Unbiased Loss Development Factors

by Daniel M. Murphy

UNBIASED LOSS DEVELOPMENT FACTORS

By Daniel M. Murphy, FCAS, MAAA

UNBIASED LOSS DEVELOPMENT FACTORS

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 presents a mathematical framework to answer those questions for the class of linear link ratio estimators used in practice. A more accurate method of calculating link ratios is derived based on classical regression theory. The circumstances under which the traditional methods could be considered optimal are discussed. It is shown that two traditional estimators may in fact be least squares estimators depending on the set of assumptions one believes governs the process of loss development. Formulas for variances of, and confidence intervals around, point estimates of ultimate loss and loss reserves are derived. A triangle of incurred loss dollars is analyzed to demonstrate the concepts and techniques. A summary of a simulation study is presented and suggests that the performance of the incurred loss development technique based on the more general least squares estimator may approach that of the Bornhuetter-Ferguson and Stanard-Buhlmann techniques in some situations. The requisite mathematics is within the reach of the actuarial student equipped with the first three exams.

1. INTRODUCTION

Three standard methods of estimating link ratios in practice 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 even whether the methods are biased or unbiased.[1] 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 leads to discover systematic or operational reasons why a development triangle may appear as it does, and the only concern now is how to deal with the remaining noise.

Proofs of the technical theorems are relegated to the Appendix. The mathematics within the body of the paper is intended to motivate discussion and application.

An example will help motivate the exposition, so consider the accident year incurred loss development triangle and its triangle of link ratios in Figures 1A and 1B. The specific content of the example triangle is incidental to the purpose of this paper. It is hoped that the data is sufficiently realistic to exemplify adequately the application of these results. The extension of the results to other kinds of triangles should be self-evident.

Denote the link ratio as $\mathbf{b}$, and the SAD, WAD, and GAD estimates of $\mathbf{b}$ as $\mathbf{b}_{SAD}$, $\mathbf{b}_{WAD}$, and $\mathbf{b}_{GAD}$ respectively. For 12-24 months of development in the example triangle, these statistics evaluate to $\mathbf{b}_{SAD}=3.953$, $\mathbf{b}_{WAD}=2.480$, and $\mathbf{b}_{GAD}=3.129$. To determine which estimate is best, we must first unveil the hidden assumptions implicit in the actuarial technique called loss development.

[1] See, for example, James N. Stanard, "A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques," and John P. Robertson's discussion in the *Proceedings of the Casualty Actuarial Society*, LXXII, 1985.

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 multiplied by 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 multiplied by an unknown constant. Symbolically,

$$y = \mathbf{b}\, x + e$$

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

Expected Value of the Link Ratio

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

Model I $\quad y = \mathbf{a} + \mathbf{b}\, x + e$

$E(e)=0$, $\mathrm{Var}(e)$ is constant across accident years, 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: $E(y \mid x)=\mathbf{a}+\mathbf{b}x$. 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.[2] For the remainder of this paper all expectations are conditional on the current evaluation.

It is a well known theorem, the Gauss-Markoff Theorem, that the "best estimates" of $\mathbf{a}$ and $\mathbf{b}$ are the least squares estimates, denoted $\hat{a}$ and $\hat{b}$:

$$\hat{b} = \frac{\Sigma (x-\bar{x})\, y}{\Sigma (x-\bar{x})^2} \quad \text{and} \quad \hat{a} = \bar{y} - \hat{b}\bar{x}\ .$$

For example, the least squares estimates $\hat{a}$ and $\hat{b}$ for the 12-24 month development period in the triangle of Figure 1 are $\hat{a}$ = \$373.63 (all amounts will be given in thousands of dollars) and $\hat{b}$ = 2.027. These estimates were calculated using a popular spreadsheet software package.

[2] See, for example, Henry Scheffé, *The Analysis of Variance*, Wiley, 1956, p. 195.

The indicated regression line is shown in Figure 2A.

The method of estimating link ratios[3] by least squares under the assumptions of Model I will be called the Least Squares Linear (LSL) method. The least squares estimators of the line's parameters will be denoted $\mathbf{a}_{LSL}$ and $\mathbf{b}_{LSL}$.

Five properties of the least squares estimates are particularly appealing.[4]

1. The least squares estimates are linear functions of the variables $y \mid x$.
2. They are unbiased; i.e., $E(\mathbf{a}_{LSL})=\mathbf{a}$ and $E(\mathbf{b}_{LSL})=\mathbf{b}$.
3. Within the class of all linear unbiased estimates of **a** and **b**, the least squares estimates have the smallest variance. Least squares estimators are therefore called B.L.U.E.: Best Linear Unbiased Estimators.
4. The vertical deviations of the (x,y) observations from the regression line sum to zero; in other words, the average residual is zero.
5. The least squares regression line passes through the sample average $(\bar{x},\bar{y})$.

Before continuing, glance again at Figure 2A. By visual inspection one might say that the y-intercept is close enough to zero that it could reasonably be ignored in the predicted relationship between y and x.[5] If one believes the y-intercept should truly be zero in the first place, perhaps the model to use is

Model II $y = \mathbf{b}\, x + e$
$E(e)=0$, $Var(e)$ is constant across accident years, the e's are uncorrelated between accident years and are independent of x.

This model would be inappropriate if there were a significant probability that $x = 0$.

The BLUE estimator for **b** under Model II is

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

The method of estimating link ratios by least squares under this strictly multiplicative development model will be called the Least Squares Multiplicative (LSM) method. The least squares estimator of the line's parameter will be denoted $\mathbf{b}_{LSM}$.

---

[3] The estimate $\hat{a}$ of the constant term can be considered a "link ratio" if the link ratio function is viewed as being vector valued $(\hat{a},\hat{b})$.

[4] These results can be found in many introductory texts on statistical regression. Property 3 is the Gauss-Markoff Theorem.

[5] Although it will be demonstrated that the y-intercept is significantly different from zero.

In the example triangle the 12-24 month LSM link ratio is $\mathbf{b}_{LSM}$ = 2.204. Figure 2B illustrates the difference between the LSL and LSM indicated regression lines.

Does $\mathbf{b}_{LSM}$ satisfy the five properties of the LSL estimator above? Obviously, $\mathbf{b}_{LSM}$ is a linear function of the y's (again, conditional on the known x values). The fact that it is unbiased is easy to prove. It has minimum variance within the class of linear unbiased estimators by virtue of the Gauss-Markoff Theorem because it is the least squares estimator. But $\mathbf{b}_{LSM}$ does not necessarily satisfy Properties 4 and 5. At first, the fact that $\mathbf{b}_{LSM}$ does not zero out the sum of the residuals nor determine a regression line passing through $(\bar{x}, \bar{y})$ may seem to be a drawback. But on second thought, it must be inevitable. Indeed, a least squares regression line is required to satisfy two conditions: it must be close to the data and it must zero out the residuals. A two parameter line is free to satisfy two conditions. But a one-parameter line has the ability to satisfy only one condition. LSM satisfies the first, so it cannot be expected to satisfy the second as well.

If one were to define a "good" linear unbiased estimator as one which satisfies Properties 4 and 5, but not necessarily Property 3, then $\mathbf{b}_{WAD}$ would be best (Theorem 1). However, the price of adopting $\mathbf{b}_{WAD}$ rather than $\mathbf{b}_{LSM}$ is an increase in the probability that the prediction of losses as of the next evaluation would be off the mark because the variance of $\mathbf{b}_{WAD}$ is greater than the variance of $\mathbf{b}_{LSM}$.[6] Such are the standards by which $\mathbf{b}_{WAD}$ may be considered "optimal."

In the example, with $\mathbf{b}_{LSM}$= 2.204 for 12-24 months of development, the average residual is \$227.9 and the standard deviation of the residuals is \$876.5. With $\mathbf{b}_{WAD}$=2.480, the average residual is \$0.4 and the standard deviation of the residuals is \$953.1.

Let us continue now to attack the assumptions of LSL and LSM to discover what we can about $\mathbf{b}_{SAD}$ and $\mathbf{b}_{GAD}$. Take the constant variance assumption for example. The impact of trend would imply that the variance of e is not constant across accident years. On-leveling the loss triangle may adjust for such heteroskedasticity but in addition may introduce unwelcome side effects. A model that speaks directly to the issue of non-constant variances is

Model III $\quad y = \mathbf{b}\, x + x\, e$

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

This model differs from Model II in that it explicitly postulates a dependent relationship between the current evaluation and the error term, xe. By dividing both sides of this equation by x we see that 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 $\mathbf{b}$ is $\mathbf{b}_{SAD}$ (Theorem 3). The technique of estimating link ratios under the assumptions of Model III will be called the SAD method.

[6] Again, the Gauss-Markoff Theorem. This fact is proved directly for this actuarial problem as Theorem 2. Intuitively, Var($\mathbf{b}_{LSM}$)≤Var($\mathbf{b}_{WAD}$) because $\mathbf{b}_{LSM}$ gives more weight to the larger values of x.

Another model that can adjust for trend is

Model IV $y = \mathbf{b}xe$

$E(e)=1$, $Var(e)$ is constant across accident years, the e's are uncorrelated between accident years and are 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 IV is the geometric average of the link ratios, $\mathbf{b}_{GAD}$ (Theorem 4). The technique of estimating link ratios under the assumptions of Model IV will be called the GAD method.

For the remainder of the paper, results will be stated in terms of the LSL and LSM methods. Results for SAD and GAD, which are left to the reader, can be derived directly or by applying the results below to the transformed SAD and GAD models on which Theorems 3 and 4, respectively, depend.

Estimate of the Next Evaluation

The point estimate of the expected value of incurred losses as of the next evaluation given the current evaluation is

LSL	LSM
$y_{LSL} = \mathbf{a}_{LSL} + \mathbf{b}_{LSL}\,x$	$y_{LSM} = \mathbf{b}_{LSM}\,x$

The estimates are unbiased under the assumptions of their respective models (Theorem 5). For the example triangle the LSL and LSM estimates of the 24-month evaluation of accident year 1991 are, respectively, \$2983 = \$374 + 2.027 x \$1287 and \$2837 = 2.204 x \$1287.

Estimated Ultimate Loss: A Single Accident Year

The Chain Ladder Method states that if $\mathbf{b}_1$ is a link ratio from 12 to 24 months, $\mathbf{b}_2$ is a link ratio from 24 to 36 months, etc., and if U is the number of links required to reach ultimate, then $\mathbf{B}_U=\mathbf{b}_1\mathbf{b}_2\cdots\mathbf{b}_U$ is the (to ultimate) loss development factor (LDF). The implicit assumption is that future development is independent of prior development. This assumption implies a type of "transitive" property of loss development: if the conditional expectation of y given x is $\mathbf{b}_1x$ and the conditional expectation of z given y is $\mathbf{b}_2y$ then the conditional expectation of z given x is $\mathbf{b}_1\mathbf{b}_2x$.[7]

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

[7] See Lemma 1 in Appendix A. This assumption may not hold in practice, for example, when a claims department issues orders to "strengthen reserves" after having operated for some time under a less conservative strategy.

that an unbiased estimate of a to-ultimate loss development factor is the product of the unbiased link ratio estimates; i.e., $\hat{B}_U = \hat{b}_1 \hat{b}_2 \cdots \hat{b}_U$.

The very simplicity of the closed form LDF is one of the beauties of the multiplicative development approach. A closed form expression for the intercept term of the more general LSL approach is not nearly as simple, but this should not be considered a deterrent because a closed form, to-ultimate expression is unnecessary. 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 easy 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 (Section 3), so the point estimates may as well be calculated in the same step.

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 $\mu_n = E(x_n \mid x_0)$. The unbiased chain estimate is built from the individual links $x_n \mid x_{n-1}$ of losses as of one age conditional on losses at the previous age.

Under the more general LSL model, it is assumed that for each n there exist constants $\mathbf{a}_n$ and $\mathbf{b}_n$ such that the random variable x_n conditional on x_{n-1} can be expressed as

$$x_n \mid x_{n-1} = \mathbf{a}_n + \mathbf{b}_n x_{n-1} + e_n \ .$$

It is also assumed that $E(e_n)=0$, that $Var(e_n)=\sigma^2_n$, and that the e_n's are independent of all the x's and, by the CLIA, of each other. Theorem 6 proves that the following recursive formulas yield unbiased estimates of future evaluation.

LSL	LSM
$\hat{\mu}_1 = \hat{a}_1 + \hat{b}_1 x_0$	$\hat{\mu}_1 = \hat{b}_1 x_0$
$\hat{\mu}_n = \hat{a}_n + \hat{b}_n \hat{\mu}_{n-1}$	$\hat{\mu}_n = \hat{b}_n \hat{\mu}_{n-1}$

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

For the example, the LSM estimate for 24-36 months of development is $\mathbf{b}_{LSM}=1.133$. Therefore, the prediction of accident year 1991 losses evaluated as of 36 months would be \$3380 = \$2983 x 1.133 if LSL had been used for the 12-24 development period; if LSM had been used, the estimate would be \$3214 = \$2837 x 1.133. The LSM prediction of accident year 1990 losses as of 36 months would be \$3167 = \$2795 x 1.133.

