Refining Reserve Runoff Ranges

Gary G. Venter, FCAS, MAAA

Abstract

Reserve runoff ranges are often wider than they need to be. This paper applies some practical tools used by regression modelers to find ways to reduce the ranges. Four approaches are explored: finding better-fitting models; getting rid of insignificant parameters; using exposure information; and considering whether some part of the triangle should be ignored.

Keywords. Loss reserving; regression modeling; range estimates; parameter reduction.

1. INTRODUCTION

Techniques that can reduce the runoff variance and reserve ranges are outlined and illustrated through three examples of fitting models to development triangles. Two basic paradigms for development models are used:

[1] Future development is a proportion of losses emerged to date, plus a random error.

[1] Future development is a proportion of the as yet unknown ultimate, plus a random error.

The chain ladder method is the paragon of the first paradigm, and the Bornhuetter-Ferguson (BF) method is an early example of the second. Multiplicative effects models, where the mean of each cell is a product of a row and column parameter, are also of the second type, as the row parameters can be scaled to be expected ultimate.

The factors estimated for both model types can be distorted if there are diagonal (calendar-year) influences in the data. It is possible to identify and take into account such influences in either of the modeling paradigms. This is investigated in all of the examples.

Exposure information, if available, can also improve the model fit and reduce the variance and ranges. Also there are situations where the common models fit better to a portion of the triangle than to the whole triangle, and this is explored as well.

The first example is a triangle whose development pattern is much better explained as a factor times ultimate than a factor times already emerged, but the multiplicative effects model has so many parameters that the estimated variance and runoff ranges are higher than for the chain ladder, despite the better fit, due to greater parameter uncertainty. Ways to maintain a good fit while

eliminating insignificant parameters are explored, and lead to a lower variance. These are somewhat ad hoc methods out of the regression modelers' tool bag. Their application is more of an art than a science but they can produce better models in many cases. The multiplicative effects model can easily handle calendar-year influences by including row, column, and diagonal factors.

The second example is one in which development factors appear to provide a reasonable fit to the data, at least at the early lags. The chain ladder is often presented in a regression context, where factors are calculated using some form of regression on the previous cumulative losses. That gives a separate variance for each factor. It is possible to include diagonal effects in the chain ladder, but the factors have to be computed in a single overall regression. This can get into problems with heteroscedasticity, where a single variance is assumed for each cell but latter lags in fact have lower variances. This does not usually affect the parameter estimates very much, but it does distort the estimated runoff variance. A heteroscedasticity adjustment is introduced and applied to this case. Further use of parameter-reduction techniques are also illustrated.

The third example is of a triangle that exhibits a good deal of change in development patterns over time, and ways to test for that are explored. It also has exposure information available, and using that improves the model. Parameter reduction by fitting a distribution to the emergence lag pattern is applied to this triangle as well.

Section 2 reviews some details of the two modeling paradigms and provides a common notation to discuss them. Section 3 addresses how to compare fits of alternative models. Sections 4, 5, and 6 are the three examples. Section 7 concludes. Standard assumptions, discussed in each case, are used for the distributions of the error terms, but other distributions could be used. These are beyond the scope of this paper, but should not be ignored in application.

2. BACKGROUND ON DEVELOPMENT TRIANGLE MODELS

Mack (1993) [13] presents statistical assumptions and criteria under which the chain-ladder estimate is optimal, and shows how to calculate the implied variance. Mack's assumptions are intuitive from the viewpoint of what actuaries might imagine development factors are doing. Basically they postulate that the incremental losses at a given lag are a factor times the previous cumulative, plus a random innovation.

Having a model like Mack's allows for testing how well the chain-ladder assumptions apply to

specific triangles[1]. Which model works best for a given data set is an empirical matter, but when the chain-ladder assumptions fail it is often because incremental losses are not fit well as a factor times previous cumulative. Then the losses at each lag might be modeled as a fraction of the yet-unknown ultimate losses. This is an element of the Bornheutter-Ferguson approach, so all such models can be regarded as formalized versions of BF. Typically these take the form of multiplicative fixed-effects models (MFE), where each cell's expected loss is a product of row and column (and perhaps diagonal) factors.

2.1 Variants of Chain Ladder

Murphy (1994) [16] 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 and a constant variance, standard 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. Using transformed regressions Murphy shows that the factor estimators when the variance is proportional to losses or losses squared are $\Sigma y/\Sigma x$ and average(y/x), respectively. $\Sigma y/\Sigma x$ is typical in actuarial applications and is the same estimator as Mack's. It is the regression estimator for a no-constant regression of $y/x^{1/2}$ on $x^{1/2}$ that converts the constant variance to a variance proportional to x. Unfortunately it is difficult to tell which behavior of the variance best holds for a single column, so judgment is often needed.

2.2 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 innovation. Some notation is needed to discuss this.

