The Application of Cumulative Distribution Functions in the Stochastic Chain Ladder Model

by Son T. Tu, Ph.D., ACAS, MAAA

THE APPLICATION OF CUMULATIVE DISTRIBUTION FUNCTIONS IN THE STOCHASTIC CHAIN LADDER MODEL

SON T. TU
SCRUGGS CONSULTING
ARGYLE, TEXAS

ABSTRACT

A new stochastic model based on the traditional chain ladder is introduced. It makes explicit use of cumulative distribution functions and payment patterns. It incorporates a mathematical rationale for non-stochastic variations in the age-to-age factors. Perturbation methods are used to obtain and justify the solution. Estimation of liabilities in the tail is a natural product of the model. All stochastic variables are assumed to be normally distributed, and the assumption is then confirmed with the chi square goodness-of-fit test. Extensive numerical solutions of an actual problem are given. Several new avenues of related research are suggested.

KEYWORDS

Chain ladder; Loss reserving; Cumulative distribution functions; Tail factor; Stochastic models; Perturbation theory.

Problem #2: This problem is one of tail development. One common solution is to consider the incurred and paid amounts for a number of accident years preceding the earliest year in the triangle and to assume that the incurred-to-paid ratio for these years constitutes an accurate reflection of the tail development. But development for some long-tailed lines such as Workers' Compensation can continue over thirty or more years. Data may not be available or credible over such a long period of time.

A second common technique to account for tail development is Bondy development. However, it is difficult to estimate the errors associated with this technique. Furthermore, it does not have a very firm mathematical or practical foundation.

Our model gives a mathematical rationale for non-stochastic variations in the ata factors. It estimates tail factors, and, whenever warranted by the data, allows these factors to vary by accident years. It also emphasizes the tendency of the traditional approach to rely on unnecessarily limited information for the estimation of critical quantities. As a case in point, take the 108:120 factor of 1.017 in Table 2. This is in fact only a realization of a random variable. But the traditional approach would carry it over to all accident years. In contrast, based on all the information available in the triangle, our model estimates the mean and standard deviation at each individual point.

In section 2, we relate how cumulative distribution functions with time-varying parameters can lead to non-stochastic variations. In section 3, we introduce the scaling and proportionality functions, and motivate and present the stochastic model. To obtain the solution, we use perturbation methods, in section 4, to develop an iterative regression procedure. The mathematical formulas for the estimates of the ultimate liability amounts and the standard deviations are derived in section 5; we also show that the losses are in theory normally distributed, to the leading order.

In section 6, numerical solutions to the Table 1 triangle are given. We use the chi square test, in section 7, to show that the numerical results are indeed normally distributed. In section 8, we discuss the different orders of error magnitudes apparent in any loss reserving exercise. In section 9, we compare the effects of using different scaling and proportionality functions. We define and discuss parametrization in section 10. We discuss some related topics in the concluding section.

2. CUMULATIVE DISTRIBUTION FUNCTIONS

Let the cumulative distribution function $F(t)$ represent the payment pattern for an accident year, with t in years. Figure 1 illustrates:

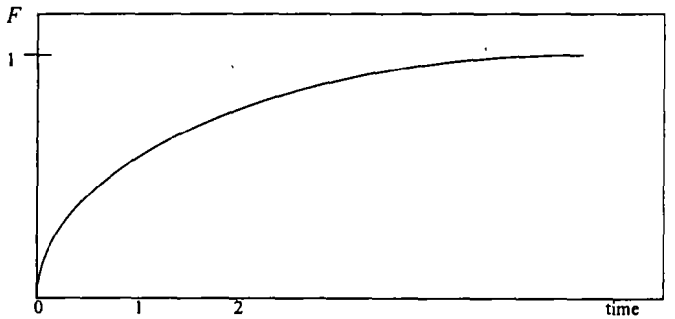

Figure 1: The cumulative distribution function $F(t)$.

At $t = 0$, $F = 0$ when no payments have been made. At $t = \infty$, $F = 1$ when all payments have been made. At any intervening time, the distribution gives the fraction that has been paid. Any ata factor

1. INTRODUCTION

In this paper, we introduce an innovation into the traditional chain ladder by making explicit use of cumulative distribution functions. We show how that can broaden the scope and usefulness of the chain ladder.

We begin with an example in Table 1, which contains the paid losses and allocated loss adjustment expenses (in millions of dollars) for Workers' Compensation from the 1992 Best's Aggregates and Averages. Best's data include most of the net insurance volume in the United States. The accident years are from 1982 to 1991.

AY	1	2	3	4	5	6	7	8	9	10
1	2409	4825	6173	7003	7578	7990	8295	8551	8747	8894
2	2602	5429	7010	8037	8738	9230	9589	9866	10060	
3	3105	6495	8617	9978	10886	11503	11959	12296		
4	3316	7241	9634	11202	12217	12915	13439			
5	3416	7701	10439	12216	13330	14105				
6	3831	8934	12063	14023	15266					
7	4527	10547	14309	16588						
8	4934	11881	16070							
9	5300	12901								
10	5488									

Table 1: Workers' Compensation paid loss triangle.

Table 2 contains the corresponding age-to-age (ata) factors, also called the linked ratios:

AY	1	2	3	4	5	6	7	8	9
1	2.003	1.279	1.134	1.082	1.054	1.038	1.031	1.023	1.017
2	2.086	1.291	1.147	1.087	1.056	1.039	1.029	1.020	
3	2.092	1.327	1.158	1.091	1.057	1.040	1.028		
4	2.184	1.330	1.163	1.091	1.057	1.041			
5	2.254	1.356	1.170	1.091	1.058				
6	2.332	1.350	1.162	1.087					
7	2.330	1.357	1.159						
8	2.408	1.353							
9	2.434								
slope	0.055	0.011	0.004	0.001	0.001	0.001	-0.001	-0.003	

Table 2: The actual age-to-age factors.

We discuss in general two problems evident in Table 2, and how the integration of cumulative distribution functions into the reserving scheme alleviates them.

Problem #1: The factors in the first few columns exhibit marked variations, as indicated by the regression slopes in the last row. The fact that most of the changes in the factors in the first few columns are increases raises the possibility that at least some of the variations are non-stochastic. We mention two common techniques to deal with this problem.

The first technique is to consider only the factors in the last few available years. For instance, the 12:24 estimate may be the average of the last four factors in the first column. The assumption is that the recent past is more reflective of the future than the distant past. While the assumption is generally sound, there are a few deficiencies with the approach. First, not all data, and information, are being used. Secondly, averaging precludes the continuation of a trend, and this preclusion may not be desirable.

The second common technique to resolve the trending problem is similar to the methods in BERQUIST and SHERMAN (1977). This approach is the regression fitting of the factors to a line or curve, from which the missing factors can be estimated. One difficulty of this approach is that the last several columns have few factors, so any single one, which after all is just a realization of a random variable, has an inordinately large amount of influence on the process. Secondly, fitting to a line compels the continuation of a trend. It cannot address a situation where the factors in a column attain either a minimum or maximum. As will be shown, that is a common occurrence for this set of data. Fitting to a curve on the other hand requires much information and therefore can result in high variability.

can be represented in terms of the distribution values. For example, the factor between the development periods j and $j+1$ is:

$$(2.1) \qquad r_j = F(j+1)/F(j).$$

We work with one function in particular, the transformed log-normal:

$$(2.2) \qquad F(t) = F(t;\mu,\sigma,\tau) = \Phi\left[\mathrm{sgn}(\ln t)\left|\ln t\right|^{\tau};\mu,\sigma\right]; \quad t > 0,$$

where Φ is the normal distribution with mean μ and standard deviation σ. The parameter τ is the exponent of the function. Our selection of this function is dictated by the fact that, of all the functions that we tested, it best describes the WC payment patterns. We have more to say on this in section 6.

