TESTING THE ASSUMPTIONS OF AGE-TO-AGE FACTORS

GARY G. VENTER

Abstract

The use of age-to-age factors applied to cumulative losses has been shown to produce least-squares optimal reserve estimates when certain assumptions are met. Tests of these assumptions are introduced, most of which derive from regression diagnostic methods. Failures of various tests lead to specific alternative methods of loss development.

ACKNOWLEDGEMENT

I would like to thank the Committee on Review of Papers for comments leading to numerous expository improvements over previous drafts. Any obscurity that remains is of course my own.

INTRODUCTION

In his paper "Measuring the Variability of Chain Ladder Reserve Estimates" Thomas Mack presented the assumptions needed for least-squares optimality to be achieved by the typical age-to-age factor method of loss development (often called "chain ladder"). Mack also introduced several tests of those assumptions. His results are summarized below, and then other tests of the assumptions are introduced. Also addressed is what to do when the assumptions fail. Most of the assumptions, if they fail in a particular way, imply least-squares optimality for some alternative method.

The organization of the paper is to first show Mack's three assumptions and their result, then to introduce six testable implications of those assumptions, and finally to go through the testing of each implication in detail.

PRELIMINARIES

Losses for accident year w evaluated at the end of that year will be denoted as being as of age 0, and the first accident year in the triangle is year 0. The notation below will be used to specify the models. Losses could be either paid or incurred. Only development that fills out the triangle is considered. Loss development beyond the observed data is often significant but is not addressed here. Thus age ∞ will denote the oldest possible age in the data triangle.

Notation

$c(w,d)$:	cumulative loss from accident year w as of age d
$c(w,\infty)$:	total loss from accident year w when end of triangle reached
$q(w,d)$:	incremental loss for accident year w from $d-1$ to d
$f(d)$:	factor applied to $c(w,d)$ to estimate $q(w,d+1)$
$F(d)$:	factor applied to $c(w,d)$ to estimate $c(w,\infty)$

Assumptions

Mack showed that some specific assumptions on the process of loss generation are needed for the chain ladder method to be optimal. Thus if actuaries find themselves in disagreement with one or another of these assumptions, they should look for some other method of development that is more in harmony with their intuition about the loss generation process. Reserving methods more consistent with other loss generation processes will be discussed below. Mack's three original assumptions are slightly restated here to emphasize the task as one of predicting future incremental losses. Note that the losses $c(w,d)$ have an evaluation date of $w+d$.

1. $\mathrm{E}[q(w,d+1) \mid \text{data to } w+d] = f(d)c(w,d)$.

   In words, the expected value of the incremental losses to emerge in the next period is proportional to the total losses emerged to date, by accident year. Note that in Mack's definition of the chain ladder, $f(d)$ does not depend on w, so the factor for a given age is constant across accident years. Note also that this formula is a linear relationship with no constant term. As opposed to other models discussed below, the factor applies directly to the cumulative data, not to an estimated parameter, like ultimate losses. For instance, the Bornhuetter-Ferguson method assumes that the expected incremental losses are proportional to the ultimate for the accident year, not the emerged to date.

2. Unless $v = w$, $c(w,d)$ and $c(v,g)$ are independent for all v, w, d and g.

   This would be violated, for instance, if there were a strong diagonal, when all years' reserves were revised upwards. In this case, instead of just using the chain ladder method, most actuaries would recommend eliminating these diagonals or adjusting them. Some model-based methods for formally recognizing diagonal effects are discussed below.

3. $\mathrm{Var}[q(w,d+1) \mid \text{data to } w+d] = a[d,c(w,d)]$.

   That is, the variance of the next increment observation is a function of the age and the cumulative losses to date. Note that $a(\cdot,\cdot)$ can be any function but does not vary by accident year. An assumption on the variance of the next incremental losses is needed to find a least-squares optimal method of estimating the development factors. Different assumptions, e.g., different functions $a(\cdot,\cdot)$ will lead to optimality for different methods of estimating the factor f. The form of $a(\cdot,\cdot)$ can be tested by trying different forms, estimating the f's, and seeing if the variance formula holds. There will almost always be some function $a(\cdot,\cdot)$ that reasonably accords with the observations, so the issue with this assumption is not its validity but its implications for the estimation procedure.

Results (Mack)

In essence what Mack showed is that under the above assumptions the chain ladder method gives the minimum variance unbiased linear estimator of future emergence. This gives a good justification for using the chain ladder in that case, but the assumptions need to be tested. Mack assumed that $a[d,c(w,d)] = k(d)c(w,d)$, that is, he assumed that the variance is proportional to the previous cumulative loss, with possibly a different proportionality factor for each age. In this case, the minimum variance unbiased estimator of $c(w,\infty)$ from the triangle of data to date $w+d$ is $F(d)c(w,d)$, where the age-to-ultimate factor $F(d) = [1 + f(d)][1 + f(d+1)]\cdots$, and $f(d)$ is calculated as:

$$f(d) = \sum_w q(w,d+1) \Big/ \sum_w c(w,d),$$

where the sum is over the w's mutually available in both columns (assuming accident years are on separate rows and ages are in separate columns). Actuaries often use a modified chain ladder that uses only the last n diagonals. This will be one of the alternative methods to test if Mack's assumptions fail. Using only part of the data when all the assumptions hold will reduce the accuracy of the estimation, however.

Extension

In general, the minimum variance unbiased $f(d)$ is found by minimizing

$$\sum_w [f(d)c(w,d) - q(w,d+1)]^2 k(d)/a[d,c(w,d)].$$

This is the usual weighted least-squares result, where the weights are inversely proportional to the variance of the quantity being estimated. Because only proportionality, not equality, to the variance is required, $k(d)$ can be any convenient function of d—usually chosen to simplify the minimization.

For example, suppose $a[d,c(w,d)] = k(d)c(w,d)^2$. Then the $f(d)$ produced by the weighted least-squares procedure is the average of the individual accident year d to $d+1$ ratios, $q(w,d+1)/c(w,d)$. For $a[d,c(w,d)] = k(d)$, each $f(d)$ regression above is then just standard unweighted least squares, so $f(d)$ is the regression coefficient $\sum_w c(w,d)q(w,d+1)/\sum_w c(w,d)^2$. (See Murphy [8].) In all these cases, $f(d)$ is fit by a weighted regression, and so regression diagnostics can be used to evaluate the estimation. In the tests below just standard least-squares will be used, but in application the variance assumption should be reviewed.

Discussion

Without going into Mack's derivation, the optimality of the chain ladder method is fairly intuitive from the assumptions. In particular, the first assumption is that the expected emergence in the next period is proportional to the losses emerged to date. If that were so, then a development factor applied to the emerged to date would seem highly appropriate. Testing this assumption will be critical to exploring the optimality of the chain ladder. For instance, if the emergence were found to be a constant plus a percent of emergence to date, then a different method would be indicated—namely, a factor plus constant development method. On the other hand, if the next incremental emergence were proportional to ultimate rather than to emerged to date, a Bornhuetter-Ferguson type approach would be more appropriate.

To test this assumption against its alternatives, the development method that leads from each alternative needs to be fit, and then a goodness-of-fit measure applied. This is similar to trying a lot of methods and seeing which one you like best, but it is different in two respects: (1) each method tested derives from an alternative assumption on the process of loss emergence; (2) there is a specific goodness-of-fit test applied. Thus the fitting is a test of the emergence patterns that the losses are subject to, and not just a test of estimation methods.

