UNBIASED LOSS DEVELOPMENT FACTORS

DANIEL M. MURPHY

Abstract

Casualty Actuarial Society literature is inconclusive regarding whether the loss development technique is biased or unbiased, or which of the traditional methods of estimating link ratios is best. This paper frames the development process in a least squares regression model so that those questions can be answered for link ratio estimators commonly used in practice, and for two new average development factor formulas. As a byproduct, formulas for variances of point estimates of ultimate loss and loss reserves are derived that reflect both parameter risk and process risk. An approach to measuring confidence intervals is proposed. A consolidated industry workers' compensation triangle is analyzed to demonstrate the concepts and techniques. The results of a simulation study suggest that in some situations the alternative average loss development factor (LDF) formulas may outperform the traditional estimators, and that the performance of the incurred loss development technique can approach that of the Bornhuetter-Ferguson and Stanard-Bühlmann techniques.

1. INTRODUCTION

Three common methods of estimating link ratios are the Simple Average Development (SAD) method—the arithmetic average of the link ratios; the Weighted Average Development (WAD) method—the sum of losses at the end of the development period divided by the sum of the losses at the beginning; and the Geometric Average Development (GAD) method—the n^{th} root of the product of n link ratios. Casualty actuarial literature is inconclusive regarding which method is "best" or indeed whether the methods are biased or unbiased. See, for example, Stanard [9] and Robertson's discussion [7]. The purpose of this paper is to present a mathematical framework for evaluating the accuracy of these methods; to suggest alternatives; and to unearth valuable information about the variance of the estimates of developed ultimate loss.

It is assumed that the actuary has exhausted all adjustments for systematic or operational reasons why a development triangle may appear as it does, and the only concern left is how to deal with the remaining noise. Although the paper uses accident year to refer to the rows of the triangle, the theory also applies to policy year and report year triangles.

2. POINT ESTIMATES

When we say that we expect the value of incurred losses as of, say, 24 months to equal the incurred value as of 12 months times a link ratio, it is possible that what we really mean is this: the value of incurred losses as of 24 months is a random variable whose expected value is conditional on the 12 month incurred value, and equals that 12 month value times an unknown constant. Symbolically,

$$y = bx + e,$$

where x and y are the current and next evaluations, respectively; b is the unknown constant development factor, called the age-to-age factor or link ratio; and e represents random noise. The first step in developing losses is estimating the link ratios.

Expected Value of the Link Ratio

Let us first generalize, and suppose that the relationship between x and y is fully linear rather than strictly multiplicative. The more general model is

Model I $y = a + bx + e.$

$E(e) = 0$; $\text{Var}(e)$ is constant across accident years; and the e's are uncorrelated between accident years and are independent of x.

This model is clearly a regression of 24-month losses y on 12-month losses x. Although x is *a priori* a random variable, once an evaluation is made it is treated as a constant for the purpose of loss development. More precisely, the model says that the expected value of the random variable y conditional on the random variable x is linear in x: $E(y \mid x) = a + bx$. With this understanding of the relationship between x and y, all classical results of least squares regression may be brought to bear on the theory of loss development. See, for example, Scheffé [8]. For the remainder of this paper, all expectations are conditional on the current evaluation.

The well known Gauss-Markoff Theorem says that the Best Linear Unbiased Estimates (BLUE) of a and b are the least squares estimates, denoted $\hat{a}$ and $\hat{b}$:

$$\hat{b} = \frac{\Sigma (x - \bar{x})\, y}{\Sigma (x - \bar{x})^2}$$

and

$$\hat{a} = \bar{y} - \hat{b}\,\bar{x} \ .$$

This model will be referred to as the Least Squares Linear (LSL) model.

Section 5 presents an argument that claim count development may follow the LSL model, supported by the simulation study of Appendix B. However, if one believes the y-intercept should truly be zero, perhaps the model to use is

Model II $y = bx + e.$

$E(e) = 0$; $\mathrm{Var}(e)$ is constant across accident years; and the e's are uncorrelated between accident years and are independent of x.

This model would not be appropriate if there were a significant probability that y should not equal zero when x does.

It is well known that the BLUE estimator for b under Model II is

$$\hat{b} = \frac{\Sigma xy}{\Sigma x^2}. \tag{2.1}$$

This model will be referred to as the Least Squares Multiplicative (LSM) model.

Can the LSL or LSM assumptions be revised to say something about the more common development factor averages? Take the assumption of constant variance across accident years. Triangles of incurred or paid dollars under the force of trend may not conform to this assumption. On-leveling the loss triangle may try to adjust for such heteroskedasticity, but may introduce unwelcome side effects as well. A model that speaks directly to the issue of non-constant variances is:

Model III $y = bx + xe.$

$E(e) = 0$; $\mathrm{Var}(e)$ is constant across accident years; and the e's are uncorrelated between accident years and independent of x.

This model differs from Model II in that it explicitly postulates a dependent relationship between the current evaluation, x, and the error term, xe. Divide both sides of this equation by x. This model also says that the ratio of consecutive evaluations is constant across accident years. In other words, it is the development percent, not the development dollars, and the random deviation in that percent that behave consistently from one accident year to the next.

This model's BLUE for b is the simple average development (SAD) factor, denoted b_{SAD}. This is easy to see. Transform Model III as follows:

Model III′ $\quad y/x = b + e$

or $\quad u = bv + e,$

where v is identically equal to unity. Formula 2.1 says that

$$\hat{b} = \frac{\Sigma vu}{\Sigma v^2} = \frac{\sum \frac{y}{x}}{\Sigma\, 1} = \frac{1}{n} \sum \frac{y}{x}$$

which is b_{SAD} .

One may object that the proportionality of the error term to the full value of x overemphasizes the true relationship. It may seem more plausible that the variance of y, or the square of the error term, is proportional to x. The model[1] that describes this relationship is:

Model IV $\quad y = bx + \sqrt{x}\, e.$

E $(e) = 0$; Var (e) is constant across accident years; and the e's are uncorrelated between accident years and independent of x.

This model's BLUE for b is the weighted average development (WAD) factor, denoted b_{WAD}. This is also easy to see. Transforming

[1] This model was inspired by Dr. Thomas Mack at the presentation of his 1993 Theory of Risk prize paper "Measuring the Variability of Chain Ladder Reserve Estimates."

Model IV by dividing both sides by $\sqrt{x}$ turns it into a simple regression of $u = y/\sqrt{x}$ onto $v = \sqrt{x}$. Formula (2.1) becomes:

$$\hat{b} = \frac{\Sigma uv}{\Sigma v^2} = \frac{\sum \frac{y}{\sqrt{x}} \sqrt{x}}{\Sigma x} = \frac{\Sigma y}{\Sigma x},$$

which is b_{WAD}. Thus, the weighted average is the best estimator if the variance of the development error is proportional to the beginning evaluation.

A fifth model that can also adjust for trend is:

Model V $\qquad y = bxe.$

E $(e) = 1$; Var(e) is constant across accident years; and the e's are uncorrelated between accident years and independent of x.

This model says that random noise shocks the development process multiplicatively, and may be appropriate in those situations in which the random error in the percentage development is itself expected to be skewed. The BLUE for b under Model V is the geometric average development (GAD) factor, denoted b_{GAD}. Indeed, transform Model V by taking the logarithm of both sides:

$$\ln y = \ln b + \ln x + \ln e$$

or

$$\ln y - \ln x = \ln b + \ln e$$

which is of the form

$$u = b'v + e'$$

where $b' = \ln b$, $v = 1$, and E $(e') = 0$. Then Formula (2.1) simplifies to:

$$\hat{b}' = \frac{\Sigma uv}{\Sigma v^2} = \frac{1}{n}\Sigma u = \frac{1}{n}\Sigma(\ln y - \ln x) = \frac{1}{n}\Sigma \ln\frac{y}{x}.$$

Therefore, the least squares estimator of the "untransformed" parameter b is:

$$\hat{b} = e^{\hat{b}'} = \exp\left(\frac{1}{n}\sum \ln\frac{y}{x}\right) = \left(\exp\sum \ln\frac{y}{x}\right)^{1/n} = \left(\Pi\frac{y}{x}\right)^{1/n}$$

which is b_{GAD}.

For the remainder of the paper, the Linear model will refer to LSL. The Multiplicative models will refer to Models II to V—LSM, SAD, WAD, and GAD—unless otherwise noted.

Estimate of the Next Evaluation

The following point estimates of the expected value of incurred losses as of the next evaluation given the current evaluation are unbiased under the assumptions of their respective models:[2]

Linear	Multiplicative
$\hat{y} = \hat{a} + \hat{b}x$	$\hat{y} = \hat{b}x$.

Estimated Ultimate Loss: A Single Accident Year

The Chain Ladder Method states that if b_1 is a link ratio from 12 to 24 months, b_2 is a link ratio from 24 to 36 months, etc., and if U is the number of links required to reach ultimate, then $B_U = b_1\, b_2 \ldots b_U$ is the (to-ultimate) loss development factor (LDF). The implicit assumption is that future development is independent of prior development. This assumption may not hold in practice when, for example,

[2] Theorem 1 in Appendix C proves this for the linear model. The proof for the multiplicative models is similar.

management issues orders for a one-time-only strengthening in case reserves.

This all-important Chain Ladder Independence Assumption (CLIA) says that the relationship between consecutive evaluations does not depend on the relationship between any other pair of consecutive evaluations. In mathematical terms, the random variable corresponding to losses evaluated at one point in time *conditional on the previous evaluation* is independent of any other evaluation *conditional on its previous evaluation*. A direct result of this assumption is the fact that an unbiased estimate of a loss development factor is the product of the unbiased link ratio estimates; symbolically, $\hat{B}_U = \hat{b}_1 \hat{b}_2 \ldots \hat{b}_U$.

The very simplicity of the closed form LDF is one of the beauties of the multiplicative chain ladder method. But a closed form, to-ultimate expression is not necessary, and quite cumbersome for the more general LSL approach. Instead, this paper proposes the use of a recursive formula. A recursive estimate of developing ultimate loss illuminates the missing portion of the triangle (clarifying the communication of the analysis to management and clients), enables the actuary to switch models mid-chain, and is straightforward to program, even in a spreadsheet. Perhaps the most compelling reason, however, is that a recursive estimate is invaluable for calculating variances of predicted losses. (See Section 3.)

The mathematical theory for developing recursive estimates of ultimate loss conditional on the current evaluation proceeds as follows. Consider a single fixed accident year. Let x_0 denote the (known) current evaluation and let $x_n \mid x_0$ denote the random variable corresponding to the n^{th} subsequent (unknown) evaluation conditional on the current evaluation. The goal is to find an unbiased estimator for $x_n \mid x_0$.

By definition, an unbiased estimate of $x_n \mid x_0$ is one which estimates $\mu_n = E(x_n \mid x_0)$. Let $\hat{\mu}_n$ denote such an estimate of μ_n. Theorem 2 (Appendix C) proves that the $\hat{\mu}_n$ defined according to the recursive formulas in Table 2.1 are unbiased under the assumptions of their respective models.

TABLE 2.1

POINT ESTIMATE — $\hat{\mu}_n$

FUTURE VALUE OF A SINGLE ACCIDENT YEAR
n TIME PERIODS IN THE FUTURE

Model	$n = 1$	$n > 1$
Linear	$\hat{a}_1 + \hat{b}_1 x_0$	$\hat{a}_n + \hat{b}_n \hat{\mu}_{n-1}$
Multiplicative	$\hat{b}_1 x_0$	$\hat{b}_n \hat{\mu}_{n-1}$

An unbiased estimate of ultimate loss conditional on the current evaluation is therefore $\hat{\mu}_U$.

Estimated Total Ultimate Loss: Multiple Accident Years

An estimate of total ultimate loss for more than one accident year combined could be obtained by simply adding up the separate accident year $\hat{\mu}_U$'s. However, a recursive expression is preferred primarily for the purpose of calculating variances because development estimates of ultimate loss for different accident years are not independent.

Notation quickly obscures the derivation, but the idea of a recursive estimate of total ultimate loss for multiple accident years is this. Start at the bottom left corner of the triangle and develop the youngest accident year to the next age. Then, add that estimate to the current evaluation of the second youngest accident year, and develop the sum to the next age. Continue recursively. An unbiased estimate of total losses at ultimate will be the final sum.

The formulas are developed as follows. To keep the indices from becoming too convoluted, index the rows of the triangle in reverse order so that the youngest accident year is the zero row, the next youngest is row 1, and so on. Next, index the columns so that the 12 month column is the zero column, the 24 month column is column 1, etc. A full triangle of $N + 1$ accident years appears in Figure 1. Let

$$S_n = \sum_{i=0}^{n-1} x_{i,n} \mid x_{i,i}$$

denote the sum of the accident years' future evaluations conditional on the accident years' current evaluations, and set $M_n = \mathrm{E}(S_n)$. We are looking for an unbiased estimate $\hat{M}_n$ of M_n. Recursive formulas for $\hat{M}_n$ are given in Table 2.2. (See Theorem 9 in Appendix C.)

