Distribution and Value of Reserves Using Paid and Incurred Triangles

Gary G. Venter, FCAS, MAAA

Abstract

Many loss reserving models are over-parameterized yet ignore calendar-year (diagonal) effects. Venter [1] illustrates techniques to deal with these problems in a regression environment. Venter [2] explores distributional approaches for the residuals. Gluck [3] shows that systematic effects can increase the reserve runoff ranges by more than would be suggested by models fitted to the triangle data alone. Quarg and Mack [4] show how to get more information into the reserve estimates by jointly using paid and incurred data.

This paper uses the basic idea and data from [4] and the methods of [1] to build simultaneous regression models of the paid and incurred data, including diagonal effects and eliminating non-significant parameters. Then alternative distributions of the residuals are compared in order to find an appropriate residual distribution. To get a runoff distribution, parameter and process uncertainty are simulated from the fitted model. The methods of Gluck [3] are then applied to recognize further effects of systematic risk.

Once the final runoff distribution is available, a possible application is estimating the market value pricing of the reserves. Here this is illustrated using probability transforms, as in Wang [5].

Keywords. Reserving Methods; Reserve Variability; Uncertainty and Ranges, Fair Value, Probability Transforms, Bootstrapping and Resampling Methods, Generalized Linear Modeling.

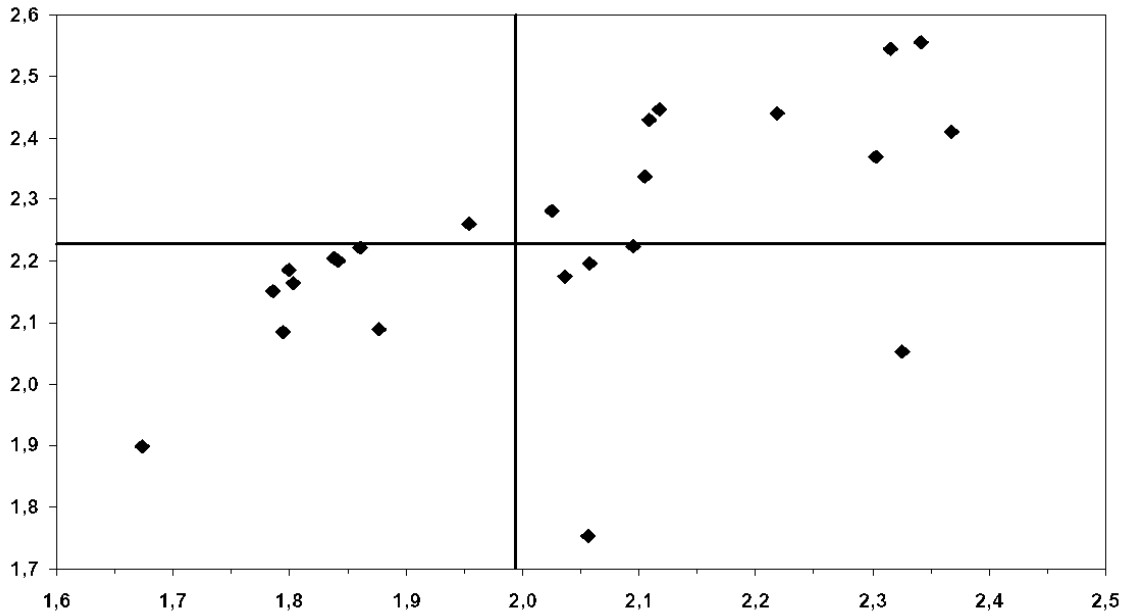
1 INTRODUCTION

Actuaries have used many methods for reconciling reserve estimates from paid and incurred triangles for decades, but formal modeling of paid and incurred simultaneously appears to have begun with Halliwell [6]. His approach was to fit regression models to both data triangles with constraints on the coefficients of both models. More recently Quarg and Mack [4] argue that a high paid-to-incurred ratio for an accident year/lag combination is suggestive of higher-than-average incurred development and lower-than-average paid development in the next period. For instance, some paid factors compared to incurred/paid ratios from http://www.actuaries.org/ASTIN/Colloquia/Zurich/Mack_presentation.pdf are reproduced in Figure 1. In [4] the development factors for paid and incurred are adjusted using these ratios. The formulas are available in Mack [7], who also provides a comparable adjustment for multiplicative cross-classified models.

Verdier and Klinger [8] suggest a modified scheme that recognizes that the impact of the incurred/paid ratios reduces in later stages of development. They also calculate the variance of the re-

sult, and suggest a multi-line extension. Jedlicka [9] studies alternative estimation procedures, and outlines a model of paid and unpaid losses instead of paid and incurred, where unpaid = incurred – paid.

Figure 1 – Individual Paid Development Factors as a Function of Incurred/Paid Ratio



The current paper models paid and incurred triangles using a regression framework, but does not fix the explanatory variables in advance. Rather it is left to the modeler to decide, based on regression diagnostics, which variables best explain each triangle's observations. The regressions are set up with incremental losses as the dependent variable, since these are the new elements that need explanation at each lag. Previous incurred, paid, and unpaid losses, cumulative or incremental, are allowed as independent variables for both triangles. Also diagonal dummies are allowed, in case there are diagonal (i.e., calendar year) effects in the triangles. None of the papers cited above include calendar-year effects, although these are common in development triangles.

Regression modeling is both an art and a science. It is not a model, but a way to build models. Here it is applied to building models of loss development triangles, but many of the issues are more general. The key issue in building regression models is what variables to include. With generalized linear models, another issue becomes what distribution best describes the residuals, and non-linear functions of the regression result become possible.

One criterion for evaluating regression models is the significance of the variables. Typically significance at the 5% level is sought, which is often close to requiring that the estimate be at least

twice its standard error. Sometimes this is relaxed a bit, perhaps to the 10% level. Another useful statistic is the standard error of the regression. That incorporates a penalty for additional parameters, so can increase when insignificant variables are added. Usually variables significant at even the 10% level will improve the standard error. The adjusted-R^2 is similarly penalized but it can be difficult to tell if a slight increase is worthwhile. Also there seems to be some ambiguity as to how it is defined for no-constant regressions, which are common in reserve analysis.

These ideas are used to build and apply regression models for reserves when both paid and incurred triangles are available. The data for the continuing example for this paper consists of Tables 1 and 2 from Quarg and Mack [4], and Table 3, which is their difference.

Table 1 – Paid Cumulative Losses

Acc. / Dev	**0**	**1**	**2**	**3**	**4**	**5**	**6**
0	576	1804	1970	2024	2074	2102	2131
1	866	1948	2162	2232	2284	2348	
2	1412	3758	4252	4416	4494		
3	2286	5292	5724	5850			
4	1868	3778	4648				
5	1442	4010					
6	2044						

Table 2 – Incurred Cumulative Losses

Acc. / Dev	**0**	**1**	**2**	**3**	**4**	**5**	**6**
0	978	2104	2134	2144	2174	2182	2174
1	1844	2552	2466	2480	2508	2454	
2	2904	4354	4698	4600	4644		
3	3502	5958	6070	6142			
4	2812	4882	4852				
5	2642	4406					
6	5022						

Quarg and Mack suggest that using paid and incurred triangles together can help reconcile their differences and improve the reserve estimates from both. In his discussion at the 2003 ASTIN Colloquium in Berlin, Mack suggested that this could also be done in a regression setting, where both the paid and incurred losses could be used in the regressions for either. This paper follows up on that suggestion, also incorporating the methods of Venter [1] to eliminate statistically insignificant variables and to incorporate any diagonal effects that may be in the data. Alternative distributions

for the residuals are also fit. These are the topics of section 2.

