

Refining Reserve Runoff Ranges

Gary G Venter

ABSTRACT

Actuaries who use the well-known reserve variance formulas of Mack and Murphy find that these tend to give quite high variances and, when distributional assumptions are made, reserve ranges. This paper looks at ways to increase the accuracy of reserve estimates and reduce their variances and ranges through three basic methods: finding better fitting models; avoiding using too many parameters; using exposure information. After reviewing some reserve modeling background, a series of examples illustrates the use of the three methods. This provides a way to build better models, but a comprehensive approach to finding the best model is not attempted.

1. BACKGROUND ON DEVELOPMENT TRIANGLE MODELS

Thomas Mack (1993) presented a set of statistical assumptions and criteria under which the chain-ladder estimate is optimal and showed how to calculate the variance of that estimate. The assumptions he outlined are intuitive from the viewpoint of what actuaries might imagine development factors are doing. Basically this model postulates that the numbers that appear in each cell of a triangle are a factor times the previous cell, plus a random element. Venter (1998) and others have argued that this is best arranged as factors applying to previous cumulative losses to produce new incremental losses. After all, it is the incrementals that need predicting. Having a good prediction of new cumulatives can be misleading, as they consist in large part of old cumulatives¹.

¹ Panning (2006) shows that modeling cumulatives from cumulatives can, in some models, accumulate correlated residuals that violate regression assumptions.

Having a model like Mack's allows for testing how well the chain-ladder assumptions apply to specific triangles². While which model works best for a given data set is an empirical matter, when the chain-ladder assumptions fail it is often because incremental losses are not plausible as a percent of the previous cumulative. An alternative modeling paradigm is that losses at each lag are a fraction of the yet-unknown ultimate losses. This is an element of the Bornheutter-Ferguson approach, so such models can be regarded as parameterized versions of BF. Here they are called multiplicative fixed-effects models (MFE), since each cell's expected loss is a product of row and column (and perhaps diagonal) factors.

1.1 Variants of Chain Ladder

Murphy (1994) gives several versions of the chain ladder in a regression setting. Losses at one age are expressed as a factor times the cumulative losses at a previous age plus a random error, plus possibly a constant term. For each age the variance of the random error could be constant, or it could be proportional to the level of the previous cumulative losses, or to the square of the previous cumulative. Murphy shows that for the model with no additive term in the constant variance case, regression theory gives the estimator $\Sigma xy / \Sigma x^2$, where y represents the current losses and x the previous. He calls this the LSM model, for least-squares multiplicative. The factor estimators for the other variance assumptions are $\Sigma y / \Sigma x$ and $\text{average}(y/x)$, respectively. The middle one is the same estimator as Mack's. Unfortunately it is often difficult to tell which behavior of the variance best holds for each column.

² See for example Mack(1994), Venter (1998), Barnett and Zehnwrith (2000).

1.2 Multiplicative Fixed Effects Models

These models express the losses in a cell in a triangle as a product of a row constant and a column constant, which are the fixed effects, plus a random error.

Some notation is needed to discuss this in any detail.

The $n+1$ columns of a triangle are numbered $0, 1, \dots, n$ and denoted by the subscript d , for delay. The rows are also numbered from 0 and denoted by w , for when. The last observation in each row of a full triangle will then have $w+d=n$. The cumulative losses in cell w,d are denoted $c_{w,d}$ and the incrementals by $q_{w,d}$.

For the MFE model $E[q_{w,d}]$ is $U_w g_d$, where U_w and g_d are the row and column parameters, respectively. Note that increasing each g by the same factor and dividing each U by that factor does not change the mean for any cell. To have specificity, it is often convenient to have the g 's sum to 1 . Then U_w can be interpreted as the ultimate loss for year w and g_d the fraction that appear at lag d .

Assuming that the distribution around the cell mean is lognormal, each cell's observation is $\log [q_{w,d}] = \log U_w + \log g_d + \varepsilon_{w,d}$, which is a linear model with a normal error term, and so estimable by regression. This was already studied by Kremer (1982). On the other hand, if the distribution is normal, so $q_{w,d} = U_w g_d + \varepsilon_{w,d}$, the model is non-linear. Mack (1991) linked this model of development triangles to MFE models in classification ratemaking, discussed in Bailey (1963), Bailey-Simon (1960), etc. These models are known to be estimable by a generalization of fixed-point iteration called Jacobi iteration for the parameters, using

$$g_d = \frac{\sum_{w=0}^{n-d} U_w q_{w,d}}{\sum_{w=0}^{n-d} U_w^2} \text{ and } U_w = \frac{\sum_{d=0}^{n-w} g_d q_{w,d}}{\sum_{d=0}^{n-w} g_d^2} .$$
 This is just the result of alternatively treating the g 's and the U 's as known constants, so the model temporarily becomes a simple factor model in the other parameter.

1.3 Poisson - Constant Severity Distribution

A convenient starting point for multiplicative fixed-effects models is to assume the error terms follow the Poisson - constant severity (PCS) distribution. This is the aggregate loss distribution consisting of a Poisson frequency and a constant severity. In this context that assumes all claims or payments in all cells are the same size, call it b . This of course is rarely the case, but the model has some advantages. First, it is an aggregate claims distribution, which most triangles consist of. However its main historical appeal is that in an MFE model estimated by MLE it gives the same reserve estimate as the chain ladder.

In the pure Poisson case, the agreement of methods was shown by Hachemeister and Stanard (1975) although that finding was not published formally until Kremer (1985) in German (translated into Russian as well) and Mack (1991) in English. Renshaw and Verrall (1998) extend this to include a Poisson frequency with constant severity. They called this over-dispersed Poisson not PCS. A good presentation is Clark (2003), who in addition uses a parameterized distribution for the payout pattern. He also discusses the Cape Cod version, for which all years are at the same level. None of the cited papers compare the MFE - PCS variance to the chain ladder's, however.

Giving the same answer as the chain ladder is not a particularly useful criterion

for evaluating models, but it starts from a familiar base. Thus PCS will be the starting point for MFE models here. Using the MLE approach abandons the non-parametric framework. This is not losing much, however, as adopting least-squares estimation is equivalent to MLE with a normal distribution assumption.

There are some disadvantages to MFE models. First of all, they assume all observations are independent, which could easily fail. For PCS it is not possible to estimate the severity parameter b by MLE, as the likelihood function is increasing in b , so does not converge. Thus b has to be assumed known in advance and in fact estimated separately later. Also the PCS variance is proportional to its mean. This may be appropriate for some data, but the variance could be proportional to the square of the mean. The lognormal distribution has this property and is a limiting case for products of random effects, via a multiplicative version of the central limit theorem, and so is also a logical distributional assumption.

1.4 Adding in Calendar-Year Effects

Diagonal effects can be a result of accelerated or stalled claim department activity in a calendar year. Such a departure would often be made up for in a later year or years, so more than one diagonal can be affected. A similar pattern can arise from inflation operating on calendar years. Inflation operating on accident year is in the factor approach, as each year gets its own level. But there can appear to be inflation by accident year that is actually generated by calendar year. If that inflation varies by year, high and low residuals can show up by diagonal. Large differences in residuals among diagonals would suggest that either calendar year inflation or claim department variation is operating. In many cases there are diagonal effects in triangles, and modeling them can provide better fits and

remove errors from the estimates of the other parameters in the model.

Taylor (1977) , following Verbeek (1972) discusses a method for estimating calendar-year effects, which he calls the separation method. For some decades after that, models of calendar year effects were informally called separation models, even when they did not use that technique.

In the chain-ladder model, calendar-year effects can be included by using dummy variables. This is most convenient as a multiple regression, building from the constant-error-variance version of the Murphy factor model. Murphy's model can be estimated by doing separate regressions for each column of incremental losses on the previous column of cumulative losses. Alternatively these regressions can be put together into a single multiple regression. The dependent variables of incremental losses to be fit can be strung out into a single column vector, and each column of previous cumulatives can be put into an independent column vector that is mostly zeros but has the cumulatives at the positions where the dependent variable has the corresponding next incrementals. If the multiple regression is done without a constant, then the coefficient for each vector of previous cumulatives is the development factor of the Murphy model. This assumes a single variance for all the cells, however, while doing separate regressions for each column, as Murphy does, allows a different variance for each column. Violation of the single variance assumption does not usually bias the parameter estimates to any significant degree, but it does change the variance estimates.

Diagonal, or calendar-year, dummy variables can be put into the multiple regression as additional independent variables. The dummy variable for a particular

diagonal would be 0 next to every dependent variable (incremental loss) observation that is not from that diagonal, and it would be set to 1 at the positions corresponding to loss observations that are on that diagonal. Such dummies are not needed for every diagonal, as that would over-determine the model, but only for those diagonals that appear to be high or low. The coefficient in the multiple regression for a diagonal dummy would be an additive term that applies after the development factor. It would thus raise or lower the diagonal by a constant amount. An alternative would be to replace the 1's in the dummy with the appropriate previous cumulative loss. The coefficient of this variable would then represent a multiplicative factor for the diagonal.

In the MFE models, factors can be included for the diagonals of interest. Here the incremental losses $q_{w,d}$ are on the $w+d^{\text{th}}$ diagonal, and so a factor h_{w+d} can be used so that $E[q_{w,d}] = U_w g_d h_{w+d}$.

In the lognormal MFE model $q_{w,d} = U_w g_d h_{w+d} (1 + \eta_{w,d})$, taking logs gives $\log q_{w,d} = \log U_w + \log g_d + \log h_{w+d} + \varepsilon_{w,d}$, which is a linear multiple regression model. It can be set up as a regression with the log incrementals in a single column as the dependent variable, and with dummy variables with 1's picking out the observations on the w^{th} row, d^{th} column, and $w+d^{\text{th}}$ diagonal. The regression coefficient for the w^{th} row vector would then be the estimate for $\log U_w$, the coefficient for the d^{th} column vector would estimate $\log g_d$, and that of the $w+d^{\text{th}}$ diagonal vector would be the estimate of $\log h_{w+d}$.

Barnett and Zehnwrith (2000) set up a model framework of this type, but they

denote $\log U_w$ by α_w and they express $\log g_d = \sum_{k=1}^d \gamma_k$ and $\log h_{w+d} = \sum_{t=1}^{w+d} t_\tau$. This makes $\gamma_d = \log[g_d/g_{d-1}]$, for instance. Thus they call γ a trend. If the g 's are trending upwards or downwards by a power curve for several columns, the same γ can be used for those columns. This has potential for reducing the number of parameters in the model. Similarly the t 's are trends over calendar years and may be constant for a few years, reducing the number of diagonal parameters.

2. COMPARING MODELS

This paper's goal is finding ways to increase the accuracy and reduce the variance and ranges of reserve estimates. The methods explored for doing this are finding better fitting models, reducing the number of parameters, and using exposure information where available. A lower predictive variance is suggestive but not absolutely definitive for having the best model. Calculating variances can also be tedious. Thus when searching for models, variances are calculated only for a few models, and comparison of fits are based on other criteria, from information theory. The original information criterion, by Akaike, called the AIC, can be interpreted as imposing a penalty of 1 to the maximized loglikelihood for each parameter in the model. This is often regarded as too low a penalty, however.