To simulate the non-stochastic variations in the ata factors, we allow the three function parameters to vary by accident year. Let k denote an accident year and (μ_k,σ_k,τ_k) denote the corresponding function parameters. We assume that these parameters can be expressed in the following polynomial forms:

$$(2.3) \qquad \begin{aligned} \mu_k &= \mu + (k-1)\alpha_1 + (k-1)^2\alpha_2 + (k-1)^3\alpha_3 + \mathrm{K} \\ \sigma_k &= \sigma + (k-1)\beta_1 + (k-1)^2\beta_2 + (k-1)^3\beta_3 + \mathrm{K} \\ \tau_k &= \tau + (k-1)\chi_1 + (k-1)^2\chi_2 + (k-1)^3\chi_3 + \mathrm{K} \end{aligned}$$

The right hand side (RHS) of (2.3) has the following interpretations: (μ,σ,τ) are the base parameters, $(\alpha_1,\beta_1,\chi_1)$ are the linear annual changes, $(\alpha_2,\beta_2,\chi_2)$ are the quadratic annual changes, and so on. We also refer to the first set as the base coefficients, the second the linear coefficients, the third the quadratic coefficients, and so forth. All coefficients are assumed independent. For any given problem, only some are statistically significant. A major part of the analysis is the determination of all those.

If all annual changes are statistically insignificant, the parameters and the ata factors do not vary by accident years. In that case, we retrieve somewhat the traditional chain ladder, but the methodology to estimate the factors differs from the traditional approach.

If any of the annual changes is statistically significant, the parameters and the ata factors vary by accident years, and thus exhibit non-stochastic variations. In such case, each accident year in effect has its own payment pattern, different from those of any other year.

There are many reasons for non-stochastic variations in the parameters. A major one is that the insurance operations are changing. Another is that the environmental climate in which the insurance operates is changing. Clearly, there can be many other factors. Equations (2.3) are simply our attempt to measure the extent to which all these factors affect the payment patterns. It is important to realize that (2.3) does not compel the accident years to have different payment patterns. It simply allows that possibility. If it turns out that the accident years have a statistically similar payment pattern, then the annual changes should be statistically insignificant.

We define the partial derivatives as follows:

$$(2.4) \qquad G = \frac{\partial F}{\partial \mu}, \quad H = \frac{\partial F}{\partial \sigma}, \quad K = \frac{\partial F}{\partial \tau}.$$

3. THE GOVERNING EQUATION

Let y_{kj} be the cumulative paid loss amount for accident year k at development year j. Then the *loss* factor is:

$$(3.1) \qquad q_{kj} = \frac{y_{k,j+1}}{y_{kj}},$$

and the *development* factor is:

$$(3.2) \qquad r_{kj} = \frac{F(j+1;\mu_k,\sigma_k,\tau_k)}{F(j;\mu_k,\sigma_k,\tau_k)}.$$

We define the *governing* equation as:

$$(3.3) \qquad q_{kj} = r_{kj} + \ln(r_{kj})\varepsilon_{kj}, \qquad \varepsilon_{kj} \sim N(0,s^2).$$

The ε_{kj}'s are assumed independent and normally distributed with mean zero and standard deviation s. We call ε_{kj} and $\ln(r_{kj})\varepsilon_{kj}$ the error and the error term, respectively. Equation (3.3) simply says that successive payments should obey the payment pattern, with some stochastic error. As its name implies, a governing equation contains the basic assumptions and governs the behavior of the model. Given it, everything else should logically follow.

We note that the development factors are in essence a proxy for the payment pattern. If all actual and estimated ata factors closely match, then we infer that the distribution accurately depicts the real payment pattern. If there is a consistent mismatch in some of the factors, then we cannot make that inference.

Undoubtedly, the most unusual feature of (3.3) is the scaling function $\ln(r_{kj})$. It is needed because the magnitudes of the error terms change drastically throughout the development. The magnitude is large at the beginning of development, it is small near full development, it is zero at full development, and it goes through the whole continuum in between. In section 8, we discuss this subject in detail with numerical examples.

We mention two mathematical anomalies which the scaling function prevents. First, suppose that we are very far in the tail where all development has definitively ceased. Therefore, $q_{kj} = 1$. If our model is any good, it would also predict $r_{kj} = 1$. Hence the error term must be zero, and the presence of the scaling function ensures that equality.

Secondly, suppose we want to compute the variance of the ultimate loss amount. As will be shown in section 5, that includes the sum of an infinite series, each term of which corresponds to the error term in (3.3). If the scaling function were absent, the infinite series and the variance would have no finite limits. But if it were present, the terms in the series would approach zero asymptotically, and the series would have a finite limit.

In section 7, we show that, without a scaling function, the model cannot satisfy the normality assumption. In section 8, we demonstrate that, in such case, the error terms are not properly scaled.

We will primarily work with (3.3), but the general form of the governing equation is:

$$(3.4) \qquad q_{kj} = r_{kj} + w(y_{kj})b(r_{kj})\varepsilon_{kj}.$$

ε_{ij} is the error, $b(r_{ij})$ is the scaling function, $w(y_{ij})$ is the proportionality function, and the product of all three, $w(y_{ij})b(r_{ij})\varepsilon_{ij}$, is the error term. The scaling function must satisfy the following conditions:

(3.5) $\quad b(r) > 0, \text{ for } r > 1 \text{; and } b(1) = 0.$

We call $w(y_{ij})$ the proportionality function, because it dictates the loss amount proportionality of the error term. If we multiply (3.4) by y_{ij}, then:

(3.6) $\quad y_{k,j+1} = y_{ij}\left[r_{ij} + w(y_{ij})b(r_{ij})\varepsilon_{ij}\right].$

For the particular form of (3.3), $w(y_{ij}) = 1$, and the error term is proportional to y_{ij}. We therefore call that the linear proportionality function. Similarly, when $w(y_{ij}) = 1/y_{ij}^{1/2}$ and $\ln(y_{ij})/y_{ij}$, the proportionality functions are square root and logarithmic, respectively.

A priori, we do not have any reference to prefer one set of scaling and proportionality functions over another. In section 9, we test a number of them and compare their numerical results, with the deviations and chi square values as the measuring sticks. Our conclusion is that the most appropriate model has the logarithmic scaling function and the linear proportionality function, as in (3.3).

MURPHY (1995) presented three models which can be written as follows:

(3.7) $\quad y_{k,j+1} = y_{ij}r + \varepsilon_{ij}$, Least Squares Multiplicative (LSM);

$y_{k,j+1} = y_{ij}(r + \varepsilon_{ij})$, Simple Average Development (SAD); and

$y_{k,j+1} = y_{ij}r + y_{ij}^{1/2}\varepsilon_{ij}$, Weighted Average Development (WAD).

There are two major differences between (3.6) and (3.7). Murphy's models do not have a scaling function. And they assume the development factor r to be constant in any given development period, whereas we allow r_{ij} to vary within a development period. The forms of (3.7) have different proportionality functions. With our terminology, SAD takes the linear function, and WAD takes the square root function. For LSM, $w(y_{ij}) = 1/y_{ij}$.

For their chain-ladder stochastic models, many authors (VERRALL, 1990; ZEHNWIRTH, 1990) have assumed that the loss amounts are log-normally distributed. STANARD (1985) and HALLIWELL (1996) have shown that such models have inherent upward bias. MACK (1995) argued that they suffer higher variability. Our model bypasses these difficulties, because (3.3) implies that the loss quantities are normally distributed, as will be shown in section 5.

The normal distribution for the loss amounts has two additional advantages. First, if the liability for an accident year is normally distributed, the sum for all accident years is also normally distributed., and the variance of the sum can be calculated. Secondly, suppose we have another model which also gives normally distributed estimates, the combination of estimates from the two models is normally distributed.

The governing equation (3.3) is to be used in two ways, matching and estimation:

a) Matching. There are forty-five points (ata factors) in Table 2. We apply (3.3) to every point. From this matching, we obtain estimates for the n variables so as to minimize the sum of squares of errors.

b) Estimation. For the particular case of (3.3), (3.6) becomes:

(3.8) $\quad y_{k,j+1} = y_{ij}\left[r_{ij} + \ln(r_{ij})\varepsilon_{ij}\right].$

Equation (3.8) gives the estimate at the next period based on the actual or estimated value at the previous period. If y_{kj} is an actual amount, we assume that there is no error associated with it; actually, this assumption is a direct consequence of the governing equation itself. And all the variance of the estimate $y_{k,j+1}$ comes from the parameter error in r_{kj} and the process error in ε_{kj}. If y_{kj} is an estimated amount, then its variance also contributes to that of the next estimate. Our convention is that the RHS of (3.6) and (3.8) should take the actual y_{kj} whenever available.

MACK (1995) made an important distinction. In many models (VERRALL, ZEHNWIRTH), y_{kj} in (3.8) is the expected value; whereas, in the traditional chain ladder, it is the actual value. In this paper, the latter is the case.

