Two Approaches to Calculating Correlated Reserve Indications Across Multiple Lines of Business

by
Gerald S. Kirschner, FCAS, MAAA
Colin Kerley, FIA
Belinda Isaacs, FIAA

Abstract

As reserving actuaries focus more on reserve ranges and less on point estimates, the question of how to develop a reasonable reserve range in the aggregate becomes more and more relevant. When working with a single set of "best estimates", the answer is simple - assuming all the best estimates are the mean values for each block of business being analyzed, the best estimate for the total is equal to the sum of the parts. However, if the by line best estimates are other than the mean values, the sum of the parts is not the same as the best estimate for the aggregate.

This paper presents two possible approaches to developing aggregate reserve indications when looking at results other than the mean value. The approaches both rely on a simulation model. One takes in the actuary's judgment as to the correlations between the different underlying blocks of business and the second uses bootstrapping to eliminate the need for the actuary to make judgment calls about the nature of the correlations.

I. Introduction

The bar continues to be raised for actuaries performing reserve analyses. For example, the approval of Actuarial Standard of Practice # 36 for United States actuaries clarifies and codifies the requirements for actuaries producing "written statements of actuarial opinion regarding property/casualty loss and loss adjustment expense reserves."[1] A second example in the United States is the National Association of Insurance Commissioners requirement that companies begin booking management's best estimate of reserves by line and in the aggregate, effective January 2001. A third example, this one from Australia, is contained in the Australian Prudential Regulation Authority's (APRA) General Insurance Prudential Standards, applicable from July 2002

[1] Actuarial Standard of Practice # 36, page 1.

onwards. In these regulations, APRA specifically states "the Approved Actuary must provide advice on the valuation of insurance liabilities at a given level of sufficiency – **that level is 75%**."[2]

In this environment, it is clear that actuaries are being asked to do more than ever before with regard to reserve analyses. One set of techniques that has been of substantial interest to the paper-writing community for quite some time is the use of stochastic analysis or simulation models to analyze reserves. Stochastic methods[3] are an appealing approach to answering the questions currently being asked of reserving actuaries. One might ask, "Why? What makes stochastic methods more useful in this regard than the traditional reserving methods that I've been using for years?"

The answer is not that the stochastic methods are *better* than the traditional methods.[4] Rather the stochastic methods are more informative about more aspects of reserve indications than traditional methods. When all an actuary is looking for is a point estimate, then traditional methods are quite sufficient to the task. However, when an actuary begins developing reserve ranges for one or more lines of business, and trying to develop not only ranges on a by-line basis, but in the aggregate, the traditional methods quickly pale in comparison to the stochastic methods. The creation of reserve ranges from point estimate methods is often an ad-hoc one, such as looking at results using different selection factors or different types of data (paid, incurred, separate claim frequency and severity development, etc.) or judgmentally saying something like "my best estimate plus or minus ten percent." When trying to develop a range in the aggregate, the ad-hoc decisions become even more so, such as "I'll take the sum of my individual ranges less X% because I know the aggregate is less risky than the sum of the parts."

Stochastic methods, by contrast, provide actuaries with a structured, mathematically rigorous approach to quantifying the variability around a best

---

[2] Australia Prudential Regulation Authority, Prudential Standard GPS 210, "Liability Valuation for General Insurers", July 2002 p. 2.

[3] In this paper we use the word stochastic to mean frameworks that are not deterministic, i.e. have a random component. This is typically done by creating a framework for the reserving technique where many previously fixed quantities are represented by random variables. Probability distributions may then be generated for claims reserves, either analytically or by Monte-Carlo simulation.

[4] When we talk about "traditional methods", we mean the time-honored tradition of analyzing a triangle of paid or incurred loss data by looking at different averages of age-to-age development factors, selecting one for each development age and projecting paid or incurred losses to "ultimate" using the selected factors. There are many variations on this basic approach that can be applied, including data adjustments (like Berquist-Sherman), factor modifications (like Bornheutter-Ferguson), trend removal, but at the end of the day, the traditional methods all produce one reserve indication with no information as to how reality might differ from that single indication .

estimate. This is not meant to imply that all judgment is eliminated when a stochastic method is used. There are still many areas of judgment that remain, such as the choice of stochastic method and / or the shape of the distributions underlying the method, and the number of years of data being used to fit factors. What stochastic methods do provide is (a) a consistent framework and a repeatable process in which the analysis is done and (b) a mathematically rigorous answer to questions about probabilities and percentiles. Now, when asked to set reserves equal to the 75th percentile, as in Australia, the actuary has a mechanism for identifying the 75th percentile. Moreover, when the actuary analyzes the same block of business a year later, the actuary will be in a position to discuss how the 75th percentile has changed, knowing that the changes are driven by the underlying data and not the application of different judgmental factors (assuming the actuary does not alter the assumptions underlying the stochastic method being used).

It cannot be stressed enough, though, that stochastic models are not crystal balls. Quite often the argument is raised that the promise of stochastic models is much greater than the benefit they provide. The arguments typically take on one or both of the following forms:

1. Stochastic models do not work very well when data is sparse or highly erratic. Or, to put another way, stochastic models work well when there is a lot of data and it is fairly regular – exactly the situation in which it is easy to apply a traditional point-estimate approach.
2. Stochastic models overlook trends and patterns in the data that an actuary using traditional methods would be able to pick up and incorporate into the analysis.

England and Verrall, in their 2002 paper, addressed this sort of argument with the response "It is sometimes rather naïvely hoped that stochastic methods will provide solutions to problems when deterministic methods fail. Indeed, sometimes stochastic models are judged on whether they can help when simple deterministic models fail. This rather misses the point. The usefulness of stochastic models is that they can, in many circumstances, provide more information which may be useful in the reserving process and in the overall management of the company."[5] This, in our opinion, is the essence of the value proposition for stochastic models. They are not intended to replace traditional techniques. There will always be a need and a place for actuarial judgment in reserve analysis that stochastic models will never supplant. Even so, as the bar is raised for actuaries performing reserve analyses, the additional information inherent in stochastic models makes the argument in favor of adding them to the standard actuarial repertoire that much more compelling.

---

[5] England & Verrall (2002), pp. 2-3.

Having laid the foundation for why we believe actuaries ought to be incorporating stochastic models into their everyday toolkit, let us turn to the actual substance of this paper – using a stochastic model to develop an aggregate reserve range for several lines of business with varying degrees of correlation between the lines.

II. Correlation – mathematically speaking and in lay terms

Before jumping into the case study, we will take a small detour into the mathematical theory underlying correlation.

Correlations between observed sets of numbers are a way of measuring the "strength of relationship", between the sets of numbers. Broadly speaking, this "strength of relationship" measure is a way of looking at the tendency of two variables, X and Y, to move in the same (or opposite) direction. For example, if X and Y were positively correlated then if X gives a higher than average number, we would expect Y to give a higher than average number as well.

It should be mentioned that there are many different ways to measure correlation, both parametric (for example, Pearson's R) and non-parametric (Spearman's Rank Order, or Kendall's tau). It should also be mentioned that these statistics only give a simple view of the way two random variables behave together – to get more detailed picture, we would need to understand the joint probability density function (pdf) of the two variables.

As an example of correlation between two random variables we will look at the results of flipping two coins, and look at the relationship between correlation coefficients and conditional probabilities.

Example 1

We have two coins, each with an identical chance of getting heads (50%) or tails (50%) with a flip. We will specify their joint distribution, and so determine the relationship between the outcomes of both coins. Note that in our notation, 0 signifies a Head, 1 a Tail.

Case 1:

Case 1 Joint Distribution Table

		Coin B		
		0	**1**	**Marginal**
Coin A	**0**	0.25	0.25	0.5
	1	0.25	0.25	0.5
	Marginal	0.5	0.5	

The joint distribution table shows the probability of all the outcomes when the two coins are tossed. In the case of two coin tosses there are 4 potential outcomes, hence, there are 4 cells in the joint distribution table. For example, the probability of Coin A being a head (0) and Coin B a tail (1) can be determined by looking at the 0 row for Coin A and the 1 column for Coin B, in this example 0.25. In this case, our coins are independent. The correlation coefficient is zero, where we calculate the correlation coefficient by:

*Correlation Coefficient = Cov(A, B) / (Stdev(A) * Stdev(B))* *(2.1)*

and

*Cov(A, B) = E[(A – mean(A)) * (B – mean(B))] = E(AB) – E(A)E(B)* *(2.2)*

We can also see that the outcomes of coin B are not linked in any way to the outcome of coin A. For example

P(B = 1 | A = 1) = P(A = 1, B = 1) / P(A = 1)
= 0.25 / 0.5
= 0.50
= P(B = 1)

Case 2:

Case 2 Joint Distribution Table

		Coin B		
		0	**1**	**Marginal**
Coin A	**0**	0.3125	0.1875	0.5
	1	0.1875	0.3125	0.5
	Marginal	0.5	0.5	

From this distribution we calculate the correlation coefficient to be 0.25[6].