The Hannan-Quinn information criterion (HQIC) has a per-parameter penalty of the log of the log of the number of observations N . For instance for a 10x10 triangle with 55 observations, this gives a penalty of 1.388 for each parameter. The Schwartz-Bayesian information criterion is higher, at the log of the square root of N , which is per-parameter penalty of 2 for 55 observations. This seems to be regarded as a bit high, however. Growing more popular with information theorists is the small sample AIC, denoted by AIC_c . Its per-parameter penalty with p pa-

rameters is $N/[N - p - 1]$, which increases with the number of parameters. The penalty is a bit less than that of the HQIC when there are not too many parameters, but is higher with over-parameterized models. A typical standard for what is a small sample is anything less than 40 times the number of parameters, so would include virtually all loss-development triangles.

Here the AIC_c is favored but the HQIC also used. The formal criteria are actually double what is stated above, but dividing by 2 is convenient in that it directly penalizes the loglikelihood. Since the MFE - PCS loglikelihood increases with b , as does the variance, worse fitting models with a higher variance can have a higher loglikelihood. Thus comparing likelihoods across PCS models requires fixing a value of b and using it for different models. The choice of b affects the scale of the loglikelihood and so the meaning of the parameter penalties, so these can only be regarded as general guidelines and not strict cutoffs for this model.

3. MODELING DETAILS

For the PCS model, a cell with frequency λ has mean $b\lambda$ and variance $b^2\lambda$. For the MFE implementation then $b\lambda_{w,d} = U_w g_d$.

This model is applied here to incremental losses, so that the observation $q_{w,d}/b$ is Poisson with mean $U_w g_d/b$. The loglikelihood function³ can be shown to be a constant plus a weighted sum of these observed values minus the fitted means. The weight applied to each observed value is the log of its fitted mean, and the

³ Note that this requires not fitting just one Poisson distribution but $(n/2 + 1)(n+1)$ of them, defined by $2n+1$ row-column parameters plus b . But MLE applies to fitting multiple distributions with the same parameters. This is noted in the *Loss Models* textbook, for instance.

additive constant is $C = -\sum \ln \Gamma(1 + q_{w,d}/b) \equiv -\sum \ln [(q_{w,d}/b)!]$. Thus:

$l = C + \sum \left(\frac{q_{w,d}}{b} \ln \frac{U_w g_d}{b} - \frac{U_w g_d}{b} \right)$. Taking derivatives, the MLE estimates can be expressed as:

$$g_d = \sum_{w=0}^{n-d} q_{w,d} / \sum_{w=0}^{n-d} U_w \quad \text{and} \quad U_w = \sum_{d=0}^{n-w} q_{w,d} / \sum_{d=0}^{n-w} g_d, \text{ which do not depend on } b.$$

These can be iterated, starting with some values then solving alternatively for the g 's and U 's until the results converge. If the resulting g 's do not sum to 1, it is easy to just divide each by their sum and multiply each U by the same sum. With all the rows and columns getting their own parameters, starting at the upper right and working back can show that these estimates correspond to the chain-ladder calculation. Essentially the U 's are the last diagonal grossed up to ultimate by the development factors and the g 's are the factors converted to a distribution of ultimate. The fitted incrementals are then the g 's applied to the U 's, and can be calculated by using the development factors to back cumulatives down from the last diagonal and then differencing to get the incrementals.

From the chain-ladder viewpoint the fits so calculated use future information to predict the past. But this is not the chain-ladder paradigm. Sometimes incremental losses are better fit as a fraction of ultimate (MFE model) than as a factor times previous cumulative (chain-ladder model). The drawback is that there are more parameters needed for MFE. The chain ladder estimates each subsequent column conditionally on the current column and does not estimate the first column of the triangle. It requires the calculation of n factors to do this. The PCS model does estimate the first column but uses $2n+1$ parameters. Comparing the

fits of the two models is thus a bit awkward. Perhaps comparing the estimated variances is the best way to do this. The process variances can be thought of as measuring the accuracy of the models, and the parameter variance is the parameter penalty.

Clark (2003) discusses calculating the MFE – PCS variance. First an estimate of b is needed. Since the variance of each cell is b times its mean, he suggests estimating b by the sum over the cells of the ratios of cell squared residual to cell fitted mean, all divided by (observations – parameters). That is, with N observations and p parameters, the estimate of b is:

$$\hat{b} = \frac{1}{N - p} \sum_{w,d} \frac{(q_{w,d} - U_w g_d)^2}{U_w g_d}.$$

Then the estimated variance of each projected incremental cell is the cell's mean times this b , and so the reserve variance is the reserve times b . But this assumes all the parameters are known. Since in fact they are estimated, there is another element of reserve variance usually called parameter variance. Clark suggests estimating the parameter variance by the delta method. The delta method (see *Loss Models*) starts with the usual covariance matrix of the parameters, calculated as the inverse of the MLE information matrix (matrix of 2nd derivatives of the negative loglikelihood wrt the parameters). The delta method calculation of the parameter variance of a function of the parameters is the covariance matrix left and right multiplied by the vector of the derivatives of the function wrt the parameters. In this case the function of the parameters is the reserve. For the PCS model, the 2nd derivatives of the loglikelihood function wrt the parameters are:

$$\frac{\partial^2 l}{\partial U_w^2} = -\sum_{d=0}^{n-w} \frac{q_{w,d}}{b U_w^2} ; \quad \frac{\partial^2 l}{\partial g_d^2} = -\sum_{w=0}^{n-d} \frac{q_{w,d}}{b g_d^2} ; \quad \frac{\partial^2 l}{\partial U_w \partial g_d} = -\frac{1}{b}, \text{ otherwise } 0.$$

The derivative of the reserve wrt g_d is $\sum_{w>n-d} U_w$ and wrt U_w is $\sum_{d>n-w} g_d$.

However with g_n set to $1 - \sum_{d<n} g_d$, these have to be adjusted. First $\frac{\partial^2 l}{\partial U_0 \partial g_d} = 0$.

Also now $\frac{\partial^2 l}{\partial g_d^2} = -\frac{q_{0,n}}{bg_n^2} - \sum_{w=0}^{n-d} \frac{q_{w,d}}{bg_d^2}$ and for $d \neq j$, $\frac{\partial^2 l}{\partial g_d \partial g_j} = -\frac{q_{0,n}}{bg_n^2}$. The derivative of

the reserve wrt U_w is the same, but wrt g_d it is now $-\sum_{w=1}^{n-d} U_w$.

4. EXAMPLE 1

In this example the MFE - PCS model is fit to a triangle that has often been used as an example and for which the Mack estimates are known. This is first fit by the MFE - PCS model, then some diagonal parameters are added in, and then ways to reduce the number of parameters used are explored.

The starting point in Table 1 is the incremental development triangle for years 1972 - 81 from Taylor and Ashe (1983) that has been used by Mack, Clark, and many other authors. The first column is estimated ultimate counts.

Often dividing the losses by exposure information like counts produces a more stable triangle, but preliminary analysis suggests that in this case it does not. The source of the data has not been identified, but it is consistent with excess losses with an increasing retention, which with inflation can make the losses more stable than average claim size. Thus exposure information is not necessarily useful in every case, and will not be used here, but is included for reference.

#	<u>Lag 0</u>	<u>Lag 1</u>	<u>Lag 2</u>	<u>Lag 3</u>	<u>Lag 4</u>	<u>Lag 5</u>	<u>Lag 6</u>	<u>Lag 7</u>	<u>Lag 8</u>	<u>Lag 9</u>
40	357,848	766,940	610,542	482,940	527,326	574,398	146,342	139,950	227,229	67,948
37	352,118	884,021	933,894	1,183,289	445,745	320,996	527,804	266,172	425,046	
35	290,507	1,001,799	926,219	1,016,654	750,816	146,923	495,992	280,405		
41	310,608	1,108,250	776,189	1,562,400	272,482	352,053	206,286			
30	443,160	693,190	991,983	769,488	504,851	470,639				
33	396,132	937,085	847,498	805,037	705,960					
32	440,832	847,631	1,131,398	1,063,269						
43	359,480	1,061,648	1,443,370							
17	376,686	986,608								
22	344,014									

Table 1 – Taylor Ashe triangle with ultimate claim counts (#)

Mack’s methods lead to a reserve estimate of 18,681,000 to the end of the triangle and a prediction standard error of 2,447,000. The MFE – PCS model calculated as outlined above gives the same reserve estimate but a prediction standard error of 2,827,000. The difference is due to the combination of a much better fit from the MFE – PCS model, with an almost 50% reduction in process standard deviation, and a parameter standard deviation greater by almost 70% due to the greater number of parameters.

To illustrate the difference in fits, Figures 1 and 2 graph the delay 1 incremental losses as a function of the delay 0 losses and as a function of the estimated ultimate losses. A factor times ultimate losses looks like a much better explanation of the incremental losses than does a factor times losses at 0.

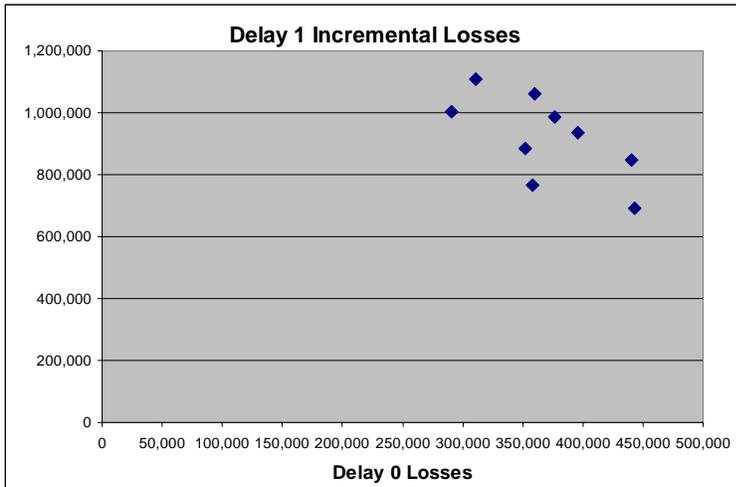


Figure 1

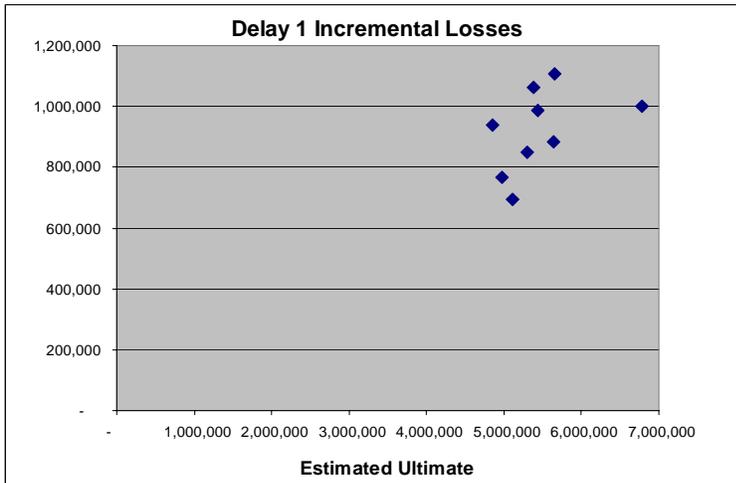


Figure 2

There are of course assumptions that need to be verified for either model. For instance in MFE all of the observations are assumed independent, while for Mack at least the rows should be independent. Both assumptions are violated when there are strong calendar year (diagonal) effects, which there are in this triangle.

Table 2 shows the residuals by diagonal for the MFE - PCS model. Diagonals 2, 3, 4, 6, and 7 are all suspicious, with 7 being the most problematic.