Estimated Total Ultimate Loss: Multiple Accident Years

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

The idea behind the recursive estimate for multiple accident years is this. Starting at the bottom left corner of the triangle, add up columns of estimated future evaluations. Find a recursive unbiased estimate of those column sums. Then an unbiased estimate of total losses at ultimate will be the final sum.

The formulas are developed as follows. To keep the notation from becoming too convoluted, index the rows of the triangle in reverse order so that the youngest accident year is the zeroth row, the next youngest is row 1, and so on. Next, index the columns so that the 12 month column is the zeroth column, the 24 month column is column 1, etc. A full triangle of N+1 accident years appears as in Figure 3. If

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

denotes the sum of the accident years' future evaluations conditional on the accident years' current evaluations, then an unbiased estimate of the future evaluation of multiple accident years is an estimate of $E(S_n)$. Let M_n denote this expectation. Recursive formulas for estimates of M_n are:

<u>LSL</u>	<u>LSM</u>
$\hat{M}_1 = \hat{a}_1 + \hat{b}_1 x_{0,0}$	$\hat{M}_1 = \hat{b}_1 x_{0,0}$
$\hat{M}_n = n\hat{a}_n + \hat{b}_n (\hat{M}_{n-1} + x_{n-1,n-1})$	$\hat{M}_n = \hat{b}_n (\hat{M}_{n-1} + x_{n-1,n-1})$

Stop when n=U, the age at which all accident years are assumed to have reached ultimate. These estimates are unbiased under the assumptions of their respective models. See Theorem 10.

The completed triangle of Figure 1A is shown in Figure 4 where it was assumed that LSL is appropriate through 84 months of development, LSM thereafter, and that losses are fully mature (i.e., case reserves are adequate, on average) after 108 months. Then, for example, $\hat{M}_1$ =\$2,982 because the 1991 accident year is the only one for which 24 months is a future development point. Accident years 1991 and 1990 are the only years which have yet to reach the age of 36 months, so $\hat{M}_2$ = \$3,268 + \$3,470 = \$6,738. And so on. Accident years 1984 through 1991 have yet to reach ultimate (108 months) so $\hat{M}_8$ = \$47,554.

Estimated Reserves for Outstanding Losses

Unbiased estimates of outstanding losses are

$$\hat{\mu}_U - \text{paid to date}$$

for a single accident year and

$$\hat{M}_U - \text{Total Paid To Date}$$

for multiple accident years.

Estimated Pure Premiums and Loss Ratios

Assuming exposures and earned premiums are static variables,[8] unbiased estimates of the pure premium rate and of the loss ratio for a single accident year are

$$\frac{\hat{\mu}_U}{\text{exposure}} \quad \text{and} \quad \frac{\hat{\mu}_U}{\text{earned premium}},$$

respectively. For multiple accident years, the estimates are

$$\frac{\hat{M}_U}{\text{Total Exposure}} \quad \text{and} \quad \frac{\hat{M}_U}{\text{Total Earned Premium}}$$

Of course, the latter statistics are most useful when all quantities are brought onlevel.

[8] Audit and reinsurance exposures and premiums may be random variables.

3. VARIANCE

The least squares point estimators of Section 2 are functions of random variables. As such, they are themselves random variables with their own inherent variances. Exact formulas for, and estimates of, these variances will be addressed in turn.

Variance of the Link Ratio Estimates

It is well known[9] that the exact variances of the link ratio estimators are

LSL	LSM
$\mathrm{Var}(\hat{a}) = \dfrac{\Sigma x^2}{n\Sigma(x-\bar{x})^2}\sigma^2$	
$\mathrm{Var}(\hat{b}) = \dfrac{\sigma^2}{\Sigma(x-\bar{x})^2}$	$\mathrm{Var}(\hat{b}) = \dfrac{\sigma^2}{\Sigma x^2}$ (1)

where

$$\bar{x} = \frac{1}{I}\sum x_i$$

and I is the number of accident years used in the estimate of the link ratio. Unbiased estimates of these variances are obtained by plugging in the unbiased estimate s^2 of σ^2 where s^2 is the Mean Square for Error (MSE) of the link ratio regression. The MSE or its square root s (the standard error of the estimate) is a standard statistic produced in the output of regression software. Most regression software will calculate an estimate of the square root of the variance in equation (1), sometimes called the standard error of the coefficient.

For 12-24 months of development in the example triangle $s^2_{LSL}=848.8^2$. Estimates of the standard deviations of the 12-24 month LSL intercept and slope factor are 77.35 and 0.194, respectively. For the LSM model the MSE is 876.5^2 and the standard error of the coefficient is 0.157. The spreadsheet software used to calculate these statistics automatically generates s_{LSL} and the standard error of the coefficient. The average x^2 value had to be calculated "by hand" to derive the estimate of $\mathrm{Var}(\hat{a})$.

Variance of Estimated Ultimate Loss: A Single Accident Year

Before continuing, it is time to make an important distinction. The point estimate of ultimate loss $\hat{\mu}_U$ calculated recursively above is an estimate of the expected value of the (conditional on

---

[9] See for example Robert B. Miller and Dean W. Wichern, *Intermediate Business Statistics*, Holt, Rinehart and Winston, 1977.

x_0) ultimate loss x_U.[10] 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 estimate $\hat{\mu}_U$ is comprised of two components: the variance in the estimate of the expected value of $x_U \mid x_0$ — Parameter Risk — and the inherent variability of ultimate loss about its mean μ_U — Process Risk.[11] 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

$$x_U \mid x_0 = \mu_U + \epsilon_U$$

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

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

The following recursive formulas for exact values of these two variance components are derived in Theorems 8 and 9.

Parameter Risk

LSL

For n=1:

$$\mathrm{Var}\,\hat{\mu}_1 = \frac{\sigma_1^2}{I_1} + (x_0 - \bar{x}_0)^2\,\mathrm{Var}\,\hat{b}_1$$

For n>1:

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

LSM

For n=1:

$$\mathrm{Var}\,\hat{\mu}_1 = x_0^2\,\mathrm{Var}\,\hat{b}_1$$

For n>1:

$$\mathrm{Var}\,\hat{\mu}_n = \mu_{n-1}^2\,\mathrm{Var}\,\hat{b}_n + b_n^2\,\mathrm{Var}\,\hat{\mu}_{n-1} + \mathrm{Var}\,\hat{b}_n\,\mathrm{Var}\,\hat{\mu}_{n-1}$$

---

[10] For better or for worse, it is usually the expected value of an unknown quantity – e.g., rates or reserves – that actuaries are called upon to produce. The "Statement of Principles Regarding Property and Casualty Loss and Loss Adjustment Expense Reserves" is rather vague on that issue, but "The "Statement of Principles Regarding Property and Casualty Insurance Ratemaking" (Principle 1) and, for example, "Actuarial Standard of Practice No. 7: Performing Cash Flow Testing for Insurers" (section 5.5) are quite explicit.

[11] This process risk is the conditional variance of developing losses about the conditional mean. As pertaining to triangles of incurred loss dollars, it includes, but is not limited to, 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.

Process Risk

$$\mathrm{Var}(x_1 \mid x_0) = \sigma_1^2$$

and

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

The equation for Process Risk is the same under both models. Unbiased estimates of these variances are obtained by plugging in unbiased estimates s_n^2 for σ_n^2, $\hat{\mathbf{b}}_n$ for $\mathbf{b}_n$, and $\hat{\mu}_n$ for μ_n.

Parameter Risk and Total Risk are illustrated in the familiar graphs of Figures 5A and 5B where ± 2 standard deviation prediction bands are drawn around the LSL and LSM estimates, respectively, of 12-24 months of development from the example triangle. First, Parameter Risk is represented by two curves ± 2 standard deviations (the square root of the estimated Parameter Risk) away from the least squares line. Total prediction risk is represented by two curves ± 2 standard deviations (the square root of the sum of estimated Parameter Risk plus estimated Process Risk) away from the least squares line. The actuary may represent Process Risk to the layman as the distance between the Total Risk and Parameter Risk bands; of course, this is technically incorrect.

Notice that in Figure 5A the Parameter Risk bands widen in both directions as x moves away from its average value of $824 and that in Figure 5B the bands widen as x moves away from zero. This occurs because the equation for parameter risk is a function of distance of x_0 from the average value of x for the LSL model and a function of the absolute value of x_0 for LSM.

There is a subtle difference between a "prediction band" which measures the error one would expect in a prediction based on the regression, and the more common "confidence band" which measures the fit of the regression relative to the sample data. The concept of the confidence band is illustrated in Figure 5C where, for example, a one-standard-deviation confidence band is drawn around the LSL regression of 12-24 months of development. The radius of the confidence band is the square root of the MSE, 848.8. Using the techniques of the next section, it can be shown that one should expect about 34% of the data points to fall outside the confidence band. In other words, one should expect about six outliers. In this case, there are only four. The identification of outliers can provide the actuary with useful information before he or she enters into fact-finding interviews with the claims and underwriting departments. The identification of outliers provides information of a more technical nature as well. Indeed, note that the outliers in Figure 5C occur at the higher values of x. This suggests that the variance of y is not independent of x. The assumptions of the SAD or GAD methods, or a variant, may more appropriately describe the random processes underlying these particular data.

As a final note, ultimate loss is not ultimate until the final claim is closed. Suppose it takes C development periods, $C \geq U$, 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 estimated ultimate loss through U development periods may be the same as estimated ultimate loss through C development periods, the variances of the two estimates are not the same. Even if it is true that $\mathbf{b}_n = 1$ for $n > U$, whereby parameter risk halts at $n = U$, process risk continues to add up, so recursive estimates of $\mathrm{Var}(x_n \mid x_0)$ should be carried out beyond $n = U$.

In the example, it was assumed that an accident year will be closed after 144 months based on a visual inspection of Figure 1B (accident year 1974 was considered a data anomaly). The recursive projection of ultimate accident year 1991 loss was already displayed in Figure 4. The detailed calculation of the variance (Total Risk) is shown in Figure 6A.

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 for a given accident year the variance in the prediction $PRED_U$ of ultimate loss is

$$\begin{aligned} \mathrm{Var}(PRED_U) &= \mathrm{Var}(\hat{M}_U) + \mathrm{Var}(E_U) \\ &= \text{PARAMETER RISK} + \text{PROCESS RISK} \\ &= \text{TOTAL RISK}\,. \end{aligned}$$

In Theorems 10 and 11 are derived the following recursive formulas for exact values of these two variance components.

PARAMETER RISK

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

where

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

is the average "x value," and I_n=N-n+1 (assuming a full column in the triangle) is the number of data points, in the regression estimate of the n^{th} link ratio.

PROCESS RISK

$$\mathrm{Var}(E_1) = \sigma_1^2$$

and

$$\mathrm{Var}(E_n) = n\sigma_n^2 + b_n^2\,\mathrm{Var}(E_{n-1}) \quad .$$

The equation for Process Risk is the same under both models. Unbiased estimates of these variances are obtained by plugging in unbiased estimates s_n^2 for σ_n^2, $\hat{b}_n$ for $\mathbf{b}_n$, and $\hat{M}_n$ for M_n.

For the example, Figure 6B shows the calculation of the estimate of the variance of the estimate of total ultimate loss for accident years 1984 through 1991 combined. Most of the basic statistics are the same those appearing in Figure 6A.

Variance of Estimated Outstanding Losses: Single or Multiple Accident Years

Assuming paid losses are constant at any given evaluation,[12] it is obvious that the variance of a reserve equals the variance of total ultimate losses:

$$\begin{aligned}\mathrm{Var}(\textit{Estimated Reserve}) &= \mathrm{Var}(\textit{Estimated Ultimate Loss} - \text{Paid Loss})\\ &= \mathrm{Var}(\textit{Estimated Ultimate Loss}) \quad .\end{aligned}$$

This equality holds for estimated reserves for a single accident year and for multiple accident years.

Variance of Estimated Pure Premiums and Loss Ratios: Single or Multiple Accident Years

Assuming static exposures and pure premiums, the variances of the estimated pure premium rate and of the estimated loss ratio are

$$\frac{\mathrm{Var}(\textit{Estimated Ultimate Loss})}{\text{exposures}^2} \quad \text{and} \quad \frac{\mathrm{Var}(\textit{Estimated Ultimate Loss})}{\text{earned premiums}^2} \quad .$$

Again, these formulas hold for single or multiple accident years.