The n+1 columns of a triangle are numbered 0, 1, … n and denoted by the subscript d. The rows are also numbered from 0 and denoted by w. 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,

[1] See for example Mack (1994) [14], Venter (1998) [21], Barnett and Zehnwirth (2000) [3].

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 are 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) [9]. 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) [12] linked this model of development triangles to MFE models in classification ratemaking, such as those in Bailey (1963) [1], Bailey-Simon (1960) [2], etc. These models can be estimated by a generalization of fixed-point iteration called Jacobi iteration, using $g_d = \sum_{w=0}^{n-d} U_w q_{w,d} \Big/ \sum_{w=0}^{n-d} U_w^2$ and $U_w = \sum_{d=0}^{n-w} g_d q_{w,d} \Big/ \sum_{d=0}^{n-w} g_d^2$. 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.

2.3 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 a distribution of aggregate claims, which most triangles consist of. However its historical appeal is that an PCS model estimated by MLE 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) [6] although that finding was not published formally until Kremer (1985) [10] in German (translated into Russian as well) and Mack (1991) [12] in English. Renshaw and Verrall (1998) [17] extend this to the over-dispersed Poisson, which in generalized linear model terminology is defined as any member of the exponential family whose variance is proportional to its mean. However the only distribution meeting this criterion is the PCS. A good presentation is Clark (2003) [4], who in addition uses a parameterized distribution for the payout pattern. 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 the error terms will be assumed approximately PCS distributed for MFE models here.

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 function[2] can be shown to be:

$$l = C + \sum\left(\frac{q_{w,d}}{b}\ln\frac{U_w g_d}{b} - \frac{U_w g_d}{b}\right)$$, where $C = -\Sigma \ln \Gamma(1 + q_{w,d}/b) \equiv -\Sigma \ln [(q_{w,d}/b)!]$. Taking derivatives, the MLE estimates can be expressed as: $g_d = \sum_{w=0}^{n-d} q_{w,d} \Big/ \sum_{w=0}^{n-d} U_w$ and $U_w = \sum_{d=0}^{n-w} q_{w,d} \Big/ \sum_{d=0}^{n-w} g_d$, which do not depend on b. Technically, the Poisson probabilities are zero unless $q_{w,d}$ is an integral multiple of b. However Mack (2002) [15], chapter 1.3.7, shows that there is a continuous analogue of the Poisson that can be scaled by b and gives estimates close to the PCS. When the PCS is applied in a continuous setting it can be thought of as using this distribution. For more details see Venter (2007) [22].

The MLE formulas can be solved by iteration, starting with some values then solving alternatively for the g's and U's until the results converge. If then the g's do not sum to 1, just divide each by their sum and multiply each U by the same sum. Starting at the upper right corner of the triangle 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.

From the chain-ladder viewpoint these 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

[2] 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 [8], for instance.

there are more parameters needed for MFE. The chain ladder models each column conditionally on the previous column and does not estimate the first column of the triangle. It requires the calculation of n factors. The PCS model does estimate the first column but uses $2n+1$ parameters. Comparing the fits of the two models is thus 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) [4] 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:

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

This is a kind of moment matching, but it is not clear how good an estimate of b this might be. 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. This is the process variance, assuming all the parameters are known. Since in fact they are estimated, another element of reserve variance is the parameter variance. Clark suggests estimating this 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 calculates the parameter variance of a function of the parameters by 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:

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

The derivative of the reserve wrt g_d is $\sum_{w>n-d} U_w$ and wrt U_w is $\sum_{d>n-w} g_d$. But 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 not affected by this adjustment, but wrt

g_d it is $-\sum_{w=1}^{n-d} U_w$.

2.4 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. Not accounting for such effects when they are present can lead to misestimating row and column parameters.

Taylor (1977) [18], following Verbeek (1972) [23], 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 lognormal MFE model given by $q_{w,d} = U_w g_d h_{w+d}(1+\eta_{w,d})$, 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.

Barnett and Zehnwirth (2000) [3] set up a model framework of this type, but in a way that facilitates parameter reduction. They denote $\log U_w$ by α_w and express $\log g_d = \sum_{k=1}^{d} \gamma_k$ and $\log h_{w+d} = \sum_{t=1}^{w+d} \iota_t$. This makes $\gamma_d = \log[g_d/g_{d-1}]$, for instance. Thus it makes sense to 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, reducing the number of parameters in the model. Similarly the ι's are trends over calendar years and may be constant for a few years, reducing the number of diagonal parameters.

3. COMPARING MODELS

This paper's goal is finding ways to increase the accuracy and reduce the variance and ranges of reserve estimates. 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, Akaike's information criterion, or 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 10×10 triangle with 55 observations, this gives a penalty of 1.388 for each parameter. The Schwartz-Bayesian information criterion penalty is higher, at the log of the square root of N, which is per-parameter penalty of 2 for 55 observations. This may be a bit high, however. An alternative is the small sample AIC, denoted by AIC_c. Its per-parameter penalty with p parameters 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 most loss-development triangles.

