AN ESTIMATE OF STATISTICAL VARIATION IN DEVELOPMENT FACTOR METHODS

ROGER M. HAYNE

Abstract

This paper explores some properties of the lognormal distribution. It is possible that these properties can provide information not only regarding the variability of age-to-age development factors but also regarding that of age-to-ultimate factors, if the actuary is willing to assume that these factors are lognormally distributed. Considered are problems of parameter estimation and uncertainty under the assumption of independence of the age-to-age factors. Some results are generalized by weakening the independence hypothesis, and a method of parameter estimation with missing observations is presented. This paper is intended as a starting point, indicating useful results if factors are assumed to be lognormally distributed. Still an open question is the specific situations where such is the case.

1. INTRODUCTION

Development factor techniques have long been in the casualty actuary's repertoire of projection methods and are used extensively in both ratemaking and in the estimation of loss reserves. There are, however, numerous sources of variability in such results. For this reason an actuary often applies several techniques to obtain several estimates of ultimate losses. The actuary then selects a "best estimate" which reflects his or her best judgment of the amount of those losses.

The complex interactions in the data and the influence of non-random events (such as changes in claims practices) add to the variability inherent in the results of any projection technique. This makes it difficult to assess whether random fluctuations alone can be responsible for a range of estimates provided by various techniques or whether the various methods detect different patterns actually present in the data, or whether some combination of the two exists.

This paper will not present a loss projection technique but rather will explore properties of the lognormal distribution which will allow for some estimation

of the statistical variability inherent in development factor projections if certain specific assumptions are satisfied. This can be useful in judging the differences among several projection techniques. For example, if wide fluctuations can be expected in projections (evidenced by wide confidence intervals), then variations in projections using different methods can be expected. If, on the other hand, there is little evidence of statistical variability and if the results of two methods are "far" apart, there is reason for further investigation to determine the cause of such differences.

Since the objective here is to study variability, the results depend on the underlying probability distributions and not on the particular age-to-age factors selected in practice. Thus, the resulting estimates of variability can be used as a measure of a "range of reasonableness" for various development factor selections and projections of other methods.

There are several useful properties of the lognormal probability distribution which motivated its choice as the statistical model here. First, the lognormal is defined for positive values of the random variable (development factors, except for extreme situations, are positive). Next, the distribution is skewed to the right but retains positive probabilities for large factors. This also has intuitive appeal for development factors which can be very large and experience anomolous swings. A third, and most useful, property of the lognormal distribution which suggested its consideration is its reproductive property. As will be stated in more precise form and greater generality below, the product of two lognormal random variables is, under certain assumptions, itself lognormal. In addition, the parameters of the product distribution can be determined easily from those of the two distributions. In terms of development factors this allows inferences regarding the age-to-ultimate factors to be made based on assumptions on the age-to-age factors.

The purpose of this paper is to explore the consequences of the assumption that the development factors are lognormally distributed. As stated above, the lognormal was chosen due to its useful properties and not on the basis of empirical data. Just as the normal distribution has been used to derive approximations in other areas of statistical work, it is possible that the lognormal model here may provide useful approximations in practice.

2. NOTATION AND BASIC CONCEPTS

This paper will deal with development factor techniques (see, for example, [1]). For this purpose, let $L_{i,j}$ denote data for exposure period i valued j valuation

periods from the start of the exposure period i. Exhibit 1 gives a few of the possible choices for each of the parameters. For simplicity, $L_{i,j}$ will generally be referred to here as incurred losses for accident year i valued j years from the beginning of year i. This is by no means an attempt to limit the results shown here.

Let D_j denote the factor to develop losses valued at j years into losses valued at $j + 1$ years (commonly called the age-to-age development factor). Thus, if the loss data strictly followed this model, the following relation would hold:

$$L_{i,j+1} = D_j L_{i,j} \tag{2.1}$$

Let D_j^* denote the factor to develop losses from age j years to their ultimate value (commonly called the age-to-ultimate development factor). From the above definition of D_j the following formula holds:

$$D_j^* = \prod_{k=j}^{\infty} D_k \tag{2.2}$$

Extensive use will be made here of the lognormal probability distribution which depends on two parameters, denoted here by μ and σ^2. As used in this paper the probability density function is defined as:

$$f(x) = \exp\{-[(\ln(x) - \mu)^2]/2\sigma^2\}/x\sigma\sqrt{2\pi} \tag{2.3}$$

Here $\ln(x)$ denotes the natural logarithm and $\exp(x)$ its inverse. This distribution has been used rather extensively in actuarial work especially in the modeling of size-of-loss distributions (see, for example, [2], [3] and [4]) and has many useful properties.