One final note before leaving this section. Aggregate losses are often expressed as the compound product of a frequency distribution (e.g., Poisson or negative binomial) and a severity distribution (e.g., lognormal or Pareto). In practice, parameters for those distributions are estimates, the result being that the variance of the aggregate loss distribution depends not only on the inherent variance of the postulated frequency and severity distributions but on the variance of the parameter estimates. The parameter error of the frequency distribution could be estimated by applying the above techniques to the frequency triangle, defined to be the triangle of claim counts per exposure. The parameter error of the severity distribution could be estimated by applying the above techniques to the incurred (or paid) severity triangle, defined as the triangle of cumulative incurred (paid) dollars divided by cumulative incurred (paid) claims. Furthermore, since it is the mean of the distributions that are usually sought, only the Parameter Risk above need be considered.

---

[12] Salvage and subrogation could be handled as a separate category.

4. CONFIDENCE INTERVALS

Confidence intervals necessarily are phrased in terms of a probability measure. As a result, 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 (lognormally distributed under GAD which may be a bit more believable).

Confidence Intervals Around the Link Ratios

Let α be the probability measurement of the width of the confidence interval. Then $100\alpha\%$ confidence intervals around the true LSL link ratios $(\mathbf{a}_n, \mathbf{b}_n)$ are:

$$\hat{a}_n \pm t_{\frac{\alpha}{2}}(I_n-2)\sqrt{\hat{Var}(\hat{a}_n)}$$

and

$$\hat{b}_n \pm t_{\frac{\alpha}{2}}(I_n-2)\sqrt{\hat{Var}(\hat{b}_n)}$$

where $t_\alpha(I_n-2)$ denotes the two-tailed α point (the "t-value")[13] of Student's t distribution with I_n-2 degrees of freedom and where I_n is the number of accident years used in the estimate of the n^{th} link ratio. The degrees of freedom under LSL is I_n-2 because two parameters are estimated under that model. These formulas may be used for the LSM model as well; in that case the degrees of freedom are I_n-1.

To demonstrate how these formulas can be used, suppose we want to test the hypothesis that the 12-24 month LSL constant term is not significantly greater than zero. Recall that this constant term was estimated to be \$373.63. Refer to Figure 7. There are 18 data points and two parameters, so the degrees of freedom equals 16. At the 99% confidence level, the one-tailed t-value is 2.62. It was shown above that the estimated variance of the constant term is 77.35^2. Then, if the constant term were truly zero, there would be a 99% chance that the estimated intercept would be less than or equal to 202.66 = 77.35 x 2.62. Since the estimated value of the intercept falls outside the confidence interval, it appears that LSL is an appropriate model for this young stage of development. In fact, it appears that LSL is appropriate for the youngest six stages of development. The confidence of that statement is $94\% = .99^6$.

As another example, the decision to assume that case reserves are reasonably adequate by 108 months is based on the apparent random nature of the link ratios thereafter. Notice in Figure 7 that the LSM link ratios are either at, or well within, one standard deviation ("Std(b)") of unity for 120-132 months and beyond, but the 108-120 link ratio (.992) is more than one

---

[13] This assumes that the available t-table is presented in terms of a one-tailed test, or, if not, that the actuary is able to look up the appropriate value accordingly.

standard deviation away from unity. Somewhat subjectively, it was deemed appropriate to ignore this significant average negative development, as well as the relatively insignificant positive development thereafter. If the actuary were to set the 108-120 factor to, say, an interpolated value between the 96-108 and 120-132 factors, it may generally be considered a not unreasonable application of actuarial judgment and may just so happen to reflect an amount of conservativism consistent with the risk posture of the owners of the enterprise. However, in the end, the ability of that actuary to convince management that this judgment is appropriate depends on the level of trust established between the parties.

It is clear that 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 the variances of the residuals in the development model are the same for all development periods (i.e., $\sigma_i = \sigma_j$ for all i and j), then a single estimate of the MSE can be obtained by solving for all link ratios simultaneously. The result is that the t-value should become reasonably small[14] and can make for "tighter" inferences for all development periods.

Confidence Intervals Around Estimated Ultimate Loss

This section will begin with the GAD model because all results are exact.[15] Under the transformed GAD model

$$\ln(x_n) = \ln(\mathbf{b}_n) + \ln(x_{n-1}) + \ln(e_n)$$

or

$$x_n' = b_n' + x_{n-1}' + e_n'$$

the point estimate of ultimate transformed loss is

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

and the estimate of the variance of the prediction is

[14] For an NxN triangle, df=(N-1)(N-2)/2 under LSM if no data points are discarded. For example, with a moderately-sized 5x5 triangle the two-tailed 90%-ile t-value is only 18% greater than the smallest possible 90%-ile t-value, namely the 90%-ile 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.

[15] Commonly used probability distributions are location oriented, so additive models such as the transformed GAD model are quite tractable. The use of scale-oriented probability distributions may yield results more directly applicable to the multiplicative models actuaries favor.

$$\hat{Var}(pred') = (C + \sum_{j=1}^{\mu} \frac{1}{I_j}) s'^2 \qquad \text{(Theorem 13)}$$

where we assume under transformed GAD that all σ_n's are equal. It is well known that in this case the MSE is proportional to a chi-square random variable with degrees of freedom equal to the number of data points less the number of estimated parameters. Therefore, a one-sided $100\alpha\%$ confidence interval[16] for ultimate transformed losses x_C' given the current transformed evaluation x_0' is exactly equal to

$$\hat{\mu}_C' \pm t_\alpha(df)\sqrt{\hat{Var}(pred')} \quad .$$

The corresponding $100\alpha\%$ confidence interval around the "untransformed" prediction of ultimate loss x_C given x_0 is

$$\exp\left(\hat{\mu}_C' \pm t_\alpha(df)\sqrt{\hat{Var}(pred')}\right) \quad .$$

If df is large enough, $t_\alpha(df)$ may be replaced by z_α, the standard normal point, without significant loss in accuracy.[17]

With this justification, an *approximate* $100\alpha\%$ confidence interval around a prediction under any of the models is

$$pred \pm t_{\frac{\alpha}{2}}(df)\sqrt{\hat{Var}(pred)} \quad .$$

Figures 6A and 6B show how this approach is used to derive estimates of ultimate loss at the 80% confidence level.

<u>*Confidence Intervals around Reserves*</u>

Confidence intervals around reserves are obtained by subtracting paid dollars from the endpoints of the confidence intervals around ultimate loss. This is simply due to the fact that if

---

[16] At the risk of pedantry, "prediction interval" is more correct.

[17] This is often done in practice, particularly in time series analysis, even when df is not 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 Everette S. Gardner Jr., "A Simple Method of Computing Prediction Intervals for Time Series Forecasts," *Management Science*, Vol. 34, No. 4, April 1988, p. 541-546.

$$\alpha = P(\text{lower bound} \leq \text{ultimate loss} \leq \text{upper bound})$$

then it is also true that

$$\alpha = P(\text{lower bound - paid} \leq \text{outstanding loss} \leq \text{upper bound - paid}).$$

Figures 6A and 6B demonstrate the application of this concept as well. The 80% risk load for all outstanding losses for accident years 1984 through 1991 is about 27% of the expected value. It would be interesting to see how much this load is reduced for the same level of confidence when an analysis of paid dollar triangles is also conducted. Incurred and paid estimates should be negatively correlated, therefore the variance of their average should be reduced even more than if independence were simply assumed.

Confidence Intervals around Pure Premiums and Loss Ratios

Confidence intervals around pure premiums and loss ratios are obtained by dividing the endpoints of the confidence intervals around ultimate loss by exposures or premiums, respectively. This scale shift is akin to the location shift for the confidence intervals around reserves.

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

When the current evaluation is zero, the practice is to abandon the multiplicative loss development methods and adopt an alternative, e.g., Bornhuetter-Ferguson, Stanard-Buhlmann, or a variation on frequency-severity. LSL may be a fourth possibility.

To elaborate, consider the development of reported claim counts. Let N be the true ultimate number of claims for a given accident year. Let r_i be the random report year of the i^{th} claim. Assume that the r_i are independent and identically distributed for all claims so that if p_n is the probability that a claim is reported before the end of the n^{th} year, then p_n is independent of i. Based on these assumptions it is not difficult to show that if x_n is the number of reported claims at the n^{th} evaluation then

$$E(x_n | x_{n-1}) = N\frac{p_n - p_{n-1}}{1-p_{n-1}} + \frac{1-p_n}{1-p_{n-1}}x_{n-1} \qquad (2)$$

which is of the form $\mathbf{a}_n+\mathbf{b}_n x_{n-1}$. Clearly the constant term $\mathbf{a}_n$ is non-zero until all claims are reported.

Figure 8A shows the true development line for evaluation 1 to evaluation 2 when N=40 and the p_n's are 1/2, 3/4, 7/8, ... , $1-½^n$, along with a scattering of ten random data points.

Equation (2) becomes even more interesting when the reporting pattern is exponential, as might be expected from a Poisson frequency process. In that case it is straightforward to prove that the LSL coefficients $(\mathbf{a}_n,\mathbf{b}_n)$ are identical for every age n. This somewhat surprising result can be put to good use when the triangle is too small to give stable LSL estimates of individual link ratios, as will be demonstrated in the following section.

From Equation (2) one can see that the slope factor $\mathbf{b}_n$ does not depend on the exposure (N) but only on the reporting pattern, and that the constant term $\mathbf{a}_n$ is proportional to the exposure. An increase in exposure from one accident year to the next will cause an upward, parallel shift in the development regression line. Equation (2) may also be used as a paradigm for loss dollars, although the bias of case reserves complicates the analysis, and systematic factors such as trend can change expected ultimate loss dollars from one accident year to the next. Development triangles, therefore, can be expected to display data samples randomly distributed about not a single regression line but about multiple parallel regression lines as claim frequency increases, as the volume of business expands, or simply through the impact of trend. This is pictured in Figure 8B where a random sample is displayed about the regression line of Figure 8A and about a parallel line determined by N=80. The estimated regression line based on all the points combined will indicate a less significant constant term.

6. COMPARING THE MODELS USING SIMULATION

In the 1985 *Proceedings* Mr. James N. 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[18] it was demonstrated that WAD loss development was clearly inferior to three additive methods — Bornhuetter-Ferguson (BF), Stanard-Buhlmann[19] (SB), and a little-used method called the Additive Model (ADD) — because it had greater average bias and a larger variance. The 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 based on inverted LDFs and are therefore nonlinear functions of the y's.

Stanard's simulation was replicated here to test additionally the accuracy of LSM, LSL, SAD and GAD.[20] 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, by "ultimate loss" is meant 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 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

---

[18] 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.

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

[20] For the details behind the computer model, the reader is referred to Mr. Stanard's published results. The simulation was reproduced in C on an IBM PS/2 Model 70 with a math coprocessor. The most complicated scenarios requiring 15000 iterations took about an hour and a half to process.

becomes a better predictor.

In Exhibits I through IV, the average bias and standard deviation of the first accident year are zero because the simulation defines ultimate to be the current age of that accident year.

Exhibit I: Claim Counts Only

In this case, 5000 claim count triangles were simulated, the "actual ultimate" as of the last column was simulated as well, accident year ultimates were estimated using the various methods, and averages and standard deviations of the prediction errors were calculated.

Of the multiplicative estimators, LSM has the smallest bias and the smallest variance for every accident year. As can be expected, WAD is close behind. The remaining methods could perhaps be ordered BF, SB, ADD, and LSL, in increasing order of accuracy as measured by the standard deviation of the accident-year-total projection.

Consider first the average bias. In Figure 9A is graphed the relationship between incurred counts at 12 months, x, with incurred losses at 24 months, y, which we know from the previous section 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 9A, 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 I.

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 $E(y)-E(x)$ via $\bar{y}-\bar{x}$ calculated from prior accident years. Since the environment in Exhibit 1 — exposure, frequency, trend, etc. — does not change by accident year, the average of 5000 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 $\frac{\bar{y}}{\bar{x}}$ is infinity. The average of 5000 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.[21]

[21] 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

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 9A 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 average and overstate expected development when the current evaluation is greater than average.

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 9A 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:

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

WAD yields biased estimates, see [Stanard].

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 counterintuitive 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 II: Random Severity, No Trend

In this case, 5000 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 III: Random Severity, 8% Severity Trend Per Year

This is where it gets interesting. This could be considered the standard 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 with $\alpha = \frac{1}{2}$,[22] 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 9B 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. The upward trend makes it more likely that the estimated LSL intercept will be less than zero, 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 IV: Random Severity, 8% Trend, On-Level Triangle

[22] Robert P. Butsic 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, p. 58.

In this case, rows of the triangle were trended to the level of the most recent accident year assuming that the research department is perfectly prescient in its estimate of past trend. For most of the models the total bias decreases 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.