Diagonal	Average Residual	Fraction Positive
0	87,787	1 of 1
1	35,158	1 of 2
2	(76,176)	0 of 3
3	(74,853)	1 of 4
4	100,127	4 of 5
5	(26,379)	2 of 6
6	103,695	5 of 7
7	(115,163)	1 of 8
8	(17,945)	3 of 9
9	38,442	6 of 10

Table 2

A related issue is correlation of residuals between columns. This can be a result of diagonal effects that have not been modeled. Table 3 shows the correlation of the MFE - PCS residuals from one column to the next for the first four columns. All the correlations are negative and two are quite significant.

Columns	0-1	1-2	2-3	3-4
Correlation	-21.5%	-89.5%	-48.9%	-85.4%
Significance	0.289	0.001	0.133	0.015

Table 3

4.1 Incorporating Diagonal Effects

Factors can be put into the model for diagonal effects. Denoting the factor for the j^{th} diagonal as h_j , then instead of the cell expected loss being given by $b\lambda_{w,d} = U_w g_d$, it becomes $b\lambda_{w,d} = U_w g_d h_{w+d}$. Still assuming that the λ 's are Poisson means, the likelihood function is:

$$l = C + \sum \left(\frac{q_{w,d}}{b} \ln \frac{U_w g_d h_{w+d}}{b} - \frac{U_w g_d h_{w+d}}{b} \right)$$

The unconstrained parameter estimates still have an iterative formulation:

$$g_d = \frac{\sum_{w=0}^{n-d} q_{w,d}}{\sum_{w=0}^{n-d} U_w h_{w+d}}, U_w = \frac{\sum_{d=0}^{n-w} q_{w,d}}{\sum_{d=0}^{n-w} g_d h_{w+d}}, \text{ and } h_j = \frac{\sum_{w+d=j} q_{w,d}}{\sum_{w+d=j} U_w g_d}.$$

These converge a bit slowly, but 50 or so iterations often suffice. This can be done in a spreadsheet without programming any functions. Again the g 's can be made to sum to 1, and so represent a payout pattern, but with the calendar-year factors the U 's are then no longer the ultimate losses.

Two models with calendar year effects were fit to the Taylor-Ashe data, adding diagonal parameters for the 7th diagonal, and for the 6th and 7th. To compare the loglikelihoods, b was fixed at 37,183.5. This is the estimated value for another MFE - PCS model, discussed below. With this value, the maximum loglikelihood values for zero, one, and two diagonal factors are:

$$-149.11, -145.92, -145.03.$$

With 55 observations, the HQIC penalty for an additional parameter is 1.388. According to this, the model with both diagonals is better than the one with no diagonal parameters, but not as good as the one with only the 7th diagonal. The AIC_c strongly penalizes having so many parameters (up to 21) with only 55 observations, and charges the first diagonal parameter 2.5 and the second 2.65. This says that the 2nd parameter is clearly helping, but the first one still is. The factors for the 6th and (in both models) 7th diagonal are 1.136 and 0.809.

Including these parameters corrects for potential random errors in the row and column parameter estimates from ignoring diagonal effects. Recall that the chain ladder and original PCS reserves were 18,681,000. Adding one diagonal parameter increases this to 19,468,000 and having them both increases it further to 19,754,000. Thus it appears that the original reserve estimates were too low.

4.2 Reducing the Number of Parameters

The number of parameters in the MFE – PCS model is uncomfortably high and usually not all are significant. There are a few methods for eliminating insignificant parameters without hurting goodness of fit too greatly. First, parameters that are fairly close to each other can be set equal. Also, when changes are systematic, a parameter for a year or delay could be set to the average of the parameters before and after it. More generally, several parameters in a row could be expressed as a trend, which can reduce the number of parameters further. Reducing the parameters in these ways can eliminate distinctions that are not supported by the data. The differences between some years could be small compared to the variability in the parameters for each year, so these years could be given the same parameter. The same can apply to the by-lag and the diagonal effects.

To illustrate these methods, the following model was fit to the Taylor-Ashe data. Accident year 0 is low so gets its own parameter U_0 . Accident year 7 gets its own parameter U_7 as it is high. All the other years get the same parameter U_a , except year 6 which is a transition and gets the average of U_a and U_7 . Thus there are three accident year parameters. This is between the original PCS and Cape Cod models, which get 10 and 1 accident year parameters, respectively.

The fraction paid is divided into high and low payment years with parameters g_a and g_b . Delay 0 is a low year as payments start slowly. Delays 1, 2, and 3 are the higher payment lags and all get g_b . Delays 5, 6, 7, and 8 are again low getting g_a , but delay 4 is a transition and gets the average of g_a and g_b . Finally delay 9 gets the left-overs, i.e., $1 - 5.5g_a - 3.5g_b$. Thus there are only two delay parameters.

Clark suggested using parameterized distributions to describe the payout pattern. Weibull and loglogistic distributions conditional on being less than or equal to 9 were fitted. The loglogistic was better than the Weibull but not as good as the high-low model (all with two parameters) in terms of loglikelihood.

Three of the diagonals were specified as high or low diagonals, getting factors $1+c$ or $1-c$. The 7th diagonal is low and the 4th and 6th are high. Thus only one diagonal parameter c is used.

The loglikelihood for this six-parameter model is -146.66. This is not as good as the twenty-parameter model above, with a loglikelihood of -145.92, but it gets an HQIC penalty that is less by 19.4 and an AIC_c penalty that is lower by 25.5. These clearly overwhelm the difference in loglikelihood of 0.74.

The resulting parameters and their standard errors are:

Parameter	U_0	U_7	U_a	g_a	g_b	c
Estimate	3,810,000	7,113,775	5,151,180	0.0678751	0.1739580	0.1985333
Std Error	372,849	698,091	220,508	0.0034311	0.0056414	0.0568957

Table 4

The parameter variances came from the information matrix. The 2nd derivatives of the unconstrained loglikelihood wrt U_w and g_d do not change with the inclusion of diagonal parameters. The other 2nd partials are:

$$\frac{\partial^2 l}{\partial h_j^2} = - \sum_{w+d=j} \frac{q_{w,d}}{bh_j^2}, \quad \frac{\partial^2 l}{\partial U_w \partial g_d} = - \frac{h_{w+d}}{b}, \quad \frac{\partial^2 l}{\partial U_w \partial h_j} = - \frac{g_{j-w}}{b}, \quad \frac{\partial^2 l}{\partial g_d \partial h_j} = - \frac{U_{j-d}}{b}.$$

The derivatives of the loglikelihood wrt U_a , g_a , g_b , and c , just use the chain rule

on the sum of the derivatives of the loglikelihood wrt the parameters above. However U_a and U_7 are now not independent, as they go into estimation of some of the same cells, and similarly for g_a and g_b . The Appendix summarizes the 2nd partials of the loglikelihood for the six-parameter model.

The correlations of adjacent residuals improve a good deal with the diagonal parameters, as shown in Table 5. This is still somewhat problematic, however, as the correlations are all negative and some are weakly significant. These correlations are still there after accounting for diagonal effects, so might indicate some degree of actual serial correlation in accident year payments. Perhaps ARIMA models could have a role in this modeling. The logic is that high development in one year would be followed by low development the next, which is possible. But forcing the column factors to sum to one would induce some degree of negative correlation across columns, so the extent of this would have to be established before any firm conclusions about auto-correlated development could be made.

Columns	0-1	1-2	2-3	3-4
Correlation	-0.9%	-58.1%	-50.7%	-74.1%
Significance	0.491	0.066	0.123	0.046

Table 5

The reserve estimate from this model is 19,334,000, which is quite close to that of the twenty-parameter model. The prediction standard error (with $b = 37,183.5$) is down to 1,350,000, compared to 2,827,000 for the full MFE - PCS and 2,447,000 for the chain ladder. The better fit from including calendar-year effects and the reduced number of parameters has decreased the standard error appreciably. The breakdown of the variance into parameter and process is in Table 6.

Model	Original 19 Parameter	6 Parameter
Parameter Variance	7,009,527,908,811	1,103,569,529,544
Process Variance	982,638,439,386	718,924,545,072
Total Variance	7,992,166,348,198	1,822,494,074,616
Parameter Std Dev	2,647,551	1,050,509
Process Std Dev	991,281	847,894
Standard Deviation	2,827,042	1,349,998

Table 6

There is a decrease in the process standard deviation of 15%, probably coming from recognizing the diagonal effects, and a 60% reduction in the parameter standard deviation in going from 19 to 6 parameters, for a total decrease in the prediction standard error of over 50%

4.3 Testing the Variance Assumption

In the PCS model the variance of each cell is b times its mean. An alternative is that the variance is proportional to the square of the mean. If that holds for a particular loss triangle, then the PCS standardized residuals (residuals divided by modeled standard deviation) would probably tend to be larger in absolute value for the cells with the larger means. A plot of standardized residuals vs. fitted values would be a way to show this up. These are graphed in Figure 3 for the six-parameter model. This effect does not appear. However the positive residuals have more extreme values than do the negative residuals, which could be indicative of a more highly skewed model.

There is a possible analogue to the PP-plot as well. A PP-plot for a probability distribution fitted to data compares the empirical cumulative probability to the

fitted cumulative probability at each sample point. Here we are fitting 55 Poisson distributions, each of which has a sample of 1, namely $q_{w,d}/b$. The typical empirical probability for the p^{th} observation out of a sample of N is $p/(N+1)$, so this would be $1/2$ for each of our 55 observations. But you could start with the fitted probability at each observation, and rank these 55 fitted values from 1 to N and then assign the empirical probability of $\text{rank}/(N+1)$ to each. This gives something like a PP-plot, and is shown in Figure 4 for the six-parameter model.

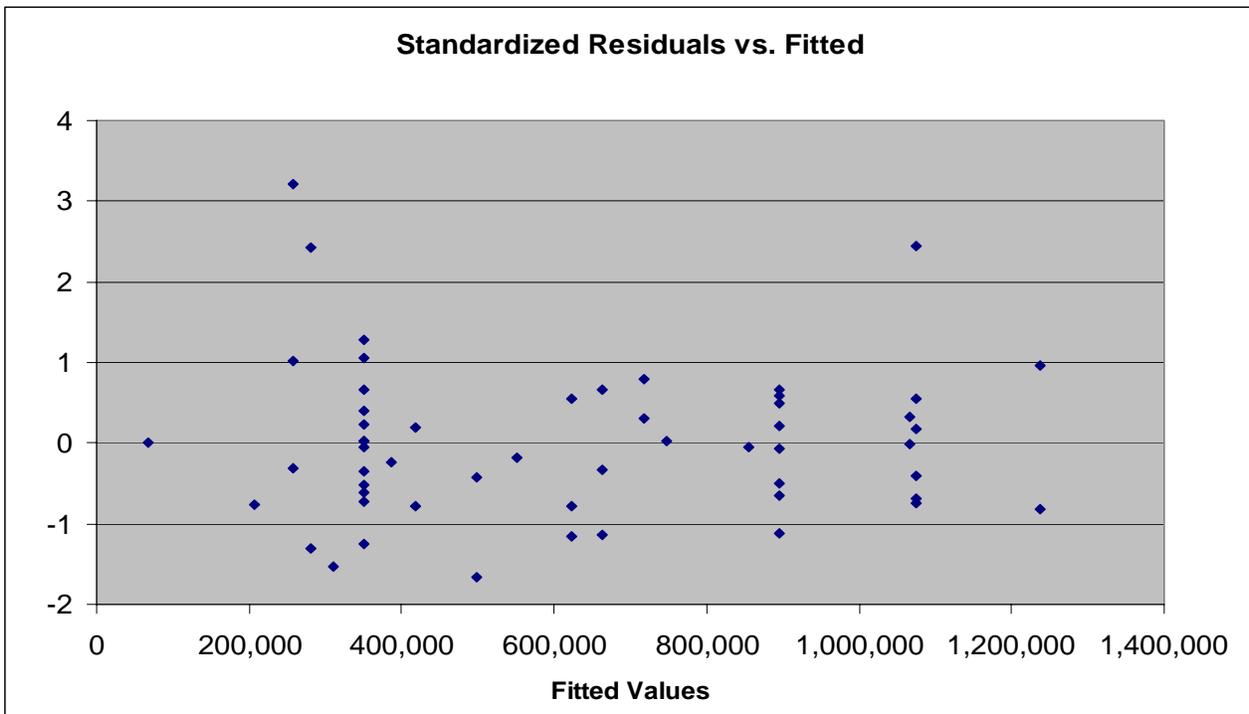


Figure 3

The fit is not too bad, but is better below the median than above. Above there are more observations below most of the probability levels than the Poissons would predict, as shown by the empirical probabilities being higher than the Poisson probabilities. That is a bit surprising, in that usually you would expect observed data to have more large observations than the Poisson. Probably overall this graph would be supportive of the distributional assumption.