Table 3 – Losses Estimated to Be Unpaid at Year-End

Acc. / Dev	0	1	2	3	4	5	6
0	402	300	164	120	100	80	43
1	978	604	304	248	224	106	
2	1492	596	446	184	150		
3	1216	666	346	292			
4	944	1104	204				
5	1200	396					
6	2978						

Section 3 addresses the issue of runoff ranges arising from the models developed in section 2. Section 4 widens the runoff ranges to include systematic risk, as discussed by Gluck [3]. Section 5 discusses uses for the resulting distribution, and in particular proposes a method to use the runoff distribution to estimate the value of the reserves.

2 BUILDING MODELS

2.1 Exploratory Analysis

To paraphrase Yogi Berra, you can see a lot about your data just by observing it. The starting point of building a regression model is to explore the relationships that may be in the data. This is what makes this approach difficult to reduce to a strict algorithm, however. Some of the steps that can be used in looking at paid and incurred development triangles are outlined below.

Modeling paid losses as a function of paid and incurred could also include using unpaid losses as an explanatory variable, as unpaid is just the difference between incurred and paid. The first step in this analysis is to look at the data and explore relationships that may exist.

One thing that stands out in the unpaid triangle is that the lag 0 loss for the most recent year is more than double that of any previous year. The incurred is also at an unprecedented level, but the paid is not. That raises a question as to whether or not the latest year represents a significant increase in exposure, or is just an unusual fluctuation. The paid chain ladder estimate for ultimate for year 6 is 6128, compared to 8429 for incurred development, or a difference of 2301. Usually an analyst would know more about the business reasons for such a difference. For instance, there could have been a significant increase in premium volume, or a major loss event, or, on the other hand, a change in

reserving methodology that does not affect paid losses. Without such background, only historical data patterns can be used on this point, even though it is quite a bit out of the range of historical observations.

The paid losses could be modeled as a function of the previous incurred, paid, or outstanding, or some combination of those. Here the incremental paid losses at each lag are modeled, as that is the new information at that lag. To start the analysis, the correlations of the paid losses with the previous unpaid and the previous cumulative paid and incurred for the continuing example are shown in Table 4.

Table 4 – Correlation % for Incremental Paid Losses with Previous Cumulative Losses

Incremental Paid at Lag with:	**Incurred**	**Paid**	**Unpaid**
Paid at Lag 1	88	84	70
Paid at Lag 2	68	57	92

Table 4 shows that the lag1 incremental paids correlate most strongly with the previous incurreds, while at lag 2, the correlation is strongest with previous unpaid. At later lags (not shown) the unpaid continue to be strong predictors of the next incremental payments, and interestingly enough, after a few years the percentage of unpaid that is paid in the next year is fairly steady, as shown in Table 5, which is calculated as the sum of paid divided by the sum of previous unpaid column by column. The high factor at lag 1 reflects the continuing reporting of claims after lag 0. The similar factors after lag 2 suggest that only one parameter will be needed for the later lags.

Table 5 – Average Paid at Each Lag as Factor Times Previous Unpaid (Sum/Sum)

Lag	**1**	**2**	**3**	**4**	**5**	**6**
Percent	1.95	0.67	0.33	0.33	0.28	0.36

Since unpaid losses are a strong predictor of the next period's paid losses, a model for projecting future unpaid is needed to fill out the triangles. Unpaid can be calculated from models for incurred and paid losses, or could be modeled directly, say by expressing expected unpaid as a factor times previous unpaid.

In the continuing example, the unpaid losses at lags 1 and 2 have a stronger correlation with previous cumulative paid losses than with previous incurred losses (61% vs. 52% for lag 1 and 47% vs. 42% for lag 2). Preliminary regressions indicated that for lag 1, current incremental paid was significant, but not for lag 2. Also a constant term was significant for lag 1.

For the later lags, the fairly constant ratio of paid to previous unpaid would suggest the same for

unpaid to previous unpaid. Due to changing incurred development, this was fairly noisy, however. Table 6 shows the unpaid losses at each later lag as a percentage of the unpaid at the previous lag. There is no clear trend, so it may work to model these all as a single constant percentage, especially given that pattern for paid losses.

Table 6 – Unpaid Losses as a Percent of Previous Unpaid

Acc. / Dev	3	4	5	6
0	73	83	80	54
1	82	90	47	
2	41	82		
3	84			

2.2 Regression Analysis

The entire triangle can be set up as a single large regression analysis, either for paid or unpaid losses. This is in effect a series of regressions combined into a single error structure. As an example, for paid losses, the dependent and independent variables for a trial regression are shown in Table 7. This is to explore the structure of the data, but depending on the patterns in the residuals other regressions may be needed to find better models. A reasonable starting point is ordinary multiple regression, which assumes constant variance of the residuals (homoscedasticity). Even though the residuals are not likely to be constant here, as the small increments at the end of the triangle will probably have smaller residuals, such heteroscedasticity usually does not affect the regression coefficients much, although it does affect the overall predictive error distribution.

The coefficients for the three variables in this regression are: 0.818, 0.696, and 0.325, with standard errors of 0.033, 0.131, and 0.264. Thus the first two variables are significant but the third is not. Even though all the lags have similar ratios of paid to previous unpaid on average, the individual ratios are enough different to reduce the significance.

Table 7 – Dependent and Independent Variables for Paid Regression

Paid Increments	Previous Incurred	Previous Unpaid	Previous Unpaid
1228	978	0	0
1082	1844	0	0
2346	2904	0	0
3006	3502	0	0
1910	2812	0	0
2568	2642	0	0
166	0	300	0

214	0	604	0
494	0	596	0
432	0	666	0
870	0	1104	0
54	0	0	164
70	0	0	304
164	0	0	446
126	0	0	346
50	0	0	120
52	0	0	248
78	0	0	184
28	0	0	100
64	0	0	224
29	0	0	80

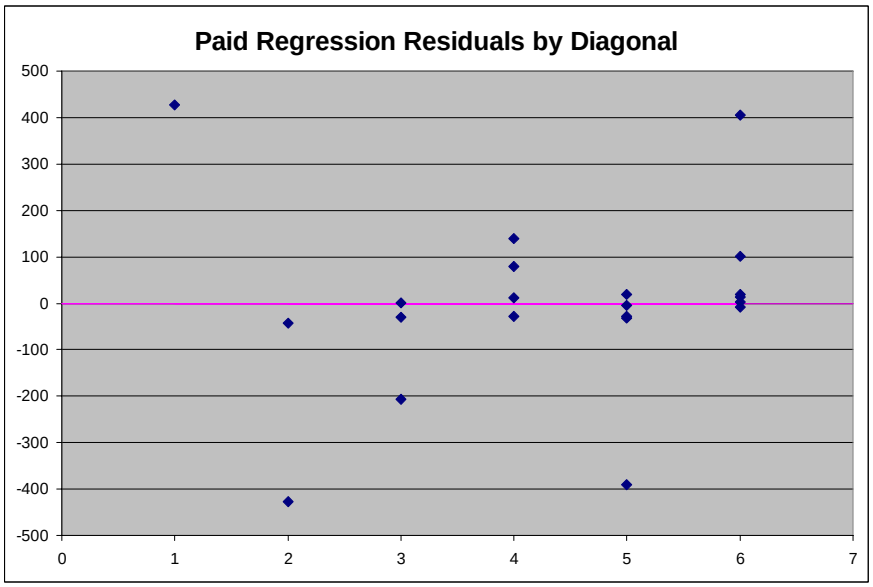
There are also diagonal effects in the residuals. The j^{th} diagonal is the one with row number plus column number = j. It also has j elements. The sum of the residuals by diagonal and number of positive residuals are in Table 8. The residuals by diagonal are graphed in Figure 2. Each diagonal can be seen to be quite biased.

Table 8 – Sum of Residuals and Number of Positive Residuals by Diagonal

Diagonal	**1**	**2**	**3**	**4**	**5**	**6**
Sum	427.5	–470.2	–236.8	200.9	–437.3	532.8
# > 0	1	0	1	3	1	5

There appear to be strong diagonal effects, coming in pairs of years, so offsetting each other over time. Dummy variables can be put in to model diagonal effects. Putting in dummies that are 0 or 1 would give additive effects for each diagonal – essentially adding or subtracting a positive constant for each cell on the diagonal. However, because the incremental paids are of such different sizes, some scaling of diagonal effects would be desirable. For modeling calendar-year effects, it is often more convenient to work with logs of losses, so the effects are automatically multiplicative, as in Barnett and Zehnwirth [10].

