

# Thinking Outside the Triangle

By

Glenn Meyers, FCAS, MAA, Ph.D.

---

## Abstract

*Overfitting is the result of having model that is too complex for the amount of data that is available. This is typically the case when for a loss reserve model which has a large number of parameters on a smallish triangle of data. This paper uses simulation methods to diagnose overfitting in estimating the predictive distribution of loss reserves by the method of maximum likelihood. This paper then shows how to use a Bayesian fitting methodology to overcome overfitting using prior information that is “outside the triangle.”*

## Key Words

Reserving Methods, Overfitting, Reserve Variability, Uncertainty and Ranges, Bayesian Estimation

## 1. Introduction

A characteristic of models that are typically fit to triangles of loss reserve data is that the models have a large number of parameters, and these models are fitted to a relatively small data set. For example, in fitting a chain ladder model to a triangle with ten years of data, one must estimate nine parameters with only 55 data points<sup>1</sup>. This paper gives an example where such fitting practices lead to overfitting and thus understates the variability or the resulting estimates of ultimate losses. It then illustrates a Bayesian method that overcomes this problem. A feature of this method is that it requires the modeler to specify a prior distribution of its parameters. To do this, one must use information that is “outside the triangle.”

A single example cannot invalidate all loss reserving methods that use models with a large number of parameters. But the methodology used to illustrate the overfitting is applicable to other examples. The fitting method is that of maximum likelihood. This is a

---

<sup>1</sup> This assumes that the losses are all paid at the end of ten years. If we drop this assumption, either additional parameters or assumptions must be added.

well established method of fitting loss reserve models<sup>2</sup>. See, for example, Clark [2003]. I would like to think that the example is sufficiently realistic and general to raise serious doubts about the practice of fitting models with a large number of parameters to small triangles without resorting to outside information.

Experienced loss reserving actuaries almost always emphasize the judgmental nature of loss reserving. A Bayesian methodology such as that illustrated in this paper makes the necessary judgments transparent.

## **2. Overfitting**

In the context of this paper, “overfitting” means that the model is too complex for the amount of data that is available. The predictions made by an overfit model can be wildly inaccurate. A common prescription for overfitting is to either use a simpler model or get more data. If a simpler model was generally applicable to the loss reserving problem, I believe that someone would have discovered it by now. Obtaining more data is problematic. There is a general belief among actuaries that each loss reserving situation is unique, and considerable judgment has to be made in deciding what, if any, additional data can be used in developing loss reserve estimates. The unattractiveness of the common remedies to overfitting leads us to a third alternative – quantify the errors made by overfitting.

There has been a lot of work done on quantifying the effects of overfitting for the normal distribution and it is instructive to review this now. While much of what follows in this review could be presented analytically, I chose to present this work in terms of a simulation algorithm since that is what follows later in the analysis of a loss reserving methodology.

---

<sup>2</sup> Section 12.3 of Klugman, Panjer and Willmot [2004] describes the asymptotic properties of the maximum likelihood estimator. A quick summary of the section is that “under mild regularity conditions” the maximum likelihood estimator works very well for large samples. But for small samples, all bets are off.

Given a sample  $x_1, \dots, x_n$  from a normal distribution, it is well known that the maximum likelihood estimates of the parameters  $\mu$  and  $\sigma$  satisfy the following equations.

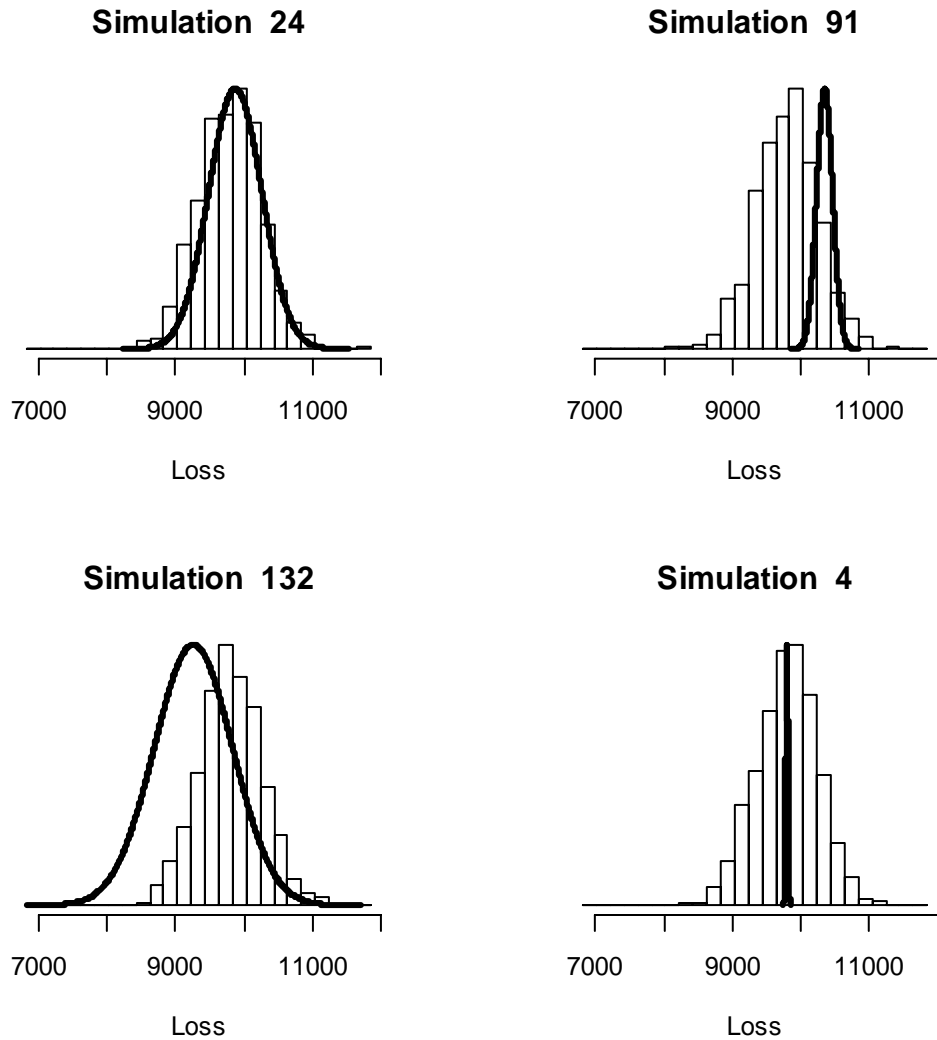
$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{and} \quad \hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2} \quad (2.1)$$