In particular, if the random variable X is lognormally distributed with parameters μ and σ^2, then the random variable $Y = \ln(X)$ is normally distributed with mean μ and variance σ^2. This fact, and reference to tables of normal probabilities, allows for easy calculation of confidence intervals for this distribution.

Probably of greatest use here is the following theorem (see p. 11 of [5]):

THEOREM 2.1

If $\{X_j\}$ is a sequence of independent variables, where X_j is lognormally distributed with parameters μ_j and σ_j^2, $\{b_j\}$ a sequence of constants and $c = \exp(a)$ a positive constant, then provided $\Sigma b_j\mu_j$ and $\Sigma\sigma_j^2 b_j^2$ both converge, the product

$$c \prod X_j^{b_j} \tag{2.4}$$

is lognormally distributed with parameters $a + \Sigma b_j \mu_j$ and $\Sigma b_j^2 \sigma_j^2$.

This theorem thus gives rise to the primary result used in this paper. In particular:

COROLLARY

If: (1) each age-to-age development factor D_j is lognormally distributed with parameters μ_j and σ_j^2 ($j = 1, 2, 3, \ldots$),

(2) all age-to-age development factors are independent, and

(3) $\sum_{j=1}^{\infty} \mu_j$ and $\sum_{j=1}^{\infty} \sigma_j^2$ both converge;

then each age-to-ultimate development factor D_j^* is lognormally distributed with parameters

$$\sum_{k=j}^{\infty} \mu_k \text{ and } \sum_{k=j}^{\infty} \sigma_k^2 \tag{2.5}$$

In most applications the third assumption above is fulfilled. Usually it is assumed that loss development stops after some finite point in time so that $\mu_j = \sigma_j^2 = 0$ for j sufficiently large.

3. SIMPLIFIED EXAMPLE

As an example of an application of these results assume that there is no development after four years (i.e. $\mu_j = \sigma_j^2 = 0$ for $j > 3$), that the age-to-age development factors D_1, D_2, and D_3 are each lognormally distributed with known hypothetical parameters given in the top half of Exhibit 2, and that all D_1, D_2, and D_3 are independent. Then the age-to-ultimate factors D_1^*, D_2^*, and D_3^* are all lognormally distributed with parameters as shown in the bottom part of Exhibit 2.

These parameters then allow for the calculation of various percentiles for the distributions of the age-to-age and age-to-ultimate development factors. To this end, let t denote the p ($0 < p < 1$) quantile of a standard normal random variable Z. That is, t satisfies the equation

$$P(Z < t) = p \tag{3.1}$$

Since D_j is assumed to be lognormally distributed with parameters μ_j and σ_j^2 the following formula holds:

$$P(D_j < \exp\{\mu_j + t\sigma_j\}) = p. \tag{3.2}$$

Formula (3.2) then allows the computation of percentiles for the age-to-age factors using μ_j, σ_j and various percentiles of the normal distribution. Some examples are presented in the top part of Exhibit 3. For example, $\exp(.175 + .674 \times \sqrt{.075}) = 1.433$, and so forth. Similar examples for the age-to-ultimate factors are shown on the bottom of that exhibit.

4. REFINEMENTS

Under the assumption of lognormality, this procedure provides a means to estimate the *statistical* uncertainty inherent in the development factor method. This method assumes the parameters μ_j and σ_j^2 are known. Yet to be addressed, however, is the uncertainty surrounding μ_j and σ_j^2. In actual practice μ_j and σ_j^2 are not known for certain. Most often the only source of knowledge lies in the historical development factors themselves.

Assume here that there are n_j accident years of incurred loss data available valued at year j and year $j + 1$ and that there are k such periods of development under consideration. Thus $L_{i,j}$ and $L_{i,j+1}$ are defined for $i = 1, 2, 3, \ldots, n_j$ and $j = 1, 2, 3, \ldots, k$. Assume further that the historical development factors at age j, defined by

$$d_{i,j} = L_{i,j+1}/L_{i,j} \tag{4.1}$$

form a random sample from a lognormal distribution of unknown parameters μ_j and σ_j^2. Moreover, define the statistics:

$$Y_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \ln(d_{i,j}) \tag{4.2}$$