Figure 2



Here another method was used to create scaling in the diagonal effects. Since there is only one positive independent observation for each dependent observation in Table 7, and the independent and dependent variables all scale in a similar way, setting the dummy for each dependent variable equal to the positive independent variable would have a scaling effect. Also making a single dummy for each pair of diagonals, with opposite signs on the two diagonals, would reduce the number of variables, possibly without harming the goodness of fit. The matrix of dependent and independent variables for this is in Table 9.

For instance, the variable "*d* 6 – 5" is the dummy variable for diagonals 5 and 6. The observations in that column consist of the value of the independent variable for diagonal 6, its negative for diagonal 5, and 0 elsewhere. The (positive) coefficient for this variable will thus produce a reduction in the fitted values for diagonal 5 and an increase in the values for diagonal 6. This will not be an additive constant, but will be to a large extent scaled to the value of the increment being fitted. The other diagonal dummies work the same way.

The coefficients (not shown) come out quite similar to those for the regression with no diagonal elements, but now all are significant. The standard error of the regression has gone down from 206.6 for the regression without the diagonals to 73.4 with the diagonal dummies. The standard error is penalized for the number of variables, so is a good test to see if adding a variable is helpful. Sometimes regression modelers will keep in a variable that is only weakly significant if it improves the overall standard error.

Table 9 – Dependent and Independent Losses for Paid Regression with Diagonal Pairs

Paid	Incurred	Unpaid	Unpaid	*d* 6 - 5	*d* 4 - 3	*d* 1 - 2
1228	978	0	0	0	0	978
1082	1844	0	0	0	0	-1844
2346	2904	0	0	0	-2904	0
3006	3502	0	0	0	3502	0
1910	2812	0	0	-2812	0	0
2568	2642	0	0	2642	0	0
166	0	300	0	0	0	-300
214	0	604	0	0	-604	0
494	0	596	0	0	596	0
432	0	666	0	-666	0	0
870	0	1104	0	1104	0	0
54	0	0	164	0	-164	0
70	0	0	304	0	304	0
164	0	0	446	-446	0	0
126	0	0	346	346	0	0
50	0	0	120	0	120	0
52	0	0	248	-248	0	0
78	0	0	184	184	0	0
28	0	0	100	-100	0	0
64	0	0	224	224	0	0
29	0	0	80	80	0	0

Separating the diagonal dummies into individual variables for each diagonal did not help the standard error except in the case of diagonals 1 and 2. Putting in individual diagonal elements for them dropped the overall standard error to 63.3. The coefficients are in Table 10. The coefficients for diagonals 1 and 2 can be seen to be quite different in magnitude, so combining them into a single variable gives a worse fit.

Table 10 – Paid Regression Model

Parameter	Estimated	St dev	*t*	Pr(>|*t*|)
Incurred 0	0.8286	0.0107	77.341	0.0000
Unpaid 1	0.6619	0.0406	16.309	0.0000
Unpaid 2 - 5	0.3342	0.0808	4.1340	0.0012
Diagonal 6 – 5	0.1378	0.0155	8.9102	0.0000
Diagonal 4 – 3	0.0326	0.0138	2.3682	0.0341
Diagonal 2	-0.2384	0.0355	-6.7189	0.0000
Diagonal 1	0.4270	0.0656	6.5056	0.0000

Without going into so much detail, a similar process for fitting a model to the unpaid losses led to a regression with independent variables the previous cumulative paid and current paid for lag 1 (with a constant term). Just previous cumulative paid was the explanatory variable for lag 2, and a single variable of previous unpaid was used for the later lags. This means that for lags beyond 2, the

expected unpaid was modeled as a constant percentage (here estimated as 66.15%) of the previous unpaid.

The only significant diagonal is diagonal 3, which was modeled with a dummy variable similar to those in Table 9. The problem is that lag 1 itself is a multiple regression with two explanatory variables, so to define the diagonal dummy the rule used for any row was to give it the largest value among the explanatory variables in that row if the row is on diagonal 3, and zero otherwise. The overall standard error of the regression is 77.0. Dropping the diagonal 3 dummy increases the standard error to 92.6, so the dummy helps a good deal. The coefficients and other statistics are in Table 11. Diagonal 4 is not significant but improves the standard error slightly to 76.7. In the end this was not included in the model.

In this model, the high incurred losses for accident year 6 at development 0 will affect the projected paid at development 1, which will go into the estimated unpaid at development 2 and so on. However this is not as dramatic an effect as in the chain ladder, where the high incurred losses in the lower left corner would be multiplied by a large cumulative factor.

Table 11 – Unpaid Regression Model

Parameter	Est value	St dev	*t* student	Prob(> | *t*|)
Paid Cum 0	0.8215	0.1036	7.9316	0.0000
Paid Incrm 1	-0.5436	0.0864	-6.2889	0.0000
Constant 1	522.68	96.860	5.3963	0.0001
Paid Cum 1	0.0766	0.0098	7.8092	0.0000
Unpaid 2 - 5	0.6615	0.0983	6.7315	0.0000
Diagonal 3	0.0800	0.0281	2.8501	0.0128

2.3 Distribution of Residuals

Various distributions can be fit to the selected models by MLE. Typically the distributions are parameterized so that the mean is one of the parameters, and for each cell that is fit as a function of the covariates. All the other parameters of the distribution are constant across all the cells. However for many distributions it can work just as well if some parameter not the mean is a function of the covariates, and the other parameters are still constant.

Typically in generalized linear models, the residuals are modeled as members of the exponential family. These distributions are characterized by expressing the variance of each cell as a function of its mean, often as proportional to the p^{th} power of the cell's mean. However the skewness of the distributions also grows with p, which is not always in accord with the data. In Appendix 1, several dis-

tributions are discussed which give the variance of each cell as a multiple of the p^{th} power of the cell's mean by making p one of the parameters of the distribution. Then even with the same value of p, the different distributions can still have heavier or lighter tails, as indicated for instance by skewness.

The Weibull can be used as well, but it is more difficult to adjust its mean-variance relationship as it involves gamma functions, so a p-version was not fit. But the Weibull is an interesting possible residual distribution as it can be fairly heavy-tailed or lighter tailed than the normal, or even negatively skewed, depending on the parameters. It is most easily expressed by its survival function $1 - F(x) = S(x) = \exp[-(x/b)^c]$, and $E[X^j] = b^j (j/c)!$, where $y!$ is short for $\Gamma(1+y)$. The skewness is negative for c above about 3.6. The variance is proportional to the square of the mean, so p is always 2. The regression fit the b for each cell, not the mean.

Table 12 shows the results of fitting several distributions to the paid model. For this data, moving to less skewed distributions increases p and at the same time improves the fit (as measured by log-likelihood, which is equivalent to any of the information criteria such as AIC as all the distributions have the same number of parameters, except the Weibull, which has one fewer but has the best fit anyway). The Weibull, with $c = 7.437$ has skewness of -0.50.

Table 12 – Paid Model Distribution Fits

	p	– Ln L	Skew
Lognormal-p	1.50	111.94	> 3CV
Gamma-p	1.57	111.23	2CV
ZMCSP-p	1.60	110.52	CV
Normal-p	1.61	109.88	0
Weibull	2	108.76	-0.50

The similar, somewhat abbreviated, results for the model of unpaid losses are in Table 13. That Weibull has $c = 6.037$ and skewness -0.38. These two models will be used to project paid and unpaid losses. This does not imply that the Weibull is better in general. Other data could give quite different distributions.

