., "Testing the Assumptions of Age-to-Age Factors," *PCAS* LXXXV, 1998, pp. 807–847.

[12] Zehnwirth, Ben, "Probabilistic Development Factor Models with Applications to Loss Reserve Variability, Prediction Intervals and Risk Based Capital," Casualty Actuarial Society *Forum*, Spring 1994, 2, pp. 447–605.

APPENDIX A

WEIGHTED LEAST SQUARES ESTIMATES WITHOUT INTERCEPT

In this appendix, we give a brief outline of the simplest case of the derivation of weighted least squares estimates. This corresponds to the no-intercept models at the start of Section 2—including the standard chain ladder.

Consider the following model:

$$y(i) = bx(i) + \varepsilon(i), \qquad \text{with} \tag{A.1a}$$

$$\varepsilon(i) \sim \text{Normal}(0, \sigma^2/w(i)). \tag{A.1b}$$

The value $w(i)$ is called the weight of observation i. Note that weights are inversely proportional to variances. Estimation of the parameters via maximum likelihood corresponds to minimizing the weighted sums of squared residuals:

$$\sum_{i=1}^{n} w(i)(y(i) - \hat{b}x(i))^2. \tag{A.2}$$

We find the minimum using calculus in a straightforward manner: taking derivatives with respect to $\hat{b}$, and setting the result equal to zero, we get

$$\sum_{i=1}^{n} -2x(i)w(i)(y(i) - \hat{b}x(i)) = 0; \tag{A.3}$$

then solving for $\hat{b}$ we obtain:

$$\hat{b} = \sum_{i=1}^{n} w(i)x(i)y(i) \Bigg/ \sum_{i=1}^{n} w(i)x(i)^2. \tag{A.4}$$

When $w(i) = x(i)^{-\delta}$, we obtain the estimates given near the start of Section 2; specifically, with $\delta = 1$ we obtain Equation 2.2; and, with $\delta = 2$, we obtain Equation 2.3.

APPENDIX B

CALCULATIONS FOR LINK RATIO MODELS WITH INTERCEPTS AND ACCIDENT-YEAR TRENDS

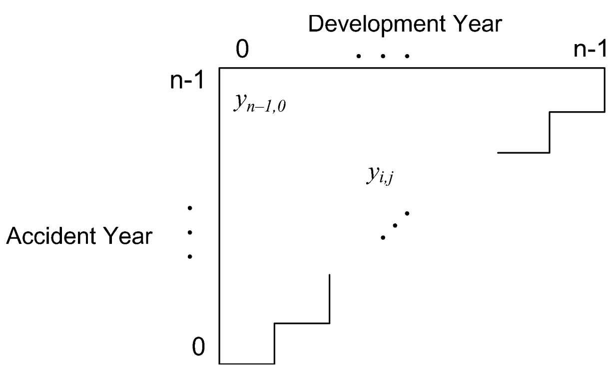
Let there be n accident years, numbering the most recent accident year as 0, and the first as $n-1$, as in Murphy [8]. Let y_{ij} be the cumulative amount paid in accident year i, development year j, $i = 0,\ldots,n-1$, $j = 0,\ldots,n-1$, as in Figure 33. This simplifies many of the formulas. Let $x_{ij} = y_{i,j-1}$, so that y_{ij}/x_{ij} is the observed development factor from $j-1$ to j in accident year i.

The only difference a more complex array shape (such as missing early payment years, or with late development years cut-off) will make is to change the limits on summations.