Here the AIC_c is favored but the HQIC also used. The formal criteria are actually double what are 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, thus, the meaning of the parameter penalties. Therefore, these can only be regarded as general guidelines and not strict cutoffs for this model.

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) [20] 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. Exposure information is not useful in every case, and will not be used here, but is included for reference.

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

#	Lag 0	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lg 9
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									

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.

Figure 1

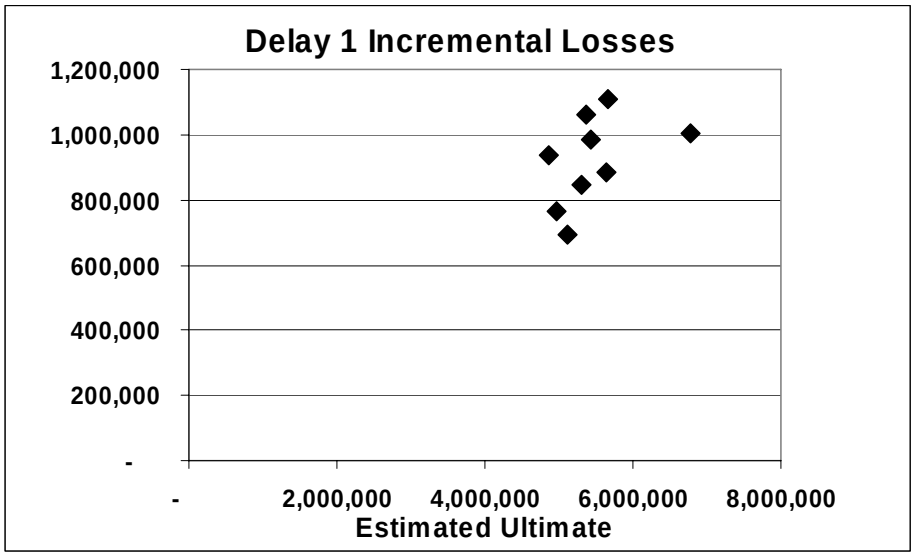
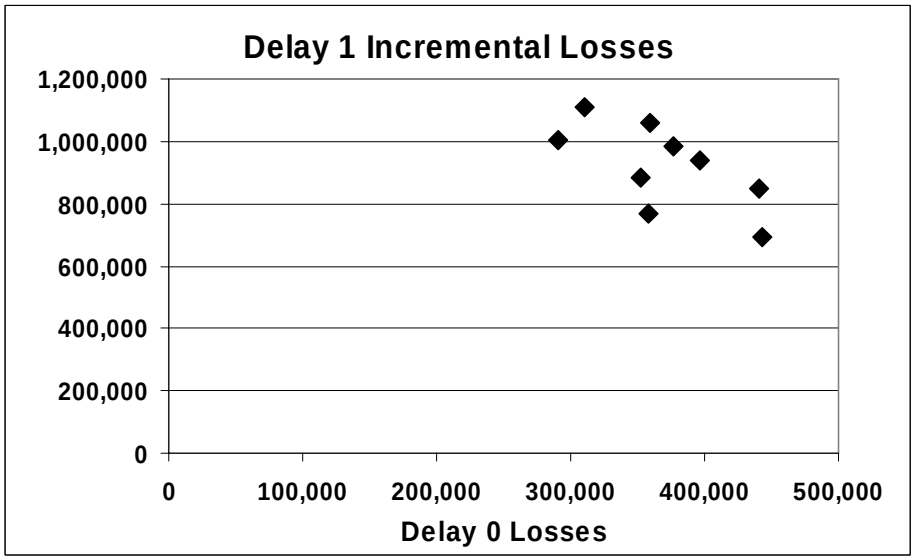


Figure 2



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.

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,

as 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. A related issue is correlation of residuals among 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.

Table 2

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 3

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

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 the cell expected loss is not given by $b\lambda_{w,d} = U_w g_d$, but by $b\lambda_{w,d} = U_w g_d h_{w+d}$. 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:

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

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.

Using this method, 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. The other *b*'s in the iteration were kept at 1. 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 makes no diagonal parameters better than two but worse than one. 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. 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 low.

4.2 Reducing the Number of Parameters

Regression modelers use various techniques to get rid of insignificant parameters without hurting the fit. Parameters that are not significantly different from 0 or 1 are sometimes defaulted to those values. Also parameters that are not significantly different from 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 linear or geometric 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. This can be done for row, column, or diagonal parameters. For instance, up to random effects, the upward and downward diagonal deviations could be indistinguishable. This could hold for many of the late small lag factors and some accident-year mean levels as well.

