DISCUSSION OF PAPER PUBLISHED IN VOLUME LXXII

A SIMULATION TEST OF PREDICTION ERRORS OF LOSS RESERVE ESTIMATION TECHNIQUES

BY JAMES N. STANARD

DISCUSSION BY EDWARD F. PECK

1. INTRODUCTION

This discussion of James Stanard's paper "A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques" will use his simulation technique to test three loss reserving methods. Two of these methods are discussed in Stanard's paper, and one is relatively new having been presented in the *Proceedings* last year by Daniel Murphy [6]. The three methods are shown to be special cases of a general weighted average approach. In addition, some of the concepts presented by Stanard concerning the expected value of a loss development factor will be analyzed in a little more detail. Please note that the results derived in this discussion are due to the assumptions made within this discussion and may not be applicable to general loss reserving situations.

2. THREE LOSS RESERVE METHODS

To describe these three methods, the following notation will be used: if $X_{i,j}$ represents a random sum of losses from accident year i, measured j years after the beginning of the accident year, then an accident year loss triangle is as shown in Table 1.

An age-to-age average loss development factor from age j to age $j+1$ can be defined as

$$\mathrm{LDF}_j = \sum_i [X_{i,j+1}/X_{i,j}]/n,$$

TABLE 1

ACCIDENT YEAR LOSS TRIANGLE

Accident Year	Age in years 1	2	3	4
1	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$
2	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	
3	$X_{3,1}$	$X_{3,2}$		
4	$X_{4,1}$			

where n is the number of accident years which have reached age $j + 1$. This is the usual average of available individual LDFs. This will be called Method I.

Another way to calculate age-to-age factors is to divide the sums:

$$\mathrm{LDF}_j = \sum_i X_{i,j+1} \Big/ \sum_i X_{i,j}.$$

This will be referred to as Method II. Both of these calculations include only those accident years where both $X_{i,j}$ and $X_{i,j+1}$ exist.

Finally, another approach is to define a proportional relationship of losses from one age to the next and find a least squares estimator. If $X_{i,j+1} = p_{i,j} X_{i,j}$, where $p_{i,j}$ is the parameter to be estimated, then an age-to-age factor can be defined as

$$p_j = \mathrm{LDF}_j = \sum_i X_{i,j+1} X_{i,j} \, / \, \sum_i X^2_{i,j}$$

So, this p_j is an estimator of the change in losses from one age to the next, just as the LDFs using the other two methods are. This calculation would again use only those available $X_{i,j}$s. This least squares technique will be called Method III.

3. WEIGHTED AVERAGE APPROACH

Suppose the observed value $X_{i,j}$ is regarded as a "fixed" or controllable value and is used to predict the random value $X_{i,j+1}$. Since $X_{i,j}$ is not considered a random variable it will be written in lowercase as $x_{i,j}$. To estimate $X_{i,j+1}$, it would make sense to use a weighted average of the available $x_{i,j}$s. The weights are given as

$$w_{i,j} = x_{i,j}^t \Big/ \sum_i x_{i,j}^t,$$

where $\sum_i w_{i,j} = 1.0$. An age-to-age link ratio is then given by

$$\mathrm{LDF}_j = \sum_i w_{i,j} X_{i,j+1} / x_{i,j}.$$

The three methods described in Section 2 can be viewed as special cases of this general weighted average. Table 2 relates the methods and weights.

TABLE 2

WEIGHTS USED

Method	"t"	Weight
I	0	$1/n$
II	1	$x_{i,j} / \sum_i x_{i,j}$
III	2	$x_{i,j}^2 / \sum_i x_{i,j}^2$

If the statistics $X_{i,j+1}/x_{i,j}$ are from the same distribution (or different distributions with the same mean), then the weighted averages will be unbiased since the weights sum to one. This may not hold for $X_{i,j+1}/X_{i,j}$, where the denominator is viewed as a random variable, as will be discussed later.