$$S_j^2 = \frac{1}{n_j} \sum_{i=1}^{n_j} (\ln(d_{i,j}) - Y_j)^2 \tag{4.3}$$

It follows that Y_j and S_j^2 are the maximum likelihood estimators of μ_j and σ_j^2 respectively (see [5], p. 39) but, as in the normal analog, S_j^2 is biased. However, the statistic

$$V_j^2 = \frac{n_j}{n_j - 1} S_j^2 \tag{4.4}$$

is an unbiased and minimum variance estimator for σ_j^2, though it is no longer a maximum likelihood estimator (see [6], p. 165). Using these statistics, con-

fidence intervals for μ_j and σ_j^2 can be obtained. This follows from the fact that, under these assumptions of lognormality and independence, for each j, ln $(d_{i,j})$ form a sample from a normal distribution and thus

$$\frac{Y_j - \mu_j}{V_j/\sqrt{n_j}} \text{ has a } t \text{ distribution with } n_j - 1 \text{ degrees of freedom, and} \tag{4.5}$$

$$\frac{(n_j - 1)V_j^2}{\sigma_j^2} \text{ has a chi-squared distribution with } n_j - 1 \text{ degrees of freedom.} \tag{4.6}$$

Exhibit 4 shows a hypothetical development factor triangle which is generated using lognormally distributed random numbers. Since it is assumed that each column represents a random sample from a lognormal distribution, Y_j and V_j^2 provide estimators for μ_j and σ_j^2, respectively, for each value of j. In addition, the above observations regarding the distributions of Y_j and V_j^2 lead to the confidence intervals for μ_j and σ_j^2 given on the bottom of that exhibit. This information is helpful in estimating the degree of parameter uncertainty contained in the various age-to-age estimates. It cannot, however, be easily extended in general to the age-to-ultimate factors without additional assumptions, usually made about σ_j^2.

Since $\ln(d_{i,j})$ are normally distributed with mean μ_j and variance σ_j^2, the Y_j values are normally distributed with mean μ_j and variance σ_j^2/n_j. Moreover, any sum, such as $Y_1 + Y_2 + \ldots + Y_k$, is also normally distributed, in this case with mean $\mu_1 + \mu_2 + \ldots + \mu_k$ and variance $\sigma_1^2/n_1 + \sigma_2^2/n_2 + \ldots + \sigma_k^2/n_k$. If σ_1, σ_2, ..., σ_k are all known then the normal distribution can be used to obtain a confidence interval for $\mu_1 + \mu_2 + \ldots + \mu_k$ of the form

$$Y_1 + \ldots + Y_k \pm t\sqrt{\sigma_1^2/n_1 + \ldots + \sigma_k^2/n_k} \tag{4.7}$$

where t is the selected percentile for the standard normal distribution.

Normal distribution theory also provides results in the case where $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \ldots = \sigma_k^2 = \sigma^2$ but all are unknown. The obvious generalizations can be made in this case; however, it is quite unlikely that this would occur in development factor applications. The author is not aware of further statistics which do not need a restrictive assumption such as those on σ_1^2 through σ_k^2 above and which can be used to obtain estimates of parameter variability.

Exhibit 5 provides an example of an application of the first of these assumptions, that the values of σ_1^2, σ_2^2, ..., σ_k^2 are all known. Here they are

assumed equal to the corresponding estimates V_j^2. The second set of assumptions, that all the σ_j^2 are equal, is not applied to this data. The fact that the 90% confidence interval for σ_1^2 does not intersect any of the confidence intervals for the remaining σ_j^2 leaves the validity of this assumption open to serious question in this case.

5. RELAXATION OF INDEPENDENCE CONDITIONS

In the results presented this far, independence has been a necessary condition. The principal result in Theorem 2.1, however, is a special case of a more general theorem where the independence assumption can be replaced with one of multivariate lognormality. For this, some additional notation is necessary.

Denote by $\mathbf{R}^{m\times n}$ the set of matrices with m rows and n columns, having real entries. Following Aitchison and Brown (see [5], p. 11) a random variable $\vec{X} \in \mathbf{R}^{n\times 1}$ is said to have a multivariate lognormal distribution with parameters $\vec{\mu} \in \mathbf{R}^{n\times 1}$ and $\mathbf{\Sigma} \in \mathbf{R}^{n\times n}$ if the variable $\vec{Y} = \ln(\vec{X}) = (\ln(X_1), \ldots, \ln(X_n))'$ has a multivariate normal distribution with mean vector $\vec{\mu}$ and covariance matrix $\mathbf{\Sigma}$. It is assumed that $\mathbf{\Sigma}$ is symmetric (i.e. $\sigma_{i,j} = \sigma_{j,i}$) and positive definite, thus assuring that its inverse, $\mathbf{\Sigma}^{-1}$, exists. If A is a matrix, denote its transpose by A'. The following result then holds:

THEOREM 5.1

If the age-to-age development factors $\vec{D} = (D_1, \ldots, D_n)' \in \mathbf{R}^{n\times 1}$ have a multivariate lognormal distribution with parameters $\vec{\mu} = (\mu_1, \ldots, \mu_n)' \in \mathbf{R}^{n\times 1}$ and $\mathbf{\Sigma} = (\sigma_{i,j}) \in \mathbf{R}^{n\times n}$, symmetric and positive definite, then, each age-to-ultimate factor

$$D_j^* = \prod_{k=j}^{n} D_k, \quad j = 1, 2, \ldots, n \tag{5.1}$$

is lognormally distributed (with a single variate lognormal distribution) with parameters given by:

$$\sum_{k=j}^{n} \mu_k \text{ and } \sum_{i,k=j}^{n} \sigma_{i,k} \tag{5.2}$$

Proof: By definition $\vec{Y} = \ln(\vec{D})$ is normally distributed with parameters $\vec{\mu}$ and $\mathbf{\Sigma}$.

By a well known result in multivariate normal analysis (see [6], p. 383), the sum $Y_j + Y_{j+1} + \ldots + Y_n$ is normally distributed with mean and variance

given respectively by the parameters in (5.2). Since $Y_j + Y_{j+1} + \ldots + Y_n = \ln(D_j) + \ln(D_{j+1}) + \ldots + \ln(D_n) = \ln(D_j \times D_{j+1} \times \ldots \times D_n) = \ln(D_j^*)$, it follows that $\ln(D_j^*)$ is normally distributed and thus D_j^* is lognormally distributed. The parameters for that distribution are then given by the sums in (5.2). This completes the proof.

From multivariate normal theory ([6], p. 382), each of the above Y_j is normally distributed with mean μ_j and variance $\sigma_{j,j}$. Thus, each D_j has a lognormal distribution with parameters μ_j and $\sigma_{j,j}$. Hence, once $\vec{\mu}$ and $\boldsymbol{\Sigma}$ are known, confidence intervals for the D_j can be determined as in the case when independence is assumed. Similarly, confidence intervals for the D_j^* can be determined.

Parameter estimation, however, is not as simply generalized. The author is unaware of any method to estimate the parameters $\vec{\mu}$ and $\boldsymbol{\Sigma}$ in the general case based on the usual triangular form of historical development factors arrays.

If $d_{i,j}$ denotes the historical age-to-age factor for accident year i from age j to age $j + 1$ and the collection of such factors is based on $n + 1$ years of annual experience, ending in the current year, then $d_{i,j}$ is not defined if $i + j$ exceeds $n + 1$. Thus, the usual estimators for $\vec{\mu}$ and $\boldsymbol{\Sigma}$, which would require data for all allowable i,j values, cannot be applied. If it is assumed that the age-to-age factors are independent for $j \geq m$ for some m then the previously stated results apply to each column for which $j \geq m$.

If, now, $m \leq (n + 1)/2$ then the array of factors $d_{i,j}$, $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$; $i + j \leq n + 1$ will have at least m observations in each column. In this case the results of Bhargava [7] are applicable. In that paper, Bhargava derives maximum likelihood estimators for $\vec{\mu}$ and $\boldsymbol{\Sigma}$ for normal distributions based on samples with data missing in particular patterns. The available data from the first m columns of a development triangle form such an array if $m < (n + 1)/2$.

Following Bhargava, set

$$\mu_k = \nu_k + \sum_{j=1}^{k-1} \beta_j^{(k)} \mu_j \tag{5.3}$$

$$\sigma_{i,k} = \sigma_{k,i} = \sum_{j=1}^{k-1} \beta_j^{(k)} \sigma_{i,j};\ i = 1, 2, \ldots, k-1 \tag{5.4}$$

$$\sigma_{k,k} = \sigma_{k(0)}^2 + \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} \beta_i^{(k)} \beta_j^{(k)} \sigma_{i,j} \tag{5.5}$$