Several of these methods were tried for the Taylor-Ashe data. A fairly extreme example gets the MFE model down to just six parameters without significantly degrading the fit. In this model,

accident year 0 is low so gets its own parameter U_0. Accident year 7 is high and also gets its own parameter U_7. 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.

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 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 rest, i.e., $1 - 5.5g_a - 3.5g_b$. This leaves only two delay parameters. 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.

This model uses the techniques of setting parameters equal if they are not significantly different and putting other parameters on trend lines – in this case averages – of other parameters. 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
StdError	372,849	698,091	220,508	0.0034311	0.0056414	0.0568957

Table 4

Estimating the parameters was done by an add-in spreadsheet optimizer on the loglikelihood. Most of the build-in spreadsheet optimizers have trouble estimating this many parameters. 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:

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

The derivatives of the loglikelihood wrt U_a, g_a, g_b, and c, 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

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.

Table 5

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 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

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. 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

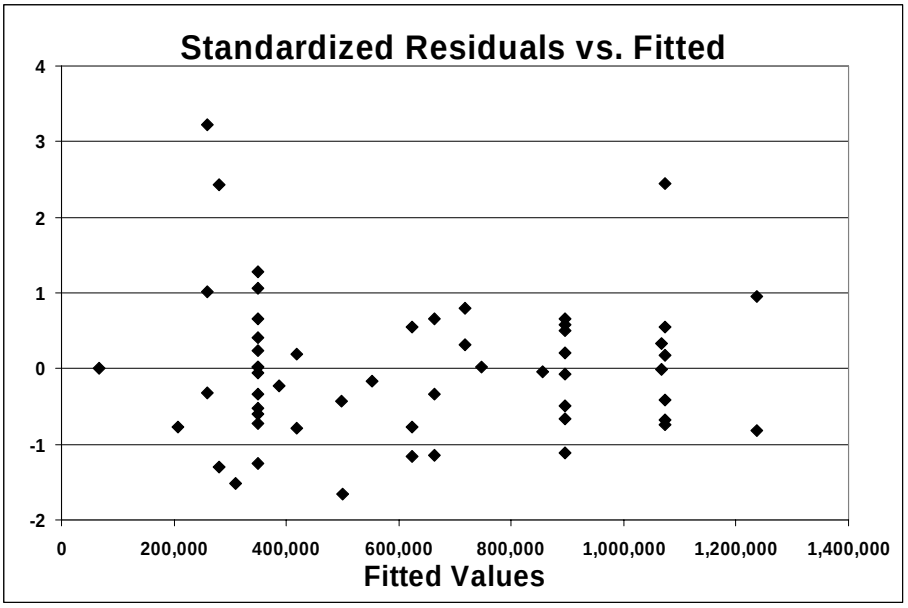
In the PCS model the variance of each cell is b times its mean. If the variance is proportional to a higher power of the mean, then the PCS standardized residuals (residuals divided by modeled standard deviation) would tend to be larger in absolute value for the cells with the larger means. A plot of standardized residuals vs. fitted values would be a way to show this. 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 there are 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 ½ for each of our 55 observations. But you could start with the fitted probability at each point, rank these 55 fitted values from 1 to N and then assign the empirical probability = rank/(N+1) to each. This gives something like a PP-plot, and is shown in Figure 4 for the six-parameter model.

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 is supportive of the distributional assumption, but Figures 3 and 4 both weakly suggest a lighter tail than the Poisson.