Now 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, x_{ij} is the cumulative paid in accident year i, up to development year $j-1$, and z_{ij} is the count of accident years from the top, starting from 0. Since we number from the bottom, in the current

FIGURE 33

TRIANGULAR LOSS DEVELOPMENT ARRAY OF SIZE n, WITH ACCIDENT YEARS LABELED IN REVERSE ORDER



notation, $z_{ij} = n-1-i$. Denote the cumulative by y_{ij}. Note that $p_{ij} = y_{ij} - x_{ij}$. Here, α_j is an intercept (level) term, λ_j represents accident-year trend, and β_j represents 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 three 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: $\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}$, where $\mathbf{y} = (y_{n-1}, y_{n-2}, \ldots, y_j)'$, $\boldsymbol{\beta} = (\alpha, \lambda, \beta)'$, $\mathbf{u} = (u_{n-1}, u_{n-2}, \ldots, u_j)'$,

$$X = \begin{bmatrix} 1 & z_{n-1} & x_{n-1} \\ 1 & z_{n-2} & x_{n-2} \\ \vdots & \vdots & \vdots \\ 1 & z_j & x_j \end{bmatrix} \qquad \text{and}$$

$$\mathrm{Var}(\mathbf{u}) = \sigma^2 U = \sigma^2 \begin{bmatrix} x_{n-1}^{\delta} & 0 & \cdots & 0 \\ 0 & x_{n-2}^{\delta} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & x_j^{\delta} \end{bmatrix}.$$

Let $\mathbf{y}^* = U^{-1/2}\mathbf{y}$, $X^* = U^{-1/2}X$, $\mathbf{e} = U^{-1/2}\mathbf{u}$. Then we have $\mathbf{y}^* = X^*\boldsymbol{\beta} + \mathbf{e}$, with the e_i's independently normal with variance σ^2. That is, $\mathbf{y}^* = (y_{n-1}x_{n-1}^{-\delta/2}, y_{n-2}x_{n-2}^{-\delta/2}, \ldots, y_j x_j^{-\delta/2})'$ and similarly with each column of X, including the column of 1s. The parameter estimates (and parameter variance-covariance matrix) for this new (unweighted) regression are the same as that of the old regression; standard regression results give $\hat{\boldsymbol{\beta}} = (X^{*\prime}X^*)^{-1}X^{*\prime}\mathbf{y}^*$ (see, for example, Cook and Weisberg [3]).

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 α and/or β as appropriate.

Residuals

Note that the residuals, $\hat{u}_{ij}$, are the same whether we consider incremental or cumulative values: $\hat{u}_{ij} = y_{ij} - \hat{y}_{ij} = p_{ij} - \hat{p}_{ij}$. We now derive the variance of the residuals. Let $H^* = X^*(X^{*\prime}X^*)^{-1}X^{*\prime}$. Note that $\hat{\mathbf{y}}^* = X^*\hat{\boldsymbol{\beta}} = X^*(X^{*\prime}X^*)^{-1}X^{*\prime}\mathbf{y}^* = H^*\mathbf{y}^*$. Then $\mathrm{Var}(\hat{\mathbf{e}}) = \mathrm{Var}(\mathbf{y}^* - \hat{\mathbf{y}}^*) = \mathrm{Var}((I - H^*)\mathbf{y}^*) = (I - H^*) \cdot \mathrm{Var}(\mathbf{y}^*)(I - H^*)' = \sigma^2(I - H^*)^2$. Note that $H^{*2} = H^*$. Hence $\mathrm{Var}(\hat{\mathbf{e}}) = \sigma^2(I - H^*)$. This is a standard regression result (see, for example, Cook and Weisberg [3]). Consequently $\mathrm{Var}(\hat{\mathbf{u}}) = U^{1/2}\mathrm{Var}(\hat{\mathbf{e}})(U^{1/2})'$, from which we can find the variance of an individual observation. Consider the regression for development year j, and suppress that subscript. Then $\mathrm{Var}(\hat{e}_i) = \sigma^2[1 - (\mathbf{x}_i^*)'(X^{*\prime}X^*)^{-1}\mathbf{x}_i^*]$, where $\mathbf{x}_i^*$ is the i^{th} row of X^*. Therefore $\mathrm{Var}(\hat{u}_i) = \sigma^2[1 - (\mathbf{x}_i^*)'(X^{*\prime}X^*)^{-1}\mathbf{x}_i^*]x_i^{\delta}$.