Table 13 – Unpaid Model Distribution Fits

	p	–Ln L	Skew
ZMCSP-p	1.96	113.30	CV
Normal-p	2.03	112.93	0
Weibull	2	111.88	-0.38

With a different distribution of residuals, the coefficients for previous unpaid, etc. change a bit

from the usual regressions. For the Weibull, the larger cells have higher variances, so higher residuals are not penalized so much there, but the fits are now closer for the smaller cells. For the paid regression this ends up with the diagonal 6 – 5 and diagonal 4 – 3 parameters almost the same. Forcing these to be the same reduces the number of parameters by one but barely affects the loglikelihood, so this change was made. This is done by making a single dummy variable that is the sum of the *d* 6 – 5 and *d* 4 – 3 variables in Table 9. As mentioned above, the Weibull fit was for the *b* parameter, not the mean, so the coefficients have to be multiplied by $(1/c)!$ to get their effect on the mean. Table 14 shows the resulting coefficients for the two models.

Table 14 – Weibull Models' Estimated Covariate Parameters

Paid Parameter	Estimate	Unpaid Parameter	Estimate
Incurred 0	0.7811	**Paid Cum 0**	0.7358
Unpaid 1	0.6854	**Paid Incr 1**	-0.4275
Unpaid 2 - 5	0.3306	**Constant 1**	388.41
Diagonal 6–5+4–3	0.0339	**Paid Cum 1**	0.0908
Diagonal 2	-0.1873	**Unpaid 2 - 5**	0.7234
Diagonal 1	0.3971	**Diagonal 3**	0.0525

The projected mean incurred in Table 15 agrees closely in total with Quarg and Mack [4] except for year 6, for which they are about 1000 higher. Their model seems to give more emphasis to the incurred value for that year than to the paid. This model leans more toward believing the paid, but still ends up higher than year 3, which had more paid at 0. The average of the paid and incurred CL estimates is 7279, halfway between this model and [4]'s.

Table 15 – Completing the Square

Incurred	0	1	2	3	4	5	6
0	978	2104	2134	2144	2174	2182	2174
1	1844	2552	2466	2480	2508	2454	2460
2	2904	4354	4698	4600	4644	4652	4658
3	3502	5958	6070	6142	6158	6169	6177
4	2812	4882	4852	4863	4871	4877	4881
5	2642	4406	4646	4665	4679	4690	4697
6	5022	6182	6656	6685	6707	6722	6733

3 RUNOFF RANGES

The sum of the Weibull estimates in the bottom triangle may be close to being normally distributed, but simulation is usually required to get a good handle on the actual distribution of the runoff

losses. The simulation can be divided into parameter risk and process risk components. Distributions for the regression coefficients and the Weibull cs can be estimated by either the Fisher information matrix or the bootstrap, as detailed below. Here the information matrix method was used. Parameters can be simulated from the estimated distributions of the parameters, and then the runoff losses can be simulated from the Weibull distributions for each cell.

At the parameter values that maximize the likelihood function, the derivative of the negative log-likelihood (NLL) with respect to each parameter should be zero, but the second derivatives should be positive. This just means that the likelihood surface is flat at the minimum NLL but is curved upwards, which is usual for a minimum value. The mixed second partial derivatives could be anything, however. As detailed in the actuarial exams, the Fisher information matrix is the matrix of all the second derivatives and mixed second partials of the NLL with respect to the parameters. Thus if there are n parameters, it is an $n \times n$ matrix. Its matrix inverse is an estimate of the covariance matrix of the parameters.

Bootstrapping could be done by resampling with replacement from the normalized residuals of the fitted triangles to generate new triangles, and refitting the models. Each resampled triangle would give a new set of fitted parameters for the paid and unpaid models. The table of parameters that results from doing this many times would be the estimated empirical parameter distribution. For these models this would probably give some correlation to some of the parameters across the paid and unpaid models, which are uncorrelated under the information matrix method since they come from different models. Also the dependent and independent variables would change with each resampling, which could end up with more parameter diversity as well.

In Tables 7 and 9, label the dependent variables y_j for $j = 1, \ldots, 21$, and label the corresponding independent variables $x_{i,j}$. In the final models i ranges from 1 to 6. Call the covariate parameters β_i, $i = 1, \ldots, 6$. The Weibull b parameter for each dependent variable is $b_j = \sum_{i=1}^{6} \beta_i x_{i,j}$. Then the derivative of b_j with respect to β_i is just $x_{i,j}$. Thus, the derivative of NLL $= -\sum_j \ln f(y_j)$ with respect to β_i is

$$\partial NLL / \partial \beta_i = -\sum_j x_{i,j} \, \partial \ln f(y_j) / \partial b_j .$$ Similarly, $$\partial^2 NLL / \partial \beta_i \partial \beta_k = -\sum_j x_{i,j} x_{k,j} \, \partial^2 \ln f(y_j) / \partial b_j^{\,2}$$ and

$$\partial^2 NLL / \partial \beta_i \partial c = -\sum_j x_{i,j} \, \partial^2 \ln f(y_j) / \partial b_j \partial c$$ The nice thing about these formulas is that the depend-

ence on i or k is only in the x factor. The rest is just a single column (function of j) that comes right from the Weibull.

The Weibull formulas (suppressing j) are:

$$\partial \ln f(y) \Big/ \partial b = \frac{c}{b}\left[\left(\frac{y}{b}\right)^{c} - 1\right]$$

$$\partial^2 \ln f(y) \Big/ \partial b \partial c = \frac{1}{b}\left[\left(\frac{y}{b}\right)^{c}\left(1 + c\ln\left(\frac{y}{b}\right)\right) - 1\right]$$

$$\partial^2 \ln f(y) \Big/ \partial b^2 = \frac{c}{b^2}\left[1 - (1+c)\left(\frac{y}{b}\right)^{c}\right]$$

$$\partial^2 \ln f(y) \Big/ \partial c^2 = -\frac{1}{c} - \left(\frac{y}{b}\right)^{c}\left(\ln\left(\frac{y}{b}\right)\right)^2$$

These give the parameter standard deviations and correlation matrices in Tables 16 - 19.

Table 16 – Paid Parameters and Standard Deviations

Paid	**Inc 0**	**Unpd 1**	**Unpd 2-5**	**Diag 6543**	**Diag 2**	**Diag 1**	***c***
Parameters	0.832	0.730	0.352	0.036	-0.200	0.423	7.427
Standard dev	0.050	0.052	0.016	0.014	0.069	0.176	1.392
Ratio	16.70	14.08	22.31	2.49	-2.87	2.40	5.33

Table 17 – Unpaid Parameters and Standard Deviations

Unpaid	**Pd Cum 0**	**Pd Inc 1**	**Const**	**Pd Cum 1**	**Unpd 2-5**	**Diag 3**	***c***
Parameters	0.793	-0.461	418.5	0.098	0.780	0.057	6.037
Standard dev	0.145	0.100	102.9	0.008	0.042	0.022	1.148
Ratio	5.48	-4.62	4.07	11.64	18.55	2.52	5.26

Table 18 – Paid Correlation Matrix

1	0.17	0.00	-0.12	-0.24	-0.28	0.11
0.17	1	0.00	-0.19	-0.62	-0.05	0.14
0.00	0.00	1	0.19	0.01	0.00	0.26
-0.12	-0.19	0.19	1	0.13	0.03	-0.03
-0.24	-0.62	0.01	0.13	1	0.07	-0.08
-0.28	-0.05	0.00	0.03	0.07	1	-0.03
0.11	0.14	0.26	-0.03	-0.08	-0.03	1

Table 19 – Unpaid Correlation Matrix

1	-0.86	0.00	0.02	0.01	-0.03	0.06
-0.86	1	-0.49	0.00	-0.01	-0.05	-0.03
0.00	-0.49	1	-0.02	-0.01	0.07	-0.03
0.02	0.00	-0.02	1	0.07	-0.29	0.29
0.01	-0.01	-0.01	0.07	1	-0.04	0.22
-0.03	-0.05	0.07	-0.29	-0.04	1	-0.09
0.06	-0.03	-0.03	0.29	0.22	-0.09	1

Two simulation steps were done with these parameters. First the parameters were simulated, then the Weibull losses were simulated for each cell in the projected lower triangles.

