ROC Working Party

Reserving Uncertainty

Paper for GIRO 2008

Working Party Members:

Neil Bruce – Chairman

Cathy Chen
Gary Dunne
Ian Hinder
Tim McMurrough
Glenn Meyers
Andy White
Tom Wright

Many thanks also to those other members of this (and other) profession(s) who have participated in the work of this working party for some of its life, and to those who have provided other useful insights.

Contents

Page

1. Summary and Introduction

1.1. A Word of Thanks

A great many people have contributed a lot of hard work to this working party – beyond the members named, including former working party members, their colleagues and other members of this and other professions. We are very grateful to all who have taken part and have helped contribute to a wide ranging paper.

1.2. Our objectives this year

The objectives we set ourselves this year were:

Objective 1: Statistical Testing When Conditions Met
To continue the consideration of the performance of statistical methods when all their underlying conditions are met. We would like to better understand to what extent and in what circumstances these methods can be reliable indicators of ultimate claims outcomes at the tails of the distribution.

Objective 2: Test More Methods
To expand the review and testing of methods on "real" data to include additional methods, if possible those which operate on transactional data. This would ideally include further quantitative and qualitative review.

Objective 3: Robustness of Methods in Real Life
To expand the work on simulated data to test the response of methods to circumstances when the underlying conditions are not met – which is the case in most real life scenarios.

We have made some progress on the first and second of these objectives. The third has been touched on with the work on assessing the ability of stochastic methods to predict the development of real data triangles, but this has not been as much of a focus as other areas.

The key area we have investigated has been that surrounding the first objective. The difficulties that the more commonly used methods exhibit in estimating the higher percentiles is concerning, given that these methods are used widely in assessing capital requirements of companies and also in providing indicative estimates of reserving uncertainty.

1.3. This report

This report has five main components:

1.3.1. Accuracy of stochastic reserving methods

Last year we presented some results that indicated methods did not predict the true extent of the variability of claims reserves even when the underlying data exactly matched the method's criteria. This became our top priority for the current year's work and we present our results relating to our continuing investigations relating to this topic.

1.3.2. Application of Bayesian techniques to range estimation

As an extension of the work above, we have tested another method that is gaining popularity in the area of reserve ranges, Bayesian modelling. Here we use a set of data from the FSA returns to test the effectiveness of one particular method – the collective risk method – compared to standard maximum likelihood methods.

1.3.3. International approaches

The CAS in the US and IAA in Australia have both been active in the area of stochastic methods recently, and we discuss their findings and approaches, and indicated how they compare to practice in the UK. We also note a number of other jurisdictions and their current status with regard to reserve range methodologies.

1.3.4. Applying Methods to Real Data

The third area of work is in the application of a numbers of methods and models to some real claims development data. This is an extension to the main thrust of the work last year. The latest results extend the results of the over-dispersed Poisson (ODP) method.

We also consider methods based on transactional level data, although the latter does not easily fit within this section as we are unable to get "real" transactional level data to test. However CAS has recently developed a data generator, which we have used to investigate this type of method.

1.3.5. Areas of concern when applying stochastic reserving

We are conscious that the application of stochastic methods is still a relatively new area for most actuaries. This section discusses a number of topics that need to be considered when applying stochastic methods, either particular to individual methods, or in general.

This discussion is very much at a high-level and "entry-level" in terms of technical content. It is designed to raise awareness of the areas where stochastic methods are weak, and where the results of such methods may need to be treated with caution. Our aim is to help prevent over-reliance on methods and their results without full understanding of the implicit assumptions being made.

1.4 Next Year

We believe the combination of our investigations this year and the work last year has given an appropriate level of basic education and guidance for the profession.

Although we believe many of the areas commented on in this paper require further research and investigation, we consider that the best method for this is through individual research as they are becoming increasingly disparate. We are therefore minded to disband this working party after this year as a ROC sponsored group, but invite others to continue this work on a more ad hoc basis through the normal GIRO working party route.

We also note that the work we have carried out recently has been increasingly focussed on the extremes of the distribution. There are many areas of work where the focus is much less extreme, and these could potentially be considered in more detail, particularly around the communication of such results.

We also reserve the right to reinstate this workstream if we believe it is appropriate, particularly if Solvency II demands become sufficiently crystallised to require interpretation and investigation.

1.5 Summary of findings

The extent to which stochastic models are able to correctly predict the underlying distribution of future claims payments is currently creating significant discussion.

We find that the work last year implying that some stochastic reserving methods understate the extremities of the predicted distribution is correct. These methods as usually applied rely on Maximum Likelihood Estimator methods to derive parameter values, which seems to exacerbate the problems. We have found that the use of Bayesian methodologies helps to reduce this effect, although there remains underestimation in the research we and others have carried out.

We find that a hybrid method using the higher of Mack and ODP provides a consistently better result at higher percentiles of the reserve distribution than using either in all cases when the underlying data exactly meets the assumptions relating to the ODP method. Note that this method does not produce significantly better results in all cases.

When investigating the effects of changing the properties of the triangles under Mack and ODP, we discovered some apparently anomalous results. These indicate that shorter tail business can be more understated at higher percentiles than longer tail business (assuming a full run-off triangle is available).

This result combined with the result indicating that applying these methods to classes with fewer expected claims also makes the estimation of higher percentiles worse lead us to infer that it is the volatility in the development patterns that is the key driver to the estimation error.

In effect, shorter tail business may well have greater volatility in the early periods of development than longer tail business when using similar development period intervals. Similarly where fewer claims are expected, the development pattern will be more volatile than where a greater number of claims will give more statistical stability to a development pattern.

Thus we anticipate that accuracy of these methods can be improved by choosing development intervals that are appropriate for the length of tail of the business being modelled to ensure that the development pattern is as stable as possible. Where such stability is not achievable through either development intervals being too long, or low frequency of claims, we expect the methods to perform less well.

We have not had time to determine the precise number of periods within a development curve to maximise the accuracy of these methods and encourage others to investigate this in future.

We therefore also expect that other methods are more suitable to the modelling of low frequency claims and hence methods based on transactional level data or operational time may be more suitable for such situations. We have investigated transactional methods briefly in this paper and note that they require extreme care when parameterising if sensible results are to be obtained.

To expand our review of common methods, we have looked at methods that use transactional level information. We believe that although these methods use more information than the traditional aggregate triangular data based methods, there are limitations to these methods that imply the results are not necessarily "better".

We have also considered changes to the standard ODP Bootstrap method, and show the results of applying these changes to real data. In addition we have put together a summary of what we believe are common areas of concern when looking at estimating reserve uncertainty and our own views on issues to consider.

Finally, we have also looked at other actuarial professions and highlight some of the work being done that may be of interest to members of the UK profession.

To put this paper into context, we note that any estimation of the uncertainty of a homogeneous book of business will only tell part of the story of the uncertainty relating to a wider portfolio.

Correlations between such books will almost certainly be a significant factor in assessing any overall portfolio based uncertainty, although the evidence presented in this paper and its predecessor indicate that the uncertainty within a single portfolio can be at least as great.

Finally, we remind the reader that the challenges of communicating the information contained within these calculations to stakeholders can be daunting, and ultimately key to a successful reserving uncertainty calculation.

2. Accuracy of stochastic reserving methods

2.1. Introduction

In our report to GIRO 2007 (Section 9 and Appendix B) we presented results of numerical simulations in which the performance of some widely used stochastic reserving methods were tested by applying the methods to large numbers of artificially generated run-off triangles. Some of our main findings were that:

- The method of Mack (1993) can significantly understate the chances of extreme outcomes, even when its assumptions are perfectly satisfied (see section B.1.3 of our report last year).
- Of the various published Bootstrap methods based on over-dispersed Poisson assumptions that we tested, the best performing seemed to be the method described by England (2001). However, this method also tended to understate the chance of extreme outcomes (see section B.2.4.3 of last year's report).
- The Mack method seemed to perform better on data that satisfies the over-dispersed Poisson assumptions (and therefore does not satisfy Mack's assumptions) than on data that does satisfy Mack's assumptions.

We have looked into these results in more detail this year, and have found that they are robust, and continue to apply even when the size of the triangles increases and the parameters used to generate the data are varied.

We have also considered (Section 2.3.7) the possibility of developing a scale of correction factors that might be applied to high percentiles given by these commonly used stochastic methods in order to remove most of the error. However, the results we have obtained this year lead us to believe that there is no consistent set of correction factors because the magnitude of the error varies substantially depending on the data.

As one possible way forward, we have considered the possibility of using multiple stochastic methods (eg Mack and ODP-Bootstrap) then forming a predictive distribution by taking each predictive percentile to be the maximum of the percentiles of the different methods. We have tested (Section 2.4) such a hybrid stochastic method on large numbers of artificial triangles and found that it performs better than each of its component methods individually.

2.2. Performance of Mack's method where its assumptions are true

In our report to GIRO 2007 (Section B.1 of Appendix B) we described tests of the performance of Mack's method on artificial run-off triangles constructed so that Mack's assumptions are perfectly satisfied. We found that, in these ideal conditions, the method has a tendency to understate the chance of extreme outcomes. We carried out these tests last year only on triangles with 10 origin years and 10 development years (55 data-points in total). This year we have carried out similar tests on larger artificial triangles.

In Table 2-1 below, the first column shows the same results as in Table B-5 (Section B1.3.4) of last year's paper. These were obtained by applying Mack's method to 10,000 artificial triangles, all with 10 origin years and 10 development years. The figure 8.4% indicates that in 840 of the 10,000 simulations, the true reserve exceeded what was supposed to the 99th percentile of the predictive distribution obtained by Mack's method. Clearly if the method correctly assesses the reliability of reserve estimates, this should occur in only 1% of simulations.

We speculated that the cause of this poor performance of Mack's method might be that the underlying theory relies on asymptotic formulas and that these give poor approximations when applied to finite datasets. If this is the explanation, then we should expect the performance to improve as the number of data-points in the triangle increases. To test this, we have carried out simulations on triangles of increasing size as shown in the Table 2-1 below. Each set of results is based on 10,000 artificial triangles of the specified size. The artificial triangles were generated using Algorithm A described in Section B.1.2.9 of last year's paper. As the triangle size was increased, the development factor parameters were changed so that the mean delay to payment was always about half the number of development periods in the triangle.

Clearly, the total reserve across all accident years increases substantially as the triangle size increases because the number of origin years increases and the length of the development pattern increases in each origin year. However the variance parameters of Algorithm A (denoted alpha-k in Mack's 1993 paper) were all set to the value one, so the amount of random variation of each individual datapoint in the traingle remains the same in this sense as the triangle size increases

These results show that the performance of Mack's method does seem to improve as the triangle size increases, particularly for percentiles away from the extremes. However, even with 100 origin periods and 100 development periods, the method still seems to materially understate the chance of extreme outcomes: what was supposed to be the 99th percentile of the predictive distribution was exceeded in 2.1% of simulations.

Table 2-1 Results for Mack's method (Algorithm A)

Triangle size	10	15	20	25	100
Mean BCL estimate	78.0				6,707,341
Proportion of sims with (BCL estimate > True reserve)	48.1%				48.6%
Mean of (BCL estimate – True reserve)	0.93				-313.5
Mean Mack standard error	29.52				87,826.5
Mean of (BCL – True) / (Mack std error)	-0.51				-0.062
Mean square of the above	4.2				1.208
1%	8.4%	6.1%	5.6%	4.7%	2.1%
5%	16.3%	14.0%	13.0%	12.3%	7.6%
10%	22.5%	21.1%	19.8%	18.7%	12.9%
20%	32.6%	32.0%	30.7%	29.0%	23.1%

30%	41.2%	41.6%	39.9%	39.0%	32.5%
50%	57.8%	58.2%	57.8%	56.7%	51.6%
70%	73.9%	75.2%	74.6%	74.3%	70.8%
80%	81.9%	83.2%	83.1%	83.5%	80.3%
90%	90.0%	91.4%	92.1%	91.9%	89.9%
95%	93.8%	95.3%	95.9%	96.1%	94.9%
99%	97.8%	98.4%	98.5%	98.8%	98.5%

2.3. ODP Bootstrap method where assumptions are true – summary of previous results

Section B.2.4 of our report last year describes the testing of several variants of the Bootstrap method based on ODP assumptions that have appeared in the literature.

Maximum likelihood estimates of reserves based on the ODP assumptions are equal to basic Chain Ladder estimates, and the original aim of the Bootstrap method based on these assumptions was (like Mack's method) to find a predictive probability distribution of reserves centred on Basic Chain Ladder (BCL) estimates. However, the mean of the Bootstrap predictive distribution is not in general equal to the BCL estimate, so the method does not achieve this objective.

Nevertheless, the results we presented last year (Section B.2.4.3) suggest that the method as described by England (2001) performs reasonably well when its assumptions are perfectly satisfied, except at extreme percentiles.

The results obtained last year (from Table B-10 of Section B.2.4.3 of last year's report) are repeated in Table 2-2. The figure 2.6% achieved in the first set of 10,000 simulations means that in 260 of these simulated triangles, the "true" reserve turned out to exceed the 99th percentile of the Bootstrap predictive distribution. Clearly if the Bootstrap predictive distribution were correct, this should occur with a probability of 1%, so the actual number of occurrences in 10,000 simulations should vary according to a Binomial distribution with $p = 1\%$ and $n = 10{,}000$. This implies that the actual number of occurrences in 10,000 simulations has approximately a 90% chance of falling in the range 80 to 120 (that is 0.8% to 1.2% of simulations).

Since the actual proportion was between 2.5% and 2.8% in all four independent sets of 10,000 simulations, we conclude that the probability of an outcome in excess of what is supposed to be the 99th percentile is significantly higher than 1%. In other words, the predictive distribution tends to be too light-tailed, and tends to understate the true chance of extremely high outcomes.

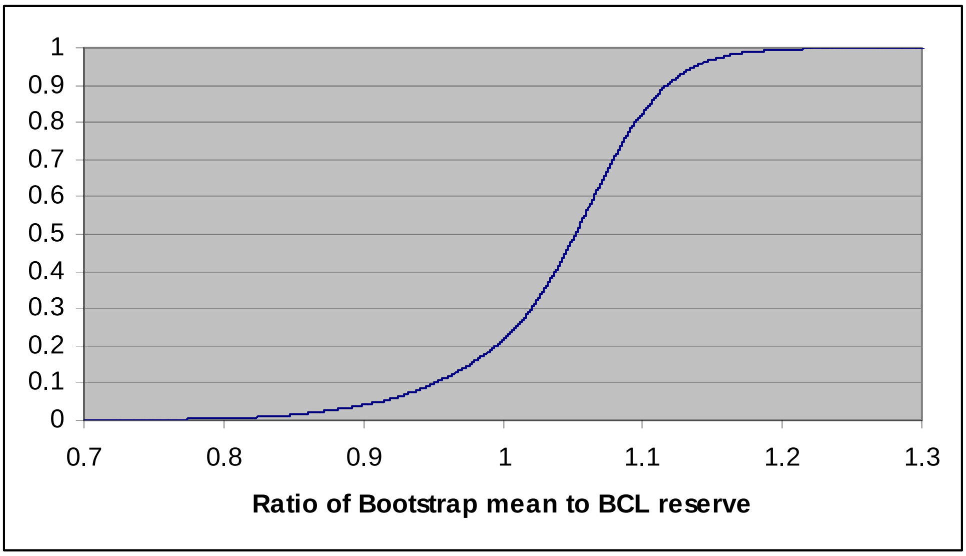
These results also show that although the BCL estimate is usually less than the true outcome (the BCL estimate is greater in 46.7% of simulations), the mean BCL estimate is higher than the mean true reserve. Also, the mean of the Bootstrap predictive distribution tends to be higher than the BCL estimate.

