Simulating Loss Development

The principal task in simulating a company's loss development is identifying the stochastic process that generates that development. Testing different processes against the historical development data is a way to approach this task. The second task is to model how the company's carried reserves respond to the loss emergence scenarios generated. One assumption for this may be that the company knows the process that produces its development, and uses a reserving methodology appropriate for that process. The simulation would proceed by generating loss emergence scenarios stochastically and then applying the selected reserving method to produce the carried reserves for each scenario. On the other hand, if the company has a fixed reserve methodology that it is going to use no matter what, then that methodology can be used to produce the carried reserves from the simulated emergence.

For this discussion, "emergence" could either mean case emergence or paid emergence, or both. The main concern here is simulating the emerging losses by period. This may or may not involve simulating the ultimate losses. For instance, one way to generate the losses to emerge in a period is to multiply simulated ultimate losses times a factor drawn from a percentage emerged distribution. This is appropriate when the process producing the losses for each period works by taking a randomized percent of ultimate losses. This method might involve some quite complicated methods of simulating ultimates, but all those that take period emergence as a percentage of ultimate will be considered to be using the same type of emergence pattern. Several other emergence patterns will be considered below, and the reserving methods appropriate for each will be discussed. Then methods for identifying the emergence patterns from the data triangles will be explored.

Types of Emergence Patterns

Six characteristics of emergence patterns will be considered here. Each will be treated as a binary choice, thus producing 64 types of emergence patterns. However there will be sub-categories within the 64, as not all of the choices are actually binary. The six basic choices for defining loss emergence processes are:

1. **Do the losses that emerge in a period depend on the losses already emerged?** Mack has shown that the chain ladder method assumes an emergence pattern in which the emerged loss for a period is a constant factor times the previous emerged, plus a random disturbance. Other methods, however, might apply factors only to ultimate losses, and then add a random disturbance. The latter is the emergence pattern assumed by the Bornheutter-Ferguson (BF) method, for example.
2. **Is all loss emergence proportional?** Both the chain ladder and BF methods use factors to predict emergence, and so are based on processes where emergence is proportional to something – either ultimate losses in the BF case or previously emerged in the chain ladder.. However, the expected loss emergence for a period could be constant – not proportional to anything. Or it could be a factor times something plus a constant. If this is the emergence pattern used, then the reserving methodology should also incorporate additive elements.
3. **Is emergence independent of calendar year events?** Losses to emerge in a period may depend on the inflation rate for the period. This is an example of a calendar year or diagonal effect. Another example is strong or weak development due to a change in claim handling methods. Thus this is not a purely

binary question – if there are diagonal effects there will be sub-choices relating to what type of effect is included. The Taylor separation method is an example of a development method that recognizes calendar year inflation. In many cases of diagonal effects, the ultimate losses will not be determined until all the development periods have been simulated.

4. **Are the parameters stable?** For instance a parameter might be a loss development factor. A stable factor could lead to variable losses due to randomness of the development pattern, but the factor itself would remain constant. The alternative is that the factor changes over time. There are sub-cases of this, depending on how they change.
5. **Are the disturbance terms generated from a normal distribution?** The typical alternative is lognormal, but the possibilities are endless. Clearly the loss development method will need to respond to this choice.
6. **Are the disturbance terms homoskedastic?** Some regression methods of development assume that the random disturbances all have the same variance, at least by development age. Link ratios are often calculated as the ratio of losses at age j+1 divided by losses at age j, which assumes that the variance of the disturbance term is proportional to the mean loss emerged. Another alternative is for the standard deviation to be proportional to the mean. The variance assumption used to generate the emerging losses can be employed in the loss reserving process as well.

Notation

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.

$c_{w,d}$: cumulative loss from accident year w as of age d
$c_{w,\infty}$: ultimate loss from accident year w
$q_{w,d}$: incremental loss for accident year w to emerge in period d
f_d: factor used in emergence for age d
h_w : factor used in emergence for year w
g_{w+d}: factor used in emergence for calendar year w+d
a_d: additive term used in emergence for age d

Question 1

The stochastic processes specified by answering the six questions above can be numbered in binary by considering yes=1 and no=0. Then process 111111 (all answers yes) can be specified as follows:

$$q_{w,d} = c_{w,d-1} f_d + e_{w,d} \qquad (1)$$

where $e_{w,d}$ is normally distributed with mean zero. Here f_d is a development factor applied to the cumulative losses simulated at age d-1. A starting value for the accident year is needed which could be called $c_{w,-1}$. For each d it might be reasonable to assume that $e_{w,d}$ has a different variance. Note that for this process, ultimate losses are generated only as the sum of the separately generated emerged losses for each age.

Mack has shown that for process 111111 the chain ladder is the optimal reserve estimation method. The factors f_d would be estimated by a no-constant linear regression. In process 111110 (heteroskedastic) the chain ladder would also be optimal, but the method of estimating the factors would be different. Essentially these would use weighted least squares for the estimation, where the weights are inversely proportional to the variance of $e_{w,d}$. If the variances are proportional to $c_{w,d-1}$, the resulting factor is the ratio of the sum of losses from the two relevant columns of the development triangle.

