Estimating Predictive Distributions for Loss Reserve Models

by Glenn G. Meyers

ABSTRACT

This paper demonstrates a Bayesian method for estimating the distribution of future loss payments of individual insurers. The main features of this method are (1) the stochastic loss reserving model is based on the collective risk model; (2) predicted loss payments are derived from a Bayesian methodology that uses the results of large, and presumably stable, insurers as its prior information; and (3) this paper tests its model on a large number of insurers and finds that its predictions are well within the statistical bounds expected for a sample of this size. The paper concludes with an analysis of reported reserves and their subsequent development in terms of the predictive distribution calculated by this Bayesian methodology.

KEYWORDS

Reserving methods, reserve variability, uncertainty and ranges, Schedule P, suitability testing, collective risk model, Fourier methods, Bayesian estimation, hypothesis testing

1. Introduction

Over the years, there have been a number of stochastic loss reserving models that provide the means to statistically estimate confidence intervals for loss reserves. In discussing these models with other actuaries, I find that many feel that the confidence intervals estimated by these methods are too wide. The reason most give for this opinion is that experienced actuaries have access to information that is not captured by the particular formulas they use. These sources of information can include intimate knowledge of claims at hand. A second source of information that many actuarial consultants have is the experience gained by setting loss reserves for other insurers.

As one digs into the technical details of the stochastic loss reserving models, one finds many assumptions that are debatable. For example, Mack [9], Barnett and Zehnwirth [1], and Clark [2] all make a number of simplifying assumptions on the distribution of an observed loss about its expected value. Now it is the essence of predictive modeling to make simplifying assumptions. Which set of simplifying assumptions should we use? Arguments based on the "reasonability" of the assumptions can (at least in my experience) only go so far. One should also test the validity of these assumptions by comparing the predictions of such a model with observations that were not used in fitting the model.

Given the inherent volatility of loss reserve estimates, testing a single estimate is unlikely to be conclusive. How conclusive is the following statement?

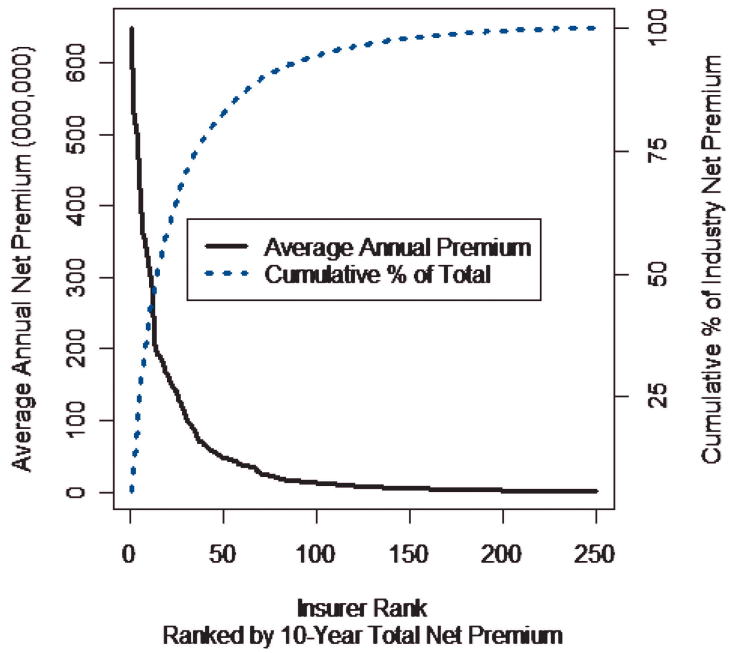
> Yes, the prediction falls somewhere within a wide range.

A more comprehensive test of a loss reserve model would involve testing its predictions on many insurers.

The purpose of this paper is to address at least some of the issues raised above.

- The methodologies developed in this paper will be applied to the Schedule P data submitted on the 1995 NAIC Annual Statement for each of 250 insurers.
- The stochastic loss model underlying the methods of this paper will be the collective risk model. This model combines the underlying frequency and severity distributions to get the distribution of aggregate losses. This approach to stochastic loss reserving is not entirely new. Hayne [5] uses the collective risk model to develop confidence regions for the loss reserve, but the regions assume that the expected value of the loss reserve is known. This paper makes explicit use of the collective risk model to first derive the expected value of the loss reserve.
- Next, this paper will illustrate how to use Bayes' Theorem to estimate the predictive distribution of future paid losses for an individual insurer. The prior distributions used in this method will be "derived" by an analysis of loss triangles for other insurers. This method will provide some of the "experience gained by setting loss reserves for other insurers" that is missing from existing statistical models for calculating loss reserves. An advantage of such an approach is that all assumptions (i.e., prior distributions) and data will be clearly specified.
- Next, this paper will test the predictions of the Bayesian methodology on data from the corresponding Schedule P data in the corresponding 2001 NAIC Annual Statements. The essence of the test is to use the predictive distribution derived from the 1995 data to estimate the predicted percentile of losses posted in the 2001 Annual Statement for each insurer. While the circumstances of each individual insurer may be different, the predicted percentiles of the observed losses should be uniformly distributed. This will be tested by standard statistical methods.
- Finally, this paper will analyze the reported reserves and their subsequent development in

Figure 1. Distribution of insurer size



terms of the predictive distributions calculated by this Bayesian methodology.

The main body of the paper is written to address a general actuarial audience. My intention is to make it clear what I am doing in the main body. In the Appendix, I will discuss additional details needed to implement the methods described in some of the sections.

2. Exploratory data analysis

The basic data used in this analysis was the earned premium and the incremental paid losses for accident years 1986 to 1995. The incremental paid losses were those reported as paid in each calendar year through 1995.

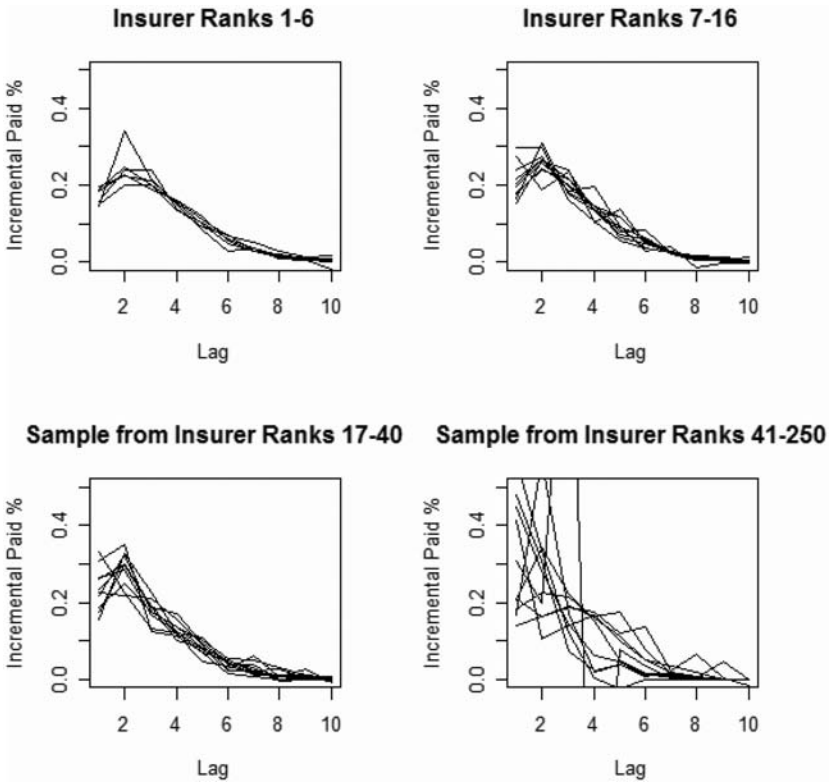
The data used in this analysis was taken from Schedule P of the 1995 NAIC Annual Statement, as compiled by the A. M. Best Company. I chose the Commercial Auto line of business because the payout period was long enough to be interesting but short enough so that ignoring the tail did not present a significant problem. The estimation of the tail is beyond the scope of this paper.

I selected 250 individual insurance groups from the hundreds that were reported by A. M. Best. Criteria for selection included: (1) at least some exposure in each of the years 1986 to 1995; (2) no abrupt changes in the exposure from year to year; and (3) no sharp decreases in the cumulative payment pattern. There were occasions when the incremental paid losses were negative, but small in comparison with the total loss. In this case, I treated the incremental losses as if they were zero. I believe this had minimal effect on the total loss reserve.

Let's look at some graphic summaries of the data. Figure 1 shows the distribution of insurer sizes, ranked by 10-year total earned premium. It is worth noting that 16 of the insurers accounted for more than half of the total premium of the 250 insurers.

Figure 2 shows the variability of payment paths (i.e., proportion of total reported paid loss segregated by settlement lag) for the accident year 1986. This figure makes it clear that payment paths do vary by insurer. How much these differences can be attributed to systematic differ-

Figure 2. Empirical payment paths for accident year 1986: Incremental paid losses as a proportion of 10-year total



ences between insurers, versus how much can be attributed to random processes, is unclear at this point.

Figure 3 shows the aggregate payment patterns for four groups, each accounting for approximately one quarter of the total premium volume.

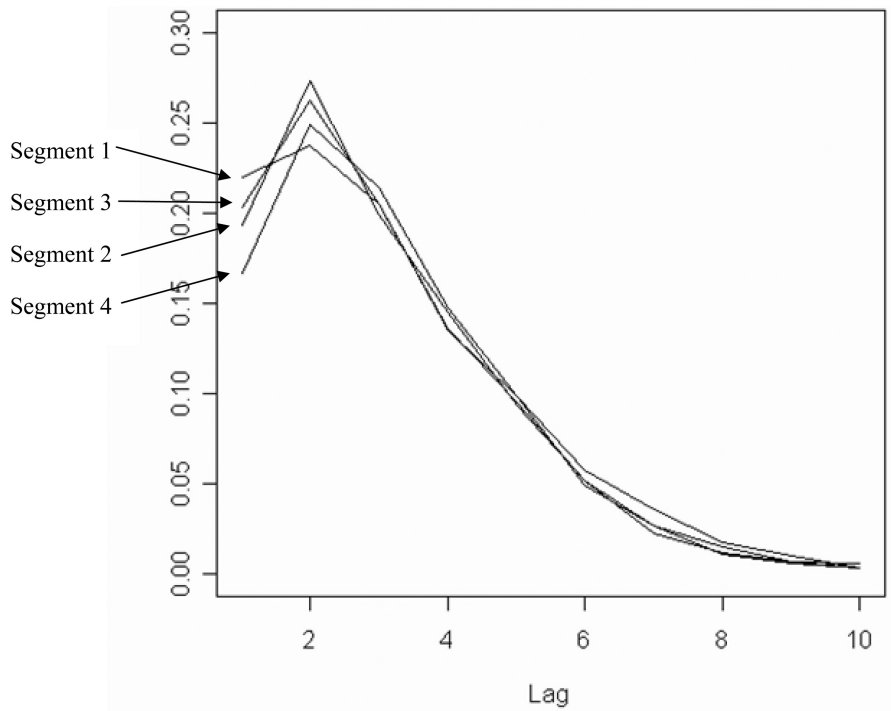
- Insurers ranked 1–6, 7–16, 17–40, and 41–250 each accounted for about one quarter of the total premium.
- Each plot represents approximately one quarter of the total premium volume.
- The variability of the incremental paid loss factors increases as the size of the insurer decreases.
- Segment 1—Insurers ranked 1–6, Segment 2—Insurers ranked 7–16, Segment 3—Insurers ranked 17–40, Segment 4—Insurers ranked 41–250.
- There is no indication of any systematic differences in payout patterns by size of insurer.

