

Back-Testing the ODP Bootstrap of the Paid Chain-Ladder Model with Actual Historical Claims Data

By Jessica (Weng Kah) Leong, Shaun Wang, Han Chen

ABSTRACT

This paper will back-test the popular over-dispersed Poisson (ODP) bootstrap of the paid chain-ladder model, as detailed in England and Verrall (2002), using real data from hundreds of U.S. companies, spanning three decades. The results show that this model produces distributions that underestimate reserve risk. Therefore, we propose two methods to increase the variability of the distribution so that it passes the back-test. In the first method, a set of benchmark systemic risk distributions are estimated by line of business that increase the variability of the bootstrapped distribution. In the second method, we show how one can apply a Wang Transform to estimate the systemic bias of the chain-ladder method over the course of the underwriting cycle.

TABLE OF CONTENTS

1. INTRODUCTION	2
2. SUMMARY OF EXISTING PAPERS	3
3. Data	4
4. THE METHOD BEING TESTED	5
5. BACK-TESTING	7
6. Analysis of the results.....	13
7. Possible adjustments to the method	15
8. Summary.....	24
Appendix A: The ODP bootstrap of the paid chain-ladder method that was tested	25
Appendix B Methodology for Reserving Back-Testing	33

1. INTRODUCTION

Reserve risk is one of the largest risks that non-life insurers face. A study done by A.M. Best (2010) identified deficient loss reserves as the most common cause of impairment¹ for the U.S. non-life industry in the last 41 years. It accounted for approximately 40% of impairments in that period. This, as well as encouragement from Solvency II regulation in Europe, has resulted in the growing popularity of the estimation of reserve distributions.

The over-dispersed Poisson (ODP) bootstrap of the chain-ladder method, as described in England and Verrall (2002) is one of the most popular methods used to obtain reserve distributions.² In the rest of this paper, we will simply refer to this as the “bootstrap model.”

Before relying on a method to estimate capital adequacy, it is important to know whether the method “works.” That is, is there really a 10% chance of falling above the method’s estimated 90th percentile? There have been many papers on different ways to estimate reserve risk, but very few papers on testing whether these methods work, and even fewer of these papers test the methods using real (as opposed to simulated) data.

In this paper, we test if the model works by back-testing the bootstrap model using real data spanning three decades on hundreds of companies. This paper differs from other papers on this topic to date because:

- We use real data, rather than simulated data.
- We test the reserve distribution in total – the sum of all future payments, not just the next calendar year’s payments.
- We test the distribution over many time periods — this is important due to the existence of the reserving cycle.
- We test multiple lines of business.

We suspect that other papers have not attempted this due to a lack of data of sufficient depth and breadth. In contrast, we have access to an extensive, cleaned U.S. annual statement database, as described in Section 3.

¹ A.M. Best defines an impairment to be when the regulator has intervened in an insurer’s business because they are concerned about its solvency.

² In 2007 a survey of the members of the Institute of Actuaries (U.K.), it was identified it as the most popular method.

2. SUMMARY OF EXISTING PAPERS

There have been a limited number of papers on the back-testing of reserve risk methods. Below is a summary of two such papers.

General Insurance Reserving Oversight Committee (2007) and (2008)

In 2007 and 2008, the General Insurance Reserving Oversight Committee, under the Institute of Actuaries in the U.K., published two papers detailing their testing of the Mack and the ODP bootstrap models. They tested these models with simulated data that complied with all the assumptions under each model. The ODP bootstrap model tested by the committee is the same as the bootstrap of the paid chain-ladder model being tested in this paper, as detailed in England and Verrall (2002). The results showed that even under these ideal conditions, the probabilities of extreme results could be under-stated using the Mack and the ODP bootstrap models. The simulated data exceeded the ODP bootstrap model's 99th percentile between 1% and 4% of the time. The 1% result was from more stable loss triangles. The simulated data exceeded the Mack model's 99th percentile between 2% and 8% of the time.

The Committee also tested a Bayesian model (as detailed in Meyers (2007)) with U.K. motor data (not simulated data). The test fitted the model on the data excluding the most recent diagonal, and the simulated distributions of the next diagonal are compared to the actual diagonal. The model allows for the error in parameter selection that can help overcome some of the underestimation of risk seen in the Mack and ODP bootstrap models. However, "it is no guarantee of correctly predicting the underlying distribution."

Meyers, G., Shi, P. (2011) "The Retrospective Testing of Stochastic Loss Reserve Models"

This paper back-tests the ODP bootstrap model as well as a hierarchical Bayesian model, using commercial auto liability data from U.S. annual statements for reserves as of December 2007. Two tests were performed. The first was to test the modeled distribution of each projected incremental loss for a single insurer. The second was to test the modeled distribution of the total reserve for many insurers, which is very similar to the test in this paper, however limited to only one time period (reserves as of December 2007). They conclude: "[T]here might be environmental changes that no single model can identify. If this continues to hold, the actuarial profession cannot rely solely on stochastic loss reserve models to manage its reserve risk. We need to develop other risk management strategies that do deal with unforeseen environmental changes."

3. DATA

The data used for the back-testing are from the research database developed by Risk Lighthouse LLC and Guy Carpenter. In the U.S., (re)insurers file annual statements each year as of December 31. The research database contains annual statement data from the 21 statement years 1989 to 2009.³ Within the annual statement is a Schedule P that, net of reinsurance, for each line of business, provides the following:

- A paid loss triangle by development year for the last 10 accident years
- Booked ultimate loss triangle by year of evaluation, for the last 10 accident years, showing how the booked ultimate loss has moved over time
- Earned premium for the last 10 accident years

Risk Lighthouse has cleaned the research database, by:

- Re-grouping all historical results to the current company grouping as of December 31, 2009 to account for the mergers and acquisitions activities over the past 31 years.
- Restating historical data (e.g., under new regulations a previous transaction does not meet the test of risk transfer and must be treated as deposit accounting).
- Cleaning obvious data errors such as reporting the number not in thousands but in real dollars.

For the purposes of our back-testing study, we refined our data as follows:

- We used company groups rather than individual companies since a subsidiary company cedes business to the parent company or sister companies and receives its percentage share of the pooled business.
- To ensure we had a reasonable quantity of data to apply the bootstrap model, we used up to 100 of the largest company groups for each line of business. For each line we began with the largest⁴ 100 company groups and removed those with experience that cannot be modeled due to size or consistency (some companies are missing random pieces of data). The resulting number of companies ranges from a high of 78 companies for Private Passenger Auto, to a low of 21 companies for Medical Professional Liability.⁵
- We used losses net of reinsurance, rather than gross. As a practical issue, the paid loss triangles are only reported net of reinsurance, and this also avoids the inter-company pooling

³ The database is updated annually

⁴ Size was determined by average premium from accident years 1989 to 2010

⁵ To clarify, there was an average of 78 companies per accident year for Private Passenger Auto and an average of 21 companies per accident year for Medical Professional Liability.

and possible double counting issue of studying gross data for company groups. Additionally, loss triangles gross of reinsurance can be created, but this data covers around half of the time span of the data net of reinsurance.

- We concentrated on the following lines of business:
 1. Homeowners (HO)
 2. Private Passenger Auto (PPA)
 3. Commercial Auto Liability (CAL)
 4. Workers Compensation (WC)
 5. Commercial Multi Peril (CMP)
 6. Medical Professional Liability — Occurrence and Claims Made (MPL)
 7. Other Liability — Occurrence and Claims Made (OL)

The Other Liability and Medical Professional Liability lines have only been split into Occurrence and Other Liability subsegments since 1993. Therefore, to maintain consistency pre- and post-1993, we have combined the two subsegments for these lines.

4. THE METHOD BEING TESTED

We are testing the reserve distribution created using the ODP bootstrap of the paid chain-ladder method, or simply the “bootstrap model” as described in Appendix 3 of England and Verrall (2002). We feel it is the most commonly used version of the model. How it specifically applies in our test is outlined below, and the steps with a numerical example are shown in Appendix A.

1. Take a paid loss and ALAE, 10 accident year by 10 development year triangle.
2. Calculate the all-year volume-weighted average age-to-age factors.
3. Estimate a fitted triangle by first taking the cumulative paid loss and ALAE to date from (1).
4. Estimate the fitted historical cumulative paid loss and ALAE by using (2) to undevelop (3).
5. Calculate the unscaled Pearson residuals, (from England and Verrall (1999)).

where

Δ
= incremental actual loss from step (1) and
= incremental fitted loss from step (4).

-
- Calculate the degrees of freedom and the scale parameter:

where

n = degrees of freedom
 N = number of incremental loss and ALAE data points in the triangle in step 1,
 and
 k = number of parameters in the paid chain-ladder model (in this case, 10
 accident year parameters and 9 development year parameters).

—.

- Adjust the unscaled Pearson residuals (— calculated in step 5:

—
— .

- Sample from the adjusted Pearson residuals — in step 7, with replacement.

- Calculate the triangle of sampled incremental loss — .

— .