TESTABLE IMPLICATIONS OF ASSUMPTIONS

Verifying a hypothesis involves finding as many testable implications of that hypothesis as possible, and verifying that the tests are passed. In fact a hypothesis can never be fully verified, as there could always be some other test you haven't thought of. Thus the process of verification is sometimes conceived as being really a process of attempted falsification, with the current tentatively-accepted hypothesis being the strongest (i.e., most easily testable) one not yet falsified. (See Popper [9].) The assumptions (1)–(3) are not directly testable, but they have testable implications. Thus they can be falsified if any of the implications are found not to hold, which would mean that the optimality of the chain ladder method could not be shown for the data in question. Holding up under all of these tests would increase the actuary's confidence in the hypothesis, still recognizing that no hypothesis can ever be fully verified. Some of the testable implications are:

1. Significance of factor $f(d)$.
2. Superiority of factor assumption to alternative emergence patterns such as:
   (a) linear with constant: $\mathrm{E}[q(w,d+1) \mid \text{data to } w+d] = f(d)c(w,d) + g(d)$;
   (b) factor times parameter: $\mathrm{E}[q(w,d+1) \mid \text{data to } w+d] = f(d)h(w)$;
   (c) including calendar year effect: $\mathrm{E}[q(w,d+1) \mid \text{data to } w+d] = f(d)h(w)g(w+d)$.

Note that in these examples the notation has changed slightly so that $f(d)$ is a factor used to estimate $q(w, d+1)$, but not necessarily applied to $c(w,d)$. These alternative emergence models can be tested by goodness of fit, controlling for number of parameters.

3. Linearity of model: look at residuals as a function of $c(w,d)$.
4. Stability of factor: look at residuals as a function of time.
5. No correlation among columns.
6. No particularly high or low diagonals.

The remainder of this paper consists of tests of these implications.

TESTING LOSS EMERGENCE—IMPLICATIONS 1 & 2

The first four of these implications are tests of assumption (1). Standard diagnostic tests for weighted least-squares regression can be used as measures.

Implication 1: Significance of Factors

Regression analysis produces estimates for the standard deviation of each parameter estimated. Usually the absolute value of a factor is required to be at least twice its standard deviation for the factor to be regarded as significantly different from zero. This is a test failed by many development triangles, which means that the chain ladder method is not optimal for those triangles.

The requirement that the factor be twice the standard deviation is not a strict statistical test, but more like a level of comfort. For the normal distribution this requirement provides that there is only a probability of about 4.5% of getting a factor of this absolute value or greater when the true factor is zero. Many analysts are comfortable with a factor with absolute value 1.65 times its standard deviation, which could happen about 10% of the time by chance alone. For heavier-tailed distributions, the same ratio of factor to standard deviation will usually be more likely to occur by chance. Thus, if a factor were to be considered not significant for the normal distribution, it would probably be even less significant for other distributions. This approach could be made into a formal statistical test by finding the distribution that the factors follow. The normal distribution is often satisfactory, but it is not unusual to see some degree of positive skewness, which would suggest the lognormal. Some of the alternative models discussed below are easier to estimate in log form, so that is not an unhappy finding.

It may be tempting to do the regression of cumulative on previous cumulative and test the significance of that factor in order to justify the use of the chain ladder. However it is only the incrementals that are being predicted, so this would have to be carefully interpreted. In a cumulative-to-cumulative regression, the significance of the difference of the factor from unity is what needs to be tested. This can be done by comparing that difference to the standard deviation of the factor, which is equivalent to testing the significance of the factor in the incremental-to-cumulative regression. Some alternative methods to try when this assumption fails are discussed below.

Implication 2: Superiority to Alternative Emergence Patterns

If alternative emergence patterns give a better explanation of the data triangle observed to date, then assumption (1) of the chain ladder model is also suspect. In these cases development based on the best-fitting emergence pattern would be a natural option to consider. The sum of the squared errors (SSE) would be a way to compare models (the lower the better) but this should be adjusted to take into account the number of parameters used. Unfortunately it appears that there is no generally accepted method to make this adjustment. One possible adjustment is to compare fits by using the SSE divided by $(n-p)^2$, where n is the number of observations and p is the number of parameters. More parameters give an advantage in fitting but a disadvantage in prediction, so such a penalty in adjusting the residuals may be appropriate. A more popular adjustment in recent years is to base goodness of fit on the Akaike Information Criterion, or AIC (see Lütkepohl [5]). For a fixed set of observations, multiplying the SSE by $e^{2p/n}$ can approximate the effect of the AIC. The AIC has been criticized as being too permissive of over-parameterization for large data sets, and the Bayesian Information Criterion, or BIC, has been suggested as an alternative. Multiplying the SSE by $n^{p/n}$ would rank models the same as the BIC. As a comparison, if you have 45 observations, the improvement in SSE needed to justify adding a 5th parameter to a 4 parameter model is about 5%, $4\frac{1}{2}$%, and almost 9%, respectively, for these three adjustments. In the model testing below the sum of squared residuals divided by $(n-p)^2$ will be the test statistic, but in general the AIC and BIC should be regarded as good alternatives.

Note again that this is not just a test of development methods but is also a test to see what hypothetical loss generation process is most consistent with the data in the triangle.

The chain ladder has one parameter for each age, which is less than for the other emergence patterns listed in implication 2. This gives it an initial advantage, but if the other parameters improve the fit enough, they overcome this advantage. In testing the various patterns below, parameters will be fit by minimizing the sum of squared residuals. In some cases this will require an iterative procedure.

Alternative Emergence Pattern 1: Linear with Constant

The first alternative mentioned is just to add a constant term to the model. This is often significant in the age 0 to age 1 stage, especially for highly variable and slowly reporting lines, such as excess reinsurance. In fact, in the experience of myself and other actuaries who have reported informally, the constant term has often been found to be more statistically significant than the factor itself. If the constant is significant and the factor is not, a different development process is indicated. For instance in some triangles earning of additional exposure could influence the 0-to-1 development. It is important in such cases to normalize the triangle as much as possible, e.g., by adjusting for differences among accident years in exposure and cost levels (trend). With these adjustments a purely additive rather than a purely multiplicative method could be more appropriate.

Again, the emergence assumption underlying the linear with constant method is:

$$\mathrm{E}[q(w,d+1) \mid \text{data to } w+d] = f(d)c(w,d) + g(d).$$

If the constant is statistically significant, this emergence pattern is more strongly supported than that underlying the chain ladder.

Alternative Emergence Pattern 2: Factor Times Parameter

The chain ladder model expresses the next period's loss emergence as a factor times losses emerged so far. An important alternative, suggested by Bornhuetter and Ferguson (BF) in 1972, is to forecast the future emergence as a factor times estimated ultimate losses. While BF use some external measure of ultimate losses in this process, others have tried to use the data triangle itself to estimate the ultimate (e.g., see Verrall [13]). In this paper, models that estimate emerging losses as a percent of ultimate will be called parameterized BF models, even if they differ from the original BF method in how they estimate the ultimate losses.

The emergence pattern assumed by the parameterized BF model is:

$$\mathrm{E}[q(w,d+1) \mid \text{data to } w+d] = f(d)h(w).$$

That is, the next period expected emerged loss is a lag factor $f(d)$ times an accident year parameter $h(w)$. The latter could be interpreted as expected ultimate for the year, or at least proportional to that. This model thus has a parameter for each accident year as well as for each age (one less actually, as you can assume the $f(d)$'s sum to one—which makes $h(w)$ an estimate of ultimate losses; thus multiplying all the $f(d)$'s, $d > 0$, by a constant and dividing all the h's by the same constant will not change the forecasts). For reserving purposes there is even one fewer parameter, as the age 0 losses are already in the data triangle, so $f(0)$ is not needed. Thus, for a complete triangle with n accident years the BF has $2n-2$ parameters, or twice the number as the chain ladder. This will result in a penalty to goodness of fit, so the BF has to produce much lower fit errors than the chain ladder to give a better test statistic.

Testing the parameterized BF emergence pattern against that of the chain ladder cannot be done just by looking at the statistical significance of the parameters, as it could with the linear plus constant method, as one is not a special case of the other. This testing is the role of the test statistic, the sum of squared residuals divided by the square of the degrees of freedom. If this statistic is better for the BF model, that is evidence that the emergence pattern of the BF is more applicable to the triangle being studied. That would suggest that loss emergence for that book can be more accurately represented as fluctuating around a proportion of ultimate losses rather than a percentage of previously emerged losses.