In all the processes 1111xx Mack showed that some form of the chain ladder is the best linear estimate, but when the disturbance term is not normal, linear estimation is not necessarily optimal.

Processes of type 0111xx do not generate emerged losses from those previously emerged. A simple example of this type of process is:

$$q_{w,d} = h_w f_d + e_{w,d} \qquad (2)$$

Here h_w can be interpreted as the ultimate losses for year w, with the factors f_d summing to unity. For this process, reserving would require estimation of the f's and h's. I call this method of reserving the parameterized BF, as Bornheutter and Ferguson estimated emergence as a percentage of expected ultimate. The method of estimating the parameters would depend on the distribution of the disturbance term $e_{w,d}$. If it is normal and homoskedastic, a regression method can be used iteratively by fixing the f's and regressing for the h's, then taking those h's to find the best f's, etc. until both f's and h's converge. If heteroskedastic, weighted regressions would be needed. If a lognormal disturbance is indicated, the parameters could be estimated in logs, which is a linear model in the logs.

Question 2

Additive terms can be added to either of the above processes. Thus an example of a 0011xx process would be:

$$q_{w,d} = a_d + h_w f_d + e_{w,d} \qquad (3)$$

If the f's are zero, this would be a purely additive model. A test for additive effects can be made by adding them to the estimation and seeing if significantly better fits result.

Question 3

Diagonal effects can be added similarly. A 0001xx model might be:

$$q_{w,d} = a_d + h_w f_d g_{w+d} + e_{w,d} \qquad (4)$$

Again this can be tested by goodness of fit. There may be too many parameters here. It will usually be possible to reasonably simulate losses without using so many distinct parameters. Specifying relationships among the parameters can lead to reduced parameter versions of these processes. For instance, some of the parameters might be set equal, such as h_w=h for all w. Note that the 0111xx process $q_{w,d} = hf_d + e_{w,d}$ is the same as the 0011xx process $q_{w,d} = a_d + e_{w,d}$, as a_d can be set to hf_d. The resulting reserve estimation method is an additive version of the chain ladder, and is sometimes called the Cape Cod method.

Another way to reduce the number of parameters is to set up trend relationships. For example, constant calendar year inflation can be specified by setting $g_{w+d}=(1+j)^{w+d}$. Similar trend relationships can be specified among the h's and f's. If that is too much parameter reduction to adequately model a given data triangle, a trend can be established for a few periods and then some other trend can be used in other periods.

Question 4

Rather than trending, the parameters in the loss emergence models could evolve according to some more general stochastic process. This could be a smooth process or one with jumps. The state-space model is often used to describe parameter variability. This model assumes that observations fluctuate around an expected value that itself changes over time as its parameters evolve. The degree of random fluctuation is measured by the variance of the observations around the mean, and the movement of the parameters is quantified by their variances over time. The interplay of these two variances determines the weights to apply, as in credibility theory.

To be more concrete, a formal definition of the model follows where the parameter is the 2nd to 3rd development factor. Let:

β_i=2nd to 3rd factor for ith accident year
y_i=3rd report losses for ith accident year
x_i=2nd report losses for ith accident year

The model is then:

$$y_i=x_i\beta_i+\varepsilon_i. \qquad (5)$$

The error term ε_i is assumed to have mean 0 and variance σ_i^2.

$$\beta_i=\beta_{i-1}+\delta_i. \qquad (6)$$

The fluctuation δ_i is assumed to have mean 0 and variance ν_i^2, and to be independent of the ε's.

In this general case the variances could change with each period i. Usually some simplification is applied, such as constant variances over time, or constant with occasional jumps in the parameter – i.e., occasional large ν_i's.

If this model is adopted for simulating loss emergence, the estimation of the factors from the data can be done using the Kalman filter.

Questions 5 and 6

The error structure can be studied and usually reasonably understood from the data triangles. The loss estimation method associated with a given error structure will be assumed to be maximum likelihood estimation from that structure. Thus for normal distributions this is weighted least squares, where the weights are the inverses of the variances. For lognormal this is the same, but in logs.

Identifying Emergence Patterns

Given a data triangle, what is the process that is generating it? This is useful to know for loss reserving purposes, as then reserve estimation is reduced to estimation of the parameters of the generating process. It is even more critical for simulation of company results, as the whole process is needed for simulation purposes.

Identifying emergence patterns can be approached by fitting different ones to the data and then testing the significance of the parameters and the goodness of fit. As more parameters often appear to give a better fit, but reduce predictive value, a method of penalizing over-parameterization is needed when comparing competing models. The method proposed here is to compare models based on sum of squared residuals divided by the square of the degrees of freedom, i.e., divided by the square of observations less parameters.

This measure gives impetus to trying to reduce the number of parameters in a given model, e.g., by setting some parameters the same or by identifying a trend in the parameters. This seems to be a legitimate exercise in the effort of identifying emergence patterns, as there are likely to be some regularities in the pattern, and simplifying the model is a way to uncover them.

Fitting the above models is a straightforward exercise, but reducing the number of parameters may be more of an art than a science. Two approaches may make sense: top down, where the full model is fit and then regularities among the parameters sought; and bottom up, where the most simplified version is estimated, and then parameters added to compensate for areas of poor fit.

