Modeling Variance in Reserves

Gary Venter

Abstract Typically very simple single-parameter models are used for cell variances in reserving. This could be contributing to misstatement of runoff risk. We look at more choices available for single-parameter models and try a two-factor model of variance by cell for greater specificity.

1 Introduction

Each cell in a reserve dataset usually gets its own mean m_i in reserve models, but usually only one parameter is used for the variance of all the cells. For instance if the cells are modeled by a normal distribution, the distribution for cell i typically would have $\mu_i = m_i$ and $\sigma_i^2 = s^2$. If this does not work well, the model is called heteroskedastic. A possible response in that case could be to assign somewhat more variance to the cells with larger means, for example.

In the case where the cell means m_i are all positive, one way to do this could be to set $\mu_i = m_i$ and $\sigma_i^2 = s\sqrt{m_i}$. Here s is still constant across all the cells but now the cell variances are not.

The same method could be applied to other distributions. Take the case of a gamma distribution with mean = $\alpha\beta$ and variance = $\alpha\beta^2$. If for cell i you set $\alpha_i = m_i^{1.5}/s$ and $\beta_i = sm_i^{-0.5}$, you get a gamma distribution for that cell with mean = m_i and variance = $s\sqrt{m_i}$.

The gamma has variance = mean$^2/\alpha$. This relationship still holds in every cell, but α varies across the cells. What also holds for the gamma, and in every cell, is skewness = 2CV, where CV is standard deviation divided by mean. But now the cells with lower mean have higher CV and so higher skewness.

This modeling of variances uses common distributions like the normal or gamma for each cell, but instead of making some parameter of the distribution constant

Gary Venter
University of New South Wales, e-mail: gary.venter@gmail.com

across all the cells, both parameters vary by cell, with a specified mean – variance relationship constant across the cells.

2 What is a reasonable mean – variance relationship across cells in a loss triangle?

A typical loss triangle has aggregate losses in the cells, so is a combination of number of claims and claim size. In the case of a paid loss triangle that would be the number of claims with payments in the cell and the size of the payments. It is pretty common for larger claims to pay later. So if there are not too many partial payments, the later cells in the triangle would have much lower frequency but higher severity than the earlier cells.

The variance of aggregate losses is generally proportional to the mean of frequency and the square of the mean of severity. Thus the later cells would have both aggregate mean and variance going down by the frequency mean, and mean going up by the increase in severity, and variance going up by the square of the mean of severity. Thus the variance would not be going down as fast as the mean would be.

One way to model this would be to make the variance of the cells proportional to a power of the mean that is less than unity.

3 Two-parameter variance model

As seen in the introduction, the variance can be modeled as proportional to a selected power of the mean. But what exactly that power should be is not clear. To address this, that power itself can be specified as a parameter. This would model the variance of a cell using two parameters r and s, with the variance of cell i:

$$variance_i = sm_i^r \tag{1}$$

Modeling the mean – variance relationship in this way allows the choice of distribution to be based on other shape characteristics. For several distributions the skewness highlights the shape differences. Especially for one or two-parameter distributions, the skewness can be expressed as a function of the CV, as follows.

- Normal: skewness = 0
- Poisson: skewness = CV
- Gamma: skewness = 2CV
- Inverse Gaussian: skewness = 3CV
- Lognormal: skewness = (3+CV^2)CV
- Inverse gamma: skewness = 4CV/(1–CV^2) for CV < 1 and infinite otherwise

For CV < 1, the lognormal skewness is always less than 4CV, and the inverse gamma skewness is always greater than this. For higher CV, the lognormal skewness can be quite large, but is always finite. Thus a lognormal always has a lower skewness than the inverse gamma with the same mean and variance.

For these distributions, solving for the cell parameters to get $variance_i = sm_i^r$, and simulation, are discussed next.

3.1 Normal

For the normal for cell i, set the parameters as $\mu_1 = m_i$ and $\sigma_i^2 = sm_i^r$. Simulation is easy with inverse normal functions, which are now ubiquitous.

