

*A Comparative Study of the Performance of  
Loss Reserving  
Methods Through Simulation*  
by Prakash Narayan, ACAS, and  
Thomas V. Warthen, III, FCAS

# A COMPARATIVE STUDY OF THE PERFORMANCE OF LOSS RESERVING METHODS THROUGH SIMULATION

## ABSTRACT

Actuaries are often asked to provide a range or confidence level for the loss reserve along with a point estimate. Traditional methods of loss reserving do not provide an estimate of the variance of the estimated reserve and actuaries use various ad hoc methods to derive a range for the indicated reserve. Regression models for loss reserving are getting increasing attention in actuarial research as they provide an estimate of the variance of the loss reserve, along with a point estimate. However, these methods are rarely used in practice, both because of their complexity and the lack of their historical use. In this paper we use a Monte Carlo simulation method to compare loss reserve estimation methods, including traditional methods, and regression based methods of loss reserving. Our approach is similar to that of Stanard [5], where he compares several traditional actuarial methods using simulation techniques. However, we use different methods for simulating loss triangles, and compare the estimated reserve based on several characteristics.

# A COMPARATIVE STUDY OF THE PERFORMANCE OF LOSS RESERVING METHODS THROUGH SIMULATION

## I. Introduction

Loss reserving, or projecting losses to their ultimate value, is an important actuarial function. The loss development factor (LDF) method attempts to estimate the pattern with which losses for a given cohort of claims changes over time. This method produces a point estimate of the required reserve, and is probably the most commonly used actuarial technique for projecting losses to their ultimate value. However, actuaries are often asked to provide a range or the variability associated with the point estimate of the loss reserve.

In the last decade, a variety of methods have been suggested by actuaries for this purpose. Thomas Mack [3] developed a methodology to estimate the variability of the estimated loss reserves when the LDF method is used. However, his method may not be appropriate in many situations as the selection of the development factors is often judgmental. Randall Holmberg [2] has also presented a model by which actuaries can estimate the variability of their loss reserve estimates. Regression modeling of the loss triangle, which can provide both a point estimate and the variability associated with the point estimate, is getting increasing attention from actuaries. Regression methods provide an estimator of the variance more directly. These methods, however, are rarely used by actuaries both because of their complexity and the reluctance to use a methodology that is not widely used. It is desirable to thoroughly test a new methodology before it can be accepted as an appropriate technique and used in practice. Comparisons of forecasting methods based on historical data are not generally considered an objective method for testing forecasting methods. Such studies are likely to be biased by the preference of the investigator.

Alternatively, statistical simulation is a well-accepted technique for comparing various methods of estimation when the properties of the estimators cannot be studied analytically. Stanard [5] used this technique to compare various traditional methods of loss reserving. We shall apply the same technique to compare the traditional methods with the regression method of loss reserving. Our study is more comprehensive in the sense that it uses a variety of methods to simulate the loss triangles.

We have selected the LDF method as one of the methods to compare because it is the most commonly used traditional actuarial method. We have included the Buhlmann complementary loss ratio method (which Stanard refers to as the additive model), because this method was the best of the tested methods per the Stanard [5] study. We have compared these loss reserve estimation methods and regression methods. The various loss reserve estimation regression models considered in this study differ in the number of the parameters used in modeling the loss triangle.

Our approach is to simulate random loss triangles by a variety of methods, and estimate

the corresponding loss reserves using the traditional, Buhlmann complementary loss ratio model, and log-regression models. We assume that the ultimate losses, hence the reserves, are known with certainty. We compute the deviations of the estimated reserves from the actual reserves derived by various methods. We expect this deviation to be small for a good reserving method. We use several criteria to compare the estimated deviations of actual versus estimated reserves under the various reserving methods.

In Section II, the particular methods of simulating random loss triangles are described. We do not claim that these methods capture all the intricacies of the claims process. However, we do believe that they generate loss data triangles that are stochastic and do not provide an apparent advantage to any particular method of loss reserve estimation. A particular method of reserve estimation may incorporate some underlying assumptions about the claims process and will obviously provide a better estimate of the loss reserve if those assumptions are satisfied. In practice it may not be possible to test the assumptions underlying a particular loss reserve estimation method. If a statistical test is applied, it can only detect a gross violation of the assumptions and cannot confirm that those assumptions are in fact true.

Loss development factor methods have an extensive history of use in actuarial practice, which preceded the investigation and documentation of the assumptions underlying these methods. Given the current and historical familiarity with loss development factor methods, the assumptions underlying these methods are in some sense secondary to the methods themselves. Or, in other words, given their widespread historical use and technical adequacy as loss development estimation methods, actuaries would presently be using loss development factor methods even if no studies about the underlying assumptions were ever published. This is a major consideration which led us to use a variety of methods to simulate the random loss triangles.

We are comparing a traditional loss development factor loss reserve estimation method, the Buhlmann complementary loss ratio method, and three fixed regression loss estimation models to estimate the loss reserves. These methods and models are briefly described in Section III. Section IV discusses the criteria used to compare the results of the simulations. One can definitely define comparison criteria other than those used here. However, the criteria used are comprehensive, and an estimator performing better in the criteria considered will likely be a good estimator with respect to other reasonable criteria. We have also provided a brief summary of the results of the simulations for the aggregate loss reserves in this section. Appendix B provides the individual accident year results of our computations. Section V makes several observations on the results and draws some conclusions based on this simulation study.