Stanard [10] assumed a loss generation scheme that resulted in the expected loss emergence for each period being proportional to the ultimate losses for the period. This now can be seen to be the BF emergence pattern. Then by generating actual loss emergence stochastically, he tested some loss development methods. The chain ladder method gave substantially larger estimation errors for ultimate losses than his other methods, which were basically different versions of BF estimation. This illustrates how far off reserves can be when one reserving technique is applied to losses that have an emergence process different from the one underlying the technique.

A simulation in accord with the chain ladder emergence assumption would generate losses at age j by multiplying the simulated emerged losses at age $j-1$ by a factor and then adding a random component. In this manner the random components influence the expected emergence at all future ages. This may seem an unlikely way for losses to emerge, but it is for the triangles that follow this emergence pattern that the chain ladder will be optimal. The fact that Stanard used the simulation method consistent with the BF emergence pattern, and this was not challenged by the reviewer, John Robertson, suggests that actuaries may be more comfortable with the BF emergence assumptions than with those of the chain ladder. Or perhaps it just means that no one would be likely to think of simulating losses by the chain ladder method.

An important special case of the parameterized BF was developed by some Swiss and American reinsurance actuaries at a meeting in Cape Cod, and is sometimes called the Cape Cod method (CC). It is given by setting $h(w)$ to just a single h for all accident years. CC seems to have one more parameter than the chain ladder, namely h. However, any change in h can be offset by inverse changes in all the f's. CC thus has the same number of parameters as the chain ladder, and so its fit measure is not as heavily penalized as that of BF. However a single h requires a relatively stable level of loss exposure across accident years. Again it would be necessary to adjust for known exposure and price level differences among accident years, if using this method. The chain ladder and BF can handle changes in level from year to year as long as the development pattern remains consistent.

The BF model often has too many parameters. The last few accident years especially are left to find their own levels based on sparse information. Reducing the number of parameters, and thus using more of the information in the triangle, can often yield better predictions, especially in predicting the last few years. It could be that losses follow the BF emergence pattern, but this is disguised in the test statistic due to too many parameters. Thus, testing for the alternate emergence pattern should also include testing reduced parameter BF models.

The full BF not only assumes that losses emerge as a percentage of ultimate, but also that the accident years are all at different mean levels and that each age has a different percentage of ultimate losses. It could be, however, that several years in a row, or all of them, have the same mean level. If the mean changes, there could be a gradual transition from one level to another over a few years. This could be modeled as a linear progression of accident year parameters, rather than separate parameters for each year. A similar process could govern loss emergence. For instance, the 9th through 15th periods could all have the same expected percentage development. Finding these relationships and incorporating them in the fitting process will help determine what emergence process is generating the development.

The CC model can be considered a reduced parameter BF model. The CC has a single ultimate value for all accident years, while the BF has a separate value for each year. There are numerous other ways to reduce the number of parameters in BF models. Simply using a trend line through the BF ultimate loss parameters would use just two accident year parameters in total instead of one for each year. Another method might be to group years using apparent jumps in loss levels and fit an h parameter separately to each group. Within such groupings it is also possible to let each accident year's h parameter vary somewhat from the group average, e.g., via credibility, or to let it evolve over time, e.g., by exponential smoothing.

Alternative Emergence Patterns Example

Table 1 shows incremental incurred losses by age for some excess casualty reinsurance. As an initial test, the statistical sig-

TABLE 1

INCREMENTAL INCURRED LOSSES

	Age									
Year	0	1	2	3	4	5	6	7	8	9
0	5,012	3,257	2,638	898	1,734	2,642	1,828	599	54	172
1	106	4,179	1,111	5,270	3,116	1,817	−103	673	535	
2	3,410	5,582	4,881	2,268	2,594	3,479	649	603		
3	5,655	5,900	4,211	5,500	2,159	2,658	984			
4	1,092	8,473	6,271	6,333	3,786	225				
5	1,513	4,932	5,257	1,233	2,917					
6	557	3,463	6,926	1,368						
7	1,351	5,596	6,165							
8	3,133	2,262								
9	2,063									

TABLE 2

STATISTICAL SIGNIFICANCE OF FACTORS

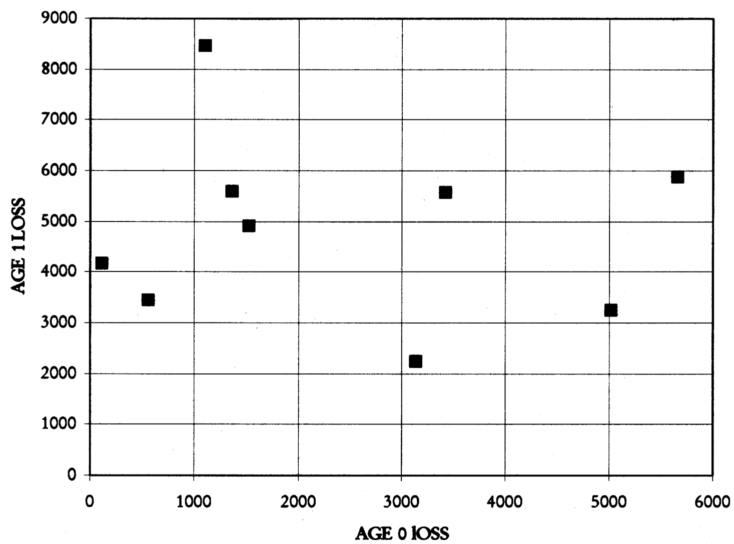
	0 to 1	1 to 2	2 to 3	3 to 4	4 to 5	5 to 6	6 to 7	7 to 8
a	5,113	4,311	1,687	2,061	4,064	620	777	3,724
Std. Dev. a	1,066	2,440	3,543	1,165	2,242	2,301	145	0.000
b	−0.109	0.049	0.131	0.041	−0.100	0.011	−0.008	−0.197
Std. Dev. b	0.349	0.309	0.283	0.071	0.114	0.112	0.008	0.000

nificance of the factors was tested by regression of incremental losses against the previous cumulative losses. In the regression the constant is denoted by a and the factor by b. This provides a test of implication 1—significance of the factor, and also one test of implication 2—alternative emergence patterns. In this case the alternative emergence patterns tested are factor plus constant and constant with no factor. Here they are being tested by looking at whether or not the factors and the constants are significantly different from zero, rather than by any goodness-of-fit measure.

Table 2 shows the estimated parameters and their standard deviations. As can be seen, the constants are usually statistically

FIGURE 1

AGE 1 VS. AGE 0 LOSSES



significant (parameter nearly double its standard deviation, or more), but the factors never are. The chain ladder assumes the incremental losses are proportional to the previous cumulative, which implies that the factor is significant and the constant is not. The lack of significance of the factors and the significance of many of the constants both suggest that the losses to emerge at any age $d+1$ are not proportional to the cumulative losses through age d. The assumptions underlying the chain ladder model are thus not supported by this data. A constant amount emerging for each age usually appears to be a reasonable estimator, however.

Figure 1 illustrates this. A factor by itself would be a straight line through the origin with slope equal to the development factor, whereas a constant would give a horizontal line at the height of the constant. As an alternative, the parameterized BF model was fit to the triangle. As this is a non-linear model, fitting is a little more involved. A statistical package that includes non-linear regression could ease the estimation. A method of fitting the parameters without such a package will be discussed, followed by an analysis of the resulting fit.

To do the fitting, an iterative method can be used to minimize the sum of the squared residuals, where the (w,d) residual is $[q(w,d) - f(d)h(w)]$. Weighted least squares could also be used if the variances of the residuals are not constant over the triangle. For instance, the variances could be proportional to $f(d)^p h(w)^q$ for some values of p and q, usually 0, 1, or 2, in which case the regression weights would be $1/f(d)^p h(w)^q$.