Let us perform the following simulation.

1. Select a random training sample of size  $n$  from a normal distribution with parameters  $\mu$  and  $\sigma$ .
2. Find the maximum likelihood estimates,  $\hat{\mu}$  and  $\hat{\sigma}$  from Equation 2.1 based on this sample.
3. Select a random testing sample of size  $m$  from the same normal distribution with parameters  $\mu$  and  $\sigma$ .
4. Test to see if a normal distribution with parameters  $\hat{\mu}$  and  $\hat{\sigma}$  predicts the distribution of the second sample.

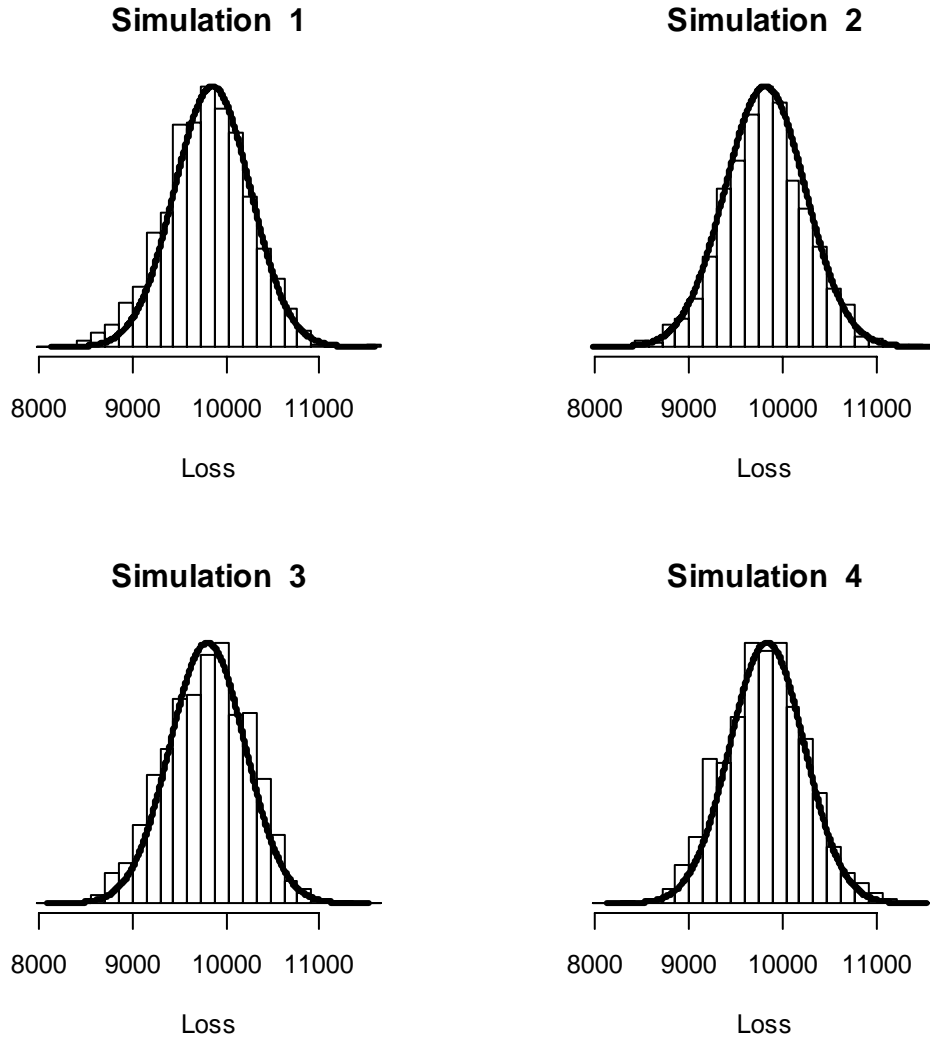
Since the testing sample size,  $m$ , can be arbitrarily large, the standard tests of statistical significance that are typically used for real data are not particularly helpful. Thus this paper uses graphical tests. For a single estimate, the test will consist of a plot of the fitted density function from Step 2 and a histogram from the sample taken in Step 3. Figure 1 illustrates possible effects of overfitting when the training sample is small. Figure 2 illustrates what happens when the training sample size,  $n$ , is sufficiently large.