3. A stochastic loss reserve model

The goal of this paper is to develop a loss reserving model that makes testable predictions, and to then actually perform the tests. Let's start with a more detailed outline of how I intend to reach this goal.

1. The model for the expected payouts will be fairly conventional. It will be similar to the "Cape Cod" approach first published by Stanard [12]. This approach assumes a constant expected loss ratio across the 10-year span of the data.
2. Given the expected loss, the distribution of actual losses around the expected will be modeled by the collective risk model—a compound frequency and severity model. As mentioned above, this approach has precedents with Hayne [5]. This will conclude Section 3.

Figure 3. Empirical payment paths for the four industry segments



3. In Section 4, I will turn to estimating the parameters for the above models. The initial estimation method will be that of maximum likelihood.
4. I will then discuss testing the predictions of the model in Section 5. Initially, the tests will be on the same data that was used for fitting the models. (The tests on data in the 2001 Annual Statements will come later.) As mentioned above, the test will consist of calculating the percentiles of each of the observed loss payments and testing to see that those predictions are uniformly distributed.

As we proceed, I will focus on the 40 largest insurers. I do this because, in my judgment, the models are responding mainly to random losses for the smaller insurers. As we shall see, the results of the fitted models for the 40 largest insurers will form the basis for the Bayesian analysis that will be applied to each insurer, large and small. Implicit in this approach is the assumption that main systematic differences in the loss payment paths are somehow captured by the largest 40 insurers.

Let's proceed.

Assume that the expected losses are given by the following model.

$$\mathrm{E}[Paid\ Loss_{AY,Lag}] = Premium_{AY} \times ELR \times Dev_{Lag}, \qquad (1)$$

where

- $AY(1986 = 1, 1987 = 2, \ldots)$ is an index for accident year.
- $Lag = 1, 2, \ldots, 10$ is the settlement lag reported after the beginning of the accident year.
- *Paid Loss* is the incremental paid loss for the given accident year and settlement lag.
- *Premium* is the earned premium for the accident year.
- *ELR* is an unknown parameter that represents the expected loss ratio.
- Dev_{Lag} is an unknown parameter that depends on the settlement lag.

As with Stanard's "Cape Cod" method, the *ELR* parameter will be estimated from the data.

The "Cape Cod" formula that I used to estimate the expected loss is by no means a necessary feature of this method. Other formulas, like the chain ladder model, can be used.

A common adjustment that one might make to Equation 1 is to multiply the *ELR* by a premium index to adjust for the "underwriting cycle." I tried this, but it did not appreciably increase the accuracy of the predictions *for this data and time period*. Thus I chose to use the simpler model in this paper. But one should consider using a premium index in other circumstances.

Let $X_{AY,Lag}$ be a random variable for an insurer's incremental paid loss in the specified accident year and settlement lag. Assume that $X_{AY,Lag}$ has a compound negative binomial (CNB) distribution, which I will now describe.

- Let Z_{Lag} be a random variable representing the claim severity. Allow each claim severity distribution to differ by settlement lag.
- Given $\mathrm{E}[\mathit{Paid\ Loss}]_{AY,Lag}$, define the expected claim count $\lambda_{AY,Lag}$ by

$$\lambda_{AY,Lag} \equiv \mathrm{E}[\mathit{Paid\ loss}_{AY,Lag}]/\mathrm{E}[Z_{Lag}]. \quad (2)$$

This definition of the expected claim count may not correspond to the way an insurer actually counts claims. What is important is that it is consistent with the way the claim severity distribution is constructed. Our purpose is to describe the distribution of $X_{AY,Lag}$.

- Let $N_{AY,Lag}$ be a random variable representing the claim count. Assume that the distribution of $N_{AY,Lag}$ is given by the negative binomial distribution with mean $\lambda_{AY,Lag}$ and variance $\lambda_{AY,Lag} + c \cdot \lambda^2_{AY,Lag}$.
- Then the random variable $X_{AY,Lag}$ is defined by

$$X_{AY,Lag} = Z_{Lag,1} + Z_{Lag,2} + \cdots + Z_{Lag,N_{AY,Lag}}.$$

While the above defines how to express the random variable $X_{AY,Lag}$ in terms of other random variables $N_{AY,Lag}$ and Z_{Lag}, later on we will need to calculate the likelihood of observing $x_{AY,Lag}$ for various accident years and settlement lags. The details of how to do this are in the technical appendix. Here is a high-level overview of what will be done below.

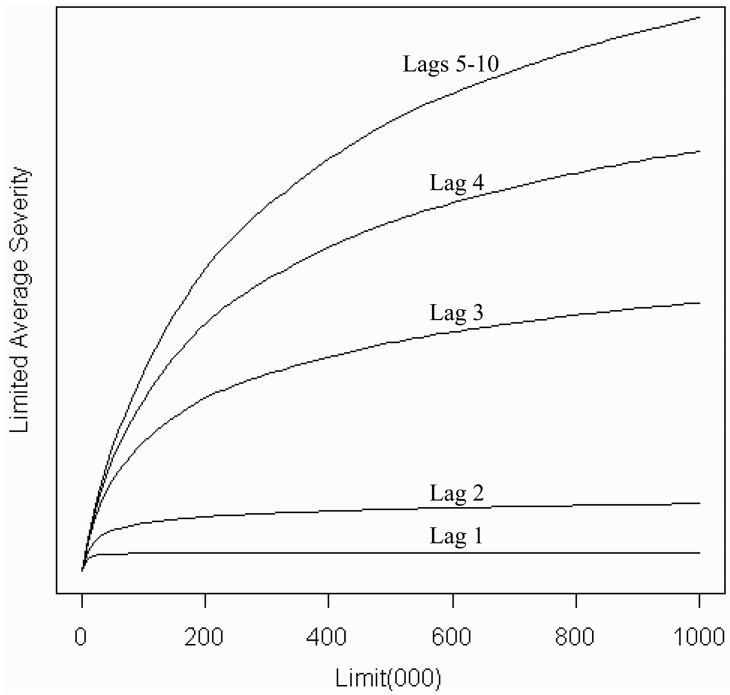
1. The distributions of Z_{Lag} were derived from data reported to ISO as part of its regular increased limits ratemaking activities. Like the substantial majority of insurers that report their data to ISO, the policy limit will be set to $1,000,000. The distributions varied by settlement lag with lags 5–10 having the highest severity. See Figure 4 below. For this application I discretized the distributions at intervals h, which depended on the size of the insurer. I chose h so the 2^{14} (16,384) values spanned the probable range of losses for the insurer.
2. I selected the value of 0.01 for the negative binomial distribution parameter, c. The paper by Meyers [10] analyzes Schedule P data for Commercial Auto and provides justification for this selection.
3. Using the Fast Fourier Transform (FFT), I then calculated the entire distribution of a discretized $X_{AY,Lag}$, rounded to the nearest multiple of h. The use of Fourier Transforms for such calculations is not new. References for this in CAS literature include the papers by Heckman and Meyers [6] and Wang [13].
4. Whenever the probability density of a given observation $x_{AY,Lag}$ given $E[\mathit{Paid\ Loss}_{AY,Lag}]$ was needed, I rounded the $x_{AY,Lag}$ to the nearest multiple of h and did the above calculation.

The resulting distribution function is denoted by:

$$\mathrm{CNB}(x_{AY,Lag} \mid \mathrm{E}[\mathit{Paid\ loss}_{AY,Lag}]). \quad (3)$$

This specifies the stochastic loss reserving model used in this paper. The parameters that depend on the particular insurer are the *ELR* and the 10 Dev_{Lag} parameters. I will now turn to showing how to estimate these parameters, given the earned premiums and the Schedule P loss triangle.

Figure 4. Limited average severity by settlement lag



Clark [2] has taken a similar approach to loss reserve estimation. Indeed, I credit Clark for the inspiration that led to the approach taken in this section and the next. Clark used the Weibull and loglogistic parametric models where I used Equation 1 above. In place of the CNB distribution described above, Clark used what he calls the "overdispersed Poisson" (ODP) distribution.[1] He then estimated the parameters of his model by maximum likelihood. This is where I am going next.

4. Maximum likelihood estimation of model parameters

The data for a given insurer consists of earned premium by accident year, indexed by $AY = 1, 2, \ldots, 10$, and a Schedule P loss triangle with losses $\{x_{AY,Lag}\}$ and $Lag = 1, \ldots, (11 - AY)$. With this data, one can calculate the probability, conditional on the parameters ELR and $Dev_{AY,Lag}$, of obtaining the data by the following equation.

$$L(\{x_{AY,Lag}\}) = \prod_{AY=1}^{10} \prod_{Lag=1}^{11-AY} CNB(x_{AY,Lag} \mid E[Paid\ Loss_{AY,Lag}]). \tag{4}$$

Generally one calls $L(\cdot)$ the likelihood function of the data.

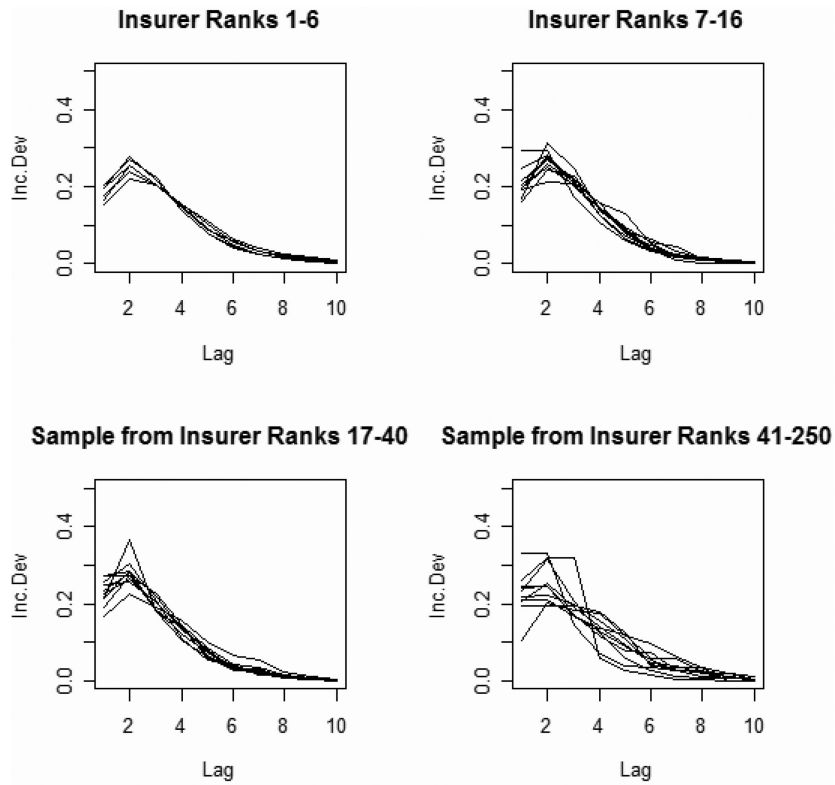
For this model, maximum likelihood estimation refers to finding the parameters ELR and Dev_{Lag} that maximize Equation 4 (indirectly through Equation 1). There are a number of mathematical tools that one can use to do this maximization. The particular method I used is described in the Appendix.