Given the parameters $\vec{\mu}$ and $\mathbf{\Sigma}$, the equations in (5.4) form a set of $k - 1$ linear equations in $k - 1$ unknowns, $\beta_1^{(k)}$, $\beta_2^{(k)}$, ..., $\beta_{k-1}^{(k)}$. Once these values are determined, ν_k and $\sigma_{k(\theta)}^2$ can be found from (5.3) and (5.5), respectively. Conversely, given ν_k, $\vec{\beta}^{(k)}$ and $\sigma_{k(\theta)}^2$, these equations give the parameters $\vec{\mu}$ and $\mathbf{\Sigma}$. Bhargava then determined maximum likelihood estimates for the parameters ν_k, $\vec{\beta}^{(k)}$, and $\sigma_{k(\theta)}^2$, and, using (5.3), (5.4) and (5.5), derived maximum likelihood estimates for $\vec{\mu}$ and $\mathbf{\Sigma}$.

To state those results some further notation is necessary. Suppose that $\vec{y}_1$, $\vec{y}_2$, $\vec{y}_3$, ..., $\vec{y}_n$ are n independent observations from a population that has a multivariate normal distribution with m variates with $m \leq (n + 1)/2$. The sample will satisfy Bhargava's definition of a monotone sample if observations for the i^{th} variate are available in $\vec{y}_1$, $\vec{y}_2$, ..., $\vec{y}_{n+1-i}$. Since $m \leq (n + 1)/2$, there will be at least m complete vectors. Note, this merely formalizes the situation that exists in a development factor matrix showing annual development for $n + 1$ accident years if the vector y_j is thought of as the first m elements of the j^{th} row. Though independence of the various age-to-age factors is no longer assumed, independence of the rows (accident year observations) is.

Given this sample, define the matrix $\mathbf{y}_{(1,\,k-1)}$ as the matrix composed of a column of 1's, followed by the first $k - 1$ elements of the observations $\vec{y}_1$, $\vec{y}_2$, ..., $\vec{y}_{n+k-1}$. This is a matrix with a column of 1's followed by the first $k - 1$ columns of the largest matrix containing observations for all of the first k variates in the data triangle. Let $\vec{y}_{(k)}$ denote the column matrix composed of the $n + k - 1$ observations of the k^{th} variate.

With this notation, Bhargava presents the following result:

THEOREM 5.2

Assume that $\vec{\mu} \in \mathbf{R}^{m\times 1}$, $\mathbf{\Sigma} \in \mathbf{R}^{m\times m}$ is symmetric and positive definite, and that $\vec{y}_1$, $\vec{y}_2$, ..., $\vec{y}_n$ is an independent, monotone sample from the multivariate normal population with mean vector $\vec{\mu}$ and covariance matrix $\mathbf{\Sigma}$. If ν_k, $\vec{\beta}^{(k)}$ and $\sigma_{k(\theta)}^2$ are defined as in (5.3), (5.4) and (5.5) then the maximum likelihood estimators for ν_k, $\vec{\beta}^{(k)}$ and $\sigma_{k(\theta)}^2$ are given by:

$$(\hat{\nu}_k, \hat{\vec{\beta}}^{(k)})' = (\mathbf{y}'_{(1,\,k-1)}\mathbf{y}_{(1,\,k-1)})^{-1}\mathbf{y}'_{(1,\,k-1)}\vec{y}_{(k)} \tag{5.6}$$

$$\begin{aligned}(n + 1 - k)\hat{\sigma}_{k(\theta)}^2 &= \sum_{j=1}^{n+1-k} (y_{j,\,k} - \hat{\nu}_k - \sum_{i=1}^{k-1} \hat{\beta}_i^{(k)} Y_{j,\,i})^2 \\ &= \vec{y}'_{(k)}(\mathbf{I} - \vec{y}_{(k)}(\mathbf{y}'_{(1,\,k-1)}\mathbf{y}_{(1,\,k-1)})^{-1}\vec{y}_{(k)})\vec{y}_{(k)}\end{aligned} \tag{5.7}$$

Though not immediately obvious, the value of $\hat{\nu}_k$ is the constant coefficient of the least squares multiple linear regression of $\vec{y}_{(k)}$ against the first $k - 1$ variates, based on the observations in the first $n + 1 - k$ rows of the matrix. Similarly $\hat{\vec{\beta}}^{(k)}$ are the coefficients of each of the first $k - 1$ variates. Finally, $\hat{\sigma}^2_{k(\theta)}$ is the conditional variance of the fit. It denotes the amount of variance which remains unexplained by the regression. Thus, estimation of the ν_k, $\vec{\beta}^{(k)}$ and $\sigma^2_{k(\theta)}$ can be accomplished using multiple regression for $k = 2, 3, \ldots, m$ while $\hat{\nu}_1$ and $\hat{\sigma}^2_{1(\theta)}$ are the sample mean and variance of the first column of the matrix.