**Figure 1**  
**Simulation of Maximum Likelihood Fits**  
**With Small Training Samples**



- These plots were selected from a set of 150 fits of simulated data with  $\mu = 10,000$  and  $\sigma = 500$ . The maximum likelihood estimators were based on a training sample consisting of three observations. The histograms were based on the testing sample of 1,000 observations simulated from the same model.
- The particular plots were selected to illustrate what can, and frequently does, happen when overfitting. In general, the estimated means are highly volatile and the standard deviation is often underestimated.

**Figure 2**  
**Simulation of Maximum Likelihood Fits**  
**With Large Training Samples**



- These plots were selected from a set of 150 fits of simulated data with  $\mu = 10,000$  and  $\sigma = 500$ . The maximum likelihood estimators were based on a training sample consisting of 100 observations. The histograms were based on the testing sample of 1,000 observations simulated from the same model.
- The remaining 146 plots looked similar to the first four. With a sufficiently large sample, maximum likelihood estimation does not lead to overfitting.

At this point let us introduce a second test of the predicted distribution known as a p-p plot<sup>3</sup>. This test consists of calculating the sorted percentiles of the testing sample implied by the estimate of the distribution, and plotting these percentiles against the points  $\left(\frac{1}{m+1}, \frac{2}{m+1}, \dots, \frac{m}{m+1}\right)$ . If this plot lies along the line with slope one and passing through the origin, one concludes that the percentiles are uniformly distributed and that the sample passed the test.

Figures 1P and 2P give the p-p plots using the same simulated data that was used in Figures 1 and 2 above.

Now rather than test the predictions of a series of individual model fits, let us do a combined test of the fitting procedure itself. To do this, one puts the predicted percentiles from all the fits into a single vector and constructs a p-p plot. Figure 3 contains a combined p-p plot for simulated data underlying Figures 1 and 1P.

Figure 3 also contains an attempt to fix the downward bias of the maximum likelihood estimator for the variance. When fitting a normal distribution to a small sample, statisticians often recommend replacing the maximum likelihood estimate of the variance with its unbiased estimate given by:

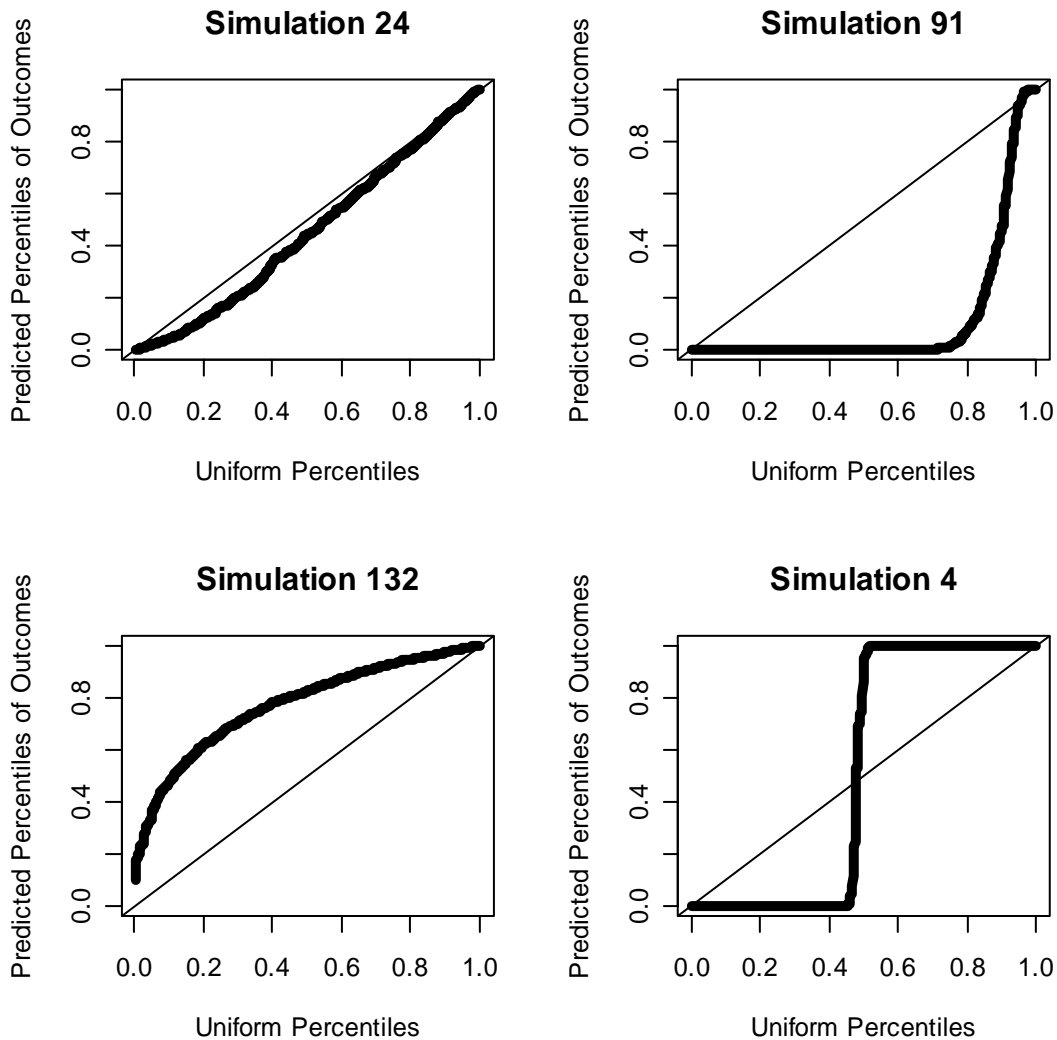
$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\mu})^2 = \frac{n}{n-1} \hat{\sigma}^2. \quad (2.2)$$

The second p-p plot in Figure 3 tests the combined fits using the parameters  $\hat{\mu}$  and  $s$  for the mean and standard deviation of the fitted distribution.