3.2 Continuous Scaled Poisson (CSP)

[**?**] and [**?**] discuss a continuous version of the Poisson, with a scale parameter. This is more useful for reserve applications than the ODP from the exponential family, which has positive probability only at integer multiples of the scale parameter. The CSP density is:

$$f(x;\mu,\theta) = \frac{e^{-\mu/\theta}(\mu/\theta)^{x/\theta}}{\theta\Gamma(1+x/\theta)} \tag{2}$$

It also has a probability mass at zero which is positive but less than the corresponding Poisson probability at zero. Its mean and variance are very close to μ and $\theta\mu$, with unnoticeable deviations from these values except for very small means. The skewness is also very close to the CV of $(\theta/\mu)^{0.5}$.

For cell i, setting $\mu_i = m_i$ and $\theta_i = sm_i^{r-1}$ gives mean m_i and variance sm_i^r.

Actuaries sometimes simulate the ODP using the gamma with the same mean and variance, but it would probably be more accurate to average a normal and a gamma percentile with the same probability.

3.3 Gamma

The gamma distribution is commonly parameterized to have mean = $\alpha\beta$ and variance = $\alpha\beta^2$. Setting $\alpha_i = m_i^{2-r}s^{-1}$ and $\beta_i = sm_i^{r-1}$ gives mean m_i and variance sm_i^r. For simulation, the inverse gamma function is now widely available.

3.4 Inverse Gaussian

The inverse Gaussian is basically a heavier-tailed gamma distribution. It's density is often parameterized as:

$$f(x;\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} e^{\frac{-\lambda(x-\mu)^2}{2\mu^2 x}} \tag{3}$$

Then it has mean = μ and variance = μ^3/λ. Setting $\mu_i = m_i$ and $\lambda_i = m_i^{3-r}s^{-1}$ gives mean = m_i and variance = sm_i^r.

Simulation takes two random draws and is a bit intricate. See Wikipedia's article on the inverse Gaussian for details.

3.5 Lognormal

The lognormal mean for cell i is $m_i = e^{\mu_i + 0.5\sigma_i^2}$. Thus set $\mu_i = log(m_i) - 0.5\sigma_i^2$. The variance is $m_i^2\left(e^{\sigma_i^2} - 1\right) = sm_i^r$. From this it follows that $\sigma_i^2 = log(1 + sm_i^{r-2})$. Simulation of the lognormal is done by exponentiating a normal simulation with the same μ and σ.

3.6 Inverse Gamma

The inverse gamma is the distribution of $1/X$ where X is gamma distributed. Its density is:

$$f(x;\theta,\alpha) = \frac{\theta^\alpha e^{-\theta/x}}{x^{\alpha+1}\Gamma(\alpha)} \tag{4}$$

For any $z < \alpha$, including negative numbers and not limited to integers, the z^{th} moment is:

$$E(X^z) = \frac{\theta^z \Gamma(\alpha - z)}{\Gamma(\alpha)} \tag{5}$$

This makes the mean = $\frac{\theta}{\alpha-1}$, the variance = $\frac{\theta^2}{(\alpha-1)^2(\alpha-2)}$, and the CV = $\frac{1}{\sqrt{\alpha-2}}$.

Taking $\alpha_i = 2 + m_i^{2-r}s^{-1}$ and $\theta_i = m_i + m_i^{3-r}s^{-1}$ gives mean m_i and makes the variance = sm_i^r for the i^{th} cell.

The survival function $S(x) = 1 - F(x)$ is given by the gamma distribution function at $1/x$ with parameters α and θ. In Excel, then, $S(x)$ can be calculated as gamma.dist(1/x, α, θ, 1), and simulation a of x is 1/gamma.inv(rand(), α, θ). (Simulation of $S(x)$ is equivalent to simulation of $F(x)$ because for a uniform(0,1) variable U, $1-U$ is also uniform(0,1)).

4 Distribution Shapes

Some things that characterize the shapes of distributions include the heaviness of the right tail, behavior near zero, and the mode. The skewness of a distribution here is an indication of the heaviness of the right tail. All of the distributions except the CSP and gamma always have a non-zero mode. Behavior near zero can be classified in three classes by (the limit of) the density at zero – whether it is zero, positive, or goes to infinity. The case with density zero can be further subdivided into density asymptotic to the x-axis, to the y-axis, or neither.

The inverse Gaussian, lognormal, and inverse gamma distributions always are zero at zero with density asymptotic to the x-axis. So they always have positive modes as well. Thus their basic shapes do not depend on the parameters, and letting both distributional parameters vary across cells does not make the shapes different cell by cell. This is basically true of the normal and Poisson as well. However the gamma is an exception.