## II. Simulating Random Loss Triangles:

Modeling a claims process to generate the random elements of a loss triangle is quite complicated. There does not appear to be any study that derives a severity distribution for losses where the individual loss amount may change over time. Stanard [5] and Pentikainen and Rantala [4] describe methods of simulating random loss triangles. Their methods are fundamentally different. The Stanard method is based on a loss severity distribution of individual claim amounts whereas Pentikainen and Rantala [4] use an aggregate stochastic claim process.

The various methods of loss triangle simulation used here do not satisfy the assumptions underlying the various methods of loss reserving compared. For example, for log-regression modeling, it is assumed that the incremental losses are independent. However, this assumption is violated by all the methods used for simulating the random loss triangles. Similarly the random loss triangle simulation methods do not satisfy the basic requirement of the LDF method that the future development is completely determined by the latest available data. One can then infer that our study tests the robustness of the various methods of loss reserving.

While we have defined and used particular methods of random loss triangle simulation in this analysis, similar methods may have been used in earlier actuarial studies. We have not performed an exhaustive literature search, and therefore make no specific attribution, believing that our application of these methods of generating random loss triangles is original.

We have used four different techniques for simulating the loss triangles. The Pentikainen and Rantala [4] method is one of them. As we shall see later, the log-regression method of loss reserving requires that the incremental losses be positive. If this is not the case, some subjective judgments need to be made. One way to treat such incidences is to delete such observations from the data set. To be uniform and consistent, we have selected loss triangle simulation methods that will generate positive incremental losses. Stanard's method does not satisfy this requirement and is not used.

For all the methods utilized in this study, eleven accident years are considered. It is further assumed that the losses completely mature at the 11th year of development, i.e., the first accident year is at the ultimate loss level and no further development is expected. Since we require complete knowledge of the ultimate losses for a proper comparison of the reserve estimation results of the different estimation methods, we generate a complete history for each accident year. However, in estimating the reserves, only the top half of the loss triangle is available to the actuary as data, and is used to estimate the lower half of the triangle, particularly the last (rightmost) column, which represents the projection of ultimate losses. We shall denote by  $L_{i,j}$  the losses for the accident year  $i$  at the end of the period  $i + j - 1$  and by  $S_{i,j}$  the corresponding incremental losses. For simplicity,  $i$  and  $j$  are assumed to vary from 1 to 11.

## 1. Random Reporting Factor:

The steps of this random loss triangle generation method are:

- (i) Generate  $N$ , the frequency of losses, as Poisson with mean 100.
- (ii) Generate  $N$  log-normally distributed random variables with parameters  $\mu$  (7.3659) and  $\sigma$  (1.517427) and add them together. This is our ultimate loss for accident year 1 and let us denote it by  $S_1$ . These parameters correspond to a severity mean of 5000 and a coefficient of variation of 3.
- (iii) Generate 10 uniform random numbers  $X_j$  and compute
$$T_j = 0.1 + 0.5 * X_j + 0.5 * \ln(j), \quad j = 1, 2, \dots, 10.$$
$$U_j = T_1 + T_2 + \dots + T_j.$$
- (iv) The loss for accident year 1 at delay  $j$ ,  $L_{1,j}$  is given by
$$L_{1,j} = S_1 * (1 - \text{EXP}(-U_j)) \text{ and}$$
$$L_{1,11} = S_1.$$
- (v) Repeat steps (i) through (iv) for the accident years 2 through 11 except that the ultimate losses for the accident year  $i$  are multiplied by  $1.06^{(i-1)}$  or accident year losses are inflated 6% a year.

This method may look like a development factor model. However, this method does not strictly satisfy the assumptions of the loss development factor model. The ratio of the expected losses  $EL_{i,j+1} / EL_{i,j}$  is a constant, not the conditional expectation. It also does not satisfy the assumption of independence of incremental losses.

## 2. Random Backward Development Factor:

This method is similar to method 1 except that the factors are computed backward. The steps of the method are:

- (i) Generate  $N$ , the frequency of losses as Poisson with mean 100.
- (ii) Generate  $N$  log-normally distributed random variables with parameters  $\mu$  (7.3659) and  $\sigma$  (1.517427) and add them together. This is our ultimate loss for accident year 1 and let us denote it by  $S_1$ .
- (iii) Generate lognormal variates  $Y_{11..j}$  for  $j = 1, 2 \dots 10$  with parameters  $a_j$  and  $b_j$  where
$$a_j = (j + (j - 1)^2) / 100 \text{ and}$$
$$b_j = (j + (j - 1)^2) / 500.$$

$Y_j$  is actually a randomly generated development factor for the period  $j$  to  $j+1$ .

- (iv) Losses reported at the end of year  $j$  for the accident year 1,  $L_{1, 10}$ , are  $S_1/ Y_{10}$ . The reported losses at earlier valuation dates are computed by dividing by  $Y_j$  successively.
- (v) Repeat the steps (i) through (iv) for the accident years 2 through 11 except that ultimate losses for the accident year  $j$  are multiplied by the  $1.06^{(i-1)}$  or accident year losses are inflated 6% a year.

The  $a_j$  and  $b_j$  parameter are selected so that  $Y_j$  is greater than one with very high probability.

