Reducing the Variance of Reserve Estimates

Gary G Venter

ABSTRACT

Actuaries who use the well-known reserve variance formulas of Mack and Murphy find that they tend to give quite high variances. Here we look at ways to reduce the variance through three basic methods: find better fitting models; reduce the number of parameters; use exposure information.

INTRO

Thomas Mack (1993) introduced the world to finding statistical modeling assumptions for the chain-ladder method and calculating the variance of the resulting estimate. The assumptions he outlined are quite intuitive from the viewpoint of what actuaries might imagine is happening behind the scenes when they are applying development factors. Basically, 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 suggested that this is best arranged as factors applying to previous cumulatives to produce new incrementals. 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 cumulatives[1].

Having a model like Mack's allows for testing of the assumptions, which several authors have discussed[2]. While which model works best for a particular

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

[2] Mack(1994), Venter (1998), Barnett and Zehnwirth (2000) among others.

data set is an empirical matter, the chain-ladder often fails 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. We will call them multiplicative fixed-effects models, since each cell's expected loss is a product of row and column (and perhaps diagonal) factors.

POISSON – CONSTANT SEVERITY MODEL

A convenient starting point for multiplicative fixed-effects models is the Poisson – constant severity model (PCS). This model postulates that each cell has an aggregate loss distribution consisting of a Poisson frequency and a constant severity – that is all claims or payments in all cells are of the same size, call it b. This of course is rarely the case, but the model has some advantages. First of all, aggregate claims distributions often have gamma-like tails, which this does as well. However its main historical appeal is that it gives the same reserve estimate as the chain ladder, and through the same calculation steps.

For MLE 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 and Mack (1991) in English. Renshaw and Verrall (1998) extended this to the model that agrees with a Poisson aggregate with constant severity . A good presentation is Clark (2003), who assumes 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 resulting variance to that from Mack, however.

MODELING APPROACH

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 our starting point here. We will keep to the MLE approach, even though this abandons the non-parametric and regression frameworks. This is not losing much, however, as adopting least-squares estimation is equivalent to MLE with a normal distribution assumption, and so is regression.

There are some disadvantages to the PCS model. First of all, it assumes all observations are independent, which could easily fail. Second, 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 to be known in advance and in fact estimated separately later. Also the PCS variance is proportional to its mean. Often having the variance proportional to the square of the mean is thought to be more reasonable for loss models. The lognormal distribution has this property and also is a limiting case for products of random effects, via a multiplicative version of the central limit theorem, and so is a more logical distributional assumption. But start where you are, they say, so we will start with the PCS.

COMPARING MODELS

Our stated goal is reducing the variance of reserve estimates, and so increasing their accuracy. The methods we will explore for doing this are finding better fitting models, reducing the number of parameters, and using exposure information where available. Having a lower predictive variance is useful but not absolutely definitive as being the best model. Calculating variances can

also be tedious. Thus when searching for models, comparison of fits will be based on information-theoretic criteria, and variances only calculated for a few models. The original information criterion, by Akaike, called the AIC, penalizes the loglikelihood by 1 for each parameter in the model. This is probably not enough of 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.

Even more popular with the information guys 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 standard for what is a small sample might be less than 40 times the number of parameters, so would include virtually all loss-development triangles.

We will favor the AIC_c but also check the HQIC. However since the 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.

MODELING DETAILS

The n+1 columns of a triangle will be numbered 0, 1, … 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 typical 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 PCS model, a cell with frequency λ has mean $b\lambda$ and variance $b^2\lambda$. Initially we will assume there is a separate parameter for each row and each column, so $b\lambda_{w,d} = U_w g_d$. The capitalization is rather arbitrary for historical reasons. 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, we often adopt the convention that 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 appears at lag d.

We apply this model to incremental losses, so that the observation $q_{w,d}/b$ is Poisson with mean $U_w g_d/b$. The loglikelihood function[3] 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 additive constant is $C = -\Sigma \ln \Gamma(1 + q_{w,d}/b)$. 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:

---

[3] Note that we are 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.

$g_d = \sum_{w=0}^{n-d} q_{w,d} \Big/ \sum_{w=0}^{n-d} U_w$ 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.