Table 2-2 – Results for 2001 Bootstrap ODP method (Algorithm B)

	Results from different sets of 10,000 simulations			
	Set 1	Set 2	Set 3	All
Mean BCL estimate	3,668	3,654	3,632	3,651
% of triangles with BCL estimate greater than true reserve	46.9%	46.8%	46.4%	46.7%
% of triangles with BS mean greater than true reserve	51.8%			
Mean of (BCL estimate – True reserve)	303.7	292.7	268.4	288.3
Mean Bootstrap mean	3,727	3,719	3,695	3,714
Mean Bootstrap standard error	1,307	1,309	1,299	1,305
Mean of (BCL - True) / (BS std error)	-0.245	-0.237	-0.255	-0.246
Mean square of the above	1.701	1.692	1.670	1.688
1%	2.6%	2.8%	2.5%	2.6%
5%	8.3%	8.0%	8.4%	8.2%
10%	14.3%	13.7%	14.5%	14.2%
20%	24.8%	24.4%	25.3%	24.8%
30%	34.8%	34.2%	35.3%	34.8%
50%	53.9%	53.4%	54.4%	53.9%
70%	71.8%	71.2%	71.9%	71.6%
80%	80.1%	79.8%	80.8%	80.2%
90%	88.6%	88.4%	89.0%	88.7%
95%	93.0%	93.1%	93.2%	93.1%
99%	97.3%	97.5%	97.6%	97.5%

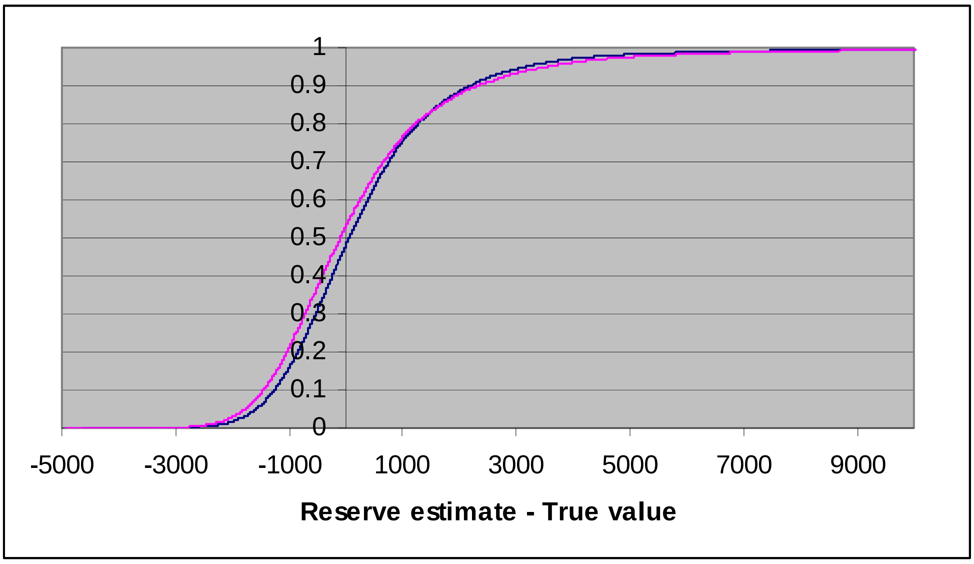
These results are more easily understood by reference to some graphs. The graphs below are all cumulative distribution functions for the first set of 10,000 simulations. That is, the vertical axis shows the proportion of simulations in which the quantity plotted on the horizontal axis takes a value less than the value on the horizontal axis.

The first graph below shows the cumulative distribution function for the ratio of the mean of the Bootstrap distribution to the BCL reserve estimate. Reading off from the value 1.0 on the horizontal axis, we see that in only about 22% of simulations was the mean of the Bootstrap distribution less than the BCL estimate. This confirms that in general, the method does not give a predictive distribution that is centred on the BCL estimate.

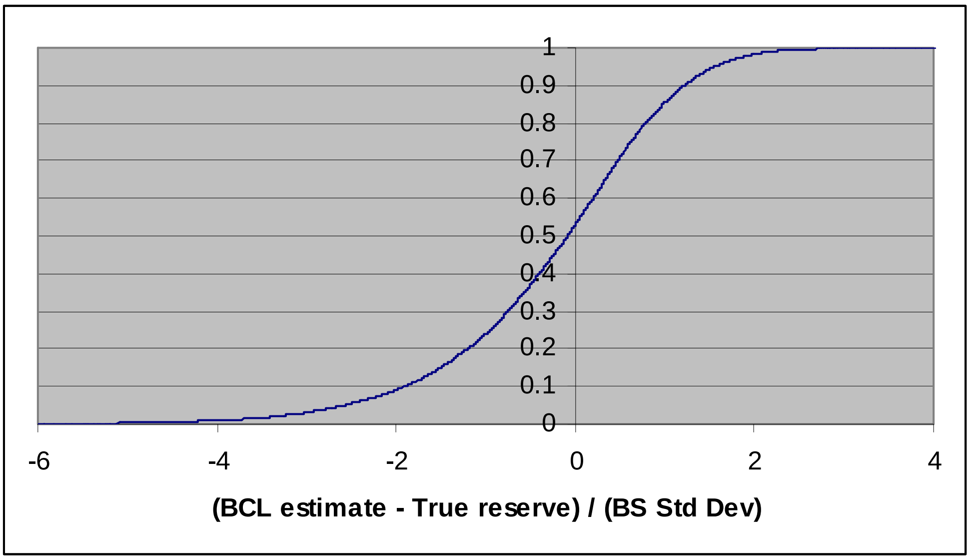


The next graph shows the cumulative distribution functions of the quantities (BCL estimate – True reserve) and (Bootstrap mean – True reserve). The curve that lies further to the right for probabilities (vertical axis) between 0 and 0.8 relates to the Bootstrap mean. From where this curve crosses the vertical axis we see that 48.2% of simulations have Bootstrap mean less than true reserve.

The other curve relates to the BCL reserve, and from where this crosses the vertical axis we see that 53.1% of simulations have BCL reserve less than true reserve. Both curves are positively skewed (that is, the right tail is longer than the left tail) from which it is clear that the mean value of the difference is positive (that is, both the BCL estimate and the Bootstrap mean are positively biased estimates of the true reserve). This is confirmed by the results shown in the Table 2-2.

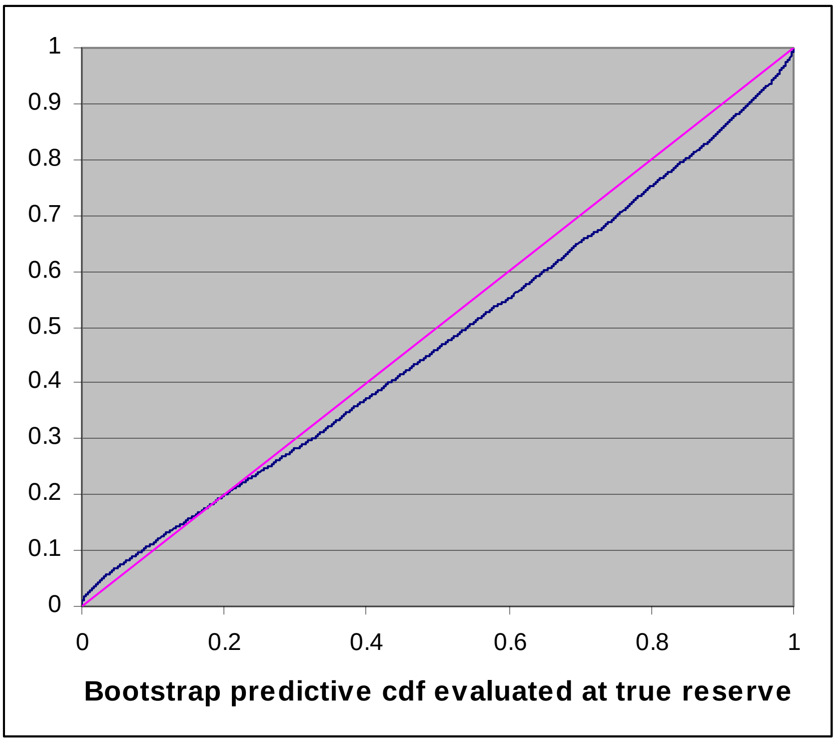


The third graph shows the cumulative distribution function of the quantity (BCL reserve – True reserve) / (Bootstrap Standard Error). Comparing this to the graph above, we see that when the estimation error of the BCL is standardised by dividing by the Bootstrap standard error, the cdf changes from being positively skewed to negatively skewed. This implies that when the BCL underestimates the true reserve (left tail), the Bootstrap standard error also tends to be understated.



Finally, the fourth graph shows the empirical cumulative distribution function of the Bootstrap predictive distribution evaluated at the true reserve. For example, reading off from the value 0.8 on the horizontal axis gives the proportion of simulations (vertical axis) in which the true reserve was smaller than the 80th percentile of the Bootstrap predictive distribution. The corresponding value on the vertical axis is 0.752, which means that in 75.2% of simulations (7,520 out of 10,000 simulated triangles) the true reserve was smaller then the 80th percentile (as shown in Table 2-2) of the Bootstrap predictive distribution. This implies that in the other 24.8% of simulations the true reserve was greater than the 80th percentile of the predictive distribution. In other words, something that should really have a probability of 20% occurred in 24.8% of simulations.

The graph also shows the straight line from (0, 0) to (1, 1): ideally, the cumulative distribution function would not deviate significantly from this straight line.



2.4. ODP Bootstrap method applied where its assumptions are true – new results

The results obtained last year, and summarised in Section 2.3 above, relate to a particular set of parameter values used to generate the artificial triangles. This year, we have carried out further tests with the aim of establishing how generally these results hold. The tests we have carried out are as follows:

(a) Same artificial triangles as last year but an increased number of Bootstrap simulations on each triangle (2,000 instead of 1,000).

(b) More stable triangles than tested last year. That is, triangles with a smaller degree of random deviation from the underlying development pattern.

(c) Triangles of the same size as tested last year (10 origin years and 10 development years) but with a shorter-tailed development pattern than used last year.

(d) Larger triangles than tested last year (15x15 and 20x20).

The results are summarised in the following sub-sections.

2.4.1. Increased number of Bootstrap simulations

Table 2-3 below shows the results obtained by applying the ODP Bootstrap method described by England (2001) to the same artificial triangles as used in the first set of 10,000 simulations carried out last year. The only difference is that the number of Bootstrap simulations carried out on each triangle has been increased from 1,000 to 2,000. The results show that this increase in the number of Bootstrap simulations has no significant effect.

Table 2-3 Results for 2001 Bootstrap ODP method (Algorithm B) – Increasing the number of Bootstrap simulations

Number of Bootstrap simulations for each of 10,000 triangles	1,000	2,000
Mean BCL estimate	3,668	
% of triangles with BCL estimate greater than true reserve	46.9%	
% of triangles with Bootstrap mean greater than true reserve	51.8%	51.9%
Mean of (BCL estimate – True reserve)	303.7	
Mean Bootstrap mean	3,727	3,727
Mean Bootstrap standard error	1,307	1,308
Mean of (BCL - True) / (BS std error)	-0.245	-0.245
Mean square of the above	1.701	1.698
1%	2.6%	2.6%
5%	8.3%	8.4%
10%	14.3%	14.1%
20%	24.8%	24.7%
30%	34.8%	34.9%
50%	53.9%	54.0%
70%	71.8%	71.9%
80%	80.1%	80.1%
90%	88.6%	88.6%
95%	93.0%	93.0%
99%	97.3%	97.4%

2.4.2. More stable triangles

Each of the 10,000 artificial triangles analysed in the preceding results was generated using Algorithm B from Section B.2.2 of last year's report. This algorithm uses a compound Poisson/Log-Normal distribution for each incremental amount Y_{jk} of the triangle. In other words, each Y_{jk} is constructed as the sum of a number (N_{jk} say) of independent Log-Normal amounts, where the number N_{jk} is generated from a Poisson distribution.

Algorithm B for artificial run-off data:

1. The ultimate number of claims in an origin year is generated by random sampling from a Poisson distribution (same parameters for each origin year, but independent sampling).
2. Each claim is assumed to be settled by a single payment, and the development year of the payment determined by independent random sampling from a Multinomial distribution (same parameters for each origin year).
3. The amount of each individual claim payment is determined by independent random sampling from a Log-Normal distribution (same parameters in every cell of the triangle).
4. The amounts of claims settling in the upper left triangle of the run-off array are accumulated to create this run-off triangle, and all claim amounts (regardless of the development year when settled) are accumulated to obtain the 'true' ultimate position for each origin year.

Using the same parameters in all cells (j,k) for the Log-Normal distribution (at Step 3) ensures that the quantity $Var(Y_{jk}) / E(Y_{jk})$ is the same in all cells of the triangle, which is required by the ODP assumption. In last year's report (Section B6.2) it is proved that this ratio (denoted φ) is related to the mean μ and variance σ^2 of the loss distribution used at Step 3 by $\varphi = (\mu^2 + \sigma^2) / \mu$. (Note that μ and σ^2 here are the mean and variance of the Log-Normal distribution used at Step 3, not the mean and variance of the related Normal distribution.)

The parameters used in Algorithm B to generate the artificial triangles analysed in last year (and in the results described in previous sections of this report) are as follows:

1. Poisson mean = 100 claims in each origin year.
2. Given a claim, the probabilities of it being settled in each development year are as given in the middle row of Table 2-5.
3. Log-Normal with mean $\mu = 10$ and coefficient of variation $\sigma / \mu = 2.528$. (Note that these are parameters of the Log-Normal distribution, not parameters of the underlying Normal distribution.)

To investigate the effect of having a more stable run-off triangle, we have increased the Poisson parameter at Step 1 from 100 to 1,000. In every cell (j,k) of the triangle, this has the effect of increasing both $E(Y_{jk})$ and $Var(Y_{jk})$ by a factor of 10. The ODP dispersion parameter (φ) is therefore unchanged, but in each cell of the triangle (each aggregate paid amount Y_{jk}) the coefficient of variation is decreased by the factor 0.316 (reciprocal of the square root of 10) and the skewness coefficient is decreased by the factor 0.1, so the triangle will typically show considerably less variation from the underlying run-off pattern. Results based on 10,000 triangles generated in this way are shown in the final column of the Table 2-4 below.

Table 2-4 – Results for 2001 Bootstrap ODP method (Algorithm B) – Increased number of claims in each origin year

Expected number of claims in each origin year	100	1,000
Mean BCL estimate	3,668	34,046
% of triangles with BCL estimate greater than true reserve	46.9%	48.4%
% of triangles with Bootstrap mean greater than true reserve	51.8%	50.6%
Mean of (BCL estimate – True reserve)	303.7	368.9
Mean Bootstrap mean	3,727	34,331
Mean Bootstrap standard error	1,307	5,154
Mean of (BCL - True) / (BS std error)	-0.245	-0.090
Mean square of the above	1.701	1.052
1%	2.6%	1.1%
5%	8.3%	5.1%
10%	14.3%	10.7%
20%	24.8%	21.6%
30%	34.8%	32.2%
50%	53.9%	52.7%
70%	71.8%	71.8%
80%	80.1%	80.9%
90%	88.6%	89.9%
95%	93.0%	94.4%
99%	97.3%	98.1%

These results show that the mean of the Bootstrap distribution is proportionately much closer to the true reserve than previously, and percentiles of the predictive distribution are much more accurate.