### 3. Individual Losses With Changing Severity:

This method is based on the ideas of Stanard [5] and Buhlmann, Schnieper and Straub [1]. As in Stanard, we assume an exponential delay in reporting and settlement with the added assumption that the severity distribution varies with delay. The actual method is described below:

- (i) Generate  $N$ , the frequency of losses as a Poisson variate with mean 100.
- (ii) For each  $N$ , generate a uniform variate  $X_1$  for the occurrence date, and exponential variates  $X_2$  and  $X_3$ , mean 2 and 5 respectively, for the reporting delay and the settlement delay. Based on these values, individual claims amounts are determined as follows.
- (iii) If  $X_1 + X_2$  is greater than  $j$ , the loss amount for this claim by delay  $j$  is zero.
- (iv) The claim amounts for a claim follow a Pareto distribution. The percentile level of an individual loss remains constant over time but the shape and scale parameters change until the claim is settled. The Pareto distribution function and the parameter values by delay are given in Appendix A. After settlement, the claim value or the loss amount does not change. As we require that the ultimate values of the claim be known, we truncate  $X_1 + X_2 + X_3$  at 11. This provides loss amounts for each claim for delays 1 to 11. These are added together appropriately to get the aggregate losses at each delay period.
- (v) Repeat the steps (i) through (iv) for the accident years 2 through 11 except that ultimate losses for the accident year  $i$  are multiplied by the  $1.06^{(i-1)}$  or accident year losses are inflated 6% a year.

#### 4. Pentikainen and Rantala Method:

This method is based on the procedure described by Pentikainen and Rantala [4]. Our implementation may differ slightly from theirs. We shall describe the computational steps of this method very briefly and the reader is encouraged to review the original Pentikainen/Rantala paper for a complete understanding of the method. The computational steps of this method are:

(i) We assume a reporting pattern for a cohort of aggregate losses. This pattern is constant over time and includes pure IBNR. The pattern used is 0.220, 0.180, 0.150, 0.120, 0.100, 0.080, 0.060, 0.040, 0.027, 0.016, 0.007 and will be denoted by  $X(j)$ ,  $j = 1, 2 \dots 11$ .

(ii) Claims for the accident year  $i$  reported at delay  $j$  are given by

$$S_{i,j} = K * X(j) * XP(i) * q(i, j) * \text{infl}(t)$$

where

$K$  is a parameter and related to the overall loss level for accident year  $i$ ,  
 $X(j)$  is  $j$  th element of the reporting pattern of aggregate losses discussed earlier,  
 $XP(i) = (1.01^{i-1}) * (1.06^{i-1})$  are exposure and inflation growth,  
 $q(i, j) = 0.4 + 0.6 * q(i, j-1) + \epsilon$ ,  $q(i, 0) = 1$  and  $\epsilon$  is distributed  $N(0, 0.05)$ ,  
 $t = i + j - 1$ ,

$$\text{infl}(t) = \prod_{k=1}^t (1 + \text{infl}(k)) \text{ and}$$

$$\text{infl}(k+1) = \max(0.06 + 0.7 * (\text{infl}(k) - 0.06) + \epsilon, 0.03), \text{ infl}(1) = 0.06, \text{ and } \epsilon \sim N(0, 0.015).$$

This method is based on randomizing the aggregate losses of all the claims for an accident year. Claims reporting and inflation are modeled by auto-regressive processes. We further restrict the inflation rate to a minimum of 3%. This method also has an exposure growth of 1 percent.

One can see that each of the four methods of generating loss triangles has several parameters. As these parameters are changed, the simulated triangles may exhibit significantly different development patterns. A particular method of loss reserving, considered best with a selected set of loss triangle generation parameters, need not be better for any other set of loss triangle generation parameters. The simulation conducted here emphasizes more on a variety of methods of loss triangle generation than on the sensitivity of the loss triangle generation methods over a range of possible parameters.

We note here, that in the simulation of random loss triangles by methods 1, 2, and 3, as described above, individual claim severity is unlimited. In actual practice individual losses will have an upper limit in most cases. Occurrence of an individual very large loss in the

simulation process may cause an individual accident year loss to be out of line with other accident year losses in an individual loss triangle.

It is worth stating that the computations for the simulations were performed in EXCEL. We have, however, implemented our own module to generate the uniform random variate.

### III. Methods of Loss Reserving:

We assume that there is no further development beyond year  $n$  or, equivalently,  $L_{i,n}$  is the ultimate loss value for the accident year  $i$ . We further assume that all  $S_{i,j}$  are positive and let  $Z_{i,j}$  denote the  $\ln(S_{i,j})$ . To simplify the later exposition of our estimation process, we further assume that the accident year loss inflation rate is 6% and there is no exposure growth except for method 4 in which constant exposure growth of 1 percent is assumed. We know  $L_{i,j}$ ,  $i = 1, \dots, n$ , and  $j = 1, \dots, n + 1 - i$  and are interested in estimating  $L_{i,j}$  for  $i = 2, \dots, n$  and  $j = n + 2 - i, \dots, n$ .

We shall use two traditional methods of loss reserving and three regression models. The two traditional methods are the loss development factor method and the Buhlmann complementary loss ratio method. The loss development factor method is the most commonly used actuarial technique and the Buhlmann loss ratio method was chosen for this analysis because this method outperformed other actuarial methods in the simulation study by Stanard [5].

The three regression models we have selected for comparison are similar and the differences among them lie in the number of parameters fitted. These methods are described next.