The classes of shapes at zero for any positive distribution actually correspond to which negative moments exist, i.e., $E(X^z)$ for $z < 0$. If the -2^{nd} moment exists, the distribution is zero at zero and asymptotic to the x-axis. If the -1^{st} moment does not exist, the density goes to infinity at zero. The other shapes at zero are between these in terms of existence of moments.

For the gamma, the $-z^{th}$ moment does not exist for $z > \alpha$. This then leaves open the possibility that when the gamma distribution is used, different cells in the data could end up with different distribution shapes, some with a positive mode and some with a zero mode. For a specific degree of skewness the gamma could be the best fit, so in that case this is something to be aware of.

Also note that in each cell the standard moment relationships hold. For instance for an inverse Gaussian distribution, the variance is μ_i^3/λ_i in every cell. This can be useful for computing moments, for instance.

However the relationship of moments across cells is a different matter. There is no clear justification for this to follow the moment relationships within the cells. Before modern computing it was helpful to assume simple variance relationships across cells, for instance to estimate by quasi-likelihood, but there is not much reason to maintain such limitations these days. The method outlined here allows empirical estimation of variances across cells.

5 Example

We illustrate the methodology by fitting the model above for the six distributions separately for a reserve triangle from the literature. The data in Table 1 is from [**?**] and has been used as sample data in a number of papers.

Taylor and Ashe themselves fit a six-parameter model using factors by accident year, lag, and diagonal. Because they had claim count information for each cell (only shown here for row totals), they were able to use operational time (fraction of

claims settled) instead of lag. We have found that to be quite useful in other studies. However here we present a more usual reserving model for the sake of example.

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									

The model we use is from [**?**], and also has six parameters which are factors by accident year, lag, and diagonal. The parameters are used to create 10 row factors $U_1,...,U_{10}$, 10 column factors $g_0,...,g_9$, and 10 diagonal factors $d_1,...,d_{10}$. The mean of the i,j cell is $m_{i,j} = U_i g_j d_{i+j}$. The factors are computed by:

- $U_1,...,U_{10} : U_1, U_a, U_a, U_a, U_a, U_a, (U_a+U_8)/2, U_8, U_a, U_a$
- $g_0,...,g_9 : g_a, g_b, g_b, g_b, (g_a+g_b)/2, g_a, g_a, g_a, g_a, 1-5.5g_a-3.5g_b$
- $d_1,...,d_{10} : 1,1,1,1,1+c,1,1+c,1-c,1,1$

[**?**] assumes a CSP distribution for residuals and finds this model to be better than other ones tried, as measured by parameter-penalized loglikelihood. This is due to a reasonably good fit and the small number of parameters. Here we assume cell variances of $sm_{i,j}^r$ and so fit two more parameters s,r for each of the six distributions detailed above.

With only eight parameters and good starting values, Excel's Solver was able to fit these six models by MLE. The fitted parameters and the resulting negative loglikelihoods for each distribution are:

.

Distribution	U_1	U_8	U_a	g_a	g_b	c	s	r	NLL
Normal	3,720,805	7,028,052	5,106,002	0.0671	0.178	0.204	108,302,367	0.383	725.64
CSP	3,717,265	7,199,259	5,180,590	0.0669	0.174	0.206	27,466,624	0.482	723.81
Gamma	3,695,077	7,288,171	5,209,799	0.0667	0.173	0.210	29,043,757	0.477	722.36
Inverse Gaussian	3,673,158	7,356,946	5,221,355	0.0668	0.172	0.215	53,984,497	0.432	721.55
Lognormal	3,665,649	7,336,085	5,210,018	0.0668	0.172	0.214	49,661,850	0.439	721.60
Inverse Gamma	3,650,508	7,384,349	5,216,918	0.0672	0.172	0.218	272,803,497	0.317	721.44

5.1 Findings

The parameters for the cell means are pretty constant across the distributions. All the *r* power parameters are between 0.3 and 0.5, suggesting that there is indeed a tendency for the larger losses to pay later. While a power below unity is not unexpected, this is surprisingly low. Perhaps this dataset has a quite strong tendency for large losses to pay later.