---

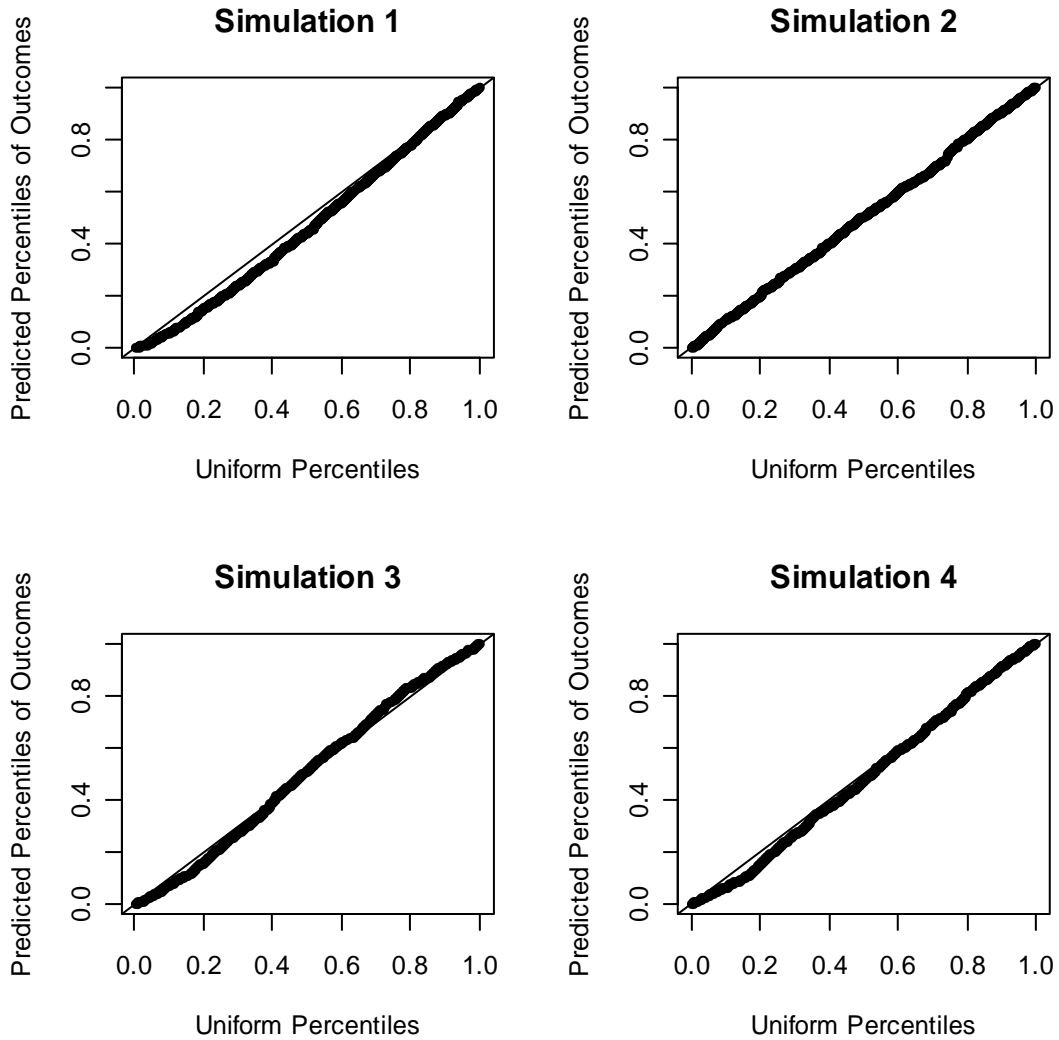
<sup>3</sup> pp plots are discussed more fully in Klugman *et. al.* [2004], p. 424.