FIGURE 1
NOTATION FOR THE KNOWN AND UNKNOWN
PORTIONS OF A LOSS TRIANGLE

Age of Accident Year

A/Y	0	1	2		n-1	n	n+1		N-1	N
N	$x_{N,0}$	$x_{N,1}$	$x_{N,2}$	...	$x_{N,n-1}$	$x_{N,n}$	$x_{N,n+1}$	...	$x_{N,N-1}$	$x_{N,N}$
N-1	$x_{N-1,0}$	$x_{N-1,1}$	$x_{N-1,2}$	...	$x_{N-1,n-1}$	$x_{N-1,n}$	$x_{N-1,n+1}$	...	$x_{N-1,N-1}$	$x_{N-1,N} \mid x_{N-1,N-1}$
N-2	$x_{N-2,0}$	$x_{N-2,1}$	$x_{N-2,2}$	...	$x_{N-2,n-1}$	$x_{N-2,n}$	$x_{N-2,n+1}$	...	$x_{N-2,N-1} \mid x_{N-2,N-2}$	$x_{N-2,N} \mid x_{N-2,N-2}$
	⋮	⋮	⋮		⋮	⋮	⋮		⋮	⋮
n	$x_{n,0}$	$x_{n,1}$	$x_{n,2}$	...	$x_{n,n-1}$	$x_{n,n}$	$x_{n,n+1} \mid x_{n,n}$	...	$x_{n,N-1} \mid x_{n,n}$	$x_{n,N} \mid x_{n,n}$
n-1	$x_{n-1,0}$	$x_{n-1,1}$	$x_{n-1,2}$	...	$x_{n-1,n-1}$	$x_{n-1,n} \mid x_{n-1,n-1}$	$x_{n-1,n+1} \mid x_{n-1,n-1}$	...	$x_{n-1,N-1} \mid x_{n-1,n-1}$	$x_{n-1,N} \mid x_{n-1,n-1}$
n-2	$x_{n-2,0}$	$x_{n-2,1}$	$x_{n-2,2}$	...	$x_{n-2,n-1} \mid x_{n-2,n-2}$	$x_{n-2,n} \mid x_{n-2,n-2}$	$x_{n-2,n+1} \mid x_{n-2,n-2}$	...	$x_{n-2,N-1} \mid x_{n-2,n-2}$	$x_{n-2,N} \mid x_{n-2,n-2}$
	⋮	⋮	⋮		⋮	⋮	⋮		⋮	⋮
1	$x_{1,0}$	$x_{1,1}$	$x_{1,2} \mid x_{1,1}$	...	$x_{1,n-1} \mid x_{1,1}$	$x_{1,n} \mid x_{1,1}$	$x_{1,n+1} \mid x_{1,1}$	...	$x_{1,N-1} \mid x_{1,1}$	$x_{1,N} \mid x_{1,1}$
0	$x_{0,0}$	$x_{0,1} \mid x_{0,0}$	$x_{0,2} \mid x_{0,0}$	...	$x_{0,n-1} \mid x_{0,0}$	$x_{0,n} \mid x_{0,0}$	$x_{0,n+1} \mid x_{0,0}$	...	$x_{0,N-1} \mid x_{0,0}$	$x_{0,N} \mid x_{0,0}$

TABLE 2.2

POINT ESTIMATE — $\hat{M}_n$

TOTAL FUTURE VALUE OF MULTIPLE ACCIDENT YEARS
n TIME PERIODS IN THE FUTURE

Model	$n = 1$	$n > 1$
Linear	$\hat{a}_1 + \hat{b}_1 x_{0,0}$	$n\hat{a}_n + \hat{b}_n (\hat{M}_{n-1} + x_{n-1,n-1})$
Multiplicative	$\hat{b}_1 x_{0,0}$	$\hat{b}_n (\hat{M}_{n-1} + x_{n-1,n-1})$

Estimated Reserves for Outstanding Loss

Assuming paid dollars to date are not expected to be adjusted significantly,[3] an unbiased estimate of outstanding loss for a single accident year is $\hat{\mu}_U$ – paid to date. For multiple accident years, an unbiased estimate is $\hat{M}_U$ – total paid to date.

3. VARIANCE

The least squares point estimators of development factors, ultimate losses, or reserves are functions of random variables. As such, they are themselves random variables with their own inherent variances. Estimates of these variances will be addressed in turn.

Variance of the Link Ratio Estimators

For the LSL or LSM models, the formula for the variance of the link ratio estimator is a straightforward result of least squares theory. For the other models, one must first transform the data so that the model takes on the usual regression form (i.e., the error term does not involve x).[4] Once the regression theory yields up the estimate of

[3] Which is not true if salvage, subrogation, or deductible recoveries could be significant.

[4] Model III, for example.

$\text{Var}(\hat{b})$, one applies that to the original, "untransformed" data in the formulas for estimated future losses (below).

We will adopt the convention that a "hat" (^) over a quantity denotes an unbiased estimate of that quantity. Unbiased estimates of the variances of the link ratio estimators are given in Table 3.1. These formulas can be found in many statistics texts. See Miller and Wichern [6], for example.

TABLE 3.1

ESTIMATES OF THE VARIANCES OF THE LINK RATIO ESTIMATORS

Model	$\hat{\text{Var}}(\hat{a})$	$\hat{\text{Var}}(\hat{b})$
LSL	$\frac{\Sigma x^2}{n\Sigma (x-\bar{x})^2}\hat{\sigma}^2$	$\frac{\hat{\sigma}^2}{\Sigma (x-\bar{x})^2}$
LSM	n/a	$\frac{\hat{\sigma}^2}{\Sigma x^2}$

The average "x value" $\bar{x}=\frac{1}{I}\sum x_i$ is the average of the known evaluations of prior accident years as of the age of the link ratio being estimated; I is the number of accident years used in the average. The unbiased estimate $\hat{\sigma}^2$ of the variance σ^2 of the error term e, sometimes denoted s^2, is the Mean Square Error (MSE) of the link ratio regression. The MSE, or its square root s (sometimes referred to as the standard error of the y estimate), can be found in regression software output. Most regression software will also calculate $\hat{\text{Var}}(\hat{b})$, or its square root (sometimes referred to as the standard error of the coefficient).

Variance of Estimated Ultimate Loss: A Single Accident Year

It is time to make an important distinction. The point estimate of ultimate loss $\hat{\mu}_U$ from Section 2 above is an estimate of the expected value of the (conditional on x_0) ultimate loss x_U. Actual ultimate loss will vary from its expected value in accordance with its inherent variation about its developed mean μ_U . As a result, the risk that actual ultimate loss will differ from the prediction $\hat{\mu}_U$ is comprised of two components.

The first component, Parameter Risk, is the variance in the estimate of the expected value of $x_U \mid x_0$. The second component, Process Risk,[5] is the inherent variability of ultimate loss about its conditional mean μ_U. Symbolically, if (conditional on x_0) ultimate loss for a given accident year is expressed as the sum of its (conditional) mean plus a random error term ε_U,

$$x_U \mid x_0 = \mu_U + \varepsilon_U,$$

then the variance in the prediction of ultimate loss $pred_U$ is

$$\begin{aligned} \mathrm{Var}(pred_U) &= \mathrm{Var}(\hat{\mu}_U) + \mathrm{Var}(\varepsilon_U) \\ &= \text{Parameter Risk} + \text{Process Risk} \\ &= \text{Total Risk}. \end{aligned}$$

Tables 3.2 and 3.3 give recursive formulas for estimates of Parameter Risk and Process Risk, respectively.[6]

[5] This Process Risk is the conditional variance of developing losses about the conditional mean. As pertains to triangles of incurred loss dollars, it includes the unconditional *a priori* process risk of the loss distribution (mitigated by the knowledge of losses emerged to date), the random variation of the claims occurrence and reporting patterns, and the random variation within case reserves.

[6] The Parameter Risk formulas are derived in Theorem 6. The Process Risk formulas are derived in Theorem 7A for LSL and LSM, Theorem 7B for WAD, and Theorem 7C for SAD. See Appendix C.

TABLE 3.2
PARAMETER RISK ESTIMATE — $\hat{\text{Var}}(\hat{\mu}_n)$
A SINGLE ACCIDENT YEAR

Model	$n = 1$	$n > 1$
Linear	$\frac{\hat{\sigma}_1^2}{I_1} + (x_0 - \bar{x}_0)^2 \hat{\text{Var}}(\hat{b}_1)$	$\frac{\hat{\sigma}_n^2}{I_n} + (\hat{\mu}_{n-1} - \bar{x}_{n-1})^2 \hat{\text{Var}}(\hat{b}_n) +$
		$\hat{b}_n^2 \hat{\text{Var}}(\hat{\mu}_{n-1}) + \hat{\text{Var}}(\hat{b}_n) \hat{\text{Var}}(\hat{\mu}_{n-1})$
Multiplicative	$x_0^2 \hat{\text{Var}}(\hat{b}_1)$	$\hat{\mu}_{n-1}^2 \hat{\text{Var}}(\hat{b}_n) +$
		$\hat{b}_n^2 \hat{\text{Var}}(\hat{\mu}_{n-1}) + \hat{\text{Var}}(\hat{b}_n) \hat{\text{Var}}(\hat{\mu}_{n-1})$

The average "x value"

$$\bar{x}_{n-1} = \frac{1}{I_n} \sum_{I=0}^{N} x_{i,n-1}$$

is the average of the known evaluations of prior accident years as of age $n - 1$; I_n is the number of data points in the regression estimate of development from age $n - 1$ to age n. Each of the other quantities in Table 3.2 come from the loss triangle, from x_0, from Section 2, from the regression output ($\hat{\sigma}^2$, $\hat{\text{Var}}(\hat{b})$), or from the prior recursion step ($\hat{\text{Var}}(\hat{\mu}_{n-1})$). The Multiplicative models refer to LSM, WAD, and SAD, but not GAD.[7]

[7] The regression calculation on the logarithm-transformed data will provide an estimate of the variance of the transformed parameter b', but there is no easy translation to an estimate of the variance of the original parameter b. The best way to work with the GAD model is in its transformed state. See Section 4 and Theorem 8 of Appendix C. Similarly, Tables 3.3, 3.4, and 3.5 exclude mention of the GAD model.

TABLE 3.3
PROCESS RISK ESTIMATE — $\hat{\text{Var}}(x_n \mid x_0)$
A SINGLE ACCIDENT YEAR

Model	$n = 1$	$n > 1$
LSL, LSM	$\hat{\sigma}_1^2$	$\hat{\sigma}_n^2 + \hat{b}_n^2 \hat{\text{Var}}(x_{n-1} \mid x_0)$
WAD	$x_0 \hat{\sigma}_1^2$	$\hat{\mu}_{n-1}\hat{\sigma}_n^2 + \hat{b}_n^2 \hat{\text{Var}}(x_{n-1} \mid x_0)$
SAD	$x_2^2 \hat{\sigma}_1^2$	$(\hat{\mu}_{n-1}^2 + \hat{\text{Var}}(x_{n-1} \mid x_0))\sigma_n^2 + \hat{b}_n^2 \hat{\text{Var}}(x_{n-1} \mid x_0)$

Each of the quantities in Table 3.3 come from the loss triangle, Section 2, the regression output, or the prior recursion step.

Note that ultimate loss is not ultimate until the final claim is closed. Suppose it takes C development periods, $C > U$, to close out the accident year. Then the estimate of ultimate loss is not of $x_U \mid x_0$ but of $x_C \mid x_0$. Although the point estimate would be the same at age C as at age U, the variances will not be the same. Even if b_n is not significantly different from unity for $n > U$, whereby parameter risk halts at age U, process risk continues to build up, so recursive estimates of $\text{Var}(x_n \mid x_0)$ should be carried out beyond $n = U$.

Variance of Estimated Ultimate Loss: Multiple Accident Years

Actual total ultimate loss S_U for multiple (open) accident years will vary from the estimate $\hat{M}_U$ as a result of two sources of uncertainty: Parameter Risk—the variance in the estimate of M_U—and Process Risk—the inherent variance of S_U about its developed mean M_U. Symbolically, if we express total ultimate loss for multiple accident years (conditional on the current evaluation of all accident years) as the sum of its mean M_U plus a random error term E_U,

$$S_U = M_U + E_U,$$

then the variance in the prediction of total ultimate loss $pred_U$ is

$$\begin{aligned}\operatorname{Var}(pred_U) &= \operatorname{Var}(\hat{M}_U) + \operatorname{Var}(E_U)\\ &= \text{Parameter Risk} + \text{Process Risk}\\ &= \text{Total Risk.}\end{aligned}$$

Tables 3.4 and 3.5 give recursive formulas for estimates of Parameter Risk and Process Risk, respectively.[8]

TABLE 3.4
PARAMETER RISK ESTIMATE — $\hat{\operatorname{Var}}(\hat{M}_n)$
MULTIPLE ACCIDENT YEARS

Model	$n = 1$	$n > 1$
Linear	$\frac{\hat{\sigma}_1^2}{I_1} + (x_0 - \bar{x}_0)^2 \hat{\operatorname{Var}}(\hat{b}_1)$	$n^2 \frac{\hat{\sigma}_n^2}{I_n} + (\hat{M}_{n-1} + x_{n-1,n-1} - n\bar{x}_{n-1})^2 \hat{\operatorname{Var}}(\hat{b}_n) + \hat{b}_n^2 \hat{\operatorname{Var}}(\hat{M}_{n-1}) + \hat{\operatorname{Var}}(\hat{b}_n) \hat{\operatorname{Var}}(\hat{M}_{n-1})$
Multiplicative	$x_0^2 \hat{\operatorname{Var}}(\hat{b}_1)$	$(\hat{M}_{n-1} + x_{n,n})^2 \hat{\operatorname{Var}}(\hat{b}_n) + \hat{b}_n^2 \hat{\operatorname{Var}}(\hat{M}_{n-1}) + \hat{\operatorname{Var}}(\hat{b}_n)\hat{\operatorname{Var}}(\hat{M}_{n-1})$

TABLE 3.5
PROCESS RISK—$\hat{\operatorname{Var}}(S_n)$
MULTIPLE ACCIDENT YEARS

Model	$n = 1$	$n > 1$
LSL, LSM	$\hat{\sigma}_1^2$	$n\hat{\sigma}_n^2 + \hat{b}_n^2 \hat{\operatorname{Var}}(S_{n-1})$
WAD	$x_0 \hat{\sigma}_1^2$	$(\hat{M}_{n-1} + x_{n-1,n-1}) \hat{\sigma}_n^2 + \hat{b}_n^2 \hat{\operatorname{Var}}(S_{n-1})$
SAD	$x_2^2 \hat{\sigma}_1^2$	$\left(x_{n-1,n-1}^2 + \sum_{i=0}^{n-2} \hat{\mu}_{i,n-1}^2 + \hat{\operatorname{Var}}(S_{n-1})\right)\sigma_n^2 + b_n^2 \operatorname{Var}(S_{n-1})$

[8]Theorems 6 and 7 of Appendix C.

Variance of Estimated Outstanding Losses: Single or Multiple Accident Years

Assume paid losses are constant at any given evaluation. Then the variance of loss reserves equals the variance of ultimate losses.

4. CONFIDENCE INTERVALS

Confidence intervals are phrased in terms of probabilities, so this discussion can no longer avoid making assumptions about the probability distribution of the error terms, e_n. The traditional assumption is that they are normally distributed or, under GAD, lognormally distributed.

Confidence Intervals Around the Link Ratios

Let α be the probability measurement of the width of the confidence interval. Table 4.1 gives two-sided $100\alpha\%$ confidence intervals around the true LSL link ratios (a_n, b_n), where $t_\alpha(df)$ denotes Student's t distribution with df degrees of freedom and where I_n is the number of data points used in the estimate of the n^{th} link ratio. The degrees of freedom under the linear model are $I_n - 2$ because two parameters are estimated; $df = I_n - 1$ under the multiplicative models because only the single parameter b_n need be estimated.