- Using the sampled triangle created in 9, project the future paid loss and ALAE using the paid chain-ladder method.

- Include process variance by simulating each incremental future loss and ALAE from a Gamma distribution with the following:

mean = projected incremental loss in step 10, and
 variance = mean x scale parameter from step 6.

We assume that each future incremental loss is independent from each other. Note that theoretically we assume an over-dispersed Poisson distribution; however, we are using the Gamma distribution as a close approximation.

- Estimate the unpaid loss and ALAE by taking a sum of the future incremental losses from step (10).

Back-Testing Bootstrapping Method

-
- Repeat steps 8 to 12 (in our case, 10,000 times to produce 10,000 unpaid loss and ALAE estimates resulting in a distribution).

It is important to note that we are only testing the distribution of the loss and ALAE that is unpaid in the first 10 development years. This is to avoid the complications in modelling a tail factor.

5. BACK-TESTING

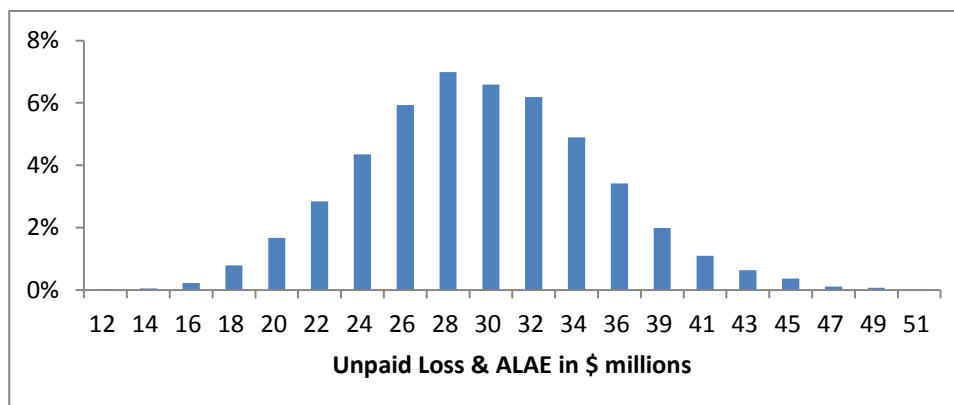
5.1 Back-testing as of December 2000

The steps in our back-testing are detailed below. First, we detail the steps for one insurer at one time period and then expand this to multiple insurers over many time periods.

- Create a distribution of the unpaid loss and ALAE by using the bootstrap model as of December 2000, as detailed in the prior section 4, using Schedule P paid loss and ALAE data for a particular company A's homeowners book of business.
- Isolate the distribution of unpaid loss and ALAE for the single accident year 2000, as shown in Figure 1. We do this so that we can test as many time periods as possible.

Figure 1

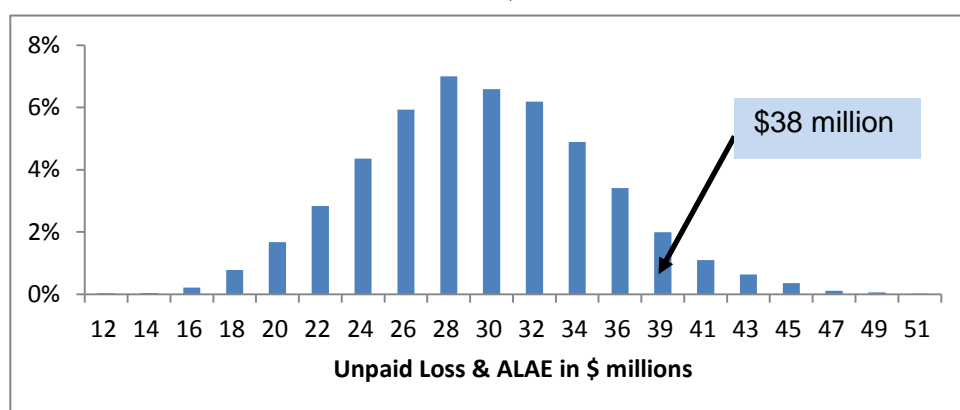
Company A's distribution of unpaid loss & ALAE, net of reinsurance as of 12/2000
Data in \$ millions



- The unpaid loss and ALAE is an estimate of the cost of future payments. Eventually, we will know how much the actual payments cost. In this case, the actual payments made total \$38 million⁶ (sum of the payments for accident year 2000 from development periods 24 to 120). We call this the “actual” unpaid – what the reserve should have been, with perfect hindsight. This falls at the 91st percentile of the original distribution.

⁶ We have scaled all the numbers shown in this example by the same factor, to disguise the company.

Figure 2
Company A's distribution of unpaid loss & ALAE, net of reinsurance as of 12/2000
Data in \$ millions



4. We can repeat steps 1 to 3 for another 74 companies. Some of the percentiles for these companies are listed in Figure 3.

Figure 3
Percentile where the actual unpaid falls in the distribution created as of 12/2000, by company.

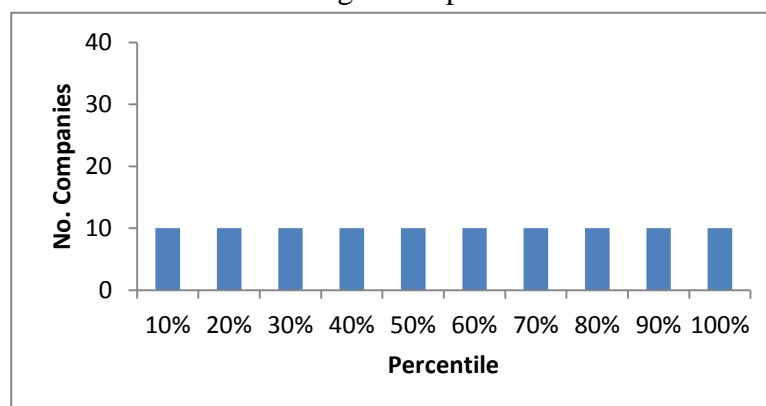
Company	Percentile
Company A	91%
Company B	55%
Company C	88%
Company D	92%
Company E	39%
Company F	75%
Company G	67%
...	...

Results

If the bootstrap model gives an accurate indication of the probability of the actual outcome, we should find a uniform distribution of these 75 percentiles. For example, the 90th percentile is a number that the insurer expects to exceed 10% of the time (that is the definition of the 90th percentile). Therefore, we should find 10% of the companies have an actual outcome that falls above the 90th percentile. And similarly, there should be a 10% chance that the actual reserves fall in the 80th to 90th percentile and so on. That is, ideally we should see Figure 4 when we plot these percentiles.

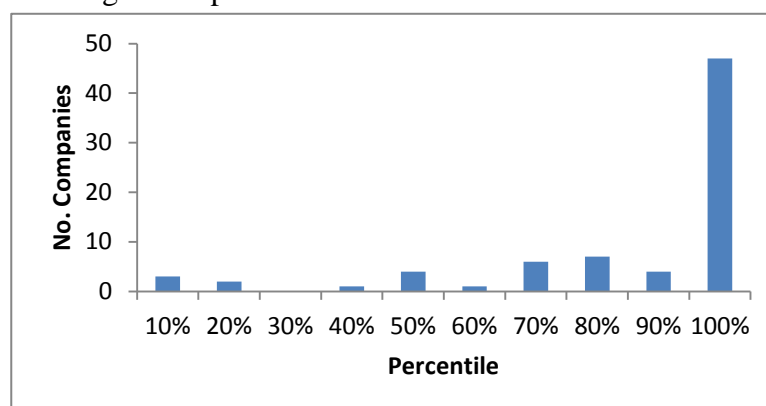
Back-Testing Bootstrapping Method

Figure 4
Ideal histogram of percentiles



When we plot the percentiles in Figure 3, what we actually see is shown in Figure 5:

Figure 5
Histogram of percentiles for Homeowners as of 12/2000



This shows that 46 out of 75 companies had actual reserves that fell above the 90th percentile of the original distributions created in 12/2000. For 46 out of 75 companies, the reserve was much higher than they initially expected.

5.2 Back-Testing as of December 1996

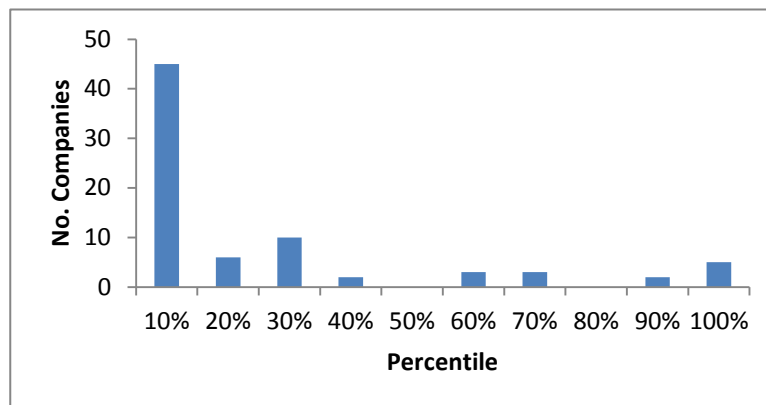
The test can be repeated at another time period — instead of December 2000, we can try December 1996. That is, we repeat steps 1 to 4, but this time, we are creating reserve distributions for 76 companies as of December 1996.