**Figure 1P**  
**P-P Plots of Maximum Likelihood Fits**  
**With Small Training Samples**



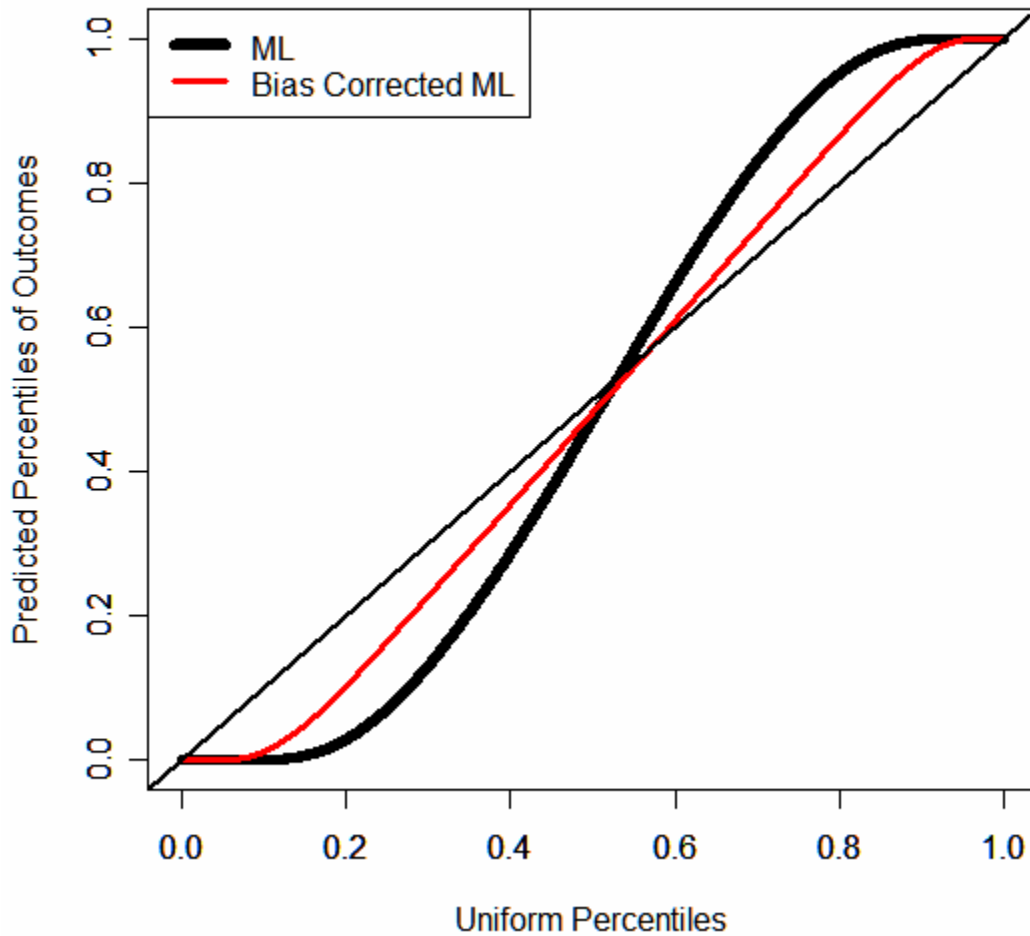
- Simulation 24 is close to what we should expect of a good fit. The remaining simulations are far from what we should expect for an adequate fit. Most of the data is: (1) at the very low predicted percentiles for Simulation 91; (2) at the high predicted percentiles for Simulation 132; and (3) at the extreme tails for Simulation 4.

**Figure 2P**  
**P-P Plots of Maximum Likelihood Fits**  
**With Large Training Samples**



- The p-p plots line up pretty much as expected. There is some random variation in the fit for a sample size of size one hundred.

**Figure 3**  
**Combined P-P Plots of Maximum Likelihood Fitting Methodologies**  
**With the Small Training Samples**



- This plot shows that both the maximum likelihood and the bias corrected maximum likelihood fitting methods underestimate the tails of the testing sample since there are too many observations from the testing samples out in the tails. For example, about 25% of the test observations are above the predicted 90<sup>th</sup> percentile.
- One could consider the “tilted S curve” such as we see on the above p-p plot, as a signature of overfitting.

### 3. Bayesian Model Fitting

This section continues with the example of fitting normal distributions discussed in the last section. The philosophy underlying the last section could be stated as follows. “We believe that the data comes from a single underlying model, but we don’t know the parameters. Let us choose the model with the single set of parameters that best describes the data we have.” In contrast, the philosophy of this section is stated as follows. “We believe that the data could come from a number of plausible models. Let us choose our model to be a mixture of all the plausible models, with the mixing probabilities being influenced by the data we have.”

The formula for calculating the mixing probabilities is Bayes’ Theorem, which in general terms can be stated as follows<sup>4</sup>.

$$\Pr\{\text{Model}|\text{Data}\} \propto \text{Likelihood}\{\text{Data}|\text{Model}\} \cdot \Pr\{\text{Model}\} \quad (3.1)$$

The term “ $\Pr\{\text{Model}\}$ ” is called the prior distribution and represents information that the modeler has in addition to the data. The term “ $\Pr\{\text{Model}|\text{Data}\}$ ” is called the posterior distribution and represents the information that the model has after allowing the influence of the data.

Let us now continue with the normal example of the previous section. In place of the single  $\mu = 10,000$ , let  $\mu_j$  run from 9,000 to 11,000 in increments of 200. In place of the single  $\sigma = 500$ , let  $\sigma_j$  run from 250 to 750 in increments of 50. This makes up the set of 121 “plausible models.” Since this is an illustrative example, we can be lazy and use a prior distribution that assigns equal probability to each model. Given a training sample,  $x_1, \dots, x_n$ , the likelihood of the of the data is given by:

$$L_{ij}\{x_1, \dots, x_n\} = \prod_{k=1}^n \phi(x_k | \mu_j, \sigma_j) \quad (3.2)$$

---

<sup>4</sup> In many texts, the statement of Bayes’ Theorem has a denominator on the right hand side that contributes to the constant of proportionality in this statement of Bayes’ Theorem. This paper uses the more compact statement with the understanding that at the end, the posterior probabilities must be normalized so that they add to one.

where  $\phi(x_k | \mu_i, \sigma_j)$  is the density function of the normal distribution with mean  $\mu_i$  and standard deviation  $\sigma_j$ .