TABLE 4.1

100 α% CONFIDENCE INTERVALS AROUND THE LINK RATIO PARAMETERS

Model	a_n	b_n
Linear	$\hat{a}_n \pm t_{\alpha/2}(I_n - 2)\sqrt{\hat{V}\text{ar}(\hat{a}^n)}$	$\hat{b}_n \pm t_{\alpha/2}(I_n - 2)\sqrt{\hat{V}\text{ar}(\hat{b}^n)}$
Multiplicative	n/a	$\hat{b}_n \pm t_{\alpha/2}(I - 1)\sqrt{\hat{V}\text{ar}(\hat{b}^n)}$

These formulas could be used, for example, to test the hypothesis that a_n is not significantly different from zero or that b_n is not significantly different from unity. If the first hypothesis were true, then a multiplicative model may be preferred over the more general linear model. The second test of hypothesis would give an objective means of selecting U.

Near the tail of the triangle, the degrees of freedom drop prohibitively. Inferences about the link ratios become less precise. If it can be assumed that beyond a certain age the variances of the residuals in the development model are identical (i.e., $\sigma_i^2 = \sigma_j^2$ for all i and j greater than some value), then a single estimate of that MSE can be obtained by solving for all link ratios simultaneously.[9]

Confidence Intervals Around Estimated Ultimate Loss

This section is motivated by the GAD model because all results are exact.[10] Under the transformed GAD model (and assuming identically distributed e_n's),

$$\ln(x_n) = \ln(b_n) + \ln(x_{n-1}) + \ln(e),$$

or

$$x_n' = b_n' + x'_{n-1} + e'.$$

[9]With a moderately-sized 5 x 5 triangle the two-tailed 90 percentile t-value is only 18% greater than the smallest possible 90 percentile t-value, namely the 90 percentile point on the standard normal curve. This can be especially important for the small triangles that consultants or companies underwriting new products are wont to see. For an example of this, see the case study in Appendix A.

[10]See Theorem 8 in Appendix C. The multiplicative chain ladder method makes the probability distribution of the error term of the compound process rather intractable. The logarithmic transformation turns the GAD compound multiplicative process into a compound additive process in which case regression theory yields exact results.

The point estimate of ultimate transformed loss for a single accident year is:

$$pred' = \hat{\mu}_C' = \hat{\mu}_U' = x_0' + \sum_{j=1}^{u} \hat{b}_j' \ .$$

An unbiased estimate of the Total Error = Parameter Error + Process Error of the (transformed) prediction is:

$$\hat{V}\text{ar}\,(pred') = \left(C + \sum_{j=1}^{U} \frac{1}{I_j} \right) \hat{\sigma}'^2 \ .$$

Therefore, assuming one only wants to limit the downside risk, a one-sided 100α% confidence interval for ultimate loss is:

$$\hat{\mu}_C' - t_\alpha(df)\sqrt{\hat{V}\text{ar}\,(pred')}$$

where df equals the number of data points in the multiple regression less the number of estimated link ratios, u. Finally, the corresponding 100α% confidence interval around the "untransformed" prediction of ultimate loss is:

$$e^{\hat{\mu}'_C - t_a(df)\sqrt{\hat{V}\text{ar}\,(pred')}} \ .$$

With this motivation, an approximate 100α% one-sided confidence interval around a recursive ultimate loss prediction using any of the models is:

$$pred - t_a\,(df)\sqrt{\hat{V}\text{ar}\,(pred)} \, ,$$

where df equals the total number of data points used in all link ratio estimates less the total number of estimated parameters. Two-sided confidence intervals are similarly defined, using $\pm\, t_{\alpha/2}\,(df)$. If df is large enough, $t_\alpha(df)$ may be replaced by z_α, the standard normal point, without significant loss of accuracy. This is often done in practice, particularly in time series analysis, even when *df* is not particularly large. The *t* distribution is preferred, however, because the thinner tails of the standard normal will understate the radius of the confidence interval. For another perspective on this subject, see Gardner [3].

Confidence Intervals Around Reserves

Confidence intervals around reserves are obtained by subtracting paid dollars from the endpoints of the confidence intervals around ultimate loss, because if:

$$\alpha = \text{Prob}\{\text{lower bound} \le \text{ultimate loss} \le \text{upper bound}\},$$

then as well,

$$\alpha = \text{Prob}\{\text{lower bound} - \text{paid} \le \text{outstanding loss} \le \text{upper bound} - \text{paid}\}.$$

5. AN ARGUMENT IN SUPPORT OF A NON-ZERO CONSTANT TERM

When the current evaluation is zero but the next evaluation is not expected to be, the loss development method is abandoned. Three alternatives might be Bornhuetter-Ferguson, Stanard-Bühlmann, or a variation on frequency-severity. LSL might be a fourth possibility.

Consider the development of reported claim counts. Let *exposure* be the true ultimate number of claims for a given accident year. Assume that the reporting pattern is the same for all claims. That is, if p_n is the probability that a claim is reported before the end of the n^{th} year, then the p_n's are independent and identically distributed for all claims. Based on these assumptions, it is not difficult to show that if x_n is the cumulative number of reported claims as of the n^{th} evaluation then

$$E(x_n \mid x_{n-1}) = exposure \frac{p_n - p_{n-1}}{1 - p_{n-1}} + \frac{1 - p_n}{1 - p_{n-1}} x_{n-1} \qquad (5.1)$$

which is of the form $a_n + b_n x_{n-1}$. Clearly the constant term a_n is nonzero until all claims are reported.

Equation 5.1 becomes surprisingly simple when the reporting pattern is exponential, as might be expected from a Poisson frequency process. In that case the LSL coefficients (a_n,b_n) are identical for every age n. This fact can be put to good use for small claim count triangles, as demonstrated in Appendix B.

The constant term a_n of Equation 5.1 is proportional to *exposure*. The slope factor b_n does not depend on *exposure* but only on the reporting pattern (the p's). Therefore, an increase in *exposure* from one accident year to the next will result in an upward, parallel shift in the development pattern. Claim count triangles, therefore, can be expected to display development samples randomly distributed about not a *single* regression line but *multiple parallel* regression lines.

Equation 5.1 may also be used as a paradigm for loss dollars, where trend may provide an upward force on *exposure*.

6. CONCLUSION

The traditional methods of calculating average development factors are the least squares estimators of an appropriately framed mathematical model. The conclusion is that link ratio averages are unbiased if the development process conforms to the specified model. If the independence assumption of the chain ladder method holds as well, the loss development method is unbiased.

A happy byproduct of the least squares perspective is that formulas for the variances of estimated ultimate loss and reserves drop right out. The formulas are particularly easy to apply if ultimate loss by accident year is estimated through an iterative procedure, rather than through a single, closed-form expression. Confidence intervals around ultimate loss and reserves can be estimated easily, although the suggested approach yields only approximate results (with a special case exception).

The simulation study in Appendix B suggests that, in some situations, the performance of the more general linear model may exceed that of the multiplicative models and may even rival that of the nonlinear Bornhuetter-Ferguson and Stanard-Bühlmann methods.

Some questions for further research come to mind. Can the formulas for parameter error be used in conjunction with the collective risk model? Is there a simple way to estimate the correlation between paid and incurred triangles, and how can that information be used to derive optimal, variance-minimizing weights for making final selections from the paid and incurred development estimates? Can the theory be used to find credibility formulas for averaging link ratios from small triangles with link ratios from larger triangles? Finally, can the Chain Ladder Independence Assumption be relaxed, to allow, say, for higher-than-expected development in one period to be followed by less-than-expected development the next?

REFERENCES

[1] A. M. Best Company, *Aggregates & Averages*, 1992 Edition.

[2] Butsic, Robert P., and Rafal J. Balcarek, "The Effect of Inflation on Losses and Premiums for Property-Liability Insurers," *Inflation Implications for Property-Casualty Insurance*, 1981 Casualty Actuarial Society Discussion Paper Program, pp. 58-102.

[3] Gardner, Everette S., Jr., "A Simple Method of Computing Prediction Intervals for Time Series Forecasts," *Management Science*, Vol. 34, No. 4, April 1988, pp. 541-546.

[4] Hogg, Robert V., and Allen T. Craig, *Introduction to Mathematical Statistics*, Fourth Edition, Macmillan Publishing Co., Inc., 1978.

[5] Hossack, I. B., John H. Pollard, and Ben Zehnwirth, *Introductory Statistics with Applications in General Insurance*, Cambridge University Press, 1983.

[6] Miller, Robert B., and Dean W. Wichern, *Intermediate Business Statistics*, Holt, Rinehart and Winston, 1977.

[7] Robertson, John P., Discussion of "A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques," by James N. Stanard, *PCAS* LXXII, 1985, pp. 149-153.

[8] Scheffé, Henry, *The Analysis of Variance*, Wiley, 1956, p. 195.

[9] Stanard, James N., "A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques," *PCAS* LXXII, 1985, pp. 124-148.

APPENDIX A

A CASE STUDY OF INDUSTRYWIDE WORKERS' COMPENSATION

The methods of this paper are applied to the consolidated industry workers' compensation incurred loss triangle as of December 31, 1991 [1]. The data and link ratios are displayed in Exhibits A-1 and A-2. Bulk plus IBNR reserves are removed from the incurred loss and ALAE triangles of Schedule P-Part 2. We will use the loss development method based on five-year weighted average (WAD) link ratios to estimate total ultimate loss for accident years 1982 through 1991. Then we will calculate the variance of that estimate, and use it to estimate the confidence level of industry reserves for those years.

Per the text, to estimate variances for the WAD method we must first transform the data by taking the square root of all "current evaluations" x, then dividing all "future evaluations" y by $\sqrt{x}$. We will model the data in two parts: 1) for the 12:24 month link ratios, and 2) for all other link ratios simultaneously. We shall see that there are justifiable statistical reasons for splitting the triangle this way. In addition it helps demonstrate the methodology.

Exhibit A-3 runs the regression for the 12:24 month link ratios. The original data evaluated as of 12 and 24 months for the five most recent accident years—1986 through 1990—are shown, as well as the transformed data. Using a popular spreadsheet package, the regression was run on the transformed data. The regression output indicates a good fit ($R^2 = 95\%$). Note that the "x coefficient" agrees with the average link ratio in Exhibit A-2; the variance of that estimated parameter is $0.01487^2 = 0.00022$. The MSE is 13.6272, which drives not only the variance of that estimated link ratio parameter but also the process error in the development of losses from age 12 to age 24.

For setting up the multiple regression solution of the remaining link ratios—24:36 months through 108:120 months—refer to Exhibit A-4. We first build the y vector by stacking the "next" evaluations of those link ratios on top of each other. Then we create the x matrix by placing the "current" evaluation in the same row as the corresponding y value. For each successive age of development, the x values are placed in successive columns. The transformed data are shown in Exhibit A-5, and the regression output is shown in Exhibit A-6. The R^2 value is extremely high. The MSE is much lower (0.3545) than it was for 12:24 development, which suggests that it was indeed prudent to split up the triangle into two regressions. Again, note that the x coefficients correspond to the original five-year weighted averages in Exhibit A-2.

These parameters and variances are almost all that is needed to complete the triangle in Exhibit A-8. In fact, these factors will square the triangle to 120 months, but not to ultimate. Since Part 2 of Schedule P does not include a tail factor, we will estimate a tail from Part 1 as follows.

For the five oldest accident years, we will compare developed 120-month losses (actuals for accident year 1982) with ultimate losses per industry estimates as reported in Schedule P-Part 1. Under the assumption that industry ultimate losses for those relatively mature years are reasonably accurate, we will use the weighted average of that ratio as the 120:ultimate tail factor. This weighted average is subject to random variation, so we will use the techniques of the paper to estimate the MSE and variance of that tail factor estimate. This is done in Exhibit A-7.

Exhibit A-8 shows the completed triangle, followed by the variance calculations using the formulas of Tables 3.4 and 3.5. For example, the Table 2.2 recursive formula calculates the 48-month future value M_3 of accident year 1989 through 1991 losses in total as $72{,}731 = (47{,}611+21{,}624) \times 1.05051$. The Table 3.4 recursive formula calculates the Parameter Risk of that estimate as:

$$103{,}328 = (47{,}611 + 21{,}624)^2 \times 0.00216^2 + 1.05051^2 \times 73{,}370 + 0.00216^2 \times 73{,}370.$$

The Table 3.5 formula calculates the Process Risk of the projection as:

$$323{,}963 = (47{,}611 + 21{,}624) \times 0.3545 + 1.05051^2 \times 271{,}435.$$

The estimates of ultimate loss using this procedure are compared with the consolidated industry estimates in Exhibit A-9. Total projected ultimate loss and ALAE using the five year weighted averages of the link ratios, and the tail factor as estimated above, is (in millions) $191,509. The industry carried ultimate is $188,251, or about 1.7% less than indicated, a seemingly small difference. However, the standard deviation of the projection is only $1,840. So the carried ultimate is about 1.77 standard deviations less than the projection. Therefore, using the Student *t* distribution with 30 degrees of freedom,[11] the estimated one-sided confidence level for industry reserves is about 4%.

[11] Add up the *df*'s in Exhibits A-3, A-6, and A-7.

EXHIBIT A-1

CONSOLIDATED INDUSTRY WORKERS' COMPENSATION
REPORTED INCURRED LOSSES AND ALLOCATED EXPENSES BY AGE
(EXCLUDING BULK + IBNR)
($ 000,000 OMITTED)

Accident	Age									
Year	12	24	36	48	60	72	84	96	108	120
1982	6,174	8,061	8,639	8,951	9,207	9,363	9,464	9,559	9,634	9,725
1983	6,891	9,117	9,682	10,136	10,464	10,651	10,774	10,893	11,025	
1984	8,048	10,761	11,937	12,656	13,023	13,285	13,449	13,615		
1985	8,796	12,050	13,287	14,060	14,572	14,835	15,109			
1986	9,450	13,086	14,552	15,334	15,797	16,144				
1987	10,953	15,074	16,699	17,485	17,961					
1988	12,776	17,600	19,519	20,299						
1989	13,600	19,677	21,624							
1990	14,890	21,268								
1991	15,497									

Source: Best's *Aggregates & Averages*, 1992 Edition.