2.4.3. Shorter development pattern

For the following results, we return to an expected number of 100 claims in each origin year, but increase the speed of development. The multinomial probabilities used at Step 2 of Algorithm B are given in the following table:

Table 2-5 – Parameters used at Step 2 of Algorithm B

Dev year	1	2	3	4	5	6	7	8	9	10
Previous probs	.043	.143	.198	.193	.155	.110	.072	.044	.026	.015
New probs	.280	.378	.216	.0864	.0285	.0083	.0022	.0006	.0001	.00003

Table 2-6 – Results for 2001 Bootstrap ODP method (Algorithm B) – Shorter development pattern

Development pattern	Long	Short
Mean BCL estimate	3,668	1,265
% of triangles with BCL estimate greater than true reserve	46.9%	48.3%
% of triangles with Bootstrap mean greater than true reserve	51.8%	49.9%
Mean of (BCL estimate – True reserve)	303.7	22.0
Mean Bootstrap mean	3,727	1,284
Mean Bootstrap standard error	1,307	405.2
Mean of (BCL - True) / (BS std error)	-0.245	-0.236
Mean square of the above	1.701	1.996
1%	2.6%	4.3%
5%	8.3%	11.3%
10%	14.3%	17.8%
20%	24.8%	28.3%
30%	34.8%	37.5%
50%	53.9%	53.8%
70%	71.8%	69.8%
80%	80.1%	78.1%
90%	88.6%	86.7%
95%	93.0%	91.5%
99%	97.3%	96.6%

These results show proportionately less bias in the BCL estimates and the Bootstrap mean, but extreme percentiles are less accurate for the shorter development profile. This may be because the expected number of claims shows more variation across development years, decreasing rapidly from a high number in early development years to a relatively low number in later years. As a consequence, the key assumption of the Bootstrap method (that the distribution of residuals has the same shape in all development years) becomes less reliable: the skewness of the residuals will actually vary substantially across development years, being much higher in the later development years where there are relatively few claim settlements.

2.4.4. Larger triangles

Next we consider results for triangles of increasing size. We have repeated the analysis using 10,000 triangles with 15 origin years, and 10,000 triangles with 20 origin years. In each case, the delay to settlement follows approximately a Gamma distribution, with mean delay being equal to one quarter of the number of development periods in the triangle. The multinomial probabilities (used at Step 2 of Algorithm B) with these properties are shown in Table 2-7.

Table 2-7 – Parameters used at Step 2 of Algorithm B

	Size of triangle		
Dev Year	10	15	20
1	0.27974	0.09052	0.03609
2	0.37824	0.23094	0.12437
3	0.21575	0.24856	0.18083
4	0.08644	0.18790	0.18466
5	0.02853	0.11703	0.15537
6	0.00833	0.06449	0.11566
7	0.00224	0.03266	0.07913
8	0.00056	0.01555	0.05088
9	0.00014	0.00706	0.03121
10	0.00003	0.00309	0.01844
11		0.00131	0.01058
12		0.00054	0.00592
13		0.00022	0.00324
14		0.00009	0.00174
15		0.00003	0.00092
16			0.00048
17			0.00025
18			0.00013
19			0.00006
20			0.00003

Results based on triangles with these parameters are shown in the Table 2-8 below. (Results for triangles with 10 origin years are repeated from the previous sub-section for convenience.) Although the accuracy of extreme percentiles seems to improve as the triangle size increases, the accuracy of less extreme percentiles deteriorates.

Table 2-8 – Results for 2001 Bootstrap ODP method (Algorithm B) – Increasing size of triangle

Number of origin years and development years	10	15	20
Mean BCL estimate	1,265	2,588	3,874
% of triangles with BCL estimate greater than true reserve	48.3%	46.9%	43.3%
% of triangles with Bootstrap mean greater than true reserve	49.9%	49.6%	46.4%
Mean of (BCL estimate – True reserve)	22.0	95.2	123.4
Mean Bootstrap mean	1,284	2,635	3,900
Mean Bootstrap standard error	405.2	798.5	1,156
Mean of (BCL - True) / (BS std error)	-0.236	-0.261	-0.344
Mean square of the above	1.996	1.836	1.871
1%	4.3%	3.1%	2.9%
5%	11.3%	9.8%	10.5%
10%	17.8%	17.2%	17.9%
20%	28.3%	28.9%	30.4%
30%	37.5%	39.0%	41.1%
50%	53.8%	55.3%	59.6%
70%	69.8%	71.0%	74.9%
80%	78.1%	78.9%	81.6%
90%	86.7%	86.5%	88.9%
95%	91.5%	91.3%	92.9%
99%	96.6%	96.4%	97.1%

2.4.5. Conclusions for ODP Bootstrap method

Our main conclusion from the results presented in previous sub-sections is that there appears to be no simple but widely applicable correction that can be made to extreme percentiles obtained from the ODP Bootstrap method when it is applied to triangles that satisfy the ODP assumptions.

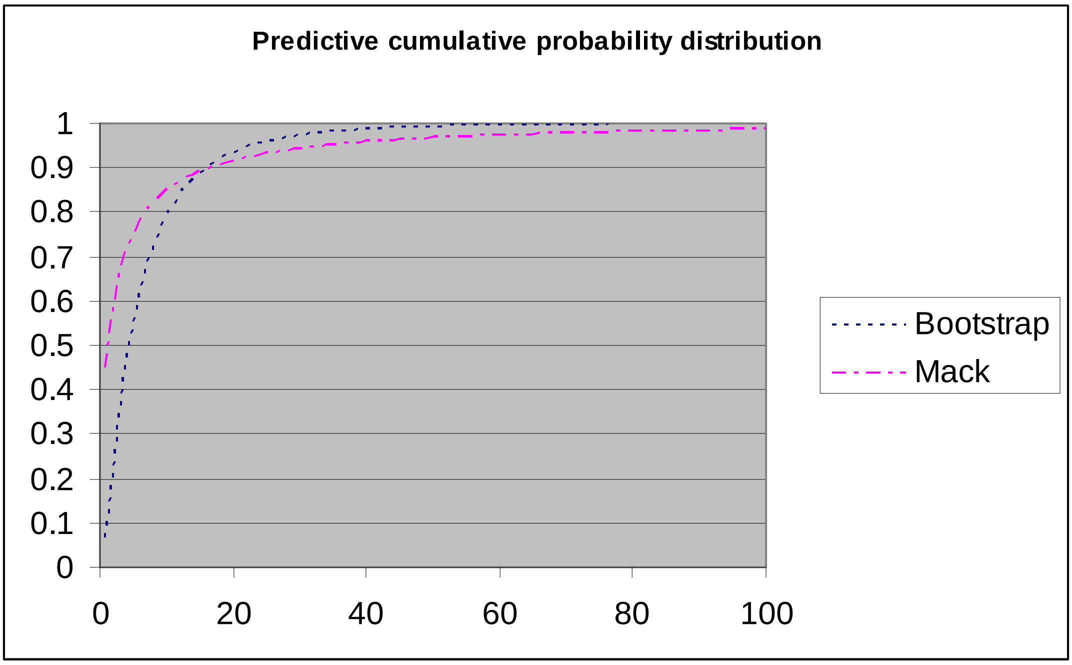
If we had found, under a wide range of parameter values, that what was supposed to be the 99th percentile was usually close to being the true 97th percentile, then we would have a simple correction formula for this percentile. However, our results show that what is supposed to be the 99th percentile can be close to the true 99th percentile in some cases, but can be close to the true 96th percentile in other cases.

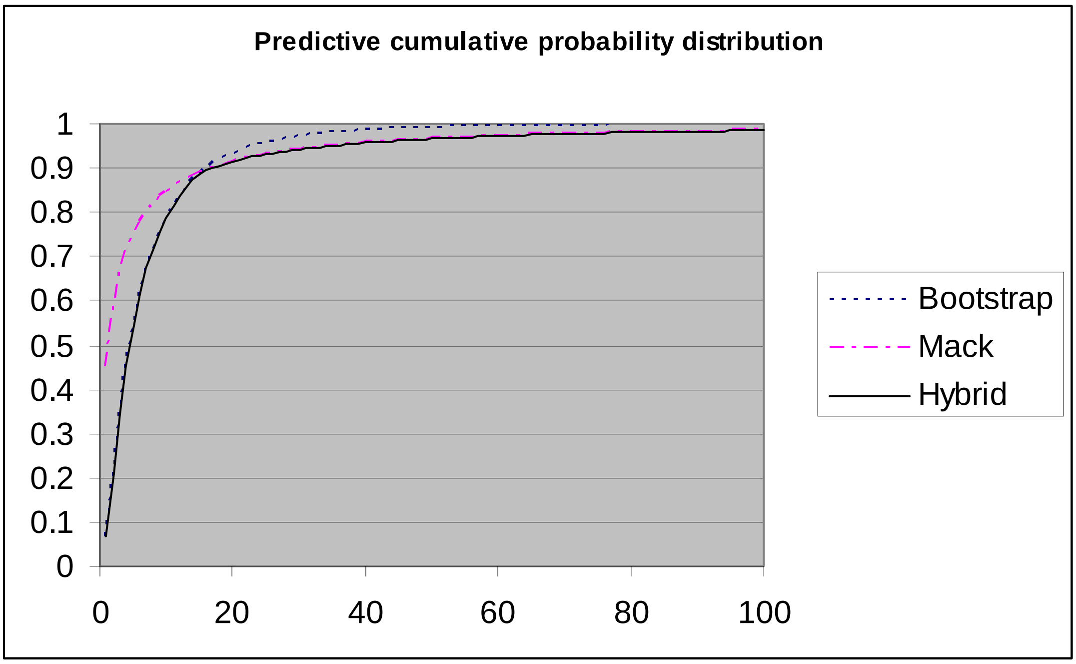
In practice there is no easy way to tell which of these cases applies so we are unable to recommend a simple rule of thumb for correcting extreme percentiles obtained by this method.

2.5. Hybrid Mack/ODP-Bootstrap method

2.5.1. Introduction

Since our results show that both the ODP-Bootstrap method and Mack's method tend to understate the chance of extremely high outcomes, one possible approach is to apply both methods, then define the predictive distribution function as the minimum of the Mack and ODP predictive distribution functions. This is equivalent to defining any required percentile of the predictive distribution to be the maximum of the corresponding percentiles from the Mack and ODP predictive distributions. This is illustrated below.





The example illustrated above is a case in which the Mack method gives higher values for percentiles of extreme adverse events than the Bootstrap method. (For example, the 95th percentile is about 23 according to the Bootstrap, and about 34 according to the Mack method). However in some cases, the Bootstrap might give higher extreme percentiles. It is also possible that the Mack and ODP distribution functions cross more than once in the right tail. Regardless of which method gives the higher percentiles, the hybrid distribution function is defined, at each possible reserve value on the horizontal axis, as the minimum of the two distribution functions

We have tested this hybrid method on triangles that satisfy the ODP assumptions: the same sets of triangles as used in the previous section for testing the ODP-Bootstrap method.

2.5.2. Results for hybrid method

Results obtained by applying this hybrid method to the first set of 10,000 artificial triangles analysed in Table 2-2 are given in the final column of the Table 2-9 below. The other two columns give results obtained from the ODP and Mack methods separately, for comparison.

To explain these results, consider the first row of the table. This shows that, in 2.6% of simulations (260 out of 10,000 triangles) the true reserve exceeded the 99th percentile of the ODP Bootstrap predictive distribution, and in 1.4% (140 triangles), the true reserve exceeded the 99th percentile of the Mack predictive distribution. The final column shows that in 1.3% of simulations (130 triangles) the true reserve exceeded the maximum of the two 99th percentiles (that is, the true reserve exceeded both the 99th percentile of the ODP Bootstrap distribution and the 99th percentile of the Mack predictive distribution).

Table 2-9 – Results of applying hybrid method to data generated using Algorithm B

	Standard algorithm			More stable triangles			Shorter development pattern		
	ODP	Mack	Hybrid	ODP	Mack	Hybrid	ODP	Mack	Hybrid
1%	2.6%	1.4%	1.3%	1.1%	1.1%	0.9%	4.3%	1.6%	1.6%
5%	8.3%	6.8%	6.1%	5.1%	5.6%	4.7%	11.3%	6.9%	6.8%
10%	14.3%	13.1%	11.8%	10.7%	11.5%	10.2%	17.8%	14.1%	14.0%
20%	24.8%	27.2%	23.8%	21.6%	22.9%	21.1%	28.3%	26.9%	26.6%
30%	34.8%	41.2%	34.6%	32.2%	34.0%	31.9%	37.5%	39.2%	37.5%
50%	53.9%	64.7%	53.6%	52.7%	55.0%	52.5%	53.8%	59.8%	53.8%
70%	71.8%	81.4%	71.5%	71.8%	73.5%	71.5%	69.8%	77.4%	69.8%
80%	80.1%	87.8%	79.7%	80.9%	82.2%	80.3%	78.1%	84.2%	78.1%
90%	88.6%	93.4%	88.1%	89.9%	90.5%	89.3%	86.7%	90.9%	86.7%
95%	93.0%	96.0%	92.5%	94.4%	94.7%	93.8%	91.5%	94.4%	91.5%
99%	97.3%	98.3%	96.9%	98.1%	98.1%	97.7%	96.6%	97.5%	96.4%

The results in Table 2-9 show that, in these particular triangles, the hybrid method gives a more accurate assessment of the chance of extreme adverse outcomes than either the ODP method or the Mack method used alone. To investigate whether this is true more generally, we have also applied the hybrid method to all other sets of triangles considered in Section 2.3.

Table 2-9 also gives results for the 10,000 more stable triangles from Section 2.4.2 and for the 10,000 triangles with a shorter development pattern analysed in Section 2.4.3 Table 2-10 gives results for the triangles of increasing size analysed in Section 2.4.4.

Table 2-10 – Increasing size of triangle

	Bootstrap		Mack		Hybrid	
Size	15	20	15	20	15	20
1%	3.1%	2.9%	1.0%	0.9%	1.0%	0.9%
5%	9.8%	10.5%	6.0%	5.1%	5.9%	5.1%
10%	17.2%	17.9%	12.6%	11.5%	12.4%	11.4%
20%	28.9%	30.4%	26.2%	25.8%	25.9%	25.6%
30%	39.0%	41.1%	39.9%	41.6%	38.3%	39.6%
50%	55.3%	59.6%	61.9%	69.3%	55.3%	59.4%
70%	71.0%	74.9%	79.8%	86.2%	71.0%	74.9%
80%	78.9%	81.6%	86.5%	91.8%	78.8%	81.5%
90%	86.5%	88.9%	92.5%	96.2%	86.5%	88.9%
95%	91.3%	92.9%	95.7%	97.9%	91.3%	92.9%
99%	96.4%	97.1%	98.4%	99.2%	96.3%	97.1%

2.5.3. Conclusions

The methods described above all have difficulty in estimating the more extreme percentiles of the distribution of reserves.

The hybrid method, when applied to triangles that satisfy the ODP assumptions, generally seems to give a more reliable assessment of the chances of extreme adverse outcomes than either the Mack method or the ODP-Bootstrap method alone.

Further investigations into the effects have shown that shorter tail business and classes with fewer claims are more difficult to model using these techniques.

We infer that the key determinant for such methods to produce reasonable results is the stability of the development pattern within the dataset. This stability may be improved by increasing the frequency of timesteps within the development triangle for shorter tail business. Similarly for classes where numbers of claims are small, grouping together similar classes of business, or working from larger datasets (e.g. market data) may also reduce the errors involved.

Note that all results in Section 2 relate to uncertainty in the total reserve (the sum of all origin years). We cannot be certain that the same conclusions would hold for origin years considered separately.

3. Application of Bayesian techniques to range estimation

3.1. Introduction

In response to our work last year on the accuracy of the predictions of standard stochastic models, we have investigated an alternative method of estimating the variability of future reserves.

This method allows explicitly for the parameter error within the estimation procedure by using a number of parameter sets when estimating the uncertainty within the reserves. The validity of each of these parameter sets is estimated by calculating the probability of the data coming from the model with that set of parameters, and the overall distribution of reserves being the convolution of the results of each of the parameter sets weighted by these probabilities.

3.2. Methodology

This section fits a model to paid losses in the UK Motor data as held in UK FSA returns. We split this data into training and test data. The test data consists of the last diagonal of the data, and the training data consisted of everything else. Using this model we use the training data to predict the distribution of the sum of losses in the test data. We then calculate the percentile of the observed loss in the test data.

This is done for each of the 34 insurers in the UK Motor data. We will validate the model by testing the hypothesis that the thirty-four percentiles are uniformly distributed.