These can be put into a fixed-point iteration, 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, just divide each by the sum and multiply each U by the same sum. However with all the rows and columns getting their own parameters, starting at the upper right and working back can show that these estimates come from 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 essentially a different model. Sometimes incremental losses are better fit as a fraction of ultimate than as a factor times previous cumulative. The drawback is that there are more parameters needed. 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+1 parameters 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.

Clarke (2003) discusses calculating the 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:

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

Then the estimated variance of each projected incremental cell is the cell's mean times this *b*. For the reserve estimate the variance is thus 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. Clarke 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:

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

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

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 $\frac{\partial^2 l}{\partial g_d^2} = -\frac{q_{0,n}}{b g_n^2} - \sum_{w=0}^{n-d} \frac{q_{w,d}}{b g_d^2}$ and for d ≠ j, $\frac{\partial^2 l}{\partial g_d \partial g_j} = -\frac{q_{0,n}}{b g_n^2}$. The derivative of the reserve wrt U_w is the same, but wrt g_d it is now $-\sum_{w=1}^{n-d} U_w$.

EXAMPLE 1

This is a development triangle from Taylor and Ashe (1983) which has been used by Mack, Clarke, and many other authors.

357,848	766,940	610,542	482,940	527,326	574,398	146,342	139,950	227,229	67,948
352,118	884,021	933,894	1,183,289	445,745	320,996	527,804	266,172	425,046	
290,507	1,001,799	926,219	1,016,654	750,816	146,923	495,992	280,405		
310,608	1,108,250	776,189	1,562,400	272,482	352,053	206,286			
443,160	693,190	991,983	769,488	504,851	470,639				
396,132	937,085	847,498	805,037	705,960					
440,832	847,631	1,131,398	1,063,269						
359,480	1,061,648	1,443,370							
376,686	986,608								
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 PCS model 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 PCS model, indicated by an almost 50% reduction in process standard deviation, and a parameter standard deviation greater by almost 70%.

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.