To simulate the parameters, MLE parameters are asymptotically multivariate normal with the derived correlation matrices. However with small samples like these, the normal approximation might not hold. Simulation experiments have found that lognormal distributions are more realistic for small samples. As one example, the simple Pareto shape parameter, given a known location parameter, is inverse gamma distributed. This gets normal-like for large samples, but is heavier-tailed, as is the lognormal, for small samples. For the lognormal assumption, the absolute value of negative parameters could be assumed to be lognormal. One advantage of the lognormal over the normal is that the simulated parameters will not change signs from the mean parameter, even for remote points in their distributions.

For these reasons the lognormal was used here. To simulate the multivariate lognormal, the correlation matrices were input into a normal copula, and then lognormal marginal distributions applied. This maintains the Kendall's tau and rank correlation, but not the linear correlation, of the parameters. The needed reserve position, calculated as ultimate incurred less current incurred, from the mean parameters is 6212. 10,000 simulations had mean runoff of 6203, with a standard deviation of 801. This gives a CV of 13%. The coefficient of skewness is 3.4%, which is closer to a normal

distribution than the 26% for a gamma with the same CV. Thus, the normal might provide a reasonable approximation in this case.

4 OTHER SYSTEMATIC RISK

Loss reserves are subject to inflation and trends in the lawsuit environment that happen between occurrence and payment. Some degree of such trend may be built into the accident year level changes, but this is hardly a full reflection of the risk. The average level of future inflation built into the projections could be off, and in addition there are likely to be year-to-year changes in inflation, perhaps correlated one year to the next. Usually the data in the triangle itself is not sufficient to estimate these systematic risks, so they have to be superimposed afterwards.

An internal study of historical variability in trends and actual runoff, based on U.S. annual statement data for a number of companies and inflation variability, suggested that there is quite a bit more variability in actual runoff than standard reserving models would predict. Also Wright [11] found in a simulation test that runoff ranges from typical methods tend to be too narrow. Gluck [3] proposes ways to incorporate systematic risk elements into insurer financial models in general and loss reserve runoff risk in particular. The model used below is roughly consistent with his approach but the numerical values are for illustration only.

In a single simulation of the runoff, the simulated value for losses paid in accident year w at lag d, and thus in calendar year $w+d$, is multiplied by a simulated factor $H_{w,d}$ given by:

$$H_{w,d} = BD^{w+d-n}E_{w+d}\text{, where:}$$

B is a mean 1 factor for all calendar years that can be thought of as frequency risk; a normal distribution with a standard deviation of 10% is assumed in the example.

D is a lognormal mode 1 draw for all calendar years in the simulation to represent an overall trend error that compounds; n is the last diagonal in the data; a standard deviation of 2% is used for D.

E_{w+d} is generated from an AR-1 model, to represent (ii). The process for E is as follows: The X_is are independent $N(0, \sigma^2)$ random draws, and $\rho \in [0,1]$ is the autocorrelation coefficient. Let $t_1 = X_1$, and $t_{i+1} = \rho t_i + X_{i+1}$. Then $E_{w+d} = \exp\left(\sum_{j=1}^{w+d-n} t_j\right)$. The values $\sigma = 2.5\%$ and $\rho = 70\%$ are used in the

numerical example.

Using lognormal mode 1 factors gives an increase in the mean reserve. Actually something similar happens in just multiplying normal mean 1 factors that are positively correlated. This is justified in a few ways. First, an error in trend would compound and the effects on each year would be correlated. Second, new claim types and other superimposed changes tend to have an upward drift. Third, many reserve models, including this one, do not project ongoing calendar-year trends, but these often do affect open claims from previous years.

In a sense, this approach incorporates a degree of model risk, in addition to process and parameter risk. For instance, it is difficult in the fitting to distinguish calendar-year trends from upward and downward individual calendar-year gyrations. This is a model-risk issue. Even if the fitted model has calendar-year trends in it, there is still a question of which trend to project going forward. Thus putting trends into the model does not automatically solve the model-risk problem, and systematic projection risk still needs to be incorporated. The simulated distributions, with annual discount factor 0.96 (a rate of $4\,{}^{1}/_{6}\,\%$), are in Table 20.

Table 20 – Simulated Moments and Percentiles of Runoff Distribution

Probability	Model	+ Systematic	Discounted
0.4%	4,089	3,766	3,507
1.0%	4,359	4,054	3,760
5.0%	4,881	4,614	4,289
10.0%	5,174	4,930	4,586
25.0%	5,665	5,506	5,119
50.0%	6,203	6,186	5,746
75.0%	6,734	6,941	6,441
90.0%	7,228	7,663	7,107
95.0%	7,515	8,101	7,507
99.0%	8,078	9,056	8,396
99.6%	8,389	9,714	8,924
Mean	6,203	6,258	5,808
Std. Dev.	801	1,076	991
CV	0.13	0.17	0.17
Skewness	0.03	0.40	0.38

Including systematic risk at this level slightly increased the mean, but approximately doubled the variance, increasing the spread both upward and downward. Even the discounted losses with systematic risk were higher than the original model above the 95th percentile. Both systematic risk distributions were slightly more skewed than the gamma with the same CV, which would have skew-

ness of about 0.34, but were less skewed than the lognormal or inverse Gaussian. Possible distributional assumptions that could match these distributions are discussed in Appendix 2.

If several lines are being modeled, it would be reasonable to assume that the systematic risk elements were highly correlated across lines, since they arise largely from external influences. Thus even if the development patterns themselves are not highly correlated, including systematic risk could produce a higher correlation.

5 VALUE OF RESERVES

Once a distribution of reserves has been estimated, what do you do with it? One application is to estimate the financial value of the reserves, which could be useful for market-value accounting or valuation of the entire company. There are a few alternatives for how to do this. For instance, in Australia insurers post the 75th percentile of the distribution as the balance-sheet value. To try to get to a market value, there are two prevailing financial theories: the capital asset pricing model (CAPM) and its generalizations, and arbitrage-free pricing. Typically CAPM-like approaches price only the systematic risk, while arbitrage-free pricing looks at the whole distribution of possible outcomes. There are also traditional actuarial pricing methods, like mean plus a percentage of standard deviation.

Balance-sheet items need to be additive, as users of financial statements like to add and subtract assets and liabilities. If there really were a market for reserve risk, prices would be additive also. Otherwise traders could buy risk, pool it, and sell it for no-risk profits. Arbitrage-free and CAPM prices are additive, and standard deviation loads can be made additive, as shown below. These methods can be subdivided at will and still maintain additivity. Lines can be allocated by state and accident year, and summed to by-state totals, etc. Here only methods that use the entire distribution will be used, but having a model for systematic risk would allow using CAPM-type approaches as well.

Arbitrage-free pricing uses probability transforms of possible events, putting more weight on adverse outcomes, and takes the transformed mean as the price. This is where having a distribution of reserve runoff could be applied. One well-known transform is the Wang [5] transform. This transform applies to the survival function $S(x) = 1 - F(x)$ to produce a transformed survival function $S^*(x)$. In its original form it just translated normal percentiles, so $S^*(x) = \Phi[\lambda+\Phi^{-1}(S(x))]$, where Φ is the standard normal distribution function. A bit different form, first suggested by John Major, is

$S^*(x) = Q_\nu[\lambda+\Phi^{-1}(S(x))]$, where Q_ν is the *t* distribution with ν degrees of freedom. This puts more weight into the tails of the distribution.

The original normal-normal version will be called the NN transform here. The NN transform moves the probability away from the lower percentiles towards the higher percentiles. The Wang transform generally does this as well, but the heavier *t* tails can also put more probability into the extreme left tail, even after translating by λ. This is a stronger effect with lower values of ν, because the t approaches the normal for high value of ν. In this transform ν does not have to be an integer, as the beta distribution can be used to calculate the *t* even for non-integer degrees of freedom. Using the function betadist as defined in Excel, the calculation is $Q_\nu(x) = ½ + ½ \text{ sign}(x)\text{betadist}[x^2/(\nu+x^2), ½, \nu/2]$.