Back-Testing Bootstrapping Method

Results

The histogram of the resulting percentiles is shown in Figure 6.

Figure 6
Histogram of percentiles for Homeowners as of 12/1996



In this case, 45 out of 76 companies had actual reserves that fell below the 10th percentile of the original distributions created in 12/1996. For 45 out of 76 companies, the reserve was much lower than they initially expected. This is the opposite of the result seen in 12/2000.

These results, where most insurers are either under- or over-reserved at each point in time, are perhaps not surprising. At any one point in time, the reserve is estimated with what is currently known. As the future unfolds and the claims are actually paid, some systemic effect can cause the claims environment to move away from the historical experience, causing most insurers to be either under- or over-reserved. For example, at the time of writing (2012) inflation has been historically low, and actuaries set their current reserves in this environment. If, as the future unfolds, claims inflation increases unexpectedly, then the reserves for most insurers will be deficient, similar to what is seen in Figure 5.

Therefore, testing one time period at a time may not result in a uniform distribution of percentiles. However, if many time periods are tested, and all the percentiles are plotted in one histogram, this may result in a uniform distribution.

5.3 Back-Testing Multiple Periods

For this test, repeat steps 1 to 4 in section 5.1, each time estimating the reserve distribution as of 12/1989, 12/1990, 12/1991... to 12/2002. Note that at the time of testing, we only had access to data as of 12/2009. The actual unpaid for accident year 2002 should be the sum of the payments for that accident year from the 24th to the 120th development period. However, as of 12/2009 we only

Back-Testing Bootstrapping Method

have payments from the 24th to the 96th development period. The remaining payments had to be estimated. A similar issue exists for the test as of 12/2001. We use the average of company's accident year 1998 to 2000, 96th to 120th development factors to estimate the remaining payments.

Results

The test for an average of 74 companies for 14 accident years results in 1,038 percentiles, shown in a histogram in Figure 7.

Figure 7

Histogram of percentiles for Homeowners as of 12/1989, 12/1990... and 12/2002

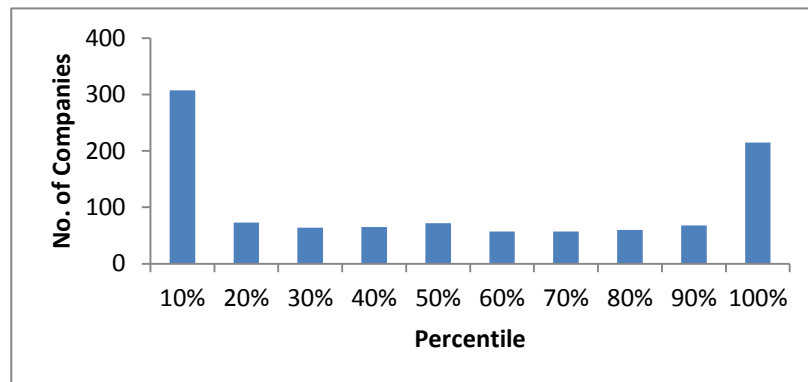


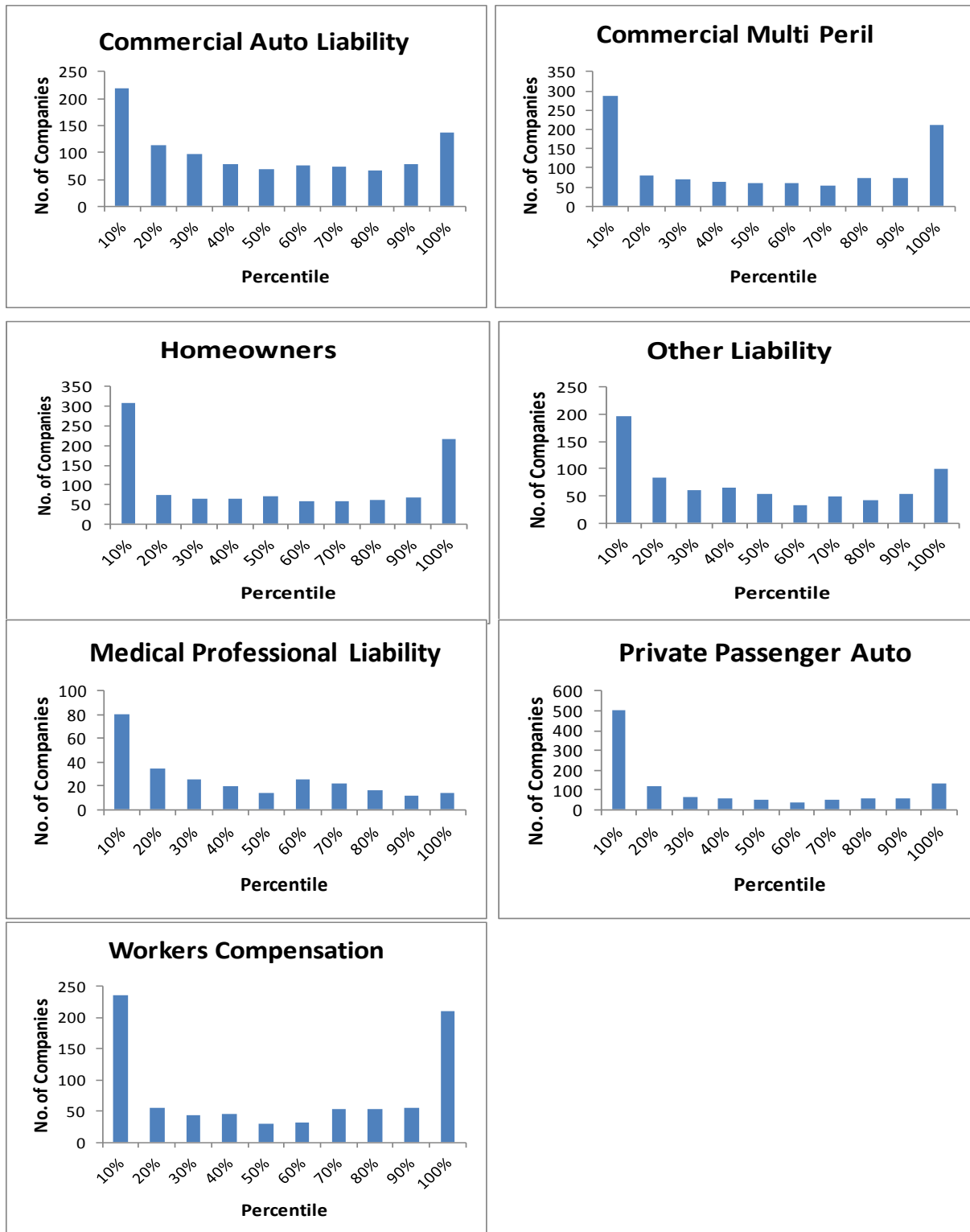
Figure 7 shows that, around 20% of the time, the actual reserve is above the 90th percentile of the bootstrap distribution, and 30% of the time the actual reserve is below the 10th percentile of the distribution. When you tell management the 90th percentile of your reserves, this is a number they expect to be above 10% of the time. Instead, when using the bootstrap model, we find that companies have exceeded the modeled 90th percentile, 20% of the time. In this test, the bootstrap model appears to be underestimating reserve risk.

5.4 Back-Testing of Other Lines of Business

The test can be repeated for the other lines of business. When this is done, the results are shown in Figure 8.

Figure 8

Histogram of percentiles as of 12/1989, 12/1990... and 12/2002



Back-Testing Bootstrapping Method

The histograms above do not follow a uniform distribution. In this test, for most of these lines of business, using the bootstrap model has produced distributions that underestimate reserve risk. However, for medical professional liability and private passenger auto in particular, it appears that the paid chain-ladder method is producing reserve estimates that are biased high.

6. ANALYSIS OF THE RESULTS

A reserve distribution is a measure of how the actual unpaid loss may deviate from the best estimate. By applying the ODP bootstrap to the paid chain-ladder method, we can get such a reserve distribution around a paid chain-ladder best estimate.

