Correlation and the Measurement of Loss Reserve Variability

by Randall D. Holmberg

CORRELATION AND THE MEASUREMENT OF LOSS RESERVE VARIABILITY

Randall D. Holmberg, FCAS

Loss reserves are the largest liability on the balance sheet of an insurance company, yet they are only estimates. Even the actuary responsible for making the estimates is often unable to quantify the inherent uncertainty. This is partly a consequence of the complexity of estimating the variability of the reserve estimates. Correlation across several dimensions makes statistical measurement of uncertainty difficult. Most insurers have only a limited number of historical data points available with which to make estimates of the multiple correlations, making estimation of correlation problematic. This paper presents a mathematically simple model of loss development variability which allows the inclusion of several types of correlation. It can also be adapted to deal with other complexities which may arise in the analysis of reserves. The paper also presents methods which make it easier to estimate correlations in practical applications.

CORRELATION AND THE MEASUREMENT OF LOSS RESERVE VARIABILITY

Loss reserves, generally the largest liability on the balance sheet of an insurer, are only estimates of ultimate loss payments. Even if these estimates as carried on the balance sheet are unbiased, neither deliberately redundant nor inadequate, they are subject to uncertainty. Furthermore, the magnitude of the uncertainty of these estimates is generally unknown, even to the actuary who reviews reserves and certifies their adequacy in statutory financial statements. Considering the importance of reserve estimates to an insurer's reported net worth, it is important to quantify the uncertainty of these estimates.

Attempts to quantify the uncertainty of loss reserve estimates can easily get stymied by the complexity of reserve issues. There is potential for substantial correlation across many dimensions. There are usually relatively few historical data points from which to estimate the multiple correlations that are possible. Yet it is unlikely that correlation is insignificant in the variability of the total reserve estimate. Therefore, actuaries need a model which can deal with correlation but which allows reasonable estimation of the correlations involved.

This paper presents a model for measuring the uncertainty of loss reserve estimates. Its main virtue is the directness and simplicity of the approach. It includes adjustments to account for many of the kinds of correlation effects which arise in analyzing reserves. The data for this measurement will be available in one form or another at any insurance company. The relevance of the items used in the measurement to the question being asked is easy to see. The model is simple enough that it is relatively easy to add features to cope with complications that the model as presented herein does not consider.

While estimation of the correlations involved is difficult in practice, this paper presents several approaches which have proved helpful in making such estimates.

Even when the parameters required by the model are difficult to estimate, the model may be used in sensitivity testing to get a greater appreciation of the importance (or lack of importance) of correlation to the accuracy of reserve estimates. The parameters have clear-cut intuitive interpretations, so sensitivity testing should prove fruitful to a knowledgeable reserve actuary.

The paper will present the model in a relatively simple form and then suggest adaptations to deal with situations of greater complexity. An example of applying the approach is integrated into the description of the model.

BASIC APPROACH

In the property-casualty insurance industry in the United States, actuaries generally rely on a link ratio loss development approach to determine their estimates of accident-year ultimates and hence the adequacy of carried loss reserves. It seems natural to consider the way reserve adequacy is estimated in determining the variance of the resulting estimate. We take a very direct approach. We measure the variance of historical link ratios which the actuary examines when determining projections of future development patterns. From these variances, the variance of the resulting estimate of ultimate is computed. The variance of the estimated ultimate for a single exposure period is equal to the variance of the estimate of reserve shortage or redundancy for that period. The exposure period variance for a single period is then combined with those of other periods to arrive at the variance of total reserve need at a valuation date.

Variance of Link Ratios

In this description of the model, we phrase all discussion in terms of incurred loss development methods. However, this same approach works in a paid loss development context. Similarly, although all references are to "accident year" this model can be used with other exposure periods such as report years or accident quarters.

We will first establish some notation. Let R denote the total IBNR reserve need as of the valuation date in question. In our formulation, R includes provision for adverse or favorable development on known cases. Case reserves are treated as a constant. Therefore, the variability of R is equal to the variability of total reserve need. Let n be the number of accident years and the maximum number of valuations included in our development triangle. Define $l_{i,j}$ as the incurred loss for accident year i as valued j years after the beginning of the accident year. Both i and j are numbered sequentially beginning with 1. Let L_i be the ultimate loss for accident year i. Let $d_{i,j}$ be the link ratio for accident year i between valuations j and $j+1$. Finally, define $D_{i,j}$ as the development factor for accident year i from age j to ultimate. In this formulation, $d_{i,n} = D_{i,n}$ is the tail factor for accident year i. The latest available historical valuation of year i is $l_{i,n-i+1}$. Note the following:

$$d_{i,j} = \frac{l_{i,j+1}}{l_{i,j}} \qquad (1)$$

$$D_{i,j} = \prod_{k=j}^{n} d_{i,k} \qquad (2)$$

$$D_{i,j} = d_{i,j} \cdot D_{i,j+1} \text{ for } j < n \qquad (3)$$

$$L_i = D_{i,n-i+1} \cdot l_{i,n-i+1} \qquad (4)$$

$$E(L_i) = E(D_{i,n-i+1}) \cdot l_{i,n-i+1} \qquad (5)$$

$$R = \sum_i L_i - \sum_i l_{i,n-i+1} \qquad (6)$$

$$E(R) = \sum_i E(L_i) - \sum_i l_{i,n-i+1} \qquad (7)$$

In the traditional link ratio approach,

$$E(D_{i,j}) = \prod_{k=j}^{n} E(d_{i,k}), \text{ and}$$

$$E(d_{i,j}) = E(d_{k,j}) \text{ for all } i, k \text{ for which } j > n-i \text{ and } j > n-k.$$

We will not require that these two relationships hold in our model.

$E(L_i)$ is the estimated ultimate loss. The variance of L_i around this mean is what we will measure to arrive at the variability of loss reserve estimates for a single accident year.