7. CONCLUSION

Loss development predictions can be improved by the use of least squares estimators. In certain situations the least squares estimators coincide with the more traditional simple average development and geometric average development estimators. Under the four sets of assumptions about the loss development process considered here, the weighted average link ratio estimator is always inferior to an alternative, least squares estimator.

If the assumptions of a given model considered here can be married with the independence assumption that forms the basis of the Chain Ladder Method, the developed estimates of ultimate loss are unbiased. The variance of estimated ultimate loss can itself be estimated through relatively straightforward application of recursive formulas. A range of estimates can be given with associated approximate levels of confidence if one is willing to make some assumptions about the probability distribution of the error terms.

At this point, statistical techniques may be of some guidance in selecting one model over another, but the final choice of the most appropriate set of assumptions will probably be a judgment call depending on, among other things, the exposure and the claims operation of the book of business.

The simulation study suggests that the performance of the more general Least Squares Linear method exceeds that of the multiplicative development methods and may, in some situations, rival that of the nonlinear additive methods in common use today. It would be interesting to investigate the correlation between development estimates of ultimate loss based on incurred and paid triangles, and use that information to derive optimal, variance-minimizing weights for making final selections.

Figure 1A

Auto Liability
Incurred Loss + ALAE ($000)

	Age of Accident Year (Months)																		
A/Y	12	24	36	48	60	72	84	96	108	120	132	144	156	168	180	192	204	216	228
1973	45	111	134	149	163	164	165	165	165	165	165	165	165	165	165	165	165	165	165
1974	38	92	105	123	315	315	338	340	341	345	350	442	518	558	536	536	536	536	
1975	79	195	240	447	513	738	757	767	792	797	797	797	797	797	797	797	797		
1976	103	442	901	1,480	2,325	2,372	2,378	2,549	2,566	2,466	2,469	2,480	2,480	2,480	2,480	2,480			
1977	58	274	706	1,555	1,605	1,719	1,785	1,741	1,604	1,604	1,604	1,604	1,604	1,604	1,604				
1978	50	405	1,371	1,193	1,229	1,371	1,524	1,518	1,510	1,516	1,517	1,517	1,517	1,517					
1979	65	1,228	1,371	1,277	1,378	1,398	1,552	2,011	2,026	2,012	2,020	2,020	2,020						
1980	561	1,210	1,501	1,641	1,870	1,945	1,967	1,985	2,174	2,175	2,175	2,175							
1981	555	1,469	1,517	1,626	1,652	1,749	1,764	1,764	1,764	1,764	1,764								
1982	444	856	1,107	1,404	1,730	1,948	2,029	2,085	2,103	2,122									
1983	441	977	834	988	1,122	1,297	1,476	1,476	1,480										
1984	458	1,393	1,790	2,149	2,387	2,530	2,492	2,446											
1985	1,262	4,472	5,758	6,239	6,481	6,718	6,856												
1986	3,128	6,399	8,675	9,390	9,780	10,204													
1987	3,556	6,099	7,139	7,766	7,947														
1988	1,969	3,010	3,928	3,927															
1989	1,260	3,374	3,930																
1990	764	2,795																	
1991	1,287																		

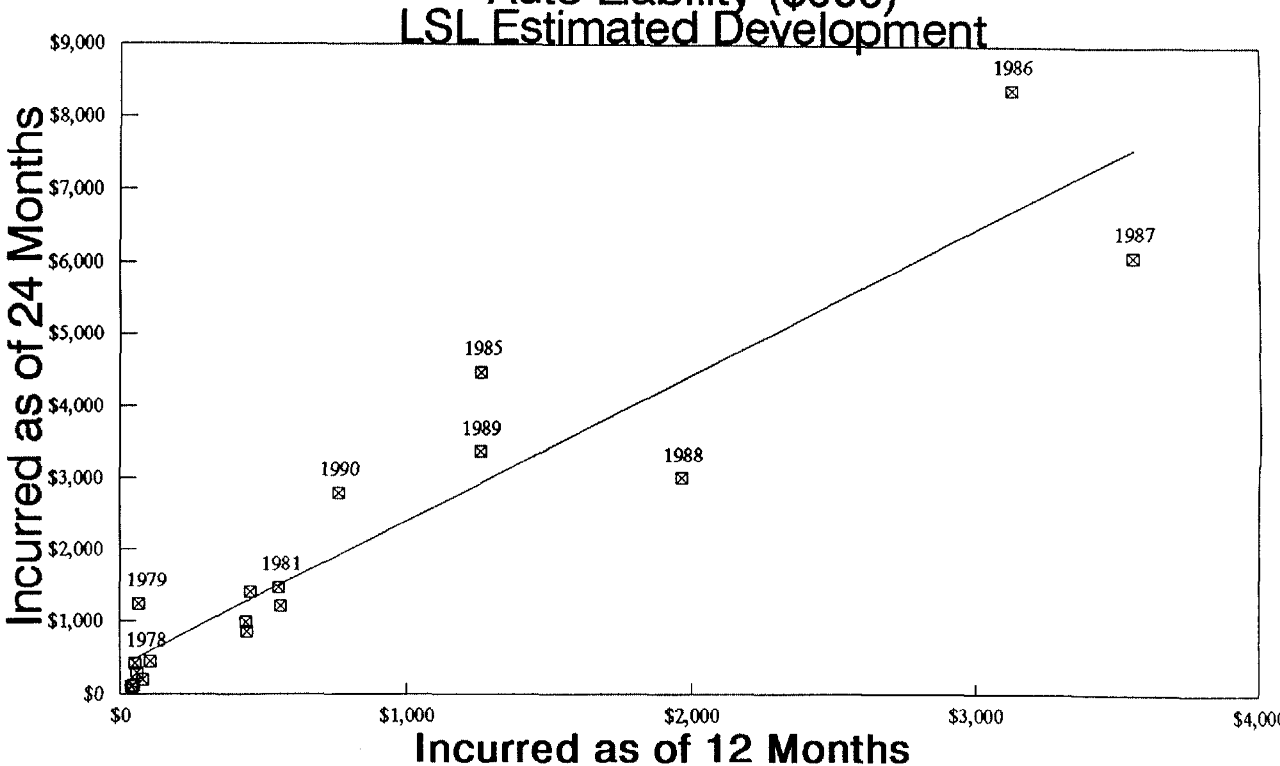
Figure 1B

Auto Liability
Incurred Loss + ALAE Development

Development Period

A/Y	12-24	24-36	36-48	48-60	60-72	72-84	84-96	96-108	108-120	120-132	132-144	144-156	156-168	168-180	180-192	192-204	204-216	216-228
1973	2.467	1.207	1.112	1.094	1.006	1.006	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1974	2.421	1.141	1.171	2.561	1.000	1.073	1.006	1.003	1.012	1.014	1.263	1.172	1.077	0.961	1.000	1.000	1.000	
1975	2.468	1.231	1.863	1.148	1.439	1.026	1.013	1.033	1.006	1.000	1.000	1.000	1.000	1.000	1.000	1.000		
1976	4.291	2.038	1.643	1.571	1.020	1.003	1.072	1.007	0.961	1.001	1.004	1.000	1.000	1.000	1.000			
1977	4.724	2.577	2.203	1.032	1.071	1.038	0.975	0.921	1.000	1.000	1.000	1.000	1.000	1.000				
1978	8.100	3.385	0.870	1.030	1.116	1.112	0.996	0.995	1.004	1.001	1.000	1.000	1.000					
1979	18.892	1.116	0.931	1.079	1.015	1.110	1.296	1.007	0.993	1.004	1.000	1.000						
1980	2.157	1.240	1.093	1.140	1.040	1.011	1.009	1.095	1.000	1.000	1.000							
1981	2.647	1.033	1.072	1.016	1.059	1.009	1.000	1.000	1.000	1.000								
1982	1.928	1.293	1.268	1.232	1.125	1.043	1.028	1.009	1.009									
1983	2.215	0.854	1.185	1.136	1.156	1.138	1.000	1.003										
1984	3.041	1.285	1.201	1.111	1.060	0.985	0.982											
1985	3.544	1.288	1.084	1.039	1.037	1.021												
1986	2.685	1.033	1.082	1.042	1.043													
1987	1.715	1.171	1.088	1.023														
1988	1.529	1.305	1.000															
1989	2.678	1.165																
1990	3.658																	
<u>Averages</u>																		
SAD	3.953	1.433	1.242	1.217	1.085	1.044	1.031	1.007	0.999	1.002	1.033	1.025	1.013	0.992	1.000	1.000	1.000	1.000
WAD	2.480	1.206	1.116	1.082	1.059	1.034	1.035	1.008	0.996	1.002	1.010	1.010	1.007	0.998	1.003	1.007	1.013	1.000
GAD	3.129	1.340	1.203	1.177	1.080	1.043	1.028	1.006	0.998	1.002	1.030	1.023	1.012	0.992	1.000	1.000	1.000	1.000

Figure 2A



Auto Liability ($000)
LSL Estimated Development
Incurred as of 24 Months
Incurred as of 12 Months
$9,000
$8,000
$7,000
$6,000
$5,000
$4,000
$3,000
$2,000
$1,000
$0
$1,000
$2,000
$3,000
$4,000
1986
1987
1985
1989
1988
1990
1981
1979
1978

Figure 2B

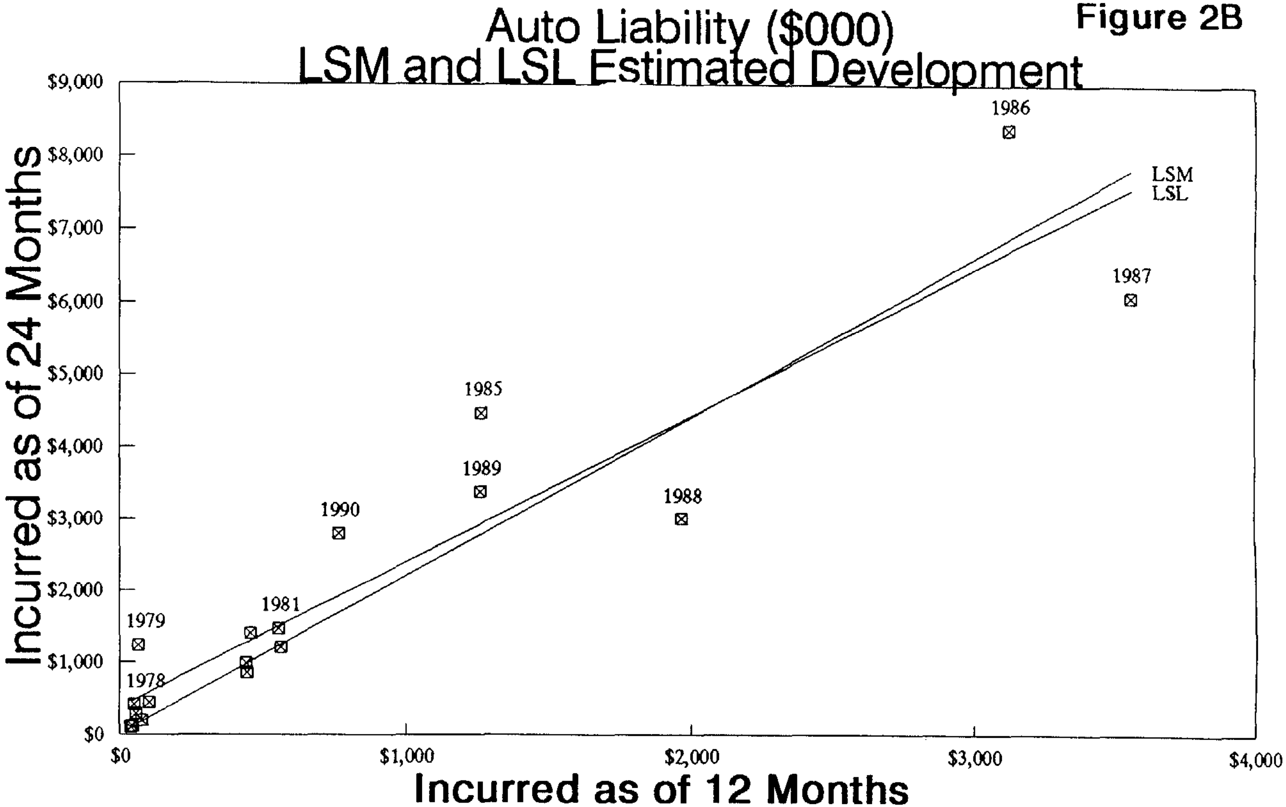

Auto Liability ($000)
LSM and LSL Estimated Development
Incurred as of 24 Months
Incurred as of 12 Months
$0
$1,000
$2,000
$3,000
$4,000
$5,000
$6,000
$7,000
$8,000
$9,000
$0
$1,000
$2,000
$3,000
$4,000
1978
1979
1981
1985
1986
1987
1988
1989
1990
LSM
LSL

Figure 3

Notation for the "Known" and "Completed" 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	$x_{0,0}$	$x_{0,1} \mid x_{0,0}$	$x_{0,2} \mid x_{0,0}$	⋯	$x_{0,n-1} \mid x_{0,0}$	$x_{0,n} \mid x_{0,0}$	$x_{0,n+1} \mid x_{0,0}$	⋯	$x_{0,N-1} \mid x_{0,0}$	$x_{0,N} \mid x_{0,0}$

The shaded area highlights $S_n = \sum_{i=0}^{n-1} x_{i,n} \mid x_{i,i}$. $M_n = E(S_n)$.