Assuming for the time being that $x_{i,j}$ is fixed, it could be helpful to consider the variance of $X_{i,j+1}$ in deciding which set of weights to use. In some cases, the variance of $X_{i,j+1}$ for a given $x_{i,j}$ may depend on the size of $x_{i,j}$. For example, a "large" value of $x_{i,j}$ could typically be followed by a small variance in $X_{i,j+1}$.

If the $X_{i,j+1}$s are independent and their variances for a given $x_{i,j}$ are given by $s_{i,j+1}^2$, then define the random variable

$$K_j = \sum_i w_{i,j} X_{i,j+1} / x_{i,j}, \qquad \text{and}$$

$$\text{Var}(K_j) = \sum_i w_{i,j}^2 / x_{i,j}^2 s_{i,j+1}^2. \tag{3.1}$$

If the variance of $X_{i,j+1}$ for a given $x_{i,j}$ depends on the size of $x_{i,j}$, one possible way to relate the two is to consider $s_{i,j+1}^2$ to be proportional to $x_{i,j}^r$:

$$s_{i,j+1}^2 \propto x_{i,j}^r. \tag{3.2}$$

Note that $r < 0$ is possible and would imply an inverse relationship between the size of loss and the subsequent variance.

Substituting the right side of Equation 3.2 in Equation 3.1 yields

$$\text{Var}(K_j) \propto \sum_i x_{i,j}^{r-2} w_{i,j}^2. \tag{3.3}$$

The variance of K_j as a function of $x_{i,j}$ is developed here to help choose weights and therefore a reserving method. As Stanard points out, an estimator should be unbiased and have a minimum variance.

It can be shown (see the appendix) that the weight structure that minimizes the variance of K_j is

$$w_{i,j} = x_{i,j}^{2-r} \Big/ \sum_i x_{i,j}^{2-r}.$$

This leads to choosing the usual arithmetic averages (Method I) if $r = 2$, Method II if $r = 1$, and Method III, the least squares estimator, if $r = 0$.

Applying all of this to a loss development triangle, the question is whether the variance of the sum of losses at a particular point in time is dependent on a previous measure of losses. One way to check differing variances at various levels of a predictor variable $x_{i,j}$ is to plot the residuals. Unfortunately, there aren't enough points to look at in most loss reserving situations even if a consistent relationship between accident years is assumed. In some cases, however, one may believe that greater early development of losses commonly reduces the variance of the next period loss level. If this is the case, it would make sense to choose r less than zero.

Exhibit 1 displays the results of applying the Methods I, II, and III using the simulation procedure outlined by Stanard. Recall that Methods I, II, and III correspond to r values of 2, 1, and 0, respectively, depending on the variance assumption. Also tested are weighting schemes where r is set equal to -1 and -2. This would correspond to the case where there is an inverse relationship between the variance and the previous size of loss as discussed above. It is interesting to note that the mean prediction error decreases as r decreases.

These results show that $r = 0$ (Method III) produces the smallest prediction error for the current accident year, but the prediction of previous accident years can be improved by using r less than zero. Given a knowledge of the underlying structure of loss development, as is the case in this simulation model, it would be possible to choose an optimal value of r for the specific structure. In fact, r doesn't have to be restricted to integers; it could take on any real value and even vary by accident year. Finding an optimal r would be nearly impossible with actual loss data due to the lack of sufficient data and changes in underlying reporting patterns. But it could be possible to find a range of r values that would improve estimates.

4. AGE-TO-AGE FACTORS—LOG-NORMAL MODEL

If we regard $x_{i,j}$ more realistically as an observation of a random sum $X_{i,j}$ at time j, followed next period by loss $X_{i,j+1}$, then pairing them, $(X_{i,j}, X_{i,j+1})$, adds another dimension to evaluating their relationship.