To illustrate this approach, the data triangle of reinsurance loss data first introduced by Thomas Mack will be the basis of model estimation.

Questions 1 & 2 – Factors and Constant Terms

Table 1 shows incremental incurred losses by age for some excess casualty reinsurance. As an initial step, the statistical significance of link ratios and additive constants was tested by regressing 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 question 1 – dependence of emergence on previous emerged, and also one of question 2 – proportional emergence. 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 1 - Incremental Incurred Losses

0	1	2	3	4	5	6	7	8	9
5012	3257	2638	898	1734	2642	1828	599	54	172
106	4179	1111	5270	3116	1817	-103	673	535	
3410	5582	4881	2268	2594	3479	649	603		
5655	5900	4211	5500	2159	2658	984			
1092	8473	6271	6333	3786	225				
1513	4932	5257	1233	2917					
557	3463	6926	1368						
1351	5596	6165							
3133	2262								
2063									

Table 2 - Statistical Significance of Link Ratios and Constants

	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’	5113	4311	1687	2061	4064	620	777	3724
Std a	1066	2440	3543	1165	2242	2301	145	0
\`b’	-0.109	0.049	0.131	0.041	-0.100	0.011	-0.008	-0.197
std b	0.349	0.309	0.283	0.071	0.114	0.112	0.008	0

Table 2 shows the estimated parameters and their standard deviations. As can be seen, the constants are usually statistically significant (parameter nearly double its standard deviation, or more), but the factors never are. The lack of significance of the factors shows 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 met 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.

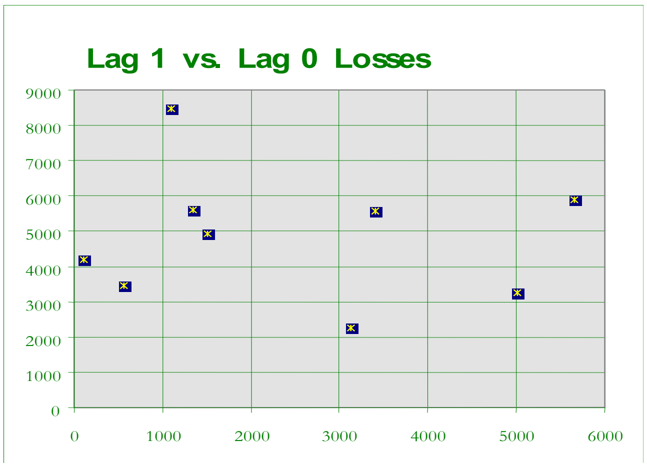

Figure 1

Although emerged losses are not proportional to previous emerged, they could be proportional to ultimate incurred. To test this, the parameterized BF model (2) was fit to the triangle. As this is a non-linear model, fitting is a little more involved. A method of fitting the parameters 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 $f_d{}^p h_w{}^q$, 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:

$$h_w = \sum_d f_d q_{w,d} / \sum_d f_d^2 \qquad (7)$$

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:

$$f_d = \sum_w h_w q_{w,d} / \sum_w h_w^2 \qquad (8)$$

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. Ten iterations were used in this case, but substantial onvergence 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. A statistical package that includes non-linear regression could ease the estimation.

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:

$$h_w^2 = \sum_d[q_{w,d}^2/f_d] \;/\; \sum_d f_d \quad \text{and}$$

$$f_d^2 = \sum_w[q_{w,d}^2/h_w] \;/\; \sum_w h_w$$

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	17401	15729	23942	26365	30390	19813	18592	24154	14639	12733
h_w ult	15982	16501	23562	27269	31587	20081	19032	25155	13219	19413

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.

Table 4 - Development 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	8 to 9
Incremental	1.22	0.57	0.26	0.16	0.10	0.04	0.03	0.02	0.01
	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
Ultimate	6.17	2.78	1.77	1.41	1.21	1.10	1.06	1.03	1.01
0.162	0.197	0.204	0.147	0.115	0.082	0.037	0.030	0.015	0.009

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 sum of squared errors adjusted for number of parameters is the comparison measure, where the parameter adjustment is made by dividing the sum of squared errors by the square of [the number of observations less the number of parameters], as discussed earlier. Here there are 45 observations, as only the predicted points count as observations. The adjusted sum of squared residuals is 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).

The Cape Cod (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:

$$h = \sum_{w,d} f_d q_{w,d} \;/\; \sum_{w,d} f_d^2 \qquad (9)$$

The estimated h is 22,001, and the final factors f are shown in Table 5. The adjusted sum of squared errors for this fit is 75,409. Since the CC is a special case of the BF, the unadjusted fit is of course worse than that of the BF method, but with fewer parameters in the CC, the adjustment makes them similar. This formula for h is the same as the formula for h_w except the sum is taken over all w.

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.

Table 5 - Factors in CC Method

0	1	2	3	4	5	6	7	8	9
0.109	0.220	0.213	0.148	0.124	0.098	0.038	0.028	0.013	0.008

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 penalization in the fit measure adjusts for this problem, and shows 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.

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, a purely 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.