Since each model is equally likely, the posterior probability,  $p_{ij}$ , of each model is given by:

$$p_{ij} = \frac{L_{ij}(x_1, \dots, x_n)}{\sum_{i=0}^{10} \sum_{j=0}^{10} L_{ij}(x_1, \dots, x_n)} \quad (3.3)$$

The posterior probability density function, also known as the predictive distribution,  $f(x)$ , is a posterior probability weighted mixture of the normal distributions that make up the original set of “plausible models.”

$$f(x) = \sum_{i=0}^{10} \sum_{j=0}^{10} p_{ij} \cdot \phi(x | \mu_i, \sigma_j) \quad (3.4)$$

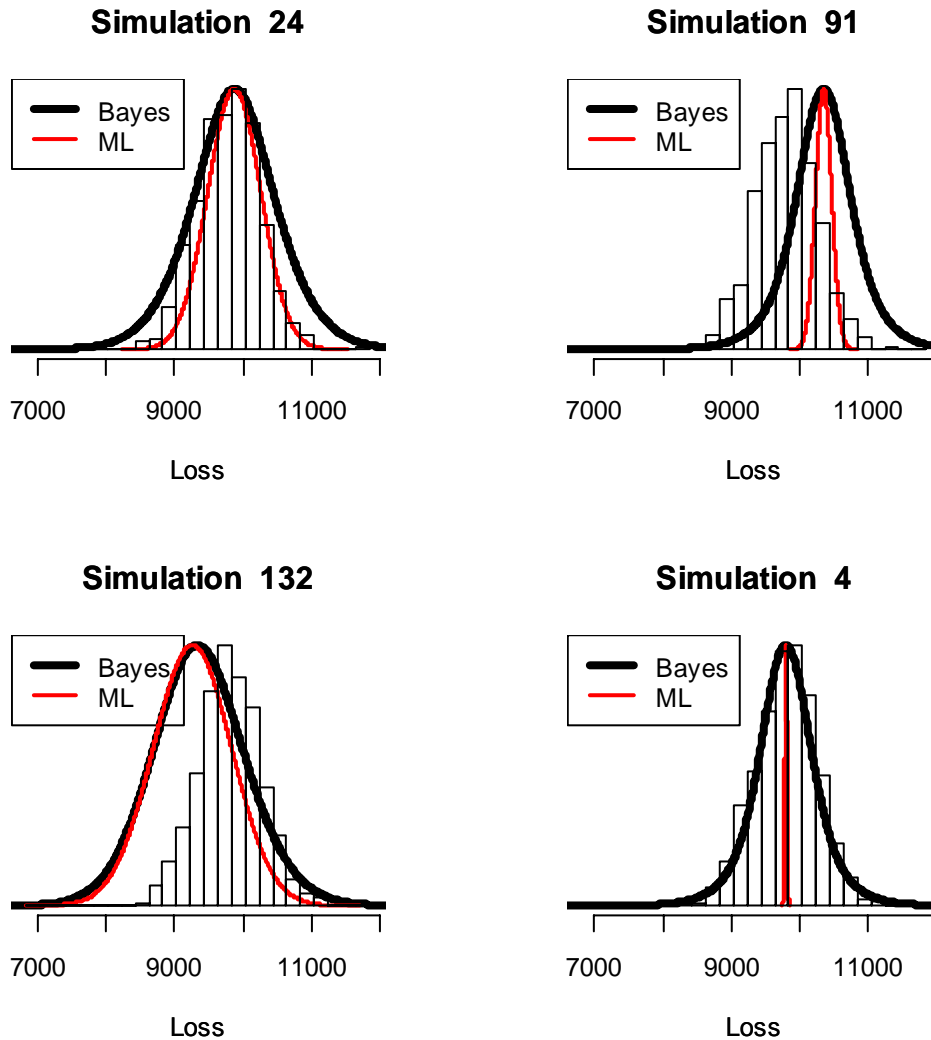
Similarly, one can calculate the cumulative predictive distribution function:

$$F(x) = \sum_{i=0}^{10} \sum_{j=0}^{10} p_{ij} \cdot \Phi(x | \mu_i, \sigma_j) \quad (3.5)$$

where  $\Phi(x_k | \mu_i, \sigma_j)$  is the cumulative distribution function of the normal distribution with mean  $\mu_i$  and standard deviation  $\sigma_j$ .