To recapitulate, our entire model consists of equations (2.2), (2.3), (3.1), (3.2) and (3.3). For a given set of data, we have to find all the statistically significant coefficients of (2.3) such that the sum of squares of errors in (3.3) are minimized, given that the payment patterns are specified by the cumulative distribution function in (2.2). The governing equation of (3.3) deserves its name because it has the central role of linking together all the different elements of the system.

4. THE ITERATIVE REGRESSION PROCESS

As described in the previous section, the system is a highly non-linear one; therefore it is impossible to obtain the solution in closed form or in one step. Instead, we apply the methods of perturbation theory to derive an iterative regression process, the application of which systematically leads to the solution.

To minimize the algebra, all derivations in this section are for the model in which only the base coefficients (μ, σ, τ) are variables. In the general model, we have to solve the regression system for n variables.

We begin by perturbing every variable:

$$\mu \to \mu + \Delta\mu, \quad \sigma \to \sigma + \Delta\sigma, \quad \tau \to \tau + \Delta\tau \,. \tag{4.1}$$

We may think of a perturbation as the replacement of a value (μ, for instance) by the sum of that value and an infinitesimal increment ($\Delta\mu$). The value is a known quantity, and the increment is an unknown quantity to be found. The reason why a perturbation is helpful is that, since the increment is assumed infinitesimal, we may retain only the linear terms in the Taylor's series expansions. Instead of a non-linear system, we in effect solve a series of linear systems. The successive solutions of the linear systems lead us closer and closer to the solution of the non-linear system.

First, we supply a guess (μ, σ, τ). Based on that guess, the regression process gives us the incremental $(\Delta\mu, \Delta\sigma, \Delta\tau)$. The sum of the guess and the increment provides the next guess. We keep up the iteration process until it converges to the solution.

Using the definition of the derivatives in (2.4), the perturbation of F has the following form:

$$\begin{aligned} F(j;\mu+\Delta\mu,\sigma+\Delta\sigma,\tau+\Delta\tau) = F(j;\mu,\sigma,\tau) + \\ G(j;\mu,\sigma,\tau)\Delta\mu + H(j;\mu,\sigma,\tau)\Delta\sigma + K(j;\mu,\sigma,\tau)\Delta\tau \,. \end{aligned} \tag{4.2}$$

In a more general case where, for instance, α_1 and α_2 are also variables, the RHS of (4.2) would include the terms $(k-1)G\Delta\alpha_1$ and $(k-1)^2 G\Delta\alpha_2$.

Using (4.2) in (3.2), we have the following for the perturbation of the development factor:

$$(4.3)\qquad r_{kj}(\mu+\Delta\mu,\sigma+\Delta\sigma,\tau+\Delta\tau)=r_{kj}(\mu,\sigma,\tau)\left[1+\left(\frac{G_{k,j+1}}{F_{k,j+1}}-\frac{G_{kj}}{F_{kj}}\right)\Delta\mu+\left(\frac{H_{k,j+1}}{F_{k,j+1}}-\frac{H_{kj}}{F_{kj}}\right)\Delta\sigma+\left(\frac{K_{k,j+1}}{F_{k,j+1}}-\frac{K_{kj}}{F_{kj}}\right)\Delta\tau\right].$$

where $F_{kj}=F(j;\mu_k,\sigma_k,\tau_k)$, and so on. We give the full derivation of (4.3) in Appendix A. If we substitute (4.3) into (3.3), the result is:

$$(4.4)\ r_{kj}\left(\frac{G_{k,j+1}}{F_{k,j+1}}-\frac{G_{kj}}{F_{kj}}\right)\Delta\mu+r_{kj}\left(\frac{H_{k,j+1}}{F_{k,j+1}}-\frac{H_{kj}}{F_{kj}}\right)\Delta\sigma+r_{kj}\left(\frac{K_{k,j+1}}{F_{k,j+1}}-\frac{K_{kj}}{F_{kj}}\right)\Delta\tau+\ln(r_{kj})\varepsilon_{kj}=q_{kj}-r_{kj}\ .$$

After division by the scaling function, (4.4) yields exactly a regression system for the increment variables $(\Delta\mu,\Delta\sigma,\Delta\tau)$. In matrix form, we have:

$$(4.5)\qquad M\Delta+\varepsilon=b,$$
$$\Delta=\left(\Delta\mu,\Delta\sigma,\Delta\tau\right)^T,\quad \varepsilon=\left\{\varepsilon_{kj}\right\},\quad b=\left\{(q_{kj}-r_{kj})/\ln(r_{kj})\right\}.$$

M is a $45\,x\,3$ coefficient matrix. This matrix changes after every iteration, since after each one, we have a new set of coefficients. For example, the row of M for accident year k and development period j has the following elements:

$$(4.6)\qquad M_{kj}^1=\frac{r_{kj}}{\ln(r_{kj})}\left[\frac{G_{k,j+1}}{F_{k,j+1}}-\frac{G_{kj}}{F_{kj}}\right],$$
$$M_{kj}^2=\frac{r_{kj}}{\ln(r_{kj})}\left[\frac{H_{k,j+1}}{F_{k,j+1}}-\frac{H_{kj}}{F_{kj}}\right],$$
$$M_{kj}^3=\frac{r_{kj}}{\ln(r_{kj})}\left[\frac{K_{k,j+1}}{F_{k,j+1}}-\frac{K_{kj}}{F_{kj}}\right].$$

In the more general case of n variables, M would be a $45\,x\,n$ matrix, and Δ an n-dimensional vector.

The solution of (4.5) so as to minimize the sum of squares of errors is well known in multiple regression analysis. It is,

$$(4.7)\qquad \Delta=\left(M^TM\right)^{-1}M^Tb\,.$$

The sum of (4.7) and the current guess constitutes the succeeding guess. When the process is stable and leads toward the solution, the sum of squares of errors of the succeeding guess is always smaller than that of the current guess. Therefore, if we continue the iteration until the guesses no longer vary, the resultant solution is guaranteed to have the smallest possible sum. We can see from (4.7) that the solution must satisfy:

$$(4.8) \qquad \left(M^T M\right)^{-1} M^T b = 0.$$

This actually may only be a local solution. Globally, the possibility of multiple solutions cannot be discounted. In practice, however, we have never encountered multiple solutions.

The vector in (4.7) indicates the general direction in which the solution lies. When the initial guess is very far from the solution, if we take full steps as indicated by (4.7), the guesses may quickly become unstable. In such cases, we should take steps that are fractions of the full steps. The steps should be sufficiently small until the iterative process enters some stable mode, then the step size may be increased. We have even encountered situations in which, by taking full steps, the guesses spiral stably toward the solution, but very slowly. In such cases, the full steps overstep the solution, and the convergence can be accelerated by taking smaller steps.

Finally, there exists the possibility of no convergence at all. This may be the result of either of two scenarios. First, the distribution being used may not be stable in the iterative regression scheme. For instance, we find any Pareto-type distribution to be highly unstable. Secondly, the distribution may not be the right one for the loss data being considered.

When the distribution is the transformed log-normal, the convergence is quite fast, and the initial guess need not be close to the solution.