EXHIBIT A-2
CONSOLIDATED INDUSTRY WORKERS' COMPENSATION LINK RATIOS

Accident	Development Period (Months)								
Year	12:24	24:36	36:48	48:60	60:72	72:84	84:96	96:108	108:120
1982	1.30566	1.07167	1.03614	1.02859	1.01694	1.01086	1.00995	1.00785	1.00949
1983	1.32298	1.06201	1.04683	1.03238	1.01788	1.01153	1.01108	1.01214	
1984	1.33712	1.10933	1.06023	1.02896	1.02013	1.01240	1.01234		
1985	1.36995	1.10269	1.05812	1.03641	1.01807	1.01851			
1986	1.38472	1.11204	1.05372	1.03020	1.02194				
1987	1.37619	1.10786	1.04703	1.02722					
1988	1.37757	1.10906	1.03996						
1989	1.44687	1.09892							
1990	1.42837								
Five Year Weighted Average									
	1.40597	1.10576	1.05051	1.03080	1.01927	1.01379	1.01127	1.01014	1.00949

EXHIBIT A-3

ESTIMATING THE 12:24 MONTH PARAMETER USING REGRESSION

Accident	y	x	$y/\sqrt{x}$	$\sqrt{x}$
Year	24 months	12 months	24 months	12 months
1986	13,086	9,450	134.61	97.21
1987	15,074	10,953	144.03	104.66
1988	17,600	12,776	155.71	113.03
1989	19,677	13,600	168.73	116.62
1990	21,268	14,890	174.30	122.02

Regression Output:

Constant		0	
Std Err of y Est		3.6915	MSE = 13.6272
R Squared		95.03%	
Number of Observations		5	
Degrees of Freedom		4	
x Coefficient	1.40597		
Std Err of Coef.	0.01487		

EXHIBIT A-4

ESTIMATING THE 24:36 THROUGH 108:120 MONTH PARAMETERS USING REGRESSION
STEP 1: ARRANGING THE DATA BEFORE TRANSFORMATION ($ 000,000 OMITTED)

Accident		x							
Year	y	24 mos	36 mos	48 mos	60 mos	72 mos	84 mos	96 mos	108 mos
1985	13,287	12,050							
1986	14,552	13,086							
1987	16,699	15,074							
1988	19,519	17,600							
1989	21,624	19,677							
1984	12,656		11,937						
1985	14,060		13,287						
1986	15,334		14,552						
1987	17,485		16,699						
1988	20,299		19,519						
1983	10,464			10,136					
1984	13,023			12,656					
1985	14,572			14,060					
1986	15,797			15,334					
1987	17,961			17,485					
1982	9,363				9,207				
1983	10,651				10,464				
1984	13,285				13,023				
1985	14,835				14,572				
1986	16,144				15,797				
1982	9,464					9,363			
1983	10,774					10,651			
1984	13,449					13,285			
1985	15,109					14,835			
1982	9,559						9,464		
1983	10,893						10,774		
1984	13,615						13,449		
1982	9,634							9,559	
1983	11,025							10,893	
1982	9,725								9,634

EXHIBIT A-5
ESTIMATING THE 24:36 TO 108:120 MONTH PARAMETERS USING REGRESSION
STEP 2: TRANSFORMING THE DATA FOR THE REGRESSION

Accident		$\sqrt{x}$							
Year	$y/\sqrt{x}$	24 mos	36 mos	48 mos	60 mos	72 mos	84 mos	96 mos	108 mos
1985	121.04	109.77							
1986	127.21	114.39							
1987	136.02	122.77							
1988	147.13	132.66							
1989	154.15	140.27							
1984	115.84		109.26						
1985	121.97		115.27						
1986	127.11		120.63						
1987	135.30		129.23						
1988	145.29		139.71						
1983	103.94			100.68					
1984	115.76			112.50					
1985	122.89			118.57					
1986	127.57			123.83					
1987	135.83			132.23					
1982	97.58				95.95				
1983	104.12				102.29				
1984	116.41				114.12				
1985	122.89				120.71				
1986	128.44				125.69				
1982	97.81					96.76			
1983	104.39					103.20			
1984	116.69					115.26			
1985	124.05					121.80			
1982	98.25						97.28		
1983	104.95						103.80		
1984	117.40						115.97		
1982	98.54							97.77	
1983	105.64							104.37	
1982	99.08								98.15

EXHIBIT A-6
ESTIMATING THE 24:36 TO 108:120 MONTH PARAMETERS USING REGRESSION
STEP 3: RUNNING THE REGRESSION

Regression Output:		
Constant	0	
Std Err of y Est	0.5954	MSE =0.3545
R Squared	99.9%	
Number of Observations	30	
Degrees of Freedom	22	

	24:36	36:48	48:60	60:72	72:84	84:96	96:108	108:120
x Coefficient	1.10576	1.05051	1.03080	1.01927	1.01379	1.01127	1.01014	1.00949
Std Err of Coef.	0.00214	0.00216	0.00226	0.00237	0.00271	0.00324	0.00416	0.00607

EXHIBIT A-7

ESTIMATING THE TAIL FACTOR USING REGRESSION

Accident Year	Developed Losses to Age 120 (y)	Carried Ultimate (x)	Tail Factor
1982	9,725	9,966	1.02482
1983	11,130	11,355	1.02019
1984	13,884	14,081	1.01422
1985	15,581	15,720	1.00889
1986	16,877	17,141	1.01561
Wtd Avg	67,197	68,263	1.01586

Regression Matrix

Accident Year	$y/\sqrt{x}$	$\sqrt{x}$
1982	101.06	98.615
1983	107.63	105.500
1984	119.51	117.830
1985	125.93	124.820
1986	131.94	129.910

Regression Output:

Constant	0	
Std Err of y Est	0.6680	MSE = 0.4462
R Squared	99.7%	
Number of Observations	5	
Degrees of Freedom	4	

x Coefficient(s)	1.01586
Std Err of Coef.	0.00258

EXHIBIT A-8
CONSOLIDATED INDUSTRY WORKERS' COMPENSATION
COMPLETED LOSS DEVELOPMENT TRIANGLE
($ 000,000 OMITTED)

Accident	Age										
Year	12	24	36	48	60	72	84	96	108	120	Ultimate
1982	6,174	8,061	8,639	8,951	9,207	9,363	9,464	9,559	9,634	9,725	9,879
1983	6,891	9,117	9,682	10,136	10,464	10,651	10,774	10,893	11,025	11,130	11,307
1984	8,048	10,761	11,937	12,656	13,023	13,285	13,449	13,615	13,753	13,884	14,104
1985	8,796	12,050	13,287	14,060	14,572	14,835	15,109	15,280	15,435	15,581	15,828
1986	9,450	13,086	14,552	15,334	15,797	16,144	16,366	16,551	16,719	16,877	17,145
1987	10,953	15,074	16,699	17,485	17,961	18,307	18,559	18,768	18,959	19,138	19,442
1988	12,776	17,600	19,519	20,299	20,924	21,328	21,622	21,865	22,087	22,296	22,650
1989	13,600	19,677	21,624	22,716	23,415	23,866	24,196	24,468	24,716	24,951	25,346
1990	14,890	21,268	23,518	24,706	25,467	25,957	26,315	26,612	26,881	27,136	27,567
1991	15,497	21,789	24,093	25,310	26,089	26,592	26,959	27,263	27,539	27,800	28,241

n	1	2	3	4	5	6	7	8	9	10
M_n	21,789	47,611	72,731	95,896	116,050	134,017	150,806	166,088	178,794	191,509
Parameter Risk	53,070	73,370	103,328	153,825	232,678	367,838	610,182	1,091,197	2,266,302	2,574,752
Process Risk	211,184	271,435	323,963	377,671	433,552	493,096	557,499	627,671	703,340	810,521
Total Risk	264,254	344,805	427,291	531,496	666,231	860,934	1,167,681	1,718,868	2,969,642	3,385,272
Standard Deviation	514	587	654	729	816	928	1,081	1,311	1,723	1,840

EXHIBIT A-9

CONSOLIDATED INDUSTRY WORKERS' COMPENSATION
ESTIMATED REDUNDANCY/(DEFICIENCY) IN CARRIED RESERVES AND ASSOCIATED LEVEL OF CONFIDENCE
ACCIDENT YEARS 1982-1991
($ 000,000 OMITTED)

Accident Year	Estimated Ultimate	Carried Ultimate	Redundancy/ (Deficiency)
1982	9,879	9,966	87
1983	11,307	11,355	48
1984	14,104	14,081	(23)
1985	15,828	15,720	(109)
1986	17,145	17,141	(4)
1987	19,442	19,304	(138)
1988	22,650	22,217	(433)
1989	25,346	24,645	(702)
1990	27,567	26,710	(856)
1991	28,241	27,112	(1,129)
Total	191,509	188,251	(3,258)

Standard Deviation	1,840
Degrees of Freedom	30
Deficiency Ratio to Standard Deviation	-1.77
Approximate Confidence Level	4%

APPENDIX B

COMPARING THE MODELS USING SIMULATION

In the 1985 *Proceedings*, Mr. James Stanard published the results of a simulation study of the accuracy of four simple methods of estimating ultimate losses using a 5x5 incurred loss triangle. For the exposure tested[12] it was demonstrated that WAD loss development was clearly inferior to three additive methods, Bornhuetter-Ferguson (BF), Stanard-Bühlmann (SB)[13], and a little-used method called the Additive Model (ADD), because it had greater average bias and a larger variance. The three additive methods differ from the multiplicative methods in that they adjust incurred losses to date by an estimated dollar increase to reach ultimate, whereas the multiplicative methods adjust by an estimated percentage increase. ADD's estimated increase is a straightforward calculation of differences in column means, $\bar{y}-\bar{x}$. BF and SB estimated increases are more complicated functions of the data.

Stanard's simulation was replicated here to test additionally the accuracy of LSM, LSL, SAD, and GAD. The model does not attempt to predict "beyond the triangle," which is to say that the methods project incurred losses to the most mature age available in the triangle, namely the age of the first accident year. In the discussion below, "ultimate loss" refers to case incurred loss as of the most mature available age.

The LSL method was modified to use LSM in those instances when the development factors were "obviously wrong," defined to be

---

[12]Normally distributed frequency with mean = 40 and standard deviation $=\sqrt{40}$ claims per year, uniform occurrence date during the year, lognormal severity with mean = \$10,400 and standard deviation = \$34,800, exponential report lag with mean = 18 months, exponential payment lag with mean = 12 months, and case reserve error proportional to a random factor equal to a lognormal random variable with mean = 1 and variance = 2, and to a systematic factor equal to the impact of trend between the date the reserve is set and the date the claim is paid.

[13]Mr. Stanard called this the "Adjustment to Total Known Losses" method, a.k.a. the "Cape Cod Method."

when either the slope or the constant term was negative. In real-life situations, this rudimentary adjustment for outliers can be expected to be improved upon with more discerning application of actuarial judgment. The reason this modification was necessary is due to the fact that a model that fits data well does not necessarily predict very well. As an extreme example, LSL provides an exact fit to the sample data for the penultimate link ratio (two equations, two unknowns), but the coefficients so determined reveal nothing about the random processes that might cause another accident year to behave differently. It is not possible to identify every conceivable factor that could explain the otherwise "unexplained" variance of a model. Such unidentified variables are reflected through the averaging process of statistical analysis: as the number of data points minus the number of parameters (the definition of degrees of freedom) increases, the model captures more of the unexplained factors and becomes a better predictor.

In Exhibits B-1 through B-4, the average bias and standard deviation of the first accident year are zero because, as stated above, the simulation defines "ultimate" to be the current age of that accident year.

Exhibit B-1: Claim Counts Only

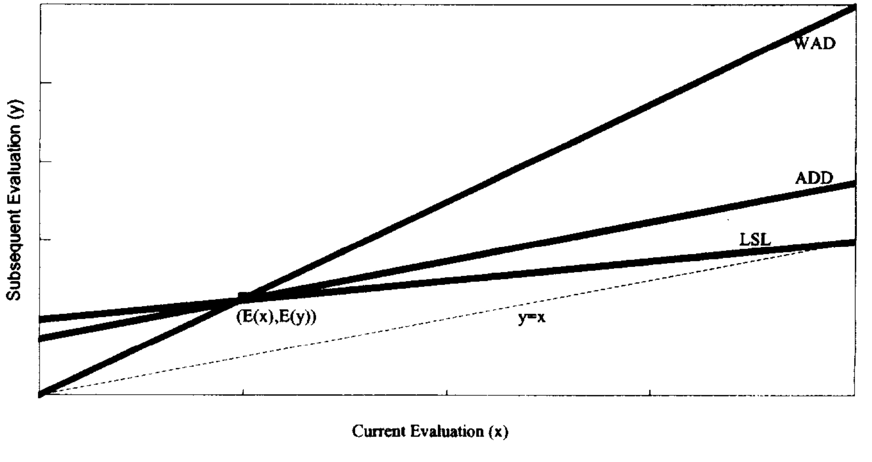
In this case, 5,000 claim count triangles were simulated; the "actual ultimate" as of the last column was simulated; accident year ultimates were predicted using the separate methods; and averages and standard deviations of the prediction errors were calculated.

LSL is the best performer, as measured by the standard deviation of the accident-year-total projection. The additive models—ADD, SB, and BF—are not far behind. Of the multiplicative estimators, LSM has the smallest bias and the smallest variance for every accident year. WAD is almost as accurate.

Why should these results not be surprising? Consider first the average bias. In Figure B-1 is graphed the relationship between incurred counts at 12 months, x, with incurred losses at 24 months, y, which we know from Section 5 of the text must be a linear relationship with a positive constant term. The ADD and WAD estimates are also shown. All relationships are shown in their idealized states where LSL is collinear with the true relationship and where the point $(\bar{x}, \bar{y})$ coincides with its expectation (E (x), E (y)). Note that the ADD model is parallel to the line $y = x$ because it adds the same amount for every value of x. The conditional (on x) bias is the signed, vertical distance from the estimated relationship to the true relationship. As is clear from Figure B-1, WAD and ADD can be expected to overstate y for $x >$ E (x) and understate y for $x <$ E (x). The weighted average of the conditional bias across all values of x, weighted by the probability density $f(x)$, is simulated by the average bias that appears in Exhibit B-1.