Table 6 - Terms in Additive Chain Ladder

1	2	3	4	5	6	7	8	9
4849.3	4682.5	3267.1	2717.7	2164.2	839.5	625	294.5	172

As discussed above, the additive chain ladder is the same as the Cape Cod method, although it is parameterized differently. The exact same goodness of fit is thus not surprising.

Finally, an intermediate BF-CC pattern was fit as an example of reduced parameter BF's. 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 accident year factor for the last 5 years 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, but that the CC is too much of a simplification. 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.

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	14829	14829	20962	25895	30828	20000	20000	20000	20000	20000

Calendar Year Impacts – Testing Question 3

One type of calendar year impact is high or low diagonals in the loss triangle. Mack suggested a high-low diagonal test which 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. 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 is found to be 10% higher than normal. Then an adjusted BF estimate of a cell might be:

$$q_{w,d}=1.1f_dh_w \text{ if } w+d=7, \text{ and } q_{w,d}=f_dh_w \text{ otherwise} \qquad (10)$$

1	2	5	4
3	8	9	
7	10		
7			

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. Then columns for the independent variables are added. The second column is the cumulative losses at age 0 for 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 the same for the 1 to 2 and 2 to 3 factors. The last two columns are the diagonal dummies. They pick out the elements of the last

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

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 significant in chain ladder models, they probably are needed in the BF models of the same data, so goodness-of-fit tests should be done with those diagonal elements included. Some examples are given in Appendix 2.

Another popular modeling approach is to consider diagonal effects to be a measure of inflation (e.g., see Taylor 1977). 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 what in company operations is driving the diagonal effects would help address these issues.

As with the BF model, the parameters of the model with inflation effects, $q_{w,d} = h_w f_d g_{w+d} + e_{w,d}$, can be estimated iteratively. With reasonable starting values, fix two of the three sets of parameters, 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 models, modeling $q_{w,d}$ as just $6756(0.7785)^d$ gives an adjusted sum of squares of 57,527 for the reinsurance triangle above. This is not the best fitting model, but is better than some, and has only two parameters. Adding more parameters to this would be an example of the bottom up fitting approach.

Testing Question 4 - Stability of parameters

If a pattern of sequences of high and low residuals is found when plotted against time, instability of the parameters may be indicated. This can be studied and a randomness in the parameters incorporated into the simulation process, e.g., through the state-space model.

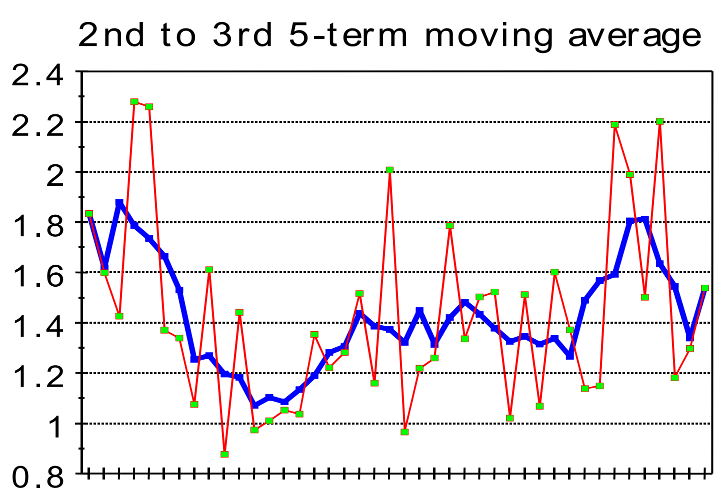

Figure 2

Figure 2 shows the 2nd to 3rd factor by accident year from a large development triangle (data in Exhibit 2) 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 mean factor over time as well as a good deal of random fluctuation around it. 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.

The state-space model assumes that observations fluctuate around a mean that itself changes over time. The degree of random fluctuation is measured by variance around the mean, and the movement of the mean by its variance over time. The interplay of these two variances determines the weights to apply, as in credibility theory.

The state-space model thus provides underlying assumptions about the process by which development changes over time. With such a model, estimation techniques that minimize prediction errors can be developed for the changing development case. This can result in estimators that are better than either using all data, or taking the average of the last few diagonals. An introductory discussion is included in Appendix 1, but for more detail see the Verrall and Zehnwirth references.

Questions 5 & 6: Variance Assumptions

Parameter estimation changes depending on the form of the variance. Usually in the chain ladder model the variance will plausibly be either a constant or proportional to the previous cumulative or its square. Plotting or fitting the squared residuals as a function of the previous cumulative will usually help decide which of these three alternatives fits better. If the squared residuals tend to be larger when the explanatory variable is larger, this is evidence that the variance is larger as well.

Another variance test would be for normality of the residuals. Normality is often tested by plotting the residuals on a normal scale, and looking for linearity. This is not a formal test, but it is often considered a useful procedure. If the residuals are somewhat positively skewed, a lognormal distribution may be reasonable. The non-linear models discussed are all linear in logs, and so could be much easier to estimate in that form. However, if some increments are negative, a lognormal model becomes awkward. The right distribution for the residuals of loss reserving models seems an area in which further research would be helpful.

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, it may be proportional to ultimate losses, as in the BF pattern, or it may depend on future inflation.