Forecasts and Standard Errors

Note that all forecasts and standard error calculations given here are conditional on the observed data.

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

$$\begin{aligned} \mathrm{Var}(\hat{y}_{i,i+k} - y_{i,i+k}) &= \mathrm{Var}(\hat{y}_{i,i+k} - \mu_{i,i+k} + \mu_{i,i+k} - y_{i,i+k}) \\ &= \mathrm{Var}(\hat{y}_{i,i+k} - \mu_{i,i+k}) + \mathrm{Var}(y_{i,i+k} - \mu_{i,i+k}) \\ &= v_{i,i+k}^{p} + v_{i,i+k}^{e}. \end{aligned} \tag{B.1}$$

The first term is what Murphy [8] calls the parameter variance, and the second the process variance. These may be thought of as the variability of the *predictions* about the true model and the variability of the *data* about the true model, respectively. Both must be estimated. Now:

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

Note that if X and Y are independent random variables, with means μ_X and μ_Y, respectively, then $\mathrm{E}[(X-\mu_X)^2(Y-\mu_Y)^2] = \mathrm{E}[(X-\mu_X)^2]\mathrm{E}[(Y-\mu_Y)^2]$. Expanding the left hand side, using the elementary properties of expectation and rearranging, we readily obtain $\mathrm{Var}(XY) = \mathrm{Var}(X)\mathrm{Var}(Y) + \mathrm{Var}(X)\mu_Y^2 + \mathrm{Var}(Y)\mu_X^2$.

Consequently,

$$\begin{aligned} \mathrm{Var}(\hat{\beta}_{i+k}\hat{y}_{i,i+k-1}) &= \beta^2_{i+k}\mathrm{Var}(\hat{y}_{i,i+k-1}) + y^2_{i,i+k-1}\mathrm{Var}(\hat{\beta}_{i+k}) \\ &\quad + \mathrm{Var}(\hat{\beta}_{i+k})\mathrm{Var}(\hat{y}_{i,i+k-1}), \end{aligned} \tag{B.3}$$

which we estimate by

$$\hat{\beta}^2_{i+k}\,\hat{\mathrm{Var}}(\hat{y}_{i,i+k-1}) + \hat{y}^2_{i,i+k-1}\,\hat{\mathrm{Var}}(\hat{\beta}_{i+k}) + \hat{\mathrm{Var}}(\hat{\beta}_{i+k})\,\hat{\mathrm{Var}}(\hat{y}_{i,i+k-1}). \tag{B.4}$$

Hence we estimate

$$\begin{aligned}\hat{v}^p_{i,i+k} = \hat{\text{Var}}(\hat{\alpha}_{i+k}) + 2z_{i,i+k}\,\hat{\text{Cov}}(\hat{\alpha}_{i+k},\hat{\lambda}_{i+k}) + z^2_{i,i+k}\,\hat{\text{Var}}(\hat{\lambda}_{i+k})\\ + 2\hat{y}_{i,i+k-1}\,\hat{\text{Cov}}(\hat{\alpha}_{i+k},\hat{\beta}_{i+k}) + 2z_{i,i+k}\hat{y}_{i,i+k-1}\,\hat{\text{Cov}}(\hat{\lambda}_{i+k},\hat{\beta}_{i+k})\\ + \hat{y}^2_{i,i+k-1}\,\hat{\text{Var}}(\hat{\beta}_{i+k}) + [\hat{\beta}^2_{i+k} + \hat{\text{Var}}(\hat{\beta}_{i+k})]\hat{v}^p_{i,i+k-1}. \qquad \text{(B.5)}\end{aligned}$$