The first step in calculating the variance of accident year ultimates is to calculate the variance of the historical link ratios for each stage of development. Exhibit 1 shows the triangle of incurred losses we will use in our examples. Exhibit 2 is the resulting triangle of link ratios. This is a ten-year triangle, so from it we calculate the variance of all $d_{i,j}$ for a fixed j, for values of $i \le 10-j$. These variances, as well as average link ratios and standard deviations, are also displayed in Exhibit 2. Note that since the variance of the link ratio at age j, $Var(d_{i,j})$, is calculated across all $i \le 10-j$ for a fixed j, we have $Var(d_{i,j}) = Var(d_{k,j})$ even if $i \ne k$ so the first subscript is not needed. In our example, the same is true of the mean at age j, $E(d_{i,j})$. However, we will carry the first subscript throughout for consistency with other notation.

The model treats historical link ratios at a given stage of development as a sample from independent identically distributed random variables. The sample variance calculated from this sample is used as an estimator of the variance of the random variable's distribution.

The actuary's chosen projection of loss development may not match historical averages. Even in such instances, the sample variance is used to estimate the variance of future development, as it represents our best estimate of the variability of future development. However, the chosen development factor is treated as the expected value of that development. The example used in this paper includes some selected link ratios which are not equal to the historical average, in order to illustrate how these selections are treated in this method.

An issue that arises is what variance to assign to the link ratios where there are few historical points, and to the tail development factor. Unless there is reason to do otherwise, in practice we usually rely on the sample variance for all ages where there are two or more historical link ratios. In many cases, assigning zero variance to the last one-year link ratio and to the tail factor is reasonable. In other cases, regressing the standard deviation of link ratios against the quantity $|1-E(d_{i,j})|$ gives a fitted line which can be used to read off the standard deviation of the link ratio or tail factor (limiting standard deviations to non-negative values, of course). Alternate approaches could be used, such as using a parallelogram of link ratios rather than a triangle, or regressing standard deviations against the stage of development j, or simply judgmentally choosing a number for these stages of development. If the data used produce standard deviations which are sufficiently "bumpy" some of these techniques might be required even for stages of development with relatively many historical link ratios available. In the example used here, we have used sample variances where available and have assumed zero variance for the last stage of development seen in the experience and for the tail factor. This is seen in Exhibits 2 and 3.

Following equation (4), since $l_{i,n-i+1}$ is a known constant, we have:

$$Var(L_i) = Var(D_{i,n-i+1}) \cdot (l_{i,n-i+1})^2. \qquad (8)$$

Therefore, determining $Var(D_{i,n-i+1})$ for each i will determine the variance of each accident year ultimate. $Var(D_{i,n-i+1})$ is a function of the $Var(d_{i,j})$ for all $j \geq n-i+1$. However, $Var(D_{i,n-i+1})$ also depends in part on the correlation between link ratios at different stages of development within a single accident year.

Correlation Between Stages of Development

There are different reasons we might expect development at different stages to be correlated. For instance, if unusually high loss development in one period were the result of accelerated reporting, subsequent development would be lower than average as the losses that would ordinarily be reported in those later periods would have already been reported. In this instance, correlation between one stage and subsequent stages would be negative. Positive correlation would occur if there were a tendency for weaker-than-average initial reserving to be corrected over a period of several years. In that case, an unusually high degree of development in one period would be a warning of more to come. These examples do not exhaust the possible reasons for correlation.

The usual link ratio approach, constructing development factors to ultimate as $E(D_{i,j}) = \prod_{k=j}^{n} E(d_{i,k})$, implicitly assumes the stages of development are uncorrelated. If the $d_{i,j}$ were independent for different values of j within a fixed i, we would have the following:

$$Var(D_{i,j}) = E^2(d_{i,j}) \cdot Var(D_{i,j+1}) + Var(d_{i,j}) \cdot E^2(D_{i,j+1}) + Var(d_{i,j}) \cdot Var(D_{i,j+1}) \qquad (9)$$

Since for the tail factor ($d_{i,10}$ in our example) $d_{i,n} = D_{i,n}$, we could start with the tail factor and use equation (9) with our variances of link ratios to chain backward and build $Var(D_{i,n-i+1})$ for all i. Here we generalize to a situation where correlation among stages of development may exist.

Hayne approached the problem of correlated link ratios using an assumption of multivariate lognormality. We propose an approach which is perhaps more intuitive and which is certainly simpler. As a consequence, our approach lacks much of the elegance of Hayne's, but it provides a model which can yield significant insights into the effects of correlation between link ratios.

Our mathematically convenient model for correlation treats a single link ratio, $d_{i,j}$, and the following age-to-ultimate factor, $D_{i,j+1}$, as correlated. We postulate a distribution for $d_{i,j}$ and a relationship between $d_{i,j}$, and $D_{i,j+1}$ which allow relatively easy calculation of means and variances while still permitting the inclusion of various correlation effects. For mathematical tractability, we assume $d_{i,j}$ is uniformly distributed with known mean and variance. Assume a uniformly distributed random variable $X_{i,j}$, stochastically independent of $d_{i,j}$ with a relation as follows:

$$D_{i,j+1} = a \cdot d_{i,j} + b \cdot X_{i,j} \qquad (10)$$

$$\text{where } a + b = 1 \qquad (11)$$

If the correlation coefficient between $d_{i,j}$ and $D_{i,j+1}$, ρ, is known[1], we have enough information to solve for a and b, and for the lower and upper bounds of the range of $X_{i,j}$,, which we will call $A_{X_{i,j}}$ and $B_{X_{i,j}}$. We can then calculate $E(D_{i,j})$ and $Var(D_{i,j})$.

[1]We will discuss the estimation of ρ later. For the moment, assume the value of ρ is known.