By looking at the conditional distributions, it is clear that there is a link between the outcome of coin B and coin A :

P(B = 1 | A = 1) = P(A = 1, B = 1) / P(A = 1)
= 0.3125 / 0.5 = 0.625

P(B = 0 | A = 1) = 0.375

So we can see that with the increase in correlation, there is an increase in the chance of getting heads on coin B given coin A shows heads.

With this 2 coin example it turns out that if we want the marginal distributions of each coin to be the standard 50% heads, 50% tails then, given the correlation coefficient we want to produce, we can uniquely define the joint pdf for the coins.

We find that, for a given correlation coefficient of ρ,

P(A = 1, B = 1) = P(A = 0, B = 0) = (1 + ρ) / 4
P(A = 1, B = 0) = P(A = 0, B = 1) = (1 - ρ) / 4

We can then recover the conditional probabilities :

P(B = 1 | A = 1) = (1 + ρ) / 2
P(B = 0 | A = 1) = (1 - ρ) / 2

So for example, we can see that

ρ = 0.50 gives P(B = 1 | A = 1) = 0.750.
ρ = 0.75 gives P(B = 1 | A = 1) = 0.875.
ρ = 1.00 gives P(B = 1 | A = 1) = 1.000.

---

[6] Proof that the correlation coefficient for case 2 is 0.25:

E(A,B) = $\sum_{i=0}^{1}\sum_{j=0}^{1} i * j * P(A_i B_j)$ = 0 + 0 + 0 + 1*1*0.3125 = 0.3125
E(A) = 0.5 = E(B)
Cov(A,B) = E(A,B) – E(A)E(B) = 0.3125 – 0.25 = 0.0625
Var(A) = $\sum_{i=0}^{1} (i - E(A))^2 * P(A_i)$ = 0.25 = Var(B)
StDev(A) = 0.5 = StDev(B)
Correlation Coefficient = 0.0625 / (0.5 * 0.5) = 0.25

As expected, an increase in the correlation coefficient means the higher the chance of throwing heads on coin B, given coin A shows heads.

In lay terms, then, we would repeat our description of correlation at the start of this section, that correlation, or the "strength of relationship" is a way of looking at the tendency of two variables, X and Y, to move in the same (or opposite) direction. As the coin example shows, the more positively correlated X and Y are, the greater our expectation that Y will be higher than average if X is higher than average.

It should be noted, however, that the expected value of the sum of two correlated variables is exactly equal to the expected value of the sum of the two uncorrelated variables with the same means.

III. Significance of the existence of correlations between lines of business

Suppose we have two or more blocks of business for which we are trying to calculate reserve indications. If all we are trying to do is determine the expected value of the reserve run-off, we can calculate the expected value for each block separately and add all the expectations together. However, if we are trying to quantify a value other than the mean, such as the 75th percentile, we cannot simply sum across the lines of business. If we do so, we will overstate the aggregate reserve need. The only time the sum of the 75th percentiles would be appropriate for the aggregate reserve indication is when all the lines are fully correlated with each other – a highly unlikely situation! The degree to which the lines are correlated will influence the proper aggregate reserve level and the aggregate reserve range. How significant an impact will there be? That primarily depends upon two factors – how volatile the reserve ranges are for the underlying lines of business and how strongly correlated the lines are with each other. If there is not much volatility, then the strength of the correlation will not matter that much. If, however, there is considerable volatility, the strength of correlations will produce differences that could be material. This is demonstrated in the following example.

Example 2: The impact on values at the 75th percentile as correlation and volatility increase

Table 1 shows some figures relating the magnitude of the impact of correlations on the aggregate distribution to the size of the correlation. In this example, we have modeled two lines of business (A and B) assuming they were Normally distributed with identical means and variances. The means

were assumed to be 100 and the standard deviations were 25. We are examining the 75th percentile value derived for the sum of A and B. The tables below show the change in the 75th percentile value between the uncorrelated situation and varying levels of correlation between lines A and B. Reading down the column shows the impact of an increasing level of correlation between lines A and B, namely that the ratio of the correlated to the uncorrelated value at the 75th percentile increases as correlation increases.

Table 1: Comparison of values at the 75th percentile as correlation increases

Correlation	Values at 75th percentile	Ratio of Values at 75th percentile
0.00	223.8	0.0%
0.25	226.7	1.3%
0.50	229.2	2.4%
0.75	231.5	3.4%
1.00	233.7	4.4%

Now let's expand the analysis to see what happens as the volatility of the underlying distributions increase. Table 2 shows a comparison of the sum of lines A and B at the 75th percentile as correlation increases and as volatility increases. The ratios in each column are relative to the value for the zero correlation value at each standard deviation value. For example, the 5.8% ratio for the rightmost column at the 25% correlation level means that the 75th percentile value for lines A + B with 25% correlation is 5.8% higher than the 75th percentile of $N(100,200)_A + N(100,200)_B$ with no correlation. As can be seen from this table, the greater the volatility, the larger the differential between the uncorrelated and correlated results at the 75th percentile.

Table 2: Comparison of values at the 75th percentile as both correlation and volatility increase

	Standard Deviation Value			
	25	50	100	200
Value for 0.00 correlation at the 75th percentile	223.8	247.7	295.4	390.8
Correlation	Ratio of values at 75th percentile			
0.25	1.3%	2.3%	3.8%	5.8%
0.50	2.4%	4.3%	7.3%	11.0%
0.75	3.4%	6.2%	10.4%	15.8%
1.00	4.4%	8.0%	13.4%	20.2%

This effect is magnified if we look at similar results but further out on the tails of the distribution, for example looking at the 95th percentiles, as is shown in Table 3.

Table 3: Comparison of values at the 95th percentile as both correlation and volatility increases

	Standard Deviation Value			
	25	50	100	200
Value for 0.00 correlation at the 75th percentile	258.1	316.3	432.6	665.2
Correlation	Ratio of values at 75th percentile			
0.25	2.7%	4.3%	6.3%	8.3%
0.50	5.1%	8.3%	12.1%	15.7%
0.75	7.3%	11.9%	17.4%	22.6%
1.00	9.3%	15.2%	22.3%	29.0%

Note that these results will also depend on the nature of the underlying distributions – we would expect different results for lines of business that were Lognormally distributed for example.

IV. Case Study

A. Background

The data used in this case study is fictional. It describes three lines of business, two long tail and one short tail. All three produce approximately the same mean reserve indication, but with varying degrees of volatility around their respective means. By having the three lines of approximately equal size, we can are able to focus on the impact of correlations between lines without worrying about whether the results from one line is overwhelming the results from the other two lines.

Appendix I contains the data triangles.

The examination of the impact of correlation on the aggregated results will be done using two methods. The first assumes the person doing the analysis can provide a *positive-definite* correlation matrix (see section B below). The relationships described in the correlation matrix are used to convert the uncorrelated aggregate reserve range into a correlated aggregate range. The process does not affect the reserve ranges of the underlying lines of business. It just influences the aggregation of the reserve indications by line so that if

two lines are positively correlated and the first line produces a reserve indication that is higher than the expected reserve indication for that line, it is more likely than not that the second line will also produce a reserve indication that is higher than its expected reserve indication. This is exactly what was demonstrated in the coin examples in Section II.

The second method dispenses with what the person doing the analysis knows or thinks he or she knows. This method relies on the data alone to derive the relationships and linkages between the different lines of business. More precisely, this method assumes that all we need to know about how related the different lines of business are to each other is contained in the historical claims development that we have already observed. This method uses a technique known as bootstrapping to extract the relationships from the observed claims history. The bootstrapped data is used to generate reserve indications that inherently contain the same correlations as existed in the original data. Therefore, the aggregate reserve range is reflective of the underlying relationships between the individual lines of business, without first requiring the potentially messy step of requiring the person doing the analysis to develop a correlation matrix.