Note that v^p_{ii} is zero, since we are conditioning on the data. Further,

$$\begin{aligned}v^e_{i,i+k} &= \text{Var}(y_{i,i+k} - \mu_{i,i+k})\\ &= \text{Var}(\alpha_{i+k} + \lambda_{i+k}z_{i,i+k-1} + \beta_{i+k}y_{i,i+k-1} + u_{i,i+k})\\ &= \beta^2_{i+k}\text{Var}(y_{i,i+k-1}) + \text{Var}(u_{i,i+k})\\ &= \beta^2_{i+k}\text{Var}(y_{i,i+k-1} - \mu_{i,i+k-1}) + \sigma^2_{i+k}x^\delta_{i,i+k}\\ &= \beta^2_{i+k}v^e_{i,i+k-1} + \sigma^2_{i+k}x^\delta_{i,i+k}, \qquad \text{(B.6)}\end{aligned}$$

which we estimate as:

$$\begin{aligned}\hat{v}^e_{i,i+k} &= \hat{\text{Var}}(y_{i,i+k})\\ &= \hat{\beta}^2_{i+k}\,\hat{\text{Var}}(y_{i,i+k-1}) + \hat{\sigma}^2_{i+k}\hat{\text{E}}(y^\delta_{i,i+k-1})\\ &= \hat{\beta}^2_{i+k}\hat{v}^e_{i,i+k-1} + \hat{\sigma}^2_{i+k}\hat{f}^\delta_{i,i+k-1}, \qquad \text{(B.7)}\end{aligned}$$

where $f^\delta_{i,j} = \text{E}(y^\delta_{i,j})$. We have

$$\hat{f}^\delta_{ij} = \begin{cases} 1, & \delta = 0\\ \hat{y}_{ij}, & \delta = 1\\ \hat{y}^2_{ij} + \hat{\text{Var}}(y_{ij}), & \delta = 2\end{cases}, \qquad \text{(B.8)}$$

just as with Murphy [8]. With the normality assumption we can obtain estimates at other values of δ, but we omit details here.

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

Forecasts and Standard Errors of Development-Year Totals

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

$$D_j = \sum_{i=0}^{j-1} y_{ij}, \tag{B.9}$$

and

$$\hat{D}_j = \sum_{i=0}^{j-1} \hat{y}_{ij}. \tag{B.10}$$

Note that

$$\begin{aligned}\mathrm{Var}(\hat{D}_j - D_j) &= \mathrm{Var}\left(\sum_{i=0}^{j-1} \hat{y}_{ij} - y_{ij}\right)\\ &= \mathrm{Var}\left(\sum_{i=0}^{j-1} \hat{y}_{ij} - \mu_{ij}\right) + \mathrm{Var}\left(\sum_{i=0}^{j-1} y_{ij} - \mu_{ij}\right)\\ &= V_j^p + V_j^e. \end{aligned} \tag{B.11}$$

Taking $Z_j = \sum_{i=0}^{j-1} z_{ij}$, then

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

$$
\begin{aligned}
&= n_j^2\mathrm{Var}(\hat{\alpha}_j) + 2n_jZ_j\mathrm{Cov}(\hat{\alpha}_j,\hat{\lambda}_j) + Z_j^2\mathrm{Var}(\hat{\lambda}_j)\\
&\quad + 2n_j[\hat{D}_{j-1} + y_{j-1,j-1}]\mathrm{Cov}(\hat{\alpha}_j,\hat{\beta}_j)\\
&\quad + 2Z_j[\hat{D}_{j-1} + y_{j-1,j-1}]\mathrm{Cov}(\hat{\lambda}_j,\hat{\beta}_j)\\
&\quad + \mathrm{Var}(\hat{\beta}_j[\hat{D}_{j-1} + y_{j-1,j-1}])\\
&= n_j^2\mathrm{Var}(\hat{\alpha}_j) + 2n_jZ_j\mathrm{Cov}(\hat{\alpha}_j,\hat{\lambda}_j) + Z_j^2\mathrm{Var}(\hat{\lambda}_j)\\
&\quad + 2[\hat{D}_{j-1} + y_{j-1,j-1}][n_j\mathrm{Cov}(\hat{\alpha}_j,\hat{\beta}_j) + Z_j\mathrm{Cov}(\hat{\lambda}_j,\hat{\beta}_j)]\\
&\quad + [D_{j-1} + y_{j-1,j-1}]^2\mathrm{Var}(\hat{\beta}_j) + [\beta_j^2 + \mathrm{Var}(\hat{\beta}_j)]\mathrm{Var}(\hat{D}_{j-1}).
\end{aligned}
\tag{B.12}
$$