Stanard points out in his appendix that, in general,

$$\mathrm{E}[Y/X] \neq \mathrm{E}[Y]/\mathrm{E}[X].$$

In the case of losses emerging and or developing and the notation used here,

$$\mathrm{E}[X_{i,j+1}/X_{i,j}] \neq \mathrm{E}[X_{i,j+1}]/\mathrm{E}[X_{i,j}].$$

So, using the average of development factors to develop ultimate losses could lead to incorrect conclusions.

For ease of presentation, the random variables $X_{i,j}$ and $X_{i,j+1}$ will be represented by X_1 and Y_1, respectively, from here on in this section. Using this notation, the issue is, what is the expected value of the statistic $Z_1 = Y_1/X_1$? To investigate Z_1, the pair of losses (X_1, Y_1) will be modeled as an element of the joint bivariate log-normal distribution where X_1 and Y_1 are possibly related via a correlation coefficient. Other joint distributions may be appropriate, and the choice depends on the characteristics of the data in question. The log-normal leads to very convenient computations, as will be seen.

If and only if X_1 and Y_1 are jointly log-normal, then $X = \ln(X_1)$, and $Y = \ln(Y_1)$ would be joint normal variables. In this case, a loss development factor is given by the statistic

$$\begin{aligned} Z_1 &= Y_1/X_1 \\ &= \exp(Y)/\exp(X) \\ &= \exp(Y - X). \end{aligned}$$

This form is convenient due to the fact that the expected value of Z_1 is easy to find using the moment generating function of the bivariate normal. $M(t_1, t_2)$ will denote the moment generating function of the bivariate normal with the following parameters:

$$\mu_x = \text{mean of } X,$$

$$\mu_y = \text{mean of } Y,$$

$$\sigma_x = \text{standard deviation of } X,$$

$$\sigma_y = \text{standard deviation of } Y, \text{ and}$$

$$\rho = \text{correlation coefficient of } X \text{ and } Y,$$

where $X = \ln(X_1)$ and $Y = \ln(Y_1)$.

$$M(t_1, t_2) = \exp[t_1\mu_x + t_2\mu_y + (t_1^2\sigma_x^2 + 2\rho t_1 t_2 \sigma_x \sigma_y + t_2^2\sigma_y^2)/2].$$

So,

$$\begin{aligned} \mathrm{E}[Z_1] &= \mathrm{E}[Y_1/X_1] = \mathrm{E}[\exp(Y - X)] \\ &= M(-1, 1) \\ &= \exp[\mu_y - \mu_x + (\sigma_x^2 - 2\rho\sigma_x\sigma_y + \sigma_y^2)/2]. \end{aligned}$$

Since

$$\begin{aligned} \mathrm{E}[X_1] &= M(1) \\ &= \exp(\mu_x + \sigma_x^2/2), \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}[Y_1] &= M(1) \\ &= \exp(\mu_y + \sigma_y^2/2), \end{aligned}$$

then

$$\mathrm{E}[Y_1]/\mathrm{E}[X_1] = \exp[\mu_y - \mu_x + (\sigma_y^2 - \sigma_x^2)/2].$$

Getting back to the question of whether $\mathrm{E}[Y_1/X_1] \neq \mathrm{E}[Y_1]/\mathrm{E}[X_1]$, define the ratio

$$\begin{aligned} d &= \mathrm{E}[Y_1/X_1]/[\mathrm{E}[Y_1]/\mathrm{E}[X_1]] \\ &= \frac{\exp[\mu_y - \mu_x + (\sigma_x^2 - 2\rho\sigma_x\sigma_y + \sigma_y^2)/2]}{\exp[\mu_y - \mu_x + (\sigma_y^2 - \sigma_x^2)/2]} \\ &= \exp(\sigma_x^2 - \rho\sigma_x\sigma_y). \end{aligned}$$

But

$\rho = \sigma_{xy}/(\sigma_x\sigma_y)$ where σ_{xy} is the covariance of X and Y.