Once the transformed events probabilities are calculated, the value of the reserves are estimated as the transformed mean. The mean of each accident year can be calculated with the same transformed probabilities. Then the resulting accident-year values add up to the total value. If several lines are being done simultaneously, the transform is done on the aggregate loss probabilities. That gives probabilities for each simulated scenario, then they are applied to the losses for each line and accident year in that scenario. Thus, any correlations gets into the overall value and the individual line values reflect the correlations.

Another transform with theoretical and empirical support (for example, see Venter [12]) is the Esscher transform. While the Wang transform is defined on the aggregate distribution, the Esscher transform is defined on the density or discrete probability function. For density $g(x)$, the transformed density with parameter c is defined by $g^*(x) = g(x)\exp(cx/EX)/\ E[\exp(cX/EX)]$. This transform depends on the distribution being transformed, as the transform at x depends on x. The Wang transform, on the other hand, depends only on $S(x)$. Thus with given parameters ν and λ, any simulation of 10,000 equally-likely events will get the same transformed probabilities.

Since the Wang transform is done on the survival function, a couple of steps are needed to apply it to the scenario probabilities from a simulation. Some of these are a bit arbitrary. The survival function at the k^{th} simulation here is calculated as $k/10{,}001$. This keeps the survival function in the range (0,1), although there are other ways to do that. Then S^* has to be translated back to individual scenario probabilities. To do this, the lowest point was considered to represent the range from zero to half the way, in probability, between it and the second lowest point. Thus, it was assigned prob-

ability $1 - [S^*(x_1)+S^*(x_2)]/2$. Then the next point gets the average between the next two midpoints, or $[S^*(x_3) - S^*(x_1)]/2$, etc. Finally the last point gets $[S^*(x_{10,000})+S(x_{9999})]/2$. This forces the probabilities to sum to 1.

The standard deviation loading can be allocated with the Euler method. This method was used by Patrik, Burnegger, and Rüegg [13] for capital allocation and Venter, Major, and Kreps [14] used it for allocation of risk measures, and showed the steps needed to apply Euler's work to random variables. The use for risk measures can be used to allocate standard deviation loading as well. The general approach for a risk measure $\rho(Y)$, where $Y = X_1+\ldots+X_n$ is to allocate to X_k $r(X_k) = \lim_{\varepsilon \to 0} \frac{\rho(Y) - \rho(Y - \varepsilon X_k)}{\varepsilon}$.

The numerator is the reduction in the risk measure from ceding a quota share of ε of X_k. Then $r(X_k)$ is the reduction in $\rho(Y)$ from an incremental reduction in X_k scaled up by ε. Basically it is treating every increment of X_k as the last in. The result of Euler is that the sum of the allocations over all the Xs is the whole risk measure $\rho(Y)$ in the case ρ is homogeneous of degree 1, i.e., $\rho(aY) = a\rho(Y)$. When $\rho(Y)$ = standard deviation(Y), the allocation is shown in [13] to be $r(X_k) = \text{Cov}(X_k,Y)/\rho(Y)$. This is not based on the standard deviations of each component, but rather the component's contribution to the standard deviation of Y.

Even a loading based on a percentile of the distribution can be allocated in this manner. The p^{th} percentile can be expressed as $E[Y \mid F(y) = p]$. Then the marginal allocation is shown in [13] to be $E[X_k \mid F(y) = p]$. In a simulation, this would be the value of X_k for the simulation where the probability of Y is p. However this is not a very stable allocation, and in practice the average of simulations for a range around that simulation is used. This is then not truly an allocation of the percentile but an allocation of a range around it, sometimes called blurred value of risk.

All of the pricing measures discussed have a free parameter or two which have to be set to something. In practice some market benchmarking can help establish this. Unlimited portfolio transfers are not usually available, but a limit of twice the mean may be. That is over the 99.99^{th} percentile for this simulation, so may be a good approximation for unlimited. An internal study a few years ago found that many reinsurance treaties are priced at the mean plus one-third to one-half of a standard deviation. Taking the Wang parameters of $\nu = 10$ and $\lambda = 0.47$ gives a loading of close to half a standard deviation, with a discounted market value of 6303.8. This is the 70.7^{th} percentile of the dis-

counted distribution, so would produce a profit slightly more than 70% of the time. Another benchmark is how much capital above the premium could be kept at return of 15%, and the probability level of that capital. In this case, with a profit load of 495, that would be 2807 of capital, which with the premium would get to the 99.93rd percentile. That would be a fairly safe capital level. Thus this is a reasonable value by some benchmarks. Whether or not it is reasonable, in fact, would require more benchmarking against actual deals, however.

The same price would come from an Esscher transform with $c = 2.6612$. The resulting values for the individual accident years from each methodology are shown in Table 21.

Table 21 – Value of Discounted Reserves by Accident Year for Several Methodologies

	1	2	3	4	5	6	total
Esscher transform	106.2	154.8	311.0	220.5	658.7	4,852.6	6,303.8
Wang transform	106.3	154.9	311.1	221.4	662.3	4,847.6	6,303.8
Standard dev.	106.2	154.6	310.6	220.1	657.4	4,854.9	6,303.8
Percentile alloc	107.1	156.4	316.3	218.4	646.1	4,859.4	6,303.8
Mean discnted	102.1	147.9	294.9	208.8	620.7	4,434.0	5,808.5
Mean undiscnted	112.4	164.6	330.9	235.8	694.9	4,718.9	6,257.7
Cov Disc w total	8,039	13,230	31,058	22,325	72,840	834,834	982,326
Cor Disc w total	43%	40%	43%	42%	50%	97%	100%

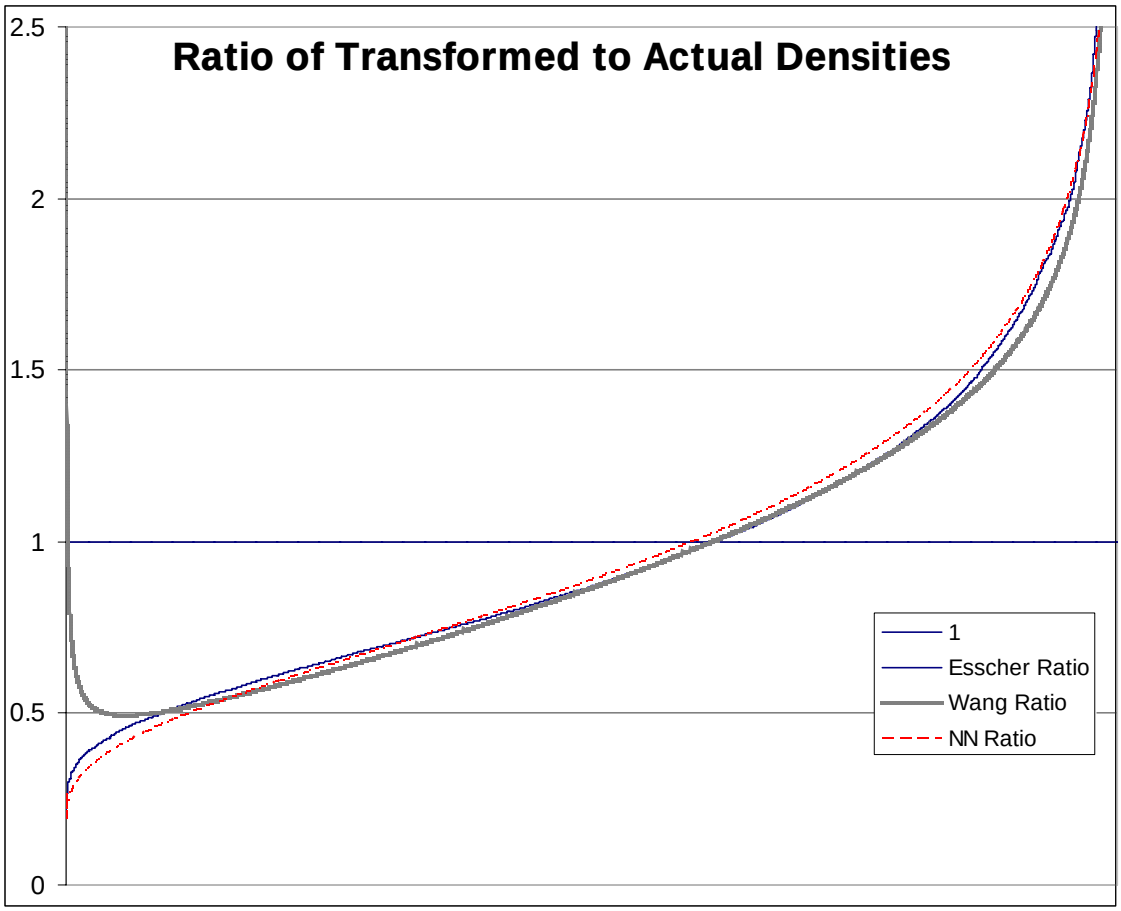
The allocations are all quite similar, but the percentile allocation is slightly different than the others. The Esscher and standard deviation values are closest overall. The percentile allocation is actually the average of 101 simulations centered at the actual percentile value adjusted slightly to balance to the mean.