Ideally, this weighted average of the bias across all values of x should be expected to be zero, which it is for the Additive Model. ADD estimates E (y) - E (x) using $\bar{y} - \bar{x}$ calculated from prior acci-

FIGURE B-1
IDEALIZED DEVELOPMENT ESTIMATORS
NO TREND



dent years. Since the environment in the first scenario—exposure, frequency, trend, etc.—does not change by accident year, the average of 5,000 simulated samples of this dollar difference across all possible values of x should get close to the true average dollar difference by the law of large numbers, so the average bias should get close to zero. For the multiplicative estimators, the average bias will probably not be zero. Take the WAD method for example. Clearly there is a positive probability (albeit small) that $\bar{x} = 0$, so the expected value of the WAD link ratio $\bar{y} / \bar{x}$ is infinity. The average of 5,000 simulations of this ratio attempts to estimate that infinite expected value, so it should not be surprising that WAD usually overstates development—and the greater the probability that $\bar{x} = 0$. the greater the overstatement.[14]

The average bias of the BF and SB methods should be greater than zero as well because the LDFs on which they rely are themselves overstated more often than not. The average LSM bias is a more complicated function of the probability distribution of x because the LSM link ratio involves x terms in the numerator and squared x terms in the denominator. The average bias appears to shift as an accident year matures. The LSL method as modified herein has residual average bias because it incorporates the biased LSM method when it detects outliers. It also seems to be the case that the bias of the estimated 4:5 year link ratio is driving the cumulative bias for the immature years.

Figure B-1 illustrates the difference between a model that is unbiased for each possible value of x, LSL, and a model which is "unbiased" only in the average, ADD. To reiterate, the purely multiplicative and purely additive estimators will understate expected development when the current evaluation is less than expected and overstate expected development when the current evaluation is greater than expected.

[14]This argument can be made more rigorous. The condition that the probability of the sample average of x be greater than zero is a sufficient but not necessary condition that $E(b_{WAD}) = \infty$. For a general, heuristic argument that WAD yields biased estimates, see Stanard [8].

Next, consider the variance. In simplified terms, the average bias statistic allows expected overstatements to cancel out expected understatements. This is not the case for the variance statistic. In Figure B-1 it is clear that, ideally, the ADD estimate of y will be closer to the true conditional expected value of y (the idealized LSL line) than will the WAD estimate for virtually all values of x. Thus, the variance of ADD should be less than the variance of WAD. The variance of LSL should be the smallest of all. However, LSL estimates twice as many parameters than do ADD and LSM, so it needs a larger sample size to do a comparable job. For the relatively small and thin triangles simulated here, a pure unmodified LSL estimate flops around like a fish out of water—the price it must pay to be unbiased for all values of x. In other words, in actual practice, the variance of an LSL method unmodified for outliers and applied to a triangle with few degrees of freedom will probably be horrendous. What is perhaps remarkable is the degree to which the rudimentary adjustment adopted here tames the LSL method.

Finally, let's look at what would happen if we estimated the LSL parameters under the assumption that all link ratio coefficients (a_n, b_n) are equal. We know from the previous section that this is true because the reporting pattern is exponential. The results of this model are:

SIMULATION RESULTS WHEN
ALL LINK RATIO PARAMETERS ARE ASSUMED EQUAL

A/Y	Average Bias	Std Dev Bias	Average % Bias	Std Dev % Bias	Age-Age Bias	Age-Age % Bias
1	0.000	0.000	0.000	0.000		
2	0.025	1.275	0.001	0.034	1.035	1.001
3	0.006	1.669	0.001	0.044	(0.019)	0.000
4	(0.034)	1.850	0.000	0.049	(0.040)	(0.001)
5	(0.006)	1.815	0.001	0.049	0.028	0.001
Total	(0.010)	5.064	0.000	0.027		

This model is the beneficiary of more degrees of freedom (eight—two parameters estimated from ten data points for each iteration) and as a result has the smallest average bias and variance yet. These results lead to a somewhat counter-intuitive conclusion: Information about development across immature ages sheds light on future development across mature ages. For example, the immature development just experienced by the young accident year 4 from age 1 to age 2 is a valuable data point in the estimate of the upcoming development of the old accident year 2 from age 4 to age 5. This should not be viewed simply as a bit of mathematical prestidigitation but as an example of the efficiencies that can be achieved if simplifying assumptions—even as innocuous as exponential reporting—can be justified.

Exhibit B-2: Random Severity, No Trend

In this case, 5,000 triangles of aggregate, trend-free incurred losses were simulated and the same calculations were performed.

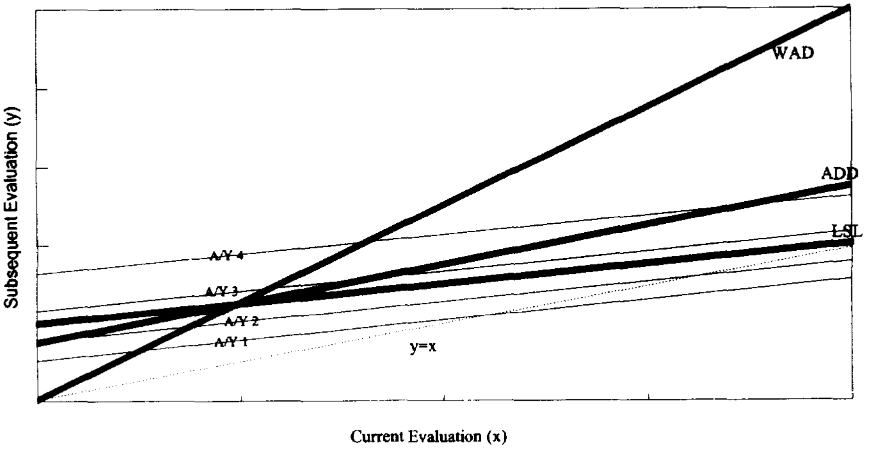
Rarely does the property/casualty actuary experience loss triangles devoid of trend, so this model is of limited interest. The introduction of uncertainty via the case reserves makes it more likely that negative development will appear, in which case LSL reverts to LSM. As a result, the additive models overtake LSL in accuracy.

Exhibit B-3: Random Severity, 8% Severity Trend Per Year

This is where it gets interesting. This could be considered the typical situation in which an actuary compiles a loss triangle that includes trend and calculates loss development factors. In this case, the environment is changing. The trending process follows the Unified Inflation Model (Butsic and Balcarek, [2]) with $\alpha = ½$, which is to say that half of the impact of inflation is a function of the occurrence date and half is a function of the transaction date (e.g., evaluating the case incurred or paying the claim).

At first, one might think that a multiplicative estimator would have had a better chance of catching the trend than would an additive estimator, but such does not appear to be the case. Consider Figure B-2 which graphs expected 12-24 month development for the first four accident years. Trend has pushed the true development line upward at an 8% clip, illustrated by four thin lines. The LSL model tries to estimate the average of the development lines, the WAD estimator tries to pass through the average ($\bar{x}, \bar{y}$) midpoint of all accident years combined, and the additive estimators try to find the line parallel to the line y=x which also passes through the average midpoint. Again, ADD will probably be closer than WAD to the average LSL line for every value of x for each accident year. The upward trend makes it more likely that the estimated LSL intercept will be less than zero,

FIGURE B-2
IDEALIZED DEVELOPMENT ESTIMATORS
WITH TREND



which makes it more likely that LSL reverts to LSM, so the modified LSL's variance gets closer yet to the variance of LSM.

Exhibit B-4: Random Severity, 8% Trend, On-Level Triangle

In this case, rows of the triangle were trended to the level of the most recent accident year assuming that the research department is perfect in its estimate of past trend. For most of the models, the total bias decreases from that of the not-on-level scenario while the total variance increases. LSM and WAD are virtually unchanged, GAD and SAD are exactly unchanged (of course), and the nonlinear estimates move in opposite directions.

For the most part, working with the on-level triangle does seem to improve the accuracy of estimated ultimate loss, but perhaps not to the degree one might hope. It would be interesting to see if working with separate claim count and on-level severity triangles would successfully decompose the random effects and further improve the predictions.

EXHIBIT B-1
Part 1

CLAIM COUNTS ONLY

	A/Y	Average Bias	Std Dev Bias	Average % Bias	Std Dev % Bias	Age-Age Bias	Age-Age % Bias
LSL							
	1	0.000	0.000	0.000	0.000		
	2	0.116	2.000	0.003	0.053	0.116	0.003
	3	0.153	2.772	0.004	0.073	0.037	0.001
	4	0.101	3.166	0.003	0.083	(0.052)	(0.001)
	5	0.080	3.780	0.003	0.100	(0.021)	0.000
	Total	0.451	8.251	0.002	0.043		
ADD							
	1	0.000	0.000	0.000	0.000		
	2	0.059	1.868	0.002	0.049	0.059	0.002
	3	0.075	2.847	0.002	0.075	0.016	0.000
	4	0.047	3.644	0.002	0.096	(0.028)	0.000
	5	0.096	3.692	0.003	0.097	0.049	0.001
	Total	0.277	8.407	0.001	0.044		
LSM							
	1	0.000	0.000	0.000	0.000		
	2	0.116	2.000	0.003	0.053	0.116	0.003
	3	0.143	3.321	0.004	0.087	0.027	0.001
	4	0.004	5.246	0.000	0.138	(0.139)	(0.004)
	5	(0.748)	10.536	(0.020)	0.277	(0.752)	(0.020)
	Total	(0.485)	14.009	(0.003)	0.074		
WAD							
	1	0.000	0.000	0.000	0.000		
	2	0.116	2.000	0.003	0.053	0.116	0.003
	3	0.203	3.336	0.005	0.088	0.087	0.002
	4	0.281	5.308	0.007	0.139	0.078	0.002
	5	0.888	11.101	0.023	0.292	0.607	0.016
	Total	1.488	14.520	0.008	0.076		

EXHIBIT B-1
Part 2

CLAIM COUNTS ONLY

	A/Y	Average Bias	Std Dev Bias	Average % Bias	Std Dev % Bias	Age-Age Bias	Age-Age % Bias
GAD							
	1	0.000	0.000	0.000	0.000		
	2	0.116	2.000	0.003	0.053	0.116	0.003
	3	0.234	3.345	0.006	0.088	0.118	0.003
	4	0.424	5.346	0.011	0.140	0.190	0.005
	5	1.873	11.585	0.049	0.305	1.449	0.038
	Total	2.647	14.943	0.014	0.079		
SAD							
	1	0.000	0.000	0.000	0.000		
	2	0.116	2.000	0.003	0.053	0.116	0.003
	3	0.265	3.354	0.007	0.088	0.149	0.004
	4	0.571	5.390	0.015	0.142	0.306	0.008
	5	2.958	12.268	0.078	0.322	2.387	0.062
	Total	3.910	15.530	0.021	0.082		
SB							
	1	0.000	0.000	0.000	0.000		
	2	0.102	1.940	0.003	0.051	0.102	0.003
	3	0.147	3.021	0.004	0.079	0.045	0.001
	4	0.137	3.997	0.004	0.105	(0.010)	0.000
	5	0.185	4.280	0.006	0.113	0.048	0.002
	Total	0.571	9.564	0.003	0.050		
BF							
	1	0.000	0.000	0.000	0.000		
	2	0.114	1.952	0.003	0.051	0.114	0.003
	3	0.184	3.064	0.005	0.081	0.070	0.002
	4	0.215	4.151	0.006	0.109	0.031	0.001
	5	0.338	5.164	0.010	0.136	0.123	0.004
	Total	0.851	10.626	0.004	0.056		

EXHIBIT B-2
Part 1

RANDOM SEVERITY, NO TREND

	A/Y	Average Bias	Std Dev Bias	Average % Bias	Std Dev % Bias	Age-Age Bias	Age-Age % Bias
LSL							
	1	0	0	0.000	0.000		
	2	9,206	193,945	0.026	0.302	9,206	0.026
	3	8,749	218,463	0.069	0.420	(458)	0.042
	4	30,028	429,112	0.138	0.650	21,279	0.065
	5	39,426	535,959	0.228	1.004	9,398	0.079
	Total	87,410	888,404	0.040	0.356		
LSM							
	1	0	0	0.000	0.000		
	2	9,206	193,945	0.026	0.302	9,206	0.026
	3	6,192	221,114	0.033	0.415	(3,015)	0.007
	4	24,331	477,371	0.052	0.742	18,140	0.018
	5	12,290	825,131	0.036	1,404	(12,042)	(0.015)
	Total	52,019	1,127,243	0.020	0.453		
WAD							
	1	0	0	0.000	0.000		
	2	9,206	193,945	0.026	0.302	9,206	0.026
	3	11,815	222,675	0.048	0.421	2,608	0.021
	4	51,641	515,997	0.119	0.807	39,826	0.068
	5	116,664	894,747	0.310	1.597	65,023	0.171
	Total	189,327	1,208,220	0.088	0.487		
GAD							
	1	0	0	0.000	0.000		
	2	9,206	193,945	0.026	0.302	9,206	0.026
	3	13,873	219,115	0.054	0.412	4,666	0.027
	4	61,706	484,892	0.147	0.763	47,833	0.088
	5	184,903	854,318	0.489	1.593	123,197	0.298
	Total	269,687	1,130,473	0.130	0.469		

EXHIBIT B-2
Part 2

RANDOM SEVERITY, NO TREND

	A/Y	Average Bias	Std Dev Bias	Average % Bias	Std Dev % Bias	Age-Age Bias	Age-Age % Bias
SAD							
	1	0	0	0.000	0.000		
	2	9,206	193,945	0.026	0.302	9,206	0.026
	3	20,621	227,597	0.072	0.440	11,415	0.045
	4	97,144	598,072	0.233	0.980	76,523	0.150
	5	405,202	1,241,904	1.063	2.516	308,058	0.673
	Total	532,174	1,552,136	0.255	0.640		
ADD							
	1	0	0	0.000	0.000		
	2	158	185,077	0.010	0.329	158	0.010
	3	(7,445)	196,201	0.023	0.472	(7,603)	0.013
	4	324	272,189	0.066	0.581	7,769	0.042
	5	(2,668)	271,443	0.140	0.680	(2,991)	0.069
	Total	(9,631)	596,942	(0.004)	0.255		
SB							
	1	0	0	0.000	0.000		
	2	6,126	184,062	0.026	0.304	6,126	0.026
	3	3,909	196,494	0.052	0.430	(2,217)	0.025
	4	15,414	291,195	0.097	0.575	11,506	0.043
	5	11,071	286,813	0.172	0.698	(4,344)	0.068
	Total	36,520	633,658	0.017	0.271		
BF							
	1	0	0	0.000	0.000		
	2	9,040	200,965	0.034	0.373	9,040	0.034
	3	10,750	221,175	0.073	0.525	1,710	0.038
	4	29,330	331,648	0.132	0.691	18,580	0.055
	5	37,124	374,743	0.225	0.886	7,794	0.082
	Total	86,244	820,177	0.040	0.342		