A starting point for the f's or the h's is needed to begin the iteration. While almost any reasonable values could be used, such as all f's equal to $1/n$, convergence will be faster with values likely to be in the ballpark of the final factors. A natural starting point thus might be the implied $f(d)$'s from the chain ladder method. For ages greater than 0, these are the incremental age-to-age factors divided by the cumulative-to-ultimate factors. To get a starting value for age 0, subtract the sum of the other factors from unity. Starting with these values for $f(d)$, regressions were performed to find the $h(w)$'s that minimize the sum of squared residuals (one regression for each w). These give the best h's for that initial set of f's. The standard linear regression formula for these h's simplifies to:

$$h(w) = \sum_d f(d)q(w,d) \Big/ \sum_d f(d)^2.$$

Even though that gives the best h's for those f's, another regression is needed to find the best f's for those h's. For this step the usual regression formula gives:

$$f(d) = \sum_w h(w)q(w,d) \Big/ \sum_w h(w)^2.$$

TABLE 3

BF PARAMETERS

Age d	0	1	2	3	4	5	6	7	8	9
$f(d)$ 1st	0.106	0.231	0.209	0.155	0.117	0.083	0.038	0.032	0.018	0.011
$f(d)$ ult.	0.162	0.197	0.204	0.147	0.115	0.082	0.037	0.030	0.015	0.009
Year w	0	1	2	3	4	5	6	7	8	9
$h(w)$ 1st	17,401	15,729	23,942	26,365	30,390	19,813	18,592	24,154	14,639	12,733
$h(w)$ ult.	15,982	16,501	23,562	27,269	31,587	20,081	19,032	25,155	13,219	19,413

Now the h regression can be repeated with the new f's, etc. This process continues until convergence occurs, i.e., until the f's and h's no longer change with subsequent iterations. It may be possible that this procedure would converge to a local rather than the global minimum, which can be tested by using other starting values.

Ten iterations were used in this case, but substantial convergence occurred earlier. The first round of f's and h's and those at convergence are in Table 3. Note that the h's are not the final estimates of the ultimate losses, but are used with the estimated factors to estimate future emergence. In this case, in fact, $h(0)$ is less than the emerged to date. As the h's are unique only up to a constant of proportionality, which can be absorbed by the f's, it may improve presentations to set $h(0)$ to the estimated ultimate losses for year 0.

Standard regression assumes each observation q has the same variance, which is to say the variance is proportional to $f(d)^p h(w)^q$, with $p = q = 0$. If $p = q = 1$ the weighted regression formulas become:

$$h(w)^2 = \sum_d [q(w,d)^2/f(d)] \Big/ \sum_d f(d) \qquad \text{and}$$

$$f(d)^2 = \sum_w [q(w,d)^2/h(w)] \Big/ \sum_w h(w).$$

TABLE 4

DEVELOPMENT FACTORS

	Incremental								
Prior	0 to 1	1 to 2	2 to 3	3 to 4	4 to 5	5 to 6	6 to 7	7 to 8	8 to 9
	1.22	0.57	0.26	0.16	0.10	0.04	0.03	0.02	0.01
	Ultimate								
	0 to 9	1 to 9	2 to 9	3 to 9	4 to 9	5 to 9	6 to 9	7 to 9	8 to 9
	6.17	2.78	1.77	1.41	1.21	1.10	1.06	1.03	1.01
	Incremental/Ultimate								
0.162	0.197	0.204	0.147	0.115	0.082	0.037	0.030	0.015	0.009

For comparison, the development factors from the chain ladder are shown in Table 4. The incremental factors are the ratios of incremental to previous cumulative. The ultimate ratios are cumulative to ultimate. Below them are the ratios of these ratios, which represent the portion of ultimate losses to emerge in each period. The zeroth period shown is unity less the sum of the other ratios. These factors were the initial iteration for the $f(d)$s shown above.

Having now estimated the BF parameters, how can they be used to test what the emergence pattern of the losses is?

A comparison of this fit to that from the chain ladder can be made by looking at how well each method predicts the incremental losses for each age after the initial one. The SSE adjusted for number of parameters will be used as the comparison measure, where the parameter adjustment will be made by dividing the SSE by the square of the difference between the number of observations and the number of parameters, as discussed earlier. Here there are 45 observations, as only the predicted points count as observations. The adjusted SSE was 81,169 for the BF, and 157,902 for the chain ladder. This shows that the emergence pattern for the BF (emergence proportional to ultimate) is much more consistent with this data than is the chain ladder emergence pattern (emergence proportional to previous emerged).

TABLE 5

FACTORS IN CC METHOD

Age d	0	1	2	3	4	5	6	7	8	9
$f(d)$	0.109	0.220	0.213	0.148	0.124	0.098	0.038	0.028	0.013	0.008

The CC method was also tried for this data. The iteration proceeded similarly to that for the BF, but only a single h parameter was fit for all accident years. Now:

$$h = \sum_{w,d} f(d)q(w,d) \Big/ \sum_{w,d} f(d)^2.$$

This formula for h is the same as the formula for $h(w)$ except the sum is taken over all w. The estimated h is 22,001, and the final factors f are shown in Table 5. The adjusted SSE for this fit is 75,409. Since the CC is a special case of the BF, the unadjusted SSE is necessarily worse than that of the BF method (in this case 59M vs. 98M), but with fewer parameters in the CC, the adjustment makes them similar. These are close enough that which is better depends on the adjustment chosen for extra parameters. The BIC also favors the CC, but the AIC is better for the BF. As is often the case, the statistics can inform decision-making but not determine the decision.

Intermediate special cases could be fit similarly. If, for instance, a single factor were sought to apply to just two accident years, the sum would be taken over those years to estimate that factor, etc.

This is a case where the BF has too many parameters for prediction purposes. More parameters fit the data better but use up information. The penalty in the fit measure adjusts for this problem, and the penalty used finds the CC to be a somewhat better model. Thus the data is consistent with random emergence around an expected value that is constant over the accident years.

TABLE 6

TERMS IN ADDITIVE CHAIN LADDER

Age d	1	2	3	4	5	6	7	8	9
$g(d)$	4,849.3	4,682.5	3,267.1	2,717.7	2,164.2	839.5	625.0	294.5	172.0

Again, the CC method would probably work even better for loss ratio triangles than for loss triangles, as then a single target ultimate value makes more sense. Adjusting loss ratios for trend and rate level could increase this homogeneity.

In addition, an additive development was tried, as suggested by the fact that the constant terms were significant in the original chain ladder, even though the factors were not. The development terms are shown in Table 6. These are just the average loss emerged at each age. The adjusted sum of squared residuals is 75,409. This is much better than the chain ladder, which might be expected, as the constant terms were significant in the original significance-test regressions while the factors were not. The additive factors in Table 6 differ from those in Table 2 because there is no multiplicative factor in Table 6.

Is it a coincidence that the additive chain ladder gives the same fit accuracy as the CC? Not really, in that they both estimate each age's loss levels with a single value. Let $g(d)$ denote the additive development amount for age d. As the notation suggests, this does not vary by accident year. The CC method fits an overall h and a factor $f(d)$ for each age such that the estimated emergence for age d is $f(d)h$. Here too the predicted development varies by age but is a constant for each accident year. If you have estimated the CC parameters you can just define $g(d) = f(d)h$. Alternatively, if the additive method has been fit, no matter what h is estimated, the f's can be defined as $f(d)h = g(d)$. As long as the parameters are fit by least-squares they have to come out the same: if one came out lower, you could have used the equations in the two previous sentences to get this same lower value for

TABLE 7

BF-CC PARAMETERS

Age d	0	1	2	3	4	5	6	7	8	9
$f(d)$	*	0.230	0.230	0.160	0.123	0.086	0.040	0.040	0.017	0.017
Year w	0	1	2	3	4	5	6	7	8	9
$h(w)$	14,829	14,829	20,962	25,895	30,828	20,000	20,000	20,000	20,000	20,000

the other. The two models have the same age and accident year relationships and so will always come out the same when fit by least-squares. They are defined differently, however, and so other methods of estimating the parameters may come up with separate estimates, as in Stanard [10]. In the remainder of this paper, the models will be used interchangeably.