The derivation of these results follows. We have the following as a consequence of (10):

$$E(D_{i,j+1}) = a \cdot E(d_{i,j}) + b \cdot E(X_{i,j}) \tag{12}$$

$$Var(D_{i,j+1}) = a^2 \cdot Var(d_{i,j}) + b^2 \cdot Var(X_{i,j}) \tag{13}$$

As a consequence of (10), we have the following for the correlation coefficient between $d_{i,j}$ and $D_{i,j+1}$ (see e.g. Sachs):

$$\rho = a \cdot \sqrt{\frac{Var(d_{i,j})}{Var(D_{i,j+1})}}$$

$$\therefore \quad a = \rho \cdot \sqrt{\frac{Var(D_{i,j+1})}{Var(d_{i,j})}} \tag{14}$$

Having determined a we can further calculate:

$$b = 1 - a \tag{15}$$

From (12):

$$E(X_{i,j}) = \frac{E(D_{i,j+1}) - a \cdot E(d_{i,j})}{b} \tag{16}$$

From (13):

$$Var(X_{i,j}) = \frac{Var(D_{i,j+1}) - a^2 \cdot Var(d_{i,j})}{b^2} \tag{17}$$

We want to know $E(d_{i,j} \cdot D_{i,j+1}) = E(D_{i,j})$, and $Var(d_{i,j} \cdot D_{i,j+1}) = Var(D_{i,j})$. We will calculate these by specifying the distributions of $X_{i,j}$ and $d_{i,j}$ and integrating the appropriate expressions over the relevant domains. First we determine these domains.

For a random variable Y, uniformly distributed on the interval between A_Y and B_Y,

$$E(Y) = \frac{A_Y + B_Y}{2} \tag{18}$$

$$Var(Y) = \frac{(B_Y - A_Y)^2}{12} \tag{19}$$

So we can derive $A_{X_{i,j}}, B_{X_{i,j}}, A_{d_{i,j}}$, and $B_{d_{i,j}}$. For notational convenience, we will use A_X to denote $A_{X_{i,j}}$, B_X to denote $B_{X_{i,j}}$, A_d to denote $A_{d_{i,j}}$, and B_d to denote $B_{d_{i,j}}$ in what follows. We can determine the bounds of $X_{i,j}$ using formulae following from (18) and (19):

$$A_X = E(X_{i,j}) - \frac{[12 \cdot Var(X_{i,j})]^{\frac{1}{2}}}{2} \tag{20}$$

$$B_X = E(X_{i,j}) + \frac{[12 \cdot Var(X_{i,j})]^{\frac{1}{2}}}{2} \tag{21}$$

Analogous formulae give the values of A_d and B_d. Now we can set up integrals and calculate $E(D_{i,j})$ and $Var(D_{i,j})$.

$$E(D_{i,j}) = E(d_{i,j} \cdot D_{i,j+1}) = \frac{1}{(B_X - A_X)\cdot(B_d - A_d)} \cdot \int_{A_d}^{B_d}\int_{A_X}^{B_X} d_{i,j} \cdot (a \cdot d_{i,j} + b \cdot X_{i,j}) dX_{i,j} dd_{i,j}$$

$$= \frac{\frac{a}{3}\cdot(B_X - A_X)\cdot(B_d^3 - A_d^3) + \frac{b}{4}\cdot(B_X^2 - A_X^2)\cdot(B_d^2 - A_d^2)}{(B_X - A_X)\cdot(B_d - A_d)} \tag{22}$$

$$E(D_{i,j}^2) = E(d_{i,j}^2 \cdot D_{i,j+1}^2) = \frac{1}{(B_X - A_X)\cdot(B_d - A_d)} \cdot \int_{A_d}^{B_d}\int_{A_X}^{B_X} d_{i,j}^2 \cdot (a \cdot d_{i,j} + b \cdot X_{i,j})^2 dX_{i,j} dd_{i,j}$$

$$= \frac{\frac{a^2}{5}(B_X - A_X)(B_d^5 - A_d^5) + \frac{a \cdot b}{4}(B_X^2 - A_X^2)(B_d^4 - A_d^4) + \frac{b^2}{9}(B_X^3 - A_X^3)(B_d^3 - A_d^3)}{(B_X - A_X)\cdot(B_d - A_d)} \tag{23}$$

$$Var(D_{i,j}) = E(D_{i,j}^2) - E^2(D_{i,j}). \tag{24}$$

We chain backward to calculate $E(D_{i,n-i+1})$ and $Var(D_{i,n-i+1})$ for all i, allowing us to calculate $Var(L_i)$. The way we do this is as follows. Start with the tail factor $D_{i,n}$ so that $E(D_{i,n}) = E(d_{i,n})$ and $Var(D_{i,n}) = Var(d_{i,n})$, quantities we have estimated or assumed. We also know $E(d_{i,n-1})$ and $Var(d_{i,n-1})$. Use (14) and (15) to calculate a and b. Use (16) and (17) to compute $E(X_{i,n-1})$ and $Var(X_{i,n-1})$. Calculate A_X, B_X, A_d, and B_d

using (20) and (21) and the analogous equations for A_d, and B_d. Finally, calculate $E(D_{i,n-1})$ and $Var(D_{i,n-1})$ using (22), (23), and (24). We can repeat this process, using $E(D_{i,n-1})$ and $Var(D_{i,n-1})$ to estimate $E(D_{i,n-2})$ and $Var(D_{i,n-2})$. We continue backward in this fashion until we reach $E(D_{i,n-i+1})$ and $Var(D_{i,n-i+1})$, allowing us to calculate $Var(L_i)$.