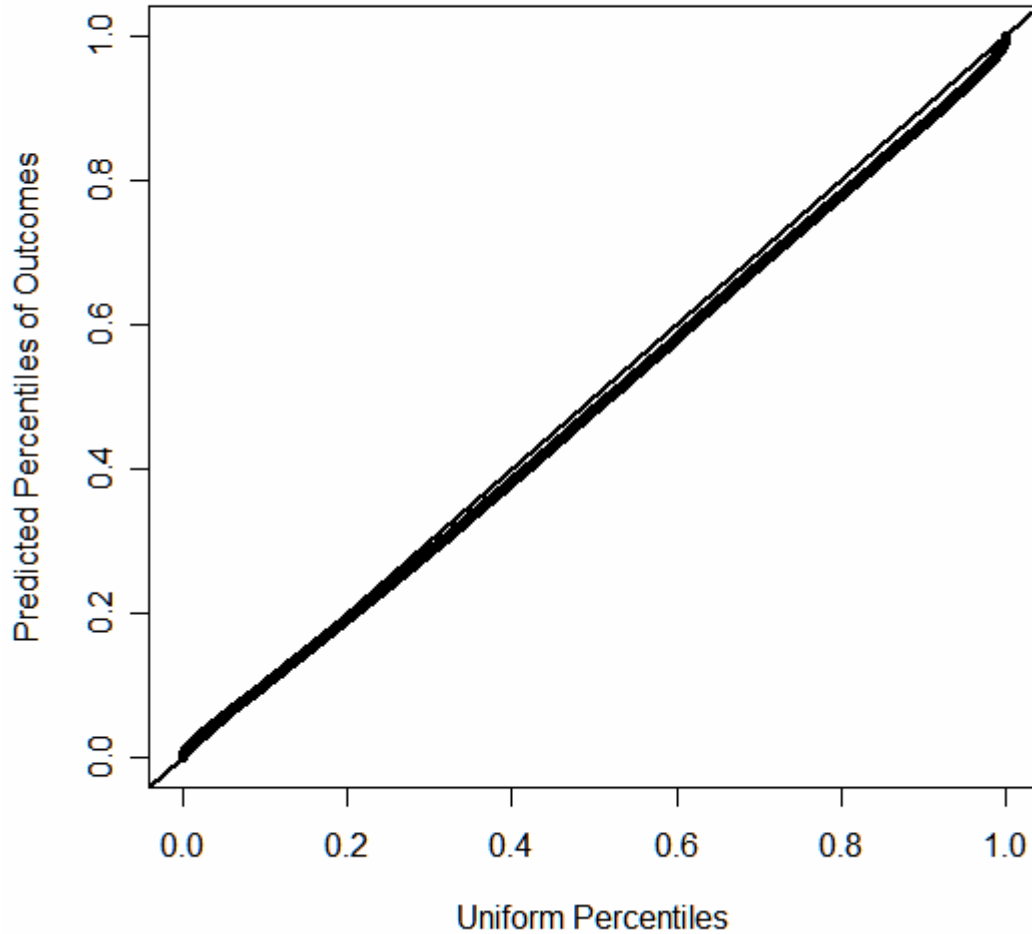
Figure 4 gives the predictive density function superimposed over the simulations that make up Figure 1 for the small training samples. Figure 5 gives the p-p plot for the combined Bayesian fits which looks very good. Let us think about what this means. One takes a training sample, denoted by  $i$ , and estimates the  $p_i^{\text{th}}$  percentile. In repeating this experiment  $t$  times, one can expect to find that the test simulations will be below the  $p_i^{\text{th}}$  percentile  $p\%$  of the time, *when calculated globally over all  $t$  experiments*. Figure 4 shows that this may not be the case for each individual experiment,  $i$ .

**Figure 4**  
**Simulation of Bayesian and Maximum Likelihood Fits**  
**With Small Training Samples**



- The simulated losses were identical to those in Figure 1.
- Over the 150 simulations, the average absolute difference between the predicted mean and the true mean was 216 for the maximum likelihood estimate of the mean and 206 for the Bayesian estimate of the mean.
- The standard deviations of the predictive distributions were less volatile than the maximum likelihood estimates of the standard deviations.

**Figure 5**  
**Combined P-P Plot of the Bayesian Fitting Methodology**  
**With the Small Training Samples on a Fixed Population**



- The simulated losses were identical to those in Figure 3.
- Comparing this plot with Figure 3 shows that the Bayesian fitting methodology yields better predictions of the testing samples. The maximum likelihood fits do not recognize uncertainty in the  $\mu$  and  $\sigma$  parameters while the Bayesian fits does recognize the uncertainty in  $\mu$  and  $\sigma$ .

Before moving on to loss reserving examples, let us consider one more example with the normal distribution. The examples above draw their simulated data from a population with  $\mu = 10,000$  and  $\sigma = 500$ . Suppose that the population is heterogeneous and consists of 121 equally likely subpopulations with  $\mu$  ranging from 9,000 to 11,000 by increments of 200 and  $\sigma$  ranging from 250 to 750 by increments of 50. This matches the range of plausible models in the Bayesian analysis above.

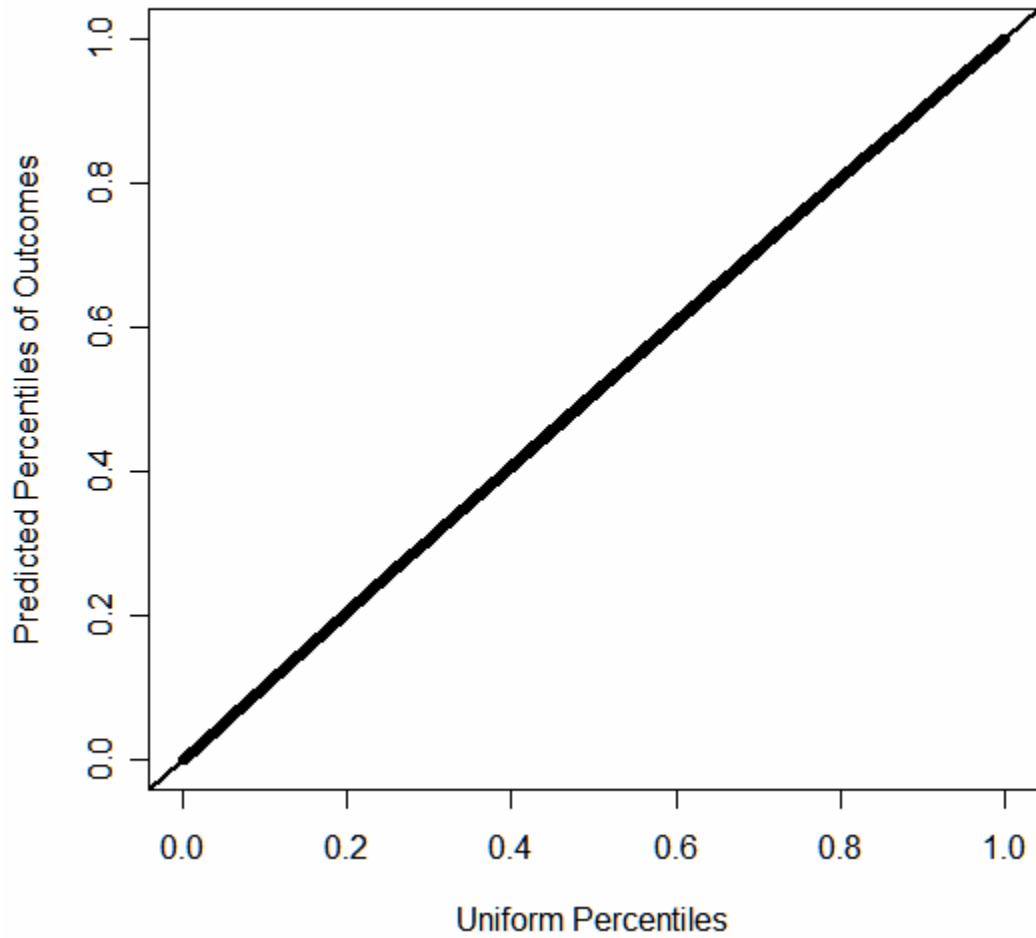
Let us pick a subpopulation at random and use a training sample to fit the distribution by the Bayesian method described above. Tests of individual fits yield results similar to Figure 4 above and will not be shown here. The salient feature of the individual fits is that they provide noticeably better estimates of the variability of the test sample. Figure 6 shows the p-p plot for the combined results which, like Figure 5, look very good. As was the case with the fixed population underlying Figure 5, the excellent fit is a global result.