Finally, an intermediate BF-CC pattern was fit as an example of the possible approaches of this type. In this case ages 1 and 2 are assumed to have the same factor, as are ages 6 and 7 and ages 8 and 9. This reduces the number of f parameters from 9 to 6. The number of accident year parameters was also reduced: years 0 and 1 have a single parameter, as do years 5 through 9. Year 2 has its own parameter, as does year 4, but year 3 is the average of those two. Thus there are 4 accident year parameters, and so 10 parameters in total. Any one of these can be set arbitrarily, with the remainder adjusted by a factor, so there are really just 9. The selections were based on consideration of which parameters were likely not to be significantly different from each other.

The estimated factors are shown in Table 7. The factor to be set arbitrarily was the accident year factor for the last 5 years, which was set to 20,000. The other factors were estimated by the same iterative regression procedure as for the BF, but the factor constraints change the simplified regression formula. The adjusted sum of squared residuals is 52,360, which makes it the best approach tried. This further supports the idea that claims emerge as a percent of ultimate for this data. It also indicates that the various accident years and ages are not all at different levels. The actual and fitted values from this, the chain ladder, and CC are in Exhibit 1. The fitted values in Exhibit 1 were calculated as follows. For the chain ladder, the factors from Table 4 were applied to the cumulative losses implied from Table 1. For the CC the fitted values are just the terms in Table 6. For the BF-CC they are the products of the appropriate f and h factors from Table 7. The parameters for all the models to this point are summarized in Exhibit 2.

Alternative Emergence Patterns-Summary

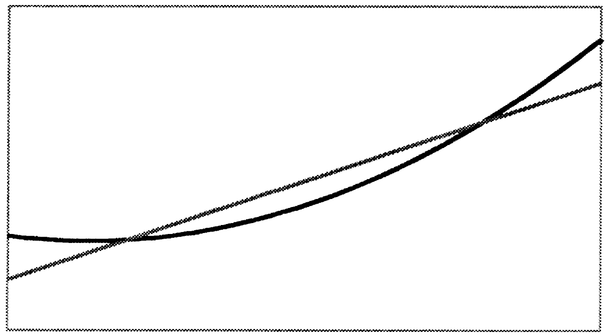
The chain ladder assumes that future emergence for an accident year will be proportional to losses emerged to date. The BF methods take expected emergence in each period to be a percentage of ultimate losses. This could be interpreted as regarding the emerged to date to have a random component that will not influence future development. If this is the actual emergence pattern, the chain ladder method will apply factors to the random component, and thus increase the estimation error.

The CC and additive chain ladder methods assume in effect that years showing low losses or high losses to date will have the same expected future dollar development. Thus a bad loss year may differ from a good one in just a couple of emergence periods, and have quite comparable loss emergence in all other periods. The chain ladder and the most general form of the BF, on the other hand, assume that a bad year will have higher emergence than a good year in most periods.

The BF and chain ladder emergence patterns are not the only ones that make sense. Some others will be reviewed when discussing diagonal effects below.

Which emergence pattern holds for a given triangle is an empirical issue. Fitting parameters to the various methods and looking at the significance of the parameters and the adjusted sum of squared residuals can test this.

FIGURE 2



RESIDUAL ANALYSIS—TESTING IMPLICATIONS 3 & 4

So far the first two of the six testable implications of the chain ladder assumptions have been addressed. Looking at the residuals from the fitting process can test the next two implications.

Implication 3: Test of Linearity—Residuals as Function of Previous

Figure 2 shows a straight line fit to a curve. The residuals can be seen to be first positive, then negative then all positive. This pattern of residuals is indicative of a non-linear process with a linear fit. The chain ladder model assumes the incremental losses at each age are a linear function of the previous cumulative losses.

A scatter plot of the incremental against the previous cumulative, as in Figure 3, can be used to check linearity; looking for this characteristic non-linear pattern (i.e., strings of positive and negative residuals) in the residuals plotted against the previous cumulative is equivalent. This can be tested for each age to see if a non-linear process may be indicated. Finding this would suggest that emergence is a non-linear function of losses to date. In

FIGURE 3

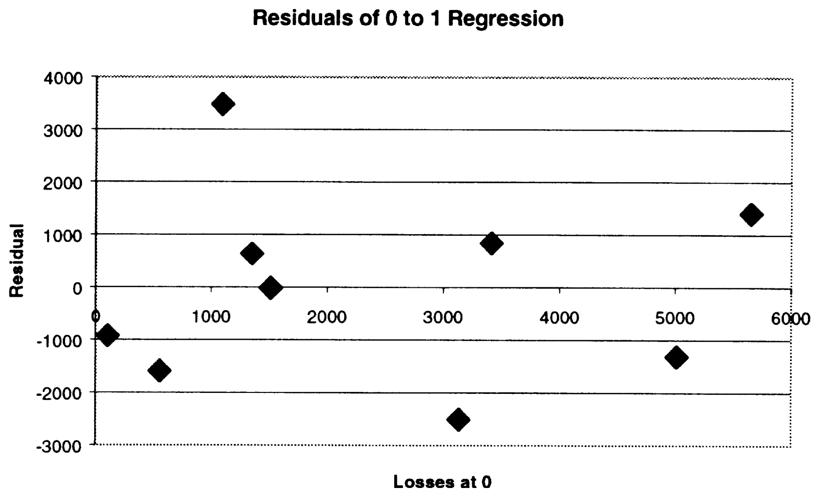


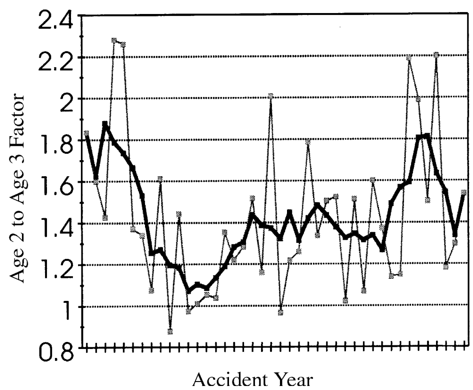
Figure 3 there are no apparent strings of consecutive positive or negative residuals, so non-linearity is not indicated.

Implication 4: Test of Stability—Residuals Over Time

If a similar pattern of sequences of high and low residuals is found when plotted against time, instability of the factors may be indicated. If the factors appear to be stable over time, all the accident years available should be used to calculate the development factors, in order to reduce the effects of random fluctuations. When the development process is unstable, the assumptions for optimality of the chain ladder are no longer satisfied. A response to unstable factors over time might be to use a weighted average of the available factors, with more weight going to the more recent years, e.g., just use the last 5 diagonals. A weighted average should be used when there is a good reason for it, e.g., when residual analysis shows that the factors are changing, but otherwise it will increase estimation errors by over-emphasizing some observations and under-emphasizing others.

FIGURE 4

2ND TO 3RD FIVE-TERM MOVING AVERAGE



Another approach to unstable development would be to adjust the triangle for measurable instability. For instance, Berquist and Sherman [1] suggest testing for instability by looking for changes in the settlement rate of claims. They measured this by looking at the changes in the percentage of claims closed by age. If instability is found, the triangle is adjusted to the latest pattern. The adjusted triangle, however, should still be tested for stability of development factors by residual analysis and as illustrated below.

Figure 4 shows the 2nd to 3rd factor by accident year from a large development triangle (data in Exhibit 3) along with its five-term moving average. The moving average is the more stable of the two lines, and is sometimes in practice called "the average of the last five diagonals." There is apparent movement of the factor over time as well as a good deal of random fluctuation. There is a period of time in which the moving average is as low as 1.1 and other times it is as high as 1.8. This is the kind of variability that would suggest using the average of recent diagonals instead of the entire triangle when estimating factors. This is not suggested due to the large fluctuations in factors, but rather because of the changes over time in the level around which the factors are fluctuating. A lot of variability around a fixed level would in fact suggest using all the data.

It is not clear from the data what is causing the moving average factors to drift over time. Faced with data like this, the average of all the data would not normally be used. Grouping accident years or taking weighted averages would be useful alternatives.

The state-space model in the Verall and Zehnwirth references provides a formal statistical treatment of the types of instability in a data triangle. This model can be used to help analyze whether to use all the data, or to adopt some form of weighted average that de-emphasizes older data. It is based on comparing the degree of instability of observations around the current mean to the degree of instability in the mean itself over time. While this is the main statistical model available to determine weights to apply to the various accident years of data, a detailed discussion is beyond the scope of this paper.