The ratios of transformed to actual probabilities for the Esscher and Wang transforms are graphed in Figure 3, along with the NN version of the Wang transform, which matches the transformed mean by setting $\lambda = 0.485$.

The Wang transform strengthens both tails, but with ν as high as 10, the left tail strengthening is not great. It also strengthens the right tail quite a bit. For the Esscher, Wang, and NN transforms, the ratios at the second to highest value are 8.4, 13.8, and 5.3. For most of the range, however, the transforms are fairly similar.

The additive methods reviewed here give similar allocations in this case. When some components are more heavy-tailed, there can be greater differences. The importance is in using some kind of additive approach. Further benchmarking would be necessary to see which make most market sense.

Figure 3 – Ratios of Transformed to Actual Probabilities



6 SUMMARY AND FURTHER POSSIBILITIES

A growing body of research is finding that paid and incurred losses can help predict each other. Here a regression approach was used to model paid and unpaid losses, with earlier paid, incurred, or unpaid losses all available as independent variables. It is not asserted that the best possible regression was found. Using paid-to-incurred ratios as independent variables could potentially be useful, for instance. In fact, after the first few lags, unpaid losses were significant in predicting future paid and unpaid, consistent with the suggestion of Jedlicka [9]. Coefficient and overall standard errors were the key regression diagnostics used for evaluating models. Barnett and Zehnwirth [10] recommend residual plots as well, which can be useful but sometimes require regression experience to evaluate.

Once a reasonable regression model was found, MLE was used to evaluate other residual distributions. This requires a non-linear optimization routine. Weibull residuals with slight negative skewness and variance proportional to the mean squared maximized the likelihood. This is a bit surpris-

ing, as if you think of the cells as compound frequency-severity processes, positive skewness and variance proportional to a lower power of the mean would be more anticipated.

Once in the world of non-linear optimization, other models become possible as well. For instance, the cell means could be linear functions of independent variables times diagonal effects times some power of the paid-to-incurred ratio, possibly with the power declining for later lags as in Verdier and Klinger [8], plus an additive residual. However this kind of modeling would not readily be able to take advantage of the ease of linear regression software for exploratory analysis. If the observations are all positive, as would be likely with paid and unpaid data, the regression steps could be done in logs, then a multiplicative model with additive residuals fit later if needed for a good residual distribution. Another modeling approach worth pursuing is the idea of Mack [7] to look and paid and incurred development in cross-classified multiplicative models.

The information matrix from MLE was used to estimate parameter uncertainty, with the selection of a lognormal distribution of parameters due to the small sample sizes. Bootstrapping is certainly an alternative here, and may be preferable in that it can pick up possible correlations among the parameters of the two different models. In fact Liu and Verrall [15] have already used bootstrapping for the model of Quarg and Mack. Bootstrapping for the model here would be a bit more computationally intensive than usual, due to the non-linear multivariate optimization at each step, but would have another advantage in that the choice of normal vs. lognormal parameter errors would not be needed.

Systematic risk, including model risk, is clearly an issue in reserve modeling, and historical loss development volatility has been substantial, with even more variability than standard models might suggest. This was reflected here with selected distributions, but should be studied in a more formal way. Similarly, reserve value was illustrated with some rough benchmarks, but more research into the market value of loss reserve risk is called for. The methods of transformed distributions and Euler for producing additive market-values were illustrated. In this case they were not so different, once an overall market value was established.

REFERENCES

[1] Venter, Gary. 2007a. "Refining Reserve Runoff Ranges," Casualty Actuarial Society *E-Forum*, Summer 2007.

[2] Venter, Gary. 2007b. "Generalized Linear Models beyond the Exponential Family with Loss Reserve Applications," CAS *E-Forum*, Summer 2007.

[3] Gluck, Spencer, "A Multiline Risk Factor Model," Paper presented at the 2007 ASTIN Colloquium, Or-

lando, Florida, http://www.actuaries.org/ASTIN/Colloquia/Orlando/Papers_EN.html

[4] Quarg, Gerhard and Thomas Mack. 2004. "Munich chain-ladder—A reserving method that reduces the gap between IBNR projections based on paid losses and IBNR projections based on incurred losses." *Blätter DGVFM,* Heft 4, 597–630.

[5] Wang, Shaun. 2004. "Cat Bond Pricing Using Probability Transforms," Geneva Papers: Etudes et Dossiers, special issue on "Insurance and the State of the Art in Cat Bond Pricing," January 2004, No. 278:19-29. http://www.gsm.pku.edu.cn/stat/public_html/ifirm/reports/Cat%20bond%20pricing%20using%20probability%20transforms%2001-2004.pdf

[6] Halliwell, Leigh J. 1997. "Conjoint Prediction of Paid and Incurred Losses," *CAS Forum*, Summer, 241-380.

[7] Mack, Thomas. 2005. "Recent Developments in Claims Reserving," Presented at the 2005 ASTIN Colloquium, Zurich, Switzerland, http://www.actuaries.org/ASTIN/Colloquia/Zurich/Mack_presentation.pdf

[8] Verdier, Bertrand and Artur Klinger. 2005. "JAB Chain: A model-based calculation of paid and incurred loss development factors," Presented at the 2005 ASTIN Colloquium, Zurich, Switzerland, http://www.actuaries.org/ASTIN/Colloquia/Zurich/Verdier_Klinger.pdf.

[9] Jedlicka, Petr. 2007. "Various extensions based on Munich Chain Ladder method," Presented at the 2007 ASTIN Colloquium, Orlando, Florida, http://www.actuaries.org/ASTIN/Colloquia/Orlando/Presentations/Jedlicka.pdf.

[10] Barnett, Glen and Ben Zehnwirth. 2000. "Best Estimates for Reserves," *Proceedings of the Casualty Actuarial Society* 87:245-321.

[11] Wright, Thomas. 2007. Chapter 9 and Appendix B of "Report of "Best Estimate and Reserve Uncertainty" Working Party (chair: Lis Gibson)," GIRO.

[12] Venter, Gary. 2004 "Discussion of 'Distribution-Based Pricing Formulas Are Not Arbitrage-Free,'" *Proceedings of the Casualty Actuarial Society* 91:25-32, http://www.casact.org/pubs/proceed/proceed04/04014-2.pdf.

[13] Patrik, Gary, Stefan Burnegger, and Marcel Beat Rüegg. 1999. "The Use Of Risk Adjusted Capital To Support Business Decision-Making," *CAS Forum*, Summer, 243-391.

[14] Venter, Gary, John Major, and Rodney Kreps. 2006. "Marginal Decomposition of Risk Measures" ASTIN Bulletin 36(2):375-413.