#### 1. Loss Development:

We compute:

$$f_{i,j} = L_{i,j+1} / L_{i,j}$$

$$f_j = \sum_{i=1}^{n-j} f_{i,j} / (n - j) \quad (1)$$

$$u_k = \prod_{j=k}^{n-1} f_j$$

and an estimated  $\bar{L}_{i,n}$  is given by

$$\bar{L}_{i,n} = L_{i, n-i+1} * u_{n-i+1}$$

#### 2. Buhlmann Complementary Loss Ratio Method:

This method of loss reserving has not been commonly applied in North America and is suitable for application to paid loss data. It is based on the presumption that the

proportion of losses paid at a particular delay remains constant over time. This proportion is estimated from the historical loss experience and is used to forecast the future. We compute:

$$M_j = \sum_{i=1}^{n-j+1} (S_{i,j} * (1 + \text{infl})^{(n-i)}) / (n-j+1) \text{ for } j = 2, 3, \dots, n \text{ and}$$

$$S_{i,j} = \bar{M}_j * (1 + \text{infl})^{(i-n)} \text{ for } j = n+2-i, \dots, n, \text{ and } i = 2, 3, \dots, n,$$

where infl is the inflation for losses and assumed to be 6%.

### 3. Regression Models

Our discussion of the regression models considered in our analysis is very brief. These models are discussed in greater detail by Zehnwirth [7] and Verral [6] among others. We have used an unbiased estimator for the loss reserves as recommended by Verral [6] rather than Bayes or maximum likelihood estimates (MLE). In these models the incremental losses are assumed to follow some stochastic distribution. Usually some transformation is applied to the incremental losses before the model parameters are estimated. Although various transformations have been investigated, the logarithmic transformation is most commonly used. Let us describe the methodology briefly with the log transformation for completeness. Readers not familiar with the methodology are encouraged to review the papers by Verrall [6] and Zehnwirth[7]. Let  $Z_{i,j}$  be the logarithm of the incremental loss for accident year  $i$  and delay  $j$ . We assume that

$$Z_{i,j} = \mu + \alpha_i + \beta_j + \varepsilon_{i,j}$$

where  $\mu$ ,  $\alpha_i$ ,  $\beta_j$  are the parameters of the model and  $\varepsilon_{i,j}$  are assumed independent identically distributed normal variates with mean 0 and variance  $\sigma^2$ , and are the error terms or the random noise. We make the usual assumption that  $\alpha_i$ ,  $\beta_j$  are zero to make the model of full rank. The parameters of the model are estimated by the least squares method. Under the assumption of the normality of the error terms, the estimates are also MLEs. We use the unbiased estimate for the forecasting and do require that the errors are independent and normally distributed.

The three regression models investigated in this paper are:

Model 1.  $\alpha_i$  and  $\beta_j$  are all different for  $i, j = 2, 3, \dots, n$ .

Model 2.  $\alpha_i = (i-1) * \alpha$  for  $i = 2, \dots, 11$  and  $\beta_j$  are all different for  $j = 2, 3, \dots, n$ .

Model 3.  $\alpha_i = (i-1) * \alpha$  and  $\beta_j = (j-1) * \beta + \gamma * \ln(j)$  for  $i, j = 2, 3, \dots, n$ .

The main difference among the three models lies in the number of parameters used to fit the data. In the actual application of regression models, one will select the model which provides the best fit to the data based on the evaluation of the residuals and other statistics of the fitted models. However, such an approach is not feasible in simulation. The number of parameters used in the three regression models are 21, 12 and 4 respectively. Zehnwirth [7] emphasizes parsimony when applying the regression models for forecasting. Obviously regression model 1 has too many parameters and model 3 too few to capture the essence of a random loss triangle. In selecting these regression models, our purpose is not to compare these models with each other but to see the effect of using lesser numbers of parameters in regression modeling.

The parameters are estimated by the method of least squares and used to forecast ultimate losses. We refer the reader to Verrall [6], who has provided a complete description of the estimation method and an unbiased estimator of the lower triangle for model 1. The other regression models require revisions to the design matrix, and modification of the appropriate equations, from those described in Verrall [6].

#### **IV. Comparison of Procedures**

We have generated 5000 realizations of hypothetical loss triangles for each of the simulation methods described earlier for generating such loss triangles. For each of the 5000 sets of hypothetical data, the reserves were estimated by the loss development method, Buhlmann complementary loss ratio method, and the regression loss reserve estimation methods as previously outlined. The deviations between the loss reserve estimates and the actual reserves were computed.

An important property of a good estimator is that it is unbiased. Stanard [5] in his study used this criterion for comparing various loss reserve estimators. If an estimator was unbiased, the average deviation of estimated versus actual reserves, over many simulations, will be negligible.

Between two unbiased estimators, statisticians prefer the estimator with the smaller variance. Between biased estimators, the estimator with the minimum mean square error is preferred. In our context, this means that the average squared deviations between the estimated and actual reserves should be small. This is an important criterion for a reserve estimation method in the insurance context. The reserves are an important component of the insurer's financial reporting. A reserving method which provides estimates with small biases, but for which the individual simulation (data set) estimates vary a lot from the actual reserves, may not be an appropriate reserve estimation method. One will prefer the reserve estimates to be closer to the true value. We have used root mean square error and the average absolute deviation of the estimated versus the actual reserve, to test the closeness of the reserve estimators to the actual reserve values. We have also computed the average percentage error. A reserve estimation method which generates a smaller percentage error in the estimate is better. Another criterion used to compare the various

loss reserving methods was to compute the correlation between the actual reserves and the estimated reserves. One would expect a high correlation for a good reserving method.

Summarizing briefly our observations based upon our analysis (refer to Table 1, below), we compare the reserve estimates for each of the loss reserve estimation methods, for each of the random loss triangle simulation methods:

#### 1. Random Reporting Factor

It is easy to see that the Buhlmann complementary loss ratio method is clearly the best loss reserve estimation method for the random reporting factor method of random loss triangle simulation. The regression models of loss reserve estimation have performed better than the loss development factor method. The correlation for all the reserving methods is low and surprisingly is smallest for the Buhlmann complimentary loss ratio method which is probably the best of all the methods.