The estimate for the variance of errors is:

$$(4.9) \qquad s^2 = \frac{1}{45-n}\sum_{k=1}^{9}\sum_{j=1}^{10-k}\varepsilon_{kj}^2, \quad \varepsilon_{kj} = \frac{q_{kj} - r_{kj}}{\ln(r_{kj})}.$$

The denominator in (4.9) is the number of degrees of freedom: 45 is the number of data points, and n is the number of variables. From (4.1), we have the following relationship for the variances of the coefficients:

$$(4.10) \qquad Var(\mu) = Var(\Delta\mu)\ ,$$

and so on. From standard regression analysis, we obtain the parameter variance matrix as:

$$(4.11) \qquad Var(\mathrm{P}) = \begin{bmatrix} Var(\mu) & Cov(\mu,\sigma) & Cov(\mu,\tau) \\ Cov(\mu,\sigma) & Var(\sigma) & Cov(\sigma,\tau) \\ Cov(\mu,\tau) & Cov(\sigma,\tau) & Var(\tau) \end{bmatrix} = s^2\left(M^T M\right)^{-1}.$$

5. THE ULTIMATE ESTIMATES

In this section, we assume that the iterative regression process has found all the coefficients and we have to obtain the estimates of the ultimate loss amounts and their variances. In particular, consider the k-th accident year, which has $y_{k,11-k}$ as the last actual cumulative paid amount. Using (3.1) and (3.3), the estimate for the loss amount at the next period is:

$$(5.1) \qquad y_{k,11-k+1} = y_{k,11-k}q_{k,11-k} = y_{k,11-k}\left[r_{k,11-k} + \ln(r_{k,11-k})\varepsilon_{k,11-k}\right].$$

After another iteration, the estimate for the succeeding period is:

(5.2) $$y_{k,11-k+2} = y_{k,11-k}q_{k,11-k}q_{k,11-k+1} = y_{k,11-k}\left[r_{k,11-k} + \ln(r_{k,11-k})\varepsilon_{k,11-k}\right]\left[r_{k,11-k+1} + \ln(r_{k,11-k+1})\varepsilon_{k,11-k+1}\right].$$

After repeated iterations, the estimate for the ultimate amount can be expressed as:

(5.3) $$y_k = y_{k,11-k}\prod_{j=0}^{\infty} q_{k,11-k+j} = y_{k,11-k}\prod_{j=0}^{\infty}\left[r_{k,11-k+j} + \ln(r_{k,11-k+j})\varepsilon_{k,11-k+j}\right].$$

In (5.3), we have an infinite variety of error terms. There are the linear error terms, containing ε_{kj}. There are the quadratic error terms, containing $\varepsilon_{kj}\varepsilon_{kl}$; and so on. A linear term is proportional to s, a quadratic term to s^2. Since s is generally small, the linear terms dominate in absolute value over the other error terms. We are thus justified in retaining only the linear terms, and (5.3) becomes:

(5.4) $$y_k = \frac{y_{k,11-k}}{F_{k,11-k}}\left[1 + \sum_{j=0}^{\infty}\frac{\ln(r_{k,11-k+j})}{r_{k,11-k+j}}\varepsilon_{k,11-k+j}\right].$$

Equation (5.4) is correct to the leading order. The ultimate loss is normally distributed, since it is the sum of normally distributed quantities. Taking the expected value of (5.4), we have:

(5.5) $$E\{y_k\} = y_{k,11-k}/F_{k,11-k}.$$

$F_{k,11-k} = F(11-k;\mu_k,\sigma_k,\tau_k)$ is the percent paid to date, and its reciprocal is the age-to-ultimate factor. Equation (5.5) says that the expected ultimate amount is the product of the paid-to-date amount and the age-to-ultimate factor, as we would expect. In the rest of this section, for the sake of brevity, we write y_k to denote the expected value of the same quantity.

To obtain the variance from (5.4), we use the following formula. Let $W = XY$ be the product of two independent stochastic quantities, then

(5.6) $$Var(W) = \overline{X}^2\left[Var(Y)\right] + \overline{Y}^2\left[Var(X)\right] + \left[Var(X)\right]\left[Var(Y)\right],$$

where the bars denote expected values. If we apply (5.6) to (5.4), then we have:

(5.7) $$Var(y_k) = y_{k,11-k}^2\left[Var\left(\frac{1}{F_{k,11-k}}\right) + \frac{s^2}{F_{k,11-k}^2}\sum_{j=0}^{\infty}\left(\frac{\ln(r_{k,11-k+j})}{r_{k,11-k+j}}\right)^2 + s^2 Var\left(\frac{1}{F_{k,11-k}}\right)\sum_{j=0}^{\infty}\left(\frac{\ln(r_{k,11-k+j})}{r_{k,11-k+j}}\right)^2\right].$$

In the derivation of (5.7), we assume that there is no error associated with the actual $y_{k,11-k}$. The variance in (5.7) is the sum of three terms. The first is the parameter error, which is just the variance of the age-to-ultimate factor.

The second term is the process error. It is the sum of an infinite series, because, at each development period, an additional amount of error contributes to the total, and theoretically there are an

infinite number of periods. In Appendix B, we prove that the series possesses a finite limit. We can also see the pivotal role of the scaling function: without it, the series would have no finite limit.

The third term is the product of the parameter and process errors. Every variance is proportional to s^2; every product of variances is proportional to s^4, and hence negligible. In the following derivations and calculations, we ignore those terms altogether. To the leading order, the variance in (5.7) is therefore the sum of the parameter and process errors.

To estimate the parameter error in (5.7), again we resort to perturbation:

$$\frac{1}{F_{k,11-k}(\mu+\Delta\mu,\sigma+\Delta\sigma,\tau+\Delta\tau)} = \frac{1}{F_{k,11-k}(\mu,\sigma,\tau)}\left[1-\frac{G_{k,11-k}}{F_{k,11-k}}\Delta\mu - \frac{H_{k,11-k}}{F_{k,11-k}}\Delta\sigma - \frac{K_{k,11-k}}{F_{k,11-k}}\Delta\tau\right]. \tag{5.8}$$

The derivation of (5.8) is in Appendix A.

Taking the variance of (5.8), we have:

$$Var\left[\frac{1}{F_{k,11-k}}\right] = \frac{1}{F_{k,11-k}^2}\left\{\left[\frac{G_{k,11-k}}{F_{k,11-k}}\right]^2 Var(\mu) + 2\frac{G_{k,11-k}H_{k,11-k}}{F_{k,11-k}^2}Cov(\mu,\sigma)+\Lambda\right\}, \tag{5.9}$$

where the dots represent the other variances and covariances. Finally, after collecting terms, (5.7) becomes:

$$Var(y_k') = y_k^2\left\{\left[\frac{G_{k,11-k}}{F_{k,11-k}}\right]^2 Var(\mu) + 2\frac{G_{k,11-k}H_{k,11-k}}{F_{k,11-k}^2}Cov(\mu,\sigma)+\Lambda + s^2\sum_{j=0}^{\infty}\left[\frac{\ln(r_{k,11-k+j})}{r_{k,11-k+j}}\right]^2\right\}. \tag{5.10}$$

We define the sum total of all the ultimate losses as:

$$y_T = \sum_{k=1}^{10} y_k\ . \tag{5.11}$$

We note that y_T is normally distributed. It can be shown that the variance is:

$$Var(y_T) = Var(\mu)\left[\sum_{k=1}^{10} y_k\frac{G_{k,11-k}}{F_{k,11-k}}\right]^2 + 2Cov(\mu,\sigma)\left[\sum_{k=1}^{10} y_k\frac{G_{k,11-k}}{F_{k,11-k}}\right]\left[\sum_{k=1}^{10} y_k\frac{H_{k,11-k}}{F_{k,11-k}}\right] + \Lambda + s^2\sum_{k=1}^{10} y_k^2\sum_{j=0}^{\infty}\left[\frac{\ln(r_{k,11-k+j})}{r_{k,11-k+j}}\right]^2. \tag{5.12}$$

Comparing (5.10) to (5.12), we observe that:

$$\sum_{k=1}^{10}\left[y_k\frac{G_{k,11-k}}{F_{k,11-k}}\right]^2 \neq \left[\sum_{k=1}^{10} y_k\frac{G_{k,11-k}}{F_{k,11-k}}\right]^2, \tag{5.13}$$

and so on. Therefore, we conclude that:

$$(5.14) \qquad \sum_{k=1}^{10} Var(y_k) \neq Var(y_T).$$

In words, the sum of the variances does not equal the variance of the sum. If the ultimate losses have negative correlation, the former is greater. If they have positive correlation, the latter is greater. We expect the second scenario, because, whatever the realization of a random variable, it most likely affects the ultimate losses in the same direction. In the next section, we show with a numerical example that such is the case. To numerically evaluate each series, we sum up the first fifty terms.

6. NUMERICAL RESULTS

In the previous two sections, we present the mathematical formulas, for the most part assuming only the base coefficients are statistically significant. In this section, we present the numerical solution to the Table 1 triangle. For this problem, four of the annual changes are significant; therefore, the reader will have to modify the formulas in the previous sections to obtain the numerical solutions in this one. The statistically significant coefficients for the Table 1 triangle are:

	μ	σ	τ	β_1	χ_1	β_2	χ_2
estimate	.7582	1.0838	.8988	-.0459	.0450	.0028	-.0057
s.d.	.0051	.0085	.01114	.0047	.0079	.0005	.0011

Table 3: Estimates and standard deviations of the coefficients.