Figure 4

Auto Liability
Incurred Loss + ALAE ($000)

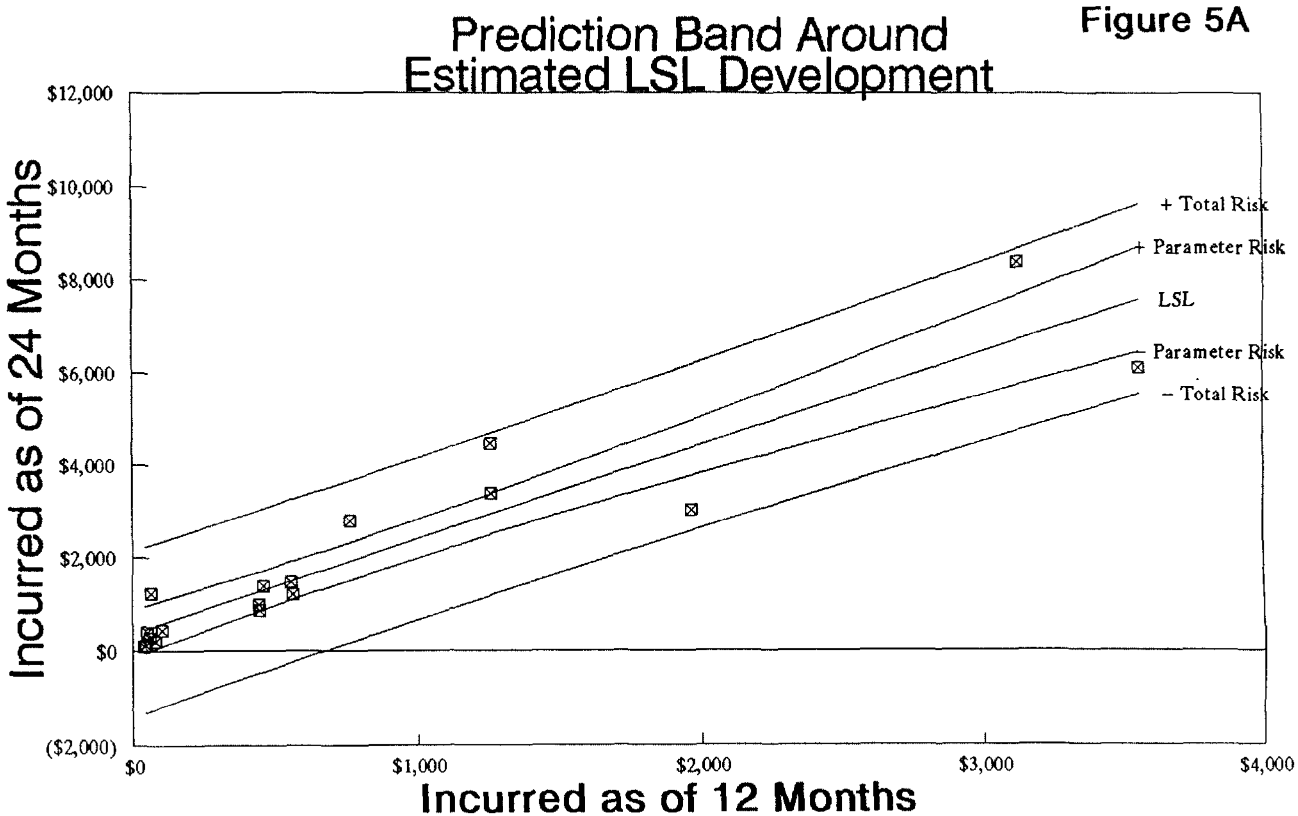
Completed Triangle

Age of Accident Year (Months)

A/Y	12	24	36	48	60	72	84	96	108	120	132	144	156	168	180	192	204	216	228
1973	45	111	134	149	163	164	165	165	165	165	165	165	165	165	165	165	165	165	165
1974	38	92	105	123	315	315	338	340	341	345	350	442	518	558	536	536	536	536	536
1975	79	195	240	447	513	738	757	767	792	797	797	797	797	797	797	797	797	797	797
1976	103	442	901	1,480	2,325	2,372	2,378	2,549	2,566	2,466	2,469	2,480	2,480	2,480	2,480	2,480	2,480	2,480	2,480
1977	58	274	706	1,555	1,605	1,719	1,785	1,741	1,604	1,604	1,604	1,604	1,604	1,604	1,604	1,604	1,604	1,604	1,604
1978	50	405	1,371	1,193	1,229	1,371	1,524	1,518	1,510	1,516	1,517	1,517	1,517	1,517	1,517	1,517	1,517	1,517	1,517
1979	65	1,228	1,371	1,277	1,378	1,398	1,552	2,011	2,026	2,012	2,020	2,020	2,020	2,020	2,020	2,020	2,020	2,020	2,020
1980	561	1,210	1,501	1,641	1,870	1,945	1,967	1,985	2,174	2,175	2,175	2,175	2,175	2,175	2,175	2,175	2,175	2,175	2,175
1981	555	1,469	1,517	1,626	1,652	1,749	1,764	1,764	1,764	1,764	1,764	1,764	1,764	1,764	1,764	1,764	1,764	1,764	1,764
1982	444	856	1,107	1,404	1,730	1,946	2,029	2,085	2,103	2,122	2,122	2,122	2,122	2,122	2,122	2,122	2,122	2,122	2,122
1983	441	977	834	988	1,122	1,297	1,476	1,476	1,480	1,480	1,480	1,480	1,480	1,480	1,480	1,480	1,480	1,480	1,480
1984	458	1,393	1,790	2,149	2,387	2,530	2,492	2,446	2,468	2,468	2,468	2,468	2,468	2,468	2,468	2,468	2,468	2,468	2,468
1985	1,262	4,472	5,758	6,239	6,481	6,718	6,858	7,075	7,139	7,139	7,139	7,139	7,139	7,139	7,139	7,139	7,139	7,139	7,139
1986	3,128	8,399	8,675	9,390	9,780	10,204	10,360	10,691	10,787	10,787	10,787	10,787	10,787	10,787	10,787	10,787	10,787	10,787	10,787
1987	3,556	6,099	7,139	7,766	7,947	8,275	8,410	8,679	8,757	8,757	8,757	8,757	8,757	8,757	8,757	8,757	8,757	8,757	8,757
1988	1,969	3,010	3,928	3,927	4,155	4,354	4,446	4,588	4,629	4,629	4,629	4,629	4,629	4,629	4,629	4,629	4,629	4,629	4,629
1989	1,260	3,374	3,930	4,288	4,522	4,734	4,829	4,984	5,028	5,028	5,028	5,028	5,028	5,028	5,028	5,028	5,028	5,028	5,028
1990	764	2,795	3,268	3,589	3,811	3,999	4,086	4,217	4,255	4,255	4,255	4,255	4,255	4,255	4,255	4,255	4,255	4,255	4,255
1991	1,287	2,982	3,470	3,802	4,028	4,223	4,313	4,451	4,491	4,491	4,491	4,491	4,491	4,491	4,491	4,491	4,491	4,491	4,491

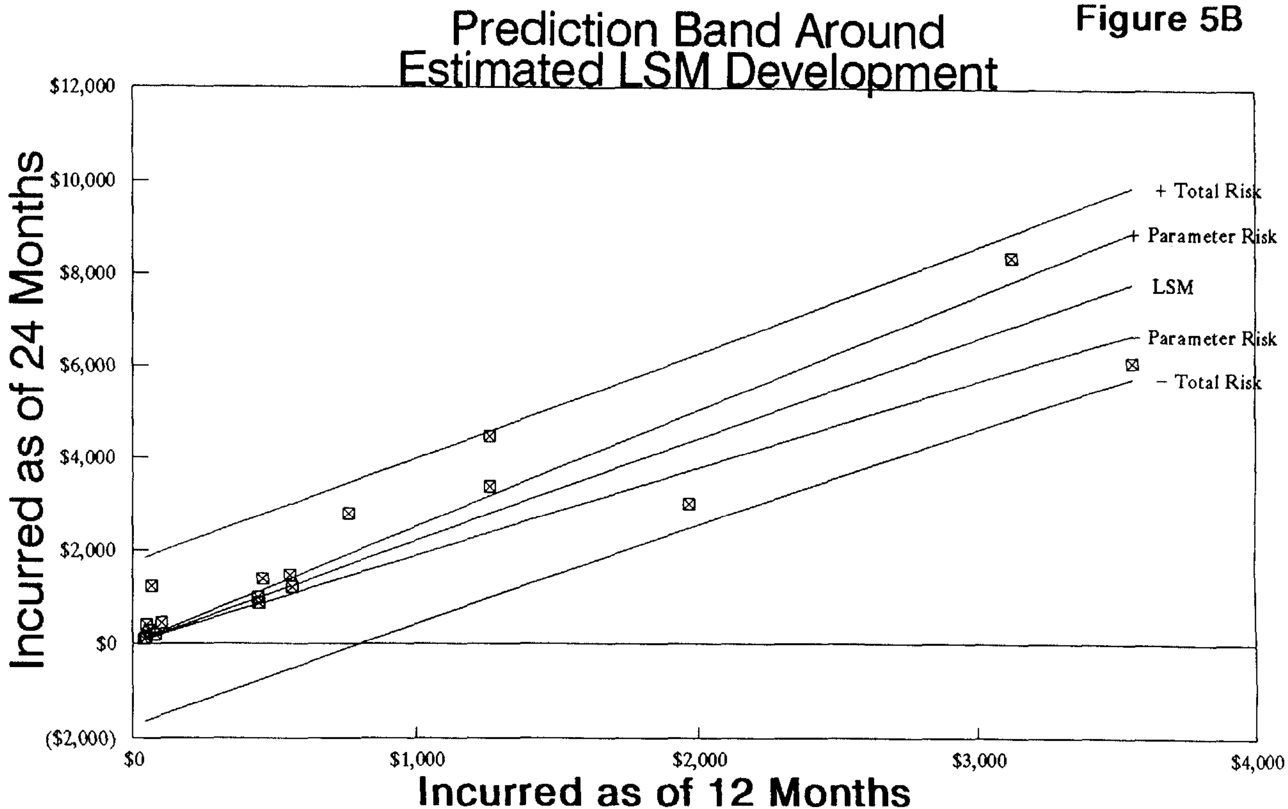
Least Squares Development Coefficients

a	373.63	255.26	137.50	161.37	58.01	43.37	0	0	0	0	0	0	0	0	0	0	0	0
b	2.027	1.078	1.056	1.017	1.034	1.011	1.032	1.009	1	1	1	1	1	1	1	1	1	1



Radius of band = 2 standard deviations
Associated confidence level = 93% (approx.)