B. A note on the nature of the correlation matrix used in the analysis

The entries in the correlation matrix used must fulfill certain requirements that cause the matrix to be what is known as *positive definite*. The mathematical description of a positive definite matrix is that, given a vector *x* and a matrix A, where

$x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$ and

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

$$x^T A x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + \cdots + a_{nn}x_n^2 \qquad (4.1)$$

Matrix A is positive definite when $x^T A x > 0$
for all *x* other than $x_1 = x_2 = \cdots = x_n = 0$. *(4.2)*

C. Correlation matrix methodology

The methodology used in this approach is that of **rank correlation**. Rank correlation is a useful approach to dealing with two or more correlated variables when the joint distribution of the correlated variables is not normal. When using rank correlation, what matters is the ordering of the simulated outcomes from each of the individual distributions, or more properly, the *re-ordering* of the outcomes.

Rank Correlation Example

Suppose we have two random variables, A and B. A and B are both defined by uniform distributions ranging from 100 to 200. Suppose we draw five values at random from A and B. They might look as shown in Table 4:

Table 4

Index	A	B
1	155	154
2	138	125
3	164	100
4	122	198
5	107	128

Now suppose we are interested in the joint distribution of A + B. We will use rank correlation to learn about this joint distribution. We will use a bivariate normal distribution to determine which value from distribution B ought to be paired with a value from distribution A. The easiest cases are when B is perfectly correlated with A or perfectly inversely correlated with A. In the perfectly correlated case, we pair the lowest value from A with the lowest value from B, the second lowest value from A with the second lowest value from B, and so on and so forth to the highest values for A and B. In the case of perfect inverse correlation, we pair the lowest value from A with the highest value from B, etc, etc, etc. The results from these two cases are shown in Table 5.

Table 5

Perfectly Correlated			**Perfect Inverse Correlation**		
Rank to Use			**Rank to Use**		
A	**B**		**A**	**B**	
1	1		1	5	
2	2		2	4	
3	3		3	3	
4	4		4	2	
5	5		5	1	
Resulting Joint Distribution			**Resulting Joint Distribution**		
A	**B**	**A+B**	**A**	**B**	**A+B**
107	100	207	107	198	305
122	125	247	122	154	276
138	128	266	138	128	266
155	154	309	155	125	280
164	198	362	164	100	264
Range of Joint Distribution			**Range of Joint Distribution**		
Low		207	**Low**		264
High		362	**High**		305

When there is no correlation between A and B, the ordering of the values from distribution B that are to be paired with values from distribution A are wholly random. The original order of the values drawn from distributions A and B is one example of the no-correlation condition. When positive correlations exist between A and B, the orderings reflect the level of correlation and the range of the joint distribution will be somewhere between the wholly random situation and the perfectly correlated one.

Application of Rank Correlation Methodology to Reserve Analysis

The application of the rank correlation methodology to a stochastic reserve analysis is done through a two-step process. In the first step, a stochastic reserving technique is use to generate N possible reserve runoffs from each data triangle being analyzed. Examples of several different techniques, including bootstrapping (England 2001), application of the chain-ladder to logarithmically adjusted incremental paid data (Christofides 1990) and application of the chain-ladder to logarithmically adjusted cumulative paid data (Feldblum 1999) can be found in articles listed in the bibliography to this paper. In this case study, 5,000 different reserve runoffs were produced using

the bootstrapping technique described in England (2001). This is the end of step one.

In step two, the user must specify a correlation matrix. We do not propose to cover how one may estimate such a correlation matrix in this paper, as we feel this is an important topic in its own right, the details of which would merit a separate paper. One such paper for readers who are looking for guidance in this area is "Correlation and the Aggregation of Unpaid Loss Distributions" by Paul Brehm, due to be published as part of the CAS 2002 Reserving Call Papers. In this paper, we will simply assume that the user has such a matrix, either calculated analytically, or estimated using some other approach, such as a judgmental estimation of correlation.

We generate 5,000 samples for each line of business from a multivariate normal distribution, with the correlation matrix specified by the user. We then sort the samples from the reserving method into the same rank order as the normally distributed samples. This ensures that the rank order correlations between the three lines of business are the same as the rank order correlations between the three normal distributions. The aggregate reserve distribution is calculated from the sum of the individual line reserve distributions. This resulting aggregate reserve range will be composed of 5,000 different values from which statistics such as the 75th percentile can be drawn. The range of aggregated reserve indications is reflective of the correlations entered into the correlation matrix at the start of the analysis.

For example, the ranked results from the multivariate normal process might be as follows:

Line 1, Rank	**Line 2, Rank**	**Line 3, Rank**
528	533	400
495	607	404
995	710	904
233	325	831
733	912	551
825	33	801
325	107	331
630	210	571
653	986	51
983	730	301
130	900	782

The first of the 5,000 values in the aggregate reserve distribution will be composed of the 528th largest reserve indication for line 1 + the 533rd largest reserve indication for line 2 + the 400th largest reserve indication for line 3.

The second of the thousand values will be composed of the 495th largest reserve indication for line 1 + the 607th largest reserve indication for line 2 + the 404th largest reserve indication for line 3. Through this process, the higher the positive correlation between lines, the more likely it is that a value below the mean for one line will be combined with a value below the mean for a second line. At the same time, the mean of the overall distribution remains unchanged and the distributions of the individual lines remains unchanged.

D. Rank correlation results

To show the impact of the correlations between the lines on the aggregate distribution, we ran the model five times, each time with a different correlation value: zero correlation, 25% correlation, 50% correlation, 75% correlation, and 100% correlation. The correlations were chosen to highlight the range of outcomes that result for different levels of correlation, not because the data necessarily implied the existence of correlations such as these. The results are shown both numerically in Table 6 and graphically in Figure 1 and

Figure 2.

Table 6: Case study results – aggregated reserve indication at different levels of correlation between underlying lines of business (all values are in thousands)

	0% corr	25% corr	50% corr	75% corr	100% corr
Mean	4,330,767	4,330,767	4,330,767	4,330,767	4,330,767
Standard Deviation	1,510,033	1,596,840	1,705,469	1,829,748	1,998,140
Minimum	2,587,213	2,293,224	2,084,841	2,086,531	1,930,725
Maximum	72,366,202	72,771,841	73,474,899	75,564,417	81,277,681
1% ile	2,995,943	2,861,958	2,695,429	2,510,514	2,408,319
5% ile	3,247,847	3,087,062	2,956,837	2,867,115	2,762,663
10% ile	3,384,401	3,241,518	3,143,080	3,033,779	2,987,948
20% ile	3,588,011	3,500,438	3,424,399	3,358,196	3,277,806
30% ile	3,782,986	3,681,105	3,615,534	3,574,383	3,522,031
40% ile	3,942,032	3,897,816	3,820,380	3,790,977	3,745,674
50% ile	4,113,146	4,078,681	4,071,349	4,027,615	3,973,908
60% ile	4,278,521	4,279,869	4,292,852	4,267,561	4,232,721
70% ile	4,493,139	4,518,971	4,547,255	4,558,175	4,560,471
80% ile	4,786,940	4,876,233	4,931,662	5,031,358	5,111,862
90% ile	5,378,096	5,475,577	5,604,519	5,679,109	5,842,125
95% ile	6,008,476	6,230,885	6,371,310	6,436,050	6,836,095
99% ile	8,286,504	8,687,785	9,310,024	10,075,891	10,322,456

	0% corr	25% corr	50% corr	75% corr	100% corr
Estimated 75th %ile	4,640,039	4,697,602	4,739,459	4,794,767	4,836,166

Figure 1: Graph of case study results – aggregated reserve indication at different levels of correlation between underlying lines of business