Correlation affects both $E(D_{i,j})$ and $Var(D_{i,j})$. Beyond this, there are important conditional expectations and variances, $E(D_{i,j}|d_{i,j-1})$ and $Var(D_{i,j}|d_{i,j-1})$ for $1 < j \leq n-i+1$. If we believe that link ratios and the following age-to-ultimate development factors are correlated, then knowledge of the last historical link ratio for each accident year should affect both our expectation of future development on that year and the variance of our estimated ultimate. It is internally consistent if in projecting ultimate losses and in estimating the variance of those ultimates we use conditional expectations and variances per the following:

from (5): $E(L_i) = l_{i,n-i+1} \cdot E(D_{i,n-i+1}|d_{i,n-i})$, if $i < n$ (25)

from (8): $Var(L_i) = (l_{i,n-i+1})^2 \cdot Var(D_{i,n-i+1}|d_{i,n-i})$, if $i < n$ (26)

from (5): $E(L_n) = l_{n,1} \cdot E(D_{n,1})$ (27)

from (8): $Var(L_n) = (l_{n,1})^2 \cdot Var(D_{n,1})$ (28)

Note that for $i < n$:

from (12): $E(D_{i,n-i+1}|d_{i,n-i}) = a \cdot d_{i,n-i} + b \cdot E(X_{i,n-i})$ (29)

from (13): $Var(D_{i,n-i+1}|d_{i,n-i}) = b^2 \cdot Var(X_{i,n-i})$ (30)

Exhibit 3 shows a calculation of conditional expectations and variances. For informational purposes, it also shows unconditional expectations and variances including correlation, and expectations and variances excluding correlation. Exhibit 3 uses the same value of ρ for all i and all j, but this is not a requirement of the model. However, when the question of estimating ρ arises, there is a benefit to having a single value to estimate, and there may be some intuitive appeal to having a single value of ρ.

At this point it is helpful to go through Exhibit 3 step by step to clarify how the model is used in practice. At the top of the exhibit, we show for each stage of development j the expected link ratio $E(d_{i,j})$ and the variance of that link ratio $Var(d_{i,j})$. $E(d_{i,j})$ is equal to our selected link ratio. $Var(d_{i,j})$ is estimated using the sample variance as shown in Exhibit 2. In this example, neither $E(d_{i,j})$ nor $Var(d_{i,j})$ vary with i, as has been noted previously. However, the model could cope with different $E(d_{i,j})$ and $Var(d_{i,j})$ for different i.

The next item in Exhibit 3 is our value of ρ, which in this example is the same for all i, j. From this point on, the exhibit is easier to interpret if we start on the right of the exhibit and work our way to the left. At each j, we determine a using equation (14) and b using (15). For $j \geq 8$, we have $a = 0$, since $Var(D_{i,j+1}) = 0$. For lesser j, we use the value of $Var(D_{i,j+1})$ from the next column to the right to calculate a and b. $E(X_{i,j})$ and $Var(X_{i,j})$, calculated using (16) and (17) respectively, are shown next, followed by A_X, B_X, which follow using (20) and (21). The values of $E(d_{i,j})$ and $Var(d_{i,j})$ from the top of the exhibit similarly allow us to calculate A_d, and B_d which are the next values shown in Exhibit 3. Then, using (22) and (23), we can calculate $E(D_{i,j})$ and $Var(D_{i,j})$. These values flow into the calculation for the next column to the left. For each j, once we have completed the column to this point we have enough information to proceed with the calculations for the column for $j-1$.

Below these unconditional means and variances, we calculate the conditional values $E(D_{n-j+1,j}|d_{n-j+1,j-1})$ and $Var(D_{n-j+1,j}|d_{n-j+1,j-1})$ using equations (25) through (28), which will be used to project the ultimate and the variance of that ultimate for accident year

$n-j+1$. These calculations require some parameters from the column for $j-1$, namely a, b, $E(X_{n-j+1,j-1})$, and $Var(X_{n-j+1,j-1})$.

Note that there are no conditional expectation and variance for $j=1$ or equivalently for $i=n$. This is of course because there is no $d_{n,0}$.

The effect of correlation on unconditional $E(D_{i,11-i})$ is relatively small, but the effect on unconditional $Var(D_{i,11-i})$ is significant, when compared to the values ignoring correlation. When conditional expectations and variances are used, both expectations and variances are significantly affected by correlation.

The assumption of a uniformly distributed $X_{i,j}$ and $d_{i,j}$ is primarily for mathematical convenience in determining variances of the product of correlated random variables. It is not intended to represent a realistic model of the probability distribution of the link ratios or age-to-ultimate development factors. Thus, the actuary may decide that using conditional probabilities and variances is putting too much reliance on a model which was chosen largely for convenience. In such a case, the actuary might base estimated ultimates on the traditional age-to-ultimate factor as the product of projected link ratios (implicitly ignoring correlation for the purpose of projecting ultimates), but use the variances including correlation, either conditional or unconditional. Alternatively, he or she might use the unconditional $E(D_{i,j})$ including correlation (recognizing correlation among future development, but ignoring the correlation to historical link ratios) instead of relying on conditional $E(D_{i,j}|d_{i,j-1})$ from this model. Note that as a consequence of (13) and (30), $Var(D_{i,j}|d_{i,j-1}) \leq Var(D_{i,j})$. Hence, the use of the correlated unconditional $Var(D_{i,j})$ would be conservative. The correlated $E(D_{i,j})$ is greater than the age-to-ultimate excluding correlation if correlation is positive, less than the uncorrelated age-to-ultimate if correlation is negative (assuming that loss development is positive). $E(D_{i,j}|d_{i,j-1})$ may be greater or smaller than $E(D_{i,j})$ depending on whether $d_{i,j-1}$ is greater or smaller than the average link ratio at that stage, and whether correlation is positive or negative.

Exhibit 4 shows a calculation where $E(D_{i,n-i+1}|d_{i,n-i})$ and $Var(D_{n-i+1}|d_{i,n-i})$ are used for calculating expected ultimates and the variance of those ultimates, based on equations (25) through (28). In our example we use conditional expectations and variances in the interest of internal consistency, as discussed earlier. For convenience in later calculations, Exhibit 4 shows variances converted to standard deviations.

Following from equation (6), we have $Var(R) = Var\left(\sum_i L_i\right)$. If the L_i for different i were independent, we could calculate the variance of the estimated total reserve need R as

$$Var(R) = \sum_i Var(L_i) \qquad (31)$$

However, the model does not require this assumption of independence, as will be seen below.