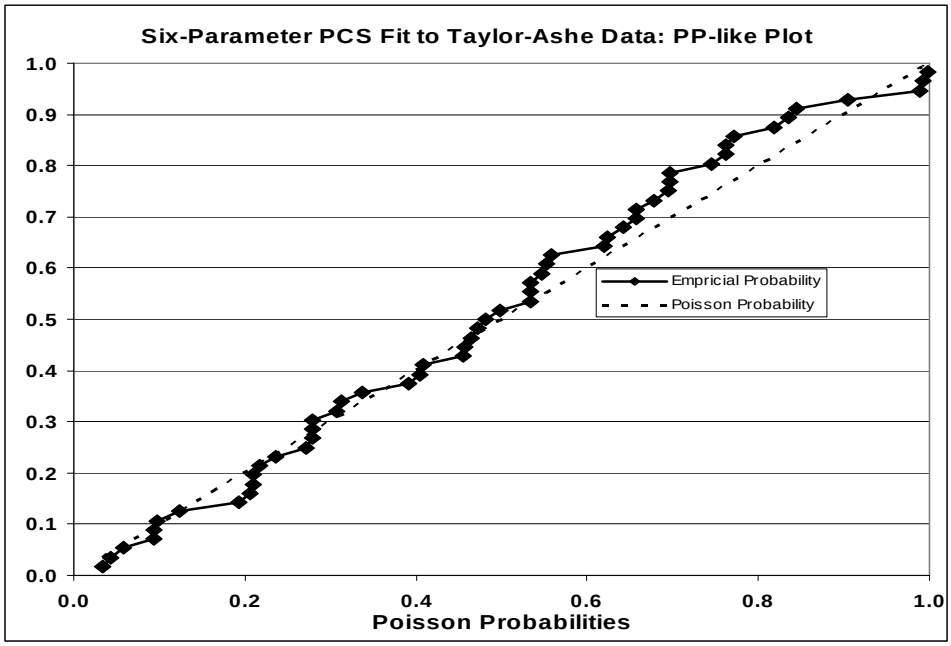
Figure 3



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 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. 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, and these could be modeled by time-series techniques. Taylor (2000) [19] and de Jong (2006) [5] explore time-series modeling for development triangles. 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.

Figure 4



5. EXAMPLE 2

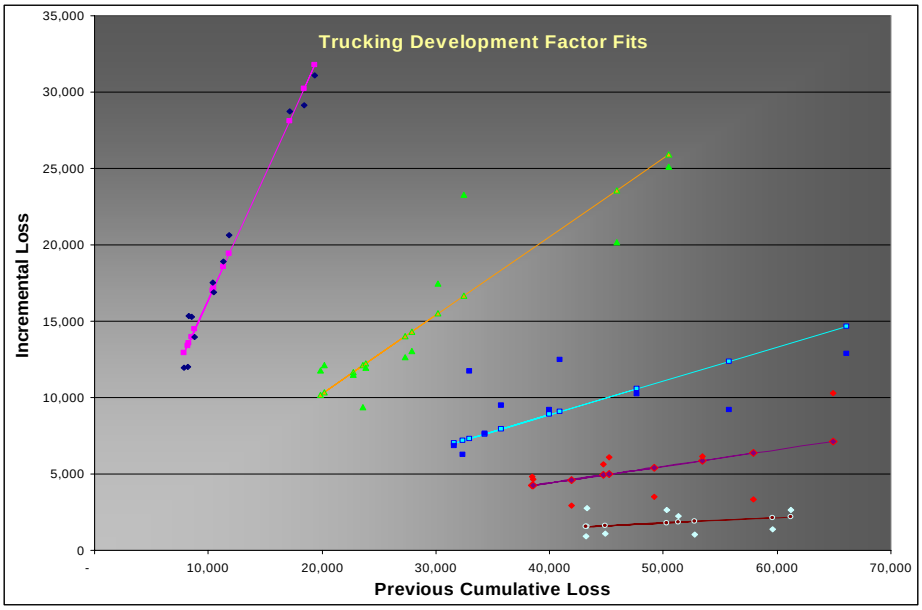
For those who like development factors, it is possible to do many of the steps of Example 1 in a factor setting. Calendar-year effects can be modeled, and parameter-reduction techniques can be applied. These can lead to better-fitting models with fewer parameters. Such ideas are illustrated in this example,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

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

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

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 incremental 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.

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 standard normal-residual regression routine provided the output in Table 9 on the 11 development factors estimated by a single no-constant multiple regression.

Table 8

Incremental	L0	L1	L2	L3	L4	D4
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	-

The first five factors are all highly significant, but none of the others are. Yet they are all positive, so some recognition of development beyond 5th is clearly needed. Since the differences among 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. For this example a constant term was included in the regression and factors f6 to f11 dropped. That reduced the number of parameters by five while still recognizing late development.

Table 9

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

5.2 Modeling Diagonal Effects

Table 10 shows the average residual from the all-factors model and the percent positive for each diagonal. The j^{th} diagonal has j+1 fitted values in it except for the 11th, which has 11 values. The 3rd, 4th, 7th, 9th and 10th diagonals are suspicious. Adding them all to the regression gives the results in Table 11. The same factors are significant but with slightly different values. The 3rd diagonal is significant at the 5% level, and the 4th and 9th at a bit weaker levels. Some combination of the diagonal adjustments might be more significant.

Table 10

Diagonal	0	1	2	3	4	5	6	7	8	9	10	11
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

This model gives separate parameters to all the development factors and the suspicious diagonals. Trying parameter reduction, a fairly minimalist model is to keep the first five factors, add a constant

to the regression for the later development, keep the 3rd diagonal, and have a common factor for the 4th, 7th, 9th, and 10th diagonals, but with the 10th 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, which is not explored here. The regression results are in Table 12. All the parameters are significant enough to keep in the model.

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
D3	-1657.7	779.5	-2.1266	0.0376
D4	1325.9	700.0	1.8941	0.0630
D9	1041.5	535.1	1.9463	0.0563
D10	-655.2	528.3	-1.2403	0.2197
D7	726.5	573.2	1.2675	0.2099

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
D3	-1832	724.59	-2.5284	0.0138
D4+D7+D9-D10	801.61	245.88	3.2601	0.0017

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:

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

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.