It is interesting to note that regression model 2 has performed slightly better than regression model 1. The main difference between these models is that regression model 2 estimates accident year inflation, and allows one parameter for that model component, whereas regression model 1 allows an inflation parameter for each accident year. Our results indicate that parsimony in the regression model is very important, and that over parametrization may provide inferior results.

#### 2. Random Backward Development Factor

Regression model 3 appears to be the best method for this loss simulation method based on aggregate combined accident years' forecast. However, examining the individual accident year forecasts, one will identify the Buhlmann complementary loss ratio method being superior. Regression model 3 does not capture the payout pattern correctly. The other regression models have performed better than the loss development factor method.

The Buhlmann method again shows poor correlation with the actual reserves while the other methods show a reasonable correlation level.

We conclude that the Buhlmann method, and regression model 2, perform better for this method of random loss triangle simulation than the other tested methods.

#### 3. Individual Losses With Changing Severity

Loss development method has performed very well for this loss simulation method. Regression model 2 appears to be better overall. Regression model 3 performs poorly, perhaps because of an insufficient number of model parameters

#### 4. Pentikainen and Rantala Method

Regression models 1 and 2 outperformed the other methods. The loss development factor method performed better than the Buhlmann complementary loss ratio method and regression model 3.

Tables 1 summarizes our results for each of the four methods of random simulation of the hypothetical loss triangles. Appendix B provides similar statistics for individual accident years. As can be seen from these tables, one or the other of the three regression models considered in this analysis is generally better than the LDF method. The Buhlmann method is slightly better in some cases. However, we had assumed that the inflation rate is known for the Buhlmann method. We are therefore utilizing additional information for this method, and obtaining slightly better answers. Such information will ordinarily not be available in actual practice. Actual loss data will be tainted by both exposure changes and the inflationary loss cost changes which will vary over time. One should note that for most of the methods of random loss generation the effect of inflation has a minimal impact on the ultimate answer derived by traditional methods. Inflation affects the weighting given to individual accident years in the total reserve. For regression models, inflation will affect the forecast in a much more complicated fashion.

Although no particular method can be identified as superior to other methods in every situation, the performance of regression models is in general quite good. It is worth noting that we have not performed a sensitivity analysis of the individual methods of simulating the loss triangles. By changing the inflation rate or the reporting pattern, for example, one may find that the performance of the individual methods of loss reserving will be different. However, we suspect that the overall performance will be similar.

**TABLE 1**

Method of Forecast:	Dev Factor	Buhlmann Loss Ratio	Model 1	Regression Model 2	Model 3
<b>Five Thousand Iterations Under Method 1</b>					
Actual Total Reserve:	Average =	1,108,298	SD=	244,287	
Bias	151,681	5,222	36,486	31,240	51,387
RMSE	466,055	266,874	395,819	328,870	341,537
Avg Abs Dev	364,628	204,674	314,829	254,069	263,444
Avg %	16.84%	4.84%	6.22%	6.75%	8.69%
Corr Actual vs Est	0.25	0.09	0.25	0.15	0.14
<b>Five Thousand Iterations by Method 2</b>					
Actual Total Reserve:	Average =	3,665,734	SD=	485,206	
Bias	157,684	(8,088)	55,356	15,393	3,125
RMSE	512,092	638,187	481,727	542,257	519,705
Avg Abs Dev	391,022	485,769	373,282	420,438	403,056
Avg %	4.38%	1.23%	1.58%	0.79%	0.47%
Corr Actual vs Est	0.70	0.11	0.70	0.57	0.58
<b>Five Thousand Iterations by Method 3</b>					
Actual Total Reserve:	Average =	1,634,559	SD=	252,631	
Bias	30,566	(83,039)	(144,192)	(52,327)	(176,089)
RMSE	413,137	441,109	375,367	299,099	340,506
Avg Abs Dev	356,932	347,340	314,629	259,057	280,243
Avg %	1.39%	-4.36%	-9.49%	-3.31%	-9.52%
Corr Actual vs Est	0.62	0.39	0.88	0.66	0.32
<b>Five Thousand Iterations by Method 4</b>					
Actual Total Reserve:	Average =	3,183,654	SD=	330,776	
Bias	10,106	(21,441)	5,326	4,789	34,136
RMSE	186,688	186,916	183,351	195,148	201,012
Avg Abs Dev	147,536	147,830	145,029	153,675	157,283
Avg %	0.23%	-0.24%	0.07%	0.06%	0.98%
Corr Actual vs Est	0.89	0.84	0.89	0.88	0.88

## V. Conclusion

It is interesting to note that regression modeling provides an appropriate tool for estimating loss reserves. Regression methods do not provide the best answers in all situations, but are stable and have the added advantage of providing directly the variance or the confidence interval for the reserve estimate. One should also note that the regression models studied are a priori fixed. In actual practice, the structure of the models will be determined from a much wider set of possible models based on an analysis of the data under review. It is expected that the testing and selection of an appropriate loss reserving regression model should improve the ultimate loss reserve forecast in actual application.