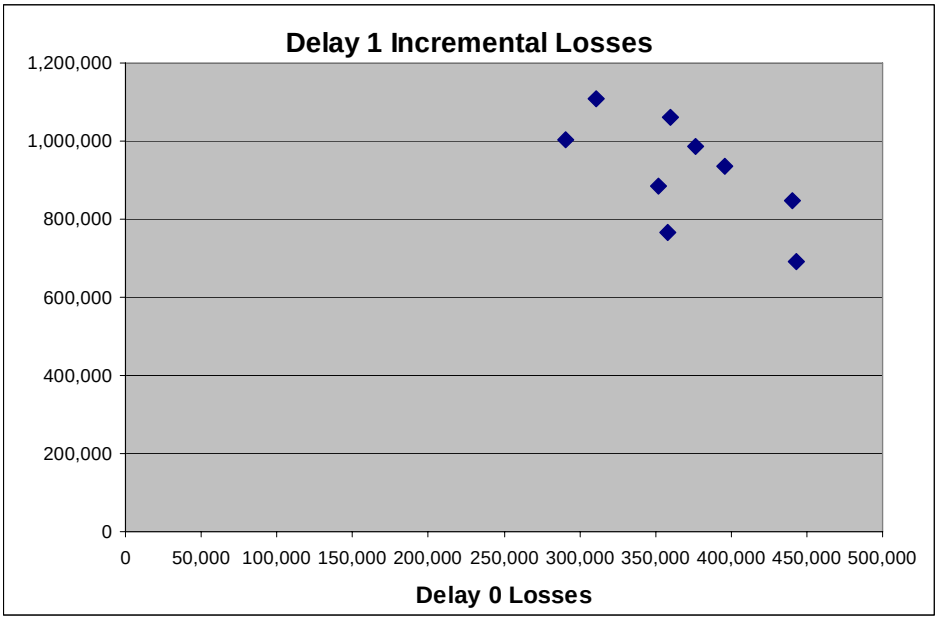

Figure 1

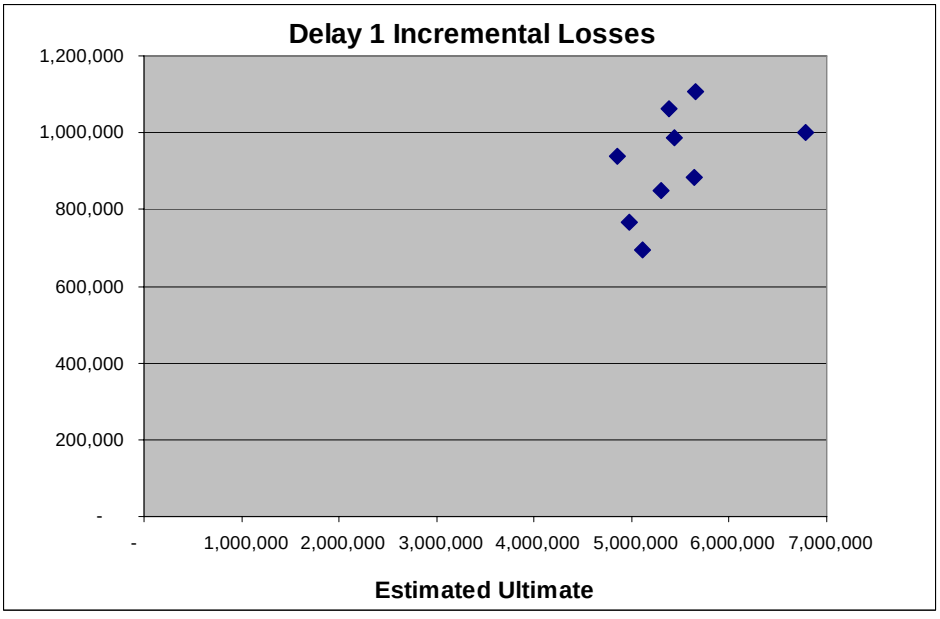

Figure 2

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

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 year of origin is built into the factor approach, as each year gets its own level. But there can appear to be inflation by year of origin that is actually generated by calendar year inflation. If the latter varies by year, a pattern of high and low residuals can show up by diagonal. A large variation in residuals among diagonals would suggest that either calendar year inflation or claim department variation is operating.

Table 1 shows the residuals by diagonal for the 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 1

A related issues is correlation of residuals between columns. This can be a result of diagonal effects that have not been modeled. Table 2 shows the correlation of the 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 2

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 becomes:

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

The unconstrained parameter estimates still have an iterative formulation:

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

These converge a bit slowly, but 50 or so iterations often suffices. 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 7^{th} diagonal, and for the 6^{th} and 7^{th}. To compare the loglikelihoods, *b* was set at 37,183.5. This is the estimated value for

another 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 is rather freaked out by having so many parameters (up to 21) with only 55 observations, and penalizes the first diagonal parameter by 2.5 and the second by 2.65. This says that the 2nd parameter is clearly not worth it, but the first one still is. The factors for the 6th and (in both models) 7th diagonal are 1.136 and 0.809.

Having the diagonal parameters corrects for random errors in the row and column parameter estimates. 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.

REDUCING THE NUMBER OF PARAMETERS

The number of parameters in the PCS model is uncomfortably high. There are a few methods available for reducing the number of parameters without hurting the goodness of fit too greatly. First, parameters that are fairly close to each other can be set equal. Also, when things are changing systematically, a parameter for one 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 could reduce the number of parameters further.

Reducing the parameters in these ways can eliminate distinctions that are not supported by the data. Every year gets its own level if the model allows it, but the differences between some years could be small compared to the variability in the possible parameters for each year. The same holds for the distribution by lag and the diagonal effects.

After reviewing some of these possibilities, we settled on the following model. Accident year 0 is low and 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 can be divided into high and low payment years with parameters g_a and g_b. Delay 0 is a low year as things are just getting warmed up. Delays 1, 2, and 3 are where most of the action is and all get g_b. Delays 5, 6, 7, and 8 are again low years 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. Clarke suggested using parameterized distributions to describe the payout pattern. We tried Weibull and loglogistic distributions conditional on being less than or equal to 9. The loglogistic was better than the Weibull but not as good as the high-low model (all with two parameters) in terms of loglikelihood.

It was efficient enough to identify three of the diagonals as high or low di-

agonals, getting factors 1 + c or 1 – c. The 7th diagonal was low and the 4th and 6th were high. Thus only one diagonal parameter was 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 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 3

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:

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

To get 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. Appendix 1 summarizes the 2nd partials of the loglikelihood for the six-parameter model.