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.

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

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.

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.

Figure 6

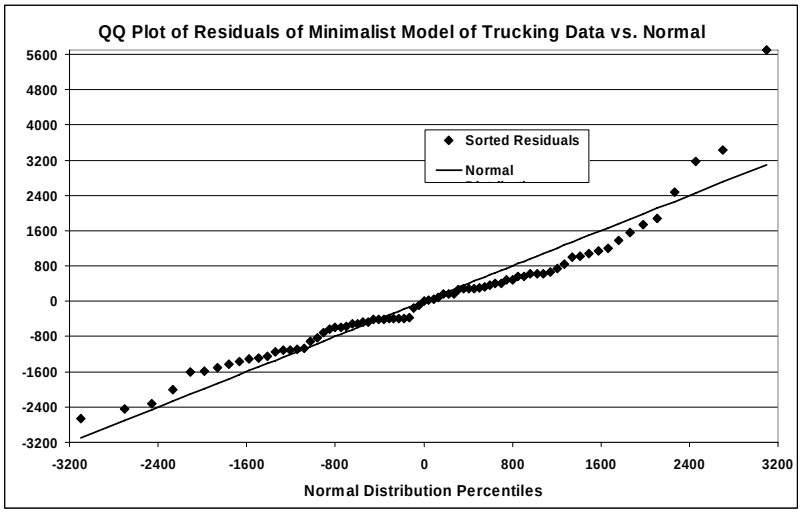
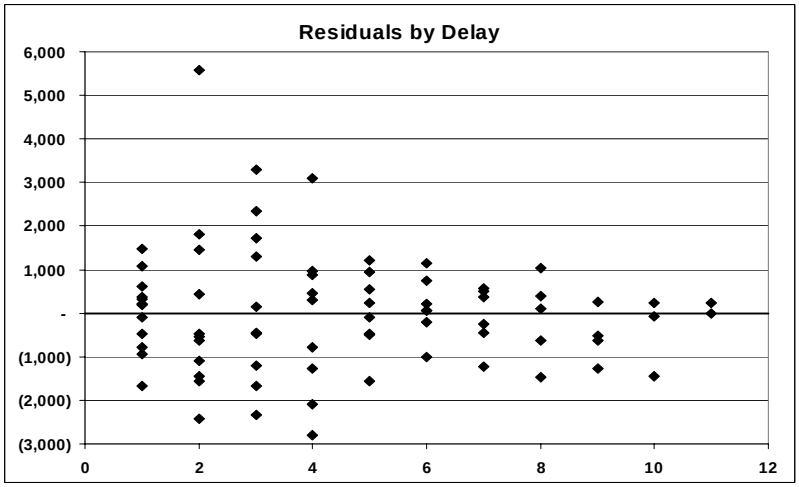


Figure 7



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) [11]. 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 **X** of independent variables, which is an $n \times p$ matrix with a row for each

observation and a column for each variable. The $p \times p$ matrix $\mathbf{Z} = (\mathbf{X'X})^{-1}$ 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 $\mathbf{\Phi}$ of the observations of the dependent variable as $\mathbf{ZX'\Phi XZ}$. When the error variances of the observations are constant and independent, i.e., $\mathbf{\Phi} = \sigma^2\mathbf{I}$, 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 $s_i = \mathbf{x_i Z x_i'}$, 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 $\mathbf{\Phi}$. Thus:

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

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. The adjusted standard deviations σ_j by column are in Table 15. Using these standard deviations, the actual residuals standardized are graphed against standard normal percentiles in Figure 8. While light in the left tail, this adjustment makes the residuals look more normal.

Table 14

h	Constant	f1	f2	f3	f4	f5	D3	D4+D7+ D9-D10
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 15