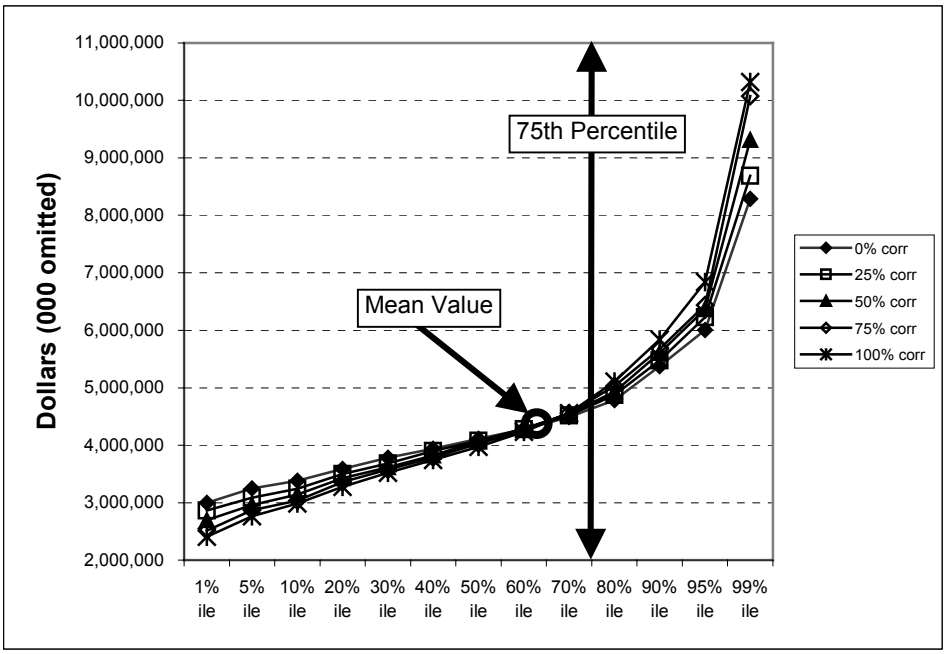
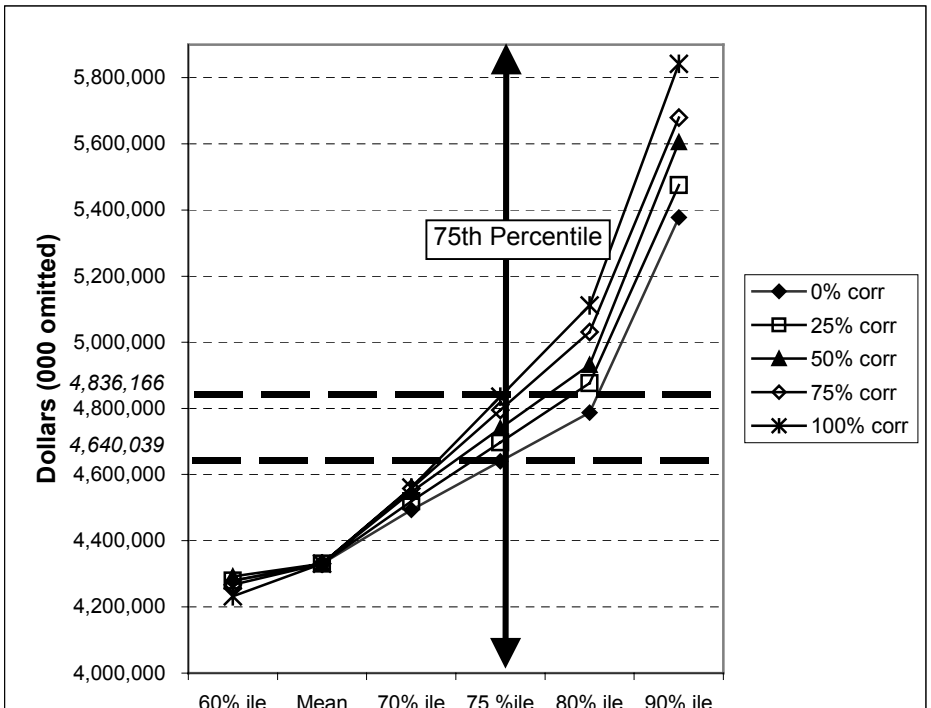


Figure 2: Graph of case study results – aggregated reserve indication at different levels of correlation between underlying lines of business – blowing up area around 75th percentile



As expected, the higher the positive correlation, the wider the aggregated reserve range. With increasingly higher positive correlations, it is less likely that a better than expected result in one line will be offset by a worse than expected result in another line. This causes the higher positive correlated situations to have lower aggregate values for percentiles below the mean and higher aggregate values for percentiles above the mean. The results of the table and graph show just this situation. For information purposes, the difference between the zero correlation situation and the perfectly correlated situation have been displayed on the graph in

Figure 2.

E. Bootstrap methodology

Bootstrapping is a sampling technique that is an alternative to traditional statistical methodologies. In traditional statistical approaches, one might look at a sample of data and postulate the underlying distribution that gave rise to the observed outcomes. Then, when analyzing the range of possible outcomes, new samples are drawn from the postulated distribution. Bootstrapping, by comparison, does not concern itself with the underlying distribution. The bootstrap says that all the information needed to create new samples lies within the variability that exists in the already observed historical data. When it comes time to create the new samples, different observed variability factors are combined with the observed data to create "pseudo-data" from which the new samples are generated. A step-by-step description of the application of the bootstrap methodology to claims reserving is contained in Appendix II.

So what is bootstrapping, then, as it is applied to reserve analysis? Bootstrapping is a resampling method that is used to estimate in a structured manner, the variability of a parameter. In reserve analysis, the parameter is the difference between observed and expected paid amounts for any given accident year / development year combination. During each iteration of the bootstrapping simulation, random draws are made from all the available variability parameters. One random draw is made for each accident year / development year combination. The variability parameter is combined with the actual observation to develop a "pseudo-history" paid loss triangle A reserve indication is then produced from the pseudo-history data triangle by applying the traditional cumulative chain-ladder technique to "square the triangle". A step-by-step walkthrough of the bootstrap process is included in Appendix II.

Note that this example is using paid amounts. The bootstrap approach can equally be applied to incurred data, to generate "pseudo-history" incurred loss

triangles, which may be developed to ultimate in the same manner as the paid data. Also, the methodology is not limited to working with just positive values. This is an important capability when using incurred data, as negative incrementals will be much more common when working with incurred data.

This approach is extended to multiple lines in the following manner. Instead of making random draws of the variability parameters independently for each line of business, the same draws are used across all lines of business. The variability parameters will differ from line to line, but the choice of which variability parameter to pick is the same across lines.

The example of Table 7 - Table 9 should clarify the difference between the uncorrelated and correlated cases. The example shows two lines of business, Line A and Line B. Both are 4x4 triangles. Table 7 shows the variability parameters calculated from the original data. We start by labeling each parameter with the accident year, development year and triangle from which the parameters are derived.

Table 8 shows one possible way the variability parameters might be reshuffled to create an uncorrelated bootstrap. For each Accident / Development year in each triangle A and B, we select a variability parameter from Table 7 at random. For example Triangle A, Accident Year 1, Development Year 1 has been assigned (randomly) the variability parameter from the original data in Table, Accident Year 2, Development Year 1. Note that each triangle uses the variability parameters calculated from that triangle's data, i.e. none of the variability parameters from Triangle A are used to create the pseudo-history in Triangle B. Also note that the choice of variability parameters for each Accident Year / Development Year in Triangle A is independent of the choice of variability parameter for the corresponding Accident Year / Development Year in Triangle B.

For the correlated bootstrap shown in Table 8, the choice of variability parameter for each Accident Year / Development Year in Triangle A is *not* independent of the choice of variability parameter for the corresponding Accident Year / Development Year in Triangle B. We ensure that the variability parameter selected from Triangle B comes from the same Accident Year / Development Year used to select a variability parameter from Triangle A.

The process shown in Table 9 implicitly captures and uses whatever correlations existed in the historical data when producing the pseudo-histories from which the reserve indications will be developed. The resulting aggregated reserve indications will reflect the correlations that existed in the actual data, without requiring the analyst to first postulate what those correlations might be. This method also does not require the second stage

reordering process that the correlation matrix methodology required. The correlated aggregate reserve indication can be derived in one step.