Correlation Between Accident Years

There are reasons that the estimated ultimates for different accident years as of a given valuation date might be positively correlated. If current case reserves are stronger (or weaker) at the valuation date than assumed implicitly in the projected development pattern for one accident year, it is likely to be true for all accident years. If claim processing has been disrupted in some way, that may very well affect all accident years. If a judicial decision changes the likelihood of paying out on certain types of claims, that could affect all accident years. There are doubtless other examples of contingencies which could cause positive correlation. It is less clear, at least to us, what realistic contingencies in property-casualty insurance could result in negative correlation, although hypothetical examples can be created.

One concern which must be noted at this point is that some of the situations which can cause correlation between accident years can also cause correlation between stages of development within an accident year. The prior section dealt with measuring the correlation between stages of development within an accident year. It is important in adding consideration of correlation between accident years that we avoid "double-counting" the correlation which results from the same cause as the correlation within an accident year. A method we propose to avoid (or at least ameliorate) the potential double-counting, without going into the complexities of multivariate analysis, is described in the next section on estimating correlation coefficients. The current section describes the mechanics of including correlation between accident years in our measure of the variability of total reserves.

We start with a formula for the variance of the sum of generalized random variables Y_k.

$$Var\left(\sum_{k=1}^{n} Y_k\right) = \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} Cov\left(Y_{k_1}, Y_{k_2}\right)$$

Note that $Cov(Y_k, Y_k) = Var(Y_k)$, and that each term $Cov\left(Y_{k_1}, Y_{k_2}\right)$, for $k_1 \neq k_2$ appears twice in this sum. Thus, for example, this formula would agree with the familiar formula for the variance of the sum of two random variables,

$$Var(Y+Z) = Var(Y) + Var(Z) + 2 \cdot Cov(Y,Z).$$

We approach the calculation of the variance of R through use of a correlation matrix. Since $Cov(L_k, L_m) = \rho_{k,m} \cdot Var(L_k)^{1/2} \cdot Var(L_m)^{1/2}$, where $\rho_{k,m}$ is the correlation coefficient between L_k and L_m, defining a matrix of the values of $\rho_{k,m}$ is the first step to calculating the variance of R.

Set up an $n \times n$ matrix C, with $c_{km} = c_{mk} = \rho_{k,m}$ for all k, m.[2] $Var(L_k)$ is known for each k based on the work done in the previous section. Therefore, we calculate

$$Var(R) = \sum_{k=1}^{n} \sum_{m=1}^{n} Cov(L_k, L_m) = \sum_{k=1}^{n} \sum_{m=1}^{n} \rho_{k,m} \cdot Var(L_k)^{1/2} \cdot Var(L_m)^{1/2} \quad (32)$$

Exhibit 5 shows the calculation of $Var(R)$ in this manner. The exhibit shows the matrix of correlation coefficients first, i.e. the matrix of $\rho_{k,m}$. It then illustrates the calculation of the matrix of covariances using the accident year standard deviations calculated in Exhibit 4. Each element in the second matrix is a term from equation (32). The sum of all elements in the second matrix is equal to the displayed "Variance of Estimated Reserve Need" at the bottom of Exhibit 5. This is converted to a standard deviation for the reader's convenience.

Estimating Correlation Coefficients

The inclusion of correlation has significant effects on both estimated IBNR reserve need and the variance of that estimate. Exhibits 4 and 5 show an estimated needed IBNR of \$69,879 and a standard deviation of that estimate of \$21,492. If we had used zero correlation everywhere in the model, the corresponding numbers would have been \$68,325 estimated IBNR need with a standard deviation of \$14,717. Clearly, the existence of correlation is an important factor in measuring the variability of reserve estimates.

---

[2]Again, take $\rho_{k,m}$ as given. The problem of estimating these correlations is treated later.

How do we determine the correlation ρ within each accident year and the correlations $\rho_{k,m}$ between accident years at a given valuation point? How do we account for the fact that some of the correlation captured in ρ is caused by the same factors that result in correlation measured in the $\rho_{k,m}$? The first step in our recommended approach is to estimate ρ, without any consideration of collinearity, as described below.

One approach which can be used to estimate ρ is an iterative approach based on the incurred loss triangle being analyzed. In our experience convergence is usually pretty rapid, taking 3 to 5 iterations.

On the first pass, treat ρ as if it were zero. Take the resulting estimated ultimates $E(L_i) = l_{i,n-i+1} \cdot \prod_{k=n-i+1}^{n} E(d_{i,k})$ and use them to calculate implied $D'_{i,j} = E(L_i)/l_{i,j}$ for $j \le n-i$ (that is for all j with at least one historical link ratio in addition to the projected development). Transform all $D'_{i,j}$ and all $d_{i,j-1}$ to random variables with mean zero and standard deviation one using these formulae:

$$D''_{i,j} = \frac{(D'_{i,j} - E(D'_{i,j}))}{(Var(D'_{i,j}))^{1/2}} \qquad (33)$$

$$d''_{i,j-1} = \frac{(d_{i,j-1} - E(d_{i,j-1}))}{(Var(d_{i,j-1}))^{1/2}}$$

where expectations and variances are calculated across varying i within a fixed j. Calculate the sample correlation coefficient between these $d''_{i,j-1}$ and $D''_{i,j}$ for all i, j such that $2 \le j \le n-i$. Use this sample correlation coefficient as ρ for all i, j in all calculations. If the ultimates are being projected using $E(L_i) = l_{i,n-i+1} \cdot \prod_{k=n-i+1}^{n} E(d_{i,k})$, this is the final estimate of ρ. If ultimates are being projected using $E(D_{i,j}|d_{i,j-1})$, or using

$E(D_{i,j})$ including the effect of correlation, as seen in Exhibits 3 and 4, we must go through the calculation of new ultimates $E(L_i)$ since these ultimates depend on ρ. For the second iteration, base $D'_{i,j}$ on the new estimate of $E(L_i)$ and calculate new $D''_{i,j}$ using (33). Calculate a new sample correlation coefficient between $d''_{i,j-1}$ and $D''_{i,j}$. Repeat this process until it converges on a value for ρ. This iterative approach maximizes use of the most relevant data for the determination of ρ, by putting all available data points into the determination.