The addition of an additive component may give an even better fit. Reduced parameter models could also give better performance, as they will be less responsive to random variation. If an additive component is significant, converting the triangle to on-level loss ratios may improve the model. Tests of stability and for calendar-year effects may lead to further improvements.

Appendix 1— State Space Model

The problem is to find a reasonable estimate for the factor at each point that recognizes changes in the factor but is not overly responsive to random fluctuations. How responsive turns out to be a function of two important variances: the variance of the movement of the factor over time, and the variance of the random fluctuation. A formal definition of the model follows in the case of 2nd to 3rd development.

β_i=2nd to 3rd factor for ith accident year

y_i=3rd report losses for ith accident year

x_i=2nd report losses for ith accident year

$y_i=x_i\beta_i+\varepsilon_i$. The error term ε_i is assumed to have mean 0 and variance σ_i^2.

$\beta_i=\beta_{i-1}+\delta_i$. The fluctuation δ_i is assumed to have mean 0 and variance ν_i^2, and to be independent of the ε's.

What needs to be estimated primarily are the factors β_i. These will be applied when y_{i-1} and x_i have been observed. The following notation will be used for the quantities to be estimated.

$\hat{\beta}_i$= estimate of β_i given $y_1, \ldots, y_{i-1}$ and $x_1, \ldots, x_{i-1}$.

$\hat{y}_i$= forecast of y_i given $y_1, \ldots, y_{i-1}$ and $x_1, \ldots, x_i$.

Γ_i= Variance of $\hat{\beta}_i$ (Γ_1 is a given parameter, later ones are estimated.)

H_i= Variance of $\hat{y}_i$ (Includes process variance σ_i and parameter variance Γ_i.)

The estimation in the state space model is an iterative process that proceeds one period at a time. The process is similar to credibility. From the signal processing background the estimation procedure is called the *Kalman filter* after its originator. The key is the so called *Kalman gain factor*, which in effect is a credibility weight to see how much the new observation should be allowed to change the previous estimate of the development factor. The iterative process at each stage has several steps as follows:

$\hat{y}_i=x_i\hat{\beta}_i$	Next observation forecast
$k_i=\Gamma_i x_i/(x_i^2\Gamma_i+\sigma_i^2)$	Update Kalman gain (credibility) factor
$\Gamma_{i+1}=\Gamma_i+\nu_{i+1}-k_i x_i\Gamma_i$	Update parameter variance - it increases because of fluctuation, but decreases due to more data
$\hat{\beta}_{i+1}=\hat{\beta}_i+k_i(y_i-\hat{y}_i)$	Update parameter estimate, using Kalman gain factor
$H_i=x_i^2\Gamma_i+\sigma_i^2$	Update forecast variance (and thus standard error)

In order to begin this process, several starting parameters are needed that may have to be estimated. They are: Γ_1, $\hat{\beta}_1$, σ_i, ν_i. Some simplifying assumptions can be made to ease this initial start-up.

Simplified State-Space Model

Assumptions that greatly simplify this model are that $\nu_i=\nu$, and $\sigma_i=x_i\sigma$, for each period i. This leads to the result:

$$\hat{\beta}_{i+1}=\hat{\beta}_i+z_i(y_i/x_i-\hat{\beta}_i), \text{ where}$$

$$z_1 = \Gamma_1/(\Gamma_1+\sigma^2) \text{ and}$$
$$z_{i+1} = 1/[1+1/(z_i+J)],\ J=\nu^2/\sigma^2$$

This is the Gerber-Jones credibility model (e.g., see *Credibility Formulas with Geometric Weights* in the 1975 Society of Actuaries Transactions or the credibility chapter of the CAS textbook). In the limit the credibility is $z_\infty=\frac{1}{2}J[(1+4/J)^{1/2}-1]$. In this formula the development factor for an accident year is the previous year's factor plus a correction term, where the correction term is the credibility weight z times the prediction error from the last year. The formula can also be interpreted as an estimation of the development factor as a weighted average of the individual year development ratios, with the weights declining for more remote periods. To see this, let $w_i=z_iy_i/x_i$. Then successive substitution yields:

$$\hat{\beta}_{i+1}=w_i+(1-z_i)w_{i-1}+(1-z_i)(1-z_{i-1})w_{i-2}+..+\hat{\beta}_1\Pi(1-z_j),$$

which is a declining weighted average of the y_j/x_j's. This still needs starting values. Assuming no prior knowledge, you could start with $z_1=1$ and $\hat{\beta}_1=y_1/x_1$. Then just J would be needed to do updates. Rather than estimating its component variances separately, J could be estimated by trying different J's to minimize the SSSSPE= $\Sigma_i[y_i/x_i-\hat{y}_i/x_i]^2$. SSSSPE stands for the sum of squares of single step prediction errors, which is a goodness-of-fit measure for varying parameter models. It quantifies how well the estimation actually performs in separating signal from noise. As an example, the fit with J=.07 is shown in Figure 3. That J yields a z_∞ of 0.23. The fit appears to capture longer term changes without following random fluctuations too greatly. The values of z_I and $\hat{\beta}_{i+1}$ that derive from the Gerber-Jones formula are shown in Exhibit 2. This starts with an initial z of unity, then updates with J=.07.