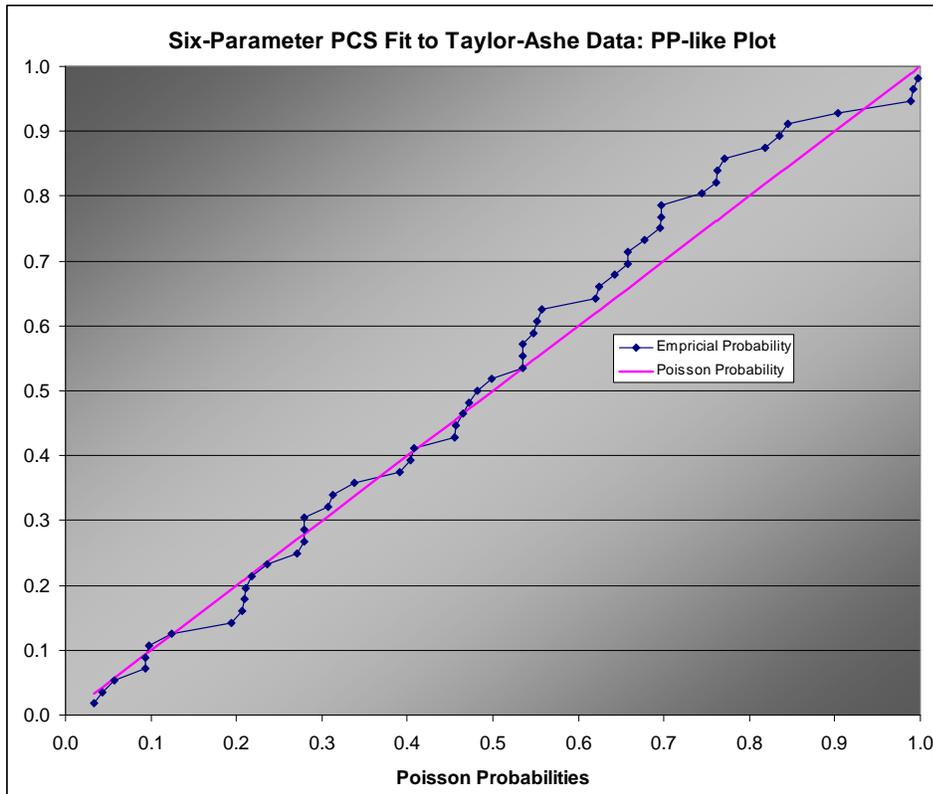


Figure 4

4.4 Example 1 Conclusions

The MFE - PCS model with one parameter for each row and column matches the chain-ladder reserve calculation but can have very different fitted values for the history in the triangle. It has more parameters so a better fit would be expected, but the variance calculation reflects the parameter uncertainty, so the chain ladder can easily give a lower variance. The fit and assumptions of both models can be strained by calendar-year effects, but these can be modeled with their own parameters in either model. As in this example, it should usually be possible to reduce the number of parameters in the models through the use of trends, combination of similar parameters, etc. The MFE models also allow for eliminating some accident year parameters, which can be reduced even to a single parameter in the Cape Cod case. In the example here, three levels sufficed for 10 years.

Many other possible models, including MFE with different distributional assumptions, have not been considered and may give better fits to this data. Negative correlations between adjacent columns might also be real. In summary, getting a better fit by recognizing calendar-year effects and then reducing the number of parameters in the model can decrease the both the process and parameter variances of the reserve estimate. The MFE paradigm is appealing when incremental losses are not well explained as a factor times previous cumulative.

5. EXAMPLE 2

For those who like development factors, it is possible to do many of the steps of Example 1 in a factor setting. Calendar-year effects can be modeled, and parameter-reduction techniques can be applied. These can lead to better fitting models with fewer parameters. Such ideas are illustrated in this example, using a triangle of long-haul trucking liability losses provided by ISO.

Lag 0	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9	Lag 10	Lag 11
11,305	30,210	47,683	57,904	61,235	63,907	64,599	65,744	66,488	66,599	66,640	66,652
8,828	22,781	34,286	41,954	44,897	45,981	46,670	46,849	47,864	48,090	48,105	48,721
8,271	23,595	32,968	44,684	50,318	52,940	53,791	54,172	54,188	54,216	54,775	
7,888	19,830	31,629	38,444	43,287	46,032	47,411	47,677	48,486	48,498		
8,529	23,835	35,778	45,238	51,336	53,574	54,067	54,203	54,214			
10,459	27,331	39,999	49,198	52,723	53,750	54,674	55,864				
8,178	20,205	32,354	38,592	43,223	44,142	44,577					
10,364	27,878	40,943	53,394	59,559	60,940						
11,855	32,505	55,758	64,933	75,244							
17,133	45,893	66,077	78,951								
19,373	50,464	75,584									
18,433	47,564										
20,640											
Factors	2.640	1.5132	1.2220	1.1102	1.0359	1.0149	1.0108	1.0093	1.0017	1.0035	1.0045

Table 7 Long-Haul Trucking Development Triangle with Murphy LSM Factors

The data is for 1984 to 1995. Recall that the LSM model calculates each factor by a least-squares regression. For this data the factors provide a believable representation of the development process for the first five lags. The actual and fitted in-

cremental losses at these lags are graphed as a function of the previous cumulative losses in Figure 5. Some of the deviations from the lines are fairly substantial, but the factors do seem to capture the general pattern of development. This is not to say that factors give the best model for this data – in fact no other models were tested. The goal is just to show how to apply the methods above to factor models.

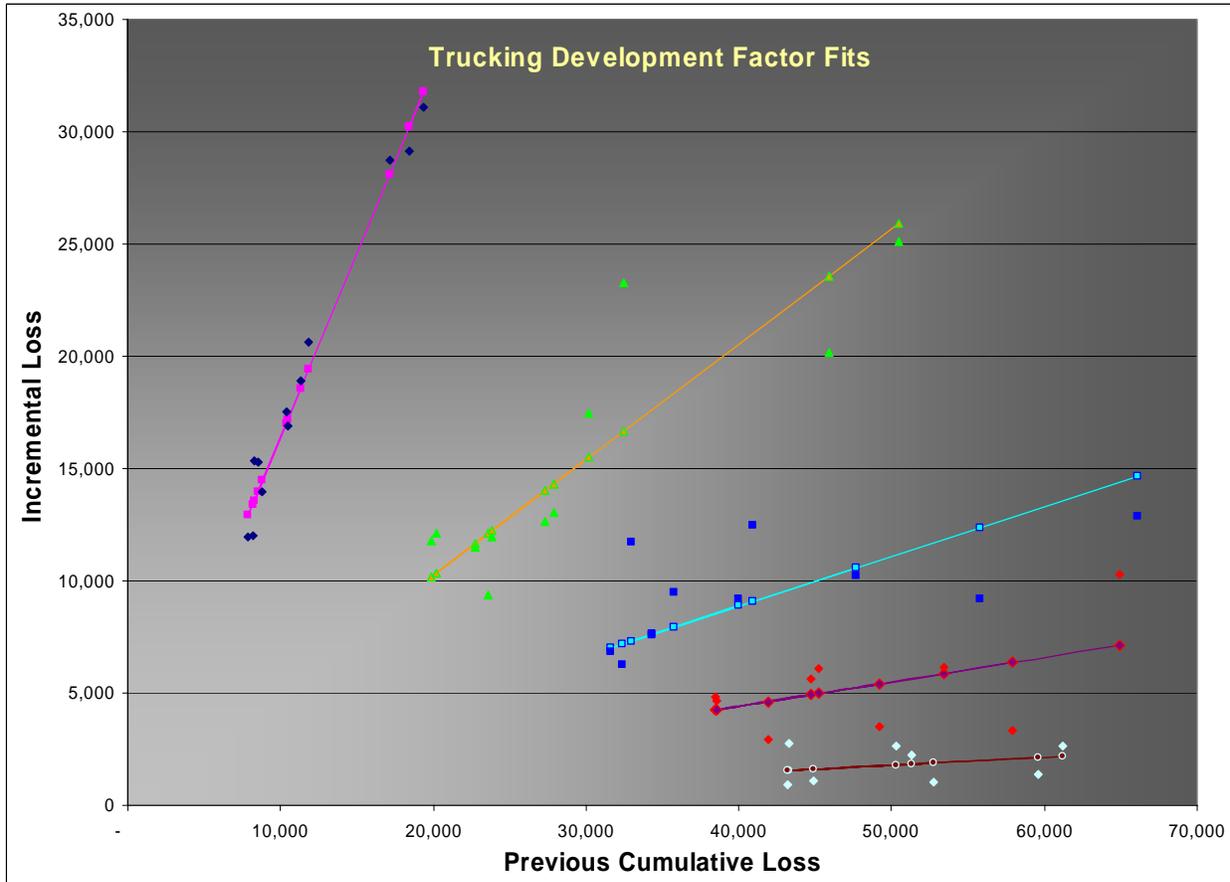


Figure 5

5.1 Multiple Regression Format

To add in diagonal elements, these regressions can be converted to a single multiple regression, and dummy variables added in for the diagonals. Table 8 shows part of the design matrix for such a regression. The incremental losses at lags 1 to 5 (partial) are strung out into the first column, then the subsequent columns are the cumulative losses at lags 0 to 4 that are to predict the next incremental losses.