[15] Liu, Huijuan, and Richard Verrall 2007. "A Dynamic Approach for Predictive Distributions of Reserves Using Paid and Incurred Claims," Presented at the 2007 IME Conference, Piraeus, Greece, http://stat.unipi.gr/ime2007/abstracts.php.

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 graduate seminars in loss modeling at Columbia University.

917.937.3277 gary.g.venter@guycarp.com

Appendix 1 – *p*-Distributions

In Venter [2] parameters are added to standard distributions to specify the relationship of variance and mean when used with covariates. Typically a distribution will be re-parameterized so that the mean is a parameter, and that parameter will be a function of the covariates. The other parameters will be constant for all the observations. Many distributions can be parameterized so that the variance of each observation will be proportional to any desired power of the mean. That power parameter can then be estimated by MLE to get an idea of how the residuals' variances relate to their means for a given data set. Not all the distributions below are in [2], and some are parameterized a bit differently here. Although they can each produce a variance proportional to any desired power of the mean, they differ in other shape features, such as skewness. Sometimes the over-dispersed Poisson (ODP) is defined as any distribution where the variance is proportional to the mean. Under that definition, any of these distributions can be an ODP just by taking p=1. However they will differ in other shape characteristics. The distributions below are by increasing skewness.

Normal-*p*

The normal distribution is typically parameterized with mean μ and variance σ^2. Introducing two new parameters k and p, it can be re-parameterized just by setting $\sigma^2 = k\mu^p$. It then has log density $\ln f(x) = -\frac{1}{2}\ln(2\pi k\mu^p) - (x-\mu)^2/(2k\mu^p)$. With k and p constant across all observations, each observation's variance will just be k times its mean raised to the p. The skewness is 0.

ZMCSP-*p*

The zero-modified continuous scaled Poisson, as discussed in Venter [2] and Mack [15], is the Poisson distribution function extended to the positive reals, plus a scaling factor, with the probability at 0 set to the value needed to bring the entire probability to 1. It has variance close to proportional to the mean and skewness close to the coefficient of variance (CV), which is the ratio of standard deviation to mean. It is a continuous form of ODP that retains much of the shape of the Poisson distribution. The density can be written as:

$$f(x) = e^{-k\mu^{2-p}}\left(k\mu^{2-p}\right)^{kx\mu^{1-p}} k\mu^{1-p} / \Gamma\left(1+kx\mu^{1-p}\right).$$

For large means, the mean, variance, and skewness are very close to μ, μ^p/k, and CV. For smaller means, a small adjustment is needed. See [2] for details.

Tweedie

The Tweedie distribution has p between 1 and 2, and the skewness is pCV. It is actually a special case of the Poisson-gamma aggregate distribution with the frequency and severity means coordinated. Starting with a Poisson in λ and a gamma in θ and α, introduce new parameters p, μ, and k with $p = 1+1/(\alpha+1)$, $\lambda = k\mu^{2-p}$, and the severity mean $\alpha\theta = \mu^{p-1}/k$. Then the aggregate mean is μ. Since α is positive, p is between 1 and 2, so both the frequency and severity means are increasing functions of μ. Thus a higher overall mean in a cell is a combination of a higher frequency mean with a higher severity mean. The aggregate variance turns out to be $\mu^p/[(2-p)k]$. Fitting by MLE is discussed in [2].

Gamma-*p*

The gamma distribution is usually parameterized $F(x,\theta,\alpha) = \Gamma(x/\theta; \alpha)$ with the incomplete gamma function Γ. This has mean $\alpha\theta$ and variance $\alpha\theta^2$. To get the mean to be a parameter, set $F(x,\mu,\alpha) = \Gamma(x\alpha/\mu; \alpha)$. Then the variance is μ^2/α and μ is still a scale parameter. For the gamma-p, take $F(x;\mu,k,p) = \Gamma[x/(k\mu^{p-1}); \mu^{2-p}/k]$, which has mean μ and variance $k\mu^p$, with skewness = 2CV.

Lognormal-*p*

The usual parameterization of the lognormal is: $F(x;\mu,\sigma) = N\left(\frac{\ln(x)-\mu}{\sigma}\right)$. This has mean $e^{\mu+\sigma^2/2}$ and variance $e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right)$. Now reparameterize with three parameters p, m and s:

$$F(x;m,s,p) = N\left(\frac{\ln\left((x/m)\sqrt{1+s^2m^{p-2}}\right)}{\sqrt{\ln\left(1+s^2m^{p-2}\right)}}\right).$$

This has mean m, variance s^2m^p, and skewness 3CV+CV^3, where CV = $sm^{p/2-1}$. Here μ has been replaced by $\ln\left(\frac{m}{\sqrt{1+s^2m^{p-2}}}\right)$ and σ^2 by $\ln(1+s^2m^{p-2})$.

Appendix 2 – Possible Distributions for Simulations

Some of the work on loss reserve risk is on moments only, so having simulated distributions can provide a test of different parameterized distributions. In this case there are three simulated distribu-

tions: the original model, that plus systematic risk, and that discounted. The CV for the first is 13% and for the other two is 17%. The skewnesses are 3%, 40%, and 38%, respectively. For two-parameter distributions, the CV and skewness are often determined by only one of the parameters, so they become functions of each other as well. The skewness for CVs of 13% and 17% for a few common distributions is shown in Table A2-1.

Table A2-1 – Skewness for CVs of 13% and 17%

Distribution CV:	**13.0%**	**17.0%**
Normal	0	0
Weibull	-45.6%	-60.1%
Poisson	13.0%	17.0%
Gamma	26.0%	34.0%
Inverse Gaussian	39.0%	51.0%
Lognormal	39.2%	51.5%

The skewness for the original model is closest to, but higher than, that of the normal. For the other models, it is closest to but somewhat higher than that of the gamma.

A convenient distribution for matching three moments is the shifted gamma. $X - a$ is gamma in θ and β, so $EX = a + \theta\beta$, $\mathrm{Var}X = \theta^2\beta$, and skewness $= 2\beta^{-1/2}$. For a positively skewed distribution it is always possible to solve for the three parameters in terms of these moments, but the shift can be negative, giving positive probability to negative values of X. If the skewness is 2 or greater, the gamma has its mode at zero, and the density declines from there, which may not be a realistic shape in some cases. Then perhaps a shifted-lognormal or power-transformed beta or gamma may work better. In terms of the moments the parameters are: $\beta = (2/\mathrm{skw})^2$; then $\theta = \mathrm{Var}/\beta = \mathrm{stdev}*\mathrm{skw}/2$; and $a = \mathrm{mean} - \theta\beta$. The parameters for the three distributions simulated are in Table A2-2.

Table A2-2 – Shifted gamma parameters for simulated distributions

	Model	**+Systematic**	**Discounted**
a	-41,581.7	910.01	580.25
θ	13.424	216.58	187.89
β	3559.6	24.69	27.83

The parameters for the original model look strange, but in fact the probability of a negative result is less than 10^{-15}. The shifted-gamma probabilities for selected percentiles of the simulated distribu-

tions are shown in Table A2-3.

Table A2-3 – Shifted gamma probabilities for simulated percentiles

Probability	Model	+Systematic	Discounted
0.40%	0.38%	0.30%	0.32%
1.00%	1.00%	0.93%	0.92%
5.00%	4.84%	4.94%	4.96%
10.00%	9.88%	9.94%	10.01%
25.00%	25.19%	25.28%	25.30%
50.00%	50.23%	50.02%	49.99%
75.00%	74.75%	75.05%	75.09%
90.00%	89.92%	89.84%	89.92%
95.00%	94.84%	94.67%	94.73%
99.00%	98.98%	98.96%	99.02%
99.60%	99.65%	99.72%	99.69%

The fits are fairly good between 1% and 99%, with a little fading off in both far tails. This is not a given from matching three moments, because other distributions matching the same moments could have fairly different shapes. The shifted gamma may or may not fit as well to other development triangle runoff distributions.