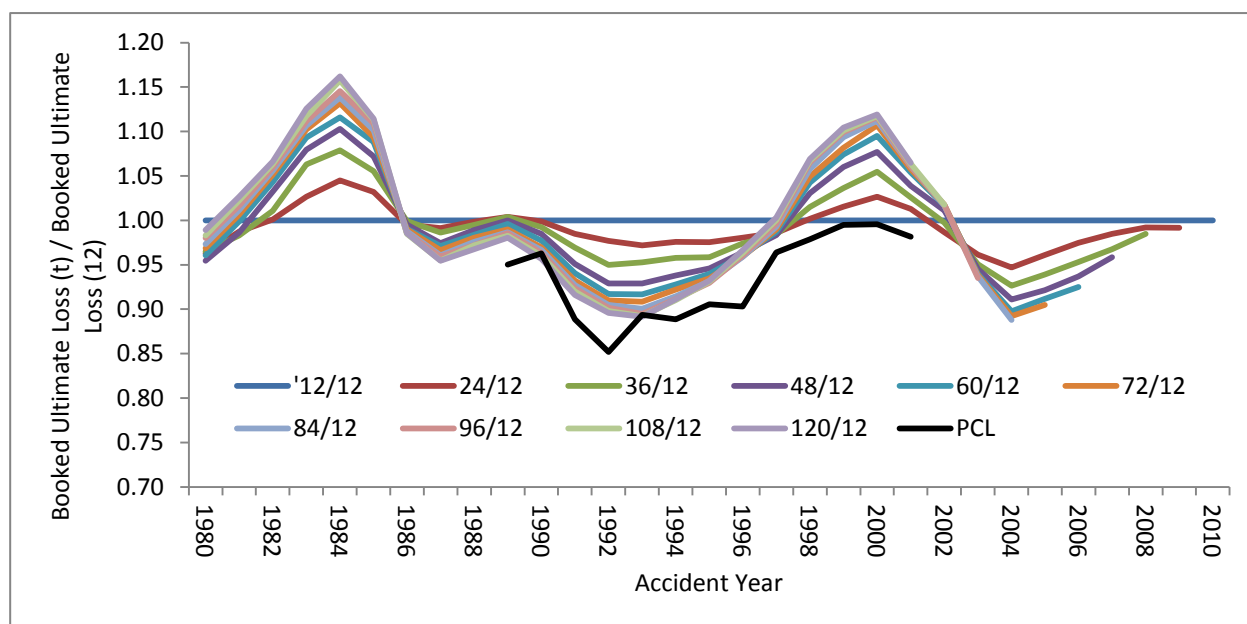
However, it is rare to rely solely on this method to determine an actuarial central reserve estimate. For better or worse, it is common practice for actuaries to estimate a distribution by using a similar ODP bootstrap of the chain-ladder method outlined here, and then “shift” this distribution by multiplication so that the mean is the same as an actuarial best estimate reserve or booked reserve.

We do not condone this practice, but it is so common that a natural question is whether the “shifting” of the distributions produced in our back-testing would result in a more uniform distribution. Booked reserve estimates use more sophisticated methods than the paid chain-ladder model, and therefore may be more accurate, so the width of the distributions being produced in this study may be perfectly suitable. If we used the booked reserve then we may not have seen the under- and over-reserving in the December 2000 and December 1996 results, respectively

In reality, the industry does under- and over-reserve, sometimes significantly. In comparison to the paid chain-ladder method, the industry is sometimes better or worse at estimating the true unpaid loss. In Figure 9, we show the booked ultimate loss at 12, 24, 36... and 120 months of evaluation, divided by the booked ultimate loss at the 12-month evaluation, for the U.S. industry, in aggregate for the seven lines of business tested in this paper.

The “PCL” line on this graph shows the cumulative paid loss and ALAE at 120 months of evaluation divided by the PCL estimate at 12 months of evaluation (using an all year weighted average on an industry 10 accident year triangle by line of business and excluding a tail factor).

Figure 9
Booked ultimate loss at t months of evaluation / Booked ultimate loss at 12 months of evaluation, in aggregate for the industry for seven lines of business, net of reinsurance



For example, for accident year 2000, the booked ultimate loss estimate as of 12/2009 ended up 12% higher than the initial booked ultimate loss as of 12/2000. In contrast, the paid chain-ladder estimate of the ultimate loss as of 12/2009 was the same as the estimate as of 12/2000 – that is, the paid chain-ladder reserve estimate as of 12/2000 was more accurate than the booked reserve.⁷

Shifting a distribution around another mean is not a sound practice. Even ignoring this, we did not feel that shifting the mean to equal the booked reserve at the time would have materially changed the broad result of our back-testing.

6.1 Why are we seeing these results?

In an Institute of Actuaries of Australia report titled, “A Framework for Assessing Risk Margins,”⁸ the sources of uncertainty in a reserve estimate are grouped into two parts: independent risk and systemic risk.

1. Independent risk = “risks arising due to the randomness inherent in the insurance process.”

⁷ Note that the comparison is not exact – the “PCL” line excludes all payments after 120 months whereas the booked reserves include those payments, which must add to the difficulty of estimation. Additionally, the booked reserve is estimated at a company level whereas the “PCL” line was estimated using a whole industry loss triangle by line of business.

⁸ IAAust Risk Margins Task Force (2008).

-
2. Systemic risk = a risk that affects a whole system. “Risks that are potentially common across valuation classes or claim groups.”⁸ Even if the model is accurately reflecting the claims process today, future trends in claims experience may move systematically away from what was experienced in the past. For example, unexpected changes in inflation, unexpected tort reform or unexpected changes in legislation.

They state that the traditional quantitative techniques, such as bootstrapping, are better at analyzing independent risk but aren’t able to adequately capture systemic risk. This is because, even if there are past systemic episodes in the data, “a good stochastic model will fit the past data well and, in doing so, fit away most past systemic episodes of risk...leaving behind largely random sources of uncertainty.”⁸

The distributions produced in this paper using the bootstrap model may be underestimating reserve risk because they only capture independent risk, not systemic risk.

7. POSSIBLE ADJUSTMENTS TO THE METHOD

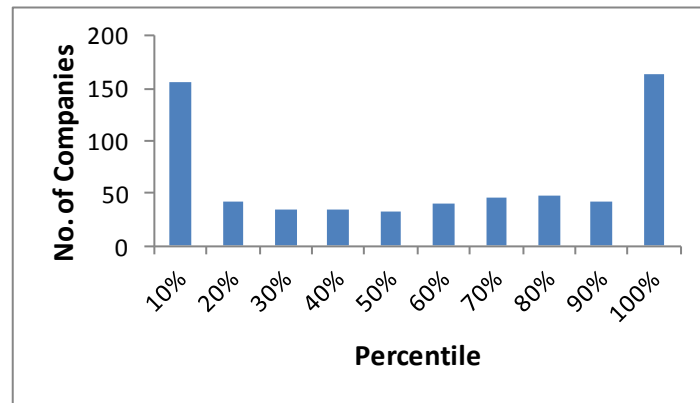
If the ODP bootstrap of the paid chain-ladder method is producing distributions that underestimate the true reserve risk, then what adjustments can be made so that it more accurately captures the risk?

7.1 *Commonly used possible adjustments*

Bootstrapping the incurred chain-ladder method

Using the method outlined in this paper on incurred loss and ALAE data instead of paid loss and ALAE data produces reserve distributions with a smaller variance. This is understandable if you assume that the case reserves provide additional information about the true cost of future payments. We back-tested the incurred bootstrap model for the Workers Compensation line of business. The process was the same as for the paid bootstrap model, but incurred loss and ALAE was substituted for the paid loss and ALAE data, resulting in distributions of IBNR. The resulting percentiles are in Figure 10.

Figure 10
Histogram of percentiles for Workers Compensation as of 12/1989, 12/1990..... 12/2002 based on incurred loss & ALAE



From Figure 10, it appears that the incurred bootstrap model is also underestimating the risk of falling in these extreme percentiles.

Using more historical years of paid loss and ALAE data

Our back-testing used paid loss and ALAE triangles with 10 historical accident and development years. A loss triangle with more historical accident years may result in a wider distribution.

We applied the bootstrap model on 20 accident year x 10 development year homeowners paid loss and ALAE triangles, but this resulted in distributions that sometimes had more and sometimes had less variability than the original distributions from the 10 x 10-year datasets.

Making Other Adjustments

There are other additions to the bootstrap model that we have not considered here, and can be areas of further study:

- Parametric bootstrapping. It is possible to simulate the accident year and development year parameters from a multivariate normal distribution using a generalized linear model, which closely follows the structure of the ODP bootstrap of the paid chain-ladder model. This is commonly called parametric bootstrapping. Re-sampling the residuals, as outlined in section 3, may be limiting, and simulating from a normal distribution may result in wider reserve distributions.
- Hat matrix. The hat matrix can be applied to standardize the Pearson residuals and make them identically distributed, as per Pinheiro, Andrade e Silva, and Centeno (2003).
- Multiple scale parameters to account for heteroskedasticity in residuals.

Back-Testing Bootstrapping Method

None of these commonly used adjustments are specifically designed to account for systemic risk. Also, the method we are testing significantly underestimates the true risk and so requires an adjustment that significantly widens the distribution. Therefore, we outline below two methods that can be applied to the bootstrap model that account for systemic risk.

7.2 Two methods to account for systemic risk

We have derived two methods to adjust the ODP bootstrap of the paid chain-ladder model so that the resulting distributions of unpaid loss and ALAE describe the true reserve risk – that is, it passes our back-testing, so that, for example, 10% of the time the actual reserve falls above the 90th percentile of our estimated distribution. These two methods are explicitly attempting to model systemic risk – the risk that the future claims environment is different from the past.