<u>Incremental</u>	<u>L0</u>	<u>L1</u>	<u>L2</u>	<u>L3</u>	<u>L4</u>	<u>D4</u>
18,904	11,305	-	-	-	-	-
13,953	8,828	-	-	-	-	-
15,324	8,271	-	-	-	-	-
11,942	7,888	-	-	-	-	1
15,306	8,529	-	-	-	-	-
16,873	10,459	-	-	-	-	-
12,027	8,178	-	-	-	-	-
17,515	10,364	-	-	-	-	-
20,650	11,855	-	-	-	-	-
28,759	17,133	-	-	-	-	-
31,091	19,373	-	-	-	-	-
29,131	18,433	-	-	-	-	-
17,474	-	30,210	-	-	-	-
11,505	-	22,781	-	-	-	-
9,373	-	23,595	-	-	-	1
11,799	-	19,830	-	-	-	-
11,943	-	23,835	-	-	-	-
12,668	-	27,331	-	-	-	-
12,150	-	20,205	-	-	-	-
13,065	-	27,878	-	-	-	-
23,253	-	32,505	-	-	-	-
20,184	-	45,893	-	-	-	-
25,120	-	50,464	-	-	-	-
10,221	-	-	47,683	-	-	-
7,668	-	-	34,286	-	-	1
11,716	-	-	32,968	-	-	-
6,815	-	-	31,629	-	-	-
9,460	-	-	35,778	-	-	-
9,199	-	-	39,999	-	-	-
6,238	-	-	32,354	-	-	-
12,451	-	-	40,943	-	-	-
9,175	-	-	55,758	-	-	-
12,874	-	-	66,077	-	-	-
3,331	-	-	-	57,904	-	1
2,943	-	-	-	41,954	-	-
5,634	-	-	-	44,684	-	-
4,843	-	-	-	38,444	-	-
6,097	-	-	-	45,238	-	-
3,524	-	-	-	49,198	-	-
4,631	-	-	-	38,592	-	-
6,165	-	-	-	53,394	-	-
10,312	-	-	-	64,933	-	-
2,671	-	-	-	-	61,235	-
1,084	-	-	-	-	44,897	-
2,623	-	-	-	-	50,318	-
2,745	-	-	-	-	43,287	-
2,238	-	-	-	-	51,336	-
1,027	-	-	-	-	52,723	-

Table 8

The last column is a dummy variable that picks out the incremental losses that are on the 4th diagonal, which are highlighted. Before looking at diagonals, a regression routine provided the output in Table 9 on the 11 development factors estimated by a single no-constant multiple regression.

Parameter	Est value	St dev	t student	Prob(> t)
f1	1.64042	0.03751	43.7337	6.2E-50
f2	0.5132	0.01564	32.8085	3.6E-42
f3	0.22199	0.0118	18.8143	5.3E-28
f4	0.11017	0.01095	10.061	7E-15
f5	0.0359	0.01111	3.23205	0.00193
f6	0.01486	0.01173	1.26635	0.20991
f7	0.01079	0.0122	0.88452	0.37968
f8	0.00931	0.01329	0.69999	0.48643
f9	0.0017	0.0147	0.1155	0.90841
f10	0.00348	0.01636	0.21279	0.83216
f11	0.00451	0.01959	0.23034	0.81855

Table 9

The first five factors are all highly significant, but none of the others are. But they are all positive, so some recognition of development beyond 5th is clearly needed. Since the differences between the factors is small compared to their standard deviations, one possibility is combining some, like 6th through 8th and 9th through 11th, or trending them, or replacing them by a constant or constants. In the end it worked well simply to add a constant to the regression and stop the factors at f5. That reduces the number of parameters by five.

5.2 Modeling Diagonal Effects

Table 10 shows the average residual from the all-factors model and the percent positive for each diagonal.

Diagonal	1	2	3	4	5	6	7	8	9	10	11	12
Avg Residual	359	721	402	(1,681)	1,226	(142)	93	599	(157)	902	(734)	(63)
% Positive	100%	50%	33%	25%	80%	17%	71%	88%	44%	50%	27%	36%

Table 10

The j^{th} diagonal has j fitted values in it except for the 12th, which has 11 values. The 4th, 5th, 8th, 10th and 11th diagonals are suspicious. Adding them in to the regression gives the results in Table 11.

Parameter	Est value	St dev	t student	Prob(> t)
f1	1.6345	0.0364	44.947	6.58E-48
f2	0.5127	0.0151	33.988	6.72E-41
f3	0.2208	0.0115	19.274	2.18E-27
f4	0.1103	0.0108	10.236	8.76E-15
f5	0.0293	0.0108	2.7165	0.0086
f6	0.0117	0.0112	1.0430	0.3011
f7	0.0080	0.0117	0.6902	0.4927
f8	0.0043	0.0130	0.3344	0.7392
f9	0.0005	0.0140	0.0359	0.9715
f10	-0.0004	0.0158	-0.0270	0.9786
f11	0.0110	0.0187	0.5855	0.5604
D4	-1657.7	779.5	-2.1266	0.0376
D5	1325.9	700.0	1.8941	0.0630
D10	1041.5	535.1	1.9463	0.0563
D11	-655.2	528.3	-1.2403	0.2197
D8	726.5	573.2	1.2675	0.2099

Table 11

The same factors are significant but with slightly different values. The 4th diagonal is significant at the 5% level, and the 5th and 10th at a bit weaker levels. Again some combination of the diagonal adjustments might be more significant.

A fairly minimalist model in this context is to keep the first five factors, add a constant to the regression for the later development, keep the 4th diagonal, and have a common factor for the 5th, 8th, 10th and 11th diagonals, but with the 11th subtracted. The constant for all development after 5th works well enough because this development is highly random and does not seem to depend on the level of previous cumulative. The late development could be due to lawsuits coming to a conclusion late in the process, with the timing being highly random. There is still a possibility of improving the model by differentiating stages of the late development, however. The regression results are in Table 12.

Parameter	Est value	St dev	t student	Prob(> t)
Constant	527.81	255.77	2.0636	0.0428
f1	1.601	0.03767	42.4984	3.23E-51
f2	0.499	0.01558	32.0293	3.77E-43
f3	0.211	0.01167	18.0798	7.01E-28
f4	0.102	0.01083	9.4008	5.59E-14
f5	0.021	0.01076	1.9818	0.0515
D4	-1832	724.59	-2.5284	0.0138
D5+D8+D10-D11	801.61	245.88	3.2601	0.0017

Table 12

5.3 Comparing Fits

The loglikelihood at the maximum for a regression with normal residuals on n observations can be expressed as a function of the SSE:

$$\log L = (n/2)\log[2\pi eSSE/n]$$

Using this, with p parameters, $AIC_c/2 = (n/2)\log[2\pi eSSE/n] + np/[n - p - 1]$.

The se of the regression is also a function of goodness of fit and number of parameters, so it is a related comparative measure. The models discussed above are compared on this basis in Table 13.

Model	p	SSE	se	AIC _c /2
All Development Factors	11	171,040,478	1609.821	684.913
All Factors and Five Diagonals	16	133,609,815	1479.975	682.907
Minimalist	8	132,867,569	1387.666	671.218

Table 13

The minimalist model is not a special case of the 16 parameter model because it has a constant term. This appears to provide a somewhat better explanation of the development than does the combination of factors even before adjusting for number of parameters.

5.4 Analysis of Residuals

Figure 6 is a QQ plot of the residuals of the minimalist model vs. the normal distribution regression assumption. The QQ plot graphs the residuals, whereas the PP plot graphs the probabilities of the residuals. In the right tail the last few residuals are much higher than the normal percentiles, while most of the positive residuals are lower than the normal would suggest. This is not very supportive of the normal assumption.

Figure 7 plots the residuals by delay. Regression assumes that all the residuals have the same distribution, but delays 2 through 4 or 5 appear to have a higher variance. Failure to have the same residual distribution is a regression problem known as heteroscedasticity. It does not necessarily affect the estimates of the coefficients, but it does require a different variance calculation.

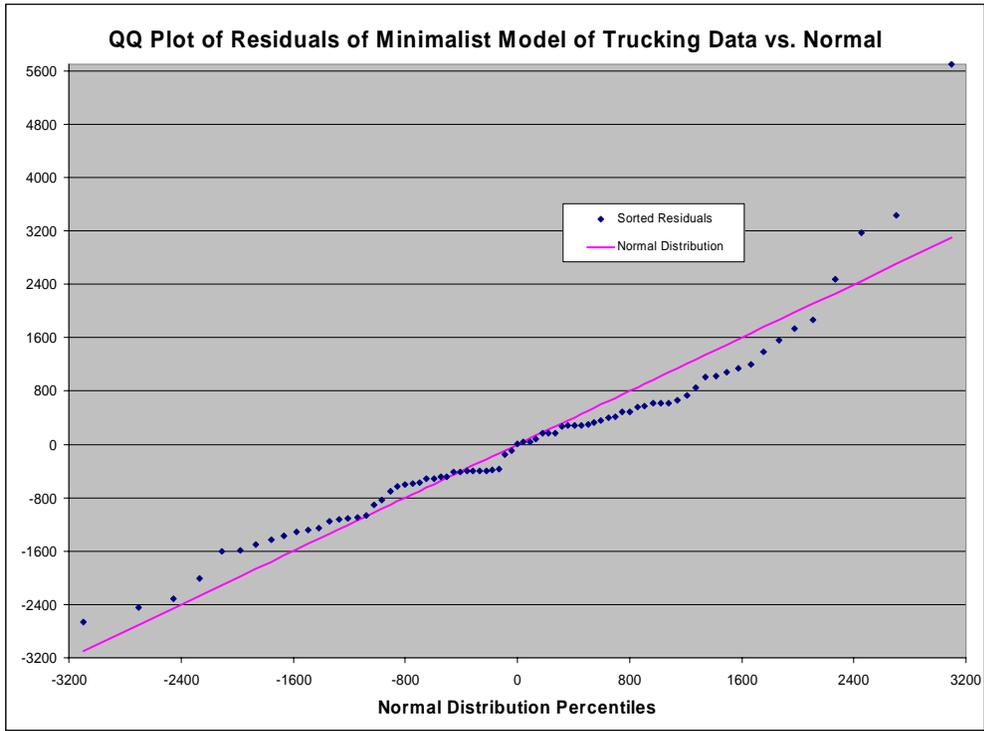


Figure 6

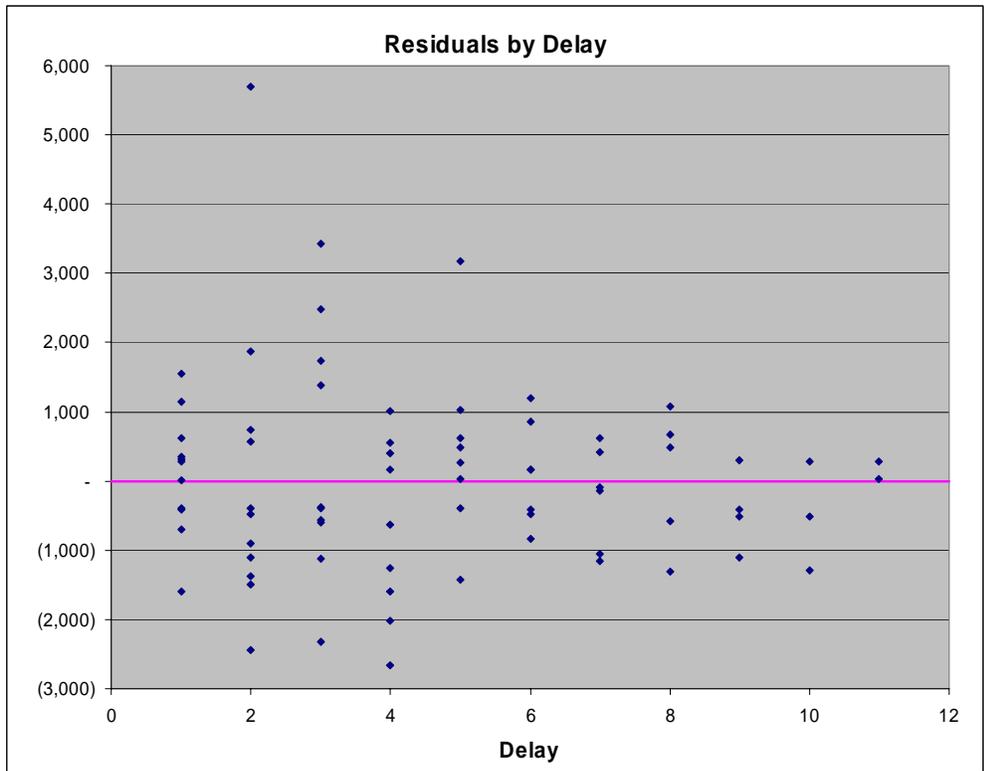


Figure 7

There is a formal test for heteroscedasticity known as White's test, which when applied to this model is ambiguous about the presence of heteroscedasticity. However White's test is not regarded as definitive. In this model heteroscedasticity would be suspected and even preferred in the sense that the smaller observed increments at later stages of development should have lower error variances than the larger increments earlier on. A preference for equalizing relative errors actually would suggest a lognormal model, which is not explored here. However there are correction methods available for adjusting the variance for heteroscedasticity in the model, and these come at little cost, because they do not change the estimate much when the variances are in fact constant. Thus such an adjustment would be appropriate for calculating the variance for this model.