Note that the last term, $\mathrm{Var}(\hat{D}_{j-1})$ is just V_{j-1}^p. We estimate V_j^p by replacing D_{j-1}, β_j, and the variance and covariance terms by their estimates, which have either been defined or are immediately available from standard regression calculations. Also,

$$
\begin{aligned}
V_j^e &= \mathrm{Var}\left(\sum_{i=0}^{j-1} y_{ij}\right)\\
&= \mathrm{Var}\left(n_j\alpha_j + Z_j\lambda_j + \beta_j\sum_{i=0}^{j-1} y_{i,j-1} + \sum_{i=0}^{j-1} u_{i,j}\right)\\
&= \mathrm{Var}(\beta_j[D_{j-1} + y_{j-1,j-1}]) + \mathrm{Var}\left(\sum_{i=0}^{j-1} u_{i,j}\right)\\
&= \beta_j^2\mathrm{Var}(D_{j-1}) + \sigma_j^2\sum_{i=0}^{j-1} x_{ij}^{\delta}\\
&= \beta_j^2 V_{j-1}^e + \sigma_j^2\left(y_{j-1,j-1}^{\delta} + \sum_{i=0}^{j-2} y_{i,j-1}^{\delta}\right).
\end{aligned}
\tag{B.13}
$$

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 C

EXPOSURES AND FORECASTING

Let us extend the notation of the previous section. Let y^o_{ij} be the observed cumulative at accident year i, development year j; let y^n_{ij} be the corresponding normalized-for-exposures cumulative. Similarly, let p^o_{ij} and p^n_{ij} be the corresponding incremental values.

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

Individual Forecasts and Standard Errors (all models)

Clearly, $\hat{y}^o_{ij} = c_i\hat{y}^n_{ij}$, and similarly for p, $\hat{p}^o_{ij} = c_i\hat{p}^n_{ij}$; also $\mathrm{Var}(\hat{y}^o_{ij} - y^o_{ij}) = c_i^2\mathrm{Var}(\hat{y}^n_{ij} - y^n_{ij})$ and $\mathrm{Var}(\hat{p}^o_{ij} - p^o_{ij}) = c_i^2\mathrm{Var}(\hat{p}^n_{ij} - p^n_{ij})$. So individual observed forecasts and standard errors are just the corresponding normalized values, multiplied by the exposure.

Cumulative Development-Year Totals

$$\mathrm{Var}(\hat{D}^o_j - D^o_j) = V^{op}_j + V^{oe}_j, \tag{C.1}$$

where

$$\begin{aligned}
V^{op}_j &= \mathrm{Var}\left(\sum_{i=0}^{j-1}\hat{y}^o_{ij}\right)\\
&= \mathrm{Var}\left(\sum_{i=0}^{j-1}c_i\hat{y}^n_{ij}\right)\\
&= \mathrm{Var}\left[\sum_{i=0}^{j-1}c_i(\hat{\alpha}_j + \hat{\lambda}_j z_{ij} + \hat{\beta}_j\hat{y}^n_{i,j-1})\right]\\
&= \mathrm{Var}\left[\hat{\alpha}_j\left(\sum_{i=0}^{j-1}c_i\right) + \hat{\lambda}\left(\sum_{i=0}^{j-1}c_i z_{ij}\right) + \hat{\beta}_j\sum_{i=0}^{j-1}\hat{y}^o_{i,j-1}\right]
\end{aligned}$$