Interestingly enough, the Taylor-Ashe dataset comes with enough detail to create count and average severity triangles. Below is the one for average severity by cell. It

E(Severity)	0	1	2	3	4	5	6	7	8	9
1	8,946	6,185	3,889	5,193	3,740	26,109	10,453	13,995	75,743	33,974
2	9,517	4,753	7,184	4,951	7,307	12,346	22,948	44,362	70,841	
3	8,300	6,341	3,812	6,645	15,642	5,651	35,428	56,081		
4	7,576	7,150	3,561	15,624	4,067	20,709	34,381			
5	14,772	3,707	5,976	6,412	9,179	36,203				
6	12,004	7,745	4,154	9,253	19,080					
7	13,776	7,371	7,749	10,323						
8	8,360	9,564	17,390							
9	22,158	10,724								
10	15,637									

does show a strong trend to larger losses in the later columns. The better fit for the more skewed distributions does not come from looking at the severity distributions. It is inferred from the degree of dispersion of cell observations from their means.

Since all the distributions are using the same model for the means, and the same mean – variance relationship, the differences in fit are only from other shapes features of the distributions. This suggests that not much difference in NLL would be anticipated. Still there is a clear trend in lower NLLs for the more skewed distributions. The combination of a high skewness with a low power *r* seems consistent with more severity at the end, but would not show up in other modeling approaches.

The last three NLLs are so close that they are virtually identical. The lognormal is the easiest to simulate, so we focus on it. The CVs by cell are shown below. The

CV	0	1	2	3	4	5	6	7	8	9
1	44%	21%	21%	21%	24%	44%	38%	53%	44%	82%
2	33%	16%	16%	14%	21%	29%	40%	33%	33%	
3	33%	16%	14%	16%	18%	40%	33%	33%		
4	33%	14%	16%	14%	25%	33%	33%			
5	29%	16%	14%	19%	21%	33%				
6	33%	14%	19%	16%	21%					
7	25%	17%	14%	14%						
8	31%	12%	12%							
9	33%	16%								
10	33%									

bigger means end up with lower CVs in this model. Lag zero has low payments in this data, so is comparable to later lags. The diagonal effects show up mildly here. The larger CVs on the right seem appropriate for the larger severities.

The average loglikelihoods are about constant across the columns. A lower loglikelihood in a cell means a higher probability for that observation, which means that the observation is in the part of the fitted distribution that has the most probability. Thus having similar values across the columns implies an equally good fit across the columns. There is a trend toward better fits in the last few columns, but with fewer cells to fit this is easier to do and so probably does not mean much.

Loglikelihood by lag: 12.7 13.2 13.1 13.6 13.2 13.5 13.8 12.4 12.7 11.8

6 Application to Reserve Distributions

Using the formulas above, it is easy enough to simulate the runoff. But first parameter uncertainty would need to be quantified. The most direct way to do this may be to estimate the parameters by MCMC. Then a set of simulated parameters comes along with the estimation, and these have a distribution and correlations that are consistent with the estimation.

When fitting by MLE, parameter estimation variances and correlations come from the information matrix, which is calculated from the 2nd partial derivatives of the loglikelihood at its maximum. Asymptotically, the estimates are multivariate normally distributed with this covariance matrix.

However if there is enough probability that a positive parameter could be negative, probably the asymptotic result does not hold. An alternative in that case could be to simulate parameters from a multivariate gamma with a normal copula. This distribution asymptotically approaches the multivariate normal.

For a two-parameter severity, the bivariate normal copula with correlation ρ can be simulated by simulating two random draws u and v, then replacing v by $w = \Phi(\rho\,\Phi^{-1}(u) + (1-\rho^2)^{0.5}\Phi^{-1}(v))$.

One possible problem with modeling the variance as sm^r is that if a cell is smaller simply because it is at a lower inflation level, it should not necessarily get a higher CV. Adjusting the data to a common price level before the analysis is an approach to this problem. Alternatively, the price level could be an explanatory variable in the model.

Using m before the application of the diagonal factor is another possibility. However the diagonal factors mostly pick up changes in inflation rates. A constant inflation rate would be projected onto the row and column parameters.

7 Conclusion

With just one more distribution parameter, cell variance and skewness can be modeled in greater detail. This probably would not change reserve estimates but could improve risk analysis of reserve outcomes.