So,

$$d = \exp(\sigma_x^2 - \sigma_{xy}).$$

This ratio d is the theoretical ratio of the expected straight average LDFs to the expected weighted average LDFs. Note that d is greater than 1.0 when $\sigma_x^2 > \sigma_{xy}$ and $\mathrm{E}[Y_1/X_1] > \mathrm{E}[Y_1]/\mathrm{E}[X_1]$.

To investigate d, the following simple model of loss development similar to Stanard's is created. Assume:

1. Losses from a Pareto severity:

$$F(x) = 1 - (15{,}000/(15{,}000 + x))^3;$$

2. A normal frequency (mean = 50, variance = 25);

3. An exponential reporting pattern:

$$P(n) = 1 - \exp(0.75n);$$

4. Five "periods" are produced (so if the report time is greater than 5 it is not in the data); and

5. 1,000 samples are produced.

The parameters of the log-normal can be estimated from the sample data using the moments of the transformed variable $\ln(X_1)$. For example,

$$m_x = \sum_i \ln(X_{1i}) \approx \mu_x, \qquad \text{and}$$

$$s_x = \left[\sum_i \ln(X_{1i})^2 - m_x^2\right]^{1/2} \approx \sigma_x.$$

Some statistics of the log transformed sample data by age of development are shown in Table 3. The correlations and covariances are between ages one and two, two and three, etc.

TABLE 3

LOG TRANSFORMED SAMPLE DATA

Age	1	2	3	4	5
Mean	12.12740	12.54222	12.68834	12.74793	12.77808
Variance	0.125657	0.086754	0.071959	0.066450	0.064028
Skew	−0.11204	−0.08357	−0.03813	−0.01074	0.032122
Correlation	0.811968	0.918299	0.970066	0.980749	
Covariance	0.084777	0.072555	0.067079	0.063972	

The next step is to calculate average loss development factors based on the loss data. These would be $\sum_i [Y_{1i}/X_{1i}]$ for straight average (Method I) LDFs and m_y/m_x for weighted average (Method II) LDFs. Four average LDFs are available linking each period:

Age-to-Age	1–2	2–3	3–4	4–5
Straight Average	1.549429	1.165989	1.063784	1.032007
Weighted Average	1.485233	1.149049	1.058583	1.029500

Now, according to the d ratio, the ratio of the straight average to weighted average LDFs from the sample data should be approximately

$$d = \exp(\sigma_x^2 - \sigma_{xy})$$

if the distributions are approximately jointly log-normal. The various values turn out to be:

Age-to-Age	1–2	2–3	3–4	4–5
Ratio	1.043223	1.014742	1.004913	1.002434
d	1.041727	1.014299	1.004891	1.002481

where, for example, Ratio 1–2 is 1.043223 = 1.549429/1.485233 and d for 1–2 is 1.041727 = exp(0.125657 − 0.084777).

Since the theoretical values and the "experimental" values are so close, it is worth the effort to check the distributions of the simulated losses at each period. The Kolmogorov–Smirnov or K–S statistic is helpful in measuring the "closeness" of an empirical distribution to a continuous assumed distribution. The hypothesis H_0 would be that the sampled distributions are normal after the $\ln(X_1)$ transformation. The statistic

$$\text{Max}[|F(x) - F_n(x)|]n^{1/2} > 1.36$$

is significant at the 95% level, where n is the number of data points. A high value indicates a poor fit and rejection of H_0.

For the standardized log transformed data:

Age	1	2	3	4	5
K–S	0.5492	0.4245	0.7075	0.6094	0.5209
Maximum Difference	0.0174	0.0134	0.0224	0.0193	0.0164

The distributions of the standardized log transformed sums of Pareto variables by period are apparently very closely approximated by a standard normal distribution, and the joint log-normal assumption appears to be valid.

The following were calculated using untransformed standardized data from the sample:

Age	1	2	3	4	5
K–S	2.1892	1.6347	1.7544	1.4142	1.6633
Maximum Difference	0.0693	0.0517	0.0555	0.0447	0.0526

These data indicate that a bivariate normal assumption would not be appropriate for this data.