927	2,46	2,13	2,01	831	713	800	919	697	808	228

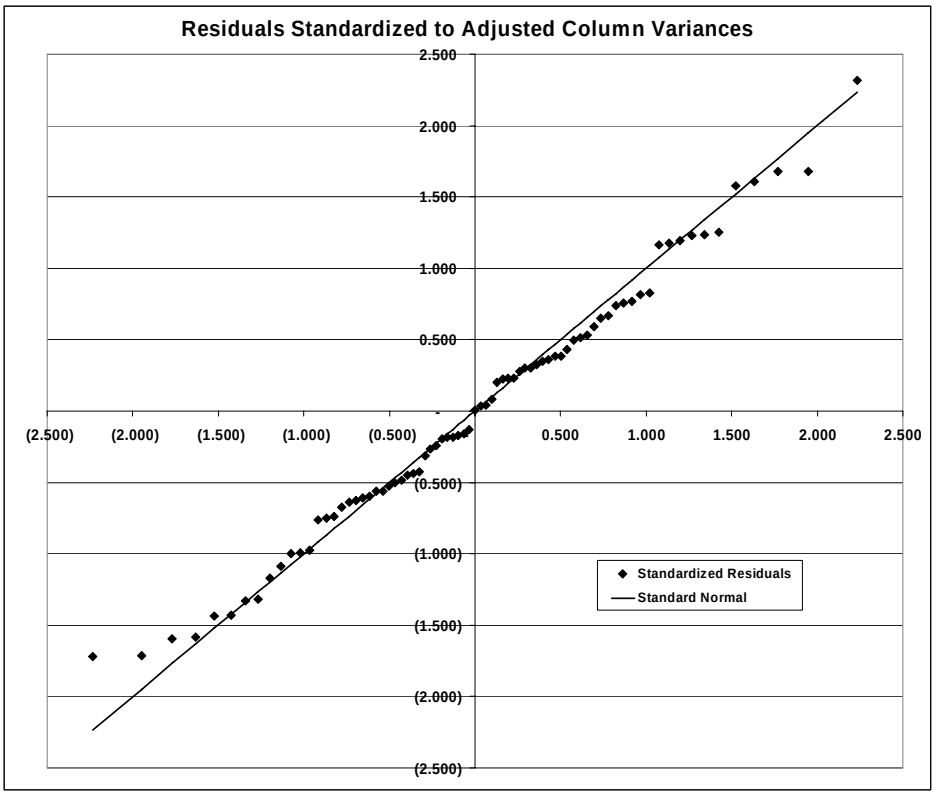


Figure 8

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 accident 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:

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

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

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

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

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

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

The results are in Table 16.

Table 16

	Minimal	Original Murphy LSM
Reserve estimate	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

The reserves corrected for calendar-year effects are lower in this case, the process variance is lower due to a bit better fit, and the parameter variance is lower because of 8 parameters vs. 11.

5.6 Variants of the Chain Ladder

Murphy considered three calculations of chain ladder factors, namely regression, ratio of sums, and average of ratios. As mentioned above, the ratio of sums is a regression for each column where the incremental losses for the column and the cumulative losses for the previous column are both divided by the square root of the previous cumulative, and the average of ratios is the regression divided by the previous cumulative itself.

These adjustments can be done for multiple regression as well. There is only one previous cumulative in each row of the design matrix, 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 cumulatives.

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. This is 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 this will bias the estimated factors downward.

The triangle in Table 17 is cumulative claim counts with exposures for 1978 - 1995 from Taylor (2000) [19]. Exposures are growing over time. The usual assumption is that this consists of 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 18. 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.

Table 17 Cumulative Claim Count Triangle with Exposures

Exposure	Lag 0	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9	Lg 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										

One approach to verifying that there actually has been a change in development is to compare the variance of the estimate using the full data and using only the more recent data that appears to be from a different population. In this case the claims through lag 6 (7^{th} 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, of which 68% are from the last accident year, and a standard deviation of 62. From just the last seven years alone these estimates are 417 claims with a standard deviation 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.

Table 18 – Development Factors for Claim Count Triangle

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									

Figure 9 graphs the 0 to 1 factors, with the groupings indicated. The last group is subdivided into two sub-groups of three years each. 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 19 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.

The last six years show what actuaries would like to see from using exposures: all the years seem to be about the same level after dividing by exposures. This allows for application of an 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. Figure 9

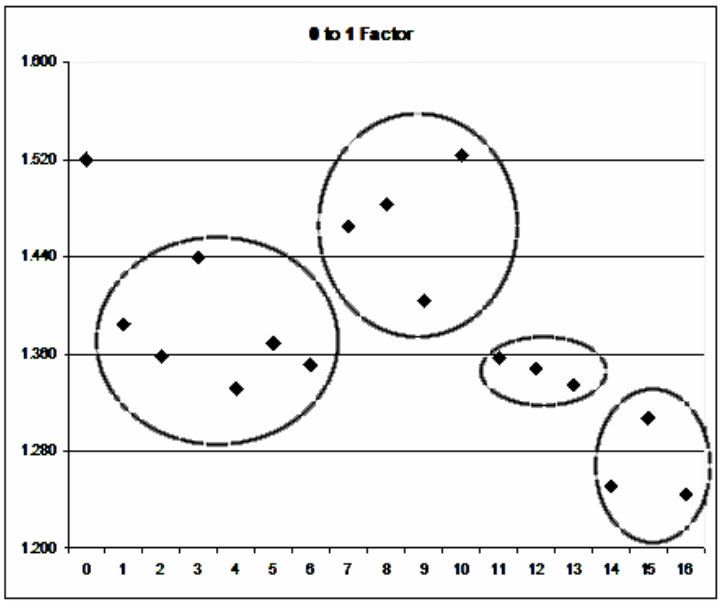


Table 19 – Cumulative and incremental claims per 10,000 exposures

Lag 0	Lag 1 cum	Lag 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		

Additive development of claims per exposure for the last six years through lag five gives an 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 incomplete older years 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 occurred 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 up on this. Taylor (2000) [19] 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) [4]. 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 $G_d = 1 - \exp[-(d/\theta)^{\omega}]$ for $d = 1, 2, \ldots 5$. Weissner (1978) [24] 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 20.