After examining the empirical paths plotted in Figures 2 and 3, I decided to put the following constraints in the Dev_{Lag} parameters.

1. $Dev_1 \leq Dev_2$.
2. $Dev_j \geq Dev_{j+1}$ for $j = 2, 3, \ldots, 9$.
3. $Dev_7/Dev_8 = Dev_8/Dev_9 = Dev_9/Dev_{10}$.
4. $\sum_{i=1}^{10} Dev_i = 1$.

[1] A random variable has an overdispersed Poisson distribution if it is an ordinary Poisson random variable times a constant scaling factor.

Figure 5. Maximum likelihood estimates of incremental paid development factors



The third set of constraints was included to add stability to the tail estimates. They also reduce the number of free parameters that need to be estimated from eleven to nine. The last constraint eliminated an overlap with the *ELR* parameter and maintained a conventional interpretation of that parameter.

Figure 5 plots the fitted payment paths for each of the 250 insurers. You might want to compare these payment paths with the empirical payment paths in Figure 2.

Figure 6 gives histograms of the 250 *ELR* estimates.

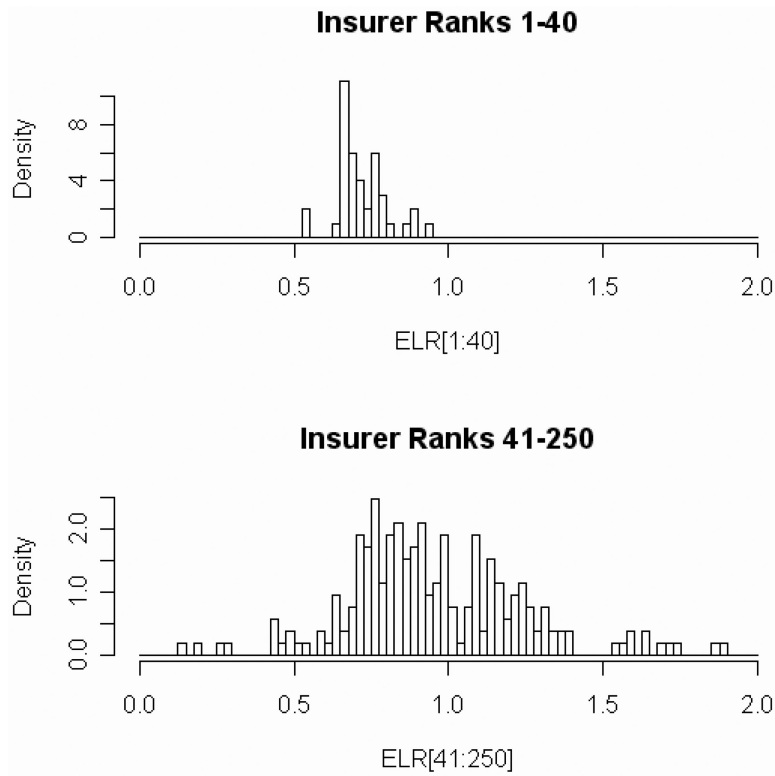
- The samples are the same insurers as in Figure 2.
- Note the wide variability of the fitted payment paths for the smallest insurers.
- Note the high variability of the ELR estimates for the smallest insurers.

5. Testing the model

Given parameter estimates *ELR* and $Dev_{AY,Lag}$, one can use the model specified by Equations 1–3 above to calculate the percentile of any observation $x_{AY,Lag}$ by first calculating the expected loss, then the expected claim count, and finally the distribution of losses about the expected loss by the CNB distribution. Whatever the expected losses, accident year or settlement lag, the percentiles should be uniformly distributed. One can also include the calculated percentiles of several insurers to give a more conclusive test of the model.

The hypothesis that any given set of numbers has a uniform distribution can be tested by the Kolmogorov-Smirnov test. See, for example, Klugman, Panjer, and Willmot (KPW) ([8], p. 428) for a reference on this test. The test is applied in our case as follows. Suppose you have a sample of numbers, $F_1, F_2, \ldots, F_n$, between 0 and

Figure 6. Maximum likelihood estimates of the ELR parameters



1, sorted in increasing order. One then calculates the test statistic:

$$D = \max_i \left| F_i - \frac{i}{n+1} \right|. \tag{5}$$

If D is greater than the critical value for a selected level α, we reject the hypothesis that the F_is are uniformly distributed. The critical values depend upon the sample size. Commonly used critical values are $1.22/\sqrt{n}$ for $\alpha = 0.10, 1.36/\sqrt{n}$ for $\alpha = 0.05$, and $1.63/\sqrt{n}$ for $\alpha = 0.01$.

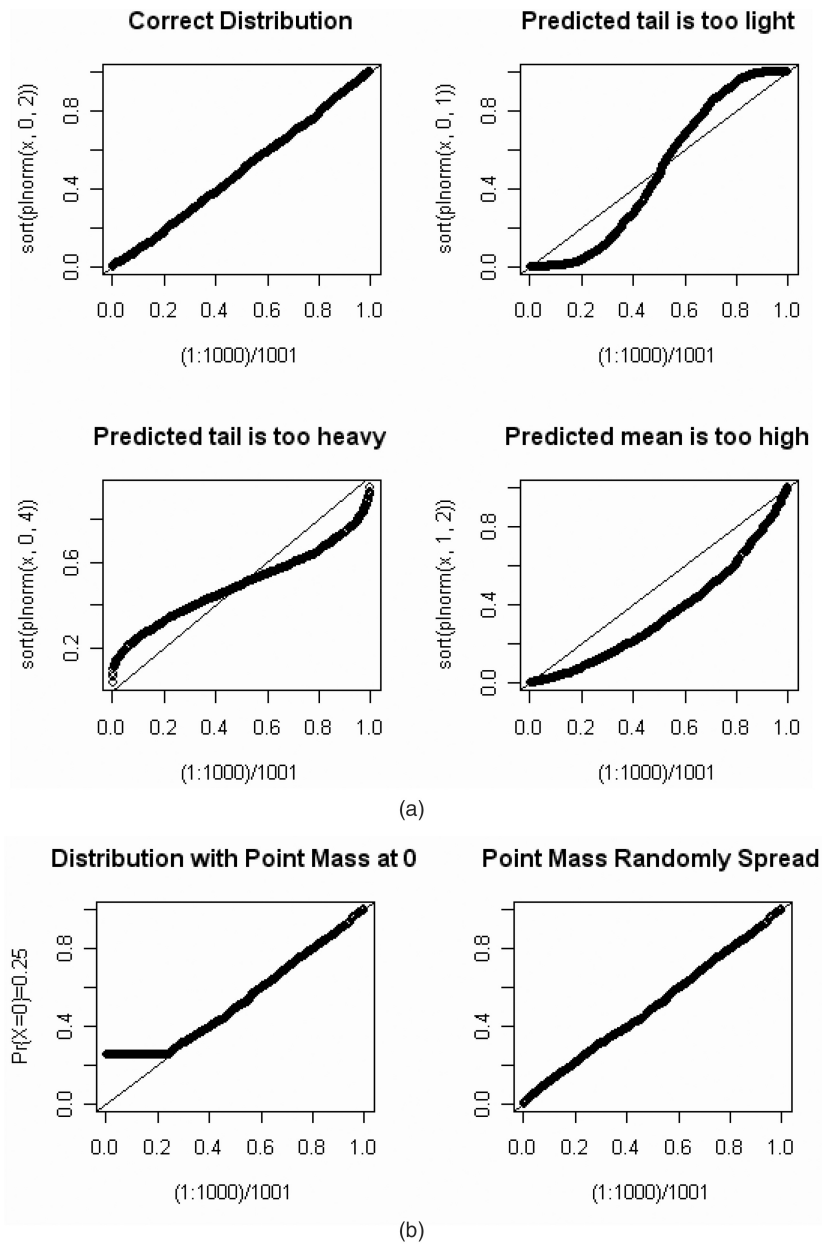
Note that the Kolmogorov-Smirnov test should not be applied when testing the model with data that was used to fit the model.

A graphical way to test for uniformity is a p-p plot, which is sometimes called a probability plot. A good reference for this is KPW ([8], p. 424). The plot is created by arranging the observations $F_1, F_2, \ldots, F_n$ in increasing order and plotting the points $(i/(n+1), F_i)$ on a graph. If the model is "plausible" for the data, the points will be near the 45° line running from $(0,0)$ to $(1,1)$. Let d_α be a critical value for a Kolmogorov-Smirnov test. Then the p-p plot for a plausible model should lie within $\pm d_\alpha$ of the 45° line.

A nice feature of p-p plots is that they provide, to the trained eye, a diagnosis of problems that may arise from an ill-fitting model. Let's look at some examples. Let x be a random sample of 1,000 numbers from a lognormal distribution with parameters $\mu = 0$ and $\sigma = 2$. Let's look at some p-p plots when we mistakenly choose a lognormal distribution with different μs and σs. On Figure 7(a), sort(plnorm(x, μ, σ)) on the vertical axis will denote the sorted F_is predicted by a lognormal distribution with parameters μ and σ.

- On the first graph, μ and σ are the correct parameters, and the p-p plot lies on a 45° line as expected.
- On the second graph, with $\sigma = 1$, the low predicted percentiles are lower than expected, while the high predicted percentiles are higher

Figure 7. Sample p-p plots

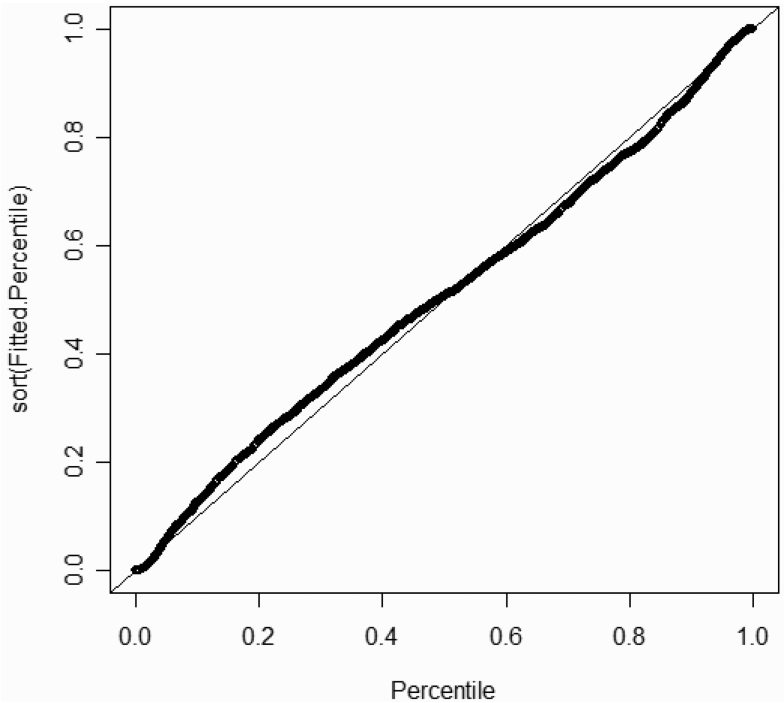


than expected. This indicates that the tails are too light.