5.5 Estimating the Variance

Again the parameter variance can be estimated by the delta method and the process variance by using the standard error. The covariance matrix of the parameters needed for the delta method is a standard output of multiple regression software. However when heteroscedasticity is suspected, an adjusted covariance matrix is appropriate.

This discussion is based on Long and Ervin (2000). They recommend a heteroscedasticity consistent covariance matrix they call HC3 whenever there is any chance of heteroscedasticity. Explaining this requires getting into the calculations underlying multiple regression. The starting point is the matrix \mathbf{X} of independent variables. This is an $n \times p$ matrix with a row for each observation and a column for each variable. The $p \times p$ matrix $\mathbf{Z} = (\mathbf{X}'\mathbf{X})^{-1}$ is widely used in regression.

The $p \times p$ covariance matrix for the parameter estimates can be expressed in terms of \mathbf{Z} and the $n \times n$ covariance matrix Φ of the observations of the dependent variable as $\mathbf{Z}\mathbf{X}'\Phi\mathbf{X}\mathbf{Z}$. When the error variances of the observations are constant and independent, i.e., $\Phi = \sigma^2\mathbf{I}$, the parameter covariance matrix simplifies to $\sigma^2\mathbf{Z}$. This is the usual parameter covariance matrix put out by regression programs. A convenient calculation of \mathbf{Z} is thus to simply divide this matrix by σ^2 .

To correct for possible heteroscedasticity, let e_i be the residual for the i^{th} observation and define $s_i = \mathbf{x}_i\mathbf{Z}\mathbf{x}_i'$, where \mathbf{x}_i is the row vector of the i^{th} observations of the independent variables. Then $e_i/(1 - s_i)$ is an adjusted residual. The adjusted parameter covariance matrix uses the diagonal matrix of squared adjusted residuals as the estimate of Φ . Thus:

$$\text{HC3} = \mathbf{Z}\mathbf{X}'\text{diag}[e_i^2/(1 - s_i)^2]\mathbf{X}\mathbf{Z}$$

is the adjusted covariance matrix of the parameters.

Since the heteroscedasticity is expected to come from differences among column variances, it would be reasonable to extend this approach to estimating adjusted column variances as well. The average of the squared adjusted residuals down a column of the triangle could be used as the estimate of the variance of the residuals for that column.

For the minimalist model this methodology was applied. The original and revised t-statistics for each parameter are in Table 14.

t's	Constant	f1	f2	f3	f4	f5	D4	D5+D8+D10-D11
Original	2.064	42.498	32.029	18.080	9.401	1.982	(2.528)	3.260
Adjusted	3.501	72.264	17.985	12.838	6.035	3.206	(1.926)	2.574

Table 14

The adjusted standard deviations σ_j by column are in Table 15.

927 2,460 2,135 2,012 831 713 800 919 697 808 228

Table 15

Using these standard deviations, the actual residuals standardized are graphed against standard normal percentiles in Figure 8.

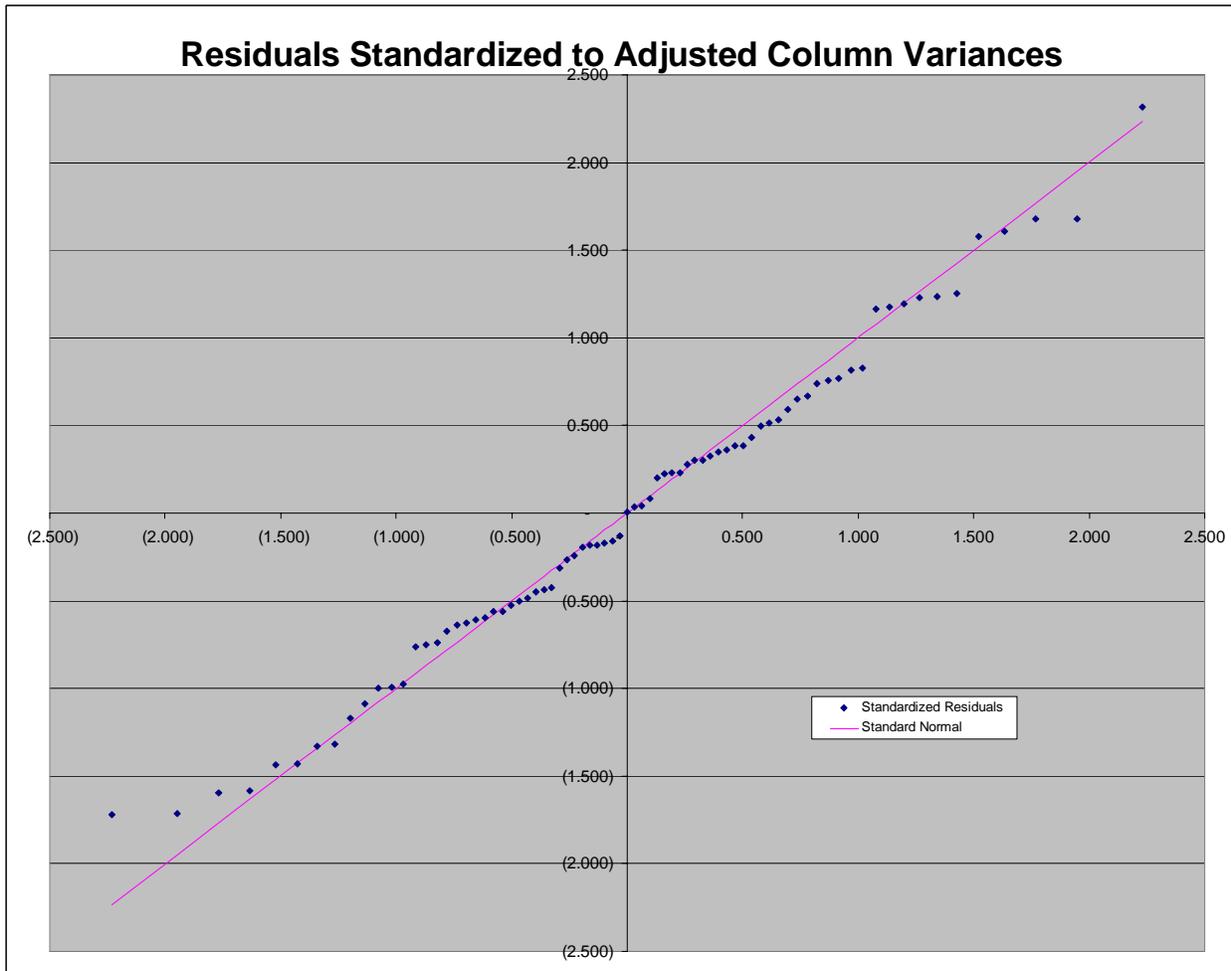


Figure 8

While light in the left tail, this adjustment makes the residuals look more normal. To calculate the variance of the projection, the recursive scheme of Murphy can be applied. First denote by S_j the cumulative losses up through lag j for all acci-

dent years in the triangle not already observed through j . The recursion begins:

$$ES_1 = c_{n,0}(1+f_1)+b$$

$$ES_j = (c_{n-j+1,j-1}+ES_{j-1})(1+f_j)+jb, \text{ where } f_j = 0 \text{ for } j > 5.$$

For the process variance given that the parameters are known:

$$\text{Var}(S_1) = \sigma_1^2$$

$$\text{Var}(S_j) = E\text{Var}(S_j | S_{j-1}) + \text{Var}E(S_j | S_{j-1}) = j\sigma_j^2 + \text{Var}[(1+f_j)S_{j-1}] = j\sigma_j^2 + (1+f_j)^2\text{Var}(S_{j-1})$$

For the delta method the derivatives of S_n can be calculated by recursion as well:

$$\partial ES_1 / \partial b = 1; \partial ES_j / \partial b = j + (1+f_j)\partial ES_{j-1} / \partial b$$

$$\partial ES_j / \partial f_j = c_{n-j+1,j-1} + ES_{j-1}$$

$$\partial ES_j / \partial f_i = 0 \text{ if } i > j \text{ and } \partial ES_j / \partial f_i = (1+f_j)\partial ES_{j-1} / \partial f_i \text{ 0 if } i < j.$$

The results are in Table 16.

	Minimal	Original Murphy LSM
Reserve	213,553	221,800
Process variance	89,501,787	92,565,591
Parameter variance	86,856,827	138,084,020
Variance	176,358,614	230,649,611
Standard deviation	13,280	15,187

Table 16

The reserves are lower in this case when corrected for calendar year effects, the process variance is lower due to a somewhat better fit, and the parameter variance is lower because of eight parameters vs. eleven.

5.6 Variants of the Chain Ladder

Murphy considered three calculations of chain ladder factors, namely regression, ratio of sums, and average of ratios. he showed that these are the least-squares estimates for the factor under three assumptions for the variance of the error term: constant, proportional to loss, proportional to loss squared, respectively. It may be difficult to clearly identify the behavior of the variance, but the three methods can have quite different impacts from extreme observations. Table 17 shows the first two columns of incremental losses from the Facultative General Liability Excess triangle of Mack (1993).

	0	1
0	5,012	3,257
1	106	4,179
2	3,410	5,582
3	5,655	5,900
4	1,092	8,473
5	1,513	4,932
6	557	3,463
7	1,351	5,596
8	3,133	2,262
sum	21,829	43,644

Table 17

Years 1 and 6 start low, so have high development factors. The respective factors from the three methods are 1.217, 1.999 and 7.206. To see their sensitivity to $q_{1,0}$, doubling it to 212 changes the factors to 1.222, 1.990 and 5.106. The implied fitted lines for the original factors along with the data points are graphed in Figure 9. The average residual is zero for the ratio of the sums, but it has a higher sum of squared residuals than the regression.

Again denoting the incrementals by y_i and the previous cumulatives by x_i , the derivatives of the respective factors f with respect to x_j are $[y_j - 2fx_j]/\Sigma x^2$, $-f/\Sigma x$,

and $-y_j/nx_j^2$. For positive incremental losses, these are negative except for the regression estimate if y is at least double the factor times x . For $q_{0,1}$ the derivatives come out $4.66E-05$, $-9.16E-05$, and $-4.13E-02$ respectively, showing greater sensitivity to the outlier as the power of x in the residual variance increases.