Table 7

Variability Parameters Calculated from Original Data

Triangle A

AY	Development Year 1	2	3	4
1	$(1_A,1_A)$	$(1_A,2_A)$	$(1_A,3_A)$	$(1_A,4_A)$
2	$(2_A,1_A)$	$(2_A,2_A)$	$(2_A,3_A)$	
3	$(3_A,1_A)$	$(3_A,2_A)$		
4	$(4_A,1_A)$			

Triangle B

AY	Development Year 1	2	3	4
1	$(1_B,1_B)$	$(1_B,2_B)$	$(1_B,3_B)$	$(1_B,4_B)$
2	$(2_B,1_B)$	$(2_B,2_B)$	$(2_B,3_B)$	
3	$(3_B,1_B)$	$(3_B,2_B)$		
4	$(4_B,1_B)$			

Calculated Variability Parameters

Uncorrelated Bootstrapping - Reshuffling of variability parameters in Triangle B is independent of the reshuffling in Triangle A

AY	Development Year 1	2	3	4
1	$(2_A,1_A)$	$(3_A,2_A)$	$(1_A,3_A)$	$(3_A,1_A)$
2	$(2_A,2_A)$	$(2_A,3_A)$	$(1_A,2_A)$	
3	$(3_A,1_A)$	$(1_A,1_A)$		
4	$(1_A,1_A)$			

AY	Development Year 1	2	3	4
1	$(2_B,2_B)$	$(3_B,2_B)$	$(1_B,3_B)$	$(2_B,2_B)$
2	$(3_B,1_B)$	$(2_B,3_B)$	$(2_B,2_B)$	
3	$(1_B,3_B)$	$(1_B,1_B)$		
4	$(1_B,2_B)$			

Randomly Selected Variability Parameters to be used in the creation of one possible pseudo-history

Table 8

Table 9

Correlated Bootstrapping - Reshuffling of variability parameters in Triangle B is identical to the reshuffling in Triangle A

AY	Development Year 1	2	3	4
1	$(2_A,1_A)$	$(3_A,2_A)$	$(1_A,3_A)$	$(3_A,1_A)$
2	$(2_A,2_A)$	$(1_A,2_A)$	$(2_A,3_A)$	
3	$(3_A,1_A)$	$(1_A,1_A)$		
4	$(1_A,1_A)$			

AY	Development Year 1	2	3	4
1	$(2_B,1_B)$	$(3_B,2_B)$	$(1_B,3_B)$	$(3_B,1_B)$
2	$(2_B,2_B)$	$(1_B,2_B)$	$(2_B,3_B)$	
3	$(3_B,1_B)$	$(1_B,1_B)$		
4	$(1_B,1_B)$			

Randomly Selected Variability Parameters to be used in the creation of one possible pseudo-history

F. Bootstrap results

The model was run one final time using the bootstrap methodology to develop an aggregated reserve range. The bootstrap results have been added to the results shown in Table 6, Figure 1 and

Figure 2. The revised results are shown in Table 10,

Figure 3, and Figure 4 where we can compare the aggregate reserve distributions generated from the two different approaches.

Table 10 (all values are in thousands)

	0% corr	25% corr	50% corr	75% corr	100% corr	Bootstrap
Mean	4,330,767	4,330,767	4,330,767	4,330,767	4,330,767	4,335,587
Standard Deviation	1,510,033	1,596,840	1,705,469	1,829,748	1,998,140	1,601,469
Minimum	2,587,213	2,293,224	2,084,841	2,086,531	1,930,725	2,250,401
Maximum	72,366,202	72,771,841	73,474,899	75,564,417	81,277,681	67,405,104
1% ile	2,995,943	2,861,958	2,695,429	2,510,514	2,408,319	2,708,420
5% ile	3,247,847	3,087,062	2,956,837	2,867,115	2,762,663	3,014,557
10% ile	3,384,401	3,241,518	3,143,080	3,033,779	2,987,948	3,194,731
20% ile	3,588,011	3,500,438	3,424,399	3,358,196	3,277,806	3,443,479
30% ile	3,782,986	3,681,105	3,615,534	3,574,383	3,522,031	3,653,888
40% ile	3,942,032	3,897,816	3,820,380	3,790,977	3,745,674	3,849,489
50% ile	4,113,146	4,078,681	4,071,349	4,027,615	3,973,908	4,043,971
60% ile	4,278,521	4,279,869	4,292,852	4,267,561	4,232,721	4,271,588
70% ile	4,493,139	4,518,971	4,547,255	4,558,175	4,560,471	4,554,548
80% ile	4,786,940	4,876,233	4,931,662	5,031,358	5,111,862	4,957,356
90% ile	5,378,096	5,475,577	5,604,519	5,679,109	5,842,125	5,691,814
95% ile	6,008,476	6,230,885	6,371,310	6,436,050	6,836,095	6,471,699
99% ile	8,286,504	8,687,785	9,310,024	10,075,891	10,322,456	9,116,338
Estimated 75th %ile	**4,640,039**	**4,697,602**	**4,739,459**	**4,794,767**	**4,836,166**	**4,755,952**

Figure 3: Graph of case study results – adding bootstrapped correlation to aggregated reserve indication at different levels of correlation between underlying lines of business

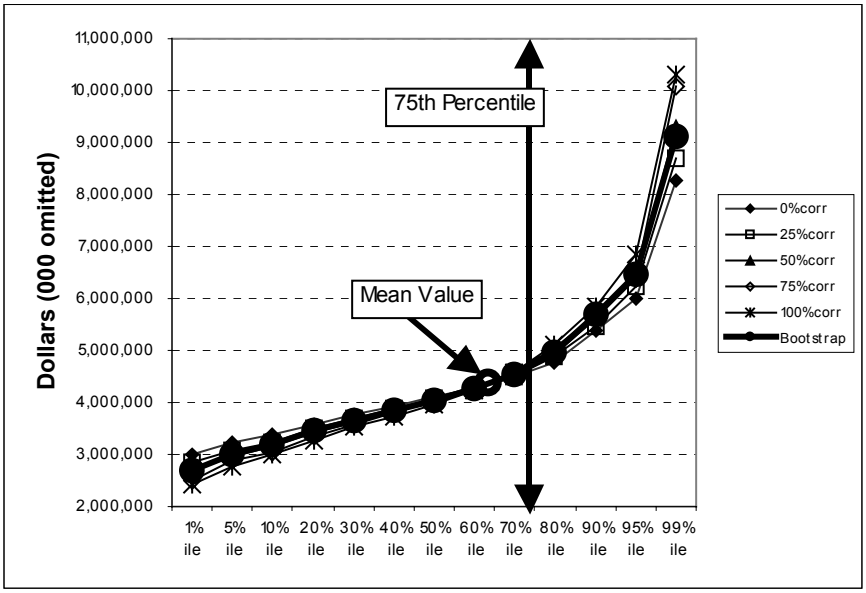
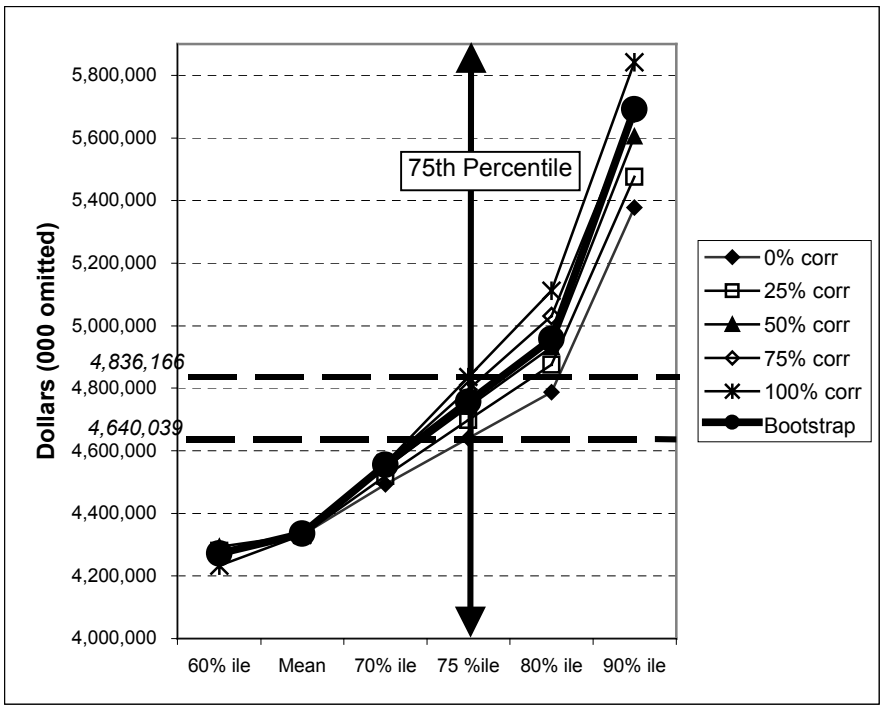


Figure 4: Graph of case study results – adding bootstrapped correlation to aggregated reserve indications at different levels of correlation between underlying lines of business – blowing up area around 75th percentile



The results shown in the preceding figures and tables provide us with the following information:

1. If we wanted to hold reserves at the 75th percentile, the smallest reserve that ought to be held is $4.640 billion and the largest ought to be $4.836 billion.
2. The maximum impact on the 75th percentile of indicated reserves due to correlation is 4.5% of the mean indication ($196 million / $4.331 billion).
3. There does appear to be correlation between at least two of the lines. The observed level of correlation is similar to what would be displayed, were there to be a 50% correlation between each of the lines. It could be that two of the lines exhibit a stronger than 50% correlation with each other and a weaker than 50% correlation with the third line so that the overall results produce values similar to what would exist at the 50% correlation level.
4. The proper reserve to book, assuming the 50% correlation is correct, is $4.75 billion.