$$= \mathrm{Var}\left[\hat{\alpha}_j C_j + \hat{\lambda}_j Z_j^* + \hat{\beta}_j \sum_{i=0}^{j-1} \hat{y}_{i,j-1}^o\right]$$

$$= C_j^2 \mathrm{Var}(\hat{\alpha}_j) + 2C_j Z_j^* \mathrm{Cov}(\hat{\alpha}_j, \hat{\lambda}_j) + (Z_j^*)^2 \mathrm{Var}(\hat{\lambda}_j)$$

$$+ 2C_j(y_{j-1,j-1}^o + \hat{D}_{j-1}^o)\mathrm{Cov}(\hat{\alpha}_j, \hat{\beta}_j)$$

$$+ 2Z_j^*(y_{j-1,j-1}^o + \hat{D}_{j-1}^o)\mathrm{Cov}(\hat{\lambda}_j, \hat{\beta}_j)$$

$$+ (y_{j-1,j-1}^o + D_{j-1}^o)^2 \mathrm{Var}(\hat{\beta}_j) + [\beta_j^2 + \mathrm{Var}(\hat{\beta}_j)]\mathrm{Var}(\hat{D}_{j-1}^o)$$

$$= C_j^2 \mathrm{Var}(\hat{\alpha}_j) + 2C_j Z_j^* \mathrm{Cov}(\hat{\alpha}_j, \hat{\lambda}_j) + (Z_j^*)^2 \mathrm{Var}(\hat{\lambda}_j)$$

$$+ 2(y_{j-1,j-1}^o + \hat{D}_{j-1}^o)[C_j \mathrm{Cov}(\hat{\alpha}_j, \hat{\beta}_j) + Z_j^* \mathrm{Cov}(\hat{\lambda}_j, \hat{\beta}_j)]$$

$$+ (y_{j-1,j-1}^o + D_{j-1}^o)^2 \mathrm{Var}(\hat{\beta}_j) + [\beta_j^2 + \mathrm{Var}(\hat{\beta}_j)]V_{j-1}^{op}, \tag{C.2}$$

and

$$V_j^{oe} = \mathrm{Var}\left(\sum_{i=0}^{j-1} y_{ij}^o\right)$$

$$= \mathrm{Var}\left(\sum_{i=0}^{j-1} c_i y_{ij}^n\right)$$

$$= \mathrm{Var}\left[\sum_{i=0}^{j-1} c_i(\alpha_j + \lambda_j z_{ij} + \beta_j y_{i,j-1}^n + u_{ij})\right]$$

$$= \mathrm{Var}\left(\sum_{i=0}^{j-1} c_i u_{ij}\right) + \beta_j^2 \mathrm{Var}\left[\sum_{i=0}^{j-1} y_{i,j-1}^o\right]$$

$$= \sum_{i=0}^{j-1} \mathrm{Var}(c_i u_{ij}) + \beta_j^2 \mathrm{Var}\left[\sum_{i=0}^{j-1} y_{i,j-1}^o\right]$$

$$= \sigma_j^2 c_{j-1}^2 (y_{j-1,j-1}^n)^\delta + \sum_{i=0}^{j-2} c_i^2 v_{i,j-1}^{ne} + \beta_j^2 V_{j-1}^{oe}$$

$$= \sigma_j^2 c_{j-1}^{2-\delta} (y_{j-1,j-1}^o)^\delta + \sum_{i=0}^{j-2} v_{i,j-1}^{oe} + \beta_j^2 V_{j-1}^{oe}, \qquad \text{(C.3)}$$

where again we estimate this variance by replacing the D's, β's, variances and covariances by their estimates.