Concluding this section, the answer to the question "What is the expected value of an LDF?" is that it depends on the joint distribution of the losses. The joint log-normal allowed for the determination of expected LDFs in terms of the parameters of the underlying variables. It would be possible to use a similar analysis on actual loss data if reasonable estimates of the distributions of losses by age could be found. Also, this analysis could be extended to the product of LDFs.

5. SUMMARY

Exhibit 1 displays the results of the three loss development methods given in Section 2 using Stanard's simulation routine. Methods II and III are clearly superior in terms of both bias and variance. To the extent that actual loss development patterns are like those simulated, Methods II and III would be preferred over Method I. As noted above, other weighting schemes may produce even better results.

Method I, the straight averaging of LDFs, shows the greatest positive bias. Part of this bias could be explained by the analysis of $E[Y/X]$ in Section 4. An obvious conclusion is that straight average LDFs will overstate projected ultimate losses, at least according to these models. However, if a selection criterion is used, such as excluding the high and low LDFs or judgment based on other information, the straight average LDFs would likely produce better results in terms of average error. The goal of the discussion here is to determine general underlying characteristics of LDFs and age-to-age methods that could possibly have a bearing on decision making.

The idea of correlation between random sums measured at successive points in time could give more insight into the selection of loss development factors and age-to-age factor methods in general. An understanding of how the aggregate distribution of losses changes with time would be a valuable tool.

REFERENCES

[1] Degroot, Morris H., *Probability and Statistics*, Addison-Wesley, 1986, pp. 300–304.

[2] Hogg, Robert and Allen Craig, *Introduction to Mathematical Statistics*, Macmillan, 1978, p. 180.

[3] Kleinbaum, David, Lawrence Krupper and Keith Muller, *Applied Regression Analysis and Other Multivariable Methods*, PWS-KENT, 1988, pp. 84–85.

[4] Mood, Alexander, Franklin Graybill and Duane Boes, *Introduction to the Theory of Statistics*, McGraw-Hill, Third Edition, 1974, p. 164.

[5] Mosteller, Frederick, Robert Rourke and George Thomas, *Probability with Statistical Applications*, Addison-Wesley, 1970, pp. 400–401.

[6] Murphy, Daniel M., "Unbiased Loss Development Factors," *PCAS* LXXXI, 1994, pp. 154–222.

EXHIBIT 1

RESULTS OF LOSS DEVELOPMENT METHODS

MEAN PREDICTION ERROR

		Accident Year			
Method	*r*	1	2	3	4
I	2	13,627	31,498	83,862	482,307
II	1	13,627	21,887	40,185	121,218
III	0	13,627	16,397	17,110	13,056
	−1	13,627	13,532	5,883	−26,840
	−2	13,627	11,958	67	−44,583

STANDARD DEVIATION OF PREDICTION ERROR

		Accident Year			
Method	*r*	1	2	3	4
I	2	170,234	285,556	391,868	2,406,638
II	1	170,234	278,987	347,260	857,741
III	0	170,234	277,716	345,466	672,590
	−1	170,234	277,909	353,319	613,091
	−2	170,234	278,408	363,256	592,641

APPENDIX

The subscript j will not be used in the appendix for clarity. The goal here is to find w_i such that $\text{Var}(K)$ is minimized. If

$$\text{Var}(K) = h(t) = \sum_i x_i^{r-2} w_i^2,$$

where

$$w_i^2 = x_i^{2t} \bigg/ \left(\sum_i x_i^t\right)^2, \quad \text{then}$$

$$h(t) = \sum_i x_i^{2t+r-2} \bigg/ \left(\sum_i x_i^t\right)^2 = f(t)/g(t),$$

$$f'(t) = 2\sum_i x_i^{2t+r-2} \ln(x_i),$$

$$g'(t) = 2\sum_i x_i^t \sum_i x_i^t \ln(x_i), \quad \text{and}$$

$$h'(t) = (g'f - f'g)/g^2.$$

Since $g^2 > 0$, we need to find t to set the numerator equal to 0 or $g'f = f'g$. With some factoring this reduces to

$$\sum_i x_i^t \sum_i x_i^{2t+r-2} \ln(x_i) = \sum_i x_i^{2t+r-2} \sum_i x_i^t \ln(x_i).$$

By inspection, $t = 2 - r$ solves this equation.