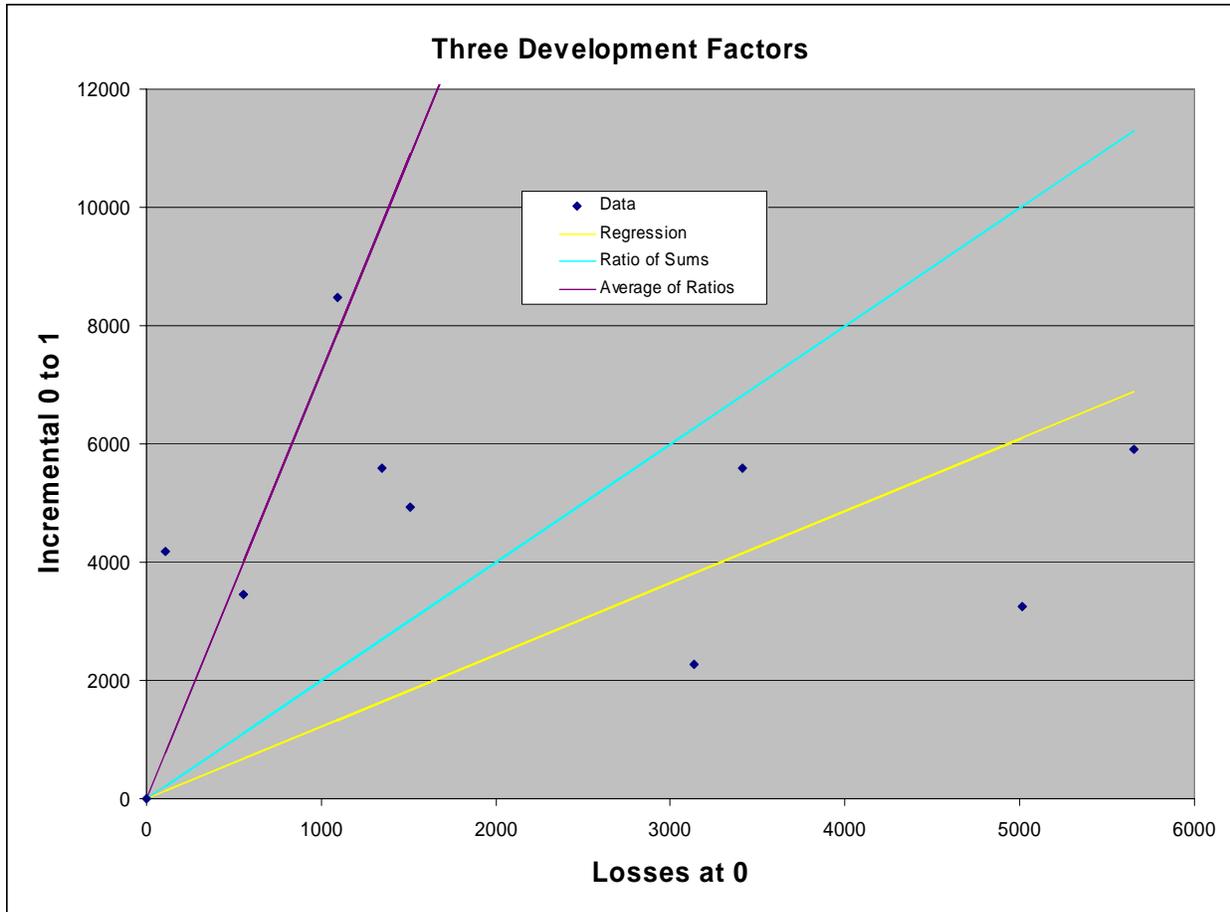


Figure 9

The ratio of sums and average of ratios estimates can be put into regression format by dividing both columns by the previous cumulative, for average of ratios, or its square root, for ratio of sums. For average of ratios, the individual ratios are modeled by a constant, and for ratio of sums the regression estimate $\Sigma xy / \Sigma x^2$

becomes $\Sigma(x/x^{1/2})(y/x^{1/2})/\Sigma x = \Sigma y/\Sigma x$.

This adjustment can be done for multiple regression as well. There is only one previous cumulative in each row of the matrix of independent and dependent variables, so the entire row, including the dummy variables and the 1 for the constant term if included, can be divided by the previous cumulative or its square root. Thus calendar-year effects can be modeled with any variant of the chain ladder. This adjustment is not likely to remove heteroscedasticity from the regressions, however, as the smallest incremental losses are still going to be factors times the largest previous cumulative.

Other variants of the chain ladder are possible. For any power s , dividing x and y by x^s gives the regression estimated factor of $\Sigma x^{1-2s}y/\Sigma x^{2-2s}$. This factor is applied to future estimates of x^{1-s} to estimate y/x^s , so would then be multiplied by x^s , but this is the same estimate as multiplying the factor by x . Even a lognormal chain ladder could be used, with $\log y = \log f + \log x + \varepsilon$, where ε is normal. Care is needed in this case, however, when exponentiating, to include the factor of $\exp(1/2\sigma^2)$ to get the expected value of the development factor. With a diagonal factor this model becomes $\log y = \log f + \log x + \log h + \varepsilon$.

Further variants of the chain ladder using generalized linear models are also possible. Generalized linear models replace the normal distribution assumption of the residuals with other distributions. The PCS could be used, for example, which would have variance proportional to mean for the entire multiple regression. This could in itself eliminate the problem of heteroscedasticity.

6. EXAMPLE 3

This example looks at using exposure data, distributions instead of lag factors, and leaving out data. Factors are sometimes based on the last five diagonals, or even last five diagonals excluding the high and low observations in each column. This example illustrates that it can sometimes be appropriate to leave out some data. However this should be done only when it is clear that there has been a change in the development process. Otherwise leaving out data will increase the variance of the estimate. Excluding high and low observations is particularly problematic in that if factors are from a skewed distribution it will bias the estimated factors downward.

The triangle for this example in Table 18 is cumulative claim counts with exposures for 1978 through 1995 from Taylor (2000).

Exposure	Lag 0	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9	Lag 10
71,543	368	559	587	595	601	606	609	610	610	610	611
75,681	393	544	569	575	579	584	588	589	591	592	592
98,960	517	702	731	748	759	769	777	778	778	778	779
102,974	578	832	881	903	920	926	929	929	930	930	930
106,810	622	828	867	883	886	893	893	894	894	894	894
110,779	660	903	931	943	955	959	963	964	964	964	964
114,307	666	900	953	963	971	975	981	982	982	982	982
117,306	573	839	901	913	918	925	931	936	937	937	938
123,304	582	863	895	922	934	947	953	955	956	956	
125,533	545	765	808	826	838	847	852	854	854		
131,265	509	775	824	846	861	865	873	873			
139,661	589	799	828	845	857	861	870				
152,895	564	760	783	795	804	809					
160,331	607	810	839	848	855						
162,900	674	843	863	875							
170,045	619	809	850								
173,248	660	821									
175,941	660										

Table 18 Cumulative Claim Count Triangle with Exposures

The exposures are growing over time. The usual assumption is that this is more units from the same population. That is not necessarily the case, however, and may not be so here. The development factors are grouped by selected accident year ranges in Table 19. The 0 to 1 factors for the four groups are 1.52, 1.37, 1.47, and 1.32, and the factors are fairly consistent within each group. Most of the development occurs from 0 to 1, so it is critical to get a good estimate for this factor.

1.519	1.050	1.014	1.010	1.008	1.005	1.002	1.000	1.000	1.002
1.384	1.046	1.011	1.007	1.009	1.007	1.002	1.003	1.002	1.000
1.358	1.041	1.023	1.015	1.013	1.010	1.001	1.000	1.000	1.001
1.439	1.059	1.025	1.019	1.007	1.003	1.000	1.001	1.000	1.000
1.331	1.047	1.018	1.003	1.008	1.000	1.001	1.000	1.000	1.000
1.368	1.031	1.013	1.013	1.004	1.004	1.001	1.000	1.000	1.000
1.351	1.059	1.010	1.008	1.004	1.006	1.001	1.000	1.000	1.000
1.464	1.074	1.013	1.005	1.008	1.006	1.005	1.001	1.000	1.001
1.483	1.037	1.030	1.013	1.014	1.006	1.002	1.001	1.000	
1.404	1.056	1.022	1.015	1.011	1.006	1.002	1.000		
1.523	1.063	1.027	1.018	1.005	1.009	1.000			
1.357	1.036	1.021	1.014	1.005	1.010				
1.348	1.030	1.015	1.011	1.006					
1.334	1.036	1.011	1.008						
1.251	1.024	1.014							
1.307	1.051								
1.244									

Table 19 – Development Factors for Claim Count Triangle

One approach to verifying that there actually has been a change in development patterns is to compare the variance of the estimate with the full data and with only the more recent data that appears to be from a different population. In this case the claims through lag 6 (7th column) were developed from all accident years and for the last seven years. Using the Mack formulas, estimating the factors from all the years combined gives an expected future claim count for the last seven years of 481, 68% from the last accident year, and a standard deviation of 62. From just the last seven years alone these are 417 claims with a standard de-

viation of 42, and 65% are from the last accident year. The estimated standard deviation is much lower with the last seven years alone, which supports the idea that there has been a change in development patterns.

Figure 10 graphs the 0 to 1 factors, with the groupings indicated. The last group is subdivided into two sub-groups of three years each.

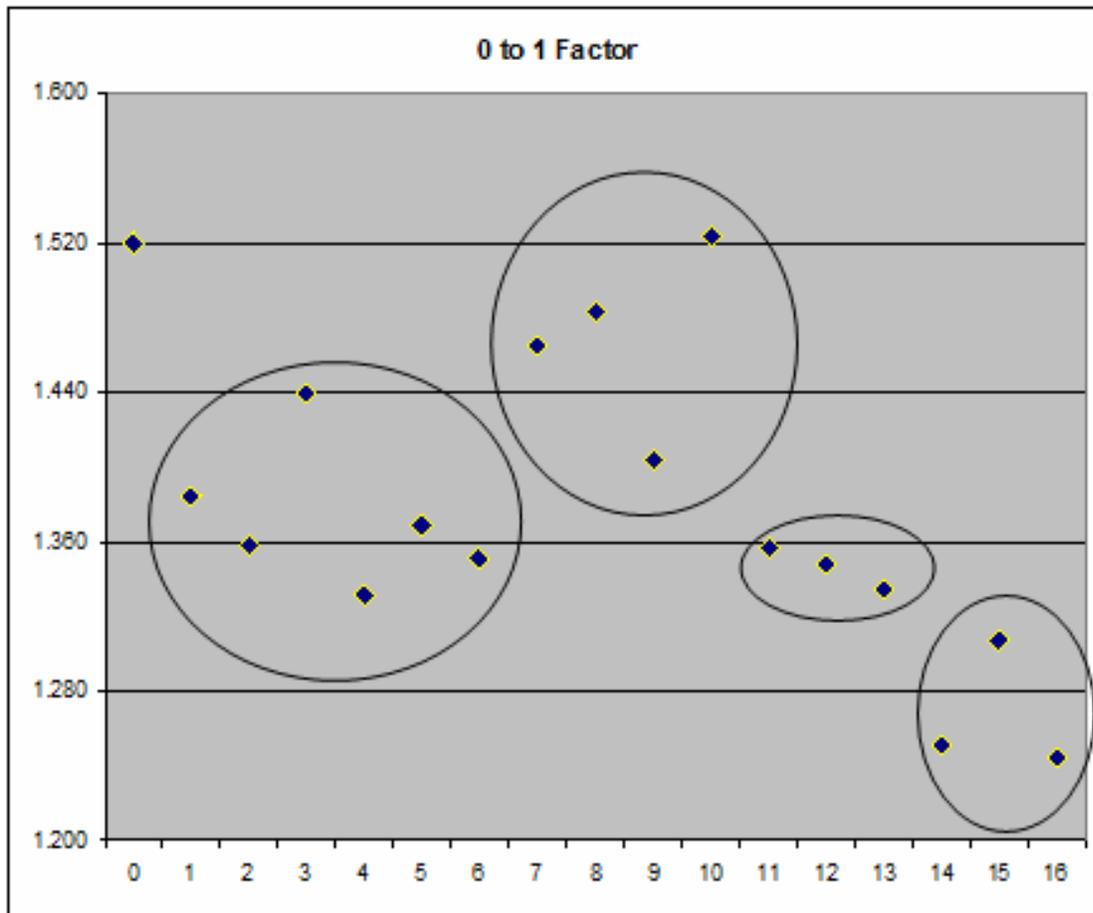


Figure 10

It appears that there have been different eras of internally consistent development factors, and that the last six factors tend to be lower than the others. This supports ignoring most of the older data, especially for the 0 to 1 factor. It raises the question of a possible continuing downward trend, however.