- On the third graph, with $\sigma = 4$, the low predicted percentiles are higher than expected, while the high predicted percentiles are lower than expected. This indicates that the tails are too heavy.
- On the fourth graph, with $\mu = 1$, almost all the predicted percentiles are lower than expected. This indicates that the predicted mean is too high.

If a random variable X has a continuous cumulative distribution function $F(x)$, the F_is associated with a sample $\{x_i\}$ will have a uniform distribution. There are times when we want to use a p-p plot with a random variable X, which we expect to have a positive probability at $x = 0$. The left side of Figure 7(b) shows a p-p plot for a distri-

Figure 8. P-P plot for the top 40 insurers



bution with $\Pr(X = 0) = 0.25$. The Kolmogorov-Smirnov test is not applicable in this case. However, we can "transform" the F_is into a uniform distribution by multiplying $F_i = F(x_i)$ by a random number that is uniformly distributed whenever $x_i = 0$. We can then use the Kolmogorov-Smirnov test of uniformity. The right side of Figure 7(b) illustrates the effect of such an adjustment. All of the p-p plots below will have this adjustment.

Now let's try this for real.

Figure 8 gives a p-p plot for the percentiles predicted for the data that was used to fit each model for the top 40 insurers. Overall there were 2,200 ($= 40 \times 55$) calculated percentiles. By examining Figure 8, we see that the fitted model has tails that are a bit too heavy.

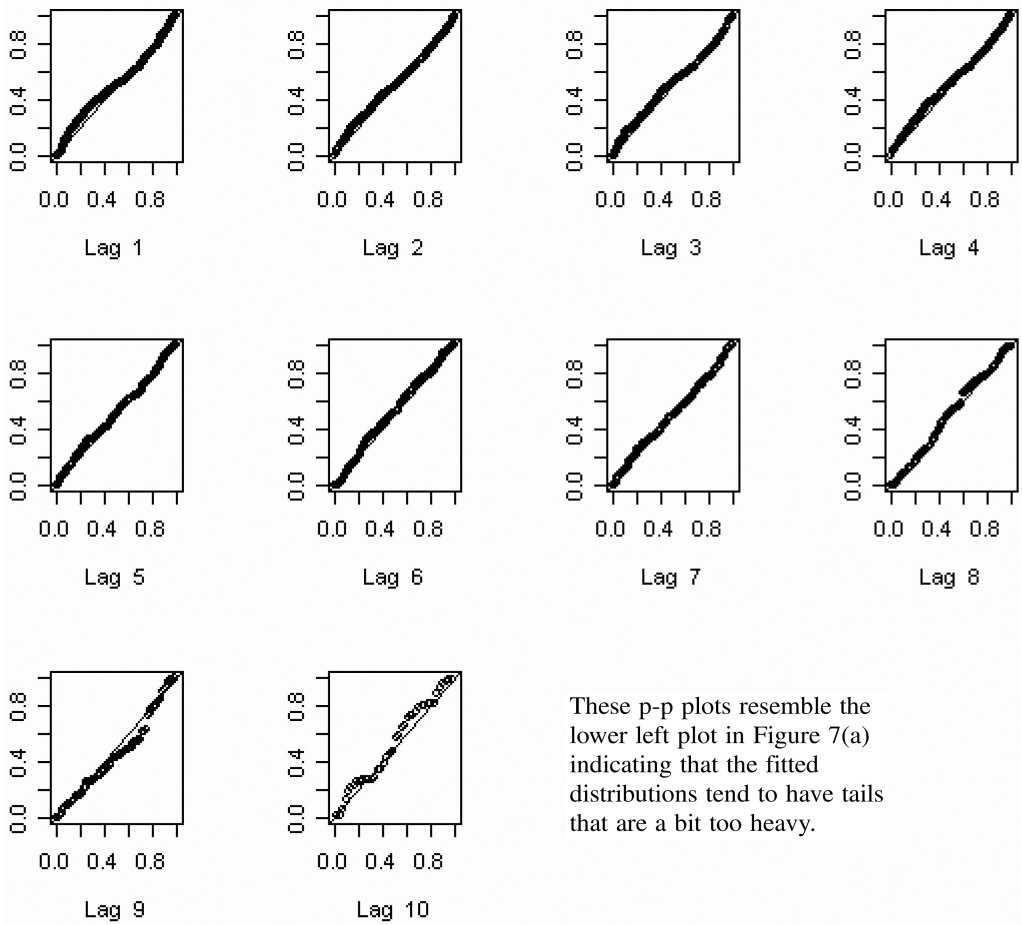
Let me make a personal remark here. In my many years of fitting models to data, it is a rare occasion when a model passes such a test with data consisting of thousands of observations. I was delighted with the goodness of fit. Nevertheless, I investigated further to see what "went wrong." Figure 9 shows p-p plots for the same data segregated by settlement lag. These plots appear to indicate that the main source of the problem is in the distributions predicted for the lower settlement lags.

Figure 10 shows p-p plots for the percentiles predicted for the data used in fitting the smallest 210 insurers. Suffice it to say that these plots reveal serious problems with using this estimation procedure with the smaller insurers. I think the problem lies in fitting a model with nine parameters to noisy data consisting of 55 observations. On the other hand, the procedure appears to work fairly well for large insurers with relatively stable loss payment patterns. See Figures 2, 3, 5 and 6. I suspect the same problem with small insurers occurs with other many-parameter models, such as the chain ladder method.

6. Predicting future loss payments using Bayes' Theorem

The failure of the model to predict the distribution of losses for the smaller insurers and

Figure 9. P-P plots for the top 40 insurers by settlement lag



the comparatively successful predictions of the model on larger insurers leads to the question: Is there any information that can be gained from the larger insurers that would be helpful in predicting the loss payments of the smaller insurers? That is the topic of this section.

Let $\omega = \{ELR(\omega), Dev_{Lag}(\omega), Lag = 1, 2, \ldots, 10\}$ be a set of vectors that determine the expected losses in accordance with Equation 1. Let Ω be the set of all ωs. These models are distinguished only by the values of their parameters, and not by the assumptions or methods that were used to generate the parameters. Using Equation 4, one can combine each expected loss model with the parameters as assumptions underlying Equations 2 and 3 to calculate the likelihood of the given loss triangle $\{x_{AY,Lag}\}$. Each likelihood can be interpreted as:

$$L = \text{Probability}\{\text{data} \mid \text{model}\} \equiv \Pr\{\{x_{AY,Lag}\} \mid \omega\}. \tag{6}$$

Then using Bayes' Theorem one can then calculate:

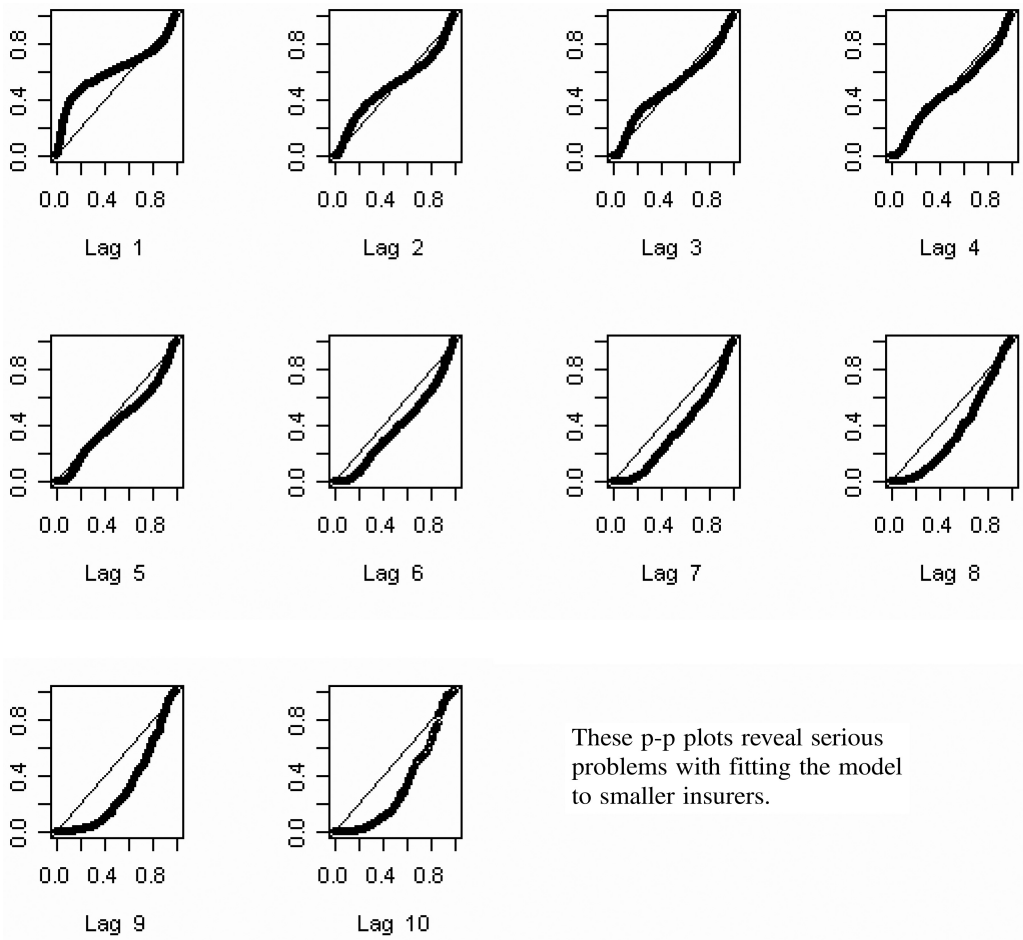
$$\text{Probability}\{\text{model} \mid \text{data}\} \propto \text{Probability}\{\text{data} \mid \text{model}\} \times \text{Prior}\{\text{model}\}.$$

Stated more mathematically:

$$\Pr\{\omega \mid \{x_{AY,Lag}\}\} \propto \Pr\{\{x_{AY,Lag}\} \mid \omega\} \times \Pr\{\omega\}. \tag{7}$$

Each $\omega \in \Omega$ will consist of forty $\{Dev_{Lag}\}$ combinations taken from maximum likelihood esti-

Figure 10. P-P plots by settlement lag for insurers ranked 41–250



mates of the top 40 insurers above. I judgmentally selected equal probabilities for each $\omega \in \Omega$. Each of the forty $\{Dev_{Lag}\}$ combinations will be independently crossed with nine potential *ELR*s starting with 0.600 and increasing by steps of 0.025 to a maximum of 0.800. Thus Ω has 360 parameter sets. I judgmentally selected the prior probability of the *ELR*s after an inspection of the distribution of maximum likelihood estimates. See Figure 11 and Table 1 below.

So we are given a loss triangle $\{x_{AY,Lag}\}$, and we want to find a stochastic loss model for our data. Here are the steps we would take to do this.

1. Using Equation 4, calculate $\Pr\{\{x_{AY,Lag}\} \mid \omega\}$ for each $\omega \in \Omega$.
2. The posterior probability of each $\omega \in \Omega$ is given by

$$\Pr\{\omega \mid \{x_{AY,Lag}\}\} = \frac{\Pr\{\{x_{AY,Lag}\} \mid \omega\} \times \Pr\{\omega\}}{\sum_{\omega \in \Omega} \Pr\{\{x_{AY,Lag}\} \mid \omega\} \times \Pr\{\omega\}}. \tag{8}$$

In words, the final stochastic model for a loss triangle is a mixture of all the models $\omega \in \Omega$, where the mixing weights are proportional to the posterior probabilities.