APPENDIX D

LIKELIHOOD AND CONDITIONAL REGRESSIONS

Let $y(j)$ be the vector of data in development year j, and $\theta(j)$ be all the parameters for that development year. Let $\mathbf{y} = (y(0)',\ldots,y(n-1)')'$, so the development years are stacked one on top of the other, and $\theta = (\theta(1)',\ldots,\theta(n-1)')'$. Then, straightforward application of conditional probability and some simplification gives us:

$$\begin{aligned}
L(\theta \mid \mathbf{y}) &\propto p[\mathbf{y} \mid \theta] \\
&\propto p[y(n-1) \mid \theta(n-1), y(n-2), y(n-3),\ldots,y(0)] \\
&\quad \cdot p[y(n-2) \mid \theta(n-2), y(n-3), y(n-4),\ldots,y(0)] \\
&\quad\ \vdots \\
&\quad \cdot p[y(1) \mid \theta(1), y(0)] \\
&\quad \cdot p[y(0)] \\
&\propto p[y(n-1) \mid \theta(n-1), y(n-2)] \\
&\quad \cdot p[y(n-2) \mid \theta(n-2), y(n-3)] \\
&\quad \cdots p[y(1) \mid \theta(1), y(0)] \cdot p[y(0)].
\end{aligned} \tag{D.1}$$

Since, for each regression, we are conditioning on the data from previous development 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 E

DESIGN MATRICES FOR THE MODELS DESCRIBED IN SECTION 3

Readers may wish to fit the regression models described in Section 3 of this paper. The models described there can be fitted to data in any of the common statistical packages, or using spreadsheet software 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, as it's overparameterized, but is the general case of which all the useful models are special cases) and then some of the more common simpler models.

Table E.1 displays the expected values in each cell in the log(incremental) array under the general model using the notation of Section 3.

The vector of observations may be produced by stacking up the development years one on top of another: $\mathbf{y} = (y(0)', y(1)', \ldots, y(n-1)')'$, as in the previous appendix. Similarly, there is 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 already an intercept parameter, so we don't add an intercept (i.e., the regression is written $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$). A good approach is to do all the α's, then all the γ's, and finally all the ι's.

For $n = 4$, this corresponds to the X-matrix in Table E.2 (the zeroes have been suppressed to make the patterns more clear).

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 1's for the columns for γ_k (where $k \leq j$), and it would have 1's 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 ex-

TABLE E.1

EXPECTED VALUES OF LOG(INCREMENTAL) UNDER THE GENERAL MODEL

Development Year

Accident Year	0	1	…	J	…	$n-1$
0	α_0	$\alpha_0+\gamma_1+\iota_1$	⋯	$\alpha_0+\sum_{k=1}^{j}\gamma_k+\sum_{r=1}^{j}\iota_r$	⋯	$\alpha_0+\sum_{k=1}^{j}\gamma_k+\sum_{r=1}^{j}\iota_r$
1	$\alpha_1+\iota_1$	$\alpha_1+\gamma_1+\sum_{r=1}^{2}\iota_t$	⋯	$\alpha_1+\sum_{k=1}^{j}\gamma_k+\sum_{r=1}^{j+1}\iota_r$	⋯	
⋮	⋮	⋮		⋮	⋰	
i	$\alpha_{iI}+\sum_{r=1}^{i}\iota_r$	$\alpha_1+\gamma_1+\sum_{r=1}^{i+1}\iota_r$		$\alpha_i+\sum_{k=1}^{j}\gamma_k+\sum_{r=1}^{i+j}\iota_r$		
⋮	⋮	⋮	⋰			
$n-1$	$\alpha_{n-1}+\sum_{r=1}^{t}\iota_r$					

TABLE E.2

DESIGN MATRIX (X-MATRIX) FOR THE FULL MODEL FOR A TRIANGLE WITH 4 YEARS' DATA

	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 E.3

Design Matrix (X-Matrix) for a Simple Model for a Triangle With 4 Years' Data

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

ample, Table E.3 shows the design matrix for the array of size 4, with one level of log payments for all years, 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.