Now suppose, however, that there is a break in the data at the jth point. By setting ν_j large in the original Kalman updating formula, Γ_j also becomes large, which leads to $k_j = 1/x_j$. This mean that the next estimate will be $\hat{\beta}_{j+1}=y_j/x_j$, that is the next observation gets full credibility. The updating starts over from there. How large do these values have to be? Just large enough so that k=1/x to the number of decimal points in the calculation.

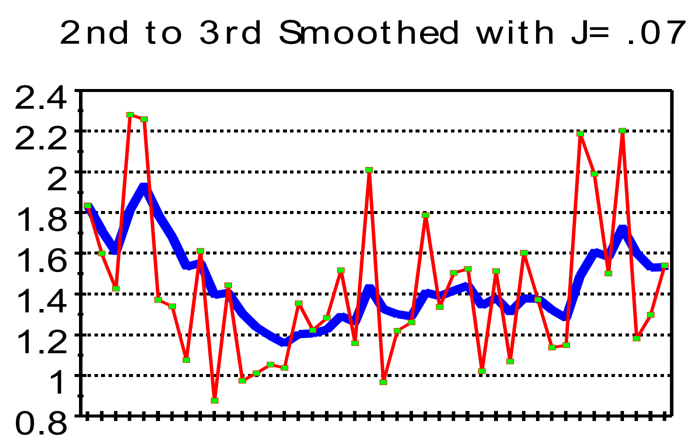

Figure 3

To illustrate this, Exhibit 2 also shows the updating calculation for the 2nd to 3rd factor above with a jump after the fifth point and another jump seven point from the end. These were chosen as points where there appears to be a change in level, not just a random fluctuation. For this example, ν=0.003 except at the two points where it is large, $\sigma_i=\sigma$ =0.3, x_I=1, and k starts at 1. Here all the x's are unity, as the parameter is the estimate. This procedure gives J=1/30 except at the two jump points, which is less than half what it was in the first example. In effect, the process is assumed to be smoother except at the jumps. The effect is shown in Figure 4.

One way of comparing such models is the SSSSPE: sum of squares of single step prediction errors $(y_i-\hat{y}_i)^2$. This is 6.08 for the first smoothing and 5.42 with the jumps. This is a good measure of fit that does not require counting degrees of freedom. If too little smoothing is done, the fit may be better but the prediction errors could be worse, as the fit would be following random fluctuations. By taking other values of the parameters, more or less smoothing can be obtained. The SSSSPE can also be used to test other smoothing techniques, such as average the last 5 diagonals. For this series that method gives an SSSSPE of 6.25.

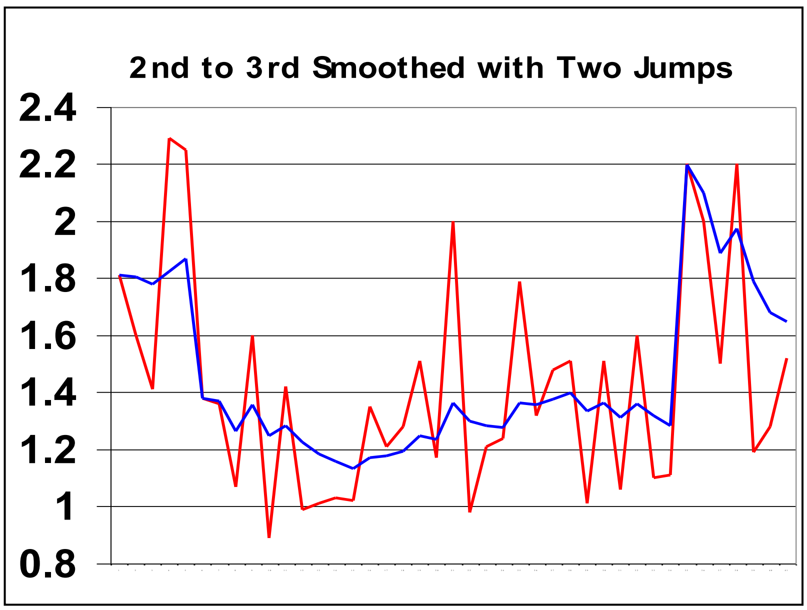

Figure 4

Appendix 2 – 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 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 previous 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 were as in Table 8:

Table 8 - 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
5569.0	5569.0	3739.2	2881.8	2361.1	993.3	993.3	993.3	993.3	-2319.9	-984.7

The sum of squared residuals for this model is 49673.4 when adjusted for seven parameters used. This is considerably better 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 dominate 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:

$$q_{w,d}=f_d g_{w+d} \qquad (11)$$

where f_d is the additive age term for the d^{th} age, 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 9. This had a slightly better adjusted sum of squared residuals of 49034.8.

Table 9 - 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
5692.3	5692.3	3823.0	2816.1	2416.7	672.1	672.1	672.1	672.1	.5598	.6684

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

$$q_{w,d}=f_d g_{w+d} h_w. \qquad (12)$$

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.

Table 10-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
5135.6	5135.6	3464.7	2730.1	1995.4	660.1	660.1	660.1	660.1	.6201	.7225	1.2672

The adjusted sum of squared residuals came down to 44700.9, 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.