Other methods which would determine a different ρ for each j would require substantially more historical points than are usually available to an insurer. We have tried using accident quarter by calendar quarter triangles to expand the number of historical points, but the data are so variable when cut this fine that the approach did not work well.

The question of determining the elements of the accident year correlation matrix C presents similar challenges to the determination of ρ. There are additional complexities due to the need to avoid the effects of collinearity.

In practice, we have found it difficult to determine values for the non-diagonal elements of this matrix from company data. When we look at homogeneous lines of insurance, calculated correlation coefficients are often not significant if a reasonable standard is used. This is particularly true for stages of development where there are relatively few data points. An approach we have used with some success is similar in respects to that outlined above for estimating ρ.

Start with our incurred loss triangle and estimated ultimates $E(L_i)$. In this instance, the $E(L_i)$ include whatever correlation effects based on ρ the actuary has decided to include. Calculate implied $D'_{i,j} = E(L_i)/l_{i,j}$ for $j \leq n-i$. Transform all $D'_{i,j}$ to random variables

$D''_{i,j}$ with mean zero and standard deviation one using the formula:

$$D''_{i,j} = \frac{(D'_{i,j} - E(D'_{i,j}))}{(Var(D'_{i,j}))^{1/2}} \qquad (34)$$

Calculate the sample correlation coefficient between $D''_{i,j}$ and $D''_{i-1,j+1}$ for all values i, j where $j \le n-i$ and both $D''_{i,j}$ and $D''_{i-1,j+1}$ exist. This represents the correlation between age-to-ultimate development at a given valuation date for accident years which are separated by one year. Use this sample correlation coefficient for all c_{km} where $|k-m|=1$. Use an analogous approach for $|k-m|=2$, basing c_{km} on the sample correlation between $D''_{i,j}$ and $D''_{i-2,j+2}$. Continue in this fashion for $|k-m|=3,4,\ldots,$ until correlation is negligible or until there are too few points with the proper spacing to calculate a sample correlation coefficient.

The adjustment to this procedure to remove collinearity is to restate (34) as:

$$D''_{i,j} = \frac{(D'_{i,j} - E(D'_{i,j}|d_{i,j-1}))}{(Var(D'_{i,j}|d_{i,j-1}))^{1/2}} \qquad (35)$$

Thus, the correlation between $D_{i,j}$ and $d_{i,j-1}$ is considered and is "reduced out" of the measure of correlation between accident years. This way, any factor which contributes to correlation both between accident years and within accident years is not double-counted. In practice, we can estimate the correlation between accident years using both (34) and (35). If the results flowing from using (34) without eliminating collinearity and ignoring correlation within accident years show a stronger correlation than the combination of using (35) with correlation within accident year, we can use (34) and ignore the correlation within accident years. Otherwise, we use the combination of correlation within accident years and the correlation between accident years measured using the results of (35).

We have found that it is often helpful to restrict these calculations to relatively low values of j, for instance, look only at pairs $D''_{i,j}$ and $D''_{i-1,j+1}$ for which $j \le 3$. It appears that in many cases the correlation between accident years becomes insignificant for accident years which are beyond the earliest few stages of development at the valuation date under consideration.

Another approach which has proved useful is looking at higher levels of aggregation for determination of this correlation, rather than looking at a homogeneous line of insurance. Combinations of lines often show more correlation between accident years than can be seen through the "noise" at a finer level of detail. If correlations from aggregated data are to be applied at a finer level of detail, the actuary should make sure that the lines combined to reach the aggregate are expected to behave similarly in terms of loss development, so that the correlations might be reasonable for use at the detail level. A further consideration is that the collinearity adjustment described above must then be done at the aggregated level. While this adjustment is possible, the description of the calculation is not given here. In practice, when we have used aggregated data to determine correlation between accident years, we have not included correlation within accident years in our variance measure. Then we can rely on equation (34) in making our estimate of the correlation.

Exhibit 5 shows a situation where correlation was believed to be significant only between years falling in the last 4 accident years as of a given valuation date, and where correlation was negligible for $|k-m|>2$.

OTHER ISSUES AND ADDITIONAL COMPLICATIONS

The model as described above deals with the variability of reserve estimates including assumptions of two varieties of correlation. This section discusses additional concerns that arise in estimating reserve need and in some cases describes how the model could be adapted to address those concerns.

Homogeneity of Data

The model as presented depends on the data in the loss development triangle being homogeneous. If the data are not homogeneous, but the mix is constant through time, the model may still provide useful information. If the data are not homogeneous and the mix is not constant through time, the model as presented will not give representative results.

Correlation Between Lines of Insurance

The model as described above deals with one homogeneous line of insurance. When analyzing reserves for an insurer, we are usually concerned with the variability of the estimated reserve need for the insurer as a whole as well as on a line-by-line basis. If the R_s are independent, the formula for the variability of the total reserve estimate is

$$Var\left(\sum_{s=1}^{r} R_s\right) = \sum_{s=1}^{r} Var(R_s) \qquad (36)$$

where R_s is the reserve estimate for line s, and r is the number of lines of insurance.

Intuitively, it would be expected that some of the R_s are not independent. However, with measurement of the correlation within accident years and between accident years as described earlier in this paper, we have generally felt comfortable that the great majority of the effects which in practical application cause correlation between lines of insurance are already captured in the measures of correlation already described. Therefore, we have used (36) to estimate the variance of reserve estimates for combinations of lines of insurance. The substantial enhancements to this model to accommodate a further measure of correlation are beyond the scope of this paper.

Effect of Inflation

Variability in loss development could be the result of changes in inflation rates. If the actuary feels the effect of inflation on the loss development triangle would distort the measurement of reserve variability described in this paper, the triangle should be adjusted to a constant dollar basis before this variability model is used. It must be noted that in order for the resulting variability measure to be complete, consideration of the variability of estimates of future inflation would have to be included separately. Such consideration is beyond the scope of this paper. Failing to remove the effects of inflation from the triangle before applying our model of variability implicitly assumes that future variability of inflation will have the same effects on reserve estimates that the historical triangle shows. The actuary may feel this assumption is justified, but at any rate the choice of such an assumption should be a conscious decision.