INDEPENDENCE—TESTING IMPLICATIONS 5 & 6

Implications 5 and 6 have to do with independence within the triangle. Mack's second assumption above is that, except for observations in the same accident year, the columns of incremental losses need to be independent. He developed a correlation test and a high-low diagonal test to check for dependencies. The data may have already been adjusted for known changes in the case reserving process. For instance, Berquist and Sherman recommend looking at the difference between paid and incurred case severity trends to determine if there has been a change in case reserve adequacy, and if there has, adjusting the data accordingly. Even after such adjustments, however, correlations may exist within the triangle.

TABLE 8

SAMPLE CORRELATION = $-1.35/(146.37 \times 0.20)^{1/2} = -.25$

Year	X = 0 to 1	Y = 1 to 2	$(X - \mathrm{E}[X])^2$	$(Y - \mathrm{E}[Y])^2$	$(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])$
1	0.65	0.32	54.27	0.14	2.78
2	39.42	0.26	986.46	0.19	−13.71
3	1.64	0.54	40.70	0.02	0.98
4	1.04	0.36	48.63	0.11	2.31
5	7.76	0.66	0.07	0.00	0.01
6	3.26	0.82	22.63	0.01	−0.57
7	6.22	1.72	3.24	1.05	−1.85
8	4.14	0.89	15.01	0.04	−0.74
Average	8.02	0.70	146.37	0.20	−1.35

Implication 5: Correlation of Development Factors

Mack developed a correlation test for adjacent columns of a development factor triangle. If a year of high emergence tends to follow one with low emergence, then the development method should take this into account. Another correlation test would be to calculate the sample correlation coefficients for all pairs of columns in the triangle, and then see how many of these are statistically significant, say at the 10% level. The sample correlation for two columns is just the sample covariance divided by the product of the sample standard deviations for the first n elements of both columns, where n is the length of the shorter column. The sample correlation calculation in Table 8 shows that for the triangle in Table 1 above, the correlation of the first two development factors is −25%.

Letting r denote the sample correlation coefficient, define $T = r[(n-2)/(1-r^2)]^{1/2}$. A significance test for the correlation coefficient can be made by considering T to be t-distributed with $n-2$ degrees of freedom. If T is greater than the t-statistic for 0.9 at $n-2$ degrees of freedom, for instance, then r can be considered significant at the 10% level. (See Miller and Wichern [7, p. 214].)

In this example, $T = -0.63$, which is not significant even at the 10% level. This level of significance means that 10% of the pairs of columns could show up as significant just by random happenstance. A single correlation at this level would thus not be a strong indicator of correlation within the triangle. If several columns are correlated at the 10% level, however, there may be a correlation problem.

To test this further, if m is the number of pairs of columns in the triangle, the number that display significant correlation could be considered a binomial variate in m and 0.1, which has standard deviation $0.3m^{1/2}$. Thus more than $0.1m + m^{1/2}$ significant correlations (mean plus 3.33 standard deviations) would strongly suggest there is actual correlation within the triangle. Here the 10% level and 3.33 standard deviations were chosen for illustration. A single correlation that is significant at the 0.1% level would also be indicative of a correlation problem, for example.

If there is such correlation, the product of development factors is not unbiased, but the relationship $\mathrm{E}[XY] = (\mathrm{E}[X])(\mathrm{E}[Y]) + \mathrm{Cov}(X,Y)$ could be used to correct the product, where here X and Y are development factors.

Implication 6: Significantly High or Low Diagonals

Mack's high-low diagonal test counts the number of high and low factors on each diagonal, and tests whether or not that is likely to be due to chance. Here another high-low test is proposed: use regression to see if any diagonal dummy variables are significant. This test also provides alternatives in case the pure chain ladder is rejected. An actuary will often have information about changes in company operations that may have created a diagonal effect. If so, this information could lead to choices of modeling methods—e.g., whether to assume the effect is permanent or temporary. The diagonal dummies can be used to measure the effect in any case, but knowledge of company operations will help determine how to use this effect. This is particularly so if the effect occurs in the last few diagonals.

A diagonal in the loss development triangle is defined by $w + d$ = constant. Suppose for some given data triangle, the diagonal $w + d = 7$ has been estimated to be 10% higher than normal. Then an adjusted BF estimate of a cell might be:

$$q(w,d) = 1.1f(d)h(w) \qquad \text{if} \quad w + d = 7, \qquad \text{and}$$

$$q(w,d) = f(d)h(w) \qquad \text{otherwise.}$$

This is an example of a multiplicative diagonal effect. Additive diagonal effects can also be estimated, using regression with diagonal dummies.

	Age			
Year	0	1	2	3
1	2	5	4	
3	8	9		
7	10			
7				

Incr. Ages 1–3	Cum. Age 0	Cum. Age 1	Cum. Age 2	Dummy 1	Dummy 2
2	1	0	0	0	0
8	3	0	0	1	0
10	7	0	0	0	1
5	0	3	0	1	0
9	0	11	0	0	1
4	0	0	8	0	1

The small sample triangle of incremental losses here will be used as an example of how to set up diagonal dummies in a chain ladder model. The goal is to get a matrix of data in the form needed to do a multiple regression. First the triangle (except the first column) is strung out into a column vector. This is the dependent variable, and forms the first column of the matrix above. Then columns for the independent variables are added. The second column is the cumulative losses at age 0 corresponding to the loss entries that are at age 1, and zero for the other loss entries. The regression coefficient for this column would be the 0 to 1 cumulative-to-incremental factor. The next two columns are cumulative losses at age 1 and age 2 corresponding to the age 2 and age 3 data in the first column. The last two columns are the diagonal dummies. They pick out the elements of the last two diagonals. The coefficients for these columns would be additive adjustments for those diagonals, if significant.

This method of testing for diagonal effects is applicable to many of the emergence models. In fact, if diagonal effects are found to be significant in chain ladder models, they probably are needed in the BF models of the same data. Thus tests of the chain ladder vs. BF should be done with the diagonal elements included. Some examples are given in the Appendix. Another popular modeling approach is to consider diagonal effects to be a measure of inflation (e.g., see Taylor [11]). In a payment triangle this would be a natural interpretation, but a similar phenomenon could occur in an incurred triangle. In this case the latest diagonal effects might be projected ahead as estimates of future inflation. An understanding of the aspects of company operations that drive the diagonal effects would help address these issues.

This approach incorporates diagonal effects right into the emergence model. For instance, an emergence model might be:

$$E[q(w,d+1) \mid \text{data to } w+d] = f(d)g(w+d).$$

Here $g(w+d)$ is a diagonal effect, but every diagonal has such a factor. The usual interpretation is that g measures the cumulative claims inflation applicable to that diagonal since the first accident year. It would even be possible to add accident year effects $h(w)$ as well, e.g.,

$$E[q(w,d+1) \mid \text{data to } w+d] = f(d)h(w)g(w+d).$$

There are clearly too many parameters here, but a lot of them might reasonably be set equal. For instance, the inflation might be the same for several years, or several accident years might be at the same level. Note that since g is cumulative inflation, a constant inflation level could be achieved by setting $g(w+d) = (1+j)^{w+d}$. Then j is the only inflation parameter to be estimated.

The age and accident year parameters might also be able to be written as trends rather than individual factors. If $f(d) = (1+i)^d$ and $h(w) = h \times (1+k)^w$, then the model reduces to four parameters h, i, j, and k. However it would be more usual to need more parameters than this, possibly written as changing trends. That is, i, j, and k might be constant for some periods, then change for others. Note that if they are constant for all periods, the estimator $h(1+i)^d(1+j)^{w+d}(1+k)^w$ is $h(1+i+j+ij)^d(1+k+j+jk)^w$, which eliminates the parameter j, as i becomes $i+j+ij$ and k becomes $k+j+jk$.

It might be better to start without the accident year trend and keep the calendar year trend, especially if the triangle has been normalized for accident year changes. The model for the (w,d) cell would then be $h(1+i)^d(i+j)^{w+d}$, which has just three parameters.