The correlations of adjacent residuals improve a great deal with the diagonal parameters, as shown in Table 4. 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.

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 4

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 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:

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 5

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%

TESTING THE VARIANCE ASSUMPTION

In the PCS model the variance of each cell is *b* times its mean. For many loss processes though 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.

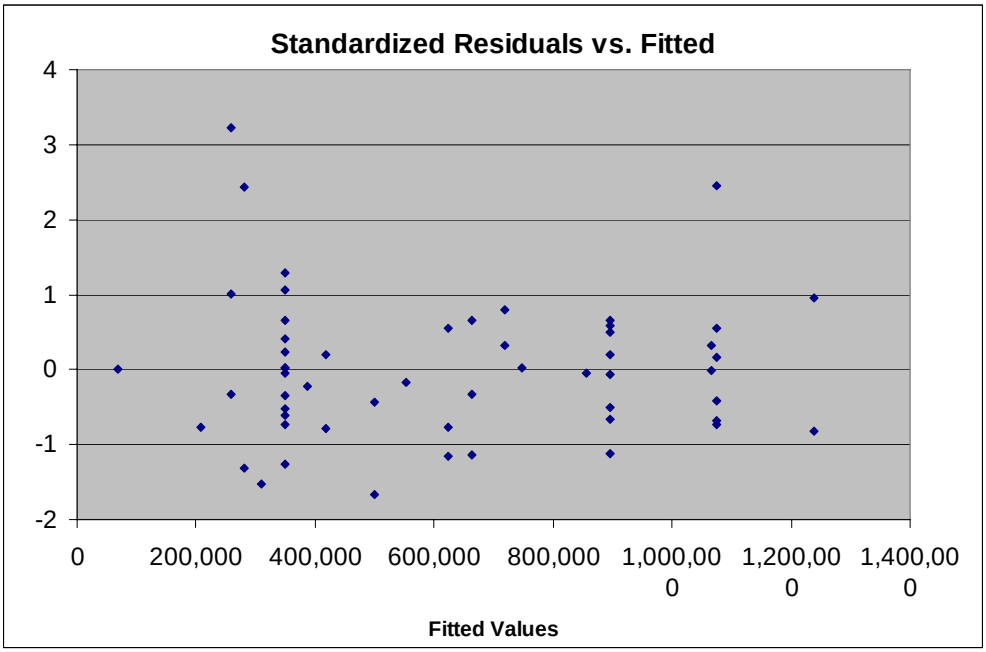

Figure 3

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 ½ 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 rank/(N+1) to each. This gives something like a PP-plot, and is shown in Figure 4 for the six-parameter model.