The two methods are:

1. The systemic risk distribution method
2. Wang transform adjustment

The two methods are based on two different assumptions. The systemic risk distribution method does not adjust the actuary's central estimate over the reserving cycle. The Wang transform adjustment does not assume that the central estimate reserve is unbiased and tries to estimate the systemic bias of the chain-ladder method over the course of reserving cycle.

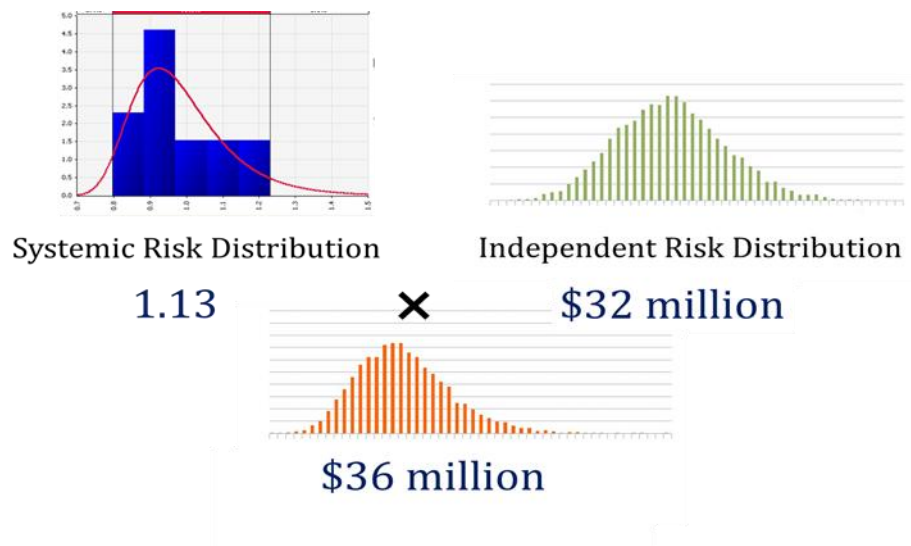
Both methods are applied and the resulting distribution is back-tested again.

Systemic Risk Distribution Method

As outlined in section 6.1, reserve risk can be broken down into two parts: (1) independent risk and (2) systemic risk. We believe that the bootstrap model only measures independent risk, not systemic risk. Systemic risk affects a whole system, like the market of insurers. It includes risks such as unexpected changes in inflation and unexpected changes in tort reform – in short, the risk that the future claims environment could be different from the past.

In this method, we estimate a benchmark systemic risk distribution by line of business, and combine this with the independent risk distribution (from the bootstrap model) to obtain the total reserve risk distribution. To combine the distributions we assume that they are independent from each other, and take one simulation from the systemic risk distribution and multiply this by one simulation from the independent risk distribution, and repeat for all 10,000 (or more) iterations.

Figure 11
Example of one iteration of the systemic risk distribution adjustment



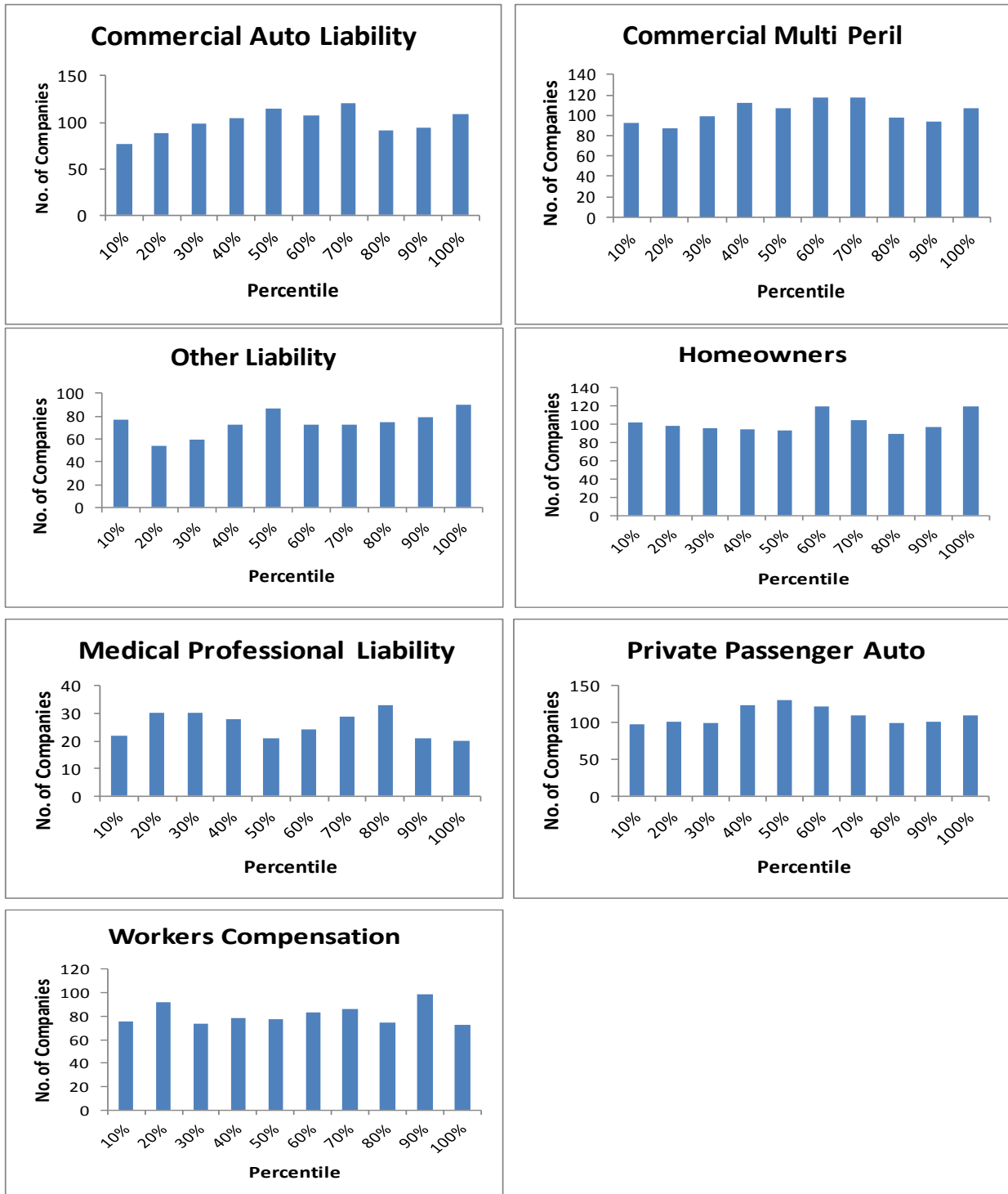
Each of the outcomes from the systemic risk distribution, such as the 1.13 in Figure 11, can be thought of as a systemic risk factor. We can calculate historical systemic risk factors for each year and for each company by the following procedure:

1. Take the mean of the bootstrap model's reserve distribution for each company, as of 12/1989 for accident year 1989.
2. Calculate what the reserve should have been for accident year 1989, back in 12/1989 (= the ultimate loss and ALAE as of 12/2008 less the paid as of 12/1989).
3. The systemic risk factor = (2)/(1).
4. Repeat 1 to 3 as of 12/1990, 12/1991,to 12/2002.

The systemic risk benchmark distribution is estimated by fitting a distribution to the historical systemic risk factors. Through a curve-fitting exercise, we found that a Gamma distribution was the best candidate to model systemic risk, with differing parameters by line of business. This results in uniform distributions of percentiles when the method outlined above is back-tested, as shown in Figure 12.

Admittedly, the fitted systemic risk distribution may differ depending on the back-testing period. However, we took care to span a time period that incorporated one upwards and one downwards period of the reserve cycle, in an attempt to not bias the results.

Figure 12
Histogram of Percentiles as of 12/1989, 12/1990... to 12/2002



As an example, for Homeowners, the systemic risk distribution is a gamma distribution with a mean of 0.98 and a standard deviation of 19%. Note that this is intended to adjust the distribution of the reserve for a single accident year at 12 months of evaluation.

Systemic risk has its roots in the reserving cycle and underwriting cycle. Indeed, there are documented evidences of the linkage between reserving cycle and underwriting cycle. Archer-Lock et al. (2003) discussed reserving cycle in the U.K. The authors conclude that the mechanical application of traditional actuarial reserving methods may be one of the causes for reserve cycle. In particular, the underwriting cycle may distort claims development patterns and that premium rates indices may understate the magnitude of the cycle. The Wang transform adjustment is an alternative systemic risk adjustment that explicitly accounts for the reserve cycle.