In the parlance of section 10, the solution has the correct parametrization. Our criterion for statistical significance is that an estimate must be at least twice as large in absolute value as its standard deviation. This criterion translates into: if the true value of a variable were indeed zero, we have a 4.6% probability of accepting it as a non-zero variable.

With the values in Table 3, the equations of (2.3) simplify to:

$$(6.1) \qquad \mu_k = \mu, \quad \sigma_k = \sigma + (k-1)\beta_1 + (k-1)^2\beta_2,$$
$$\tau_k = \tau + (k-1)\chi_1 + (k-1)^2\chi_2, \quad 1 \leq k \leq 10.$$

The process with which we obtain (6.1) is as follows. We begin with the model in which all coefficients up to and including the cubic ones are variables. In such model, we have twelve coefficients to estimate. We apply the iterative regression process to obtain the solution. If there are at least two statistically insignificant coefficients in the solution, we eliminate the most obviously insignificant one. We continue the process until all remaining coefficients are statistically significant.

In going from the estimation of twelve variables to that of seven variables, we have to examine six permutations of the model. Each permutation has a unique set of variables to be estimated. Given the assumptions that the base coefficients are always significant, which may not be true for the mean, and that all fourth- and higher-power coefficients are always insignificant, there are $2^9 = 512$ permutations, for a distribution of three parameters. In a format like EXCEL, which we use for this paper, we have to construct a separate spreadsheet for each permutation. We have fortunately systematized the process, so that a complete conversion from one permutation to another takes only a few minutes. We construct permutations as needed; we do not construct all at the same time. In some languages such as APL, which we have used in the past, one set of computer code suffices for all possible permutations, including different sizes of the data. Despite this obvious advantage of APL, we highly recommend EXCEL, given the choice between the two mediums. A programming error, especially a subtle one that does not result in

an unreasonable solution, is much more likely to escape detection in APL. And it is much easier to build additional features into an EXCEL spreadsheet than into an APL code.

From (6.1), we estimate the mean to be a constant, and both the deviation and the exponent to be quadratic curves. Substituting the values in (6.1) into the formula for the transformed log-normal of (2.2), we obtain the values of the c.d.f.'s:

AY	1	2	3	4	5	6	7	8	9	10
1	.2421	.4857	.6196	.7047	.7629	.8048	.8352	.8604	.8795	.8949
2	.2331	.4812	.6259	.7180	.7802	.8243	.8567	.8812	.9001	.9151
3	.2249	.4776	.6315	.7293	.7947	.8402	.8730	.8972	.9157	.9299
4	.2475	.4751	.6363	.7384	.8058	.8521	.8848	.9087	.9265	.9401
5	.2111	.4739	.6402	.7450	.8135	.8599	.8925	.9159	.9332	.9462
6	.2059	.4740	.6429	.7487	.8174	.8637	.8960	.9190	.9360	.9487
7	.2019	.4756	.6445	.7495	.8174	.8633	.8953	.9182	.9350	.9477
8	.1994	.4787	.6447	.7472	.8135	.8586	.8902	.9131	.9301	.9429
9	.1983	.4834	.6437	.7418	.8057	.8494	.8805	.9032	.9207	.9340
10	.1986	.4896	.6412	.7332	.7935	.8354	.8658	.8885	.9060	.9198

Table 4: The values of the cumulative distribution functions.

Let's give an example of how one value in Table 4 is calculated. From (6.1), we obtain the parameters for the fifth accident year as (.7582,.9446,.9872). From (2.2), we get that $F(4;.7582,.9446,.9872)=.7450$.

Table 4 says that, for the first accident year, 24.2% of the ultimate amount has been paid after one year, and 88.5% after 10 years. We then calculate the tail factor as the reciprocal of 88.5%. We note that the tail factors vary by accident years.

To obtain the estimated ata factors, we compute the quotients of successive values in Table 4.

AY	1	2	3	4	5	6	7	8	9
1	2.006	1.276	1.137	1.083	1.055	1.039	1.029	1.022	1.018
2	2.064	1.301	1.147	1.087	1.057	1.039	1.029	1.021	1.017
3	2.124	1.322	1.155	1.090	1.057	1.039	1.028	1.021	1.016
4	2.185	1.339	1.160	1.091	1.057	1.038	1.027	1.020	1.015
5	2.245	1.351	1.164	1.092	1.057	1.038	1.026	1.019	1.014
5	2.245	1.351	1.164	1.092	1.057	1.038	1.026	1.018	1.014
6	2.303	1.356	1.165	1.092	1.057	1.037	1.026	1.018	1.014
7	2.355	1.355	1.163	1.091	1.056	1.037	1.026	1.018	1.014
8	2.401	1.347	1.159	1.089	1.055	1.037	1.026	1.019	1.014
9	2.438	1.331	1.152	1.086	1.054	1.037	1.026	1.019	1.014
10	2.465	1.310	1.143	1.082	1.053	1.036	1.026	1.020	1.015
slope	0.055	0.011	0.004	0.002	0.001	0.000	-0.001	-0.001	

Table 5: The estimated age-to-age factors.

The match between Tables 2 and 5 is generally quite close. The slopes in the two tables match almost exactly. To make them directly comparable, those in the latter are calculated using only factors above the diagonal. The close match of the factors implies that the transformed log-normal adequately describes the payment patterns of the data.

We also consider using one of the following three distributions as the c.d.f.: the transformed normal, the transformed log-gamma, and the transformed gamma. For every of these functions, either a solution cannot be found, or there is a consistent mismatch of the ata factors somewhere in the triangle. We therefore believe that none of the three functions describes well the payment patterns of Workers' Compensation.

The table below presents the estimated cumulative paid amounts:

AY	1	2	3	4	5	6	7	8	9	10
1	2409	4833	6155	7021	7581	7994	8302	8535	8741	8900
2	2602	5370	7062	8042	8734	9232	9593	9863	10078	10227
3	3105	6595	8588	9952	10872	11509	11952	12292	12548	12744
4	3316	7245	9697	11180	12224	12918	13412	13802	14072	14278
5	3416	7669	10403	12148	13339	14091	14639	15023	15307	15520
6	3831	8821	12117	14048	15309	16131	16733	17164	17480	17717
7	4527	10662	14292	16641	18093	19107	19814	20321	20693	20973
8	4923	11846	16001	18624	20279	21401	22188	22759	23182	23503
9	5300	12921	17178	19796	21501	11668	23499	24110	24570	24925
10	5488	13528	17716	20257	21925	23082	23921	24549	25032	25412

Table 6: The estimated paid loss amounts.

Generally, we have quite good agreement between Tables 1 and 6. Let's give an example of how one value in the latter table is calculated. From Table 5, we have that $r_{54} = 1.092$. Therefore, the estimate for $y_{55} = 1.092 * 12216 = 13339$.

The estimates for the ultimate paid amounts and the corresponding standard deviations are:

AY	Ultimate	S.D.
1	9939	45
2	11176	40
3	13704	51
4	15188	62
5	16403	73
6	18676	91
7	22132	124
8	24925	166
9	26688	288
10	27629	583
total	186459	980

Table 7: Ultimate estimates.

Comparison of Tables 6 and 7 indicates that a considerable amount of liabilities lies in the tail. The standard deviation of the total is computed using (5.12). Under the assumption of mutual independence of ultimate estimates, the deviation of the total would only be 700. Since the variance of the sum is considerably greater than the sum of the variances, we infer that the ultimate estimates have a high degree of positive correlation.

7. THE CHI SQUARE TEST

In this section, we apply the chi square goodness-of-fit test to demonstrate the normality of the results. We also show that the model without the scaling function does not satisfy the normality assumption.

We define the *normalized error* as the quotient of the error and its standard deviation. From (3.3), we have:

$$(7.1) \qquad e_{kj} = \frac{\varepsilon_{kj}}{s} = \frac{q_{kj} - r_{kj}}{s\ln(r_{kj})}.$$

The normalized errors should follow the standard normal distribution. To test if that indeed is the case, we divide the real line into five intervals: $(-\infty,-.842)$, $(-.842,-.253)$, $(-.253,.253)$, $(.253,.842)$, and $(.842,\infty)$. We note that, if a random variable is normally distributed, each interval should contain 20% of the observations. We define the following two quantities: U_i as the number of observed e_{kj}'s in interval i, and V_i as the number of expected e_{kj}'s. Then the quantity χ^2 is defined as:

$$(7.2) \qquad \chi^2 = \sum_{i=1}^{5} \frac{(U_i - V_i)^2}{V_i}.$$

It is well known that χ^2 follow the chi square distribution with four degrees of freedom.