It is true that actuaries do not apply the more traditional LDF method blindly. The array of development factors is typically examined carefully before a selection of particular factors entering the reserve estimation is made. The appropriateness of the LDF method is determined for the given data set before the results of any such analysis are accepted. However, professional judgment, and the selection of an appropriate model is much more important when regression loss reserve estimation methods are used. Therefore, an important step is missing for the regression methods as applied in this study. For the Buhlmann method, we have assumed knowledge of the inflation rate in addition to what is assumed known for other methods. In practice, inflation will not be known precisely, and the loss triangle will be distorted by exposure changes and inflation. This method may therefore not be as well behaved in practice, as in the simulation studies presented here.

The point estimation of the loss reserve has been the primary focus of this study and we have not considered the variability of loss reserves around the point estimate. Verral [6] has outlined the procedures for computing the variance of the forecast including both the forecasting error and the parameter uncertainty.

One should note that the overall performance of the LDF method is quite satisfactory. The closeness of the answers of the various methods gives a comfort that the actuarial methods of loss reserve estimation are generally well behaved. These results also tell us that regression modeling provides estimates similar to traditional actuarial methods and one should not hesitate to use them. Given the advantage that regression methods also estimate the variability of the estimated reserve, it is expected that their use in the actuarial field will increase.

## References

1. Buhlmann, H., Straub, E., Schnieper, R. "Claims Reserves in Casualty Insurance based on a probabilistic model", *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* 80,1, pp 21-46.

2. Holmberg, R. D. "Correlation and the Measurement of Loss Reserve Variability", Casualty Actuarial Society Forum, Spring 1994, page 247.
3. Mack, Thomas. "Measuring the Variability of Chain Ladder Reserve Estimates" Casualty Actuarial Society Forum, Spring 1994, page 229.
4. Pentikainen, T. and Rantala, J. "A Simulation Procedure for Comparing Different Claims Reserving Methods" Casualty Actuarial Society Forum, Fall 1995, page 128.
5. Stanard, J. N. "A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques" Proceeding of the Casualty Actuarial Society LXXII, page 124.
6. Verrall, R. J. "Statistical Methods for Chain Ladder Techniques" Casualty Actuarial Society Forum, Spring 1994, page 393.
7. Zehnwirth, Ben "Probabilistic Loss Development Factor Models with Applications to Loss Reserve Variability, Prediction Intervals and Risk Based Capital" Casualty Actuarial Society Forum, Spring 1994, page 447.

## Appendix A

In this appendix a brief description of the method to generate a random loss amount at delay  $k$  for method 3 is provided. Let us define two vectors  $\lambda(i)$  and  $\theta(i)$ ,  $i = 1, 2, \dots, n$ . We generate a uniform random number  $X$  for each claim as soon as the claim is reported. If the claim is settled at delay  $k$  i.e. in between delay  $k - 1$  and  $k$  then the loss amount is given by

$$\text{Loss}(k) = \lambda(k) * ((1-X)^{-1/\theta(k)} - 1) \dots \quad \text{A.1}$$

and remains constant at later valuations. The claim amount for an open claim at delay  $k$  is computed by the equation A.1. However its value keeps changing over time as  $\lambda(k)$  and  $\theta(k)$  are different for each value of  $k$ . By keeping  $X$  constant for an individual claim, the percentile level of the loss amount is fixed. However, the claim amount increases over time as the parameter of the Pareto distribution changes. In actual practice, the claim amount may decrease from an earlier to a later valuation for some claims. The procedure used here will always increase the severity of the loss from one valuation to the next. This is done to force the incremental losses to be positive.

Parameters  $\lambda(k)$  and  $\theta(k)$  are:

$$\lambda(k) = (1000 + (k - 1) * 50) * 1.06^{(k-1)} \text{ and}$$

$$\theta(k) = 2.5 - 0.05 * (k - 1).$$

## **Appendix B**

**TABLE B1**

**Random Reporting Factor**

AY	Dev Factor	Method of Forecast:			
		Buhlmann Loss Ratio	Model 1 Bias	Regression Model 2	Model 3
2	0	0	(0)	0	(0)
3	1	1	(1)	(1)	3
4	3	3	(5)	(4)	7
5	(1)	3	(26)	(23)	(17)
6	21	12	(68)	(82)	(144)
7	60	(10)	(211)	(158)	(607)
8	371	(60)	(383)	(105)	(1,228)
9	1,365	(121)	(844)	1,135	665
10	9,755	1,141	985	7,587	14,377
11	140,106	4,253	37,041	22,871	38,312
<b>Total</b>	<b>151,681</b>	<b>5,222</b>	<b>38,486</b>	<b>31,240</b>	<b>51,367</b>
AY	RMSE				
2	7	8	4	8	5
3	25	31	17	29	25
4	101	122	72	116	107
5	388	478	291	451	425
6	1,371	1,772	1,066	1,682	1,583
7	4,478	6,041	3,722	5,660	5,390
8	13,713	18,441	12,389	17,657	16,815
9	37,972	51,449	37,716	51,583	50,046
10	110,938	117,820	118,183	125,616	126,393
11	441,193	218,265	388,764	258,292	265,187
<b>Total</b>	<b>466,055</b>	<b>266,874</b>	<b>395,819</b>	<b>328,870</b>	<b>341,537</b>
AY	Average Absolute Deviations				
2	5	6	3	6	4
3	19	23	12	21	18
4	74	90	51	85	79
5	288	356	209	332	312
6	1,018	1,295	772	1,198	1,119
7	3,373	4,430	2,750	4,150	3,876
8	10,440	13,784	9,320	13,206	12,359
9	29,290	37,826	28,973	38,281	37,003
10	86,296	88,748	91,752	96,232	97,421
11	346,382	166,051	296,479	197,807	203,655
<b>Total</b>	<b>364,628</b>	<b>204,674</b>	<b>314,829</b>	<b>254,069</b>	<b>263,444</b>
AY	Average Percentage Errors				
2	25.67%	36.00%	9.95%	35.38%	19.36%
3	22.20%	33.13%	9.27%	29.45%	35.19%
4	20.02%	29.68%	8.80%	25.10%	29.48%
5	16.10%	26.60%	7.65%	21.81%	21.16%
6	14.12%	23.31%	7.46%	19.31%	15.14%
7	11.63%	20.28%	6.92%	17.33%	11.96%
8	10.17%	17.67%	7.00%	16.12%	12.08%
9	8.67%	15.01%	6.57%	15.10%	14.19%
10	9.89%	13.16%	7.30%	14.80%	17.14%
11	30.23%	10.57%	12.90%	13.69%	16.36%
<b>Total</b>	<b>16.84%</b>	<b>4.84%</b>	<b>6.22%</b>	<b>6.75%</b>	<b>8.69%</b>