If, now, it is assumed that the first m columns of the development factor matrix have a multivariate lognormal distribution with parameters $\vec{\mu} \in \mathbf{R}^{m \times 1}$ and $\mathbf{\Sigma} \in \mathbf{R}^{m \times m}$, symmetric and positive definite, then the above procedures, applied to $y_{i,j} = \ln(d_{i,j})$, will produce the maximum likelihood estimates for ν_k, $\vec{\beta}^{(k)}$ and $\sigma^2_{k(\theta)}$ and thus $\vec{\mu}$ and $\mathbf{\Sigma}$. This result follows since, under this assumption, the values of $\ln(d_{i,j})$ form a monotone sample from a multivariate normal distribution.

As an example of these methods, Exhibit 6 shows the estimators $\hat{\nu}_k$, $\hat{\vec{\beta}}^{(k)}$ and $\hat{\sigma}^2_{k(\theta)}$ along with estimators for $\vec{\mu}$ and $\mathbf{\Sigma}$ based on the hypothetical development factors shown in Exhibit 4. In this case, the matrix is 6×6 ($n = 6$). Here it is assumed that $m = 3$, that there is no development after the sixth year, that is, $D_6 = D_7 = D_8 = \ldots = 1$, and that D_4 through D_6 are all independent and independent of the first three factors. Finally, it is assumed that D_1 through D_3 have a multivariate lognormal distribution with parameters $\vec{\mu} \in \mathbf{R}^{3 \times 1}$ and $\mathbf{\Sigma} \in \mathbf{R}^{3 \times 3}$, symmetric and positive definite.

If it is assumed that the parameters of the distributions for D_1 through D_6 are equal to their maximum likelihood estimates then Exhibit 7 shows the resulting confidence intervals for the resulting age-to-age and age-to-ultimate factors. The intervals for D_1 through D_3 are based on the fact that the natural logarithm of each is normal with mean μ_k and variance $\sigma_{k,k}$. This exhibit also compares the intervals with those derived under the assumption of independence, assuming that the parameters equal the values of Y_i and V_i^2 in Exhibit 4.

Correlation among the age-to-age development factors will, of course, impact the marginal variance of any given factor and also the variance of the resulting age-to-ultimate factors. If the various age-to-age factors are positively correlated then the resulting age-to-ultimate factors will have wider variation (and hence wider confidence intervals) when derived using the multivariate estimation than those derived using the assumption of independence. Conversely, if there is negative correlation among the age-to-age factors then the

resulting estimates of the age-to-ultimate factors derived using the multivariate techniques will probably have less variation than those derived assuming independence. This follows from the variance formula given in (5.2).

Parameter uncertainty is not as straightforward as in the completely independent case. Though the author does not know the distributions of the various estimates, Bhargava does provide likelihood ratio tests to test the hypothesis H_0: $\vec{\mu} = 0$ against H: $\vec{\mu} \in \mathbf{R}^{m \times 1}$. Those results are sufficiently complex, however, that they will not be presented here. One interesting result mentioned by Bhargava, however, is that the distribution of $(n + 1 - k)\ \hat{\sigma}^2_{k(\theta)}/\sigma^2_{k(\theta)}$, given the observations in the first $n + 1 - k$ rows and k columns, has a chi-squared distribution with $n + 1 - 2k$ degrees of freedom.

6. OBSERVATIONS

The usefulness of any theory lies in the nearness of the hypothesis of that theory to reality. In this regard, the first question that comes to mind is that of the lognormality of development factors in actual practice. The lognormal distribution has the benefit of being defined for only positive values of the random variable and does not impose an upper bound on those values. This corresponds to development factors which are generally positive and are unbounded. In practice, statistical tests such as the Kolmogorov-Smirnov Test as presented by Gary Patrik ([8], p. 65) may help in assessing the validity of the assumption of lognormality.

The independence of the various columns may also be able to be tested. Since $\ln(d_{i,j})$ are assumed to be normally distributed for $i = 1, 2, \ldots, n_j$ a test based on the sample correlation coefficient between the natural logarithms of two columns may give some insight as to the validity of this assumption. In addition, these results require the independence of the development factors of a given age from each other. Again, usual statistical tests, applied to the natural logarithms of the development factors, may be helpful in assessing the validity of this hypothesis. In any case, in actual applications, actuarial judgment is required to detect any patterns which may appear in the data (for example, correlation between columns, trend in age-to-age factors over time, etc.). Such patterns often add to the variation apparent in the factors. Actuarial judgment will thus decrease statistical variability.