Varying Volume of Data Through Time

In practice, the triangle of loss development data we are analyzing may have significant changes in the volume of business through the period of time covered. In such a situation, the calculation of sample $Var(d_{i,j})$ in the manner shown in Exhibit 2 could be distorted by points showing unusual development but backed by very little data. To cope with such situations, we have used a "dollar-weighted variance" approach which is shown in

Appendix A. The essence of this approach is defined by:

$$E_w(d_{i,j}) = \frac{\sum_{i=1}^{n-j} l_{i,j+1}}{\sum_{i=1}^{n-j} l_{i,j}} \quad \text{for } 1 \le j \le n-1 \text{ and} \tag{37}$$

$$Var_w(d_{i,j}) = \frac{\sum_{i=1}^{n-j} l_{i,j} \cdot \left(\frac{l_{i,j+1}}{l_{i,j}} - E_w(d_{i,j}) \right)^2}{\sum_{i=1}^{n-j} l_{i,j}} \quad \text{for } 1 \le j \le n-1 \tag{38}.$$

We use the weighted variance $Var_w(d_{i,j})$ in exactly the same way we used $Var(d_{i,j})$ in the earlier description of the model. Appendix A shows a revised version of Exhibits 1 through 5 (renumbered as 1A through 5A) substituting this weighted approach. In this particular example the effect is not large because the triangle used in our examples does not have extreme volume changes. In practice we have encountered many situations where the volume adjustment is important. In fact, we almost always use the weighted variance approach in practice since in situations where it is unimportant it has little effect, and in situations where it is important it gives a better representation of the variability of loss development.

Paid Versus Incurred Development

The model has been described in terms of incurred loss development and IBNR estimates. However, there is nothing in the formulation which requires that it be used in this way. All formulae and relationships would hold equally well for paid loss development analysis. Interpretation of correlation coefficients might vary, however. Depending on the reason for correlation between accident years on an incurred development basis, it might be expected that on a paid development basis there would be no correlation at all. Correlation between link ratios within a given accident year might also be zero when viewed in a paid loss development context. The actuary should consider these issues, and where possible test to see if correlation does exist. If correlation is eliminated by using paid development data, the actuary could rely on equations (8), (9) and (31) and greatly simplify the calculation of variability of reserve estimates.

Trends in Loss Development Patterns

As presented in the example in this paper, the model included no consideration of changes through time in loss development patterns. Some simple kinds of changes could be included relatively easily. For example, if a regression curve were fitted to historical values of $d_{i,j}$ and projected values were read off that curve, the appropriate adaptation of the model would be to substitute the curve values for all projected $E(d_{i,j})$ and substitute historical variance around the fitted curve for all $Var(d_{i,j})$ in the formulae describing the model. A volume-weighting scheme would be possible in this context if desired. The complications for using fitted curves in the analysis of correlation should also be considered. When normalizing the variables to arrive at $d''_{i,j-1}$ and $D''_{i,j}$ the expected values and variances should be measured considering the fitted curves rather than "raw" means and variances.

CONCLUSIONS

The model presented in this paper uses the variance of link ratios to estimate the variance of reserve estimates. It allows the inclusion of correlation effects of several varieties. It can be elaborated to cope with a number of concerns which may be important in specific situations.

While the estimation of correlation is often difficult in practice, we have presented some approaches which maximize the use of historical data in making such estimates. If estimates prove impractical, this model can be applied in a sensitivity-testing manner to demonstrate that the effects of correlation can be important as regards both the estimated reserve need and the variance of that estimate. Whether or not correlations can be estimated with much accuracy, this model gives actuaries an approach to better quantifying the uncertainty of reserve estimates.

BIBLIOGRAPHY

Roger M. Hayne, "An Estimate of Statistical Variation in Development Factor Methods", *Proceedings of the Casualty Actuarial Society,* Volume LXXII, 1985, p. 25.

C. K. Khury, "Loss Reserves: Performance Standards", *Proceedings of the Casualty Actuarial Society,* Volume LXVII, 1980, p. 1.

Lothar Sachs, Applied Statistics: A Handbook of Techniques, New York, Springer-Verlag, 1982.

Triangle of Incurred Loss Valuations

Exhibit 1A

Accident Year (i)	Stage of Development (Years) j: 1	2	3	4	5	6	7	8	9	10
1	32,223	48,439	54,284	58,146	61,305	63,739	63,604	62,721	63,247	62,159
2	42,588	65,239	77,329	82,064	85,260	85,226	80,944	79,577	80,614	
3	44,960	69,989	75,140	79,019	80,548	80,864	79,341	79,525		
4	33,145	56,088	60,732	66,551	66,857	68,395	66,806			
5	30,754	46,587	54,855	57,645	56,249	54,560				
6	33,594	47,576	52,870	59,598	58,715					
7	31,064	54,187	63,529	73,791						
8	33,831	48,453	62,742							
9	44,772	72,814								
10	48,307									

Weighted Average Development and Weighted Variance $l_{i,j} \cdot \left(\frac{l_{i,j+1}}{l_{i,j}} - E_w(d_{i,j}) \right)^2$

Exhibit 2A

Accident Year (i)	j: 1	2	3	4	5	6	7	8	9
1	97	38	13	91	64	34	1	0	0
2	29	87	50	48	5	53	5	0	
3	0	395	93	2	1	3	11		
4	597	244	5	7	16	0			
5	57	39	71	87	78				
6	676	67	87	52					
7	1078	30	355						
8	536	1035							
9	209								
$E_w(d_{i,j})$	1.558	1.149	1.087	1.015	1.007	0.975	0.991	1.011	0.983
$Var_w(d_{i,j})$	0.010029	0.004433	0.001536	0.000713	0.000472	0.000305	0.000075	0.000005	
$Var_w(d_{i,j})^{1/2}$	0.100145	0.066581	0.039193	0.026699	0.021719	0.0174634	0.008638	0.002306	