An important feature of the model is that it is a Bayesian model, with the prior distribution of parameters being derived from an analysis of fifteen of the large insurers.

Currently, there is no single paper that describes the exact methodology used in this analysis. Most of what follows is described in more detail in the following three papers by Glenn Meyers.

1. "Estimating Predictive Distributions for Loss Reserve Models"

   http://www.variancejournal.org/issues/01-02/248.pdf

2. "Thinking Outside the Triangle"

   http://www.actuaries.org/ASTIN/Colloquia/Orlando/Papers/Meyers.pdf

3. "Stochastic Loss Reserving with the Collective Risk Model"

   http://www.casact.org/pubs/forum/08sforum/11Meyers.pdf

These papers will be referred to as [1], [2], and [3] respectively.

The model, described as the "Cape Cod Model" in [3], first describes the expected paid loss in each cell indexed by the accident year (AY) and development year (Lag) by the formula:

$$\mathrm{E}\left[Loss_{AY,Lag}\right] = Premium_{AY} \cdot ELR_{AY} \cdot Dev_{Lag}$$

where $\{ELR_{AY}\}$ and $\{Dev_{Lag}\}$ are parameters to be estimated from the data. Next the model uses the collective risk model to describe how actual outcomes in each (AY, Lag) cell are distributed around the expected loss with the collective risk model. The collective risk model can be described by the following simulation algorithm.

1. Select a random claim count, $N_{AY,Lag}$ from a negative binomial distribution with mean λ and variance $\lambda + c\lambda^2$. Following [1] we set $c = 0.01$.

2. For $i = 1, 2, \ldots, N_{AY,Lag}$ select a random claim amount, $Z_{Lag,i}$.

3. If $N_{AY,Lag} > 0$, set $X_{AY,Lag} = \sum_{i=1}^{N_{AY,Lag}} Z_{Lag,i}$. Otherwise set $X_{AY,Lag} = 0$.

We use the claim severity distributions in [3], in which the expected claim severity increases with settlement lag.

Following [3], Appendix B, we use the overdispersed negative binomial to approximate the likelihood of the data. Note that [3] uses a Poisson distribution in Step 1 of the simulation algorithm above. This makes it necessary to replace the formula

$\kappa_{AY,Lag} = \dfrac{\lambda_{AY,Lag} \cdot \mu_{Lag}^2}{\sigma_{Lag}^2}$ in that appendix with $\kappa_{AY,Lag} = \dfrac{\lambda_{AY,Lag}}{\dfrac{\sigma_{Lag}^2}{\mu_{Lag}^2} + c\lambda}$.

We should stress that the claim severity distributions and the c parameter are fixed, and not estimated, based on the author's experience with US data. We suggest that the model could be improved with parameters based on UK data.

As a first step to fitting this model to the UK motor data we first obtain the maximum likelihood estimates of the parameters $\{ELR_{AY}\}$ and $\{Dev_{Lag}\}$ for each insurer. The best way to visualise the $\{ELR_{AY}\}$ and $\{Dev_{Lag}\}$ parameters is as paths along increasing AY and Lag time lines. Figures 1 and 2 below plot the paths for each of the 34 insurers.

Figure 1

$\{ELR_{AY}\}$ Paths Based on MLE Estimates

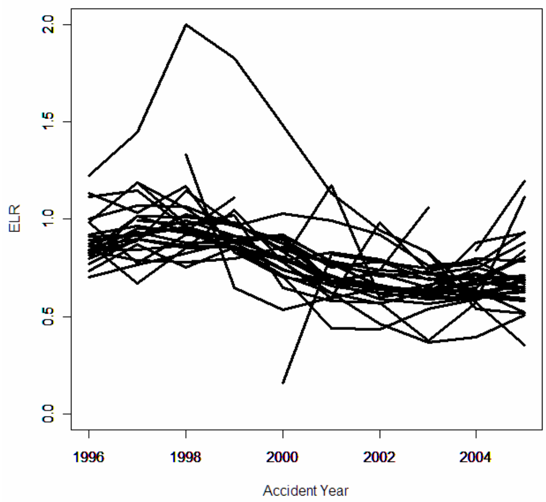
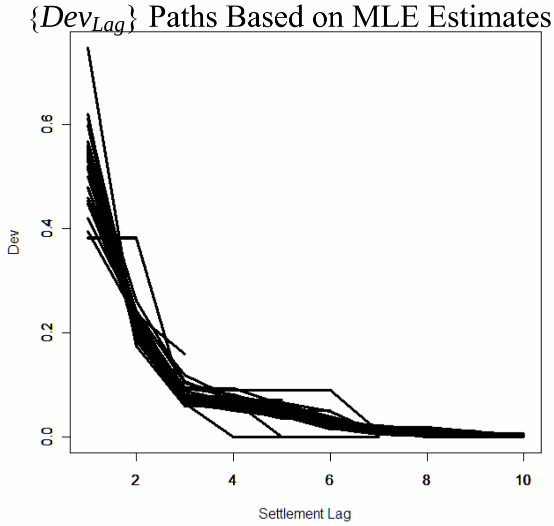


Figure 2

$\{Dev_{Lag}\}$ Paths Based on MLE Estimates

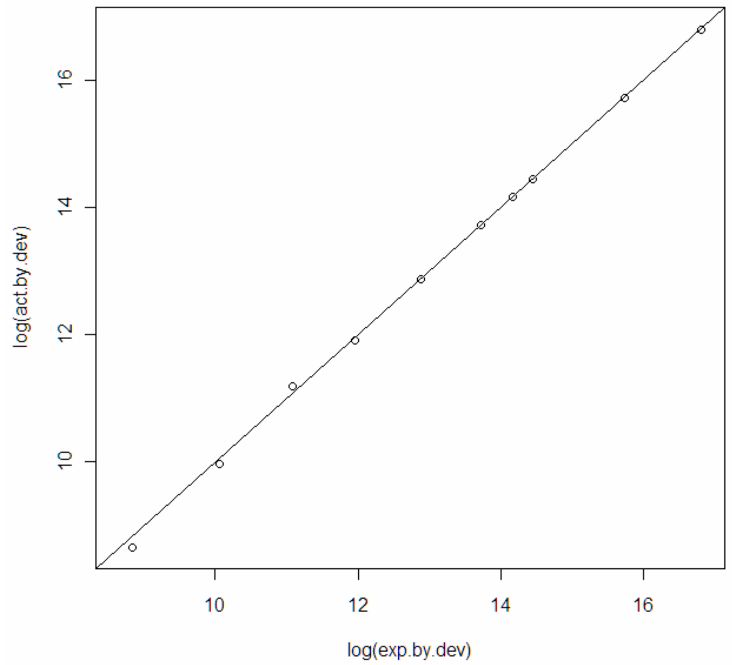


Figures 1 and 2 demonstrate two items of interest.

1. There is a fair amount of apparent random variation in the paths. As shown in Figures 4 and 5 below, restricting the plots to the 15 largest insurers reduces the random variation.
2. Beneath the random variation of the $\{ELR_{AY}\}$ paths there appears to be systematic variation (typically called the underwriting cycle) in the paths.

We now demonstrate that the MLEs of $\{ELR_{AY}\}$ and $\{Dev_{Lag}\}$ do a good job of predicting losses when we combine the results for all 34 insurers. Figure 3 shows a plot of the sum over all insurers of the expected loss against the corresponding sum of observed losses by settlement lag.

Figure 3



The sum of the paid losses in the training data was 31,274,913,000, while the sum of the expected losses determined by the MLE model was 31,670,789,000, 1.3% higher.

While, in the aggregate, the MLEs may do a good job of estimating the expected loss, paper [2] suggests by way of example that: (1) the MLEs do a poor job of estimating the tails of the distribution of outcomes; and (2) a Bayesian approach can do a good job of estimating the tails.

Papers [1] and [3] implement the Bayesian approach in different ways. Paper [1] uses a discrete prior distribution based on the MLEs of 40 large insurers. Paper [3] uses a formulaic prior assuming that each of the parameters has an independent gamma distribution and simulates the posterior distribution of parameters using the Gibbs sampler.

One criticism of the approach in [1] is that the discrete prior is too coarse and may not adequately represent the variability of the possible parameters. A criticism of the approach in [3] is that if one wants to base the prior on the experience of large insurers finding the right formula to describe the prior could be difficult. In addition, running the Gibbs sampler can be time consuming.

This paper uses a hybrid methodology that draws upon the approaches described in each paper. For each of the large insurers, we generate a set of $\{ELR_{AY}\}$ and $\{Dev_{Lag}\}$ parameters using the Gibbs sampler. The prior distribution for each parameter was a gamma distribution centred at the MLE for that parameter that had a coefficient of variation that was equal to one. From these parameter sets, a discrete prior distribution was constructed by taking a random sample of $\{ELR_{AY}\}$ and, independently, a random sample of $\{Dev_{Lag}\}$. In the end, the prior distribution of $\{ELR_{AY}\}$ and $\{Dev_{Lag}\}$ consisted of 50,000 parameter sets. Figures 4 and 5 show a sample of paths selected from the prior distribution.

Figure 4

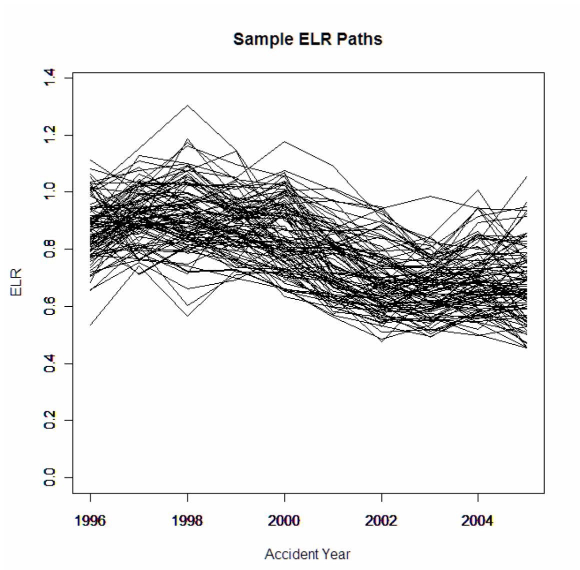
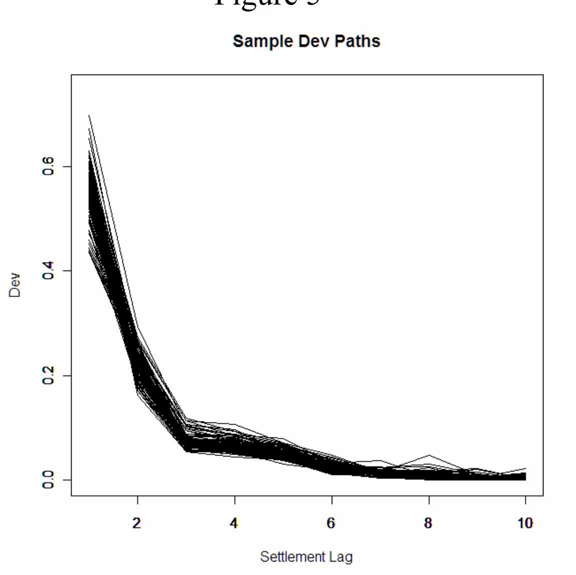


Figure 5



We then used the Bayesian methodology described in [1] and [2] to the training data and calculated posterior probabilities for each parameter set.

For each parameter set in the prior distribution[1] and each (*AY,Lag*) cell in the test data we calculated the expected loss and then used the collective risk model to calculate the distribution around the expected loss. We then calculated the distribution of the sum of the losses over each cell in the test data. The mathematical details for doing this are described in Appendix A of [3]. Since we are using the negative binomial for the claim count distribution we have to substitute $\Phi\left(\vec{\mathbf{q}}_{AY,Lag}\right)=\left(1-c\lambda\left(\Phi\left(\vec{\mathbf{p}}_{Lag}\right)-1\right)\right)^{-1/c}$ for $\Phi\left(\vec{\mathbf{q}}_{AY,Lag}\right)=e^{\left(\Phi\left(\vec{\mathbf{p}}_{Lag}\right)-1\right)}$ in Step 4 of Section

[1] To save computing time we selected those parameter sets that had a posterior probability over 0.00001. Typically this kept anywhere from several hundred to several thousand parameter sets.

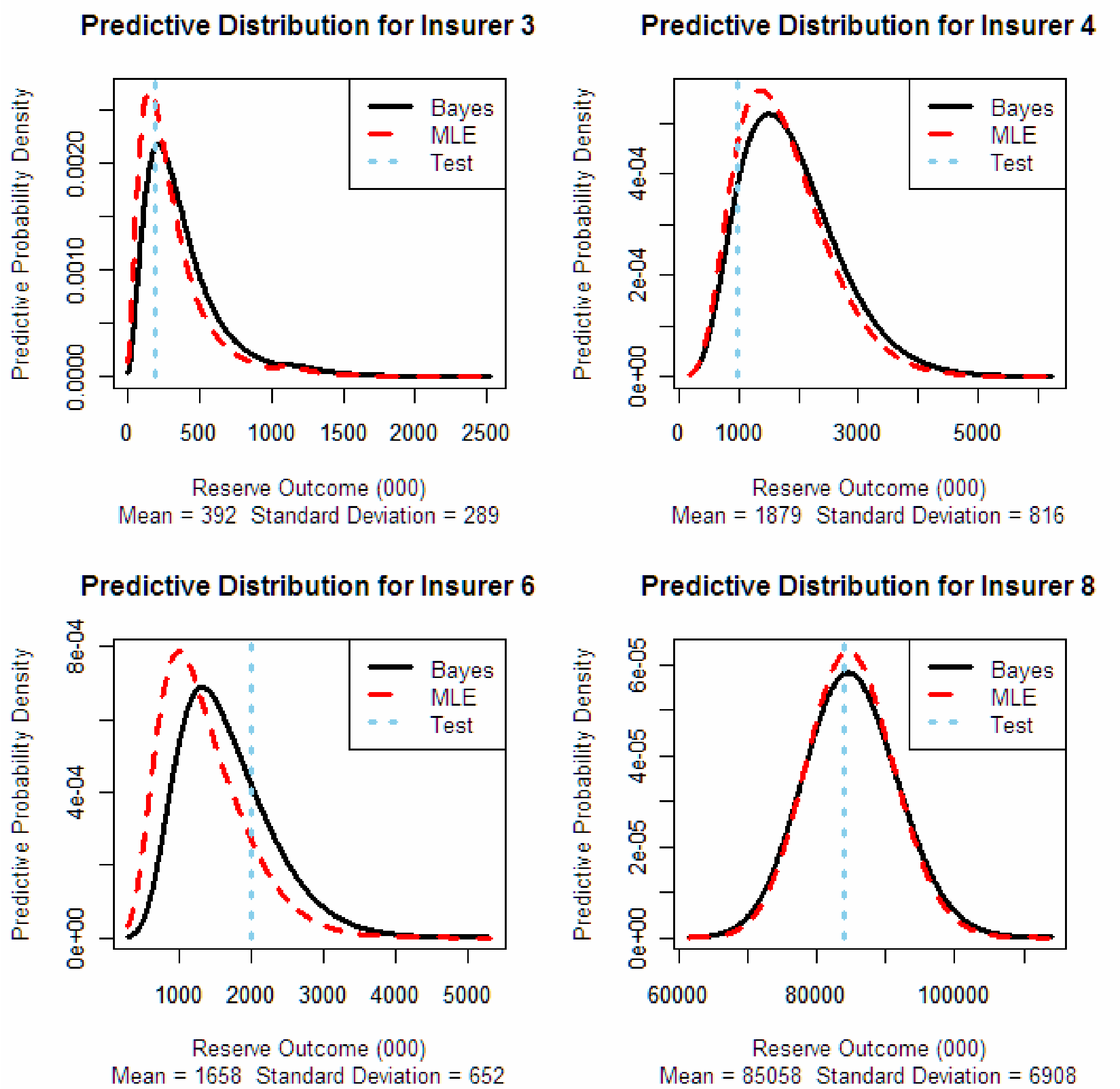
A.2. Since we are assigning posterior probabilities to predetermined discrete parameter sets, in Section A.3 we have to substitute $\Phi(\vec{\mathbf{q}}) = \sum_i \Phi(\vec{\mathbf{q}}_i) \cdot p_i$ for

$\Phi(\vec{\mathbf{q}}) = \dfrac{\sum_i \Phi(\vec{\mathbf{q}}_i)}{n}$ where p_i is the posterior probability of parameter set i.