EXHIBIT B-3
Part 1

RANDOM SEVERITY, 8% TREND

	A/Y	Average Bias	Std Dev Bias	Average % Bias	Std Dev % Bias	Age-Age Bias	Age-Age % Bias
LSL							
	1	0	0	0.000	0.000		
	2	12,848	190,771	0.030	0.300	12,848	0.030
	3	11,815	318,796	0.061	0.469	(1,034)	0.030
	4	8,339	515,561	0.080	0.629	(3,475)	0.018
	5	(23,573)	731,012	0.075	0.944	(31,912)	(0.005)
	Total	9,430	1,181,752	0.002	0.367		
LSM							
	1	0	0	0.000	0.000		
	2	12,848	190,771	0.030	0.300	12,848	0.030
	3	16,307	328,599	0.043	0.475	3,458	0.013
	4	27,133	580,424	0.057	0.728	10,826	0.013
	5	8,411	1,111,762	0.035	1.360	(18,722)	(0.021)
	Total	64,698	1,504,280	0.021	0.472		
WAD							
	1	0	0	0.000	0.000		
	2	12,848	190,771	0.030	0.300	12,848	0.030
	3	23,423	333,524	0.057	0.477	10,575	0.026
	4	62,726	608,272	0.122	0.775	39,303	0.061
	5	169,257	1,272,791	0.310	1.620	106,531	0.168
	Total	268,255	1,659,744	0.098	0.527		
GAD							
	1	0	0	0.000	0.000		
	2	12,848	190,771	0.030	0.300	12,848	0.030
	3	26,050	331,370	0.062	0.466	13,201	0.031
	4	77,169	580,779	0.149	0.755	51,119	0.082
	5	277,757	1,295,202	0.495	1.717	200,588	0.301
	Total	393,824	1,619,314	0.148	0.534		

EXHIBIT B-3
Part 2

RANDOM SEVERITY, 8% TREND

	A/Y	Average Bias	Std Dev Bias	Average % Bias	Std Dev % Bias	Age-Age Bias	Age-Age % Bias
SAD							
	1	0	0	0.000	0.000		
	2	12,848	190,771	0.030	0.300	12,848	0.030
	3	35,174	346,105	0.080	0.497	22,326	0.049
	4	124,456	685,305	0.235	0.924	89,282	0.144
	5	647,473	4,098,366	1.107	4.508	523,017	0.706
	Total	819,951	4,291,335	0.299	1.164		
ADD							
	1	0	0	0.000	0.000		
	2	(2,249)	177,229	0.008	0.337	(2,249)	0.008
	3	(15,161)	262,260	0.009	0.461	(12,912)	0.001
	4	(35,576)	335,003	0.005	0.511	(20,414)	(0.004)
	5	(92,221)	399,076	(0.028)	0.551	(56,645)	(0.033)
	Total	(145,207)	757,285	(0.053)	0.249		
SB							
	1	0	0	0.000	0.000		
	2	10,229	177,339	0.036	0.323	10,229	0.036
	3	7,628	272,101	0.055	0.456	(2,601)	0.018
	4	(5,009)	357,093	0.057	0.530	(12,637)	0.002
	5	(62,946)	420,117	0.021	0.590	(57,936)	(0.034)
	Total	(50,098)	825,565	(0.018)	0.269		
BF							
	1	0	0	0.000	0.000		
	2	16,575	212,872	0.052	0.421	16,575	0.052
	3	23,046	310,265	0.091	0.589	6,471	0.037
	4	25,574	422,741	0.114	0.668	2,529	0.021
	5	(9,528)	534,249	0.101	0.780	(35,103)	(0.012)
	Total	55,667	1,113,743	0.020	0.357		

EXHIBIT B-4
Part 1

RANDOM SEVERITY, 8% TREND, ESTIMATES BASED ON ON-LEVEL (AT 8%) TRIANGLE

	A/Y	Average Bias	Std Dev Bias	Average % Bias	Std Dev % Bias	Age-Age Bias	Age-Age % Bias
LSL							
	1	0	0	0.000	0.000		
	2	12,848	190,771	0.030	0.300	12,848	0.030
	3	19,663	321,503	0.080	0.479	6,815	0.049
	4	38,827	508,047	0.147	0.637	19,164	0.062
	5	44,325	695,596	0.216	0.928	5,498	0.060
	Total	115,663	1,148,516	0.045	0.357		
LSM							
	1	0	0	0.000	0.000		
	2	12,848	190,771	0.030	0.300	12,848	0.030
	3	16,069	326,583	0.043	0.473	3,220	0.013
	4	26,536	577,658	0.055	0.725	10,467	0.012
	5	3,262	1,070,100	0.027	1.316	(23,274)	(0.027)
	Total	58,715	1,459,667	0.019	0.460		
WAD							
	1	0	0	0.000	0.000		
	2	12,848	190,771	0.030	0.300	12,848	0.030
	3	23,310	332,453	0.057	0.476	10,461	0.026
	4	62,521	607,521	0.121	0.774	39,211	0.061
	5	166,470	1,251,178	0.305	1.598	103,950	0.164
	Total	265,149	1,635,365	0.097	0.520		
GAD							
	1	0	0	0.000	0.000		
	2	12,848	190,771	0.030	0.300	12,848	0.030
	3	26,050	331,370	0.062	0.466	13,201	0.031
	4	77,169	580,779	0.149	0.755	51,119	0.082
	5	277,757	1,295,202	0.495	1.717	200,588	0.301
	Total	393,824	1,619,314	0.148	0.534		

EXHIBIT B-4
Part 2

RANDOM SEVERITY, 8% TREND, ESTIMATES BASED ON ON-LEVEL (AT 8%) TRIANGLE

	A/Y	Average Bias	Std Dev Bias	Average % Bias	Std Dev % Bias	Age-Age Bias	Age-Age % Bias
SAD							
	1	0	0	0.000	0.000		
	2	12,848	190,771	0.030	0.300	12,848	0.030
	3	35,174	346,105	0.080	0.497	22,326	0.049
	4	124,456	685,305	0.235	0.924	89,282	0.144
	5	647,473	4,098,366	1.107	4.508	523,017	0.706
	Total	819,951	4,291,335	0.299	1.164		
ADD							
	1	0	0	0.000	0.000		
	2	(205)	182,866	0.014	0.358	(205)	0.014
	3	(4,949)	272,965	0.033	0.505	(4,744)	0.019
	4	(3,371)	352,774	0.074	0.577	1,578	0.040
	5	(7,726)	422,975	0.140	0.664	(4,335)	0.061
	Total	(16,251)	833,130	(0.003)	0.277		
SB							
	1	0	0	0.000	0.000		
	2	8,650	175,543	0.032	0.316	8,650	0.032
	3	10,927	275,491	0.063	0.471	2,277	0.030
	4	17,818	368,370	0.106	0.570	6,891	0.040
	5	12,875	440,455	0.173	0.684	(4,943)	0.061
	Total	50,271	870,120	0.021	0.284		
BF							
	1	0	0	0.000	0.000		
	2	12,243	199,536	0.041	0.382	12,243	0.041
	3	20,320	303,669	0.084	0.567	8,078	0.041
	4	38,157	423,818	0.142	0.679	17,837	0.054
	5	51,227	547,415	0.223	0.842	13,070	0.071
	Total	121,946	1,110,267	0.046	0.356		

APPENDIX C

THEOREMS

Theorem 1: Under the assumptions of Model I, $y_{LSL} = a_{LSL} + b_{LSL}x$ is an unbiased estimator of y; i.e., $E(y_{LSL}) = E(y)$. Under the assumptions of Model II, $y_{LSM} = b_{LSM}x$ is an unbiased estimator of y.

Proof: Model I assumes that $E(y) = a + bx$. Since all expectations are conditional on x and since a_{LSL} and b_{LSL} are unbiased, we have

$$\begin{aligned} E(y_{LSL}) &= E(a_{LSL} + b_{LSL}x) \\ &= E(a_{LSL}) + E(b_{LSL}x) \\ &= E(a_{LSL}) + E(b_{LSL})x \\ &= a + bx \\ &= E(y). \end{aligned}$$

The proof for LSM is similar.

Lemma 1: Under LSL, $E(x_n \mid x_0) = a_n + b_n E(x_{n-1} \mid x_0)$. Under LSM, $E(x_n \mid x_0) = b_n E(x_{n-1} \mid x_0)$.

Proof 1: The proof will be given for LSL. The proof for LSM is similar.

First,

$$f(x_n|x_0) = \frac{f(x_n, x_0)}{f(x_0)}$$

$$= \frac{\int_{x_{n-1}} f(x_n, x_{n-1}, x_0)\, dx_{n-1}}{f(x_0)}.$$

Next, the "Multiplication Rule" of conditional density functions (Hogg and Craig [4, p. 64]) states that

$$f(x_n, x_{n-1}, x_0) = f(x_n \mid (x_{n-1}, x_0))\, f(x_{n-1} \mid x_0)\, f(x_0).$$

Therefore,

$$f(x_n \mid x_0) = \frac{\int_{x_{n-1}} f(x_n \mid (x_{n-1}, x_0))\, f(x_{n-1} \mid x_0)\, f(x_0)\, dx_{n-1}}{f(x_0)}$$

$$= \int_{x_{n-1}} f(x_n \mid (x_{n-1}, x_0))\, f(x_{n-1} \mid x_0)\, dx_{n-1}\,.$$

By the CLIA, the random variable $x_n \mid x_{n-1}$ is independent of x_0. Therefore $f(x_n \mid (x_{n-1}, x_0))$ does not depend on x_0, so $f(x_n \mid (x_{n-1}, x_0)) = f(x_n \mid x_{n-1})$. The rest of the proof hinges on our ability to interchange the order of integration. We will make whatever assumptions are necessary about the form of the density functions to justify that step. Then

$$\mathrm{E}\,(x_n \mid x_0) = \int_{x_n} x_n f(x_n \mid x_0) dx_n$$

$$= \int_{x_n} x_n \left(\int_{x_{n-1}} f(x_n \mid (x_{n-1}, x_0))\, f(x_{n-1} \mid x_0)\, dx_{n-1} \right) dx_n$$

$$= \int_{x_{n-1}} \left(\int_{x_n} x_n f(x_n \mid (x_{n-1}, x_0))\, dx_n \right) f(x_{n-1} \mid x_0)\, dx_{n-1} \qquad \text{(C.1)}$$

$$= \int_{x_{n-1}} \left(\int_{x_n} x_n f(x_n \mid x_{n-1}) dx_n \right) f(x_{n-1} \mid x_0)\, dx_{n-1}$$

$$= \int_{x_{n-1}} (a_n + b_n x_{n-1}) f(x_{n-1}|x_0)\, dx_{n-1}$$

$$= a_n + b_n \int_{x_{n-1}} x_{n-1} f(x_{n-1} \mid x_0)\, dx_{n-1}$$

$$= a_n + b_n \,\mathrm{E}\,(x_{n-1} \mid x_0)\,.$$

Proof 2: Recall the well-known identity $\mathrm{E}\,(X) = \mathrm{E}_Y\,[\mathrm{E}\,(X|Y)]$ (Hossack, et al, [5, p. 63]). Consider the following variation reiterated in Equation C.1 above:

$$\mathrm{E}\,(x_n \mid x_0) = \mathrm{E}_{x_{n-1} \mid x_0}\,[\mathrm{E}\,(x_n \mid (x_{n-1}, x_0))]\,.$$

For LSL we have:

$$\begin{aligned}
\mathrm{E}\,(x_n \mid x_0) &= \mathrm{E}_{x_{n-1}|x_0}\,[\mathrm{E}\,(x_n \mid (x_{n-1}, x_0))] \\
&= \mathrm{E}_{x_{n-1}|x_0}\,[\mathrm{E}\,(x_n \mid x_{n-1})] \quad \text{by CLIA} \\
&= \mathrm{E}_{x_{n-1}|x_0}\,[a_n + b_n x_{n-1}] \\
&= a_n + b_n \,\mathrm{E}\,(x_{n-1} \mid x_0)\,.
\end{aligned}$$

Theorem 2: $\mathrm{E}\,(\hat{\mu}_n \mid x_0) = \mathrm{E}\,(x_n \mid x_0)$.

Proof: By induction on n. The proof will be given for LSL; the proof for LSM is similar.

For $n = 1$, the theorem is simply a restatement of Theorem 1.

Assume that $\mathrm{E}\,(\hat{\mu}_{n-1} \mid x_0) = \mathrm{E}\,(x_{n-1} \mid x_0)$. We have that $\hat{\mu}_n = \hat{a}_n + \hat{b}_n \hat{\mu}_{n-1}$ where $\hat{a}_n$ and $\hat{b}_n$ are functions of the random variables $x_n|x_{n-1}$, and $\hat{\mu}_{n-1}$ is a function of the random variables $x_{n-1} \mid x_{n-2}, \ldots, x_1 \mid x_0$, and x_0. The CLIA implies that $x_n|x_{n-1}$ is independent of $x_{n-1}|x_{n-2}, \ldots, x_1|x_0$, and x_0, so $\hat{a}_n$ and $\hat{b}_n$ are independent of $\hat{\mu}_{n-1}$. Therefore,

$$E(\hat{\mu}_n|x_0) = E(\hat{a}_n|x_0) + E(\hat{b}_n|x_0)E(\hat{\mu}_{n-1}|x_0) \quad \text{where } \hat{b}_n \text{ and } \hat{\mu}_{n-1} \text{ are independent}$$

$$= E_{x_{n-1}|x_0}[E(\hat{a}_n|(x_{n-1}, x_0))] + E_{x_{n-1}|x_0}[E(\hat{b}_n|(x_{n-1}, x_0))]\, E(\hat{\mu}_{n-1}|x_0)$$

$$= E_{x_{n-1}|x_0}[E(\hat{a}_n|(x_{n-1})] + E_{x_{n-1}|x_0}[E(\hat{b}_n|(x_{n-1})]\, E(\hat{\mu}_{n-1}|x_0)$$

$$= E_{x_{n-1}|x_0}[a_n] + E_{x_{n-1}|x_0}[b_n]\, [E(\hat{\mu}_{n-1}|x_0)$$

$$= a_n + b_n\, E(\hat{\mu}_{n-1}|x_0)$$

$$= a_n + b_n E(x_{n-1}|x_0) \quad \text{by the induction hypothesis}$$

$$= E(x_n|x_0) \quad \text{by Lemma 1.}$$

Theorem 3:

	Linear	Multiplicative
For $n = 1$:	$\mathrm{Var}(\hat{\mu}_1) = \dfrac{\sigma_1^2}{I_1} + (x_0 - \bar{x}_0)^2\, \mathrm{Var}(\hat{b}_1)$	$\mathrm{Var}(\hat{\mu}_1) = x_0^2\, \mathrm{Var}(\hat{b}_1)$
For $n > 1$:	$\mathrm{Var}(\hat{\mu}_n) = \dfrac{\sigma_n^2}{I_n} + (\mu_{n-1} - \bar{x}_{n-1})^2 \mathrm{Var}(\hat{b}_n) + b_n^2\, \mathrm{Var}(\hat{\mu}_{n-1}) + \mathrm{Var}(\hat{b}_n)\, \mathrm{Var}(\hat{\mu}_{n-1})$	$\mathrm{Var}(\hat{\mu}_n) = \mu_{n-1}^2 \mathrm{Var}(\hat{b}_n) + b_n^2\, \mathrm{Var}(\hat{\mu}_{n-1}) + \mathrm{Var}(\hat{b}_n)\mathrm{Var}(\hat{\mu}_{n-1})$

Proof: We will prove the multiplicative case first. We saw in Theorem 6 that $\hat{b}_n$ and $\hat{\mu}_{n-1}$ are independent random variables. The formula (Hogg and Craig, [4, p. 178, problem 4.92]) for the variance of the product of two independent random variables x and y is:

$$\text{Var}(xy) = \sigma_x^2\sigma_y^2 + \mu_x^2\sigma_y^2 + \mu_y^2\sigma_x^2 .$$

This proves the assertion because $\hat{b}_n$ is unbiased.