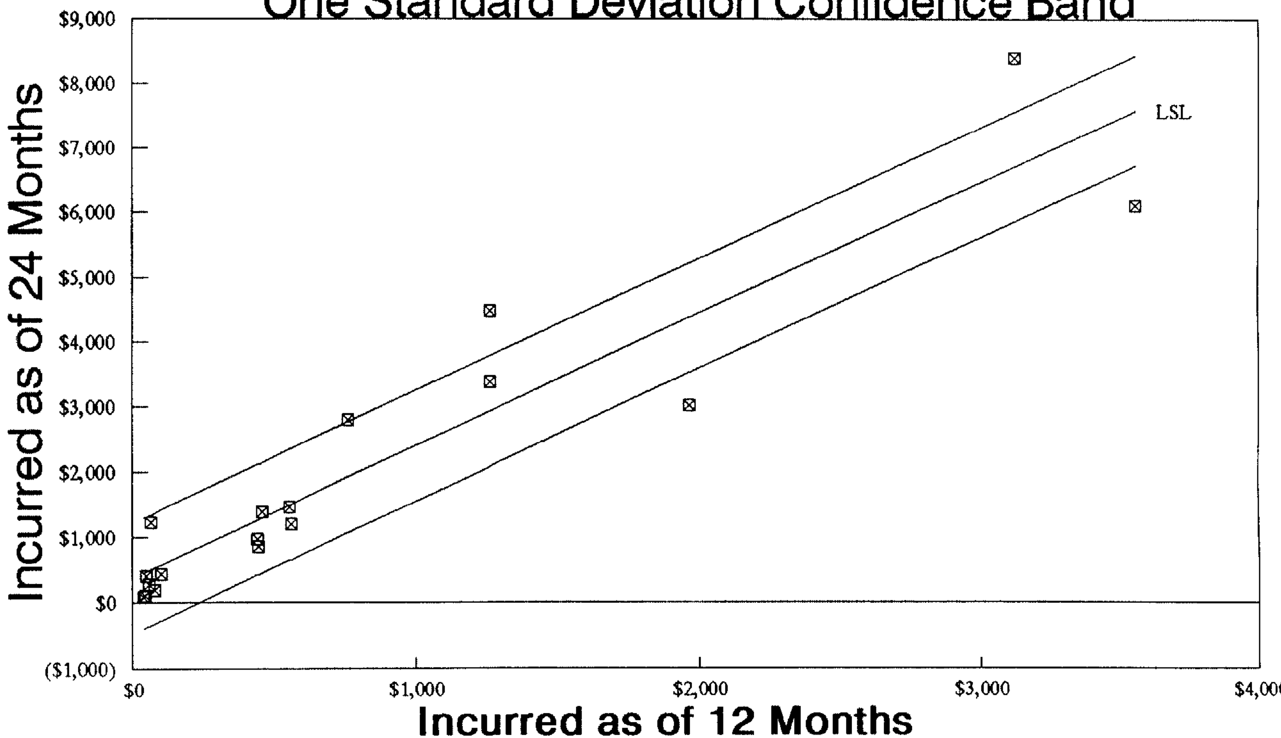
Figure 5B



Radius of band = 2 standard deviations
Associated confidence level = 93% (approx.)

Figure 5C

Identifying Outliers
One Standard Deviation Confidence Band



Radius of band = 1 standard deviation

Associated confidence level = 66% (approx.)

Figure 6A

Auto Liability
Variance of Estimated Ultimate Loss
Accident Year 1991

Age (months)	12	24	36	48	60	72	84	96	Ultimate 108	120	132	Closed 144	Paid to Date	Indicated Reserve
n	0	1	2	3	4	5	6	7	8	9	10	11		
mu hat	$1,287	$2,982	$3,470	$3,802	$4,028	$4,223	$4,313	$4,451	$4,491	$4,491	$4,491	$4,491	$202	$4,289
x bar	824	2,000	2,317	2,495	2,325	1,866								
b hat		2.027	1.078	1.056	1.017	1.034	1.011	1.022	1.032	1	1	1		
Var(b hat)		0.0377	0.0017	0.0007	0.0004	0.0001	0.0002	0.0035	0.0006	0	0	0		
s		849	384	278	212	76	72	146	139	73	31	3		
I		18	17	16	15	14	13	12	13	12	11	10		
df		16	15	14	13	12	11	10	11	10	9	8		
Parameter Risk		48,112	66,266	79,751	86,179	92,745	96,123	1.66E+05	1.88E+05	1.88E+05	1.88E+05	1.88E+05		
Process Risk		7.20E+05	9.85E+05	1.18E+06	1.26E+06	1.35E+06	1.39E+06	1.47E+06	1.59E+06	1.59E+06	1.59E+06	1.59E+06		
Total Risk		7.69E+05	1.05E+06	1.26E+06	1.35E+06	1.45E+06	1.48E+06	1.64E+06	1.77E+06	1.78E+06	1.78E+06	1.78E+06		
Std Dev		876.671	1,025.227	1,120.290	1,160.473	1,202.590	1,218.497	1,279.595	1,332.182	1,334.199	1,334.566	1,334.569		

	Closed 144	Paid to Date	Indicated Reserve
Total df	129		
One-tailed 80% t-value (same as standard normal when df=129)	0.842		
80% Confidence Risk Load (t-value x Std Dev)	$1,124		$1,124
Upper bound on 80% Confidence Interval for Ultimate Loss	$5,615	$202	$5,413
80% Confidence Risk Load as a Percent of the Expected Value	25%		26%

Figure 6B

Auto Liability
Variance of Estimated Ultimate Loss
Accident Years 1984 – 1991

Age (months)	12	24	36	48	60	72	84	96	Ultimate 108	120	132	Closed 144	Total Paid to Date	Indicated Total Reserve
n	0	1	2	3	4	5	6	7	8	9	10	11		
M hat	$1,287	$2,982	$6,739	$11,678	$16,516	$25,585	$36,443	$44,685	$47,555	$47,555	$47,555	$47,555	$37,854	$9,701
x bar	824	2,000	2,317	2,495	2,325	1,866								
b hat		2.027	1.078	1.056	1.017	1.034	1.011	1.022	1.032	1	1	1		
Var(b hat)		0.0377	0.0017	0.0007	0.0004	0.0001	0.0002	0.0035	0.0006	0	0	0		
s		849	384	278	212	76	72	146	139	73	31	3		
I		18	17	16	15	14	13	12	13	12	11	10		
df		16	15	14	13	12	11	10	11	10	9	8		
PARAMETER RISK		48,112	95,937	1.61E+05	2.27E+05	2.63E+05	3.83E+05	5.70E+06	7.35E+06	7.35E+06	7.35E+06	7.35E+06		
PROCESS RISK		7.20E+05	1.13E+06	1.49E+06	1.72E+06	1.87E+06	1.95E+06	2.18E+06	2.48E+06	2.53E+06	2.54E+06	2.54E+06		
TOTAL RISK		7.69E+05	1.23E+06	1.65E+06	1.95E+06	2.14E+06	2.33E+06	7.88E+06	9.82E+06	9.87E+06	9.88E+06	9.88E+06		
Std Dev		876.671	1,108.321	1,286.375	1,396.973	1,461.640	1,525.937	2,806.755	3,134.438	3,142.149	3,143.706	3,143.721		
Total df												129		
One–tailed 80% t–value (same as standard normal when df=129)												0.842		
80% Confidence Risk Load (t–value x Std Dev)												$2,647		$2,647
Upper bound on 80% Confidence Interval for Ultimate Loss												$50,202	$37,854	$12,348
80% Confidence Risk Load as a Percent of the Expected Value												6%		27%

Figure 7

Auto Liability
Incurred Loss + ALAE Development

Estimated Least Squares Development Coefficients

	12-24	24-36	36-48	48-60	60-72	72-84	84-96	96-108	108-120	120-132	132-144	144-156	156-168	168-180	180-192	192-204	204-216	216-228
LSL																		
a	373.63	255.26	137.50	161.37	58.01	43.37	18.67	-8.51										
b	2.027	1.078	1.056	1.017	1.034	1.011	1.022	1.013										
s	848.8	384.204	277.64	211.942	76.0792	72.0653	145.832	77.1915										
df	16	15	14	13	12	11	10	9										
Std(a)	77.35	39.7946	31.7267	25.6207	10.0074	13.4745	67.418	36.9558										
Std(b)	0.194	0.04063	0.02726	0.01978	0.00802	0.01281	0.0591	0.03224										
One-tailed 99% t-value	2.62	2.65	2.68	2.72	2.76	2.82	2.90	3.00										
Upper bound on a	202.66	105.46	85.03	69.69	27.62	38.00	195.51	110.87										
LSM																		
b	2.204	1.133	1.083	1.048	1.045	1.024	1.032	1.009	0.992	1.001	1.003	1.002	1.002	0.999	1.000	1.000	1.000	1.000
s	876.5	421.549	288.053	238.949	85.5021	74.8248	139.291	73.3336	31.2845	2.93139	34.6692	30.8364	17.687	10.822	0.000	0.000		
df	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Std(b)	0.157	0.0336	0.02092	0.01656	0.00664	0.00853	0.02397	0.01335	0.00585	0.0006	0.00764	0.00772	0.0051	0.0035	0.000	0.000		