By thinking of the MLE as a posterior distribution with a single parameter set, one can similarly calculate the distribution predicted by the MLE.

Figure 6 shows some typical plots the Bayesian and the MLE predictive distributions for the sum of the losses in the test data. It also shows observed losses from the test data.

Figure 6



As seen in Paper [2], the Bayesian predictive distribution is generally less variable than the MLE predictive distribution.

Given the predictive distributions and the test losses, we then calculated the predictive percentiles of the test losses for each of the 34 insurers. If the models are valid, the distribution of percentiles should be uniformly distributed, which is testable.

Following Papers [1] and [2], we use PP-plots to as a graphical test that the percentiles are uniformly distributed. Figures 7 and 8 show PP-plots for the Bayesian and MLE predictive distributions.

Figure 7

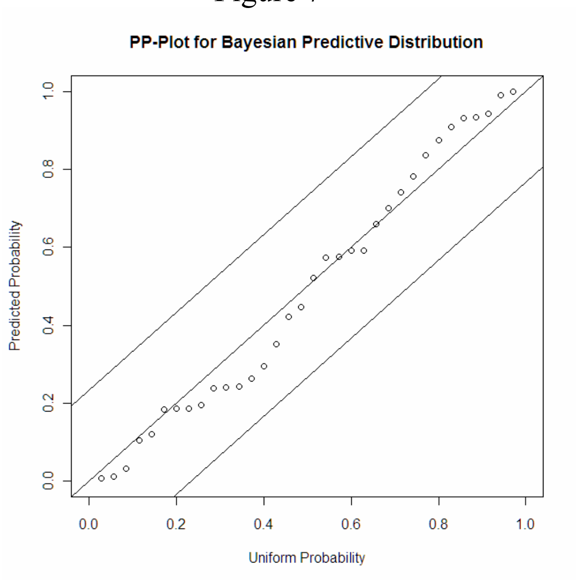
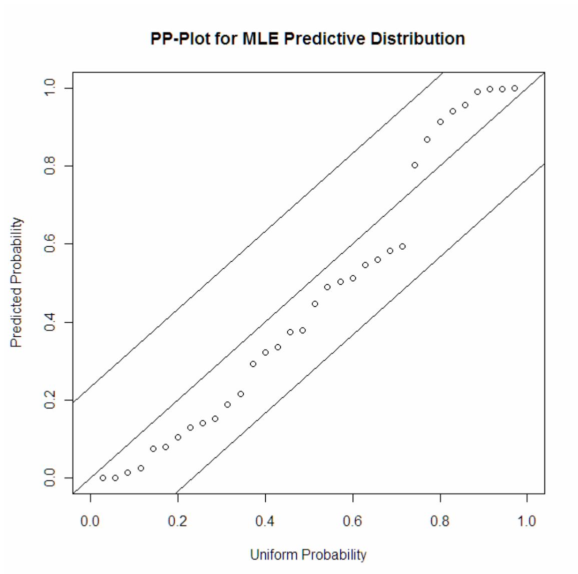


Figure 8



The PP-plot for the MLE has the elongated “S” shape that [2] suggests is a characteristic of overfitting. The Bayesian PP-plot have similar characteristics, but appears to be more consistent with what one would expect if the model were indeed correct.

We characterize these tests as a weak confirmation of the Bayesian model. The confidence bands on the PP-plot are quite wide with only 34 observations.

3.3. Summary

The sections above show that the usual methods underestimate the uncertainty within a reserve estimate given perfect data. This seems to be inherent in the use of MLE techniques to determine the parameters within the stochastic model.

Further research as described above and elsewhere (Meyers 2007) has shown that using Bayesian techniques to allow for the error in parameter selection can partially mitigate this effect, although even this enhancement is no guarantee of correctly predicting the underlying distribution.

We therefore encourage practitioners to consider the results from these models with caution when deciding on the level of uncertainty to relate to their reserves. We also emphasise the importance of correlations between reserving groups when determining any "overall" uncertainty for a book of business, the accuracy of which may have a much greater effect than that described above.

4. International approaches to stochastic reserving

4.1. Introduction

Treatment of reserve risk variability differs by territories. Some territories fully quantify reserve ranges and use them to assess risk margins and probability of adequacy whilst others may provide only qualitative commentary of uncertainty and some may not even provide this.

The trend around the globe (including the impending Solvency II reforms) is toward increased consideration of uncertainty. This section aims to summarise the situation in some of the regions most involved in this trend.

In addition we note that the International Actuarial Association are currently drafting a paper 'Measurement of Liabilities of Insurance Contracts: Current Estimates and Risk Margins'.

http://www.actuaries.org/CTTEES_RISKMARGIN/Documents/RMWG_Exposure_Draft.pdf

4.2. Australia

4.2.1. Regulatory / Accounting Framework

The Australian Prudential Regulatory Authority (APRA) reserving regime has required quantification of reserve uncertainty for many years. APRA requires that an insurer provides an actuarial assessment of its insurance liabilities at least annually.

This assessment must include an explicit risk margin in excess of the discounted mean value of reserves. The risk margin must be equal to the greater of the 75th percentile of the reserve distribution minus the mean, or half of the standard deviation. For companies in run-off the risk margin must equal the 99.5th percentile of the reserve distribution minus the mean.

Reserving is on an accident year basis and risk margins are required for both outstanding claims (OS) and unexpired risk reserve (URR).

Diversification between classes of business and between outstanding claims and unexpired risk is allowed.

Additional prudential margins may be held in the accounts and the probability of adequacy of the total booked reserves including these prudential margins is reported.

All of this requires calculation of a quantitative distribution of reserve outcomes along with correlation matrices between portfolios for both OS and URR and correlation assumptions between the OS and URR.

4.2.2. Professional Guidance

In 2001 the Institute of Actuaries of Australia commissioned Tillinghast to undertake research in respect of net risk margins as defined by the APRA legislation. This report was published in November 2001 and provides benchmark risk margins for many classes of business along with consideration of correlations.

A paper was also published independently in November 2001 by Trowbridge Consulting which presented popular reserve variability quantification methods along with benchmark risk margins and correlations.

These reports (particularly the Tillinghast report) are widely used in the Australian market, with smaller insurers in particular relying heavily on the benchmarks provided.

This perceived over-reliance on benchmarks prompted the General Insurance Practice Council (GIPC) to set up a Risk Margins Taskforce. The Taskforce is due to publish a report in November 2008 which sets out a more robust practical framework for the calculation of risk margins and a more complete suite of tools.

This paper will consider the combination of quantitative and qualitative assessment of reserve uncertainty. In particular it will focus on weaknesses of methods, sources of systemic volatility, what implications these have for the quantitative outputs and the rigorous application of judgement to address these issues.

4.2.3. Market Practice

Methods commonly used in the Australian market to calculate reserve variability include:

- Bootstrap
- Stochastic Chain Ladder
- Hindsight re-estimates (see section 4.2.4)
- Mack
- Blended quantitative / qualitative approaches
- Judgement
- Benchmarks

Typically risk margins are calculated separately by class of business for OSC and URR and the results correlated together. The correlations used tend to be Pearson (linear) correlations and based on judgement, although copulas are sometimes used.

The frequency for review of risk margin calculations (relative to the central estimate) varies from some insurers performing a new analysis at each valuation date, to some retaining the same margins for two or more years.

Reserve variability calculations are typically based on undiscounted data and adjustments are not usually made to allow for additional uncertainty due to possible variation of the discount rate or payment pattern over the runoff of the reserves. Explicit loadings for potential future inflation to be more volatile than seen in the past are not common.

4.2.4. Hindsight re-estimate method

As part of the calculation of actuarial central estimate reserves an actuary is required to compare outcomes to previous projections. This however is rarely extended to reserve variability assessments.

The hindsight re-estimate method was proposed by Andrew Houltram in his 2005 paper and is now widely used in the Australian market. This method aims to incorporate actuarial estimation error in reserve variability analysis by examining historical actuarial central estimates. Typically at least 5 years of historical ultimates are needed.

The aim is to allow for the impact of actuarial judgement (including choice of methods and the use of qualitative information) on the reserve ultimates. This approach acknowledges the fact that reserving is rarely a mechanical process and actuarial judgement typically plays a significant role in the reserving process.

The method introduced the idea of the hindsight re-estimate ie the amended amount that an actuary would have declared as the estimate of the outstanding claims liability at a prior investigation, taking into account experience that has emerged since that investigation.

The deviation, as a % of the original estimate for a given year is known as Hindsight Development Factor (HDF). Bootstrapping is used on the HDFs for each development year to simulate the 'Ultimate Ultimate' for each accident year. An empirical distribution can then be obtained.

The impact of year on year dependency of actuarial ultimates (referred to as 'longitudinal dependence') is addressed by using the method of block resampling (as opposed to point resampling) as previously described by Künsch (1989) and Efron and Tibrishani (1993). This is particularly important as the variation of actuarial ultimates often display positive correlations between years as trends in claims experience are gradually realised.

The hindsight re-estimate method relies on the assumption that past patterns of deviation will repeat themselves. Hence a limitation of the method occurs if there has been a change of actuarial staff, actuarial methods or the reserving process over time.

4.3. USA

The CAS in the US has been focussing on reserving uncertainty for a number of years. In our previous paper we referenced their 2005 Forum and a number of very useful papers arising from it. They are repeating that event this year and have a number of groundbreaking papers to be published both on this topic and others.

Papers of particular interest to the subject being discussed here include the use of paid and incurred data at the same time to improve the estimation of both the mean and uncertainty within reserves. Work has also been published on the testing of methods for robustness, which helps identify those methods that have a greater or lesser sensitivity to particular data points and parameters.

Other papers have also started the discussion on ranges for economic (or discounted) reserves. Papers are presented that address the additional drivers of uncertainty based on data not included in the past history of a triangle and hence go some way to allow for the limitations of MLE fitting discussed elsewhere in this paper.

Interested readers are directed to the CAS September 2008 eForum for these and other interesting papers.

4.4. Switzerland

Switzerland currently employs the Swiss Solvency Test, where risk margins are set on the basis of the cost of the statutory minimum capital requirement. This is designed with the aim that a third party can take over the assets and liabilities of the initial insurer in the event of insolvency. The Federal Office of Private Insurance (FOPI) argues that a third party will only be willing to do this if the cost of setting up the regulatory capital is covered by the portfolio price.

The assets are used to pay for the claims, and the excess is used to pay dividends to the investor for providing the risk capital (his investment). In this situation, it is the investor who takes the long term run-off risk. Therefore, the policyholder will not suffer a financial loss.

We note the developments in Switzerland as they reflect a different set of issues to those typically discussed. The Swiss Solvency Test is parameterised on the basis of a test over a single calendar year. Hence actuaries are required to estimate the reserve volatility over that period rather than to ultimate run-off. As most models focus on the ultimate payments this has caused a certain level of difficulty, although we note that the Hindsight Re-estimate method as discussed above may go some way to address these issues.

4.5. EU

No paper on reserving uncertainty would be complete without referring to the Solvency II developments, however as this topic is beyond the scope of this paper, we are happy to present an incomplete paper in this regard!

5. Areas of common concern when applying stochastic methods

5.1. Background

Those new to stochastic methods may not have a full picture of the areas that are still uncertain regarding the application of methods in daily work. We have therefore considered where our experience indicates that models and methods lack clarity, or can be misleading without detailed knowledge of their assumptions.

In this section we set out a rather mixed set of comments that addresses the key areas we have found challenging when confronting the issues relating to reserve uncertainty modelling and estimation.

5.2. Areas to consider

5.2.1. Check assumptions

It is important to check that the assumptions underlying the model being used are verified, or at least the assumptions that do not hold are identified. Often there are breaches such as: non-homogeneity within the portfolio; the claims triangle is not developed sufficiently and hence tail factors are required, or the claims pattern changes over time. Note that not all of these features break the assumptions of all models.

Where assumptions do not hold it is important to understand the effects this may have on the results. For example, the requirement for non-negative incremental development within the Bootstrap algorithm will tend to upwardly bias the results, and so the mean of the resulting distribution will be higher than the pure Chain-ladder result (assuming such events are treated with a default of flat development).

Thus simply because the underlying data does not fulfil all of the required assumptions for a model, it is not always necessary to completely disregard the results of that model. It can be instructive to consider the results of several imperfect models to get a better understanding of the uncertainty predicted by the data set available.

5.2.2. Consider the type of data being used

Be aware of what the reserve range is applied to. For example, if underwriting year data is used for the exercise, the resultant uncertainty measure reflects both earned and unearned portions of exposure. As the reserve uncertainty exercise is often used to indicate the uncertainty that lies within the earned claims provision only, a method of apportionment needs to be applied to the reserve range to distinguish the uncertainty that arises from the earned and unearned part of reserves.

5.2.3. Changes to claims environment

The actuary needs to be mindful of when the past variability may not fully reflect the future variability. Circumstantial changes such as change in portfolio size, mix of business, claims environment, or where the data does not include a full insurance cycle may mean judgement is required on how much credibility can be given to the historical data when projecting the future. In these instances, scenario testing may be especially helpful in sense-testing the results.

5.2.4. Reinsurance

The insurer's key concern in terms of reserving uncertainty lies within the reserves net of reinsurance. However, if the reinsurance program is not consistent over all periods due to changes in reinsurance structure, commutations, or exhaustion of the cover etc, there may be distorting effects in the underlying process that make many methods invalid. Hence caution is needed when applying any method that only relies on past data in this instance.

In addition changes to the underlying gross exposure such as line size distribution, aggregation and exposure to catastrophes will affect the drivers of uncertainty within the net claims reserve.

5.2.5. ODP methods

The pure ODP method does not allow negative incremental claims. Therefore any triangle with a mean development factors less than 1 for a given development period will not satisfy the model assumptions. Although this predominantly impacts incurred claims data, there can be instances where paid claims data also shows this feature. One should always check the residuals to see how this affects the overall validity of the method.

A potential solution is to set the development factors to a minimum of 1, which automatically increases the mean of the projected reserves and destroys the link between the ODP and Chain-ladder methods. However, if the factors are close to 1 anyway, this effect is often minimal. Otherwise one should allow for the offsetting that occurs when interpreting the overall results of the model.

Alternatively, one can use a different distribution for that particular development period. For example use a Normal with appropriate mean and standard deviation. This obviously doesn't follow the ODP methodology, but the mean result will be correct.

Another option that can be applied is to offset the ODP. This can be done on an individual development factor basis by fitting an ODP with a sensible mean and required standard deviation, but then adding a constant to the results of the distribution such that the resulting mean and standard deviation reflect that given by the residuals of the triangle.

These options are discussed in more detail in section 6.3.3

5.2.6. Incorrect mean

As the best estimate ("mean") of reserves is unlikely to be determined by a pure Chain-ladder approach, it is unlikely that the mean of the stochastic method that uses a Chain-ladder method as its base will match that generated by a full traditional reserving exercise.

Techniques that can be used to adjust for this effect include:

- Shift the model derived distribution such that the mean is reset as the best estimate, without changing the shape or size of the distribution. This may cause issues where the shift results in negative reserves at low percentiles.

- Scale the model output such that every percentile is multiplied by the ratio of the best estimate to the modelled mean. This may then understate or overstate the absolute extent of reserve movements in the extremities of the distribution.
- Choose the underlying stochastic model parameters such that the mean of the distribution reflects the best estimate. This can cause the residuals between the actual data and fitted model to be biased, leading to higher volatility than would otherwise be the case. However it could be argued that this reflects reality to some extent.