Here are some technical notes.

- In doing these calculations for the 250 insurers, it happens that almost all the weight is concentrated on, at most, a few dozen models out of the original 360. So, instead of includ-

Figure 11. Comparing the selected prior distribution of *ELR* with the maximum likelihood estimates of *ELR* for the top 40 insurers

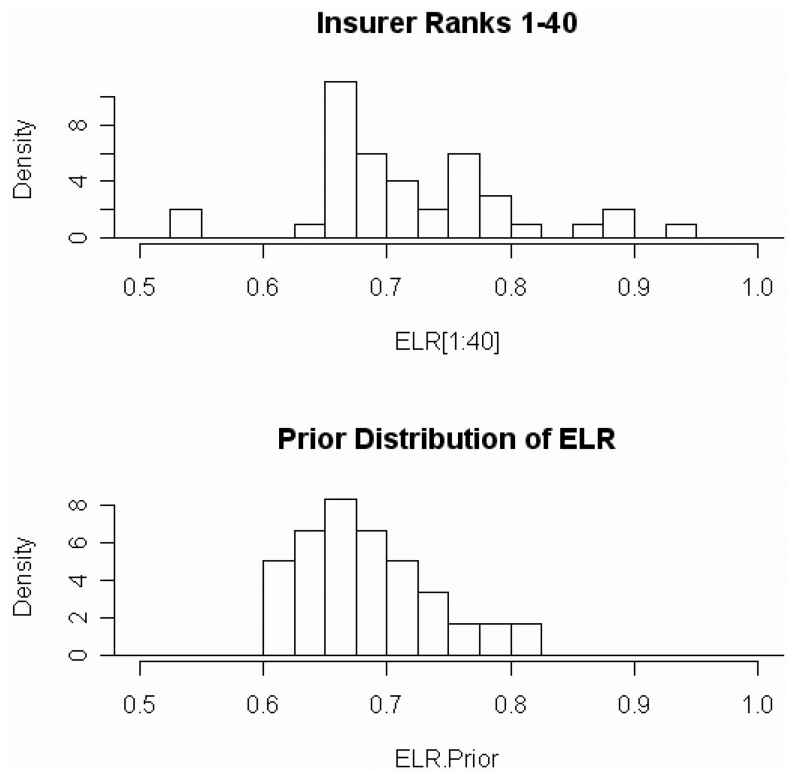


Table 1. Prior probabilities for *ELR*

ELR	Prior	*ELR*	Prior	*ELR*	Prior
0.600	3/24	0.675	4/24	0.750	1/24
0.625	4/24	0.700	3/24	0.775	1/24
0.650	5/24	0.725	2/24	0.800	1/24

ing all models in the original Ω, I sorted the models in decreasing order of posterior probability and dropped those after the cumulative posterior probability summed to 99.9%.

- When calculating the final model for any of the top 40 insurers, I excluded that insurer's parameters $\{Dev_{Lag}\}$ from Ω and added the parameters for the 41st largest insurer in its place. I did this to reduce the chance of overfitting.

The stochastic model of Equation 8 is not the end product. Quite often, insurers are interested in statistics such as the mean, variance, or a given percentile of the total reserve. I will now show how to use the stochastic model to calculate these "statistics of interest."

At a high level, the steps for calculating the "statistics of interest" are as follows.

1. Calculate the statistic conditional on ω for each accident year and settlement lag of interest, i.e., the complement of the triangle $AY = 2,\ldots,10$, $Dev = 12 - AY,\ldots,10$.
2. Aggregate the statistic over the desired accident years and settlement lags for each ω.
3. Calculate the unconditional statistic by mixing (or weighting) the conditional statistics of Step 2, above, with the posterior probabilities of each model in Ω.

These steps should become clearer as we look at specific statistics. Let's start with the expected value.

1. For each accident year and settlement lag, calculate the expected value for each ω using Equation 1.

$$\mathrm{E}[Paid\ Loss_{AY,Lag} \mid \omega] = Premium_{AY} \times ELR(\omega) \times Dev_{Lag}(\omega).$$

2. To get the total expected loss for each ω, sum the expected values over the desired accident years and settlement lags.

$$\mathrm{E}[Paid\ loss \mid \omega] = \sum_{AY,Lag} \mathrm{E}[Paid\ loss_{AY,Lag} \mid \omega].$$

3. The unconditional total expected loss is the posterior probability weighted average of the conditional total expected losses, with the posterior probabilities given by Equation 8.

$$\mathrm{E}[Paid\ loss] = \sum_{\omega \in \Omega} \mathrm{E}[Paid\ loss \mid \omega] \times \Pr\{\omega \mid \{x_{AY,Lag}\}\}.$$

Note that for each ω, the conditional expected loss will differ. Our next "statistic of interest" will be the standard deviation of these expected loss estimates. This should be of interest to those who want a "range of reasonable estimates."

The first two steps are the same as those for finding the expected loss above. In the third step we calculate E[*Paid Loss*] as above but, in addition, we calculate the second moment:

4.

$$SM[\hat{\mathrm{E}}[Paid\ loss]] \equiv \sum_{\omega \in \Omega} \mathrm{E}[Paid\ loss \mid \omega]^2 \times \Pr\{\omega \mid \{x_{AY,Lag}\}\}.$$

Then:

$$\text{Standard Deviation}[\hat{\mathrm{E}}[Paid\ loss]] = \sqrt{SM[\hat{\mathrm{E}}[Paid\ loss]] - \mathrm{E}[Paid\ loss]^2}.$$

As the second example begins to illustrate, the three steps to calculating the "statistic of interest" can get complex.

Our third "statistic of interest" is the standard deviation of the actual loss. Before we begin, it will help to go over the formulas involved in finding the standard deviation of sums of losses.

First, recall from Equation 2 that our model imputes an expected claim count, $\lambda_{AY,Lag}$, by dividing the expected loss by the expected claim severity for the settlement lag.

Next recall the following bullet from the description of the CNB distribution above.

- Let $N_{AY,Lag}$ be a random variable representing the claim count. Assume that the distribution of $N_{AY,Lag}$ is given by the negative binomial distribution with mean $\lambda_{AY,Lag}$ and variance $\lambda_{AY,Lag} + c \cdot \lambda^2_{AY,Lag}$.

The negative binomial distribution can be thought of as the following process.

1. Select the random number χ from a gamma distribution with mean 1 and variance c.
2. Select $N_{AY,Lag}$ from a Poisson distribution with mean $\chi \cdot \lambda_{AY,Lag}$.

Consider two alternatives for applying this to the claim count for each settlement lag in a given accident year.

1. Select χ independently for each settlement lag.
2. Select a single χ and apply it to each settlement lag.

If one selects the second alternative, the multivariate distribution of $\{N_{AY,Lag}\}$ is called the negative multinomial distribution. This does not change the distribution of losses of an individual settlement lag. It does generate the correlation between the claim counts by settlement lag.

I will assume that the multivariate claim count for settlement lags within a given accident year has a negative multinomial distribution. The thinking behind this is that the χ is the result of an economic process that affects how many claims occur in a given year.

Clark [3] provides an alternative method for dealing with correlation between settlement lags.

Let $F_{Z_{Lag}}$ be the cumulative distribution for Z_{Lag}. Mildenhall [11] shows that (stated in the notation of this paper) the distribution of $\sum_{Lag=12-AY}^{10} X_{AY,Lag}$ has a CNB distribution with expected claim count $\lambda_{AY,Tot} = \sum_{Lag=12-AY}^{10} \lambda_{AY,Lag}$ and claim severity distribution

$$F_{Z_{AY,Tot}} = \sum_{Lag=12-AY}^{10} \lambda_{AY,Lag} \cdot F_{Z_{Lag}} \Big/ \sum_{Lag=12-AY}^{10} \lambda_{AY,Lag}.$$

Now let's describe the three steps to calculate the standard deviation of the actual loss.

1. For each accident year and settlement lag, calculate the expected claim count $\lambda_{AY,Lag}(\omega)$ using Equation 2.
2. The aggregation for each ω takes place in two steps.
   a. Calculate the mean and variance of each accident year's actual loss.

$$E[Paid\ Loss_{AY} \mid \omega] = \lambda_{AY,Tot}(\omega) \cdot E[Z_{AY,Tot}].$$

$$\begin{aligned}\mathrm{Var}[Paid\ Loss_{AY} \mid \omega] &= \lambda_{AY,Tot}(\omega) \cdot SM[Z_{AY,Tot}] \\ &\quad + c \cdot \lambda_{AY,Tot}(\omega)^2 \\ &\quad \cdot E[Z_{AY,Tot}]^2.\end{aligned}$$

   b. Sum the first and second moments over the accident years.

$$E[Paid\ Loss \mid \omega] = \sum_{AY} E[Paid\ Loss_{AY} \mid \omega].$$

$$SM[Paid\ Loss \mid \omega] = \sum_{AY} SM[Paid\ Loss_{AY} \mid \omega].$$

3.

$$\begin{aligned}&E[Paid\ Loss] \\ &\quad = \sum_{\omega \in \Omega} E[Paid\ Loss \mid \omega] \times \Pr\{\omega \mid \{x_{AY,Lag}\}\}.\end{aligned}$$

$$\begin{aligned}&SM[Paid\ Loss] \\ &\quad = \sum_{\omega \in \Omega} SM[Paid\ Loss \mid \omega] \times \Pr\{\omega \mid \{x_{AY,Lag}\}\}.\end{aligned}$$

$$\begin{aligned}&\text{Standard Deviation}[Paid\ Loss] \\ &\quad = \sqrt{SM[Paid\ Loss] - E[Paid\ Loss]^2}.\end{aligned}$$

The final "statistic of interest" is the distribution of actual losses. We are fortunate that the CNB distribution of each individual $X_{AY,Lag}$ is already defined in terms of its Fast Fourier Transform (FFT). To get the FFT of the sum of losses, we can simply multiply the FFTs of the summands. Other than that, the three steps are similar to those of calculating the standard deviation of the actual losses. To shorten the notation, let X denote *Paid Loss*.