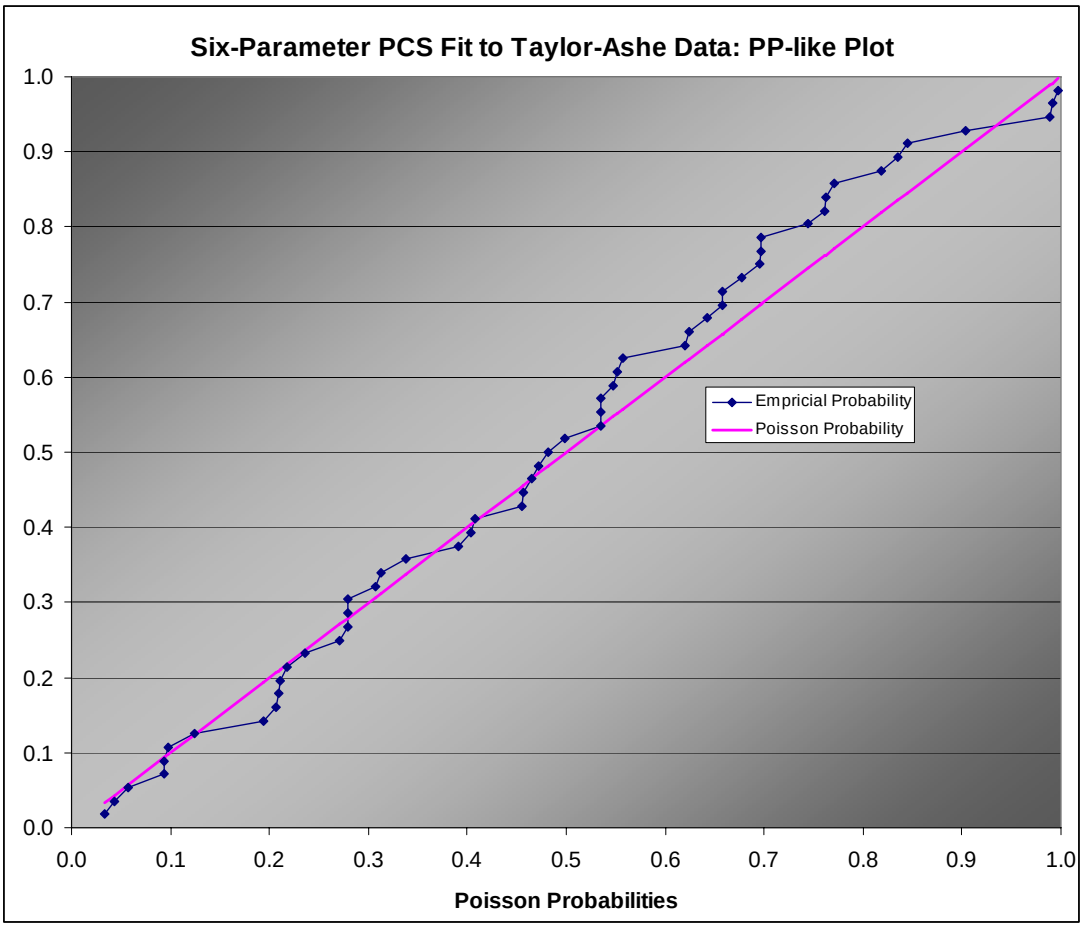

Figure 4

The fit is not too bad, but is better below the median than above. There, 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.

CONCLUSIONS

The PCS model when given 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 seems to have 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. It should be possible in most cases to reduce the number of parameters in the models through the use of trends, combination of similar parameters, etc. Although not discussed here, both calendar year and delay parameters can be reduced in the chain-ladder paradigm. The PCS model also allows for eliminating some accident year parameters, which can be reduced even to a single parameter as in the Cape Cod case. In the example here, three levels sufficed for 10 years. Many other possible models have not been considered here and may give better fits to this data. 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.

REFERENCES

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

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

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

Kremer, E 1985. *Einfuhrung in die Versicherungsmathematik* Vandenhoek & Ruprecht, Gottingen

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.

Panning, W 2006. Measuring Loss Reserve Uncertainty, PCAS, to appear.

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

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

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

Appendix 1 – 2nd Partials for Six Parameter Model of Taylor-Ashe Data

These can be written in terms of the following unconstrained second partials:

$$\frac{\partial^2 l}{\partial U_w^2} = -\sum_{d=0}^{9-w} \frac{q_{w,d}}{b U_w^2} \quad ; \quad \frac{\partial^2 l}{\partial g_d^2} = -\sum_{w=0}^{9-d} \frac{q_{w,d}}{b g_d^2} \quad ; \quad \frac{\partial^2 l}{\partial h_j^2} = -\sum_{w+d=j} \frac{q_{w,d}}{b h_j^2} \quad ;$$

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

For this model, $U_1 = \ldots = U_5 = U_8 = U_9 = U_a$ and $U_6 = ½ [U_a + U_7]$. Also $g_0 = g_5 = g_6 = g_7 = g_8 = g_a$; $g_1 = g_2 = g_3 = g_b$, $g_4 = ½ [g_a + g_b]$; and $g_9 = 1 – 5.5g_a – 3.5g_b$. Finally $h_7 = 1 – c$ and $h_6 = h_4 = 1 + c$, o.w. $h_j = 1$. For notation sake here, let $W_a = \{1,2,3,4,5,8,9\}$, $G_a = \{0,5,6,7,8\}$, $G_b = \{1,2,3\}$, $C = \{4,6,7\}$, $h_j = 0$ for $j>9$ and $U_w = g_d = 0$ for w or $d < 0$.