5.2.7. Movement of absolute risk margin when scaling variability

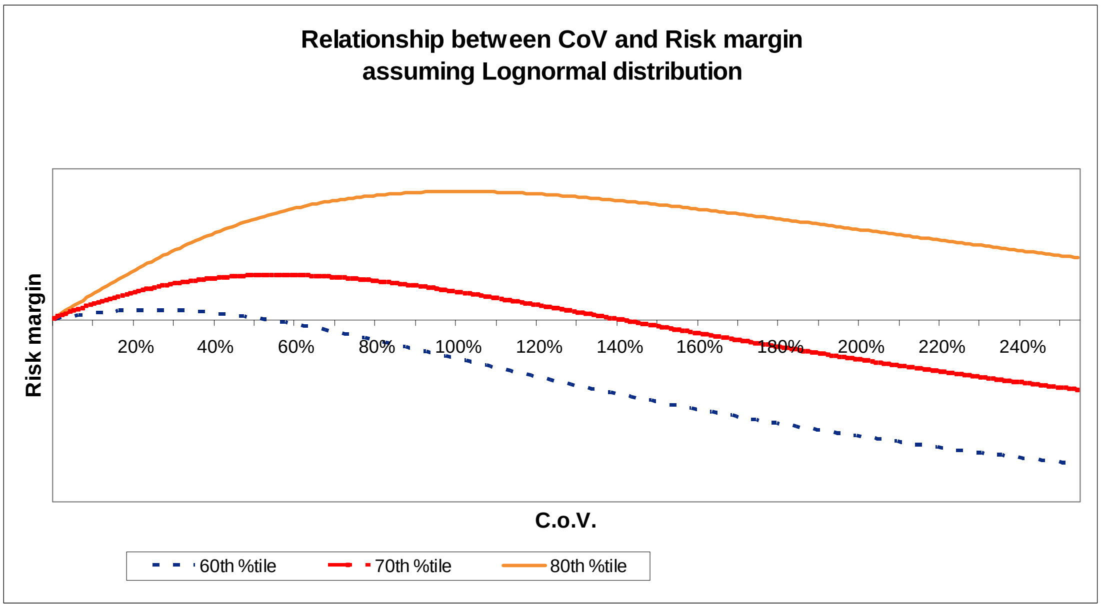
It is common to estimate reserve uncertainty by applying a derived coefficient of variation (CoV) to a mean actual reserve. One effect that this has is that as the CoV increases, the absolute difference between the mean and a high percentile decreases.

Essentially this demonstrates that for a given skewed distribution shape (eg Log-Normal), a higher CoV generally implies the mean sits at a higher percentile. Increasing the CoV whilst retaining a fixed mean can therefore reduce the absolute difference between the mean and a particular percentile. Obviously increasing the CoV also increases the "distance" between percentiles, which works in the opposite direction to the first effect.

The result of this is that if two classes have the same mean, but the "more risky" class has a higher CoV, the reserve at the 80th percentile of the less risky class could be higher than that of the more risky class.

Below is a graph showing how the risk margin, defined at the 60th, 70th, and 80th percentile, varies with CoV for a Log-Normal distribution. At all three percentiles, there is a turning point where the risk margin starts to decrease with increasing CoV and eventually becomes negative. The lower percentiles start showing this 'anomalous' effect at lower CoV's. Taylor (2006) demonstrated mathematically that that for probability p, the turning point occurs when the CoV reaches the standard Normal Z(p) for the 100p percentile of the distribution.

It is worth noting that in Australia the regulatory risk margin is the greater of the 75th percentile and half of the estimated standard deviation of the liability, which may alleviate some of this seemingly anomalous property of risk margins in the regulatory context.



Potentially this effect will increase if the selected "best" estimate has not fully taken account of the entire range of the potential outcomes. As Houtram (2003) points out, it is common for an actuary to come up with a best estimate before considering the reserving uncertainty. Although the best estimate is defined as the expected value of a distribution, a common pitfall when estimating this quantity is to use what seems to be a reasonable scenario without thinking about the full range of potential outcomes, especially the extreme tails.

If falling into this trap, the best estimate may have a tendency to be closer to the median rather than the true mean of the potential distribution, but is then treated as the mean in assessing reserve uncertainty. For a skewed distribution, the median can be significantly lower than the mean, particularly for liability related classes of business; hence one may significantly underestimate the extent of the risk margin at a given percentile.

It is not within our scope to consider further the issues around the calculation of the correct mean of the distribution and we refer the reader to the findings of the other ROC working parties for further discussion on this point.

6. Testing of methods on real data - update

6.1. Introduction

Last year we looked at a number of key methods to see how they compared when applied to real data. This included the predictive power and also ease of use and other subjective assessments.

This year we have extended this to consider transactional level methods, which although they require a more detailed dataset, they have the potential to more accurately allow for changes in the business over time using the additional information.

We have also looked at potential solutions to some of the more common problems found when applying the ODP Bootstrap method, and their effect on the results produced.

6.2. Transactional data methods

6.2.1. Introduction to transactional level methods

Transactional level methods involve building up a distribution of the reserves by examining and making use of the properties of individual claims. Claims considered include both those that have already been reported (paid, outstanding and IBNER) and those that will or may be made in the future (pure IBNR). In theory, by making appropriate use of more granular data, a higher level of the predictive efficiency can be achieved than is possible using aggregate data alone.

As transactional methods consider the underlying properties of individual claims, a more granular level of data is required than for aggregate methods. Typically this will include a full history of the individual claim payments, and/or individual case estimates, together with details about the claim status. Historical inflation figures may also be required. Although the exact data requirements vary from method to method the key requirements of most methods should not be beyond the capabilities of the average insurer.

A brief review of published papers covering individual claims reserving or transactional level methods indicates that a variety of approaches are possible to arrive at both a best estimate and the uncertainty in the reserves. Many methods involve the computations of a compound distribution which is built up from assumptions about the underlying claim numbers and severity distributions. However, some methods only model uncertainty arising from the variability in claim severity and assume that the future number of claims is known or at least relatively stable. Often the uncertainty in claim numbers is important but where an estimate of the pure IBNR is not required or is felt to be immaterial compared to the IBNER, only the claim severity needs to be modelled.

In theory the use of transactional level data in assessing reserve uncertainty has a number of advantages over the use of aggregate (triangular) data. These include:

- A higher level of predictive efficiency may be possible than achieved using aggregate models where key features of the data may be lost on aggregation;
- Simulation of individual claims can help gain a understanding as to the source of the uncertainty, and in particular whether variability is driven by uncertainty in claim numbers or claim severity;
- A more accurate netting down of the gross reserve distribution to account for the actual reinsurance programme is possible;
- The complete distribution of the unpaid liabilities can be projected. This contrasts to many aggregate methods which estimate just the first two moments;
- It should be easier to allow for changes in the nature of the account. For example if there has been a change in the development profile or the relative level of uncertainty in the reserves, it may be easier to adjust a transactional level method in a consistent way than for an aggregate method. Equally adjusting for either an unusually high or low number of very volatile claims in the historical data may be easier than making an adjustment to an aggregate level model;
- Consistency between models used to estimate underwriting and reserving uncertainty may be easier to achieve. Since many underwriting models involve simulation of individual claims, the appropriateness of reserving assumptions can be examined in light of the assumptions made in the underwriting model;
- The claims severity distribution can be tested for appropriateness both against the settled claims as well as the current outstanding case reserves. In particular examining the tail of the distribution against say the ten largest outstanding case reserves can be informative.
- Key drivers of uncertainty in the liabilities can be modelled explicitly e.g. claims inflation could be modelled as a stochastic variable. This is one possible way of providing a link and introducing correlation between the individual claims. Other key drivers of potential correlations could also be explicitly modelled e.g. the prospective Ogden discount rate in a motor book.

On the other hand, transactional level methods have several disadvantages compared to aggregate level methods. These include:

- More assumptions are often required. In particular explicit assumptions about correlations or dependencies structures within reserving classes are often required. While this flexibility may appeal to some actuaries, others may be more comfortable relying on correlations implicit in some of the aggregate methods;
- If best estimate reserves were established using the more traditional aggregate level reserving methodologies then it may be difficult to ensure consistency between the assumptions made in the best estimate and uncertainty analysis. We note however that similar difficulties often exist when trying to reconcile the actuary’s best estimate reserve to the mean of the output of reserve uncertainty models;
- Modelling claims at a transactional level typically takes much longer than modelling claims at an aggregate level. There are likely to be more parameters to estimate, programming can be more complex, and many methods require simulations to be at a granular level;

- Transactional level methods are not currently in widespread use to derive either the best estimate or investigate the uncertainty in the reserves, consequently the underlying concepts are likely to be less familiar to many actuaries;
- There is a potential danger of over fitting the model. This may result in too great a confidence about the predictive power of the model in estimating future claims development.
- Although some correlation between individual claim severities can be explicitly modelled the residual correlation can be extremely difficult to estimate, as can any potential correlation between claim numbers and severities (many methods assume that these are independent).

6.2.2. Modelling considerations

Before a transactional level model is built, a number of factors need to be considered. These include:

- Whether the modelling should be done gross or net of reinsurance and use paid or incurred data. Simulating claims at the gross level would allow precise application of the reinsurance programme so would be the typical choice. Simulating paid claims, rather than changes in incurred claims is likely to be easier but risks throwing away valuable information contained in the outstanding case estimates. If incurred data is used, the method should allow fully for the variability in the unsettled case reserves (IBNER). This is important in order to capture the uncertainty both in the outstanding case estimates themselves as well as ensuring that the claims distribution for unreported claims is sufficiently wide.
- If paid data is to be used, a further decision needs to be made as to whether partial payments should be incorporated into the model, or whether payments in respect to an individual claim are aggregated into a single claim payment assumed to be made when the claim is settled. Often it is simpler to aggregate payments together;
- Should the modelling of the future claim numbers be done in aggregate across all years of account or should simulation be performed at the individual year of account? Should we simulate the calendar year of the claim payments, and allow for the mean and variance of the claims reserves to be functions of time of settlement (often larger and more variable claims settle later), or should we use more aggregated distributions based on an appropriate weighting of more granular distributions?
- What distribution should be used for future claims and how should the parameters be estimated? Should we fit a distribution to the data or is there enough data to use an empirical distribution and Bootstrap from this?
- What distribution should be used for claim severity and how should the parameters be estimated?
- How should parameter uncertainty be incorporated into the model?

If we assume that the number of claims is unknown, the method must capture both the uncertainty in the claim numbers and severities. One possible option would be to use a compound distribution where the future number of claims is modelled using e.g. a Poisson or Negative Binomial distribution. For each non-zero claim the claim severity can then be modelled using e.g. a Log-Normal, Pareto, Gamma or Weibull distribution. For a detailed example of using such a model see the Claims reserving manual v2 (http://www.actuaries.org.uk/__data/assets/pdf_file/0020/24482/crm2-D7.pdf).

6.2.3. Example based on pseudo motor data

Method

We applied a simple transactional level mode to pseudo motor data. The data was created using a model provided by the Casualty Actuarial Society. Mack and Bootstrap methods were also used to estimate the standard error in the reserves and the results of the three methods were compared.

The data consisted of ten accident years of transactional level claims payments and case estimates. Partial payments in relation to each claim were identifiable, as was the accident date, date of payment, whether the claim was open or closed and if closed the date of settlement.

To avoid complications caused by partial settlements only payments with respect to closed claims were used to parameterise the model. To be consistent with the definition of paid claims we used settled claim numbers (including claims settled at nil cost) to be the definition of our claim numbers triangle.

Over all accident years a total of 4,993 claims had been closed, of which 589 had been settled at zero cost.

An accident year triangle of closed claim numbers (includes claims settled at zero cost) was created. The triangle is shown below.

Closed claim numbers

Accident year	Development year 1	2	3	4	5	6	7	8	9	10
1998	206	325	351	359	365	371	372	372	372	372
1999	183	329	356	362	365	367	367	367	367	
2000	206	361	382	391	394	395	395	395		
2001	235	413	437	447	451	457	458			
2002	251	419	458	461	465	468				
2003	232	436	471	479	480					
2004	338	594	635	647						
2005	378	622	674							
2006	413	718								
2007	414									

A cumulative claim paid triangle was also created, but only payments with respect to claims that were closed as at the end of 2007 were included. The triangle is shown below.

Paid claims triangle (only includes payments with respect to settled claims)

Accident year	Development year 1	2	3	4	5	6	7	8	9	10
1998	1,419,830	1,909,101	1,994,922	2,022,320	2,070,237	2,082,772	2,084,521	2,084,521	2,084,521	2,084,521
1999	1,178,858	1,658,775	1,706,496	1,715,105	1,719,856	1,722,601	1,722,601	1,722,601	1,722,601	
2000	885,946	1,508,607	1,532,175	1,553,275	1,556,839	1,557,117	1,557,117	1,557,117		
2001	1,301,764	2,072,241	2,287,300	2,312,045	2,315,920	2,322,334	2,355,494			
2002	1,723,601	3,365,806	3,426,228	3,451,681	3,451,542	3,469,816				
2003	1,508,495	2,417,997	2,508,296	2,572,407	2,579,859					
2004	1,446,536	2,860,502	2,961,968	2,981,804						
2005	1,825,774	2,892,103	3,036,670							
2006	2,410,131	4,178,555								
2007	2,374,862									

The total number of future settled claims (across all accident years) was modelled as a single Negative Binomial distribution. The mean number of future settled claims was estimated to be 509. This was based on the result of the pure Chain Ladder method applied to a triangle of settled claims. Using Bootstrap applied to the settled claims numbers triangle we estimated the standard error as 39. The proportion of nil claims was estimated from the historical claims data as 11.8%.

It would have been possible to model the claim numbers distribution as a Poisson distribution. However the results of Mack and Bootstrap on a claim numbers triangle suggested that the underlying distribution of future claims was skewed, hence a Negative Binomial was chosen as it was thought to be a better fit.

A Log-Normal distribution was fitted to the historical claim severities. The mean and the standard deviation were set equal to the mean and standard deviation of the historical settled claim sizes.

The resulting distributions were reviewed for appropriateness against the outstanding claim listing and were accepted.

@Risk was used to generate 10,000 simulations, each consisted of generating a future number of claims and subtracting an estimate of the number of future nil claims. For each non-zero claim a future claim size was generated from the selected distribution. The total future claim payment was then calculated as the sum of the individual claims estimates.

We assumed that the future proportion of nil claims was fixed and did not vary over development period – a triangle of nil claim settlements as a proportion of total settled claims numbers was used as a high level test to the reasonableness of this assumption. In this example no evidence was found of a markedly different settlement pattern for nil claims.

The mean and the standard deviations of the 10,000 simulations were calculated and compared to the results of applying Mack and Bootstrap to an aggregate paid claims triangle.

Note that future settled claim numbers were simulated in aggregate over all accident years. Some transactional level methods will model claims at the individual accident year level. If the average claim size varies with the age of development, simulation at either the individual accident year cohort or allocating future claim numbers to a cohort may be necessary in order that appropriate claim severity distributions can be applied to each year.

Results

	Mean	Standard dev.	CoV
Transactional	2.60m	395k	15%
Mack	2.29m	503k	22%
ODP Bootstrap	2.29m	384k	17%

The mean of the transactional method was substantially higher than that of the pure Chain Ladder (the method behind the mean of the Mack and Bootstrap). Although the mean of 2.60m was the based on the simulated results, it can be estimated numerically as: mean IBNR claims number x proportion of claims settled at cost x average claim size = 509 x (1-11.8%) x 5,808 =2.61m.

Differences between the means of the transactional level method and the Chain Ladder can arise for a number of reasons. They include the different weights that the methods give to data in different parts of the triangle, for example a Chain Ladder may give more weighting to a kick or a drop at the far right of the triangle than is given by the transactional method. Equally no account has been made in the transactional method for any tendency for claim size to vary with development period, but this would be captured by the Chain Ladder.