1. For each accident year and settlement lag, calculate the expected claim count $\lambda_{AY,Lag}(\omega)$ using Equation 2.
2. The aggregation for each ω takes place in three steps.
   a. Calculate the severity FFT

$$\begin{aligned}&\Phi(\vec{\mathbf{p}}_{AY,Tot|\omega}) \\ &\quad = \sum_{Lag=12-AY}^{10} \lambda_{AY,Lag(\omega)} \cdot \Phi(\vec{\mathbf{p}}_{Lag|\omega}) \Big/ \sum_{Lag=12-AY}^{10} \lambda_{AY,Lag(\omega)}\end{aligned}$$

   for each accident year, and then
   b. calculate the accident year FFT

$$\begin{aligned}&\Phi(\vec{\mathbf{q}}_{AY,Tot|\omega}) \\ &\quad = \left(1 - c \cdot \left(\sum_{Lag=12-AY}^{10} \lambda_{AY,Lag(\omega)}\right) \cdot (\Phi(\vec{\mathbf{p}}_{AY,Tot|\omega}) - 1)\right)^{-1/c}\end{aligned}$$

   for each accident year.
   c. The FFT for the sum of all accident years is given by:

$$\Phi(\vec{\mathbf{q}}_{\omega}) = \prod_{AY} \Phi(\vec{\mathbf{q}}_{AY,Tot|\omega})$$

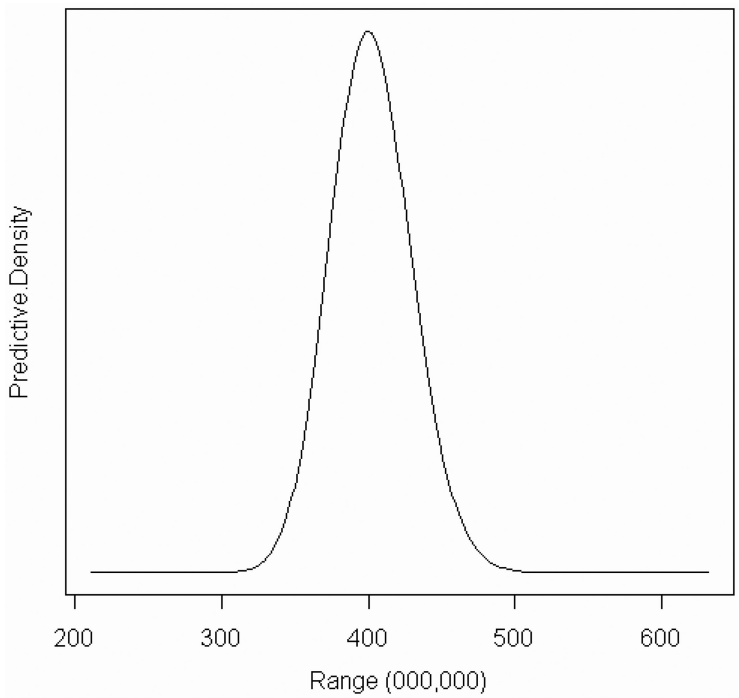
3. The distribution of actual losses is obtained by inverting the FFT:

$$\Phi(\vec{\mathbf{q}}) = \sum_{\omega \in \Omega} \Phi(\vec{\mathbf{q}}_{\omega}) \times \Pr\{\omega \mid \{x_{AY,Lag}\}\}.$$

See the Appendix for additional mathematical details of working with FFTs.

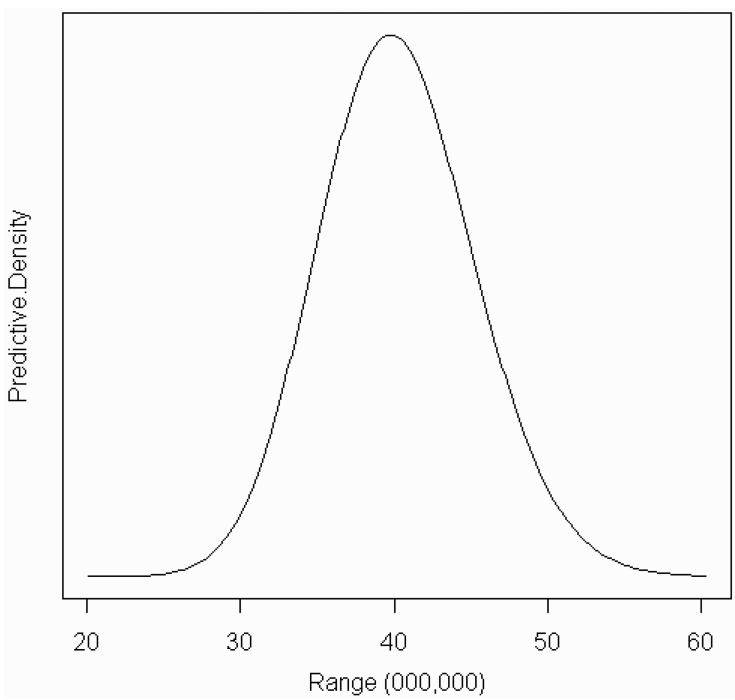
Figures 12 and 13 below show each of the three statistics for two insurers for the outstanding losses for accident years 2,…,10 up to settlement lag 10. The insurer in Figure 12 has ten times the predictive mean reserve as the insurer

Figure 12. Predictive distribution of actual losses for total reserve, insurer rank 7



- Predictive Mean = 401,951,000 (roughly ten times that in Figure 13).
- Coefficient of Variation for the Actual Loss = 6.9%.

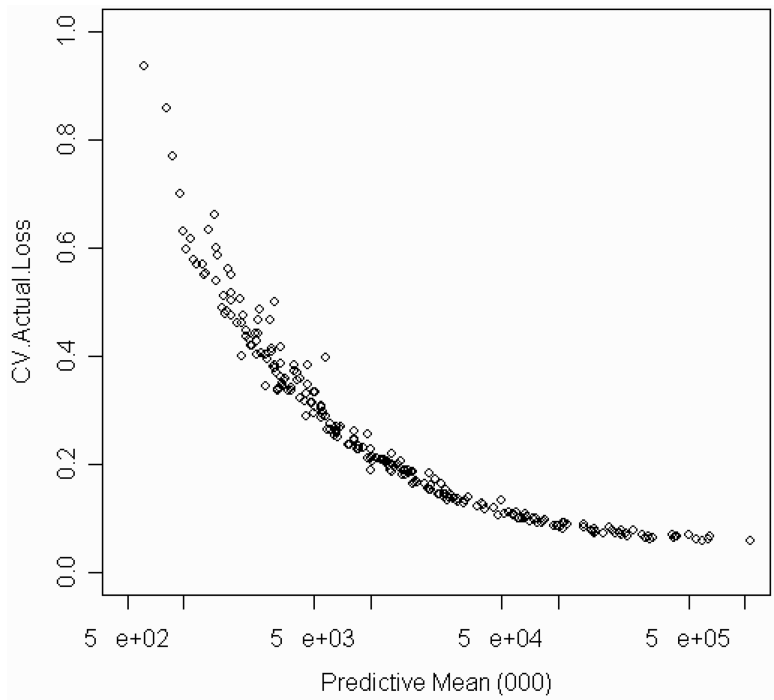
Figure 13. Predictive distribution of actual losses of total reserve, insurer rank 97



- Predictive Mean = 40,277,000 (roughly one tenth of that in Figure 12).
- Coefficient of Variation for the Actual Loss = 12.6%.

in Figure 13. Figure 14 plots the predictive coefficient of variation against the predictive mean reserve. The decreased variability that comes with size should not come as a surprise. The absolute levels of variability will be interesting only if I can demonstrate that this methodology can

Figure 14. Predictive coefficient of variation plotted with the predictive mean for 250 insurers



predict the distribution of future results. That is where I am going next.

7. Testing the predictions

The ultimate test of a stochastic loss reserving model is its ability to correctly predict the distribution of future payments. While the distribution of future payments will differ by insurer, when one calculates the predicted percentile of the actual payment, the distribution of these predicted percentiles should be uniform.

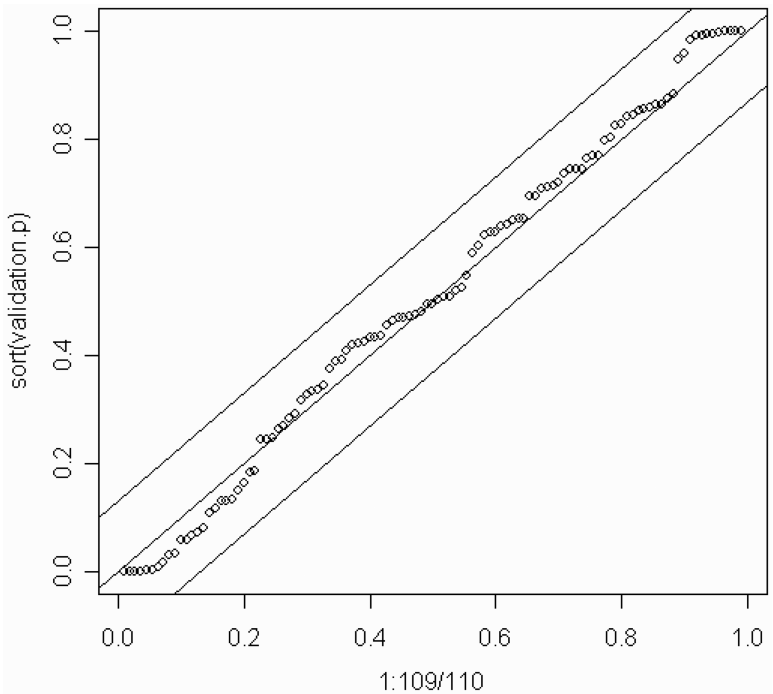
To test the model, we examined Schedule P from the 2001 NAIC Annual Statement. The losses reported in these statements contain six subsequent diagonals on the four overlapping years from 1992 through 1995. Earned premiums and losses in the overlapping diagonals for the 1995 and 2001 Annual Statements agreed in 109 of the 250 insurers, so I used these 109 insurers for the test.

Using the predictive distribution described in the last section, I calculated the predicted percentile of the total amount paid for the four accident years in the subsequent six settlement lags. These 109 percentiles should be uniformly distributed. Figure 15 shows the corresponding p-p plot and the confidence bands at the 5% level as determined by the Kolmogorov-Smirnov test. The plot lies well within that band. While we can never "prove" a model is correct with statistics, we gain confidence in a model as we fail to reject the model with such statistical tests. I believe this test shows that the Bayesian CNB model deserves serious consideration as a tool for setting loss reserves.

8. Comparing the predictive reserves with reported reserves

This section provides an illustration of the kind of analysis that can be done externally with the Bayesian methodology described in this paper. Readers should exercise caution in generalizing the conclusions of this section beyond this particular line of business in this particular time period.

Figure 15. P-P plot of predicted percentiles for paid losses from 1996 to 2001



The critical values for a Kolmogorov-Smirnov test at the 5% level are ±13.03%.

This paper makes no attempt to pin down the methods used in setting the reported reserves. However, there are many actuaries that expect reported reserves to be more accurate than a formula derived purely from the paid data reported on Schedule P. As stated in the introduction to this paper, those who set those reserves have access to more information that is relevant to estimating future loss payments.

The comparisons below will be performed to two sets of insurers—the entire set of 250 insurers and the subset of 109 insurers for which the overlapping accident years 1992–1995 agree. Testing the latter will enable us to compare the predictions based on information available in 1995 with the incurred losses reported in 2001.

The first test looks at aggregates summed over all insurers in each set. Table 2 compares the predictions of this model with the actual reserves reported on the 1995 annual statement. The "actual reserve" is the difference between the total reported incurred loss, as of 1995 for the "initial" reserve, and 2001 for the "retrospective" reserve, minus the total reported paid loss, as of 1995.

For the 250 insurers, the reported initial reserve was 9.1% higher than the predictive mean. For the 109 insurers the corresponding percentage was 9.9%. The lowering of the percentage reserves from 1995 to 2001 to 2.4% suggests that for the industry, reserves were redundant for Commercial Auto in 1995.[2]

For the remainder of this section, let's suppose that the expected value of the Bayesian CNB model described above is the "best estimate" of future loss payments. From the above, there are two arguments supporting that proposition.

1. Figure 15 in Section 7 above shows that the Bayesian CNB model successfully predicted the distribution of payments for the six years after 1995 well within the usually accepted statistical bounds of error.

[2]There are some potential biases in these figures. First, the predictive means may be somewhat understated since they ignore development after ten years. Second, the downward development from 1995 to 2001 may continue in future years.