From these we get, in terms of the unconstrained partials,

$$\frac{\partial^2 l}{\partial U_a^2} = ¼ \frac{\partial^2 l}{\partial U_6^2} + \sum_{w \in W_a} \frac{\partial^2 l}{\partial U_w^2} \quad ; \quad \frac{\partial^2 l}{\partial U_7^2} = \textit{unconstrained value} + ¼ \frac{\partial^2 l}{\partial U_6^2}$$

$$\frac{\partial^2 l}{\partial g_a^2} = ¼ \frac{\partial^2 l}{\partial g_4^2} + \sum_{d \in D_a} \frac{\partial^2 l}{\partial g_d^2} + 5.5^2 \frac{q_{0,9}}{b g_9^2} \quad ; \quad \frac{\partial^2 l}{\partial g_b^2} = ¼ \frac{\partial^2 l}{\partial g_4^2} + \sum_{d \in D_b} \frac{\partial^2 l}{\partial g_d^2} + 3.5^2 \frac{q_{0,9}}{b g_9^2}$$

$$\frac{\partial^2 l}{\partial c^2} = \sum_{j \in C} \frac{\partial^2 l}{\partial h_j^2}$$

$$\frac{\partial^2 l}{\partial U_a \partial U_7} = \tfrac{1}{4} \frac{\partial^2 l}{\partial U_6^2} \quad ; \quad \frac{\partial^2 l}{\partial g_a \partial g_b} = \tfrac{1}{4} \frac{\partial^2 l}{\partial g_4^2} + (3.5)(5.5) \frac{q_{0,9}}{b g_9^2}$$

$$\frac{\partial^2 l}{\partial U_0 \partial g_a} = -\sum_{d \in D_a} \frac{h_d}{b} - \tfrac{1}{2} \frac{h_4}{b} + 5.5 \frac{h_9}{b} \quad ; \quad \frac{\partial^2 l}{\partial U_0 \partial g_b} = -\sum_{d \in D_b} \frac{h_d}{b} - \tfrac{1}{2} \frac{h_4}{b} + 3.5 \frac{h_9}{b}$$

$$\frac{\partial^2 l}{\partial U_7 \partial g_a} = -\frac{h_7}{b} - \frac{h_6}{2b} \quad ; \quad \frac{\partial^2 l}{\partial U_7 \partial g_b} = -\frac{2}{b}$$

$$\frac{\partial^2 l}{\partial U_a \partial g_a} = -\sum_{d \in D_a, w \in W_a} \frac{h_{w+d}}{b} - \tfrac{1}{2} \left(\sum_{d \in D_a} \frac{h_{6+d}}{b} + \sum_{w \in W_a} \frac{h_{4+w}}{b} \right)$$

$$\frac{\partial^2 l}{\partial U_a \partial g_b} = -\sum_{d \in D_b, w \in W_a} \frac{h_{w+d}}{b} - \tfrac{1}{2} \left(\sum_{d \in D_b} \frac{h_{6+d}}{b} + \sum_{w \in W_a} \frac{h_{4+w}}{b} \right)$$

$$\frac{\partial^2 l}{\partial U_0 \partial c} = -\frac{g_a + g_b}{2b} \quad ; \quad \frac{\partial^2 l}{\partial U_7 \partial c} = \frac{g_a + g_b}{2b} \quad ; \quad \frac{\partial^2 l}{\partial U_a \partial c} = -\frac{3g_a + 9g_b}{2b}$$

$$\frac{\partial^2 l}{\partial g_a \partial c} = -\tfrac{1}{2} \frac{U_0 + U_a - U_7}{b} \quad ; \quad \frac{\partial^2 l}{\partial g_b \partial c} = -\tfrac{1}{2} \frac{U_0 + 7U_a - U_7}{b}$$

The diagonals are 1 in the projection period, so the unconstrained derivative of the reserve wrt g_d is still $\Sigma_{w>n-d} U_w$ and wrt U_w is $\Sigma_{d>n-w} g_d$. Thus

$\frac{\partial R}{\partial g_a} = \sum_{d=0}^{n} \left(\sum_{w=n+1-d}^{n} U_w \, \partial g_d \big/ \partial g_a \right)$ and $\frac{\partial R}{\partial U_a} = \sum_{w=0}^{n} \left(\sum_{d=n+1-w}^{n} g_d \, \partial U_w \big/ \partial U_a \right)$, etc.