As with the BF model, the parameters of models with diagonal trends can be estimated iteratively. With reasonable starting values, fix two of the three sets of parameters, and fit the third by least squares, and rotate until convergence is reached. Alternatively, a non-linear search procedure could be utilized. As an example of the simplest of these approaches, modeling $E[q(w,d+1) \mid \text{data to } w+d]$ as just $6{,}756(0.7785)^d$ gives an adjusted sum of squares of 57,527 for the reinsurance triangle above. This is not the best fitting model, but it is better than some and has only two parameters $h = 6{,}756$ and $i = -0.2215$.

Calendar year trend accounts for inflation in the time between loss occurrence and loss settlement, which many actuaries believe has an impact on ultimate losses. Whether it is influencing a given loss triangle can be investigated by testing for diagonal effects.

CONCLUSION

The first test that will quickly indicate the general type of emergence pattern faced is the test of significance of the cumulative-to-incremental factors at each age. This is equivalent to testing if the cumulative-to-cumulative factors are significantly different from unity. When this test fails, the future emergence is not proportional to past emergence. It may be a constant amount, or it may be proportional to ultimate losses, as in the BF pattern.

When this test is passed, the addition of an additive component may give an even better fit. Even when the test is failed, including an additive term may make the factor significant. In either case the BF emergence pattern may still produce a better fit. Reduced parameter BF models could also give better performance, as they will be less responsive to random variation. If an additive component is significant, then converting the triangle to on-level loss ratios may improve the forecasts.

Tests of stability and for diagonal effects may lead to further improvements in the model. However, if the emergence is stable, excluding data by using only the last n diagonals will lead to higher estimation errors on average.

An actuary might advise: "If the chain ladder doesn't work, try Bornhuetter-Ferguson." This is a reasonable conclusion, with the interpretation of "doesn't work" to mean "fails the assumptions of least-squares optimality," and "try" to mean "test the underlying assumptions of."

REFERENCES

[1] Berquist, James R., and Richard E. Sherman, "Loss Reserve Adequacy Testing: A Comprehensive, Systematic Approach," *PCAS* LXIV, 1977, pp. 123–184.

[2] Bornhuetter, Ronald L., and R. E. Ferguson, "The Actuary and IBNR," *PCAS* LIX, 1972, pp. 181–195.

[3] Gerber and Jones, "Credibility Formulas with Geometric Weights," *Society of Actuaries Transactions*, 1975.

[4] de Jong, P., and B. Zehnwirth, "Claims Reserving, State-space Models and the Kalman Filter," *Journal of the Institute of Actuaries*, 1983, pp. 157–181.

[5] Lütkepohl, *Introduction to Multiple Time Series Analysis*, Springer-Verlag, 1993.

[6] Mack, Thomas, "Measuring the Variability of Chain Ladder Reserve Estimates," Casualty Actuarial Society *Forum* 1, Spring 1994, pp. 101–182.

[7] Miller, Robert, and Dean Wichern, *Intermediate Business Statistics*, Holt, Rinehart and Winston, 1977.

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

[9] Popper, Karl R., *Conjectures and Refutations*, Routledge, 1969.

[10] Stanard, James N., "A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques," *PCAS* LXXII, 1985, pp. 124–148.

[11] Taylor, Greg C., "Separation of Inflation and Other Effects from the Distribution of Non-Life Insurance Claim Delays," *ASTIN Bulletin* 9, 1977, pp. 219–230.

[12] Verrall, Richard, "A State Space Representation of the Chain Ladder Linear Model," *Journal of the Institute of Actuaries* 116, Part III, December 1989, pp. 589–610.

[13] Zehnwirth, Ben, "A Linear Filtering Theory Approach to Recursive Credibility Estimation," *ASTIN Bulletin* 15, April 1985, pp. 19–36.

[14] Zehnwirth, Ben, "Probabilistic Development Factor Models with Applications to Loss Reserve Variability, Prediction Intervals, and Risk Based Capital," Casualty Actuarial Society *Forum* 2, Spring 1994, pp. 447–606.

EXHIBIT 1

COMPARATIVE FITS

Chain Ladder

	1	2	3	4	5	6	7	8	9
Actual	3,257	2,638	898	1,734	2,642	1,828	599	54	172
Fit	6,101	4,705	2,846	1,912	1,350	656	580	296	172
% Error	87%	78%	217%	10%	−49%	−64%	−3%	448%	0%
Actual	4,179	1,111	5,270	3,116	1,817	−103	673	535	
Fit	129	2,438	1,408	1,728	1,374	632	499	257	
% Error	−97%	119%	−73%	−45%	−24%	−714%	−26%	−52%	
Actual	5,582	4,881	2,268	2,594	3,479	649	603		
Fit	4,151	5,116	3,619	2,614	1,868	900	736		
% Error	−26%	5%	60%	1%	−46%	39%	22%		
Actual	5,900	4,211	5,500	2,159	2,658	984			
Fit	6,883	6,574	4,113	3,444	2,336	1,057			
% Error	17%	56%	−25%	60%	−12%	7%			
Actual	8,473	6,271	6,333	3,786	225				
Fit	1,329	5,442	4,131	3,591	2,588				
% Error	−84%	−13%	−35%	−5%	1,050%				
Actual	4,932	5,257	1,233	2,917					
Fit	1,842	3,667	3,053	2,095					
% Error	−63%	−30%	148%	−28%					
Actual	3,463	6,926	1,368						
Fit	678	2,287	2,856						
% Error	−80%	−67%	109%						
Actual	5,596	6,165							
Fit	1,644	3,953							
% Error	−71%	−36%							
Actual	2,262								
Fit	3,814								
% Error	69%								

CC

	1	2	3	4	5	6	7	8	9
Actual	3,257	2,638	898	1,734	2,642	1,828	599	54	172
Fit	4,364	3,746	2,287	1,631	1,082	336	188	59	17
% Error	34%	42%	155%	−6%	−59%	−82%	−69%	9%	−90%
Actual	4,179	1,111	5,270	3,116	1,817	−103	673	535	
Fit	4,364	3,746	2,287	1,631	1,082	336	188	59	
% Error	4%	237%	−57%	−48%	−40%	−426%	−72%	−89%	
Actual	5,582	4,881	2,268	2,594	3,479	649	603		
Fit	4,364	3,746	2,287	1,631	1,082	336	188		
% Error	−22%	−23%	1%	−37%	−69%	−48%	−69%		
Actual	5,900	4,211	5,500	2,159	2,658	984			
Fit	4,364	3,746	2,287	1,631	1,082	336			
% Error	−26%	−11%	−58%	−24%	−59%	−66%			
Actual	8,473	6,271	6,333	3,786	225				

EXHIBIT 1

(CONTINUED)

Fit	4,364	3,746	2,287	1,631	1,082				
% Error	−48%	−40%	−64%	−57%	381%				
Actual	4,932	5,257	1,233	2,917					
Fit	4,364	3,746	2,287	1,631					
% Error	−12%	−29%	85%	−44%					
Actual	3,463	6,926	1,368						
Fit	4,364	3,746	2,287						
% Error	26%	−46%	67%						
Actual	5,596	6,165							
Fit	4,364	3,746							
% Error	−22%	−39%							
Actual	2,262								
Fit	4,364								
% Error	93%								
BF-CC									
	1	2	3	4	5	6	7	8	9
Actual	3,257	2,638	898	1,734	2,642	1,828	599	54	172
Fit	3,411	3,411	2,373	1,824	1,275	593	593	252	252
% Error	5%	29%	164%	5%	−52%	−68%	−1%	367%	47%
Actual	4,179	1,111	5,270	3,116	1,817	−103	673	535	
Fit	3,411	3,411	2,373	1,824	1,275	593	593	252	
% Error	−18%	207%	−55%	−41%	−30%	−676%	−12%	−53%	
Actual	5,582	4,881	2,268	2,594	3,479	649	603		
Fit	4,821	4,821	3,354	2,578	1,803	838	838		
% Error	−14%	−1%	48%	−1%	−48%	29%	39%		
Actual	5,900	4,211	5,500	2,159	2,658	984			
Fit	5,956	5,956	4,143	3,185	2,227	1,036			
% Error	1%	41%	−25%	48%	−16%	5%			
Actual	8,473	6,271	6,333	3,786	225				
Fit	7,090	7,090	4,932	3,792	2,651				
% Error	−16%	13%	−22%	0%	1,078%				
Actual	4,932	5,257	1,233	2,917					
Fit	4,600	4,600	3,200	2,460					
% Error	−7%	−12%	160%	−16%					
Actual	3,463	6,926	1,368						
Fit	4,600	4,600	3,200						
% Error	33%	−34%	134%						
Actual	5,596	6,165							
Fit	4,600	4,600							
% Error	−18%	−25%							
Actual	2,262								
Fit	4,600								
% Error	103%								