Table 2. Predicted and reported loss reserves

		Reported 1995 Reserve (000)	
	Predictive Mean (000)	Initial @ 1995	Retrospective @ 2001
250 Insurers AY 1986–1995	14,873,303	16,221,998—9.1%	—
109 Insurers AY 1992–1995	1,798,794	1,976,299—9.9%	1,842,104—2.4%

2. The final row of Table 2 shows that the expected value predicted by the Bayesian CNB model, in aggregate, comes closer to the 2001 reserve than did the reported reserves for 1995.

Now let's examine some of the implications of this proposition for reported reserves.

There are many actuaries who argue that reported reserves should be somewhat higher than the mean. See, for example, Paragraph 2.17 on page 5 of Report of the Insurer Solvency Working Party of the International Actuarial Association [7]. Related to this, I recently saw a working paper by Grace and Leverty [4] that tests various hypotheses on insurer incentives.

If insurers were deliberately setting their reserves at some conservative level, we would expect to see that the reported reserves are at some moderately high percentile of the predictive distribution. Figure 16 shows that some insurers appear to be reserving conservatively. But there are also many insurers for which the predictive percentile of the reported reserve is below 50%. But by 2001, the percentiles of the retrospective reserve for 1995 were close to being uniformly distributed.

- The greater number of insurers reserved above the 50th percentile indicates that some insurers have conservative estimates of their loss reserves posted in 1995.
- The right side of this figure shows that the spread of the reserve percentiles spans all insurer sizes.

If there is a bias in the posted reserves, we would see corrections in subsequent years. The 109 insurers for which we have subsequent development provide data to test potential bias. To perform such a test, I divided the 109 insurers into two groups. The first group consisted of all insurers that posted reserves in 1995 that were lower than their predictive mean. The second group consisted of all insurers that posted reserves higher than their predictive mean.

As Figure 17 and Table 3 show, the first group shows an upward adjustment and the second group shows a more pronounced downward adjustment. The plots show that we cannot attribute these adjustments to only a few insurers. However, there are some insurers in the first group that show a downward adjustment, and other insurers in the second group that show an upward adjustment.

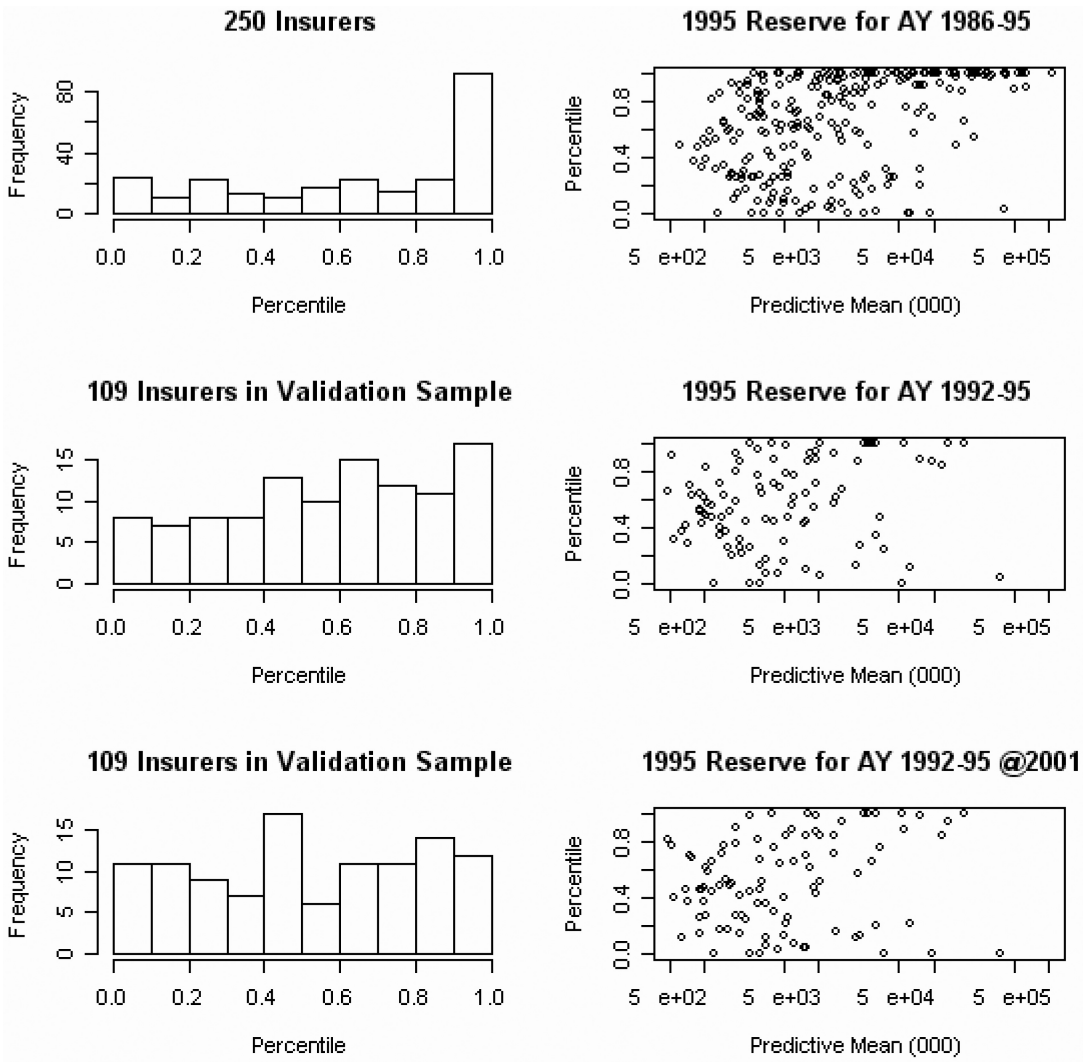
The fact that the total adjustments only go part way to the predictive mean suggests that some insurers may be able to make more accurate estimates with access to information that is not provided on Schedule P.

9. Summary and conclusions

This paper demonstrates a method, which I call the Bayesian CNB model, for estimating the distribution of future loss payments of individual insurers. The main features of this method are as follows.

- The stochastic loss reserving model is based on the collective risk model. While other stochastic loss reserving approaches make use of the collective risk model, this approach uses it as an integral part of estimating the parameters of the model.
- Predicted loss payments are derived from a Bayesian methodology that uses the results of

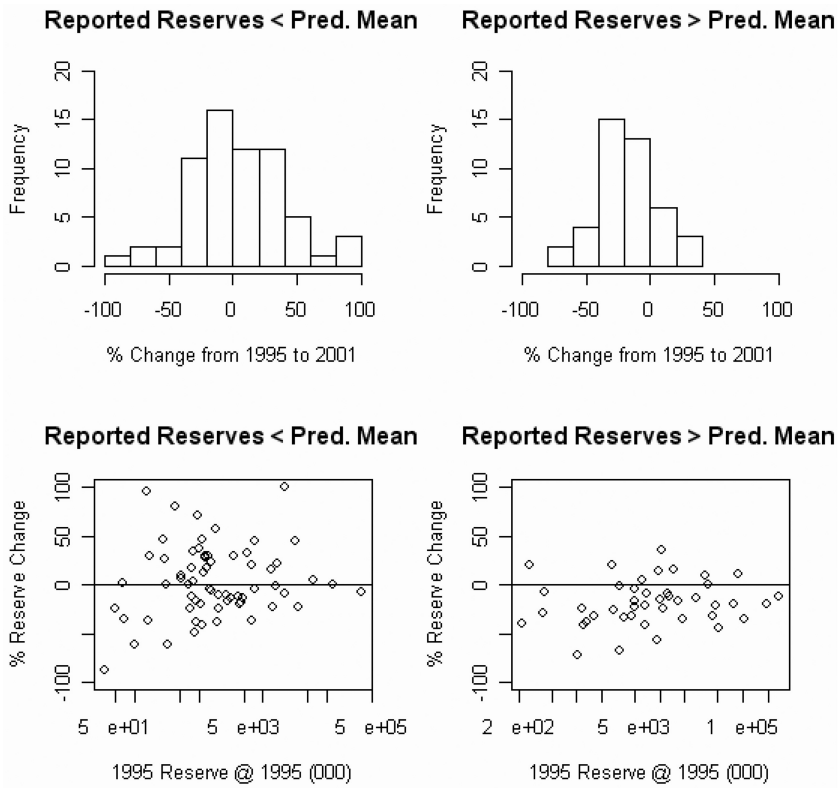
Figure 16. Predictive percentiles of reported reserves



large, and presumably stable, insurers as its "prior information." While insurers do indeed differ in their claim payment practices, the underlying assumption of this methodology is that these differences are reflected in this collection of large insurers. This paper demonstrates that using prior information derived from large insurers, together with the CNB model for the stochastic losses, makes this method applicable to all insurers, both large and small.

- Loss reserving models should be subject to testing their predictions on future payments. Tests on a single insurer are often inconclusive because of the volatile nature of the loss reserving process. But it is possible to test a stochastic loss reserving method on several insurers simultaneously by comparing its predicted percentiles of subsequent losses to a uniform distribution. This paper tests its model on 109 insurers and finds that its predictions are well within the statistical bounds expected for a sample of this size.
- By making the assumption that the Bayesian CNB model provides the "best estimate" of future loss payments, the analysis in this paper suggested that there are some insurers that post

Figure 17. Analysis of subsequent reserve changes for 109 insurers



reserves conservatively, while others post reserves with a downward bias. Readers should exercise caution in generalizing these conclusions beyond this particular line of business in this time period.

While this paper did not address the tail when a significant amount of losses are to be paid after 10 years, I do have some suggestions on how to use this approach when the tail extends beyond 10 years. First, the Dev_{Lag} parameters in each ω could be extended further based on either analysis or the judgment of the insurer. The parameters for the later lags may differ even if the parameters for the earlier lags are identical. The data in Schedule P contains the sum of losses paid for all prior years. If the insurer has premium information from the prior years, the methods described in this paper could be used to calculate the distribution of that sum conditional on each model ω, and hence contribute to the estimation of the posterior probability of each model ω.

Table 3. Summary statistics for the plots above

	Reported Reserve @ 1995	
	< Predictive Mean (000)	> Predictive Mean (000)
Number of Insurers	66	43
Total Predictive Mean	926,134	872,660
1995 Reserve @ 1995	803,175	1,173,124
1995 Reserve @ 2001	856,393	985,711

I view this paper as an initial attempt at a new method for stochastic loss reserving. To gain general acceptance, this approach should be tested on other lines of insurance and by other researchers. The data required to do such studies consist of Schedule Ps that American insurers are required to report to regulators, and claim severity distributions. This information can be obtained from vendors. AM Best compiles the

Schedule P information. ISO fits claim severity distributions for many lines of insurance.

This method requires considerable statistical and actuarial expertise to implement. It also takes a lot of work. In this paper, I have tried to make the case that we should expect that such efforts could yield fruitful results.

References

[1] Barnett, G., and B. Zehnwirth, "Best Estimate for Reserves," *Proceedings of the Casualty Actuarial Society* 87, 2000, pp. 245–321, http://www.casact.org/pubs/proceed/proceed00/00245.pdf.

[2] Clark, D. R., "LDF Curve Fitting and Stochastic Loss Reserving: A Maximum Likelihood Approach," Casualty Actuarial Society *Forum*, Fall 2003, pp. 41–91, http://www.casact.org/pubs/forum/03fforum/03ff041.pdf.