The exposure data is helpful in resolving the question of homogeneity of the last seven years. Table 20 shows the claims per 10,000 exposures for the 0 and 1 lags. The grouping of years is a bit different here. For cumulative claims, the last six years appear homogeneous and different from the years before them. This supports the idea that either the new exposures are from a different population or there has been a change in risk conditions. The claims through lag 1 have gone down from about 80 per 10,000 exposures to less than 50.

0	1 cum	1 incr
51.4	78.1	26.7
51.9	71.9	20.0
52.2	70.9	18.7
56.1	80.8	24.7
58.2	77.5	19.3
59.6	81.5	21.9
58.3	78.7	20.5
48.8	71.5	22.7
47.2	70.0	22.8
43.4	60.9	17.5
38.8	59.0	20.3
42.2	57.2	15.0
36.9	49.7	12.8
37.9	50.5	12.7
41.4	51.7	10.4
36.4	47.6	11.2
38.1	47.4	9.3
37.5		

Table 20 – Cumulative and incremental claims per 10,000 exposures lags 0 and 1

The last six years show what actuaries would like to see from using exposures: all the years appear to be at about the same level after dividing by exposures. This allows for application of a purely additive model, where each column has its own expected increment. There may still be a downward trend within these years for incremental claims at lag 1, but this will be ignored for now.

An additive development of claims per exposure for the last six years through lag five gives a outstanding reserve of 357 claims. These years can be developed through the end of the triangle using data from earlier accident years. Comparing claims per exposure at lags 0 to 5 for the first 11 years to the last 6 shows an average ratio around 1.945. Dividing the average claims per exposure by this for the older years at each lag for lags 6 and on gives a projection of the future claims for the last 6 years. This adds 35 claims to the expected emergence. Finally doing an additive development for the 4 older years that are still incomplete adds 6 more claims, for a total estimated outstanding of 398 claims.

This is considerably less than the 500 projected from the whole triangle, and can be considered an improved estimate due to the use of exposures and the changes that have taken place in the data. This shows that ignoring data can give a better and possibly significantly different estimate when there are demonstrable changes in the process. However ignoring data otherwise can degrade the estimate. It may be possible to find ways to use the older data with time-series methods instead of discarding it for the first several lags. The apparent continuing downward trend in the claims per exposure at lag 1 gives incentive for following through on this. Taylor (2000) explores some alternatives with this data.

The last 6 year triangle with exposures provides an opportunity to apply a parametric model suggested by Clark (2003). Denoting the exposures for year w by P_w and the probability of claims appearing by lag d as G_d , assume that $q_{w,d}$ is Poisson in $P_w r(G_d - G_{d-1})$, where r is an overall ratio of claims to exposures. Any distribution can be used for G , but here Weibull was assumed, with $G_d = 1 - \exp[-(d/\theta)^\alpha]$ for $d = 1, 2, \dots, 5$. Weissner (1978) suggests fitting a truncated version

of the Weibull, which is technically correct, but for simplicity that was not done here, although it does not seem to make a lot of difference in this case since claims have almost finished their development by lag 5. By starting at $d = 1$ the Weibull is fit for claim appearance after lag 0.

Clark provides the likelihood function and its first two derivatives. MLE for this triangle gives $r = 0.001525$, $\theta = 0.5637$ and $\omega = 0.4980$. The resulting outstanding through lag 5 is 354 claims, which is similar to the 357 from the additive development. However this model has only 3 parameters, while additive development has 5, so there may be a lower variance.

The sample variance for each column of claims per exposure is the sum of the squares of the deviation from the average divided by $n - 1$. This variance would apply to each projected incremental cell. In addition there is the variance of the estimated mean, which is the column variance divided by n . This all results in a factor of $(n+1)/[n(n - 1)]$ applied to the sum of squares of the column deviations. For the last column with only one observation an ad hoc variance is typically imputed, and here that was the ratio of the squares of the means applied to the previous variance. This procedure gives the variance of the ratios to exposure for each column of the triangle. In the projection period these are multiplied by the square of the exposures to give the variance of each projected cell. The sum of these through lag 5 is 1087.5, so the standard deviation is near 33.

For the Poisson-Weibull model the process variance of each cell is its mean, by the Poisson assumption. The parameter variance for each projected cell can be calculated by the delta method, using the derivatives of the loglikelihood from

Clark. The covariance matrix of the parameters is in Table 21.

	r	ω	θ
r	6.230E-09	-4.717E-06	3.605E-06
ω	-4.717E-06	6.643E-03	-2.336E-03
θ	3.605E-06	-2.336E-03	5.950E-03

Table 21 – Covariance matrix of Poisson-Weibull fit

The w, d projected cell has mean $rP_w(G_d - G_{d-1})$ and its derivatives wrt the 3 parameters are as in Clark. Summing over the projected cells gives the derivatives of the reserve wrt r, ω, θ as 231,931.82, 95.74 and 65.36. Multiplying the covariance matrix on the left and right by this as a vector gives the delta method estimate of parameter uncertainty of 292. When added to the mean this gives a total variance of 646, or standard deviation of 25.4. Going from 5 to 3 parameters is a 40% reduction in the number of parameters and not much goodness-of-fit was lost, so the standard deviation of the estimated outstanding decreased.

7. SUMMARY

Two paradigms dominate loss development triangle modeling. The conditional approach models each incremental cell's expected value as a linear function of the previous cumulative losses. The unconditional approach models the cell expected losses as a portion of an unobserved level parameter for the year. The chain ladder and BF methods are the original examples of these two paradigms. The unconditional model often fits better but since it uses more parameters (for the accident year levels), it can have higher variances and wider runoff ranges.

Alternative unconditional or conditional models can be compared on parameter-penalized maximized log-likelihood, but it is difficult to compare across the two

paradigms by this method. Perhaps the variance of the estimate is the best common comparison. How to compare models is not a settled issue, however.

Through three examples, ways of improving the estimate were explored. First it is critical to identify calendar-year effects. If these are significant, ignoring them biases the estimates of the other factors. Including them can greatly improve the fit. After that, improving the model primarily consists of getting rid of insignificant parameters. This is not a matter of simply dropping such parameters. It instead involves finding models with fewer parameters that nonetheless account for the observable features of the data.

Replacing level parameters by trends has considerable potential for reducing the number of parameters without sacrificing the fit of the model. In the examples here only linear trends were used and even then just for short periods. But non-linear trends and longer trend periods can be helpful in many cases. A related approach that helped in Example 3 is to use probability distributions for the lag factors. Exposure data when available may improve the modeling as well. When the data has undergone clearly demonstrable changes in structure, using only part of the data can improve the estimates, but otherwise ignoring data will usually increase the variance of the projection. Time series models that account for the changes in structure may be a useful alternative. These could apply vertically, to account for changes in level, horizontally, if high and low development seem to alternate, or by diagonal for evolving cost trends.

Both the conditional and unconditional models can be framed in the notation of multiple regression and put into generalized linear models for alternative resid-

ual distributions. The examples only touched on those possibilities, and many more distributions could be tried. If the normal distribution is used, a heteroscedasticity adjustment is needed. A major issue not explored is using calendar-year trends that are projected into the future instead of constants for the diagonal effects. Changing cost trends can strongly affect the projections, and could be considered a key contributor to model risk, also not addressed here.

References

Bailey, R. 1963 Insurance Rates with Minimum Bias *CAS Proceedings*, L: 4-11.

Bailey, R. and Simon, L. 1960 Two Studies In Automobile Insurance Rate Making *ASTIN Bulletin* 1, 192-217.

Barnett, G. and Zehnwirth, B. 2000. Best Estimates for Reserves. *CAS Proceedings* LXXXVII:245-321.

Clark, D. 2003. LDF Curve-Fitting and Stochastic Reserving: A Maximum Likelihood Approach. *CAS Forum* (Fall): 41-92.

Hachemeister, C. and Stanard, J. 1975. IBNR Claims Count Estimation With Static Lag Functions. *ASTIN Colloquium*, Portimao, Portugal.

Hewitt, C. 1966. Distribution by Size of Risk-A Model, *CAS Proceedings* LIII:106-114.

Klugman, S., Panjer, H. and Willmot, G. 2004. *Loss Models: From Data to Decisions*, 2nd Edition, Wiley

Kremer, E. 1982. IBNR Claims and the Two Way Model of ANOVA. *Scandinavian Actuarial Journal* 1:47-55.

Kremer, E. 1985. *Einführung in die Versicherungsmathematik* Vandenhoeck & Ruprecht, Gottingen

Long, J. and Ervin, L. 2000. "Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model." *The American Statistician* 54:217-224.

Mack, T. 1991. A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. *ASTIN Bulletin* 21:1, 93-109.

Mack, T. 1993. Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates. *ASTIN Bulletin* 23:2, 213-225.

Mack, T. 1994. Measuring the Variability of Chain Ladder Reserve Estimates, *CAS Forum* (Spring). 101-182.

Murphy, D. 1994. Unbiased Loss Development Factors, *CAS Proceedings LXXXI*: 154-222.

Panning, W. 2006. Measuring Loss Reserve Uncertainty, *CAS Forum*, Fall, 237.

Renshaw, A. and Verrall, R. 1998. A Stochastic Model Underlying the Chain Ladder Technique. *British Actuarial Journal* 4:903-23.

Taylor, G. 1977. Separation of Inflation and Other Effects From the Distribution of Non-Life Insurance Claim Delays. *ASTIN Bulletin* 9, no. 1-2:219-30.

Taylor, G. 2000. *Loss Reserving: An Actuarial Perspective*. Kluwer Academic Publishers.

Taylor, G. and Ashe, F. 1983. Second Moments of Estimates of Outstanding Claims. *Journal of Econometrics* 23, 37-61.

Venter, G. 1998. Testing the Assumptions of Age-to-Age Factors. *CAS Proceedings*. LXXXV: 807-47.

Verbeek, H. 1972. An approach to the analysis of claims experience in motor liability excess of loss reinsurance. *ASTIN Bulletin*, 6, 195-202.

Weissner, E. 1978. Estimation of the Distribution of Report Lags by the Method of Maximum Likelihood. *CAS Proceedings*. LXV: 1-9.