EXHIBIT 1

(CONTINUED)

Additive with Multiplicative Diagonals and Accident Years

	1	2	3	4	5	6	7	8	9
Actual	3,257	2,638	898	1,734	2,642	1,828	599	54	172
Fit	3,185	3,185	2,148	2,730	1,995	660	660	660	477
% Error	−2%	21%	139%	57%	−24%	−64%	10%	1,122%	177%
Actual	4,179	1,111	5,270	3,116	1,817	−103	673	535	
Fit	3,185	3,185	3,465	2,730	1,995	660	660	477	
% Error	−24%	187%	−34%	−12%	10%	−741%	−2%	−11%	
Actual	5,582	4,881	2,268	2,594	3,479	649	603		
Fit	4,036	6,508	4,390	3,460	2,529	836	604		
% Error	−28%	33%	94%	33%	−27%	29%	0%		
Actual	5,900	4,211	5,500	2,159	2,658	984			
Fit	6,508	6,508	4,390	3,460	2,529	604			
% Error	10%	55%	−20%	60%	−5%	−39%			
Actual	8,473	6,271	6,333	3,786	225				
Fit	5,136	5,136	3,465	2,730	1,442				
% Error	−39%	−18%	−45%	−28%	541%				
Actual	4,932	5,257	1,233	2,917					
Fit	5,136	5,136	3,465	1,972					
% Error	4%	−2%	181%	−32%					
Actual	3,463	6,926	1,368						
Fit	5,136	5,136	2,503						
% Error	48%	−26%	83%						
Actual	5,596	6,165							
Fit	5,136	3,710							
% Error	−8%	−40%							
Actual	2,262								
Fit	3,710								
% Error	64%								

EXHIBIT 2

SUMMARY OF PARAMETERS

	0	1	2	3	4	5	6	7	8	9
BF $f(d)$	0.162	0.197	0.204	0.147	0.115	0.082	0.037	0.030	0.015	0.009
BF $h(w)$	15,982	16,501	23,562	27,269	31,587	20,081	19,032	25,155	13,219	19,413
CC $f(d)$	0.109	0.220	0.213	0.148	0.124	0.098	0.038	0.028	0.013	0.008
Additive Chain	—	4,849.3	4,682.5	3,267.1	2,717.7	2,164.2	839.5	625.0	294.5	172.0
BF-CC $f(d)$	—	0.230	0.230	0.160	0.123	0.086	0.040	0.040	0.017	0.017
BF-CC $h(w)$	14,829	14,829	20,962	25,895	30,828	20,000	20,000	20,000	20,000	20,000

EXHIBIT 3

2ND TO 3RD FACTORS FROM LARGE TRIANGLE

2nd to 3rd→		1.81	1.60	1.41	2.29	2.25	1.38
	1.36	1.07	1.60	0.89	1.42	0.99	1.01
	1.03	1.02	1.35	1.21	1.28	1.51	1.17
	2.00	0.98	1.21	1.24	1.79	1.32	1.48
	1.51	1.01	1.51	1.06	1.60	1.10	1.11
	2.20	2.00	1.50	2.20	1.19	1.28	1.52

APPENDIX

DIAGONAL EFFECTS IN BF MODELS

As an example, a test for diagonal effects in the CC model was made in the reinsurance triangle as follows. The CC is the same as the additive chain ladder, so it can be expressed as a linear model. This can be estimated via a single multiple regression in which the dependent variable is the entire list of incremental losses for ages 1 to 9 and all accident years—45 items in all. That is, the triangle beyond age 0 is strung out into a single vector. Age and diagonal dummy independent variables can be established in a design matrix to pick out the right elements of the parameter vector of age and diagonal terms to estimate each incremental loss cell. For the additive chain ladder, the column dummy variables will be 1 or 0, as opposed to cumulative losses or 0 in the chain ladder example. Then the coefficient of that column will be the additive element for the given age.

The later columns of the design matrix would be diagonal dummies, as in the chain ladder example. By doing a multiple linear regression for the incremental loss column in terms of the age and diagonal dummies, additive terms by age and by diagonal will be estimated. The regression can tell which terms are statistically significant, and the others can be dropped from the specification.

With the reinsurance triangle tested above, the first three diagonals turned out to be lower than the others, as was the last diagonal. Also, the first two ages were not significantly different from each other, nor were the last four. This produced a model with five age parameters and two diagonal parameters—one for the first three diagonals combined, and one for the last diagonal. The parameters are shown in Table 9.

The sum of squared residuals for this model is 49,673.4 when adjusted for seven parameters used. This is considerably better

TABLE 9

TERMS IN ADDITIVE CHAIN LADDER WITH DIAGONAL EFFECTS

Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8	Age 9	Diag 1–3	Diag 9
5,569.0	5,569.0	3,739.2	2,881.8	2,361.1	993.3	993.3	993.3	993.3	−2,319.9	−984.7

than the model without diagonal effects. The multiple regression found the diagonals to be statistically significant and adding them to the model improved the fit.

A problem with the diagonal analysis is how to use them in forecasting. One reason for diagonal effects is a change in company practice, particularly in the claims handling process. If the age effects are considered the dominant influence with occasional distortion by diagonal effects, then including diagonal dummy variables will give better estimates for the underlying age terms. Then these, but not the diagonal effects, would be used in forecasting.

Having identified the significant diagonal effects through linear regression, it may be more reasonable to convert them to multiplicative effects through non-linear regression. The model could be of the form:

$$q(w,d) = f(d)g(w+d),$$

where $f(d)$ is the additive age term for age d, and $g(w+d)$ is the factor for the $w+d$th diagonal. Again this can be estimated iteratively by fixing the f's to estimate the g's by linear regression, then fixing those g's to estimate the next iteration of f's, until convergence is reached. The previous model was refit with the diagonals as factors with the result in Table 10. This had a slightly better adjusted sum of squared residuals of 49,034.8.

Diagonal factors can be used in conjunction with accident year factors as in:

$$q(w,d) = f(d)g(w+d)h(w).$$

TABLE 10

ADDITIVE CHAIN LADDER WITH MULTIPLICATIVE DIAGONAL EFFECTS

Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8	Age 9	Diag 1–3	Diag 9
5,692.3	5,692.3	3,823.0	2,816.1	2,416.7	672.1	672.1	672.1	672.1	.5598	.6684

TABLE 11

ADDITIVE CHAIN LADDER WITH MULTIPLICATIVE DIAGONAL & AY EFFECTS

Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8	Age 9	Diag 1-3	Diag 9	AY 3-4
5,135.6	5,135.6	3,464.7	2,730.1	1,995.4	660.1	660.1	660.1	660.1	.6201	.7225	1.2672

As an example, a factor was added to the above model to represent accident years 3 and 4, and the 4th age term was forced to be the average of the 3rd and 5th. The result is in Table 11.

The adjusted sum of squared residuals came down to 44,700.9, which is considerably better than the previous best-fitting model, and almost twice as good as in the original BF model, which in turn was almost twice as good as the chain ladder. It appears that accident year effects and diagonal effects are significant in this data. The fit is shown as the last section of Exhibit 1. The numerous examples fit to this data were for the sake of illustration. Some models of the types discussed may still fit better than the particular ones shown here.