[3] Clark, D. R. "Variance and Covariance Due to Inflation," Casualty Actuarial Society *Forum*, Fall 2006, pp. 61–95.

[4] Grace, M. F., and J. T. Leverty, "Property Liability Insurer Reserve Error-Motive, Manipulation, or Mistake," (working paper: Contact the authors at mgrace@gsu.edu or ty-leverty@uiowa.edu for the most recent version).

[5] Hayne, R. M., "Measurement of Reserve Variability," Casualty Actuarial Society *Forum*, Fall 2003, pp. 141–172, http://www.casact.org/pubs/forum/03fforum/03ff141.pdf.

[6] Heckman, P. E., and G. G. Meyers, "The Calculation of Aggregate Loss Distributions for Claim Severity and Claim Count Distributions," *Proceedings of the Casualty Actuarial Society* 70, 1983, pp. 41–61, addendum in 71, 1984, pp. 49–56, http://www.casact.org/pubs/proceed/proceed83/83022.pdf; http://www.casact.org/pubs/proceed/proceed84/84049.pdf.

[7] Insurer Solvency Working Party, *A Global Framework for Insurer Solvency Assessment*, 2004, International Actuarial Association, http://www.actuaries.org/LIBRARY/Papers/ Global_Framework_Insurer_Solvency_Assessment-members.pdf.

[8] Klugman, S. A., H. H. Panjer, and G. E. Willmot, *Loss Models: From Data to Decisions*, Hoboken, NJ: Wiley, 2004, http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471215775.html.

[9] Mack, T., "Distribution-Free Calculation of Chain Ladder Reserve Estimates," *ASTIN Bulletin* 23:2, November 1993, pp. 213–226, http://www.casact.org/library/astin/vol23no2/213.pdf.

[10] Meyers, G. G., "The Common Shock Model for Correlated Insurance Losses," Casualty Actuarial Society *Forum*, Winter 2006, pp. 231–254, http://www.casact.org/pubs/forum/06wforum/06w235.pdf.

[11] Mildenhall, S. J., "Correlation and Aggregate Loss Distributions with an Emphasis on the Iman-Conover Method," Casualty Actuarial Society *Forum*, Winter 2006, pp. 103–205, http://www.casact.org/pubs/forum/06wforum/06w107.pdf.

[12] Stanard, J. N., "A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques," *Proceedings of the Casualty Actuarial Society* 72, 1985, pp. 124–148, http://www.casact.org/pubs/proceed/proceed85/85124.pdf.

[13] Wang, S., "Aggregation of Correlated Risk Portfolios: Models and Algorithms," *Proceedings of the Casualty Actuarial Society* 85, 1998, pp. 848–939, http://www.casact.org/pubs/proceed/proceed98/980848.pdf.

Appendix

This appendix gives the mathematical details that implement the methodologies described in Sections 3 and 4.

A.3.1. Discretizing the claim severity distributions

The first step is to determine the discretization interval length h. I chose h so that the 2^{14} (16,384) values spanned the probable range of annual losses for the insurer. Specifically, let h_1 be the sum of the insurer's ten-year premium divided by 2^{14}. The h was set equal to 1,000 times the smallest number from the set $\{5, 10, 20, 25, 40, 50, 100, 125, 200, 250, 500, 1000\}$ that was greater than $h_1/1000$. This last step guarantees that a multiple, m, of h would be equal to the policy limit of 1,000,000.

The next step is to use the mean-preserving method (described in [8], p. 656) to discretize the claim severity distribution for each settlement lag. Let $p_{i,Lag}$ represent the probability of a claim with severity $h \cdot i$ for each settlement lag. Using the limited average severity (LAS_{Lag}) function determined from claim severity distributions

provided by ISO, the method proceeds in the following steps.

1. $p_{0,Lag} = 1 - LAS_{Lag}(h)/h$.
2. $p_{i,Lag} = (2 \cdot LAS_{Lag}(h \cdot i) - LAS_{Lag}(h \cdot (i-1)) - LAS_{Lag}(h \cdot (i+1)))/h$ for $i = 1, 2, \ldots, m-1$.
3. $p_{m,Lag} = 1 - \sum_{i=0}^{m-1} p_{i,Lag}$.
4. $p_{ik} = 0$ for $i = m+1, \ldots, 2^{14}-1$.

A.3.2. Calculating the conditional density of the CNB distribution

The purpose of this section is to show how to calculate $\mathrm{CNB}(x_{AY,Lag} \mid \mathrm{E}[Paid\ loss_{AY,Lag}])$. The calculation proceeds in the following steps.

1. Set $\vec{\mathbf{p}}_{Lag} = \{p_{0,Lag}, \ldots, p_{2^{14}-1,Lag}\}$.
2. Calculate the Fast Fourier Transform (FFT) of $\vec{\mathbf{p}}_{Lag}, \Phi(\vec{\mathbf{p}}_{Lag})$.
3. Calculate the expected claim count $\lambda_{AY,Lag}$ for each accident year and settlement lag using Equation 2, $\lambda_{AY,Lag} \equiv E[Paid\ Loss_{AY,Lag}]/E[Z_{Lag}]$.
4. Calculate the FFT of each aggregate loss random variable $X_{AY,Lag}$ using the formula

$$\Phi(\vec{\mathbf{q}}_{AY,Lag}) = (1 - c \cdot \lambda_{AY,Lag} \cdot (\Phi(\vec{\mathbf{p}}_{Lag}) - 1))^{-1/c}.$$

   This formula is derived in KPW [8, Equation 6.28]. Note the different but equivalent parameterization. The probability generating function for the negative binomial distribution is given in Appendix B of KPW [8]. It is written as $P_N(z) = (1 - \beta(z-1))^{-r}$. In this paper's notation $\lambda = \beta \cdot r$ and $c = 1/r$.
5. Calculate $\vec{\mathbf{q}}_{AY,Lag} = \Phi^{-1}(\Phi(\vec{\mathbf{q}}_{AY,Lag}))$, the inverse FFT of the expression in Step 4 above.
6. Set i equal to the multiple of h that is nearest to $x_{AY,Lag}$. Then

$$\mathrm{CNB}(x_{AY,Lag} \mid \mathrm{E}[Paid\ loss_{AY,Lag}]) = \text{the } i\text{th component of } \vec{\mathbf{q}}_{AY,Lag}.$$

Note that calculating this probability requires one to first calculate a vector of length 16,384 by inverting an FFT and reading off a single component. (To increase efficiency, one should calculate $\Phi(\vec{\mathbf{p}}_{Lag})$ for each settlement lag in advance.) Using the R computing language (www.r-project.org) on my 3 GHz personal computer with 1 GB RAM, I estimate it takes about 1/20th of a second to evaluate a single CNB probability. Evaluating a likelihood for a loss triangle with 55 $x_{AY,Lag}$s 1,000 times (typical for what follows below) takes about 45 minutes. Implementing this methodology requires the patience that I was fortunate to develop in the early days of actuarial computing.

A.4. Maximizing the likelihood for the CNB model

The purpose of this section is to show how to find the ELR and $\{Dev_i\}$ parameters that maximize the likelihood

$$L(\{x_{AY,Lag}\}) = \prod_{AY=1}^{10} \prod_{Lag=1}^{11-AY} CNB(x_{AY,Lag} \mid E[Paid\ Loss_{AY,Lag}]), \tag{4}$$

subject to the following constraints in the Dev_{Lag} parameters.

1. $Dev_1 \le Dev_2$.
2. $Dev_j \ge Dev_{j+1}$ for $j = 2, 3, \ldots, 7$.
3. $Dev_7/Dev_8 = Dev_8/Dev_9 = Dev_9/Dev_{10}$.
4. $\sum_{i=1}^{10} Dev_i = 1$.

The maximization was done using the R programming language *optim* function using the Nelder-Mead parameter search method. This method is described in KPW [8, p. 664] and is considered to be robust but slow. At this stage of the research, I value "robust" over "fast."

Primarily because of habits I developed using Excel Solver, I elected not to use standard constraints provided by the function. Instead I coded a "*tdev to Dev*" function that mapped all of $\mathcal{R}^9$ (nine dimensional unit hypercube) into a subset

of $\mathcal{R}^{11}$ that satisfied those constraints. Here is a description of *tdev to Dev*.

1. $Dev'_1 = e^{-tdev_1^2}/2$.
2. $Dev'_2 = Dev'_1 \cdot (1 + e^{-tdev_2^2})$.
3. $Dev'_i = Min\left[\left(1 - \sum_{j=1}^{i-1} Dev'_j\right), Dev'_{i-1}\right] \cdot e^{-tdev_i^2}$ for $i = 3, \ldots, 7$.
4. $Dev'_i = Min\left[\left(1 - \sum_{j=1}^{i-1} Dev'_j\right), Dev'_{i-1}\right] \cdot e^{-tdev_8^2}$ for $i = 8, 9, 10$.
5. $Dev_i = Dev'_i / \sum_{j=1}^{10} Dev'_j$.
6. $ELR = tdev_9^2 \cdot \sum_{j=1}^{10} Dev'_j$.

As noted in the previous section, the CNB model requires a lot of time to calculate. This time can be significantly reduced if one has a good set of starting values for the *optim* function. To get these starting values, I replaced the CNB distribution with the "overdispersed Poisson" (ODP) distribution given in Clark [2] to find the *ELR* and $\{Dev_i\}$ parameters that maximize the logarithm of following expression.

$$L(\{x_{AY,Lag}\}) = \prod_{AY=1}^{10} \prod_{Lag=1}^{11-AY} E[Paid\ Loss_{AY,Lag}] \cdot e^{x_{AY,Lag} \cdot E[_{AY,Lag}] - E[_{AY,Lag}]}.$$

The maximization proceeds in the following steps.

1. Pick a starting vector in $\vec{\mathbf{s}} \in \mathcal{R}^9$, e.g., $(1, 1, 1, 1, 1, 1, 1, 1, 1)$.
2. Set $\vec{\mathbf{t}} = tdev\ to\ Dev(\vec{\mathbf{s}})$ and use it to calculate $E[Paid\ Loss_{AY,Lag}]$.
3. Use $E[Paid\ Loss_{AY,Lag}]$ to calculate the ODP likelihood above.
4. Use the Nelder-Mead algorithm to calculate an updated vector $\vec{\mathbf{s}}$.
5. Return to Step 2 and repeat until convergence.
6. After convergence is obtained with the *ODP* likelihood, use the current $\vec{\mathbf{s}}$ as a starting value for the CNB likelihood in Equation 4.
7. Set $\vec{\mathbf{t}} = tdev\ to\ Dev(\vec{\mathbf{s}})$ and use it to calculate $E[Paid\ Loss_{AY,Lag}]$.
8. Use $E[Paid\ Loss_{AY,Lag}]$ to calculate the CNB likelihood above.
9. Use the Nelder-Mead algorithm to calculate an updated vector $\vec{\mathbf{s}}$.
10. Return to Step 7 and repeat until convergence.
11. Set $\vec{\mathbf{t}} = tdev\ to\ Dev(\vec{\mathbf{s}})$ to obtain the maximum likelihood estimate of *ELR* and $\{Dev_i\}$.

Run time was short for the ODP. For the CNB, I found that it generally took, on average, 1,000 iterations of Steps 7–10 to achieve R's *optim* function default convergence criteria. With the warning that individual results may vary, I felt comfortable in limiting the number of iterations to 300.