With $s = .0298$, we obtain the following normalized errors from Tables 2 and 5:

AY	1	2	3	4	5	6	7	8	9
1	-0.160	0.502	-0.740	-0.198	-0.366	-0.723	2.252	1.073	-1.349
2	1.044	-1.219	-0.158	0.199	-0.131	-0.336	0.388	-2.872	
3	-1.434	0.538	0.702	0.556	-0.345	0.547	0.431		
4	-0.056	-1.005	0.515	-0.253	-0.144	1.873			
5	0.383	0.525	1.446	-0.286	0.621				
6	1.183	-0.668	-0.451	-1.184					
7	-0.997	0.182	-0.817						
8	0.269	0.656							
9	-0.145								

Table 8: The normalized errors.

In Table 8, if there are either large positive or negative values grouped in at least one column, then the distribution does not fit well the payment patterns. We do not detect such a scenario in Table 8. The expected and observed frequencies are:

Interval	1	2	3	4	5
V_i	9	9	9	9	9
U_i	7	9	10	13	6

Table 9: The error frequencies.

The values in Table 9 give $\chi^2 = 3.33$. If the normalized errors come from the standard normal distribution, then there is a 50% probability that the chi square distribution with four degrees of freedom exceed 3.33. The normality assumption for the governing equation of (3.3) is therefore accepted.

If we use the model without a scaling function, many of the normalized errors, 26 to be exact, are bunched together in the middle interval, and $\chi^2 = 43.56$. The probability of the chi square distribution exceeding that value is nil. We therefore reject the normality assumption. In the next section, we discuss the reason why the model fails the test in that case.

8. ORDERS OF ERROR MAGNITUDES

In this section, we consider the differing orders of error magnitudes, and how the proper recognition of them is inextricably linked to the scaling function. We also indicate the reason that the model without the scaling function does not pass the normality test. We begin by considering the variances of the two points at opposite ends of the triangle, q_{19} and q_{91}.

From (3.3), the quantity q_{19}, which has the realized value of 1.0168, has the following formula:

$$(8.1) \qquad q_{19} = r_{19} + \ln(r_{19})\varepsilon_{19}.$$

Taking the expected value of (8.1), we get:

$$(8.2) \qquad E\{q_{19}\} = E\{r_{19}\} = F_{1,10}/F_{19} = 1.0175.$$

We obtain from (8.1) the variance as:

$$(8.3) \qquad Var(q_{19}) = Var(r_{19}) + s^2\left[\ln(r_{19})\right]^2.$$

The process error has the value:

$$(8.4) \qquad s^2\left[\ln(r_{19})\right]^2 = \left[.0298 * \ln(1.0175)\right]^2 = 2.7 * 10^{-7}.$$

To obtain the parameter error, we use the perturbation form for an ata factor of (4.3):

$$(8.5) \quad r_{19}(\mu + \Delta\mu, \sigma + \Delta\sigma, \tau + \Delta\tau) = r_{19}\left[1 + \left(\frac{G_{1,10}}{F_{1,10}} - \frac{G_{19}}{F_{19}}\right)\Delta\mu + \left(\frac{H_{1,10}}{F_{1,10}} - \frac{H_{19}}{F_{19}}\right)\Delta\sigma + \left(\frac{K_{1,10}}{F_{1,10}} - \frac{K_{19}}{F_{19}}\right)\Delta\tau\right].$$

Taking the variance of (8.5), we have:

$$(8.6) \qquad Var(r_{19}) = r_{19}^2\left[\left(\frac{G_{1,10}}{F_{1,10}} - \frac{G_{19}}{F_{19}}\right)^2 Var(\mu) + \Lambda\right] = 3.2 * 10^{-8}.$$

Combining terms, (8.3) gives a deviation of .0005. A normally distributed random variable with mean 1.0175 and deviation .0005 has a 10% probability of being less than 1.0168, the realized value.

The estimate for $y_{1,10}$ can be obtained from (3.1) as:

$$(8.7) \qquad y_{1,10} = y_{19}q_{19}.$$

The expected value of (8.7), given that $y_{19} = 8747$, is:

$$(8.8) \qquad E\left\{y_{1,10} \middle| y_{19} = 8747\right\} = y_{19}E\left\{q_{19}\right\} = 8747 * 1.0175 = 8900.$$

And the variance of $y_{1,10}$ is:

$$(8.9) \qquad Var\left(y_{1,10} \middle| y_{19} = 8747\right) = y_{19}^2 Var(q_{19}) = (8747 * .0005)^2 = 4.79^2.$$

A normally distributed random variable of mean 8900 and deviation 4.79 has a 10% probability of being less than 8894, the realized value.

The point q_{91}, which has the realized value of 2.434, has the following formula:

$$(8.10) \qquad q_{91} = r_{91} + \ln(r_{91})\varepsilon_{91}.$$

The mean and the variance are:

$$(8.11) \qquad E\{q_{91}\} = E\{r_{91}\} = F_{92}/F_{91} = 2.438,$$

(8.12) $$Var(q_{91}) = Var(r_{91}) + s^2\left[\ln(r_{91})\right]^2.$$

The process and parameter variances are:

(8.13) $$s^2\left[\ln(r_{91})\right]^2 = \left[.0298 * \ln(2.438)\right]^2 = 7.1 * 10^{-4},$$

(8.14) $$Var(r_{91}) = r_{91}^2\left[\left(\frac{G_{92}}{F_{92}} - \frac{G_{91}}{F_{91}}\right)^2 Var(\mu) + \Lambda + 8^4\left(\frac{K_{92}}{F_{92}} - \frac{K_{91}}{F_{91}}\right)^2 Var(\chi_2)\right] = 2.7 * 10^{-4}.$$

Combining terms in (8.12) gives a deviation of .031. A normally distributed random variable of mean 2.438 and deviation .031 has a 45% probability of being less than 2.434, the realized value.

The mean and variance of y_{92} are:

(8.15) $$E\{y_{92}|y_{91} = 5300\} = y_{91} * E\{r_{91}\} = 5300 * 2.438 = 12921.$$

(8.16) $$Var(y_{92}|y_{91} = 5300) = y_{91}^2 Var(q_{91}) = (5300 * .031)^2 = 165^2.$$

We note that the variances of q_{19} and q_{91} have very different orders of magnitudes. Their ratio is:

(8.17) $$\frac{Var(q_{91})}{Var(q_{19})} = \left(\frac{.028}{.0007}\right)^2 \approx 3200.$$

The variance of q_{91} is therefore several thousand times that of q_{19}. This is not surprising, because there is much more development, and variability, at the former than at the latter. The relative values of 2.438 and 1.018 attest to this. We can carry this further by saying that the error of a point at full development is zero. This conclusion is not only reasonable, but also inescapable, if we think with this perspective: given a realized value at full development, the estimate at the succeeding period is known with absolute certainty, namely that very same value. We note that, at full development, the parameter error is also zero, because an infinitesimal perturbation of the parameters cannot nudge the c.d.f. from unity.

The role of the scaling function is imperative in that it is the mechanism through which the different orders of errors are recognized. Without it, the process errors of (8.4) and (8.13) would be exactly equal, and the ratio in (8.17) would be very close to unity. In effect, the model would not be able to differentiate the widely divergent orders of errors. For points far from full development, the variances are understated; and for points close to full development, they are overstated. Because the normalized error is essentially the ratio of the actual error to the expected average error, the effect on it is just the opposite. For points far from full development, the normalized errors are overstated; and for points close to full development, they are understated. It is therefore not surprising that the model without the scaling function cannot pass the normality test.

9. COMPARISON OF SCALING AND PROPORTIONALITY FUNCTIONS

In this section, we want to compare the effects of different scaling functions. First, we set the proportionality function $w(y_{ij}) = 1$ in (3.4), then calculate the solutions to the Table 1 triangle using four different scaling functions, in addition to the logarithmic. In Table 10, we show the total liabilities, the deviations, the chi square values, and the implied percentages.