For the linear case,

$$\text{Var}(\hat{\mu}_n) = \text{Var}(\hat{a}_n) + 2\text{Cov}(\hat{a}_n, \hat{b}_n\hat{\mu}_{n-1}) + \text{Var}(\hat{b}_n\hat{\mu}_{n-1}) .$$

It is well known (Miller and Wichern, [6, p. 202]) that the random variables $\bar{x}_n$ and $\hat{b}_n$ are uncorrelated when $\hat{b}_n$ is determined by least squares. Since all expectations are conditional, we have that

$$\begin{aligned}\text{Var}(\hat{a}_n) &= \text{Var}(\bar{x}_n - \bar{x}_{n-1}\hat{b}_n)\\ &= \text{Var}(\bar{x}_n) + \bar{x}_{n-1}^2\text{Var}(\hat{b}_n)\\ &= \frac{\sigma_n^2}{I_n} + \bar{x}_{n-1}^2\text{Var}(\hat{b}_n). \qquad \text{(C.2)}\end{aligned}$$

Next,

$$\begin{aligned}\text{Cov}(\hat{a}_n, \hat{b}_n\hat{\mu}_{n-1}) &= \text{E}(\hat{\mu}_{n-1})\text{Cov}(\hat{a}_n, \hat{b}_n) \quad \text{where } \hat{\mu}_{n-1} \text{ is independent of } \hat{a}_n \text{ and } \hat{b}_n\\ &= \mu_{n-1}\text{Cov}(\hat{a}_n, \hat{b}_n)\end{aligned}$$

and

$$\begin{aligned}\text{Cov}(\hat{a}_n, \hat{b}_n) &= \text{Cov}(\bar{x}_n - \bar{x}_{n-1}\hat{b}_n, \hat{b}_n)\\ &= \text{Cov}(-\bar{x}_{n-1}\hat{b}_n, \hat{b}_n)\\ &= -\bar{x}_{n-1}\text{Var}(\hat{b}_n). \qquad \text{(C.3)}\end{aligned}$$

Putting these together with the formula for $\text{Var}(\hat{b}_n\hat{\mu}_{n-1})$ from the multiplicative derivation above we have:

$$\operatorname{Var}(\hat{\mu}_n) = \frac{\sigma_n^2}{I_n} + \bar{x}_{n-1}^2 \operatorname{Var}(\hat{b}_n) - 2\mu_{n-1}\bar{x}_{n-1}\operatorname{Var}(\hat{b}_n)$$

$$+ \hat{\mu}_{n-1}^2 \operatorname{Var}(\hat{b}_n) + \hat{b}_n^2 \operatorname{Var}(\hat{\mu}_{n-1}) + \operatorname{Var}(\hat{\mu}_{n-1})$$

$$= \frac{\sigma_n^2}{I_n} + (\mu_{n-1} - \bar{x}_{n-1})^2 \operatorname{Var}(\hat{b}_n)$$

$$+ b_n^2 \operatorname{Var}(\hat{\mu}_{n-1}) + \operatorname{Var}(\hat{b}_n)\operatorname{Var}(\hat{\mu}_{n-1}).$$

Theorem 4A: Under LSL and LSM,

$$\operatorname{Var}(x_n | x_0) = \sigma_n^2 + b_n^2 \operatorname{Var}(x_{n-1} | x_0).$$

Therefore, an estimate of the process risk can be had by plugging in estimates of σ_n^2, b_n^2 and the estimate of process risk from the prior recursion step.

Proof:

$$\operatorname{Var}(x_n | x_0) = \operatorname{E}_{x_{n-1}|x_0}[\operatorname{Var}(x_n | (x_{n-1}, x_0))] + \operatorname{Var}_{x_{n-1}|x_0}[\operatorname{E}(x_n | x_{n-1}, x_0))]$$

$$= \operatorname{E}_{x_{n-1}|x_0}[\operatorname{Var}(x_n | x_{n-1})] + \operatorname{Var}_{x_{n-1}|x_0}[\operatorname{E}(x_n | x_{n-1})] \qquad \text{by CLIA}$$

$$= \operatorname{E}_{x_{n-1}|x_0}(\sigma_n^2) + \operatorname{Var}_{x_{n-1}|x_0}(a_n + b_n x_{n-1}) \qquad \text{under LSL}$$

$$= \sigma_n^2 + b_n^2 \operatorname{Var}(x_{n-1} | x_0) \qquad \text{under LSL or LSM.}$$

Theorem 4B: For the WAD method, an estimate of the process variance of the prediction of the next evaluation for a single accident year is:

for $n = 1$,

$$\hat{\text{V}}\text{ar}(x_n | x_0) = x_0 \, \hat{\sigma}_1^2$$

and for $n > 1$,

$$\hat{\text{V}}\text{ar}(x_n | x_0) = \hat{\mu}_{n-1}\hat{\sigma}_n^2 + \hat{b}_n^2 \, \hat{\text{V}}\text{ar}(x_{n-1} | x_0) \, .$$

Proof: For $n = 1$, the WAD model states that

$$x_1 = x_0 b_1 + \sqrt{x_0} e_1 \, ,$$

where the variance of the random variable e_1 is σ_1^2. Therefore, the variance of x_1 given x_0 equals the variance of the error term $\sqrt{x_0} \, e_1$, or $x_0 \, \sigma_1^2$. An estimate of this process risk can be had by plugging in the estimate $\hat{\sigma}_1^2$ of σ_1^2 and the actual value of x_0.

For $n > 1$,

$$\text{Var}(x_n | x_0) = \text{E}_{x_{n-1}|x_0}[\text{Var}(x_n | (x_{n-1}, x_0))] + \text{Var}_{x_{n-1}|x_0}[\text{E}(x_n | (x_{n-1}, x_0))]$$

$$= \text{E}_{x_{n-1}|x_0}[\text{Var}(x_n | x_{n-1})] + \text{Var}_{x_{n-1}|x_0}[\text{E}(x_n | x_{n-1})] \qquad \text{by CLIA}$$

$$= \text{E}_{x_{n-1}|x_0}(x_{n-1}\sigma_n^2) + \text{Var}_{x_{n-1}|x_0}(b_n x_{n-1}) \qquad \text{under WAD}$$

$$= \text{E}(x_{n-1} | x_0)\sigma_n^2 + b_n^2 \, \text{Var}(x_{n-1} | x_0).$$

Estimates of this quantity can be had by plugging in estimates of the individual parameters: $\hat{\sigma}_n^2$ for σ_n^2, the point estimate of μ_{n-1}, $\hat{b}_n$, for b_n, and the parameter risk estimate from the previous recursion step for $\text{Var}(x_{n-1} | x_0)$.

Theorem 4C: For the SAD method, an estimate of the process variance of the prediction of the next evaluation for a single accident year is:

for $n = 1$,

$$\hat{\mathrm{V}}\mathrm{ar}\,(x_n|x_0) = x_0^2\,\hat{\sigma}_1^2,$$

and for $n > 1$,

$$\hat{\mathrm{V}}\mathrm{ar}\,(x_n|x_0) = \hat{\mu}_{n-1}^2\;\hat{\sigma}_n^2 + \hat{b}_n^2\;\hat{\mathrm{V}}\mathrm{ar}\,(x_{n-1}|x_0)\;.$$

Proof: For $n = 1$, the SAD model states that:

$$x_1 = x_0 b_1 + x_0 e_1\;,$$

where the variance of the random variable e_1 is σ_1^2. Therefore, the variance of x_1 given x_0 equals the variance of the error term $x_0 e_1$, or $x_0^2\sigma_1^2$. An estimate of this process risk can be had by plugging in the estimate $\hat{\sigma}_1^2$ of σ_1^2 and the actual value of x_0.

For $n > 1$,

$$\mathrm{Var}\,(x_n|x_0) = \mathrm{E}_{x_{n-1}|x_0}\,[\mathrm{Var}\,(x_n|(x_{n-1}, x_0))] + \mathrm{Var}_{x_{n-1}|x_0}[\mathrm{E}\,(x_n|(x_{n-1}, x_0))]$$

$$= \mathrm{E}_{x_{n-1}|x_0}\,[\mathrm{Var}\,(x_n|x_{n-1})] + \mathrm{Var}_{x_{n-1}|x_0}[\mathrm{E}\,(x_n|x_{n-1})] \qquad \text{by CLIA}$$

$$= \mathrm{E}_{x_{n-1}|x_0}(x_{n-1}^2\sigma_n^2) + \mathrm{Var}_{x_{n-1}|x_0}(b_n x_{n-1}) \qquad \text{under SAD}$$

$$= \mathrm{E}\,(x_{n-1}^2|x_0)\,\sigma_n^2 + b_n^2\,\mathrm{Var}\,(x_{n-1}|x_0)$$

$$= (\mu_{n-1}^2 + \mathrm{Var}(x_{n-1}|x_0))\,\sigma_n^2 + b_n^2\,\mathrm{Var}\,(x_{n-1}|x_0).$$

Estimates of this quantity can be had by plugging in estimates of the individual parameters: $\hat{\sigma}_n^2$ for σ_n^2, the point estimate of μ_{n-1}, $\hat{b}_n$, for b_n, and the parameter risk estimate from the previous recursion step for $\mathrm{Var}\,(x_{n-1}|x_0)$.

Lemma 2: $\mathrm{E}\,(S_n) = na_n + b_n\,(\mathrm{E}\,(S_{n-1}) + x_{n-1,\,n-1})$.

Proof:

$$
\begin{aligned}
\mathrm{E}\,(S_n) &= \mathrm{E}\left(\sum_{i=0}^{n-1} x_{i,n} | x_{i,i}\right) \\
&= \sum_{i=0}^{n-1} \mathrm{E}\,(x_{i,n} | x_{i,i}) \\
&= \sum_{i=0}^{n-1} \mathrm{E}_{x_{i,n-1}|x_{i,i}}\,[\mathrm{E}\,(x_{i,n}|(x_{i,n-1}, x_{i,i}))] \\
&= \sum_{i=0}^{n-1} \mathrm{E}_{x_{i,n-1}|x_{i,i}}\,[\mathrm{E}\,(x_{i,n}|(x_{i,n-1})] \qquad \text{by CLIA} \\
&= \sum_{i=0}^{n-1} \mathrm{E}_{x_{n-1}|x_0}(a_n + b_n\, x_{i,n-1}) \\
&= na_n + b_n \left(\sum_{i=0}^{n-2} \mathrm{E}\,(x_{i,n-1}|x_{i,i}) + x_{n-1,n-1}\right) \\
&= na_n + b_n\,(\mathrm{E}\,(S_{n-1}) + x_{n-1,n-1}).
\end{aligned}
$$

Theorem 5: Let $XD_n = (x_{0,0}, x_{1,1}, \ldots, x_{n-1,n-1})$ denote the current diagonal of the triangle for the n youngest accident years. Then

$$\mathrm{E}(\hat{M}_n | XD_n) = \mathrm{E}(S_n)\,.$$

Proof: By induction on n. The proof will be given for LSL; the proof for LSM is similar. For $n = 1$, we know that:

$$\begin{aligned} E(\hat{M}_1|XD_1) &= E(\hat{\mu}_{0,1}|x_{0,0}) \\ &= E(x_{0,1}|x_{0,0}) \qquad \text{by Theorem 2} \\ &= E(S_1). \end{aligned}$$

Now, assume

$$E(\hat{M}_{n-1}|XD_{n-1}) = E(S_{n-1}).$$

Under LSL,

$$\hat{M}_n = n\hat{a}_n + \hat{b}_n(\hat{M}_{n-1} + x_{n-1,n-1})$$

where $\hat{a}_n$ and $\hat{b}_n$ are functions of the random variables $x_{i,n}|\, x_{i,n-1}$, $i \geq n$, and $\hat{M}_{n-1}$ is a function of random variables $x_{i,j}|\, x_{i,j-1}$ and of $x_{j,j}$ for $j < n$ and $i > n$. By the CLIA, $\hat{a}_n$ and $\hat{b}_n$ are independent of $\hat{M}_{n-1}$. Therefore:

$$\begin{aligned} E(\hat{M}_n \mid XD_n) &= E(n\hat{a}_n + \hat{b}_n(\hat{M}_{n-1} + x_{n-1,n-1}) \mid XD_n) \\ &= E(n\hat{a}_n \mid XD_n) + E(\hat{b}_n \mid XD_n)\, E(\hat{M}_{n-1} + x_{n-1,n-1} \mid XD_n) \\ &= na_n + b_n(E(\hat{M}_{n-1} \mid XD_{n-1}) + x_{n-1,n-1}) \\ &= na_n + b_n(E(S_{n-1}) + x_{n-1,n-1}) \qquad \text{by the induction hypothesis} \\ &= E(S_n) \qquad \text{by Lemma 2.} \end{aligned}$$

Theorem 6: Parameter Risk

	Linear	Multiplicative
For $n = 1$:		
	$\mathrm{Var}(\hat{M}_1) = \frac{\sigma_1^2}{I_1} + (x_{0,0} - \bar{x}_0)^2 \, \mathrm{Var}(\hat{b}_1)$	$\mathrm{Var}(\hat{M}_1) = x_{0,0}^2 \mathrm{Var}(\hat{b}_1)$
For $n > 1$:		
	$\mathrm{Var}(\hat{M}_n) = n^2 \frac{\sigma_n^2}{I_n} + (M_{n-1} + x_{n-1,n-1} - n\bar{x}_{n-1})^2 \, \mathrm{Var}(\hat{b}_n) + b_n^2 \, \mathrm{Var}(\hat{M}_{n-1}) + \mathrm{Var}(\hat{b}_n) \mathrm{Var}(\hat{M}_{n-1})$	$\mathrm{Var}(\hat{M}_n) = (M_{n-1} + x_{n,n})^2 \mathrm{Var}(\hat{b}_n) + b_n^2 \, \mathrm{Var}(\hat{M}_{n-1}) + \mathrm{Var}(\hat{b}_n) \, \mathrm{Var}(\hat{M}_{n-1})$

Proof: We will prove the multiplicative case first. Since $\hat{M}_n = \hat{b}_n \, (\hat{M}_{n-1} + x_{n-1,n-1})$, the proof is immediate by virtue of the formula for the variance of the product of two independent random variables, once we note that:

$$\mathrm{Var}\,(\hat{M}_{n-1} + x_{n-1,n-1}) = \mathrm{Var}\,(\hat{M}_{n-1})$$

because $x_{n-1,\,n-1}$ can be treated as a constant with respect to this conditional variance.