Exhibit 1- Comparative Fits

Chain Ladder

	1	2	3	4	5	6	7	8	9
Actual	3257	2638	898	1734	2642	1828	599	54	172
Fit	6101	4705	2846	1912	1350	656	580	296	172
% Error	87%	78%	217%	10%	-49%	-64%	-3%	448%	0%
Actual	4179	1111	5270	3116	1817	-103	673	535	
Fit	129	2438	1408	1728	1374	632	499	257	
% Error	-97%	119%	-73%	-45%	-24%	-714%	-26%	-52%	
Actual	5582	4881	2268	2594	3479	649	603		
Fit	4151	5116	3619	2614	1868	900	736		
% Error	-26%	5%	60%	1%	-46%	39%	22%		
Actual	5900	4211	5500	2159	2658	984			
Fit	6883	6574	4113	3444	2336	1057			
% Error	17%	56%	-25%	60%	-12%	7%			
Actual	8473	6271	6333	3786	225				
Fit	1329	5442	4131	3591	2588				
% Error	-84%	-13%	-35%	-5%	1050%				
Actual	4932	5257	1233	2917					
Fit	1842	3667	3053	2095					
% Error	-63%	-30%	148%	-28%					
Actual	3463	6926	1368						
Fit	678	2287	2856						
% Error	-80%	-67%	109%						
Actual	5596	6165							
Fit	1644	3953							
% Error	-71%	-36%							
Actual	2262								
Fit	3814								
% Error	69%								

CC

	1	2	3	4	5	6	7	8	9
Actual	3257	2638	898	1734	2642	1828	599	54	172
Fit	4364	3746	2287	1631	1082	336	188	59	17
% Error	34%	42%	155%	-6%	-59%	-82%	-69%	9%	-90%
Actual	4179	1111	5270	3116	1817	-103	673	535	
Fit	4364	3746	2287	1631	1082	336	188	59	
% Error	4%	237%	-57%	-48%	-40%	-426%	-72%	-89%	
Actual	5582	4881	2268	2594	3479	649	603		
Fit	4364	3746	2287	1631	1082	336	188		
% Error	-22%	-23%	1%	-37%	-69%	-48%	-69%		
Actual	5900	4211	5500	2159	2658	984			
Fit	4364	3746	2287	1631	1082	336			
% Error	-26%	-11%	-58%	-24%	-59%	-66%			
Actual	8473	6271	6333	3786	225				
Fit	4364	3746	2287	1631	1082				
% Error	-48%	-40%	-64%	-57%	381%				
Actual	4932	5257	1233	2917					
Fit	4364	3746	2287	1631					
% Error	-12%	-29%	85%	-44%					
Actual	3463	6926	1368						
Fit	4364	3746	2287						
% Error	26%	-46%	67%						
Actual	5596	6165							
Fit	4364	3746							
% Error	-22%	-39%							
Actual	2262								
Fit	4364								
% Error	93%								

BF-CC

	1	2	3	4	5	6	7	8	9
Actual	3257	2638	898	1734	2642	1828	599	54	172
Fit	3411	3411	2373	1824	1275	593	593	252	252
% Error	5%	29%	164%	5%	-52%	-68%	-1%	367%	47%
Actual	4179	1111	5270	3116	1817	-103	673	535	
Fit	3411	3411	2373	1824	1275	593	593	252	
% Error	-18%	207%	-55%	-41%	-30%	-676%	-12%	-53%	
Actual	5582	4881	2268	2594	3479	649	603		
Fit	4821	4821	3354	2578	1803	838	838		
% Error	-14%	-1%	48%	-1%	-48%	29%	39%		
Actual	5900	4211	5500	2159	2658	984			
Fit	5956	5956	4143	3185	2227	1036			
% Error	1%	41%	-25%	48%	-16%	5%			
Actual	8473	6271	6333	3786	225				
Fit	7090	7090	4932	3792	2651				
% Error	-16%	13%	-22%	0%	1078%				
Actual	4932	5257	1233	2917					
Fit	4600	4600	3200	2460					
% Error	-7%	-12%	160%	-16%					
Actual	3463	6926	1368						
Fit	4600	4600	3200						
% Error	33%	-34%	134%						
Actual	5596	6165							
Fit	4600	4600							
% Error	-18%	-25%							
Actual	2262								
Fit	4600								
% Error	103%								

Additive with Multiplicative Diagonals and Accident Years

	1	2	3	4	5	6	7	8	9
Actual	3257	2638	898	1734	2642	1828	599	54	172
Fit	3185	3185	2148	2730	1995	660	660	660	477
% Error	-2%	21%	139%	57%	-24%	-64%	10%	1122%	177%
Actual	4179	1111	5270	3116	1817	-103	673	535	
Fit	3185	3185	3465	2730	1995	660	660	477	
% Error	-24%	187%	-34%	-12%	10%	-741%	-2%	-11%	
Actual	5582	4881	2268	2594	3479	649	603		
Fit	4036	6508	4390	3460	2529	836	604		
% Error	-28%	33%	94%	33%	-27%	29%	0%		
Actual	5900	4211	5500	2159	2658	984			
Fit	6508	6508	4390	3460	2529	604			
% Error	10%	55%	-20%	60%	-5%	-39%			
Actual	8473	6271	6333	3786	225				
Fit	5136	5136	3465	2730	1442				
% Error	-39%	-18%	-45%	-28%	541%				
Actual	4932	5257	1233	2917					
Fit	5136	5136	3465	1972					
% Error	4%	-2%	181%	-32%					
Actual	3463	6926	1368						
Fit	5136	5136	2503						
% Error	48%	-26%	83%						
Actual	5596	6165							
Fit	5136	3710							
% Error	-8%	-40%							
Actual	2262								
Fit	3710								
% Error	64%								