We also note that the standard error of the transactional level method was similar to that of the Bootstrap but quite a bit below that of Mack. The estimate of the standard deviation of the transactional method was driven by the uncertainty in the future claim sizes. The result was relatively insensitive to small changes in the assumed claim number distribution.

Where Mack and Bootstrap are used to estimate the future claim number variability, some of the problems highlighted in section 2 and in our previous paper may therefore follow through to the transactional method. However it should be noted that based on this example, the vast majority of the variability came from the claim size.

Within the transactional method, individual claim sizes were assumed to be independent. It is possible to extend the example and model future inflation or other potential drivers of correlation in claim sizes as stochastic variables. In addition claim size and numbers are also assumed to be independent.

No allowance has been made for parameter error in the transactional method. This together with assuming independence in future claims sizes is perhaps one of the reasons why the coefficient of variation is lower than that of Mack and Bootstrap.

We note that the data provided contained both property damage and bodily injury claims data. Arguably it would have been better to split this data out into property damage and Bodily Injury claims and analyse separately, however the aggregation of the results would then have required assumptions regarding correlations between the two datasets.

Possible refinements
Some claim level reserving methods use a Negative Binomial distribution, not because the underlying claims distribution is assumed to be skewed but to incorporate parameter uncertainty. When the underlying distribution is a Poisson distribution and the unknown parameter of the Poisson distribution is assumed to be from a Gamma distribution, then the posterior distribution is a Negative Binomial. In this case, allowance is being made for parameter uncertainty. The rationale for the choice of all distributions should be carefully documented, and consideration given as to whether parameter uncertainty has been allowed for.

We note that significant information is contained within the case reserves. We have not used this in the example above, but note that this could be used to assist in the parameterisation of the claim severity distribution.

Finally, in some instances it may be valuable to be able to allow for correlations between claim number and severity. This could then be used to capture effects such as seen in economic recessions within liability claims.

6.2.4. Summary

We believe that transactional level methods are a useful tool for an actuary to help in quantifying reserve uncertainty. In our opinion, there is much to be gained from further research into such methods.

Using transactional level methods alongside more traditional methods can help to improve the robustness of reserving estimates both at the best estimate as well as at the extremes of the distribution.

As with other methods, model error is hard to quantify and can be significant, however running transactional level analysis in parallel with more traditional aggregate methods should help to start to highlight the sensitivity of results to different models.

In particular transactional level could be used in relation to large claims and the results combined with Mack / Bootstrap results for attritional claims. This would allow the more volatile larger claims to be modelled in greater detail, however aggregating such results would be problematic due to the requirements of assumptions regarding correlations between the two types of claim.

6.3. Extension to ODP Bootstrap method

This section looks at developments to the basic over-dispersed Poisson (ODP) Bootstrapping method (as set out in 'Stochastic Claims Reserving in General Insurance' by Peter England and Richard Verrall, published in the British Actuarial Journal in 2002) and how these developments are implemented in practice.

These developments fall into three areas:

- Using a variable scale parameter;
- Moving beyond the Basic Chain Ladder: curve fitting and tail estimation;
- Dealing with negative development factors.

6.3.1. Using a Variable Scale Parameter

In England & Verrall (2002) , the Bootstrapping procedure was applied to an over-dispersed Poisson generalised linear model that gives the same forecasts as the traditional Chain Ladder method, using a single scale (or dispersion) parameter. The scale parameter is used for two purposes: to scale the residuals such that their variance is approximately 1, and to estimate the process variance when forecasting.

This is essentially a simplifying assumption, usually improved upon by employing a scale parameter that varies with each development period.

One of the fundamental assumptions of Bootstrapping is that the residuals should be independent and identically distributed. This is more easily achieved by using a scale parameter that varies with each development period.

It is also more consistent with the methodology employed by Mack's model. With Mack's model, the variance of cumulative payments is assumed to be proportional to the previous cumulative value, where the constant of proportionality can be viewed as a scale parameter, and has a different value for each development period in the triangle.

The only drawback to using a varying scale parameter is that there can be insufficient data to provide credible estimates for all development periods (since we have a triangle of data). This issue is particularly prominent in the tail of the triangle, and is an issue for the ODP model and Mack's model.

Further discussion of using a variable scale parameters for ODP Bootstrapping can be found in the 2006 paper 'Predictive Distributions of Outstanding Liabilities In General Insurance' by Peter England and Richard Verrall, published in the Annals of Actuarial Science in 2006.

6.3.2. Moving Beyond the Basic Chain Ladder: Curve Fitting and Tail Estimation

Bootstrapping is simply a statistical procedure that can be applied to a well-defined statistical model to obtain distributions of parameters. Forecasts can then be simulated, if required, conditional on those parameters. In England & Verrall (2002) the procedure was applied in the reserving context to a specific GLM, fitted to a triangle of data, which happens to produce the same expected predictions for future developments as would a BCL approach. Unfortunately, the term "Bootstrapping" in this context is often wrongly associated with that specific model. It is important to separate Bootstrapping as a statistical procedure from the underlying model that it is applied to.

The procedure has three steps. Step 1 is to define and fit the statistical model, and obtain parameter estimates and appropriate residuals. Step 2 is to resample the residuals (with replacement), invert the resampled residuals to give pseudo-data, and re-fit the same model that was defined at Step 1 to each set of pseudo data. This gives a distribution of parameters. Step 3 is to simulate forecasts into the future, conditional on the parameters obtained at Step 2.

For the specific model described in England & Verrall (2002), the procedure can be processed far more quickly by applying a BCL methodology to the pseudo-data instead of fitting a GLM to the pseudo-data. This is a sleight of hand that is justified on the basis that (for that specific model) the forecasts are identical.

Once we have justified the approach for a standard case (the pure Chain Ladder model with no tail), we can generalise it and include any of the common adjustments that we might make to the BCL for reserving purposes. We simply follow the three steps outlined above, and perform a sleight of hand at the appropriate point if required.

The following common extensions to the BCL methodology can be readily incorporated into the Bootstrapping process:

- Manipulating development factors;
- Curve-fitting for the estimation of tail factors;
- Smoothing the underlying data, and hence the development factors, without curve-fitting.

Incorporating these techniques into Bootstrapping, and the impact that they have, is discussed below.

Manipulating Development Factors

General development factor methods require an average development factor to be selected for each development period. Manipulating development factors covers the exclusion of certain development factors in calculating averages and the selection of which type of average to calculate for each development period.

The standard Chain Ladder method takes an average of all development factors for a given development period, weighted by the cumulative developments in the subsequent period. There are various alternative averages that might be selected, such as arithmetic averages or time-weighted averages.

When using this process in conjunction with Bootstrapping, the procedure is as follows:

1. Calculate a triangle of development ratios;
2. Choose which development factors to exclude and which form of average to calculate on the remaining development factors in the development period in question.

At this stage, we have selected development factors to create an alternative development factor method to produce a fixed point estimate of the ultimate position. In order to apply Bootstrapping to this estimate, the procedure is as follows:

3. Calculate residuals using fitted values derived from the selected alternative development factors;
4. Remove from the pool of residuals all the points that correspond to excluded development factors. (We do this because the exclusion of a development factor implies that it is sufficiently out of line with what – based on actuarial judgement – we would consider reasonable. It would therefore be inappropriate to include it as representing potential future developments.);
5. Resample with replacement from the reduced pool of residuals to produce a full triangle of pseudo-data (in the same way as with standard Bootstrapping);

6. Calculate a triangle of development ratios;
7. Exclude the same development factors in development and origin periods that were excluded in the original projection of the actual data;
8. Apply the same averaging calculations to the development factors (after exclusions) that were used in the original projection of the actual data to derive average development factors;
9. As with standard Bootstrapping, perform this process on a large number of sets of pseudo-data to produce a distribution of outputs.

Note that the exclusion of data from the triangle reduces the credibility of the fit, and also exacerbates the over-fitting effects discussed elsewhere in this paper. In addition it is debateable whether any data point should be excluded unless it is due to data errors as this is effectively ignoring valid volatility within the data.

<u>Curve-fitting</u>
Curve fitting is used in traditional reserving to smooth through the noise in the later development factors of the data and to extrapolate the data where the oldest relevant origin period is not fully run off.

The procedure is as follows:
1. Calculate a triangle of development ratios;
2. Choose which development factors to exclude and which form of average to calculate on the remaining development factors in the development period in question;
3. Choose a subset of the average development factors to which to fit curves;
4. Fit curves to these average development factors (typically exponential decay, inverse power, power, and Weibull);
5. Select which development factors are used for forecasting. That is, which are based on the average factors and which should be based on the curve-fits;
6. Select how far to extrapolate into the future and which curve should be used to do this;
7. Derive a tail factor as the product of the extrapolated development factors derived from the selected curve.

At this stage, we have used curve-fitting techniques to produce a fixed point estimate of the ultimate position. In order to apply Bootstrapping to this estimate, the procedure is as follows:
8. Calculate residuals using fitted values derived from the selected alternative development factors and curve;
9. Exclude any residuals corresponding to excluded development factors;
10. Resample with replacement from the (reduced pool of) residuals to produce pseudo-data (in the same way as with standard Bootstrapping);
11. Calculate a triangle of development ratios;
12. Exclude the same development factors in development and origin periods that were excluded in the original projection of the actual data;
13. Apply the same averaging calculations to the development factors (after exclusions) that were used in the original projection of the actual data to derive average development factors;
14. Fit the same curve to the average development factors, excluding average development factors from the same positions as the original curve fit;

15. Select which development factors are used for forecasting. That is, which are based on the average factors and which should be based on the curve-fits. Do this using the same selections for each development period that were used in the original projection;
16. Forecast into the future using the same number of future development periods and the same curve as selected in the extrapolation stage of the original projection;
17. As with standard Bootstrapping, perform this process on a large number of sets of pseudo-data to produce a distribution of outputs.

Smoothing through development factors (without curve-fitting)
This is an alternative method for dealing with development factors that are unusual, without resorting to excluding them. It uses a calculation based on the underlying data to smooth through spikes in the data.

This method is typically used when an unusually large or small development is immediately followed by an unusual movement of opposite magnitude. However, it is possible to smooth through more than two consecutive points at once. Common smoothing techniques are linear interpolation or fitting higher order polynomial splines.

The procedure is as follows:
1. Calculate a triangle of development ratios;
2. Select which contiguous development ratios exhibit spikes that are specific to data issues that should not be projected into the future;
3. Smooth through the underlying cumulative data associated with those ratios, ensuring the cumulative amounts at the start and end of the smoothing region remain unchanged;
4. Recalculate the development ratios, and then proceed in the same way as described under "Curve-fitting".

Again this method may have the drawbacks of restricting the uncertainty being modelled as discussed above.

Summary
There are a number of consequences of deviating from the basic ODP methodology in the manner described above and these should be borne in mind throughout the modelling process.

Applying a selected methodology to pseudo-datasets in the manner described above is a very literal interpretation of applying the methodology. It applies the same selections that were made in modelling the actual data (Step 1) rather than applying the same rationale. For example, when fitting a curve to the data to derive a tail factor, the initial selection may have been, say, the inverse power curve since it was the best fit to the data. However, the curve selected when deriving a tail factor for each set of pseudo-data will be the inverse power curve every time, rather than the curve that best fits the data. This is because Bootstrapping is being used to obtain a distribution of parameters for the inverse power curve.

Curve-fitting in particular can lead to very volatile outcomes, since the fit is an automatic mathematical calculation. When a curve is fitted to claims data, we need the second derivative of the curve to be negative in the tail (that is, we need the curve to flatten off). However, for some pseudo-datasets, the development factors to which a curve is fitted can produce curves with positive second derivatives in the tail. This leads to tails that increase rather than flatten off. If this occurs, it can be readily identified in the output of percentiles and the model must be refitted.

Another common problem comes about when the fitted tail is overly long. When projecting data to obtain a point estimate, there is no significant impact to fitting a long tail rather than a short tail to the data, so long as the extra development factors in the long tail are very close to 1. However, this is not the case when Bootstrapping due to the simulated process uncertainty. Even if the projected increment for a development is only slightly larger than 0, this projected value is used as the mean of a distribution, so the Bootstrapping process will continue to simulate variability in the flat section of the tail. The consequence of this is that the number of developments in a fitted tail should be kept to a realistic estimate of how long the data should take to run off in order to avoid over-estimating the variability in the tail, particularly as negative offsetting movements are not possible using the standard technique.

The better the model fits the data, the smaller the residuals will be, and hence the lower the simulated future parameter uncertainty. This means adjustments to the model such as smoothing through development factors using curve fitting may increase the volatility of the forecasts. However, a trade-off between goodness-of-fit and number of parameters (included in the calculation of the scale parameters) should balance this out (see England & Verrall, 2006).

All of the methods discussed in this section are natural extensions to carrying out Bootstrapping using development factor modelling on pseudo-data. However, they are not rigorously justified by the underlying theory since they do not project results by fitting a GLM to the data. Where a GLM would usually be fitted on theoretical grounds, a sleight of hand is being performed by using a standard actuarial deterministic method. As with any actuarial technique, care and judgement must be applied. The adjustments discussed above should not be made within a black box process, but rather should be carefully considered having Bootstrapped a standard Chain Ladder model first, then steadily deviating from that and gaining an understanding of what is driving variability.

6.3.3. Dealing with negative pseudo-development factors

Bootstrapping attempts to capture two sources of uncertainty – parameter uncertainty and process uncertainty.

Process uncertainty refers to the inherent randomness in claim development. It is captured in the Bootstrapping process by simulating future developments conditional on forecasts based firstly on the simulated parameters. When this process uncertainty is modelled using a positive distribution (such as the gamma distribution, which is a typical choice) it is not possible to use a negative mean.

A negative mean can occur when a pseudo-development factor is less than one. This can only occur when the sum of pseudo-incremental values in any development period is less than zero. Note that it does not occur simply if a pseudo-incremental value is negative, but if the sum within a development period is negative. This can occur during the residual inversion stage, and may or may not be an issue, depending on whether negative incremental values are perceived as being valid or not. This is usually less common when non-constant scale parameters are used. Note that negative incremental movements are often observed in the underlying data, for a variety of reasons. In paid data, negative increments may represent recoveries through salvage and subrogation; whilst in incurred data, negative increments may also be a consequence of claims assessors revising down case estimates.

There are three main methods to work around this problem:

- Censor the pseudo-development factor so that it is positive, but very close to 1;
- Switch to a normal distribution when required;
- Translate (shift) the distribution.

Censor the pseudo-development factor

This is a very straightforward approach, but potentially causes a bias. However, if negative increments are frequently simulated (generally because there are negative increments in the real data) then the ODP model is unsuitable in any case, so a simple work-around may be appropriate if it is only called upon infrequently.

Switch to a normal distribution when required

This has the benefit of maintaining the desired mean and variance for the simulated increment. However, the normal distribution is symmetric, so this does not allow for the positive skewness present in a distribution such as the gamma. Usually, the need to switch to a normal distribution is a rare occurrence, so this will not be noticeable. Ideally, the number of times this is required should be monitored.

Translate the distribution

For a given negative mean, simulate the increment from a positive distribution with a mean equal to the absolute value of the required mean (which is therefore positive). Negatively translate the distribution by double the positive mean, thus resulting in a distribution with the required negative mean. This method has the advantage of maintaining the positive skewness of the desired distribution, whilst also maintaining the required negative mean.

6.3.4. Worked example

We have applied the techniques above to the Employers Liability data set as used in our 2007 paper. This consists of annual data from underwriting years 1985 – 2005 as at the end of 2005.

Unless otherwise stated, all tests use a gamma distribution for process variance, switching to a normal distribution if a simulated future incremental claim movement is negative. All tests using ODP are on paid data and all tests using Mack are on incurred data.