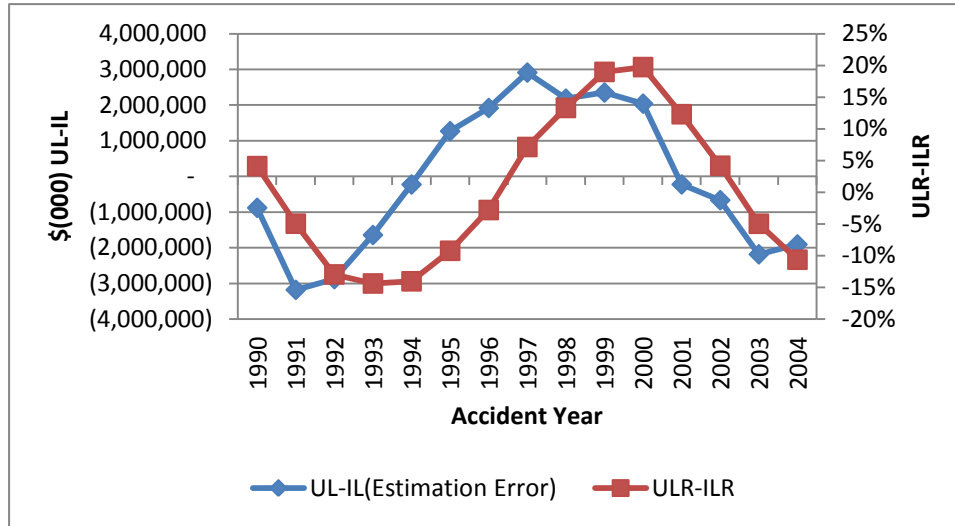
Wang-Transform Adjustment

The Wang transform adjustment method does not assume that the unpaid loss and ALAE estimate is unbiased and tries to estimate the systemic bias over the course of reserving cycle. The reserving cycle shown in Figure 13 is an interesting phenomenon that indicates that reserve risk is cyclical, and the Wang transform adjustment method tries to capture this “systemic” bias.

We use the workers compensation line of business for illustration. The data used in back-testing is the workers compensation aggregated industry data net incurred loss and ALAE. A series of back-tests (these back-tests are different from the test in section 4) are done using the chain-ladder method for industry loss reserve development. For accident years after 2001, we use the latest reported losses instead of the projected ultimate losses, since reported losses for those accident years are not yet fully developed.

In the following figures, which represent the entire non-life industry workers compensation line of business, Ultimate Losses (UL) stand for the 120-month incurred loss and ALAE for each accident year (AY) from the latest report year (RY), not including IBNR. Initial Losses (IL) represent the projected 120-month net incurred loss and ALAE for each AY from first report year, using the chain-ladder reserving method. Ultimate Loss Ratio (ULR) represents the ultimate reported incurred loss and ALAE ratio for each AY and Initial Loss Ratio (ILR) represents the initial reported incurred loss ratio for each AY. A more detailed explanation of the back-testing method is given in Appendix B.

Figure 13
Chain-Ladder Method Back-Testing Error versus (ULR-ILR) for Workers Compensation



For the workers compensation line of business, the chain-ladder reserving method has systematic errors that are highly correlated with the reserving cycle. The contemporary correlation between the estimation error and the reserve development (ULR-ILR) is 0.64 for the chain-ladder method. More noticeably, the one-year lag correlation is 0.91. The estimation error leads the loss reserve development by one-year.

In this study, we apply the Wang transform to enhance the loss reserve distribution created using the bootstrap of the chain-ladder method. Different from the systemic risk distribution method, the Wang transform adjustment method will first adjust the variability of the loss reserve then give each company group's loss reserve distribution a shift, respectively.

The procedures below describe how to apply the Wang-Transform adjustment:

1. After bootstrapping each paid loss triangle with 10,000 iterations, we apply the ratio of double exponential over normal to adjust the chain-ladder reserve distribution to be wider than the original distribution. The formula is shown below:

Where:

- 1) ϕ is a normal distribution with mean 0 and standard deviation 1.
- 2) Exponential stands for a double exponential distribution with density function
- 3) q is the quantile of each simulated reserve.

- 4) u is the median of 10,000 simulated reserves.
- 5) x is the simulated reserve.
- 6) x^* is the reserve after adjustment.

2. β is calculated for each company to measure the correlation between the company and industry. The method to calculate the β is described below:

For each AY, if the industry systemic risk factor is significantly different from one (greater than 1.01 or less than 0.99), we use an indicator of +1 if a company's corresponding AY systemic risk ratio is in the same direction as the industry systemic risk factor, otherwise we use an indicator of -1. For example, say the industry systemic risk factor for AY 2000 is 1.122, which is greater than 1. If company A's AY 2000 systemic risk factor is greater than 1, we assign an indicator of +1 to company A AY 2000. If company B's AY 2000 systemic risk factor is less than 1, we assign an indicator of -1 to company B AY 2000.

At last, for a company,

$$\beta = (\text{sum of indicators of all AYs}) / (\text{count of indicators of all AYs}).$$

If β is less than zero, we force it to zero.

3. The Wang transform is finally applied to adjust the mean of the reserve distribution. The formula is shown below:

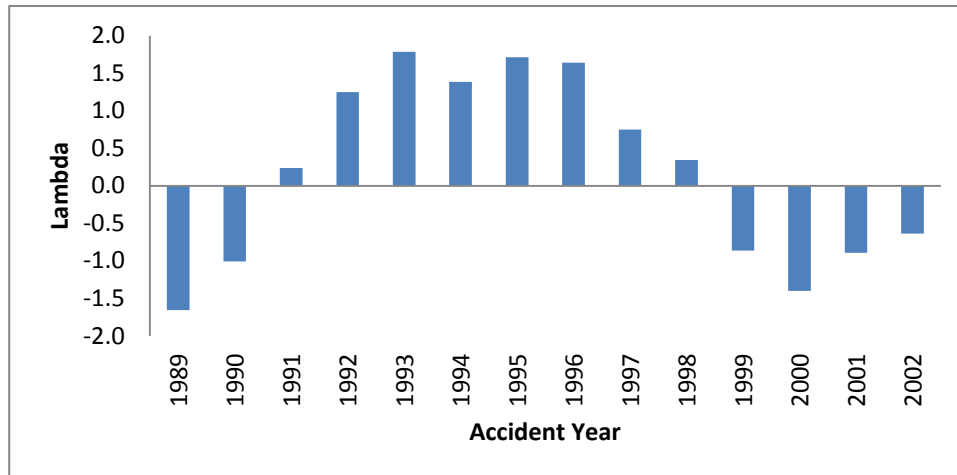
Where:

- 1) λ is reported reserve's percentile in the reserve distribution after step 1 adjustment.
- 2) Each company has its own β value.
- 3) ϕ is a normal distribution with mean 0 and stdev 1.

We change λ to result in the most uniformly distributed percentiles as measured by a chi-square test, when the adjusted reserve is back-tested. λ here is the fitted shift value of loss reserve distribution for each accident year.

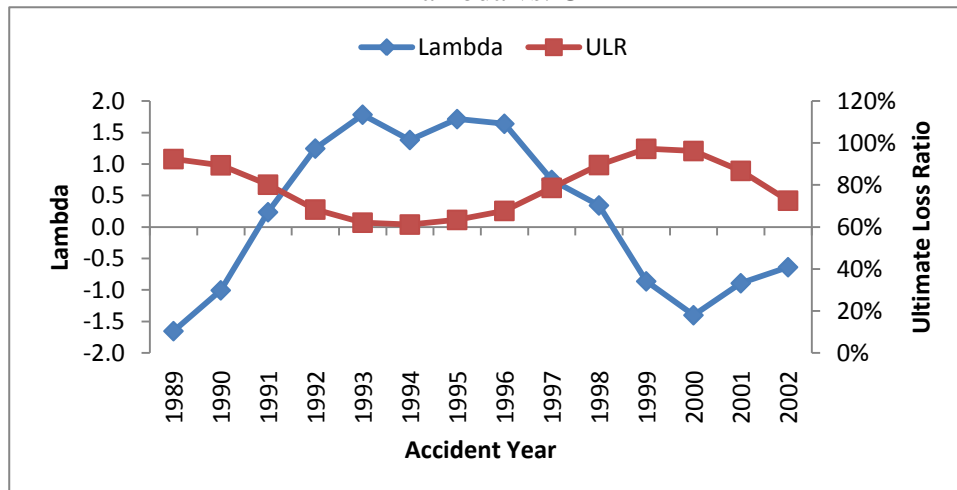
The final λ s are shown in the following figures.

Figure 14
Lambda for Each Accident Year



Comparing the fitted lambdas and the ultimate loss ratios (ULRs), we find that the loss reserve estimation error of the chain-ladder method is highly correlated with the non-life insurance market cycle. The correlation between lambda and ultimate loss ratio is -0.87.

Figure 15
Lambda vs. ULR



For other lines of business, the estimated lambdas and ULRs by year are also negatively correlated, but the magnitude of correlation is not as strong as for workers compensation.

Areas of Future Research

In this paper we have demonstrated that different lambda values can be estimated at various stages of the reserve cycle. However, more research is needed in the future to illustrate the practical

Back-Testing Bootstrapping Method

application of lambda in the Wang transform to the reserve cycle, and how well lambda along with the beta work for individual companies versus the industry as a whole.