Exhibit 2 – Simplified State-Space Example

	J=.07		Two Jumps			
2nd to 3rd	z	beta	Nu	Gamma	k	Beta
a	$b=1/[1+1/(.07+b_-)]$	$c=ab+c_-(1-b)$	d	$e=d+e_-(1-f_-)$	$f=e/(e+0.3^2)$	$g=af+g_-(1-f)$
1.81	1	1.81	0.003	0.003	1.000	1.81
1.60	0.517	1.70	0.003	0.003	0.032	1.80
1.41	0.370	1.59	0.003	0.006	0.062	1.78
2.29	0.305	1.81	0.003	0.009	0.087	1.82
2.25	0.273	1.93	0.003	0.011	0.107	1.87
1.38	0.255	1.79	1000000	1000000	1.000	1.38
1.36	0.246	1.68	0.003	0.093	0.508	1.37
1.07	0.240	1.54	0.003	0.049	0.351	1.26
1.60	0.237	1.55	0.003	0.035	0.278	1.36
0.89	0.235	1.40	0.003	0.028	0.237	1.25
1.42	0.233	1.40	0.003	0.024	0.213	1.28
0.99	0.233	1.31	0.003	0.022	0.198	1.23
1.01	0.232	1.24	0.003	0.021	0.188	1.19
1.03	0.232	1.19	0.003	0.020	0.181	1.16
1.02	0.232	1.15	0.003	0.019	0.176	1.13
1.35	0.232	1.20	0.003	0.019	0.173	1.17
1.21	0.232	1.20	0.003	0.019	0.171	1.18
1.28	0.232	1.22	0.003	0.018	0.170	1.19
1.51	0.232	1.29	0.003	0.018	0.169	1.25
1.17	0.232	1.26	0.003	0.018	0.168	1.23
2.00	0.232	1.43	0.003	0.018	0.168	1.36
0.98	0.232	1.33	0.003	0.018	0.167	1.30
1.21	0.232	1.30	0.003	0.018	0.167	1.28
1.24	0.232	1.29	0.003	0.018	0.167	1.28
1.79	0.232	1.40	0.003	0.018	0.167	1.36
1.32	0.232	1.38	0.003	0.018	0.167	1.36
1.48	0.232	1.41	0.003	0.018	0.167	1.38
1.51	0.232	1.43	0.003	0.018	0.167	1.40
1.01	0.232	1.33	0.003	0.018	0.167	1.33
1.51	0.232	1.37	0.003	0.018	0.167	1.36
1.06	0.232	1.30	0.003	0.018	0.167	1.31
1.6	0.232	1.37	0.003	0.018	0.167	1.36
1.10	0.232	1.31	0.003	0.018	0.167	1.32
1.11	0.232	1.26	0.003	0.018	0.167	1.28
2.20	0.232	1.48	1000000	1000000	1.000	2.20
2.00	0.232	1.60	0.003	0.093	0.508	2.10
1.50	0.232	1.58	0.003	0.049	0.351	1.89
2.20	0.232	1.72	0.003	0.035	0.278	1.97
1.19	0.232	1.60	0.003	0.028	0.237	1.79
1.28	0.232	1.52	0.003	0.024	0.213	1.68
1.52	0.232	1.52	0.003	0.022	0.198	1.65

References

Berquist and Sherman *Loss Reserve Adequacy Testing: A Comprehensive, Systematic Approach,* PCAS 1977

Bornheutter and Ferguson *The Actuary and IBNR*, PCAS 1972

Gerber and Jones *Credibility Formulas with Geometric Weights,* Society of Actuaries Transactions 1975

de Jong and Zehnwirth *Claims Reserving, State-space Models and the Kalman Filter*, Journal of the Institute of Actuaries 1983

Mack *Measuring the Variability of Chain Ladder Reserve Estimates,* CAS Forum Spring 1994

Murphy *Unbiased Loss Development Factors*, PCAS 1994

Popper *Conjectures and Refutations,* Poutledge 1969

Stanard *A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques*, PCAS 1985

Taylor *Separation of Inflation and Other Effects from the Distribution of Non-Life Insurance Claim Delays,* ASTIN Bulletin 1977

Verrall *A State Space Representation of the Chain Ladder Linear Model*, Journal of the Institute of Actuaries, December 1989

Zehnwirth *Linear Filtering and Recursive Credibility Estimation*, ASTIN Bulletin, April 1985

Zehnwirth *Probabilistic Development Factor Models with Applications to Loss Reserve Variability, Prediction Intervals and Risk Based Capital,* CAS Forum Spring 1994