The tests we performed are set out on the graph and tables below. The various scenarios are:

- **BCL ODP single -** This uses a straightforward ODP model on a BCL, using a single scale parameter.

- **BCL ODP single excluding ratios -** This model amends the first model by excluding a proportion of the development factors from the more recent developments in the triangle, where noise and spikes start to come about as a result of sparse data.

- **BCL ODP single Curve Fit -** This model builds on the first one by fitting an inverse power curve through the development factors. This curve is used from development year 11 onwards, including the addition of a tail from development year 20 to ultimate.

- **BCL ODP variable -** This model is the same as BCL ODP single, but uses variable scale parameters rather than a single scale parameter.

- **BCL ODP variable excluding ratios -** This model is the same as BCL ODP single excluding ratios, but uses variable scale parameters rather than a single scale parameter.

- **BCL ODP variable Curve Fit -** This model is the same as BCL ODP single Curve Fit, but uses variable scale parameters rather than a single scale parameter.

- **BCL ODP variable 1997 & prior excluded -** This model excludes all development factors from 1997 and prior, and fits an inverse power curve to the data to provide a tail factor. This curve takes over from the derived development factors from year 7 onwards.

- **BCL ODP variable smoothed -** This model is the same as BCL ODP single, but uses variable scale parameters rather than a single scale parameter. Furthermore, the variable scale parameters have been smoothed (using actuarial judgement) so that simulated future developments do not mirror the incidental jumps in the historical data in the equivalent development periods.

- **BCL ODP variable smoothed excluding ratios -** This model is the same as BCL ODP single excluding ratios, but uses variable scale parameters rather than a single scale parameter. Furthermore, the variable scale parameters have been smoothed as described above.

- **BCL ODP variable smoothed Curve Fit -** This model is the same as BCL ODP single Curve Fit, but uses variable scale parameters rather than a single scale parameter. Furthermore, the variable scale parameters have been smoothed as described above.

- **BCL ODP variable Use 0.01 -** This model is the same as BCL ODP single, but uses variable scale parameters rather than a single scale parameter. However, in this model, if a negative future development is simulated then we use a value of 0.01 to allow for the process variance.

- **BCL ODP variable smoothed excluding ratios -** This model is the same as BCL ODP single excluding ratios, but uses variable scale parameters rather than a single scale parameter. Furthermore, the variable scale parameters have been smoothed as described above.

- **BCL Mack -** This model uses a straightforward Mack model on a BCL.

- **BCL Mack excluding ratios -** This model amends a BCL by excluding some large spikes in the data in the initial development phase.

- **BCL Mack 1997 & prior excluded -** This model excludes all development factors from 1997 and prior. No tail is required as the 1997 year can be considered to be fully developed on an incurred basis.

- **BCL Mack smoothed -** This model amends a BCL by smoothing Mack's alpha (using actuarial judgement) so that simulated future developments do not mirror the incidental jumps in the historical data in the equivalent development periods.

- **BCL Mack smoothed excluding ratios -** This model is the same as BCL Mack excluding ratios, but Mack's alpha has been smoothed as described above.

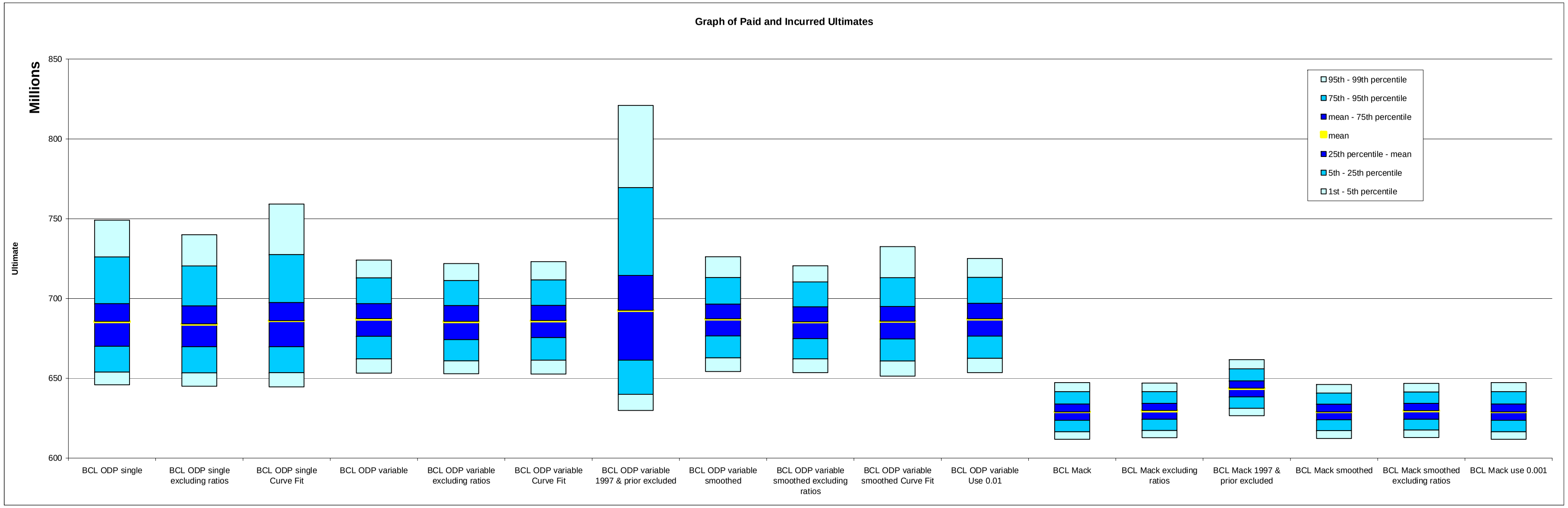
- **BCL Mack use 0.01 -** This model is the same as BCL Mack, but in this model, if a negative future development is simulated then we use a value of 0.01 for the process variance.

	CLAIMS RESERVE										
	BCL ODP single	BCL ODP single excluding ratios	BCL ODP single Curve Fit	BCL ODP variable	BCL ODP variable excluding ratios	BCL ODP variable Curve Fit	BCL ODP variable 1997 & prior excluded	BCL ODP variable smoothed	BCL ODP variable smoothed excluding ratios	BCL ODP variable smoothed Curve Fit	BCL ODP variable Use 0.01
Mean	125,198,865	123,612,811	125,871,578	126,806,599	125,201,594	125,726,748	132,234,971	126,859,917	125,139,587	125,567,878	126,879,710
CV	17.4%	16.2%	20.0%	12.0%	12.2%	12.1%	31.4%	12.0%	11.6%	13.7%	12.1%
1st Percentile	85,836,388	84,965,445	84,521,127	93,158,693	92,809,265	92,623,794	69,729,276	94,226,108	93,426,480	91,304,367	93,516,606
5th Percentile	93,795,850	93,394,188	93,487,720	102,105,671	100,912,628	101,280,417	79,882,941	102,760,692	102,042,896	100,749,894	102,411,105
25th Percentile	110,017,509	109,788,022	109,802,979	116,272,468	114,235,102	115,513,966	101,345,405	116,435,868	114,768,406	114,580,576	116,338,218
50th Percentile	123,098,114	122,173,905	122,606,813	126,558,616	124,677,231	125,283,455	124,897,137	126,162,470	124,851,041	124,367,698	126,597,903
75th Percentile	136,744,268	135,321,738	137,377,785	136,714,636	135,539,550	135,616,518	154,359,677	136,403,032	134,685,859	134,926,442	136,896,082
95th Percentile	165,944,066	160,349,439	167,428,635	152,815,239	151,150,073	151,627,938	209,457,612	153,113,745	150,314,773	153,007,256	153,141,250
99th Percentile	189,008,500	179,865,228	199,064,284	163,986,424	161,763,882	162,995,065	260,877,159	166,089,080	160,397,026	172,380,033	164,893,485

	IBNR					
	BCL Mack	BCL Mack excluding ratios	BCL Mack 1997 & prior excluded	BCL Mack smoothed	BCL Mack smoothed excluding ratios	BCL Mack use 0.01
Mean	(16,405,738)	(15,955,299)	(1,819,300)	(16,395,843)	(15,876,539)	(16,405,738)
CV	(46.5%)	(46.2%)	(410.8%)	(44.0%)	(45.8%)	(46.5%)
1st Percentile	(33,577,869)	(32,525,223)	(18,665,664)	(33,008,309)	(32,401,042)	(33,577,869)
5th Percentile	(28,705,085)	(27,963,990)	(14,055,665)	(28,086,736)	(27,631,611)	(28,705,085)
25th Percentile	(21,577,583)	(20,936,335)	(6,886,321)	(21,355,798)	(20,858,490)	(21,577,583)
50th Percentile	(16,466,246)	(16,061,280)	(1,945,550)	(16,442,351)	(15,938,878)	(16,466,246)
75th Percentile	(11,347,879)	(11,021,789)	3,125,998	(11,562,515)	(10,978,605)	(11,347,879)
95th Percentile	(3,653,734)	(3,611,790)	10,628,785	(4,541,828)	(3,796,676)	(3,653,734)
99th Percentile	1,975,576	1,653,269	16,407,705	819,966	1,522,839	1,975,576

	ULTIMATE CLAIMS (PAID PROJECTION)										
	BCL ODP single	BCL ODP single excluding ratios	BCL ODP single Curve Fit	BCL ODP variable	BCL ODP variable excluding ratios	BCL ODP variable Curve Fit	BCL ODP variable 1997 & prior excluded	BCL ODP variable smoothed	BCL ODP variable smoothed excluding ratios	BCL ODP variable smoothed Curve Fit	BCL ODP variable Use 0.01
Underlying Model	Paid Claims Ultimate	Paid Claims Ultimate excluding ratios	Paid Claims Ultimate Curve Fit	Paid Claims Ultimate	Paid Claims Ultimate excluding ratios	Paid Claims Ultimate Curve Fit	DFM Paid Claims Ultimate 1997 & prior excluded	Paid Claims Ultimate	Paid Claims Ultimate excluding ratios	Paid Claims Ultimate Curve Fit	Paid Claims Ultimate
DFM Ultimate	686,858,409	629,503,989	685,343,715	686,858,409	629,503,989	685,343,715	643,587,046	686,858,409	629,503,989	685,343,715	686,858,409
Mean	770,531,122	768,945,068	771,203,834	772,138,856	770,533,851	771,059,004	777,567,227	772,192,174	770,471,844	770,900,134	772,211,967
1st Percentile	731,168,645	730,297,701	729,853,384	738,490,949	738,141,522	737,956,051	715,061,533	739,558,365	738,758,737	736,636,623	738,848,862
5th Percentile	739,128,107	738,726,445	738,819,977	747,437,928	746,244,885	746,612,674	725,215,198	748,092,948	747,375,153	746,082,150	747,743,362
25th Percentile	755,349,766	755,120,278	755,135,236	761,604,725	759,567,358	760,846,222	746,677,661	761,768,125	760,100,663	759,912,833	761,670,475
50th Percentile	768,430,371	767,506,161	767,939,069	771,890,873	770,009,487	770,615,711	770,229,393	771,494,727	770,183,298	769,699,955	771,930,159
75th Percentile	782,076,525	780,653,995	782,710,042	782,046,893	780,871,807	780,948,775	799,691,933	781,735,289	780,018,116	780,258,699	782,228,338
95th Percentile	811,276,323	805,681,696	812,760,891	798,147,496	796,482,330	796,960,195	854,789,869	798,446,001	795,647,030	798,339,513	798,473,507
99th Percentile	834,340,757	825,197,485	844,396,541	809,318,681	807,096,138	808,327,321	906,209,416	811,421,336	805,729,283	817,712,289	810,225,742

	ULTIMATE CLAIMS (INCURRED PROJECTION)					
	BCL Mack	BCL Mack excluding ratios	BCL Mack 1997 & prior excluded	BCL Mack smoothed	BCL Mack smoothed excluding ratios	BCL Mack use 0.01
Underlying Model	Incurred Claims Ultimate	Incurred Claims Ultimate excluding ratios	Incurred Claims Ultimate 1997 & prior excluded	Incurred Claims Ultimate	Incurred Claims Ultimate excluding ratios	Incurred Claims Ultimate
DFM Ultimate	628,979,870	629,503,989	643,587,046	628,979,870	629,503,989	628,979,870
Mean	628,926,518	629,376,958	643,512,957	628,936,414	629,455,717	628,926,518
1st Percentile	611,754,387	612,807,034	626,666,593	612,323,948	612,931,215	611,754,387
5th Percentile	616,627,172	617,368,266	631,276,592	617,245,520	617,700,645	616,627,172
25th Percentile	623,754,674	624,395,922	638,445,936	623,976,459	624,473,767	623,754,674
50th Percentile	628,866,011	629,270,977	643,386,707	628,889,906	629,393,379	628,866,011
75th Percentile	633,984,378	634,310,467	648,458,254	633,769,742	634,353,652	633,984,378
95th Percentile	641,678,522	641,720,467	655,961,041	640,790,429	641,535,580	641,678,522
99th Percentile	647,307,833	646,985,525	661,739,961	646,152,223	646,855,095	647,307,833



6.4. Summary

We have presented a number of common amendments to the standard Bootstrap ODP method. These are intended to overcome some practical issues in the application of the technique to real data. In most cases these will reduce the applicability of the analytic features of the method, but do not remove the usefulness of the results.

The example data set we have applied these to show that they can have a significant effect in the calculation of both the mean and the extreme percentiles of the distribution of outcomes.

7. Areas for further research

7.1. Introduction

The scope for our work this year has largely been focussed on how to address the errors found in the methods commonly used to assess reserve uncertainty. We have gone some way to addressing this, particularly in determining some of the characteristics of a "good" triangle.

However, we have only touched on a few areas within the broader topic in the above sections. Here we set out a number of other areas where we believe that more research needs to be done in order to allow greater understanding and use of stochastic reserving methods within the Profession.

7.2. Methodologies

- Continuing investigations as to what characteristics a "good" data set has for use in one or more methods, particularly an indication of tests to possibly identify the expected level of error a given method will have on a particular dataset.
- Reserving cycle and necessary amendments to stochastic reserving methods
- Underwriting vs accident year data sets and interpretation of the results from one basis to applications in the other
- Discounting – variation between discounting stochastic reserving results of undiscounted amounts, and comparison with stochastic discounting calculations
- Reserving "catastrophes" and their parameterisation

7.3. Correlations

There are many types of correlation used within reserve uncertainty estimates. As mentioned above, these assumptions can have a dramatic effect on the anticipated uncertainty within a combined portfolio. We believe that the sensitivity of the results to these assumptions deserves detailed investigation. These include:

- Correlations between data sets
- Correlations between origin years
- Derivation of correlations through top down vs bottom up methods and the issues that arise
- Pearson (linear) vs copula correlations and the effects on the overall reserve percentiles

8. Bibliography

- Research and data analysis relevant to the development of standards and guidelines on liability valuation for general insurance (Tillinghast – 2001)

- APRA risk margin analysis (Trowbridge – 2001)

- An empirical approach to insurance liability prediction error assessment (Houltram – 2005)

- A framework for estimating uncertainty in insurance claims cost (O'Dowd, Smith & Hardy – 2005

- Linear correlation as a measure of dependency (Britt & Napoli – 2005)

- APRA general insurance risk margins (Taylor - 2006)

- Reserving judgement (Houltram – 2003)

- Assessing & monitoring insurance liability uncertainty (Collins et al. – 2005)

- "Estimating Predictive Distributions for Loss Reserve Models" Meyers, Variance Vol.1, Issue 2 (2007) p248 http://www.variancejournal.org/issues/01-02/248.pdf

- "Thinking Outside the Triangle" G. Meyers ASTIN 2007

  http://www.actuaries.org/ASTIN/Colloquia/Orlando/Papers/Meyers.pdf

- "Stochastic Loss Reserving with the Collective Risk Model" G. Meyers (2008)

  CAS eForum 2008 http://www.casact.org/pubs/forum/08sforum/11Meyers.pdf