In order to compare the results of various loss projection methods, the age-to-ultimate development factors must be multiplied by the appropriate loss amount to date. To draw statistical conclusions about the resulting loss projec-

tions, the age-to-ultimate factors must then be assumed to be independent from the amounts recorded to date.

The methods presented here can only provide estimates of *statistical* variability under very explicit assumptions. They should be looked on as providing a "range of reasonableness" of loss projections, based on such variability, rather than as a confidence interval about any specific ultimate loss estimate. In the latter case, the actuary's judgment is used to narrow a large range of possible choices (as presented by the historical development factors) in light of his or her knowledge of the underlying data.

7. CONCLUSIONS AND BEGINNINGS

This paper is presented more as an opening to further investigation than as a definitive solution to a problem. The model selected for study, that of development factor projection, is one of the simplest of the projection techniques in use by casualty actuaries and any actuary with experience in applying this technique knows its limitations and weaknesses. Hopefully the results presented here help in assessing the variability inherent in this method.

The larger challenge still facing casualty actuaries is to devise estimates of the amount of variation to be expected in the more complex projection methods used. However, a precise estimate of variability inherent in an actuary's "best estimate" probably is not possible. Actuarial judgment used to interpret diverse results of various methods, in light of the actuary's knowledge of events that may impact the patterns to be expected in the data, cannot be statistically quantified. This judgment is usually the most important aspect of the estimation of ultimate losses, but any further insight that can be gained from these techniques can be helpful in forming that judgment.

REFERENCES

[1] D. Skurnick, "A Survey of Loss Reserving Methods," *PCAS* LX, 1973, p. 16.

[2] D. R. Bickerstaff, "Automobile Collision Deductibles and Repair Cost Groups: The Lognormal Model," *PCAS* LIX, 1972, p. 68.

[3] C. Hewitt, "Credibility for Severity," *PCAS* LVII, 1970, p. 148.

[4] R. J. Finger, "Estimating Pure Premiums by Layer—An Approach," *PCAS* LXIII, 1976, p. 34.

[5] J. Aitchison and J. A. C. Brown, *The Lognormal Distribution*, Cambridge University Press, 1969.

[6] R. V. Hogg and A. T. Craig, *Introduction to Mathematical Statistics*, Third Edition, Macmillan Publishing Co., New York, 1970.

[7] R. Bhargava, "Multivariate Tests of Hypothesis with Incomplete Data," Technical Report No. 3, Applied Mathematics and Statistics Laboratories, Stanford University, California, August 1962.

[8] G. Patrik, "Estimating Casualty Insurance Loss Amount Distributions," *PCAS* LXVII, 1980, p. 57.

EXHIBIT 1

SOME CHOICES AS TO DATA ARRANGEMENT FOR DEVELOPMENT FACTOR TECHNIQUES

Type of Data (L)	Aggregation Type
Paid Losses	Report Period
Incurred Losses	Accident Period
Paid (Closed) Claim Counts	Policy Period
Reported Claim Counts	

Exposure Period (i)	Valuation Period (j)
Year	Year
Half-year	Half-year
Quarter	Quarter

EXHIBIT 2

SIMPLIFIED EXAMPLE DEVELOPMENT FACTORS

Parameters for the Age-to-Age Factors
D_1, D_2, and D_3

j	μ_j	σ_j^2
1	0.175	0.075
2	0.045	0.005
3	0.005	0.001

Parameters for the Age-to-Ultimate Factors
D_1^*, D_2^*, and D_3^*

j	μ_j^*	σ_j^{*2}
1	0.225	0.081
2	0.050	0.006
3	0.005	0.001

EXHIBIT 3

EXAMPLE PERCENTILES BASED ON SIMPLIFIED DEVELOPMENT FACTOR DATA

	Percentile				
Age	10% ($t = -1.282$)	25% ($t = -0.674$)	50% ($t = 0.000$)	75% ($t = 0.674$)	90% ($t = 1.282$)
	Percentiles for Age-to-Age Development Factors				
1	0.839	0.990	1.191	1.433	1.692
2	0.955	0.997	1.046	1.097	1.145
3	0.965	0.984	1.005	1.027	1.047
	Percentiles for Age-to-Ultimate Development Factors				
1	0.869	1.034	1.252	1.517	1.804
2	0.952	0.998	1.051	1.108	1.161
3	0.965	0.984	1.005	1.027	1.047