	Name	$b(r_{ij})$	y_T	deviation	χ^2	percent	comment
1	logarithmic	$\ln(r_{ij})$	186,459	980	3.33	50.4%	good
2	linear	$r_{ij}-1$	186,469	1095	5.11	27.6%	good
3	square root	$(r_{ij}-1)^{1/2}$	186,794	1331	11.11	2.5%	fair
4	square	$(r_{ij}-1)^2$	179,448	10146	10.00	4.0%	poor
5	no scaling	1	187,037	divergent	43.56	0.0%	invalid

Table 10: Comparison of scaling functions.

For the no scaling function, the infinite series for the process errors of (5.10) have no limits. The two best scaling functions are the logarithmic and the linear. Their estimates are identical, for all intents and purposes. The logarithmic gives the slightly lower deviation and chi square value.

Incidentally, as their argument approaches unity, these two functions have the same asymptotic behavior. Symbolically,

$$(9.1) \qquad \ln(r) \sim r-1, \quad r \to 1.$$

No other function in Table 10 shares this property. We believe that the two functions do indeed have the correct error scaling.

We have analyzed quite a number of different loss triangles. The logarithmic and the linear invariably give nearly identical estimates and deviations, but the former consistently gives the lower chi square values. We therefore select the logarithmic as the most appropriate scaling function.

In Table 11, with the logarithmic scaling function, we compare three proportionality functions: the linear, logarithmic and square root.

	Name	$y_{ij}w(y_{ij})$	y_T	deviation	χ^2	percent	Comment
1	Linear	y_{ij}	186,459	980	3.33	50.4%	good
2	Logarithmic	$\ln(y_{ij})$	186,902	1182	9.78	4.4%	fair
3	Square root	$y_{ij}^{1/2}$	186,626	1041	5.56	23.5%	fair

Table 11: Comparison of proportionality functions.

Among the three proportionality functions, the linear gives the least deviation and chi square value.

From the results in Tables 10 and 11, we choose the logarithmic scaling function and linear proportionality function as the best combination.

10. PARAMETRIZATION

A solution is overparametrized if it quantifies at least one statistically insignificant coefficient. A solution is underparametrized if it omits at least one statistically significant coefficient. A solution has correct parametrization if it is neither overparametrized nor underparametrized.

In general, overparametrization leads to a smaller sum of squares of errors. But this does not lead to greater accuracy. This is manifested in two ways. First, the number of degrees of freedom, the denominator in (4.9), decreases, counteracting the smaller numerator. Secondly and more importantly, since more variables have to be estimated, the mutual interference among them increases and the elements of the inverse matrix in (4.11) generally increase in absolute value.

Underparametrization has the reverse effects: the sum of squares of errors increases, the number of degrees of freedom increases, and the elements of the inverse matrix generally decrease.

For the Table 1 triangle, Table 12 quantifies the results of parametrization. The overparametrization quantifies all function parameters up to and including the quadratic coefficients. The underparametrized solution quantifies only the base coefficients.

	Solution	y_T	deviation
1	Correct parametrization	186,459	980
2	Overparametrization	182,097	5490
3	Underparametrization	188,852	2535

Table 12: Parametrization.

The overparametrized solution has a much larger deviation, And, in this particular example, it is consistent. That is, if the solution of mean 182,097 and deviation 5490 indeed is correct and unbiased, there is considerable probability of attaining at least the correctly parametrized value of 186,459.

Our experience indicates that overparametrization invariably leads to higher deviations. The reason for this is simple: the more variables there are to be estimated, the less accuracy with which they can be estimated. The decreased accuracy translates into higher parameter errors. In our numerical tests, we usually find the overparametrized solutions to be consistent.

In this case, the underparametrized solution yields a consistent estimate and a higher deviation. But experience tells us that underparametrization can lead to inconsistent estimates and lower deviations. An underparametrized estimate is inconsistent when the difference between itself and the correctly parametrized estimate is well outside the range of the underparametrized deviation. The reason for lower deviations due to underparametrization is: the fewer variables there are to be estimated, the more accuracy with which they can be estimated. The reason for inconsistent results is: some statistically significant variables are being omitted.

The discussion in this section indicates that, if high deviations and misleading results are to be avoided, we must insure correct parametrization.

11. DISCUSSION

A) Cumulative distribution functions. We only use functions of three parameters, because we believe only they must have at least that number of parameters to have the flexibility to describe real payment patterns. We identify four such candidates: the transformed log-normal, transformed normal, transformed log-gamma, and transformed gamma. We find all Pareto-type functions to be unstable in our iterative regression scheme.

Of the four functions, only the log-normal works well for Workers Compensation, Products Liability and Medical Malpractice, the longest tailed liability lines. But none works well for Commercial Auto Liability, Personal Auto Liability and Commercial Multiple Peril, the shorter tailed liability lines. Fortunately, we have developed a class of functions for the latter lines. We will present it in another paper.

B) Type of data. The data on which we tested these models have always been paid loss. The question is whether the model could work as well on reported data. As formulated in this paper, the answer is negative. We give two reasons for this and suggest a possible remedy. The two reasons are related.

First, a c.d.f. is by construction monotonic from zero to unity. Often a reported pattern is not, surpassing unity at some intervals. This happens because of over-reserving: case reserves were set higher than actual payments. In an ideal world with perfect case reserving, this would not happen, because, when reported, case reserves would be set at exactly the future paid amounts. Therefore the ideal incurred pattern would also be monotonic. And the flip side, under-reserving, must also be prevalent. The inference is that actual incurred amounts have errors, because the case reserves cannot be set with perfect foresight.

And that brings us to the second reason. For this model, we assume that the actual paid amounts have no errors. While this certainly is not entirely true, it is much less true of reported data. Therefore, in

working with reported data, it is imperative that we account for the errors associated with the actual data. Since the two reasons are more or less related, one remedy may rectify both.

For the y_{kj} in (3.6), we use the actual amount, and assume there is no associated error. Instead, we could use the *estimated* value, which has a quantifiable error. In effect, we are saying that since actual reported amounts have unknown errors, we should instead work with estimated amounts, for which the errors can be estimated. This also has an additional advantage that addresses the first problem: even if the actual reported pattern is not monotonic, the theoretical pattern could still be. This very distinction between using the actual and estimated values goes back to the point made by MACK (1995). The model based on estimated, as opposed to actual, estimates, is another interesting avenue of research.

With the above discussion in mind, an analysis of reported data, assuming that every theoretical obstacle can be overcome, may yield much higher parameter and process errors than that on the corresponding paid data. If such is indeed the case, there may not be much additional value in the consideration of reported data.

If paid data have considerable amounts of salvage and subrogation, they can also be non-monotonic. In such cases, it may be best to analyze the data gross of salvage and subrogation.

C) Advantages of the model. We generalize the difficulties of the traditional chain ladder fall into three categories: non-stochastic variations in the ata factors, limited information, and tail factor. We recapitulate how our model addresses each category.

i. The model simulates the non-stochastic variations in the ata factors. The statistical significance of all parameters is systematically determined. We have tested six liability lines, those mentioned in the first segment of this section. We have considered loss triangles for both individual companies and industry-wide data in the United States. And we have yet to encounter a single triangle in which only the base coefficients are statistically significant. In every case, at least some non-stochastic variations are evident.

ii. Limited information, as used in the traditional chain ladder, surfaces in a few instances. One is that averaging may only use the last few available years. Secondly, to estimate the ata factor in any development period, only information in that period is used. In contrast, to make the estimation at any single point, our regression scheme uses information available everywhere. This should decrease the parameter errors.

iii. Our model gives the tail factor for each accident year. In addition, it yields the variance of an ultimate loss, and it clearly divides that variance into parameter and process errors.

REFERENCES

BERQUIST , J.R. and SHERMAN, R.E. (1977) Loss Reserve Adequacy Testing: A Comprehensive, Systemic Approach. *Proceedings of the Casualty Actuarial Society,* LXIV.

HALLIWELL, L.J. (1996) Loss Prediction by Generalized Least Squares. Casualty Actuarial Society Annual Meeting.

MACK, T. (1995) Which Stochastic Model is Underlying the Chain Ladder Method? *Casualty Actuarial Society Forum.*

MURPHY, D.M., (1995) Unbiased Loss Development Factors. *Proceedings of the Casualty Actuarial Society,* LXXXII.

STANARD, J.N. (1985) A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques. *Proceedings of the Casualty Actuarial Society,* LXXII.