Figure 8A

Claim Count Development
Expected Number of Claims = 40

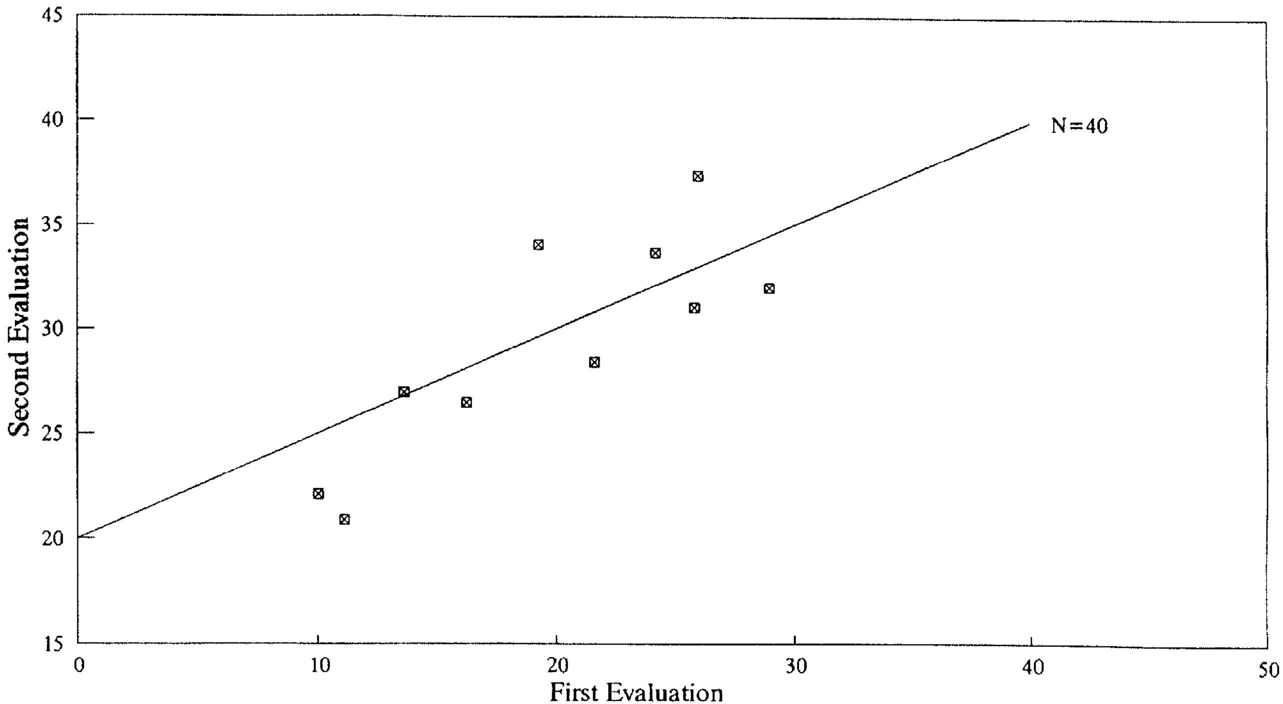


Figure 8B

Claim Count Development
Expected Number of Claims = 40 & 80

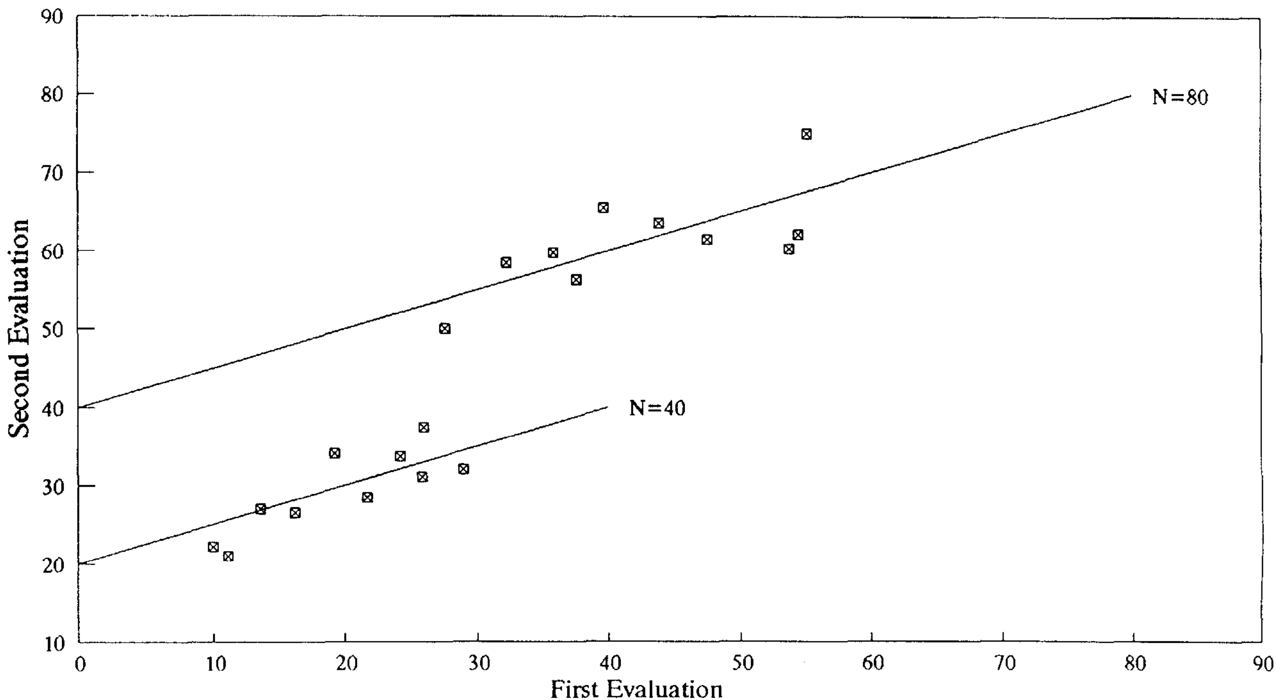


Figure 9A

Idealized Development Estimators
No Trend

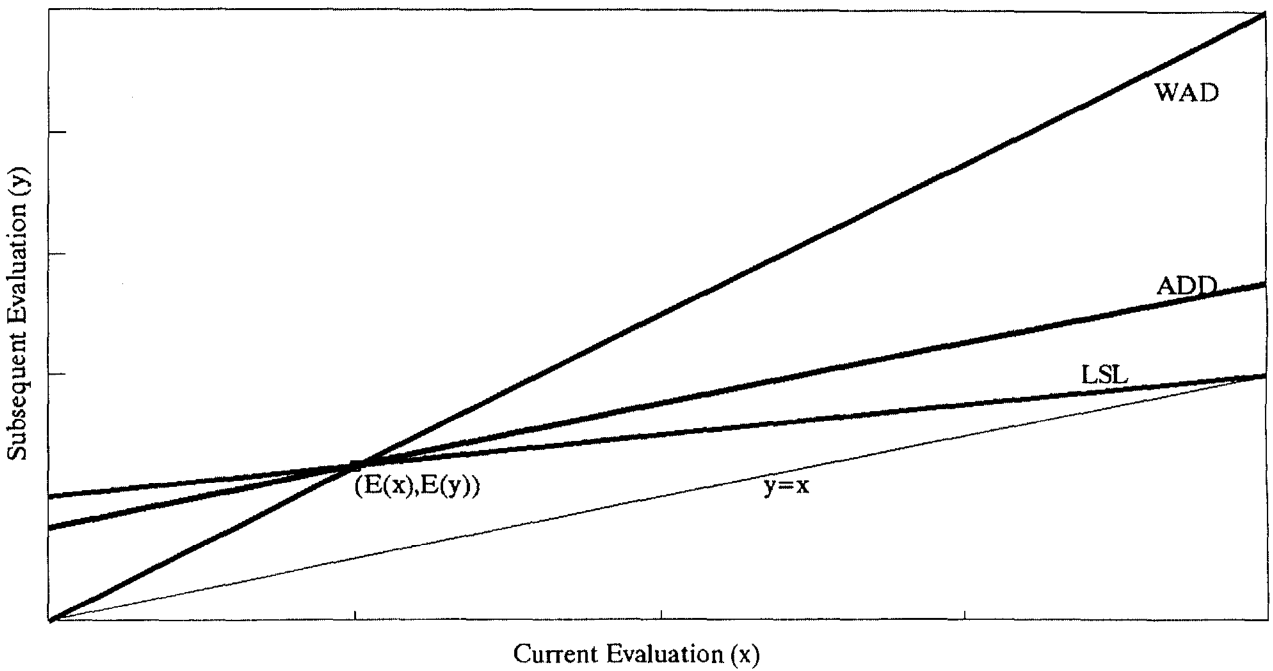


Figure 9B

Idealized Development Estimators With Trend

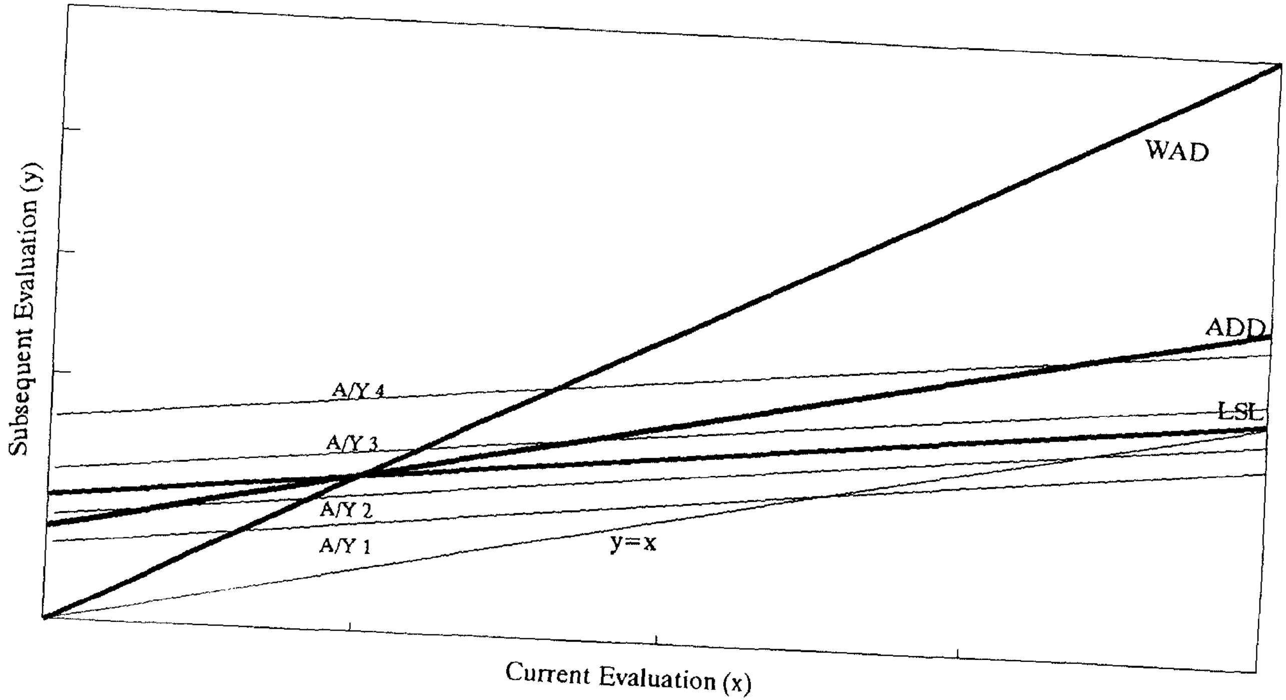


EXHIBIT I
Claim Counts Only

	A/Y	Average Bias	Std Dev Bias	Average %Bias	Std Dev %Bias	Age-Age Bias	Age-Age %Bias
Linear							
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		
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		
Nonlinear							
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 II
Random Severity, No Trend

	A/Y	Average Bias	Std Dev Bias	Average %Bias	Std Dev %Bias	Age-Age Bias	Age-Age %Bias
Linear							
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		
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		
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		
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		
Nonlinear							
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 III
Random Severity, 8% Trend

	A/Y	Average Bias	Std Dev Bias	Average %Bias	Std Dev %Bias	Age-Age Bias	Age-Age %Bias
Linear							
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		
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		
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		
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		
Nonlinear							
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 IV

Random Severity, 8% Trend, Estimates Based on On-Leveled (at 8%) Triangle

		A/Y	Average Bias	Std Dev Bias	Average %Bias	Std Dev %Bias	Age-Age Bias	Age-Age %Bias
Linear								
	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		
	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	(4,355)	0.061
		Total	(16,251)	833,130	(0.003)	0.277		
	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		
	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		
Nonlinear								
	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

Theorem 1: The $\mathbf{b}_{WAD}$ estimator satisfies Properties 4 and 5: the sum of the residuals is zero and the line through the origin with slope $\mathbf{b}_{WAD}$ passes through the sample average $(\bar{x}, \bar{y})$.

Proof:

$$\begin{aligned} \Sigma(y - b_{WAD}x) &= \Sigma y - \frac{\bar{y}}{\bar{x}}\Sigma x \\ &= n(\bar{y} - \frac{\bar{y}}{\bar{x}}\bar{x}) \\ &= 0 \quad . \end{aligned}$$

This proves Property 4. Next, $\bar{y} - \frac{\bar{y}}{\bar{x}}\bar{x} = 0$, so $\bar{y} = b_{WAD}\bar{x}$, demonstrating that the sample average is on the line through the origin with slope $\mathbf{b}_{WAD}$, Property 5.

Theorem 2: $\mathrm{Var}(\mathbf{b}_{WAD}) \geq \mathrm{Var}(\mathbf{b}_{LSM})$.

Proof: First, write $\mathbf{b}_{LSM} = \Sigma x_i y_i/\Sigma x^2 = \Sigma w_i y_i$ where $w_i = x_i/\Sigma x^2$. Recall that all expectations of y are conditional on x, including the variance, which means that expressions involving x, in particular w, may be manipulated as constants. Therefore,

$$\begin{aligned} \mathrm{Var}(b_{LSM}) &= \mathrm{Var}(\sum_i w_i y_i \mid (x_1, x_2, \dots, x_n)) \\ &= \sum_i w_i^2 \mathrm{Var}(y_i \mid x_i) \\ &= \sum_i w_i^2 \sigma^2 \\ &= \sigma^2 \sum_i \left(\frac{x_i}{\Sigma x^2}\right)^2 \\ &= \frac{\sigma^2}{\Sigma x^2} \quad . \end{aligned}$$

Next,

$$\begin{aligned}\mathrm{Var}\, b_{WAD} &= \mathrm{Var}\left(\frac{\Sigma y}{\Sigma x}\right)\\ &= \frac{1}{(\Sigma x)^2}\Sigma\,\mathrm{Var}\, y\\ &= \frac{n\sigma^2}{(\Sigma x)^2}\quad .\end{aligned}$$

To show that $\mathrm{Var}(b_{WAD}) \geq \mathrm{Var}(b_{LSD})$ we only have left to show that

$$\frac{1}{n}(\Sigma x)^2 \leq \Sigma x^2$$

or

$$\left(\frac{1}{n}\Sigma x\right)^2 \leq \frac{1}{n}\Sigma x^2 \ .$$

But the latter is just the Schwartz Inequality.[1] QED.

Theorem 3: Under Model III, the least squares estimator is b_{SAD}.

Proof: The transformed Model III

$$\frac{y}{x} = b + e$$

is of the form

$$u = bv + e$$

where the variable v is identically equal to unity. Thus, the transformed model satisfies all the assumptions of Model II. Accordingly, its least squares estimator is

$$\hat{b} = \frac{\Sigma vu}{\Sigma v^2} = \frac{\sum \frac{y}{x}}{n} = b_{SAD} \quad .$$

---

[1]See for example John F. Randolph, *Basic Real and Abstract Analysis*, p. 35.

Theorem 4: Under Model IV, the least squares estimator is $\mathbf{b}_{GAD}$.

Proof: The transformed Model IV

$$\ln(y) = \ln(\mathbf{b}) + \ln(x) + \ln(e)$$

or

$$\ln(y) - \ln(x) = \ln(\mathbf{b}) + \ln(e)$$

is of the form

$$u = b'v + ve'$$

where $b'=\ln(b)$, $v=1$, and $E(e')=0$. Thus, the transformed model satisfies the SAD assumptions. By Theorem 3

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

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

$$\hat{b} = \exp(\hat{b}') = \exp\left(\frac{1}{n}\sum \ln\frac{y}{x}\right) = \left(\exp\sum \ln\frac{y}{x}\right)^{\frac{1}{n}} = \sqrt[n]{\prod \frac{y}{x}} = b_{GAD} \quad .$$

Theorem 5: Under the assumptions of Model I, $y_{LSL} = \mathbf{a}_{LSL} + \mathbf{b}_{LSL}x$ is an unbiased estimator of y; i.e., $E(y_{LSL}) = E(y)$. Under the assumptions of Model II, $y_{LSM} = \mathbf{b}_{LSM}x$ is an unbiased estimator of y.

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

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

The proof for LSM is similar.