Some level of correlation between at least two of the lines is indicated by the bootstrapped results. This is valuable information to know, even beyond the range of reserves indicated by the bootstrap methodology. With this information, company management can assess prospective underwriting strategies that recognize the interrelated nature of these lines of business, such as how much additional capital might be required to protect against adverse deviation. If the lines were uncorrelated, future adverse deviation in one line would not necessarily be reflected in the other lines. With the information at hand, it would be inappropriate to assume that adverse deviation in one line will not be mirrored by adverse deviation in one or both of the other lines. Continuing with this thought, the bootstrapped results would have been valuable even if they had shown there to be little or no correlation between the lines – because then company management could comfortably assume independence between the lines of business and make their strategic decisions accordingly.

V. Summary and conclusions

Let us move beyond the numbers of the case study to summarize what we feel to be the important general conclusions that can be drawn. To begin, calculating an aggregate reserve distribution for several lines of business requires not only a model for the distribution of reserves for each individual line of business, but also an understanding of the dependency of the reserve amounts between each of the lines of business. To get a feel for the impact of these dependencies on the aggregate distribution, we have proposed two

different methods. One can use a rank correlation approach with correlation parameters estimated externally. However, this approach requires either calculating correlations using a method such as has been proposed by Brehm (2002) or by judgmentally developing a correlation matrix. Alternatively, one can use a bootstrap method that relies on the existing dependencies in the historic data triangles. This requires no external calibration, but may be less transparent in providing an understanding of the data. It also limits the calculations to reflecting only those relationships that have existed in the past in the projection of reserve indications.

Furthermore, it would appear that the correlation issue is not important for lines of business with non-volatile reserve ranges. However for volatile reserves, the impact of correlations between could be significant, particularly as one moves towards more extreme ends of the reserve range. If so, either correlation approach can provide actuaries with a way of quantifying the effect of correlations on the aggregate reserve range. Overall, the use of stochastic techniques adds value as such techniques can not only assess the volatility of reserves, but can also identify the significance of correlations between lines of business in a more rigorous manner than is possible with traditional techniques.

To conclude, we believe that stochastic quantification of reserve ranges, with or without an analysis of correlations between lines of business, is a valuable extension of current actuarial practice. Regulations such as those recently promulgated by APRA will accelerate the general usage of stochastic techniques in reserve analysis. An accompanying benefit to the use of stochastic reserving techniques is the ability to quantify the effects of correlations between lines of business on overall reserve ranges. This will help actuaries and company management to better understand how variable reserve development might be, both by line and in the aggregate, allowing companies to make better-informed decisions on the booking of reserves and the amount of capital that must be deployed to protect the company against adverse reserve development.

Bibliography

Brehm, P. (2002). Correlation and the aggregation of unpaid loss distributions. To be published as part of the Casualty Actuarial Society 2002 Fall Forum on Reserving Call Papers.

Christofides, S. (1990). Regression models based on log-incremental payments. *Claims Reserving Manual***, 2**, Institute of Actuaries, London.

Efron, B. & Tibshirani, R.J. (1993), *An introduction to the bootstrap*, Chapman and Hall.

England, P.D. & Verrall, R.J. (1999). Analytic and bootstrap estimates of prediction errors in claims reserving. *Insurance: Mathematics and Economics,* **25***,* pp. 281-293.

England, P.D. (2001). Addendum to 'Analytic and bootstrap estimates of prediction errors in claims reserving'. Actuarial Research Paper # 138, Department of Actuarial Science and Statistics, City University, London EC1V 0HB.

England, P.D. & Verrall, R.J. (2002). Stochastic claims reserving in general insurance. Paper presented to the Institute of Actuaries, 28 January 2002.

Feldblum, S., Hodes, D.M., & Blumsohn, G. (1999). Workers' compensation reserve uncertainty. *Proceedings of the Casualty Actuarial Society*, Volume LXXXVI, pp. 263-392.

Mack, T. (1993). Distribution-free calculation of the standard error of chain-ladder reserve estimates. *ASTIN Bulletin*, **23**, pp. 213-225.

Mack, T. (1994). Which stochastic model is underlying the chain-ladder model? *Insurance: Mathematics and Economics***, 15**, pp. 133-138.

Pinheiro, P.J.R., Silva, J.M.A., & Centeno, M. (2000). Bootstrap Methodology in Claim Reserving. Working Paper # 11/00, Centre for Applied Maths to Forecasting and Economic Decision.

Renshaw, A.E. & Verrall, R.J. (1998). A stochastic model underlying the chain-ladder technique. *B.A.J.***, 4,** pp. 903-923.

Verrall, R.J. (1991). On the estimation of reserves from loglinear models. *Insurance: Mathematics and Economics*, **10**, pp. 75-80.

Appendix I: Data sets

The data used in this case study is fictional. It describes three lines of business, two long tail and one short tail. All three produce approximately the same mean reserve indication, but with varying degrees of volatility around their respective means. The data triangles are shown in Table 11 to Table 13 below. The data is all in the format of incremental paid losses, with all dollar amounts in thousands

.

Table 11 – Line 1 (derived from Commercial Automobile business)

	1	2	3	4	5	6	7	8	9	10
1	20,513	78,579	65,886	57,537	59,293	11,338	10,815	7,811	1,117	11,792
2	13,847	39,035	39,375	29,884	32,754	10,298	6,276	6,924	3,835	0
3	15,785	49,135	42,672	27,920	36,399	27,828	9,596	6,781	0	0
4	20,784	62,266	47,120	59,331	41,672	20,726	16,790	0	0	0
5	108,531	115,103	187,886	90,515	149,616	86,813	0	0	0	0
6	26,097	59,195	1,786	19,780	22,835	0	0	0	0	0
7	64,819	142,577	100,694	34,304	0	0	0	0	0	0
8	44,065	53,039	8,975	0	0	0	0	0	0	0
9	20,022	39,276	0	0	0	0	0	0	0	0
10	37,163	0	0	0	0	0	0	0	0	0

When calculating ultimate indications from this data set, a tail extrapolation allowing for development up to thirty years was included in the calculations.

Table 12 – Line 2 (derived from Homeowners business)

	1	2	3	4	5	6	7	8	9	10
1	761,590	327,920	53,290	16,280	8,400	11,900	9,070	10,140	2,010	80
2	784,590	309,150	64,120	34,990	26,540	30,320	5,640	320	290	0
3	1,077,950	331,980	53,160	44,020	23,170	15,420	8,990	5,780	0	0
4	1,065,310	370,910	52,660	47,320	27,000	12,700	-800	0	0	0
5	1,055,040	372,020	62,250	51,310	18,710	16,970	0	0	0	0
6	1,654,920	413,100	59,920	56,950	38,050	0	0	0	0	0
7	1,326,870	440,340	129,070	58,860	0	0	0	0	0	0
8	1,875,230	465,410	96,290	0	0	0	0	0	0	0
9	1,572,510	419,950	0	0	0	0	0	0	0	0
10	1,902,050	0	0	0	0	0	0	0	0	0

When calculating ultimate indications from this data set, no tail extrapolation was used. Development was assumed to end at ten years.

Table 13 – Line 3 (derived from Workers' Compensation business)

	1	2	3	4	5	6	7	8	9	10
1	36,212	115,053	140,789	115,705	111,334	26,366	20,877	19,788	6,117	16,618
2	40,885	139,674	129,071	109,117	89,906	43,988	20,551	21,526	18,368	0
3	40,322	113,355	100,782	61,491	64,420	40,803	20,580	25,214	0	0
4	38,013	69,213	56,892	75,435	49,984	29,359	25,466	0	0	0
5	37,810	60,405	85,602	33,211	53,347	35,643	0	0	0	0
6	37,159	67,486	34,465	33,121	41,478	0	0	0	0	0
7	34,415	68,634	34,427	18,653	0	0	0	0	0	0
8	37,786	40,462	24,049	0	0	0	0	0	0	0
9	35,380	73,641	0	0	0	0	0	0	0	0
10	39,866	0	0	0	0	0	0	0	0	0

When calculating ultimate indications from this data set, a tail extrapolation allowing for development up to thirty years was included in the calculations.