Selected Link Ratios

$E(d_{i,j})$ [illegible]

Modeling Correlation Within an Accident Year

Exhibit 3A

j

Stage:	1	2	3	4	5	6	7	8	9	10
$E(d_{i,j})$	1.558	1.180	1.120	1.015	1.007	0.975	0.991	1.011	0.983	1.000
$Var_w(d_{i,j})$	0.010029	0.004433	0.001536	0.000713	0.000472	0.000305	0.000075	0.000005	0.000000	0.000000
ρ	0.100									
a	0.110096	0.095952	0.107745	0.114186	0.092599	0.0520322	0.026238	0	0	0
b	0.889904	0.904048	0.892255	0.885814	0.907401	0.9479678	0.973762	1	1	1
$E(X_{i,j})$	1.265	1.090	0.964	0.960	0.955	0.985	0.994	0.983	1.000	1.000
$Var(X_{i,j})$	0.015197	0.004944	0.002218	0.001173	0.000486	0.000091	0.000005	0.000000	0.000000	0.000000
A_X	1.051	0.968	0.883	0.901	0.917	0.968	0.990	0.983	1.000	1.000
B_X	1.478	1.212	1.046	1.020	0.993	1.001	0.998	0.983	1.000	1.000
A_d	1.385	1.065	1.052	0.968	0.970	0.945	0.976	1.007	0.983	1.000
B_d	1.731	1.295	1.188	1.061	1.045	1.005	1.006	1.015	0.983	1.000
Unconditional, including correlation:										
$E(D_{i,j})$	2.022	1.297	1.099	0.981	0.967	0.960	0.984	0.994	0.983	1.000
$Var(D_{i,j})$	0.050964	0.012156	0.004081	0.001783	0.000929	0.000404	0.000083	0.000005	0.000000	0.000000
$E(D_{i,j}^2)$	4.139	1.694	1.211	0.964	0.935	0.921	0.969			
Conditional Expectation and Variance:										
$E(D_{i,j} \mid d_{i,j-1})$	2.022	1.305	1.110	0.985	0.963	0.956	0.985	0.994	0.983	1.000
$Var(D_{i,j} \mid d_{i,j-1})$	0.050964	0.012035	0.004041	0.001765	0.000920	0.000400	0.000082	0.000005	0.000000	0.000000
If no correlation:										
$E(D_{i,j})$	2.020	1.296	1.098	0.981	0.967	0.960	0.984	0.994	0.983	1.000
$Var(D_{i,j})$	0.041337	0.010046	0.003363	0.001501	0.000810	0.000370	0.000079	0.000005	0.000000	0.000000

Projected Ultimates and Standard Deviations
By Accident Year, Including Correlation

Exhibit 4A

Accident Year	Current Valuation	Expected LDF	Variance of LDF	Expected Ultimate	Needed IBNR	Variance of Ult	Std. Dev. of Ult.
1	62,159	1.000	0.000000	62,159	0	0	0
2	80,614	0.983	0.000000	79,227	−1,387	0	0
3	79,525	0.994	0.000005	79,040	−485	32163	179
4	66,806	0.985	0.000082	65,773	−1,033	364808	604
5	54,560	0.956	0.000400	52,166	−2,394	1191987	1,092
6	58,715	0.963	0.000920	56,560	−2,155	3172039	1,781
7	73,791	0.985	0.001765	72,713	−1,078	9613035	3,100
8	62,742	1.110	0.004041	69,632	6,890	15905997	3,988
9	72,814	1.305	0.012035	94,987	22,173	63806418	7,988
10	48,307	2.022	0.050964	97,671	49,364	118928464	10,905
Total	660,033			729,929	69,896		

Projected Reserve Need -- All Years Combined
Including Correlation Between Accident Years

Exhibit 5A

Matrix of correlation coefficients $\rho_{k,m}$

	Year	m 1	2	3	4	5	6	7	8	9	10
	1	1	0	0	0	0	0	0	0	0	0
	2	0	1	0	0	0	0	0	0	0	0
	3	0	0	1	0	0	0	0	0	0	0
	4	0	0	0	1	0	0	0	0	0	0
k	5	0	0	0	0	1	0	0	0	0	0
	6	0	0	0	0	0	1	0	0	0	0
	7	0	0	0	0	0	0	1	0.5	0.2	0
	8	0	0	0	0	0	0	0.5	1	0.5	0.2
	9	0	0	0	0	0	0	0.2	0.5	1	0.5
	10	0	0	0	0	0	0	0	0.2	0.5	1

Matrix of Covariances (= Correlation Coefficient X Std Dev for Year (Horiz) X Std Dev for Year (Vert))

$$\rho_{k,m} \cdot Var(L_k)^{1/2} \cdot Var(L_m)^{1/2}$$

	Year		m 1	2	3	4	5	6	7	8	9	10
		Std. Dev.	*0*	*0*	*179*	*604*	*1,092*	*1,781*	*3,100*	*3,988*	*7,988*	*10,905*
	1	*0*	0	0	0	0	0	0	0	0	0	0
	2	*0*	0	0	0	0	0	0	0	0	0	0
	3	*179*	0	0	32163	0	0	0	0	0	0	0
	4	*604*	0	0	0	364808	0	0	0	0	0	0
k	5	*1,092*	0	0	0	0	1191987	0	0	0	0	0
	6	*1,781*	0	0	0	0	0	3172039	0	0	0	0
	7	*3,100*	0	0	0	0	0	0	9613035	6182736	4953275	0
	8	*3,988*	0	0	0	0	0	0	6182736	15905997	15928784	8698680
	9	*7,988*	0	0	0	0	0	0	4953275	15928784	63806418	43555709
	10	*10,905*	0	0	0	0	0	0	0	8698680	43555709	118928464

Variance of Estimated Reserve Need: 371653280
Standard Deviation of Estimated Reserve Need: 19,278