EXHIBIT 4

HYPOTHETICAL DEVELOPMENT FACTORS

Accident Year	Stage of Development (years) 2/1	3/2	4/3	5/4	6/5	7/6
1	1.932	1.036	1.009	1.003	1.002	1.000
2	1.975	1.038	1.013	1.006	1.001	
3	1.809	1.041	1.011	1.005		
4	1.954	1.043	1.009			
5	1.997	1.035				
6	1.932					
Estimators:						
Y	6.59×10^{-1}	3.79×10^{-2}	1.04×10^{-2}	4.66×10^{-3}	1.50×10^{-3}	
V^2	1.21×10^{-3}	1.05×10^{-5}	3.59×10^{-6}	2.31×10^{-6}	4.99×10^{-7}	
90% Confidence Intervals for:						
μ	6.30×10^{-1} to 6.88×10^{-1}	3.48×10^{-2} to 4.10×10^{-2}	8.20×10^{-3} to 1.26×10^{-2}	2.10×10^{-3} to 7.22×10^{-3}	-1.65×10^{-3} to 4.65×10^{-3}	
σ^2	5.45×10^{-4} to 5.26×10^{-3}	4.41×10^{-6} to 5.89×10^{-5}	1.38×10^{-6} to 3.06×10^{-5}	7.71×10^{-7} to 4.49×10^{-5}	1.30×10^{-7} to 1.25×10^{-4}	

EXHIBIT 5

EXAMPLE CONFIDENCE INTERVALS FOR THE PARAMETERS OF THE AGE-TO-ULTIMATE DEVELOPMENT FACTORS

Assumptions:

$$\sigma_1^2 = 1.21 \times 10^{-3} \qquad \sigma_2^2 = 1.05 \times 10^{-5}$$
$$\sigma_3^2 = 3.59 \times 10^{-6} \qquad \sigma_4^2 = 2.31 \times 10^{-6}$$
$$\sigma_5^2 = 4.99 \times 10^{-7}$$
$$\sigma_j^2 = \mu_j = 0 \text{ for } j \geq 6$$

Age	Estimator for μ_j^*	90% Confidence Interval for for μ_j^*
1	0.714	0.690 to 0.737
2	0.0554	0.0511 to 0.0578
3	0.0166	0.0143 to 0.0189
4	0.00616	0.00450 to 0.00782
5	0.00150	0.000678 to 0.00232

EXHIBIT 6

EXAMPLE PARAMETER ESTIMATES USING MULTIVARIATE SAMPLE ESTIMATION

Variable:	D_1	D_2	D_3
Estimators:			
$\hat{\nu}_k$	6.59×10^{1}	6.31×10^{2}	1.53×10^{-2}
$\hat{\beta}^{(k)}$		-3.84×10^{2}	-1.00×10^{3} -1.09×10^{-1}
$\hat{\sigma}^2_{k(\theta)}$	1.01×10^{3}	6.60×10^{6}	2.61×10^{6}
$\hat{\mu}$	6.59×10^{1}	3.79×10^{-2}	1.05×10^{-2}
$\hat{\Sigma}$	1.01×10^{3} -3.86×10^{-5} 3.20×10^{6}	-3.86×10^{5} 8.08×10^{6} -8.42×10^{7}	3.20×10^{6} -8.42×10^{7} 2.70×10^{6}

EXHIBIT 7

EXAMPLE 90% CONFIDENCE INTERVALS BASED ON MULTIVARIATE PARAMETER ESTIMATION

Intervals For Age-to-Age Factors:

D_1	D_2	D_3	D_4	D_5
Assuming Independence:				
1.825	1.033	1.007	1.002	1.000
to	to	to	to	to
2.046	1.044	1.014	1.007	1.003
Using Multivariate Estimators:				
1.834	1.034	1.008	1.002	1.000
to	to	to	to	to
2.036	1.044	1.013	1.007	1.003

Intervals for Age-to-Ultimate Factors:

D_1^*	D_2^*	D_3^*	D_4^*	D_5^*
Assuming Independence:				
1.926	1.049	1.013	1.003	1.000
to	to	to	to	to
2.161	1.063	1.021	1.009	1.003
Using Multivariate Estimators:				
1.904	1.050	1.013	1.003	1.000
to	to	to	to	to
2.147	1.062	1.021	1.009	1.003