Appendix II: A step-by-step walkthrough of the bootstrap process used for reserve simulation

Bootstrapping is a technique broadly accepted within the statistical community. It uses the noise within the historical data to make implications about both the noise in the future and about the parameter uncertainty. Since it uses the historical noise, it is not restricted to normal error structures, but rather uses the error structure implicit within the historical data. The method used is based upon the approach outlined by England and Verrall (1999) and expanded upon by England (2001). We encourage readers who want further explanation of the theory or other examples of the methodology to read both of these papers.

The theoretical model to which this bootstrapping technique is compared is a model of incremental claims known as an "over-dispersed" Poisson distribution. This model is described by Renshaw and Verrall (1998). Using the notation from England (1999), incremental claims for origin year *i* in development year *j* are denoted C_{ij}, we have:

$$E[C_{ij}] = m_{ij} \text{ and } Var[C_{ij}] = \phi E[C_{ij}] = \phi\, m_{ij} \qquad \text{(App2.1)}$$

$$\log(m_{ij}) = \eta_{ij} \qquad \text{(App2.2)}$$

$$\eta_{ij} = c + \alpha_i + \beta_j \qquad \text{(App2.3)}$$

These equations defined a generalized linear model in which the calculated value is modeled with a logarithmic link function and the variance is proportional to the mean. The proportionality of the variance to the mean instead of the equality of the variance to the mean is the reason the model is described as an "over-dispersed" Poisson. The parameter ϕ is an unknown scale parameter that is estimated as part of the fitting procedure. England (2001) notes that "with certain positivity constraints, predicted values and reserve estimates from this model are exactly the same as those from the chain ladder model."[7]

The steps undertaken to calculate runoff using the bootstrapping method are:

1. Begin with a triangle of cumulative historical payments. We will use the data from Table 12 – Line 2 (derived from Homeowners business). This is shown in Triangle 1.

[7] England (2001), p. 3.

Triangle 1: Cumulative historical data

	1	2	3	4	5	6	7	8	9	10
1	761,590	1,089,510	1,142,800	1,159,080	1,167,480	1,179,380	1,188,450	1,198,590	1,200,600	1,200,680
2	784,590	1,093,740	1,157,860	1,192,850	1,219,390	1,249,710	1,255,350	1,255,670	1,255,960	
3	1,077,950	1,409,930	1,463,090	1,507,110	1,530,280	1,545,700	1,554,690	1,560,470		
4	1,065,310	1,436,220	1,488,880	1,536,200	1,563,200	1,575,900	1,575,100			
5	1,055,040	1,427,060	1,489,310	1,540,620	1,559,330	1,576,300				
6	1,654,920	2,068,020	2,127,940	2,184,890	2,222,940					
7	1,326,870	1,767,210	1,896,280	1,955,140						
8	1,875,230	2,340,640	2,436,930							
9	1,572,510	1,992,460								
10	1,902,050									

2. Calculate factors based upon historical payments. The factors calculated are based on the cumulative chain ladder method. The factors are weighted averages.

Development Factors

	1	2	3	4	5	6	7	8	9	10
Ave	n/a	1.3088	1.0452	1.0288	1.0156	1.0124	1.0041	1.0041	1.0009	1.0001

3. Using the cumulative factors calculated in step 2, refit the original payments.

Most recent payment period equals most recent payment period cumulated payments in the actual data.

Fitted payments (accident year r, calendar year c) all other payment periods =

$$\frac{FittedPayment(r, c+1)}{ChainLadderFactor(r, c+1)} \qquad \text{(App2.4)}$$

The results of the refitting are shown in Triangle 2.

Triangle 2: Cumulative fitted values

	1	2	3	4	5	6	7	8	9	10
1	822,235	1,076,160	1,124,783	1,157,142	1,175,141	1,189,716	1,194,624	1,199,476	1,200,600	1,200,680
2	860,149	1,125,782	1,176,647	1,210,498	1,229,327	1,244,574	1,249,708	1,254,784	1,255,960	
3	1,069,695	1,400,040	1,463,298	1,505,395	1,528,811	1,547,772	1,554,158	1,560,470		
4	1,084,109	1,418,906	1,483,016	1,525,680	1,549,412	1,568,628	1,575,100			
5	1,089,411	1,425,845	1,490,269	1,533,142	1,556,989	1,576,300				
6	1,555,371	2,035,703	2,127,682	2,188,893	2,222,940					
7	1,389,272	1,818,310	1,900,466	1,955,140						
8	1,781,437	2,331,583	2,436,930							
9	1,522,331	1,992,460								
10	1,902,050									

For example, the derivation of the row 8, column 2 value of 2,331,583 = the row 8, column 3 value of 2,436,930 from Triangle 2 divided by the column 3 average of 1.0452. The derivation of the row 8, column 1 value of 1,781,437 equals the row 8, column 2 value of 2,331,583 from Triangle 2 divided by the column 2 average of 1.3088.

4. Calculate unscaled Pearson residuals. This is the residual definition chosen by England and Verrall (1999) as being suitable for a generalized linear model of the type described by formulas App2.1 through App2.3. The formula for the Pearson residual is given by formula App2.5. The calculated unscaled residuals are shown in Triangle 3.

$$Pearson\ Residual(r,c) = \frac{Actual\ Payment(r,c) - Fitted\ Payment(r,c)}{\sqrt{Fitted\ Payment(r,c)}} \qquad \text{(App2.5)}$$

The values are unscaled in the sense that they do not include the scale parameter ϕ. The scale parameter is not needed when performing the bootstrap calculations, but it will be needed to incorporate an estimate of process error in the final results. The scale parameter will be incorporated into the calculations beginning with Step 11.

Triangle 3: Unscaled Pearson Residuals

	1	2	3	4	5	6	7	8	9	10
1	-66.88	146.84	21.16	-89.38	-71.55	-22.16	59.40	75.92	26.42	0.00
2	-81.47	84.43	58.77	6.19	56.20	122.07	7.05	-66.75	-25.83	
3	7.98	2.84	-40.15	9.37	-1.61	-25.72	32.59	-6.70		
4	-18.06	62.41	-45.22	22.54	21.22	-47.01	-90.39			
5	-32.93	61.35	-8.56	40.75	-33.27	-16.84				
6	79.82	-97.01	-105.71	-17.22	21.69					
7	-52.94	17.25	163.68	17.90						
8	70.27	-114.24	-27.90							
9	40.67	-73.18								

10	0.00									

For example, the derivation of the row 8, column 1 value is:

$$Pearson\ Residual(8,1) = 70.27 = \frac{1,875,230 - 1,781,437}{\sqrt{1,781,437}}$$

5. One adjustment must be made to the unscaled Pearson residuals before they can be used in the bootstrap algorithm. This is to adjust the residuals to account for the number of degrees of freedom in the original data triangle. This step is done so as to allow the estimation variances derived from the bootstrap model to be compared to the estimation variances that can be obtained from the over-dispersed Poisson generalized linear model. The degree of freedom adjustment is accomplished by multiplying each residual by an adjustment factor equal to:

$$\sqrt{\frac{n}{n-p}} \qquad \text{(App2.6)}$$

where n = number of data points (55 in a 10 x 10 triangle) and
p = number of parameters being estimated = (2 * number of accident years)- 1

The degrees of freedom adjustment for this data triangle = $\sqrt{\frac{55}{55-19}} = 1.236.$ The adjusted residuals are shown in Triangle 4.

Triangle 4: Unscaled Pearson Residuals, adjusted for degrees of freedom in original data triangle

	1	2	3	4	5	6	7	8	9	10
1	-82.67	181.50	26.16	-110.48	-88.44	-27.38	73.42	93.83	32.66	0.00
2	-100.70	104.36	72.64	7.65	69.46	150.89	8.72	-82.51	-31.93	
3	9.87	3.52	-49.62	11.58	-1.99	-31.79	40.29	-8.28		
4	-22.32	77.14	-55.89	27.86	26.23	-58.11	-111.73			
5	-40.70	75.83	-10.58	50.36	-41.12	-20.82				
6	98.66	-119.91	-130.65	-21.29	26.81					
7	-65.44	21.33	202.31	22.13						
8	86.86	-141.21	-34.49							
9	50.27	-90.46								
10	0.00									

6. Randomly select from the adjusted Pearson residuals, excluding the cells in the top right and bottom left, as these will always be 0. An example of one possible random selection of residuals is shown in Triangle 5.