**TABLE B2**

**Random Backward Development Factor**

AY	Dev Factor	Method of Forecast:			
		Buhlmann <sub>1</sub>	Model 1	Regression	
		Loss Ratio		Model 2	Model 3
		↑	Bias		
2	2	11	(22)	(22)	16,639
3	119	208	69	40	32,153
4	222	360	156	(54)	38,008
5	694	(470)	534	(1,165)	27,398
6	1,321	(904)	664	(1,643)	2,799
7	3,705	(643)	1,683	(678)	(26,084)
8	8,940	(3,630)	3,237	(2,019)	(47,440)
9	23,597	1,501	12,390	5,744	(37,721)
10	42,679	(1,854)	16,585	5,805	(15,255)
11	76,406	(2,667)	20,062	9,384	12,628
<b>Total</b>	<b>157,684</b>	<b>(8,088)</b>	<b>55,356</b>	<b>15,393</b>	<b>3,125</b>
<b>AY</b>			<b>RMSE</b>		
2	1,564	2,837	1,591	2,786	17,107
3	4,614	9,596	4,791	9,176	33,712
4	11,084	24,442	11,681	22,880	44,274
5	22,705	50,969	24,318	46,958	52,965
6	40,045	87,472	43,509	78,937	77,117
7	64,981	133,834	70,337	121,306	120,900
8	105,093	188,472	111,145	171,808	174,206
9	154,998	222,490	157,449	210,056	208,665
10	231,105	269,200	222,925	271,969	268,947
11	320,345	287,746	293,720	316,988	317,608
<b>Total</b>	<b>512,092</b>	<b>639,187</b>	<b>481,727</b>	<b>542,257</b>	<b>519,705</b>
<b>AY</b>			<b>Average Absolute Deviations</b>		
2	1,187	2,082	1,206	2,056	16,639
3	3,553	7,119	3,681	6,848	32,192
4	8,525	18,316	8,996	17,259	39,640
5	17,267	37,913	18,525	34,859	43,380
6	30,623	63,847	33,027	57,937	57,143
7	49,699	99,623	53,807	89,929	86,860
8	78,098	133,685	82,422	122,794	119,958
9	115,161	183,563	118,176	156,441	151,540
10	167,451	193,324	163,199	200,545	196,412
11	233,783	211,057	219,467	240,241	239,523
<b>Total</b>	<b>391,022</b>	<b>485,769</b>	<b>373,282</b>	<b>420,438</b>	<b>403,056</b>
<b>AY</b>			<b>Average Percentage Errors</b>		
2	4.73%	13.54%	4.23%	13.27%	360.75%
3	3.43%	12.41%	3.18%	11.80%	171.65%
4	2.61%	11.08%	2.51%	10.42%	77.29%
5	2.30%	9.75%	2.21%	8.90%	31.55%
6	2.01%	8.80%	1.74%	7.86%	9.62%
7	2.25%	9.06%	1.74%	8.10%	0.62%
8	2.65%	8.33%	1.55%	7.37%	-2.07%
9	4.37%	8.81%	2.64%	8.09%	0.93%
10	6.14%	8.52%	2.75%	8.36%	5.41%
11	9.19%	8.02%	2.65%	8.63%	9.01%
<b>Total</b>	<b>4.38%</b>	<b>1.23%</b>	<b>1.58%</b>	<b>0.79%</b>	<b>0.47%</b>