For the linear case,

$$\begin{aligned} \mathrm{Var}\,(\hat{M}_n) &= \mathrm{Var}\,(n\hat{a}_n) + \hat{b}_n(\hat{M}_{n-1} + x_{n-1,n-1})) \\ &= \mathrm{Var}\,(n\hat{a}_n) + 2\mathrm{Cov}\,(n\hat{a}_n, \hat{b}_n(\hat{M}_{n-1} + x_{n-1,\,n-1})) + \\ &\quad \mathrm{Var}\,(\hat{b}_n(\hat{M}_{n-1} + x_{n-1,\,n-1})). \end{aligned}$$

In the proof of Theorem 3 we saw that (Equation C.2)

$$\mathrm{Var}\,(\hat{a}_n) = \frac{\sigma_n^2}{I_n} + \bar{x}_{n-1}^{\,2} \mathrm{Var}\,(\hat{b}_n),$$

and that (Equation C.3)

$$\mathrm{Cov}\,(\hat{a}_n,\hat{b}_n) = -\bar{x}_{n-1}\mathrm{Var}\,(\hat{b}_n)\,.$$

Since $\hat{M}_{n-1}$ is independent of $\hat{a}_n$ and $\hat{b}_n$ and since all expectations are conditional on the current diagonal,

$$\mathrm{Cov}\,(n\hat{a}_n,\hat{b}_n(\hat{M}_{n-1} + x_{n-1,n-1})) = n\mathrm{E}\,(\hat{M}_{n-1} + x_{n-1,n-1})\mathrm{Cov}\,(\hat{a}_n,\hat{b}_n)\,.$$

Therefore

$$\mathrm{Var}\,(\hat{M}_n) = n^2\left(\frac{\sigma_n^2}{I_n} + \bar{x}_{n-1}^2\mathrm{Var}\,(\hat{b}_n)\right) - 2n\mathrm{E}\,(\hat{M}_{n-1} + x_{n-1,n-1})\,\bar{x}_{n-1}\mathrm{Var}\,(\hat{b}_n)$$

$$+ (\mathrm{M}_{n-1} + x_{n-1,n-1})^2\,\mathrm{Var}\,(\hat{b}_n) + b_n^2\,\mathrm{Var}\,(\hat{M}_{n-1}) + \mathrm{Var}\,(\hat{b}_n)\,\mathrm{Var}\,(\hat{M}_{n-1})$$

$$= n_2\frac{\sigma_n^2}{I_n} + (M_{n-1} + x_{n-1,n-1} - n\bar{x}_{n-1})^2\,\mathrm{Var}(\hat{b}_n)$$

$$+ b_n^2\,\mathrm{Var}\,(\hat{M}_{n-1}) + \mathrm{Var}\,(\hat{b}_n)\,\mathrm{Var}\,(\hat{M}_{n-1}).$$

Theorem 7A: Process Risk for the LSL and LSM models

For $n = 1$: $\mathrm{Var}\,(S_1) = \sigma_1^2$;

for $n > 1$: $\mathrm{Var}\,(S_n) = n\sigma_n^2 + b_n^2\,\mathrm{Var}\,(S_{n-1})$.

Proof: For $n = 1$, S_1 is just the first future value of the youngest accident year conditional on its current value; i.e., $S_1 = x_{0,1}|x_{0,0}$. Therefore, $\mathrm{Var}\,(S_1) = \mathrm{Var}\,(x_{0,1}|x_{0,0}) = \sigma_1^2$ by definition of σ_1.

For $n > 1$, let X_{n-1} denote the vector of random variables $(x_{0,0}, \ldots, x_{n-2,n-1})$ corresponding to the unknown future evaluations of the $n-1$ youngest accident years as of age $n-1$. It is understood that all expectations are conditional on the current diagonal. First, recall that $S_n = \sum_{i=0}^{n-1} x_{i,n} \mid x_{i,i}$.

Next, note that

$$\text{Var}(S_n) = \text{E}_{x_{n-1}}[\text{Var}(S_n|X_{n-1})] + \text{Var}_{x_{n-1}}[\text{E}(S_n|X_{n-1})]\,. \qquad \text{(C.4)}$$

For the first term,

$$\text{Var}(S_n|X_{n-1}) = \text{Var}\left(\sum_{i=0}^{n-1} x_{i,n}|x_{i,n-1}\right)$$

$$= \sum_{i=0}^{n-1} \text{Var}(x_{i,n}|x_{i,n-1}) \quad \text{because accident years are independent}$$

$$= n\sigma_n^2$$

because σ_n^2 is constant across accident years.

For the second term of Equation C.4,

$$\text{E}(S_n|X_{n-1}) = \text{E}(a_n + b_n(S_{n-1} + x_{n-1,n-1})) \quad \text{where } a_n = 0 \text{ for LSM}$$

$$= \text{E}(a_n + b_n x_{n-1,n-1} + b_n S_{n-1}).$$

Therefore,

$$\text{Var}_{x_{n-1}}[\text{E}(S_n|X_{n-1})] = \text{Var}_{x_{n-1}}(b_n S_{n-1})$$

$$\text{because } a_n,\ b_n, \text{ and } x_{n-1,n-1} \text{ are constants}$$

$$= b_n^2\,\text{Var}(S_{n-1}).$$

Putting the two terms together, we have:

$$\mathrm{Var}(S_n) = n\sigma_n^2 + b_n^2\, \mathrm{Var}(S_{n-1})\,.$$

An unbiased estimate of this quantity can be had by plugging in unbiased estimates of σ_n^2 and b_n^2, and the Process Risk estimate from the prior recursion step.

Theorem 7B: Process Risk for the WAD model

For $n = 1$: $\mathrm{Var}(S_1) = x_{0,0}\sigma_1^2$;

for $n > 1$: $\mathrm{Var}(S_n) = (\mathrm{M}_{n-1} + x_{n-1,n-1})\, \sigma_n^2 + b_n^2\, \mathrm{Var}(S_{n-1})$.

Proof: The $n = 1$ case is just Theorem 4B. For $n > 1$, the proof follows that of Theorem 7A, with one difference; namely, $\mathrm{Var}(x_{i,n}|x_{i,n-1}) = x_{i,n-1}\sigma_n^2$. So the first term of Equation C.4 is:

$$\mathrm{E}_{x_{n-1}}[\mathrm{Var}(S_n|X_{n-1})] = \mathrm{E}_{x_{n-1}}\left[\sum_{i=0}^{n-1} x_{i,n-1}\sigma_n^2\right]$$

$$= \sigma_n^2\, \mathrm{E}\left[\sum_{i=0}^{n-1} x_{i,n-1} + x_{n-1,n-1}\right]$$

$$= \sigma_n^2\,(M_{n-1} + x_{n-1,n-1})$$

by definition of M_{n-1}. Since the second term of Equation C.4 simplifies to the same quantity as in Theorem 7A, this theorem is proved.

Theorem 7C: Process Risk for the SAD model

For $n = 1$: $\mathrm{Var}(S_1) = x_{0,0}^2\, \sigma_1^2$;

for $n > 1$: $\mathrm{Var}(S_n) = (x_{n-1,n-1}^2 + \sum_{i=0}^{n-2}\mu_{i,n-1}^2 + \mathrm{Var}(S_{n-1}))\, \sigma_n^2 + b_n^2\, \mathrm{Var}(S_{n-1})\,.$

Proof: The $n = 1$ case is just Theorem 4C. For $n > 1$, we have only to derive the first term of Equation C.4 in the proof of Theorem 7A. For SAD, $\mathrm{Var}(x_{i,n}|x_{i,n-1}) = x_{i,n-1}^2\,\sigma_n^2$, so for $i < n - 1$,

$$\begin{aligned} E_{x_{i,n-1}}[\mathrm{Var}(x_{i,n}|x_{i,n-1})] &= \sigma_n^2\, E\,(x_{i,n-1}^2) \\ &= \sigma_n^2\,[E^2(x_{i,n-1}) + \mathrm{Var}(x_{i,n-1})] \\ &= \sigma_n^2[\mu_{i,n-1}^2 + \mathrm{Var}(x_{i,n-1})]. \end{aligned}$$

Therefore

$$E_{x_{n-1}}[\mathrm{Var}(S_n|X_{n-1})] = E_{x_{n-1}}[\mathrm{Var}\left(\sum_{i=0}^{n-1} x_{i,n}|x_{i,n-1}\right)]$$

by definition of S_n

$$= \left(\sum_{i=0}^{n-2} E_{x_{i,n-1}}(\mathrm{Var}(x_{i,n}|x_{i,n-1}))\right) + \mathrm{Var}(x_{n-1,n}|x_{n-1,n-1})$$

because accident years are independent

$$= \sigma_n^2\left(\sum_{i=0}^{n-2} \mu_{i,n-1}^2 + \sum_{i=0}^{n-2} \mathrm{Var}(x_{i,n-1})\right) + x_{n-1,n-1}^2\sigma_n^2$$

$$= \sigma_n^2\left(\sum_{i=0}^{n-2} \mu_{i,n-1}^2 + \mathrm{Var}\left(\sum_{i=0}^{n-2} x_{i,n-1}\right)\right) + x_{n-1,n-1}^2\sigma_n^2$$

$$= \sigma_n^2 \left(\sum_{i=0}^{n-2} \mu_{i,n-1}^2 + \mathrm{Var}(S_{n-1}) \right) + x_{n-1,n-1}^2 \sigma_n^2 .$$

This proves the theorem.

Theorem 8: Under the transformed GAD model:

$$x'_n = b'_n + x'_{n-1} + e'_n$$

where we assume that $\sigma'^2_j = \mathrm{Var}(e'_j)$ are identical for every j, the estimate of the variance of the prediction of ultimate (transformed) loss

$$\hat{\mu}'_u = x'_0 + \sum_{j=1}^{u} \hat{b}'_j$$

is

$$\left(c + \sum_{j=1}^{u} \frac{1}{I_j} \right) \hat{\sigma}'^2$$

where $\hat{\sigma}'^2$ denotes the MSE of the simultaneous solution of the link ratios of the transformed model.

Proof: Since we assume equal variances by development age, we may solve for all parameters b_j simultaneously with the equation:

$$\begin{pmatrix} x'_{n,1} - x'_{n,0} \\ x_{n-1,1} - x'_{n-1,0} \\ \cdot \\ x'_{1,1} - x'_{1,0} \\ x'_{n,2} - x'_{n,1} \\ \cdot \\ x'_{2,2} - x'_{2,1} \\ \cdot \\ x'_{n,n-1} - x'_{n,n-2} \\ x'_{n-1,n-1} - x'_{n-1,n-2} \\ x'_{n,n} - x'_{n,n-1} \end{pmatrix} = \begin{pmatrix} 1\,0 \ldots 0\,0 \\ 1\,0 \ldots 0\,0 \\ \cdot \\ 1\,0 \ldots 0\,0 \\ 0\,1 \ldots 0\,0 \\ \cdot \\ 0\,1 \ldots 0\,0 \\ \cdot \\ 0\,0 \ldots 1\,0 \\ 0\,0 \ldots 1\,0 \\ 0\,0 \ldots 0\,1 \end{pmatrix} \times \begin{pmatrix} b'_1 \\ b'_2 \\ \cdot \\ b'_{n-1} \\ b'_n \end{pmatrix} + \begin{pmatrix} e'_1 \\ e'_1 \\ \cdot \\ e'_1 \\ e'_2 \\ \cdot \\ e'_2 \\ \cdot \\ e'_{n-1} \\ e'_{n-1} \\ e'_n \end{pmatrix},$$

or, in more concise format, $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{E}$. It is well known that the least squares estimator of β is $\hat{\beta} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}$ and that the variance-covariance matrix of this estimator is $(\boldsymbol{X'X})^{-1}\sigma'^2$. In this case, it is clear by inspection that $\boldsymbol{X'X}$ is a diagonal matrix whose j^{th} entry equals I_j, the number of data points in the estimate of the j^{th} link ratio, and whose off-diagonal elements are zero. Thus, $\mathrm{Var}(\hat{b}'_j) = \frac{\sigma'^2}{I_j}$ and $\mathrm{Cov}\,(\hat{b}'_i, \hat{b}'_j) = 0$ for $i \neq j$. Therefore, the Parameter Risk

$$\mathrm{Var}\left(C + \sum_{j=1}^{U} \hat{b}'_j\right)$$

is exactly equal to:

$$\sigma'^2 \sum_{j=1}^{U} \frac{1}{I_j}.$$

The Process Risk is equal to:

$$\sum_{j=1}^{C} \mathrm{Var}\,(e'_n) = C\,\sigma'^2\ .$$

These variances are estimated by substituting the estimate $\hat{\sigma}'^2$ for σ'^2.