Using the first derivative test, it will be shown that, as t passes through $2 - r$, the sign of $h'(t)$ changes from negative to positive, indicating that this is a minimum. That is, show

1. If $t < 2 - r$ then

$$\sum_i x_i^t \sum_i x_i^{2t+r-2} \ln(x_i) < \sum_i x_i^{2t+r-2} \sum_i x_i^t \ln(x_i) \quad \text{(A.1)}$$

   and $h'(t)$ is negative.

2. If $t > 2 - r$ then $h'(t)$ is positive.

First, let

$$t < 2 - r.$$

Then

$$2t < 2 - r + t,$$

and

$$2t + r - 2 < t.$$

Also, let

$$x_i > 1.0 \qquad \text{for all} \quad i, \text{ and } x_i \neq x_j \text{ for at least one } (i, j).$$

These two conditions are easily met for the loss data being considered. Since

$$t > 2t + r - 2,$$
$$x_i^t > x_i^{2t+r-2},$$

and

$$\sum_i x_i^t > \sum_i x_i^{2t+r-2}.$$

Equation A.1 is equivalent to the inequality

$$\sum_i x_i^{2t+r-2} \ln(x_i) \Big/ \sum_i x_i^{2t+r-2} < \sum_i x_i^t \ln(x_i) \Big/ \sum_i x_i^t. \tag{A.2}$$

For given x_is, the left side is in the form of a weighted average of $\ln(x_i)$ with weights equal to

$$x_i^{2t+r-2} \Big/ \sum_i x_i^{2t+r-2}, \tag{A.3}$$

and the right side is also a weighted average of $\ln(x_i)$ with weights

$$x_i^t \Big/ \sum_i x_i^t.$$

So, if

$$\sum_i x_i^t \ln(x_i) \Big/ \sum_i (x_i^t) \tag{A.4}$$

is a monotonically increasing function of t, Equation A.2 will be satisfied because $2t + r - 2 < t$.

Taking the first derivative of Equation A.4 with respect to t yields

$$\sum_i x_i^t \ln(x_i)^2 \Big/ \sum_i x_i^t - \left(\sum_i (x_i^t \ln(x_i^t)\right)^2 \Big/ \left(\sum_i x_i^t\right)^2 . \tag{A.5}$$

The form of Equation A.5 is algebraically identical to the variance formula

$$\text{Var} = \text{E}[X^2] - \text{E}[X]^2,$$

where the probabilities are the right side weights and the random variable is $\ln(x_i)$.

According to Mood, Graybill and Boes [4], the Jensen inequality says that if X is a random variable with mean $\text{E}[X]$, and $g(x)$ is a convex function, then $\text{E}[g(x)] \geq g(\text{E}[X])$. It follows that this will hold for Equation A.5. In this case $g(x) = x^2$ is convex, so the derivative in Equation A.5 is greater than or equal to zero. In fact, the only case where the derivative equals zero is when the probability of a given X is concentrated at a single point, or in this case $x_i = x_j$ for all (i, j), which isn't allowed. This implies that the derivative is strictly positive and Equation A.4 is monotonically increasing which, in turn, implies that Equation A.1 and Equation A.2 hold since $2t + r - 2 < t$. This means that $h'(t)$ is negative for $t < 2 - r$, which is what we meant to show.

If we now consider condition 2 from above, the same argument holds for $t > 2 - r$, implying that $h'(t)$ is positive. This shows that $h'(t)$ changes sign from negative to positive, and that $t = 2 - r$ is a minimum.