**TABLE B3**

**Individual Losses with Changing Severity**

AY	Dev Factor	Method of Forecast:			Model 3
		Buhlmann Loss Ratio	Model 1 Bias	Regression Model 2	
2	1,048	3,668	788	3,751	(9,394)
3	2,867	(5,109)	508	(5,560)	(19,900)
4	(25,022)	(28,645)	(27,849)	(28,329)	(43,090)
5	36,381	(18,287)	17,366	(17,199)	(31,036)
6	29,767	40,903	24,706	39,074	23,710
7	(84,522)	(77,462)	(72,000)	(80,096)	(93,727)
8	40,487	6,886	16,237	6,239	(2,950)
9	3,817	9,224	(7,890)	13,152	8,934
10	8,224	41,085	(9,994)	52,394	46,612
11	(2,480)	(55,304)	(86,063)	(35,753)	(55,246)
<b>Total</b>	<b>30,566</b>	<b>(83,039)</b>	<b>(144,192)</b>	<b>(52,327)</b>	<b>(176,089)</b>
<b>AY</b>			<b>RMSE</b>		
2	31,330	36,022	30,391	35,894	14,678
3	52,869	43,391	51,745	43,108	38,400
4	115,037	128,467	114,408	125,039	118,422
5	171,424	78,184	130,753	78,614	75,765
6	61,614	69,934	58,478	64,412	40,414
7	222,605	230,023	221,536	232,337	262,153
8	110,891	107,524	86,577	95,115	80,851
9	117,997	55,953	119,472	77,183	74,298
10	94,518	97,249	96,401	90,883	86,043
11	259,037	222,586	280,944	173,258	169,985
<b>Total</b>	<b>413,137</b>	<b>441,109</b>	<b>375,367</b>	<b>299,099</b>	<b>340,506</b>
<b>AY</b>			<b>Average Absolute Deviations</b>		
2	19,830	24,090	18,965	24,010	12,509
3	37,669	31,125	38,969	31,640	24,974
4	72,059	77,787	70,388	76,060	62,687
5	105,641	59,338	88,162	61,260	54,471
6	35,254	45,292	33,024	44,186	31,115
7	111,397	120,745	106,008	113,100	119,757
8	94,362	87,744	76,900	78,766	68,943
9	90,948	48,065	88,693	66,927	63,907
10	81,845	75,610	77,164	75,122	66,540
11	202,172	161,009	209,386	135,948	127,446
<b>Total</b>	<b>356,932</b>	<b>347,340</b>	<b>314,629</b>	<b>259,057</b>	<b>280,243</b>
<b>AY</b>			<b>Average Percentage Errors</b>		
2	47.87%	94.00%	44.07%	95.99%	-17.10%
3	49.86%	20.49%	42.22%	20.10%	-13.53%
4	57.18%	63.35%	48.78%	62.45%	16.07%
5	119.19%	12.26%	80.54%	16.42%	0.49%
6	56.20%	77.73%	48.35%	74.57%	48.74%
7	5.03%	-0.81%	0.06%	-3.61%	-7.82%
8	42.12%	29.37%	27.35%	25.73%	17.56%
9	15.32%	8.67%	9.34%	14.34%	13.17%
10	6.43%	21.75%	0.22%	25.76%	23.66%
11	20.04%	3.88%	-2.87%	5.32%	-1.39%
<b>Total</b>	<b>1.39%</b>	<b>-4.36%</b>	<b>-9.49%</b>	<b>-3.31%</b>	<b>-9.52%</b>

**TABLE B4**

**Pentikainen and Rantala Method**

AY	Dev Factor	Method of Forecast:			
		Buhlmann Loss Ratio	Model 1 Bias	Regression Model 2	Model 3
2	2	1	(2)	11	6,037
3	26	(44)	12	11	9,160
4	114	(52)	78	118	9,389
5	283	(217)	213	190	7,237
6	313	(880)	197	182	(755)
7	456	(1,090)	222	473	(9,710)
8	1,225	(2,113)	855	549	(14,271)
9	1,837	(3,756)	1,257	497	(7,450)
10	3,050	(5,083)	1,984	1,456	7,505
11	2,801	(8,407)	509	1,322	26,995
<b>Total</b>	<b>10,106</b>	<b>(21,441)</b>	<b>5,326</b>	<b>4,789</b>	<b>34,136</b>
AY	RMSE				
2	672	608	620	607	6,101
3	1,897	1,724	1,759	1,733	9,399
4	4,070	3,650	3,822	3,687	10,222
5	7,375	6,387	6,997	6,478	9,896
6	12,478	10,855	11,986	10,987	10,942
7	19,710	17,324	19,117	17,606	19,634
8	29,727	26,563	29,027	27,007	29,730
9	40,540	37,157	39,974	37,662	37,813
10	54,987	51,968	54,481	53,464	54,281
11	74,329	71,882	73,934	76,120	82,500
<b>Total</b>	<b>186,688</b>	<b>186,916</b>	<b>183,351</b>	<b>195,148</b>	<b>201,012</b>
AY	Average Absolute Deviations				
2	536	482	493	480	6,037
3	1,499	1,363	1,388	1,388	9,160
4	3,229	2,912	3,029	2,924	9,401
5	5,839	5,070	5,547	5,105	8,119
6	9,887	8,663	9,495	8,759	8,728
7	15,628	13,742	15,153	13,986	15,864
8	23,414	21,143	22,868	21,350	23,951
9	31,748	29,300	31,319	29,800	30,088
10	43,420	41,011	43,107	41,990	42,469
11	58,767	56,865	58,504	60,077	64,294
<b>Total</b>	<b>147,536</b>	<b>147,830</b>	<b>145,029</b>	<b>153,675</b>	<b>157,283</b>
AY	Average Percentage Errors				
2	0.42%	0.51%	0.32%	0.56%	89.77%
3	0.38%	0.25%	0.29%	0.34%	40.30%
4	0.44%	0.35%	0.35%	0.48%	18.60%
5	0.46%	0.23%	0.37%	0.38%	7.81%
6	0.30%	0.06%	0.21%	0.25%	-0.35%
7	0.25%	0.08%	0.15%	0.30%	-3.72%
8	0.37%	0.00%	0.26%	0.25%	-3.71%
9	0.34%	-0.13%	0.23%	0.15%	-1.36%
10	0.40%	-0.08%	0.25%	0.22%	1.06%
11	0.26%	-0.22%	0.02%	0.16%	2.82%
<b>Total</b>	<b>0.23%</b>	<b>-0.24%</b>	<b>0.07%</b>	<b>0.06%</b>	<b>0.98%</b>

1

2

3

4