We have also focused mostly on the paid loss development method. Another area of future research is to investigate the difference between using paid methods in isolation versus paid loss development methods in conjunction with case incurred loss development methods, exposure-based methods, and judgment.

Yet another area of future research is to investigate the reserve cycle illustrated in Figure 9.

8. SUMMARY

The genesis and popularity of a reserve risk method lies in its theoretical beauty. However, as more insurers rely on actuarial estimates of reserve risk to manage capital, estimating a reserve distribution is no longer a purely theoretical exercise. Methods should be tested against real data⁹ before they can be relied upon to support the insurance industry's solvency.

In this back-test, we see that the popular ODP bootstrap of the paid chain-ladder method is underestimating reserve risk. We believe that it is because the bootstrap model does not consider systemic risk, or, to put it another way, the risk that future trends in the claims environment — such as inflation, trends in tort reform, legislative changes, etc. — may deviate from what we saw in the past. We suggest two simple solutions to incorporate systemic risk into the reserve distribution so that the adjusted reserve distribution passes the back-test.

We hope to encourage more testing of models so that the profession has a more defensible framework for measuring risks for solvency and profitability.

⁹ We advocate the use of real versus simulated data. In this paper, we have found that the bootstrap model, for most lines, materially underestimates the probability of falling above the 90th percentile. In contrast, the same model when tested against simulated data in GIRO, found that the same bootstrap model identified a 99th percentile that was exceeded only 1% to 4% of the time by the simulated data.

APPENDIX A: THE ODP BOOTSTRAP OF THE PAID CHAIN-LADDER METHOD THAT WAS TESTED

We are testing the reserve distribution created using the bootstrap of the paid chain-ladder method. The model we are using is described in Appendix 3 of England and Verrall (2002). We detail this, with a numerical example, below:

1. Take a paid loss and ALAE, 10 accident year by 10 development year triangle

Figure A1

Company A, paid Loss & ALAE, net of reinsurance as of 12/2000
Data in \$ millions

AY	1	2	3	4	5	6	7	8	9	10
1991	94	119	124	128	130	132	133	133	133	134
1992	101	131	135	139	141	143	143	144	145	
1993	82	107	112	116	119	119	120	121		
1994	110	139	146	152	154	155	156			
1995	68	99	105	108	111	114				
1996	119	151	157	158	162					
1997	72	99	99	99						
1998	71	101	106							
1999	71	96								
2000	62									

2. Calculate the all-year, volume-weighted age-to-age factors.

Figure A2

1 - 2	2 - 3	3 - 4	4 - 5	5 - 6	6 - 7	7 - 8	8 - 9	9 - 10
1.32	1.04	1.02	1.02	1.01	1.00	1.00	1.00	1.00

3. Estimate a fitted triangle by taking the cumulative paid loss and ALAE to date from (1).

Figure A3

Company A, paid Loss & ALAE to date, net of reinsurance as of 12/2000
Data in \$ millions

AY	1	2	3	4	5	6	7	8	9	10
1991										134
1992									144	
1993								121		
1994							156			
1995						114				
1996					162					
1997				99						
1998			106							
1999		96								
2000	62									

4. Estimate the fitted historical cumulative paid loss and ALAE by using (2) to un-develop (3).

Figure A4

Company A, paid Loss & ALAE, net of reinsurance as of 12/2000
Data in \$ millions

AY	1	2	3	4	5	6	7	8	9	10
1991	90	120	125	128	130	132	132	133	133	134
1992	98	129	135	138	141	143	143	144	144	
1993	82	108	113	116	118	120	120	121		
1994	107	141	147	150	153	155	156			
1995	78	103	108	110	112	114				
1996	112	149	155	159	162					
1997	70	93	97	99						
1998	77	102	106							
1999	72	96								
2000	62									

$= 96 / 1.32$

5. Calculate the unscaled Pearson residuals, (from England and Verrall 1999).

==,

where

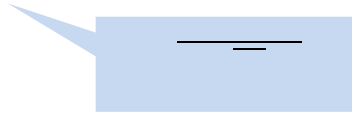
= incremental actual loss from step (1) and

= incremental fitted loss from step (4).

Figure A5

Company A, unscaled residuals, net of reinsurance as of 12/2000

AY	1	2	3	4	5	6	7	8	9	10
1991	0.35	(0.76)	0.13	0.55	(0.38)	0.14	0.44	(0.37)	(0.06)	
1992	0.35	(0.39)	(0.39)	0.12	(0.42)	0.32	(0.61)	0.05	0.05	
1993	(0.03)	(0.18)	0.07	0.55	0.73	(0.92)	0.07	0.33		
1994	0.32	(0.89)	0.52	1.21	(0.64)	(0.43)	0.12			
1995	(1.14)	1.12	1.03	0.24	0.51	0.94				
1996	0.65	(0.79)	(0.08)	(1.27)	0.31					
1997	0.16	1.00	(1.92)	(1.49)						
1998	(0.67)	0.97	0.51							
1999	(0.23)	0.40								
2000										



6. Calculate the degrees of freedom and the scale parameter:

where

= degrees of freedom,

= number of incremental loss and ALAE data points in the triangle in step 1, and

= number of parameters in the paid chain-ladder model (in this case, 10 accident year parameters and 9 development year parameters).

—

—

7. Adjust the unscaled Pearson residuals (calculated in step 5:

—
—

—
—

Figure A6

Company A, unscaled residuals, net of reinsurance as of 12/2000

AY	1	2	3	4	5	6	7	8	9	10
1991	0.43	(0.94)	0.17	0.68	(0.47)	0.17	0.55	(0.45)	(0.07)	
1992	0.43	(0.48)	(0.48)	0.15	(0.52)	0.39	(0.76)	0.07	0.07	
1993	(0.04)	(0.22)	0.09	0.68	0.90	(1.14)	0.09	0.40		
1994	0.40	(1.09)	0.64	1.49	(0.79)	(0.53)	0.15			
1995	(1.41)	1.38	1.27	0.30	0.63	1.17				
1996	0.80	(0.98)	(0.10)	(1.58)	0.39					
1997	0.19	1.23	(2.38)	(1.85)						
1998	(0.83)	1.20	0.63							
1999	(0.28)	0.50								
2000										

8. Sample from the adjusted Pearson residuals in step 7, with replacement.

Figure A7

Company A, Sampled adjusted residuals, net of reinsurance as of 12/2000

AY	1	2	3	4	5	6	7	8	9	10
1991	(1.09)	(0.28)	(0.04)	0.07	0.17	(0.07)	1.20	0.63	1.27	(0.28)
1992	(0.28)	0.43	0.39	0.80	(0.53)	1.17	0.19	(0.98)	(0.48)	
1993	0.09	0.39	(0.94)	1.20	(0.83)	(0.53)	0.63	(0.76)		
1994	0.68	(2.38)	1.23	(0.98)	0.30	0.43	(1.14)			
1995	(0.47)	(0.94)	0.50	(0.48)	0.15	(0.94)				
1996	0.63	0.15	(1.14)	1.49	1.49					
1997	0.07	0.07	0.17	(2.38)						
1998	0.15	0.63	0.39							
1999	1.20	0.50								
2000	0.50									


9. Calculate the triangle of sampled incremental loss, $\hat{\Delta}$.

—

Figure A8

Company A, sampled incremental paid Loss & ALAE, net of reinsurance as of 12/2000
Data in \$ millions

AY	1	2	3	4	5	6	7	8	9	10
1991	80	28	5	3	3	2	1	1	1	0
1992	95	34	6	5	2	4	1	(0)	0	
1993	83	28	3	5	1	1	1	(0)		
1994	114	20	9	2	3	3	(0)			
1995	74	20	5	2	2	0				
1996	119	37	3	6	6					
1997	71	23	4	(1)						
1998	78	28	5							
1999	83	26								
2000	65									



10. Project the future paid loss and ALAE, using the paid chain-ladder method.

Figure A9

Company A, cumulative paid Loss & ALAE, net of reinsurance as of 12/2000
Data in \$ millions

AY	1	2	3	4	5	6	7	8	9	10
1991	80	108	113	116	118	120	121	122	124	124
1992	95	129	136	140	142	146	146	146	146	147
1993	83	111	114	118	119	120	121	121	122	122
1994	114	134	143	145	148	151	151	151	152	152
1995	74	94	100	102	104	104	105	105	106	106
1996	119	156	160	166	172	175	175	176	177	177
1997	71	94	98	97	99	101	101	101	102	102
1998	78	106	112	114	117	118	119	119	120	120
1999	83	108	113	116	118	120	121	121	122	122
2000	65	86	89	92	93	95	95	96	96	96

Weighted average age-to-age factors

2	2 - 3	3 - 4	4 - 5	5 - 6	6 - 7	7 - 8	8 - 9	9 - 10
1.31	1.04	1.02	1.02	1.02	1.00	1.00	1.01	1.00

11. Include process variance by simulating each incremental future loss and ALAE from a Gamma distribution with:

mean = projected incremental losses in step 10,

variance = mean x scale parameter from step 6.