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

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

First,

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

Next, the "Multiplication Rule" of conditional density functions[2] states that

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

Therefore,

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

By the CLIA, the random variable $x_n|x_{n-1}$ is independent of x_0. Therefore $f(x_n|(x_{n-1},x_0))$ does not depend on x_0, so $f(x_n|(x_{n-1},x_0))=f(x_n|x_{n-1})$. 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

[2]See Robert V. Hogg and Allen T. Craig, *Introduction to Mathematical Statistics*, p. 64.

to justify that step. Then

$$
\begin{aligned}
E(x_n|x_0) &= \int_{x_n} x_n f(x_n|x_0)\,dx_n \\
&= \int_{x_n} x_n \left(\int_{x_{n-1}} f(x_n|(x_{n-1},x_0))\,f(x_{n-1}|x_0)\,dx_{n-1}\right)dx_n \\
&= \int_{x_{n-1}} \left(\int_{x_n} x_n f(x_n|(x_{n-1},x_0))\,dx_n\right) f(x_{n-1}|x_0)\,dx_{n-1} \qquad (1) \\
&= \int_{x_{n-1}} \left(\int_{x_n} x_n f(x_n|x_{n-1})\,dx_n\right) f(x_{n-1}|x_0)\,dx_{n-1} \\
&= \int_{x_{n-1}} (\mathrm{a_n}+\mathrm{b_n}x_{n-1})\,f(x_{n-1}|x_0)\,dx_{n-1} \\
&= \mathrm{a_n} + \mathrm{b_n}\int_{x_{n-1}} x_{n-1} f(x_{n-1}|x_0)\,dx_{n-1} \\
&= \mathrm{a_n} + \mathrm{b_n}\mathrm{E}(x_{n-1}|x_0) \quad .
\end{aligned}
$$

Proof 2: Recall the well-known identity $\mathrm{E(X)=E_Y[E(X|Y)]}$.[3] Consider the following variation reiterated in equation (1) above:

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

For LSL we have

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

Theorem 6: $\mathrm{E}(\beta_n|x_0) = \mathrm{E}(x_n|x_0)$.

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

[3] See for example I. B. Hossack, J. H. Pollard, and B. Zehnworth, *Introductory Statistics with Applications in General Insurance*, 1983, p. 63.

For n=1 the theorem is simply a restatement of Theorem 5.

Assume that $E(\beta_{n-1}|x_0) = E(x_{n-1}|x_0)$. We have that $\beta_n = \hat{a}_n + \hat{b}_n\beta_{n-1}$ where $\hat{a}_n$ and $\hat{b}_n$ are functions of the random variables $x_n|x_{n-1}$, and β_{n-1} is a function of the random variables $x_{n-1}|x_{n-2}, \ldots, x_1|x_0$ 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 β_{n-1}. Therefore,

$$
\begin{aligned}
E(\beta_n|x_0) &= E(\hat{a}_n|x_0) + E(\hat{b}_n|x_0)E(\beta_{n-1}|x_0) && \because \hat{b}_n \text{ and } \beta_{n-1} \text{ are independent} \\
&= E_{x_{n-1}|x_0}[E(\hat{a}_n|(x_{n-1},x_0))] + E_{x_{n-1}|x_0}[E(\hat{b}_n|(x_{n-1},x_0))]\,E(\beta_{n-1}|x_0) \\
&= E_{x_{n-1},x_0}[E(\hat{a}_n|x_{n-1})] + E_{x_{n-1}|x_0}[E(\hat{b}_n|x_{n-1})]\,E(\beta_{n-1}|x_0) \\
&= E_{x_{n-1}|x_0}[a_n] + E_{x_{n-1}|x_0}[b_n]\,E(\beta_{n-1}|x_0) \\
&= a_n + b_n E(\beta_{n-1}|x_0) \\
&= a_n + b_n E(x_{n-1}|x_0) && \text{by the induction hypothesis} \\
&= E(x_n|x_0) && \text{by Lemma 1.}
\end{aligned}
$$

Theorem 7:

For n=1:

	LSL	LSM
For n=1:	$\operatorname{Var}\beta_1 = \dfrac{\sigma_1^2}{I_1} + (x_0-\bar{x}_0)^2\operatorname{Var}\hat{b}_1$	$\operatorname{Var}\beta_1 = x_0^2\operatorname{Var}\hat{b}_1$

For n>1:

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

$$\mathrm{Var}\,\beta_n = \mu_{n-1}^2\,\mathrm{Var}\,\hat{b}_n + b_n^2\,\mathrm{Var}\,\beta_{n-1} + \mathrm{Var}\,\hat{b}_n\,\mathrm{Var}\,\beta_{n-1}$$

Proof: We will prove the LSM case first. We saw in Theorem 6 that $\hat{b}_n$ and β_{n-1} are independent random variables. The formula[4] for the variance of the product of two independent random variables x and y is

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

This proves the theorem for LSM because $\hat{b}_n$ is unbiased.

For LSL,

$$\mathrm{Var}\,\beta_n = \mathrm{Var}\,\hat{a}_n + 2\,\mathrm{Cov}(\hat{a}_n, \hat{b}_n\beta_{n-1}) + \mathrm{Var}(\hat{b}_n\beta_{n-1}).$$

It is well known[5] that the random variables $\bar{x}_n$ and $\hat{b}_n$ are uncorrelated when $\hat{b}_n$

[4]See Hogg and Craig, p. 178, problem 4.92.

[5] See R. Miller and D. Wichern, *Intermediate Business Statistics*, 1977, p.202, for example.

is determined by least squares; since all expectations are conditional, we have that

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

Next,

$$\begin{aligned}\operatorname{Cov}(\hat{a}_n, \hat{b}_n\beta_{n-1}) &= E\beta_{n-1}\operatorname{Cov}(\hat{a}_n, \hat{b}_n) \qquad \because \beta_{n-1} \text{ is independent of } \hat{a}_n \text{ and } \hat{b}_n\\ &= \mu_{n-1}\operatorname{Cov}(\hat{a}_n, \hat{b}_n)\end{aligned}$$

and

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

Putting these together with the formula for $\operatorname{Var}(\hat{b}_n\beta_{n-1})$ from the LSM derivation above we have

$$\begin{aligned}\operatorname{Var}\beta_n &= \frac{\sigma_n^2}{I_n} + \overline{x}_{n-1}^2\operatorname{Var}\hat{b}_n - 2\mu_{n-1}\overline{x}_{n-1}\operatorname{Var}\hat{b}_n + \beta_{n-1}^2\operatorname{Var}\hat{b}_n + \hat{b}_n^2\operatorname{Var}\beta_{n-1} + \operatorname{Var}\beta_{n-1}\\ &= \frac{\sigma_n^2}{I_n} + (\mu_{n-1} - \overline{x}_{n-1})^2\operatorname{Var}\hat{b}_n + b_n^2\operatorname{Var}\beta_{n-1} + \operatorname{Var}\hat{b}_n\operatorname{Var}\beta_{n-1} \quad .\end{aligned}$$

Theorem 8:

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

Proof:

$$\begin{aligned}
\mathrm{Var}(x_n|x_0) &= E_{x_{n-1}|x_0}[\mathrm{Var}(x_n|(x_{n-1},x_0))] + \mathrm{Var}_{x_{n-1}|x_0}[E(x_n|(x_{n-1},x_0))] \\
&= E_{x_{n-1}|x_0}[\mathrm{Var}(x_n|x_{n-1})] + \mathrm{Var}_{x_{n-1}|x_0}[E(x_n|x_{n-1})] && \text{by CLIA} \\
&= E_{x_{n-1}|x_0}(\sigma_n^2) + \mathrm{Var}_{x_{n-1}|x_0}(a_n + b_n x_{n-1}) && \text{under LSL} \\
&= \sigma_n^2 + b_n^2\mathrm{Var}(x_{n-1}|x_0) && \text{under LSL or LSM.}
\end{aligned}$$

Lemma 2: $E(S_n) = n\mathbf{a}_n + \mathbf{b}_n(E(S_{n-1}) + x_{n-1,n-1})$.

Proof:

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

Theorem 9: Let $XD_n=(x_{0,0},x_{1,1},...,x_{n-1,n-1})$ denote the current diagonal of the triangle for the n youngest accident years. Then

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

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

$$\begin{aligned} E(\hat{M}_1|XD_1) &= E(\hat{\beta}_{0,1}|x_{0,0})) \\ &= E(x_{0,1}|x_{0,0}) && \text{by Theorem 7} \\ &= E(S_1) \end{aligned}$$

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 $x_{i,n}|x_{i,n-1}$ 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. By the CLIA $\hat{a}_n$ and $\hat{b}_n$ are independent of $\hat{M}_{n-1}$. Therefore

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

Theorem 10: PARAMETER RISK

LSL

For n=1:

$$\mathrm{Var}\,\hat{M}_1 = \frac{\sigma_1^2}{I_1} + (x_{0,0}-\bar{x}_0)^2\,\mathrm{Var}\,\hat{b}_1$$

For n>1:

$$\mathrm{Var}\,\hat{M}_n =$$

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

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

LSM

For n=1:

$$\mathrm{Var}\,\hat{M}_1 = x_{0,0}^2\,\mathrm{Var}\,\hat{b}_1$$

For 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 LSM 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 LSL,

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

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

In the proof of Theorem 7 we saw that (equation (2))

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

and that (equation (3))

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

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,

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

therefore

$$\begin{aligned}
\operatorname{Var}\hat{M}_n &= n^2\left(\frac{\sigma_n^2}{I_n} + \bar{x}_{n-1}^2 \operatorname{Var}\hat{b}_n\right) - 2nE(\hat{M}_{n-1} + x_{n-1,n-1})\,\bar{x}_{n-1} \operatorname{Var}\hat{b}_n \\
&\quad + (M_{n-1} + x_{n-1,n-1})^2 \operatorname{Var}\hat{b}_n + b_n^2 \operatorname{Var}\hat{M}_{n-1} + \operatorname{Var}\hat{b}_n \operatorname{Var}\hat{M}_{n-1} \\
&= n^2 \frac{\sigma_n^2}{I_n} + (M_{n-1} + x_{n-1,n-1} - n\bar{x}_{n-1})^2 \operatorname{Var}\hat{b}_n \\
&\quad + b_n^2 \operatorname{Var}\hat{M}_{n-1} + \operatorname{Var}\hat{b}_n \operatorname{Var}\hat{M}_{n-1} \quad .
\end{aligned}$$

Theorem 12: $\mathrm{Var}(S_n) = n\sigma_n^2 + b_n^2\mathrm{Var}(S_{n-1})$.

Proof:

$$\mathrm{Var}(S_n) = E_{X_{n-1}|XD_n}[\mathrm{Var}(S_n | (X_{n-1}, XD_n))] + \mathrm{Var}_{X_{n-1}|XD_n}[E(S_n | (X_{n-1}, XD_n))]$$

$$= E_{X_{n-1}|XD_n}[\mathrm{Var}(\sum_{i=0}^{n-1} x_{i,n} | (x_{i,i}, x_{i,n-1}))] + \mathrm{Var}_{X_{n-1}|XD_n}(a_n + b_n(S_{n-1} + x_{n-1,n-1}))$$

$$= E_{X_{n-1}|XD_n}[\mathrm{Var}(\sum_{i=0}^{n-1} x_{i,n} | x_{i,n-1})] + b_n^2\,\mathrm{Var}_{X_{n-1}|XD_{n-1}}(S_{n-1} + x_{n-1,n-1})$$

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

Theorem 13: Under the transformed GAD model

$$x_n' = b_n' + x_{n-1}' + e_n'$$

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

$$\hat{\mu}'_U = x_0' + \sum_{j=1}^{U} \hat{b}_j' \quad .$$

is

$$(C + \sum_{j=1}^{U} \frac{1}{I_j}) s'^2$$

where s'^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 (refer to Figure 4)

$$
\begin{pmatrix} x'_{N,1}-x'_{N,0} \\ x_{N-1,1}-x'_{N-1,0} \\ \vdots \\ x'_{1,1}-x'_{1,0} \\ x'_{N,2}-x'_{N,1} \\ \vdots \\ x'_{2,2}-x'_{2,1} \\ \vdots \\ x'_{N,N-1}-x'_{N,N-2} \\ x'_{N-1,N-1}-x'_{N-1,N-2} \\ x'_{N,N}-x'_{N,N-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} \times \begin{pmatrix} b'_1 \\ b'_2 \\ \vdots \\ b'_{N-1} \\ b'_N \end{pmatrix} + \begin{pmatrix} e'_1 \\ e'_1 \\ \vdots \\ e'_1 \\ e'_2 \\ \vdots \\ e'_2 \\ \vdots \\ e'_{N-1} \\ e'_{N-1} \\ e'_n \end{pmatrix},
$$

or, in more concise format, $Y=X\beta + E$. It is well known that the least squares estimator of β is $\hat{\beta}=(X'X)^{-1}X'Y$ and that the variance-covariance matrix of this estimator is $(X'X)^{-1}\sigma'^2$. In this case, it is clear by inspection that $X'X$ 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, $\operatorname{Var}\hat{b}'_j = \frac{\sigma'^2}{I_j}$ and $\operatorname{Cov}(\hat{b}'_i,\hat{b}'_j)=0$ for $i \neq j$. Therefore, the Parameter Risk $\operatorname{Var}(C\sum_{j=1}^{U}\hat{b}'_j)$ is exactly equal to $\sigma'^2\sum_{j=1}^{U}\frac{1}{I_j}$. The Process Risk is equal to

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

These variances are estimated by substituting the estimate s'^2 for σ'^2.