Triangle 5: One possible random selection of residuals

	1	2	3	4	5	6	7	8	9	10
1	50.27	-55.89	-119.91	-31.93	-1.99	-31.93	86.86	-49.62	-22.32	69.46
2	-58.11	9.87	-119.91	-141.21	-40.70	8.72	-82.51	50.27	77.14	
3	104.36	3.52	-31.93	-40.70	-34.49	21.33	-10.58	75.83		
4	-31.93	21.33	75.83	69.46	21.33	-65.44	-130.65			
5	93.83	3.52	181.50	-27.38	-10.58	-41.12				
6	-111.73	7.65	8.72	-31.93	-31.93					
7	32.66	-41.12	-119.91	150.89						
8	22.13	-27.38	69.46							
9	-100.70	75.83								
10	73.42									

7. Calculate a "false history" based on the randomly selected residuals from step 6.

$$\begin{aligned} False\ History(r,c) &= Random\ Residual(r,c) \\ &\quad * \sqrt{Fitted\ Payment(r,c)} \\ &\quad + Fitted\ Payment(r,c) \end{aligned} \qquad \text{(App2.7)}$$

Triangle 6: False history based on random residuals in Triangle 5

	1	2	3	4	5	6	7	8	9	10
1	867,818	225,759	22,183	26,615	17,732	10,720	10,994	1,395	376	701
2	806,259	270,717	23,823	7,870	13,244	16,323	(778)	8,657	3,821	
3	1,177,634	332,366	55,226	33,746	18,138	21,898	5,540	12,337		
4	1,050,861	347,137	83,310	57,012	27,017	10,145	(4,039)			
5	1,187,350	338,473	110,492	37,203	22,213	13,596				
6	1,416,033	485,637	94,622	53,311	28,155					
7	1,427,768	402,103	47,787	89,955						
8	1,810,969	529,835	127,892							
9	1,398,085	522,125								
10	2,003,312									

An example of a false history is shown in Triangle 6, using the residuals shown in Triangle 5. For example, the derivation of the row 8, column 1 value is:

$$False\ History(8,1) = 1{,}906{,}677 = 93.83 * \sqrt{1{,}781{,}437} + 1{,}781{,}437$$

8. Recalculate the weighed average cumulative chain ladder factors using cumulated false history from Triangle 6.

Development Factors from false history in Triangle 6

	1	2	3	4	5	6	7	8	9	10
Ave	n/a	1.310	1.045	1.028	1.014	1.010	1.002	1.006	1.002	1.001

9. Use the projection ratios from Step 8 to square the triangle from Step 7 using the traditional cumulative chain ladder method, as is shown in Triangle 7.

Triangle 7: Squaring the false history triangle

	1	2	3	4	5	6	7	8	9	10
1	867,818	1,093,577	1,115,760	1,142,375	1,160,107	1,170,827	1,181,821	1,183,216	1,183,592	1,184,293
2	806,259	1,076,977	1,100,800	1,108,670	1,121,914	1,138,237	1,137,459	1,146,116	1,149,938	1,150,619
3	1,177,634	1,510,000	1,565,226	1,598,972	1,617,110	1,639,008	1,644,548	1,656,885	1,659,871	1,660,854
4	1,050,861	1,397,998	1,481,309	1,538,321	1,565,338	1,575,483	1,571,444	1,580,320	1,583,168	1,584,106
5	1,187,350	1,525,824	1,636,315	1,673,518	1,695,732	1,709,328	1,712,954	1,722,629	1,725,733	1,726,756
6	1,416,033	1,901,670	1,996,292	2,049,602	2,077,757	2,098,848	2,103,301	2,115,181	2,118,992	2,120,248
7	1,427,768	1,829,872	1,877,659	1,967,614	1,994,932	2,015,182	2,019,457	2,030,864	2,034,523	2,035,728
8	1,810,969	2,340,804	2,468,695	2,538,749	2,573,996	2,600,124	2,605,639	2,620,357	2,625,079	2,626,634
9	1,398,085	1,920,209	2,005,844	2,062,763	2,091,401	2,112,631	2,117,112	2,129,071	2,132,907	2,134,171
10	2,003,312	2,624,319	2,741,354	2,819,144	2,858,284	2,887,298	2,893,423	2,909,766	2,915,009	2,916,737

To the left of the heavy black line is the false history data from Triangle 6, to the right is the squaring of the false history data using the link ratios from Step 8.

At this point, the bootstrapping methodology has quantified a measure of the estimation error, but not the process variance. In order to obtain the full prediction error, a measure of process variance must be included in the simulation process. To incorporate process variance in the calculations, England proposes the simulation of incremental payments from a series of Gamma distributions. Each projected incremental payment is assumed to have its own Gamma distribution with mean equal to the incremental projected payments that can be derived from Step 9. The variance is equal to the incremental projected payment multiplied by the scale parameter ϕ that was previously mentioned in Step 4. As a practical measure we have extended this method to allow negative incrementals by modeling the absolute incremental projected payment with the Gamma, and then applying the appropriate sign.

10. Calculate incremental projected payments from the squared triangle. The absolute values of these incremental projected payment amounts will be used as the mean values in each Gamma distribution.

Triangle 8: Calculating incremental projected payments from the squared triangle

	1	2	3	4	5	6	7	8	9	10
1										
2										681
3									2,986	983
4								8,876	2,848	938
5							3,626	9,676	3,104	1,022
6						21,091	4,452	11,881	3,811	1,255
7					27,318	20,250	4,275	11,407	3,659	1,205
8				70,053	35,247	26,128	5,515	14,718	4,722	1,555
9			85,634	56,919	28,639	21,230	4,481	11,959	3,836	1,264
10		621,007	117,035	77,790	39,140	29,014	6,125	16,344	5,243	1,727

11. Calculate the scale parameter ϕ. The scale parameter is estimated as the Pearson chi-squared statistic divided by the degrees of freedom. The Pearson chi-squared statistic is equal to the sum of the squares of the unscaled Pearson residuals that were calculated in Step 4. The degrees of freedom equal n – p, where n and p were calculated in Step 5. The scale parameter is the same for all projected incremental payment periods.

    For the example shown here, the sum of the squares of the unscaled Pearson residuals from Triangle 3 equals 203,397, and the degrees of freedom equals (55 – 19), or 36. The scale parameter ϕ = 5,650.

12. For each incremental future payment, draw a random sample from a Gamma distribution whose mean is equal to the absolute value of the incremental payment calculated in Step 10 and whose variance equals the product of ϕ (as calculated in Step 11) and the absolute value of the incremental payment calculated in Step 10. Set the sign of the random sample so as to be the same as the original incremental payment calculated in Step 10.

Triangle 9: One possible example of random draws from Gamma distributions to simulate payments that include process error as well as parameter error

	1	2	3	4	5	6	7	8	9	10
1										
2										145
3									95	22
4								5,961	683	637
5							3,853	3,621	2,558	599
6						11,919	2,636	38,594	2,461	96
7					32,750	6,270	10,371	17,807	3,472	0
8				55,382	34,696	19,754	3,525	12,698	1,055	1,674
9			98,708	77,614	34,395	20,533	4,794	190	1,214	0
10		624,837	136,808	80,000	63,599	30,370	217	20,878	1,773	9,111

In this example, the value for row 9, column 3 was drawn from a Gamma distribution with a mean of 85,634 and a variance equal to 5,650 * 85,634. The value for row 10, column 3 was drawn from a Gamma distribution with a mean of 117,035 and a variance equal to 5,650 * 117,035.

13. Sum the incremental future payments calculated in Step 12 to arrive at the final reserve estimate for this particular simulation. In the example shown in

Triangle 9, this equals 1,478,376.

14. Repeat steps 5 through 12 N times, producing a different simulated reserve indication each time. At the end of the N simulations, examine the resulting distribution of reserves to arrive at the overall reserve range and reserve indications at different percentiles.

Results from 5,000 simulations

Mean	1,425,665
Standard Deviation	136,233
Minimum	970,631
Maximum	2,055,375
1% ile	1,125,108
5% ile	1,206,925
10% ile	1,253,563
20% ile	1,308,459
30% ile	1,353,153
40% ile	1,389,518
50% ile	1,424,384
60% ile	1,457,631
70% ile	1,492,174
80% ile	1,535,347
90% ile	1,596,733
95% ile	1,656,242
99% ile	1,760,099
Indicated reserve, based on squaring the original data triangle	1,416,460