We assume that each future incremental loss is independent from each other. Note that theoretically we assume an over-dispersed Poisson distribution, however, we are using the Gamma distribution as a close approximation.

Figure A10

Company A, incremental paid Loss & ALAE, with process variance, net of reinsurance as of 12/2000

Data in \$ millions

AY	1	2	3	4	5	6	7	8	9	10
1991	82	28	5	3	3	2	1	1	1	0
1992	96	34	6	4	2	3	1	0	0	0
1993	83	28	3	4	1	1	1	0	1	0
1994	112	23	8	2	3	3	(0)	0	2	1
1995	75	21	5	2	2	1	0	1	1	0
1996	118	37	4	6	5	4	2	0	0	0
1997	71	23	4	(1)	2	4	0	1	1	1
1998	78	27	5	4	2	0	1	0	0	0
1999	81	25	3	2	1	1	0	0	0	1
2000	65	18	5	2	1	1	1	0	0	0

12. Estimate the unpaid loss and ALAE by taking a sum of the future incremental losses from step 11.

Figure 11

Accident year	Unpaid Loss and ALAE
1991	0
1992	0
1993	1
1994	3
1995	2
1996	6
1997	9
1998	7
1999	8
2000	28
Total	64

-
13. Repeat steps 8 to 12. In our case, 10,000 times to produce 10,000 unpaid loss and ALAE estimates resulting in an unpaid loss and ALAE distribution when plotted in a histogram.

APPENDIX B METHODOLOGY FOR RESERVING BACK-TESTING

The method used in back-testing in the Wang transform adjustment uses 10 accident year by 10 development year incurred loss and ALAE triangles, net of reinsurance.

AY	NetCaseIncurred									
2001	10,597,343	15,121,875	17,148,317	18,030,720	18,315,025	18,663,494	18,963,470	19,274,423	19,492,073	19,691,532
2002	9,880,312	14,375,596	16,125,066	16,778,143	17,329,085	17,651,826	17,996,103	18,251,653	18,518,021	
2003	10,369,458	14,334,644	15,837,509	16,594,323	17,191,600	17,714,263	17,992,600	18,339,834		
2004	10,616,734	14,461,802	15,894,418	16,644,316	17,253,999	17,678,604	18,059,807			
2005	11,234,028	14,723,765	16,219,677	17,166,414	17,739,780	18,273,742				
2006	11,762,138	15,866,737	17,804,050	18,942,686	19,777,903					
2007	12,189,806	16,932,637	19,062,769	20,387,578						
2008	12,394,686	17,351,056	19,656,367							
2009	11,230,304	15,726,363								
2010	11,564,142									
2000	10,300,006	15,323,250	17,359,275	18,504,064	19,310,172	19,727,612	20,024,743	20,349,723	20,416,586	20,626,383
2001	10,597,343	15,121,875	17,148,317	18,030,720	18,315,025	18,663,494	18,963,470	19,274,423	19,492,073	
2002	9,880,312	14,375,596	16,125,066	16,778,143	17,329,085	17,651,826	17,996,103	18,251,653		
2003	10,369,458	14,334,644	15,837,509	16,594,323	17,191,600	17,714,263	17,992,600			
2004	10,616,734	14,461,802	15,894,418	16,644,316	17,253,999	17,678,604				
2005	11,234,028	14,723,765	16,219,677	17,166,414	17,739,780					
2006	11,762,138	15,866,737	17,804,050	18,942,686						
2007	12,189,806	16,932,637	19,062,769							
2008	12,394,686	17,351,056								
2009	11,230,304									

The UL is the 120-month, case-incurred loss for each AY. For example, for AY 2001, the UL is yellow highlighted cell below.

2001	10,597,343	15,121,875	17,148,317	18,030,720	18,315,025	18,663,494	18,963,470	19,274,423	19,492,073	19,691,532
2002	9,880,312	14,375,596	16,125,066	16,778,143	17,329,085	17,651,826	17,996,103	18,251,653	18,518,021	
2003	10,369,458	14,334,644	15,837,509	16,594,323	17,191,600	17,714,263	17,992,600	18,339,834		
2004	10,616,734	14,461,802	15,894,418	16,644,316	17,253,999	17,678,604	18,059,807			
2005	11,234,028	14,723,765	16,219,677	17,166,414	17,739,780	18,273,742				
2006	11,762,138	15,866,737	17,804,050	18,942,686	19,777,903					
2007	12,189,806	16,932,637	19,062,769	20,387,578						
2008	12,394,686	17,351,056	19,656,367							
2009	11,230,304	15,726,363								
2010	11,564,142									

IL stands for the projected 120-month, case-incurred loss for each AY from first RY. For example, for AY 2001, the IL is projected from the triangle below.

1992	13,577,454	18,324,228	19,318,317	19,887,417	20,329,004	20,578,733	20,829,649	21,093,165	21,251,010	21,445,507
1993	11,629,774	15,072,352	16,240,909	17,019,006	17,377,062	17,817,291	17,991,759	18,151,739	18,369,115	
1994	10,064,429	13,048,331	14,255,204	14,761,320	15,191,420	15,488,568	15,640,632	15,907,265		
1995	9,091,097	12,075,438	13,156,052	13,952,635	14,395,087	14,703,283	14,936,659			
1996	9,272,956	12,274,028	13,702,650	14,407,800	14,920,371	15,315,039				
1997	9,322,137	13,169,560	14,673,981	15,446,460	16,059,155					
1998	10,192,450	14,078,543	15,775,394	16,820,701						
1999	9,840,268	13,948,976	15,911,651							
2000	10,300,006	15,323,250								
2001	10,597,343									

The red bold cells below are projected by chain-ladder method and the yellow highlighted cell is the IL.

1992	13,577,454	18,324,228	19,318,317	19,887,417	20,329,004	20,578,733	20,829,649	21,093,165	21,251,010	21,445,507
1993	11,629,774	15,072,352	16,240,909	17,019,006	17,377,062	17,817,291	17,991,759	18,151,739	18,369,115	18,537,236
1994	10,064,429	13,048,331	14,255,204	14,761,320	15,191,420	15,488,568	15,640,632	15,907,265	16,059,355	16,206,336
1995	9,091,097	12,075,438	13,156,052	13,952,635	14,395,087	14,703,283	14,936,659	15,125,933	15,270,552	15,410,313
1996	9,272,956	12,274,028	13,702,650	14,407,800	14,920,371	15,315,039	15,493,651	15,689,983	15,839,995	15,984,968
1997	9,322,137	13,169,560	14,673,981	15,446,460	16,059,155	16,419,983	16,611,482	16,821,978	16,982,814	17,138,246
1998	10,192,450	14,078,543	15,775,394	16,820,701	17,422,664	17,814,128	18,021,886	18,250,255	18,424,746	18,593,376
1999	9,840,268	13,948,976	15,911,651	16,820,875	17,422,844	17,814,313	18,022,073	18,250,444	18,424,936	18,593,568
2000	10,300,006	15,323,250	17,243,980	18,229,336	18,881,709	19,305,957	19,531,113	19,778,606	19,967,710	20,150,461
2001	10,597,343	15,145,457	17,043,901	18,017,824	18,662,628	19,081,953	19,304,497	19,549,118	19,736,028	19,916,659

References

- [1.] A.M. Best, “1969-2009 Impairment Review,” *Best’s Special Report*, June 21, 2010.
- [2.] Archer-Lock, P., et al., “The Cycle Survival Kit—An Investigation into the Reserving Cycle and Other Issues,” paper presented at the 2003 General Insurance Convention.
- [3.] England, P., and R. Verrall, “Analytic and Bootstrap Estimates of Prediction Errors in Claims Reserving,” *Insurance: Mathematics and Economics* 25 (1999):281–293.
- [4.] England, P., and R. Verrall, “Stochastic Claims Reserving in General Insurance,” *British Actuarial Journal* 8(3), 2002: 443-518.
- [5.] General Insurance Reserving Oversight (GIRO) Committee, Institute of Actuaries (U.K.), “Best Estimates and Reserving Uncertainty,” paper presented at the 2007 GIRO conference.
- [6.] Meyers, G., and P. Shi, “The Retrospective Testing of Stochastic Loss Reserve Methods,” *2011 Summer CAS E-Forum*.
- [7.] Meyers, Glenn G., “Estimating Predictive Distributions for Loss Reserve Models,” *Variance* 1:2, 2007: 248-272.
- [8.] Pinheiro, P.J.R., J.M. Andrade e Silva, and M. De Lourdes Centeno, (2003), Bootstrap Methodology in Claim Reserving. *Journal of Risk and Insurance* 70:701–714.
- [9.] Risk Margins Task Force, Institute of Actuaries of Australia, “A Framework for Assessing Risk Margins,” 2008, paper presented at Institute of Actuaries of Australia 16th General Insurance Seminar, 9-12 November 2008.