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

Table 20 – 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. CONCLUSIONS

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 loglikelihood, 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 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 residual 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.

REFERENCES

[1] Bailey, Robert A. 1963. "Insurance Rates with Minimum Bias." *PCAS* 50:4-11.
[2] Bailey, Robert A. and Leroy J. Simon. 1960. "Two Studies In Automobile Insurance Rate Making." *ASTIN Bulletin* 1:192-217.
[3] Barnett, Glen and Ben Zehnwirth. 2000. "Best Estimates for Reserves," *PCAS* 87:245-321.
[4] Clark, David R. 2003. "LDF Curve-Fitting and Stochastic Reserving: A Maximum Likelihood Approach." *CAS Forum*, Fall:41-92.
[5] de Jong, Piet. 2006. "Forecasting Runoff Triangles." *NAAJ* 10, no. 2:28-38.
[6] Hachemeister, Charles A., and James N. Stanard. 1975. "IBNR Claims Count Estimation With Static Lag Functions." (Lecture, ASTIN Colloquium, Portimão, Portugal, September 30-October 3, 1975.
[7] Hewitt, Charles C. 1966 "Distribution by Size of Risk-A Model," *PCAS* 53:106-114.
[8] Klugman, Stuart A., Harry H. Panjer, and Gordon E. Willmot. 2004 *Loss Models: From Data to Decisions*, 2nd Ed. Hoboken, NJ: Wiley.
[9] Kremer, Erhard.1982. "IBNR Claims and the Two Way Model of ANOVA. *Scandinavian Actuarial Journal* 1:47-55.
[10] Kremer, Erhard. 1985. *Einfuhrung in die Versicherungsmathematik*. Göttingen, Ger.: Vandenhoek & Ruprecht.
[11] Long, J. Scott and Laurie H. Ervin. 2000. "Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model." *The American Statistician* 54:217-224.
[12] Mack, Thomas. 1991. "A simple parametric model for rating automobile insurance or estimating IBNR claims reserves." *ASTIN Bulletin* 21, no. 1:93-109.
[13] Mack, Thomas. 1993. "Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates." *ASTIN Bulletin* 23, no. 2:213-225.
[14] Mack, Thomas. 1994. "Measuring the Variability of Chain Ladder Reserve Estimates." *CAS Forum* (Spring):101-182.
[15] Mack, Thomas. 2002. *Schadenversicherungsmathematik*. 2nd Edition. Karlsruhe, Ger.: Verlag Versicherungs-wirtshaft.
[16] Murphy, Daniel M. 1994. "Unbiased Loss Development Factors," *PCAS* 81:154-222.
[17] Renshaw, Arthur E. and Richard J. Verrall. 1998. "A Stochastic Model Underlying the Chain Ladder Technique." *British Actuarial Journal* 4:903-923.
[18] Taylor, Greg C. 1977. "Separation of Inflation and Other Effects From the Distribution of Non-Life Insurance Claim Delays." *ASTIN Bulletin* 9, No. 1-2:219-230.
[19] Taylor, Greg C. 2000. *Loss Reserving: An Actuarial Perspective*. Boston: Kluwer Academic Publishers.
[20] Taylor, Greg C. and Frank R. Ashe. 1983. "Second Moments of Estimates of Outstanding Claims." *Journal of Econometrics* 23:37-61.
[21] Venter, Gary G. 1998. "Testing the Assumptions of Age-to-Age Factors." *PCAS* 85:807-847.
[22] Venter, Gary G. 2007. "Generalized Linear Models beyond the Exponential Family with Loss Reserve Applications," *CAS E-Forum,* Summer (forthcoming).
[23] Verbeek, Harry G. 1972. "An approach to the analysis of claims experience in motor liability excess of loss reassurance." *ASTIN Bulletin* 6, no. 3:195-202.
[24] Weissner, Edward W. 1978. "Estimation of the Distribution of Report Lags by the Method of Maximum Likelihood." *PCAS* 65:1-9.

Biography of the Author

Gary Venter is managing director at Guy Carpenter, LLC. He has an undergraduate degree in philosophy and mathematics from the University of California and an MS in mathematics from Stanford University. He has previously worked at Fireman's Fund, Prudential Reinsurance, NCCI, Workers Compensation Reinsurance Bureau and Sedgwick Re, some of which still exist in one form or another. At Guy Carpenter, Gary develops risk management and risk modeling methodology for application to insurance companies. He also teaches a graduate course in loss modeling at Columbia University.

917.937.3277 **gary.g.venter@guycarp.com**