VERRALL, R.J., (1990) Bayes and Empirical Bayes Estimation for the Chain Ladder Model. *ASTIN Bulletin,* 20.

ZEHNWIRTH, B. (1990) Stochastic Development Factor Models. Casualty Loss Reserve Seminar, Dallas, 1990.

Appendix A:
The Derivation of
a Perturbation Expression

In this appendix, we derive the perturbation expression in (4.3). All the other perturbation expressions can be obtained in a similar fashion.

In perturbation theory, our objective is to express any quantity, such as the LHS of (4.3):

$$(A.1) \qquad r_{kj}(\mu+\Delta\mu,\sigma+\Delta\sigma,\tau+\Delta\tau),$$

as a linear function of the increments $(\Delta\mu,\Delta\sigma,\Delta\tau)$.

We use (3.2) to write (A.1) as:

$$(A.2) \qquad r_{kj}(\mu+\Delta\mu,\sigma+\Delta\sigma,\tau+\Delta\tau)=\frac{F(j+1;\mu+\Delta\mu,\sigma+\Delta\sigma,\tau+\Delta\tau)}{F(j;\mu+\Delta\mu,\sigma+\Delta\sigma,\tau+\Delta\tau)}.$$

We need to express the denominator of (A.2) as a perturbation expression, which is just (4.2).

$$(A.3) \qquad \begin{aligned} F(j;\mu+\Delta\mu,\sigma+\Delta\sigma,\tau+\Delta\tau) &= F(j;\mu,\sigma,\tau)+ \\ &\quad G(j;\mu,\sigma,\tau)\Delta\mu+H(j;\mu,\sigma,\tau)\Delta\sigma+K(j;\mu,\sigma,\tau)\Delta\tau \\ &= F_{kj}+G_{kj}\Delta\mu+H_{kj}\Delta\sigma+K_{kj}\Delta\tau. \end{aligned}$$

In (A.3), we expressly recognize that the function parameters may vary by accident years. Similarly,

$$(A.4) \qquad F(j+1;\mu+\Delta\mu,\sigma+\Delta\sigma,\tau+\Delta\tau)=F_{k,j+1}+G_{k,j+1}\Delta\mu+H_{k,j+1}\Delta\sigma+K_{k,j+1}\Delta\tau.$$

If we now put (A.3) in the denominator, then we have:

$$(A.5) \qquad \begin{aligned} \frac{1}{F(j;\mu+\Delta\mu,\sigma+\Delta\sigma,\tau+\Delta\tau)} &= \frac{1}{F_{kj}+G_{kj}\Delta\mu+H_{kj}\Delta\sigma+K_{kj}\Delta\tau}= \\ \frac{1}{F_{kj}}\frac{1}{1+\left(G_{kj}/F_{kj}\right)\Delta\mu+\left(H_{kj}/F_{kj}\right)\Delta\sigma+\left(K_{kj}/F_{kj}\right)\Delta\tau} &= \\ \frac{1}{F_{kj}}\left[1-\frac{G_{kj}}{F_{kj}}\Delta\mu-\frac{H_{kj}}{F_{kj}}\Delta\sigma-\frac{K_{kj}}{F_{kj}}\Delta\tau\right]. \end{aligned}$$

In (A.5), we retain only the linear term of the following Taylor's series expansion:

$$(A.6) \qquad \frac{1}{1+x}=1-x+x^2-\Lambda\ .$$

When we substitute (A.4) and (A.5) into (A.2), we get precisely (4.3):

$$(A.7)\qquad r_{kj}(\mu+\Delta\mu,\sigma+\Delta\sigma,\tau+\Delta\tau)=r_{kj}(\mu,\sigma,\tau)\left[1+\left(\frac{G_{k,j+1}}{F_{k,j+1}}-\frac{G_{kj}}{F_{kj}}\right)\Delta\mu+\left(\frac{H_{k,j+1}}{F_{k,j+1}}-\frac{H_{kj}}{F_{kj}}\right)\Delta\sigma+\left(\frac{K_{k,j+1}}{F_{k,j+1}}-\frac{K_{kj}}{F_{kj}}\right)\Delta\tau\right].$$

Every perturbation expression can be derived in a similar manner, and is simply the result of repeated and appropriate applications of the Taylor's series expansions.

Appendix B:
The Finite Limit of the
Process-Error Infinite Series

In this appendix, we prove that the infinite series in (5.10) has a finite limit. It is sufficient to show that:

$$(B.1)\qquad \sum_{t=A}^{\infty}\left[\frac{\ln(r_t)}{r_t}\right]^2<\infty,$$

where, without loss of generality, we suppress the parameter dependency on time. A is some positive integer, which can be as large as we wish.

We rewrite the equation for the transformed log-normal of (2.2) as:

$$(B.2)\qquad F(t)=\Phi\left[(\ln t)^r;\mu,\sigma\right]=1-\int_t^{\infty}f(t)dt,$$

where we use the partial density function:

$$(B.3)\qquad f(t)=\frac{dF}{dt}=\frac{\tau}{\sqrt{2\pi}}\frac{(\ln t)^{r-1}}{t}\exp\left[-\frac{1}{2}\left(\frac{(\ln t)^r-\mu}{\sigma}\right)^2\right].$$

To obtain the asymptotic form of the development factor, we use the definition of (2.1), and perform successive approximations:

$$(B.4)\qquad r_t=\frac{F(t+1)}{F(t)}=\frac{1-\int_{t+1}^{\infty}f(t)dt}{1-\int_t^{\infty}f(t)dt}\approx\left[1-\int_{t+1}^{\infty}f(t)dt\right]\left[1+\int_t^{\infty}f(t)dt\right]$$

$$\approx 1+\int_t^{t+1}f(t)dt\le 1+f(t)$$

Taking the log of (B.4) and retaining only the first term of the resultant Taylor's series, we obtain:

$$(B.5) \qquad \ln(r_i) \le \ln\left[1 + f(t)\right] \approx f(t).$$

We simplify the elements of the series in (B.1) as:

$$(B.6) \qquad \left[\frac{\ln(r_i)}{r_i}\right]^2 \le \left[\ln(r_i)\right]^2 \le f^2(t) = \frac{\tau^2}{2\pi}\frac{(\ln t)^{2(\tau-1)}}{t^2}\exp\left[-\left(\frac{(\ln t)^\tau - \mu}{\sigma}\right)^2\right].$$

Therefore, (B.1) is satisfied if we have the following equality:

$$(B.7) \qquad \sum_{t=A}^{\infty}\frac{(\ln t)^{2(\tau-1)}}{t^2}\exp\left[-\left(\frac{(\ln t)^\tau - \mu}{\sigma}\right)^2\right] < \infty,$$

where we drop all multiple constants. Equation (B.7) is in turn satisfied if we have the following integral inequality:

$$(B.8) \qquad \int_A^{\infty}\frac{(\ln t)^{2(\tau-1)}}{t^2}\exp\left[-\left(\frac{(\ln t)^\tau - \mu}{\sigma}\right)^2\right]dt < \infty.$$

We can certainly pick an A such that:

$$(B.9) \qquad \frac{(\ln t)^{2(\tau-1)}}{t} < 1, \quad A < t.$$

Therefore, we have the following inequality:

$$(B.10) \qquad \int_A^{\infty}\frac{(\ln t)^{2(\tau-1)}}{t^2}\exp\left[-\left(\frac{(\ln t)^\tau - \mu}{\sigma}\right)^2\right]dt < \int_A^{\infty}\exp\left[-\left(\frac{(\ln t)^\tau - \mu}{\sigma}\right)^2\right]\frac{dt}{t}.$$

We make the following substitutions:

$$(B.11) \qquad x = \frac{(\ln t)^\tau - \mu}{\sigma}, \quad \frac{\sigma}{\tau}\frac{dx}{(\sigma x - \mu)^{1-1/\tau}} = \frac{dt}{t}, \quad p = 1 - 1/\tau.$$

With (B.11), equation (B.10) can be written as:

$$(B.12) \qquad \int_B^{\infty}\frac{e^{-x^2}}{(\sigma x - \mu)^p}dx < \infty, \quad B = \frac{(\ln A)^\tau - \mu}{\sigma}.$$

The inequality in (B.12) holds, irrespective of the value of p, since the exponential decays much faster than any power of x.

We thus prove the inequality of (B.1). This line of argument is applicable to any distribution, the partial density function of which decays exponentially.