Over the 150 simulations, the average absolute difference between the predicted mean and the true subpopulation mean was 240 for the maximum likelihood estimate of the mean and 197 for the Bayesian estimate of the mean.

When compared to the maximum likelihood fitting methods, the better local performance of the Bayesian fitting methods contributes to their excellent global performance.

I would like to offer some concluding remarks this and the prior section. Most actuaries will have no problem with the idea that fitting a normal distribution with only three points will lead to overfitting. Many of the ideas expressed above for normal distributions can be made more precise in an analytic framework. My purpose in discussing simulations was to show a way to graphically identify overfitting with this tool. Let us now turn to identifying overfitting for loss reserve models where the analytic results have not been worked out.

**Figure 6**  
**Combined P-P Plot of the Bayesian Fitting Methodology**  
**With Small Training Samples Taken From Random Subpopulations**



#### 4. A Loss Reserve Model

This section begins with a complete description of a loss reserve model. The loss reserve “data” used in this paper will be simulated from assumed distributions. The model specified will be similar to that described by Meyers [2006]. In that paper, Meyers fits his model to real insurer data and demonstrates that the predictive distribution of loss outcomes describes the distribution of outcomes on a testing dataset. Thus I believe it is a realistic model to use in this paper’s simulation tests.

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

$$E\left[ \text{Paid Loss}_{AY,Lag} \right] = \text{Premium}_{AY} \times ELR \times Dev_{Lag}, \quad (4.1)$$

where:

- $AY = 1, 2, \dots, 10$  is an index for accident year.
- $Lag = 1, 2, \dots, 10$  is the settlement lag reported after the beginning of the accident year.
- $\text{Paid Loss}$  is the incremental paid loss for the given accident year and settlement lag.
- $\text{Premium}$  is the earned premium for the accident year.
- $ELR$  is an unknown parameter that represents the expected loss ratio.
- $Dev_{Lag}$  is an unknown set of parameters that depend on the settlement lags. In all there are seven independent components with the last three components constrained to be equal

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 described as follows.

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

$$\lambda_{AY,Lag} \equiv E\left[ \text{Paid Loss}_{AY,Lag} \right] / E\left[ Z_{Lag} \right]. \quad (4.2)$$

- 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  $\lambda_{AY,Lag} + c \cdot \lambda_{AY,Lag}^2$ .
- Then the random variable  $X_{AY,Lag}$  is defined by

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

The  $c$  parameter of the claim count distribution is set equal to 0.01. The claim severity distribution is set as a Pareto distribution with the cumulative distribution equal to:

$$F(\tilde{x}) = 1 - \left( \frac{\theta}{\tilde{x} + \theta} \right)^\alpha \quad (4.3)$$

with  $\alpha = 2$  and  $\theta = 10, 25, 50, 75, 100, 125, 150, 150, 150$  and  $150$  respectively as the development lag runs from 1 to 10. The units are in thousands of dollars. This choice of parameters gives the model the property that the more risky claims require a longer time to settle.

To use this model, one must solve are two mathematical problems that require the use of Fast Fourier Transforms.

1. While the above defines how to express the random variable,  $X_{AY,Lag}$ , in terms of other random variables  $N_{AY,Lag}$  and  $Z_{Lag}$ , we will need to calculate the likelihood of observing  $x_{AY,Lag}$  using the triangle of data for which  $AY + Lag \leq 11$ .
2. In order to predict the distribution of outcomes for the total reserve we will need to

calculate the distribution of  $\sum_{AY=2}^{10} \sum_{Lag=AY}^{10} X_{AY,Lag}$ .

The details of how to solve these problems this are in the technical appendix of Meyers [2006].

Now that the model is specified, let us take the next steps of simulating training and test samples, and comparing the maximum likelihood and Bayesian predictive distributions with the test samples.

In order to do the Bayesian fitting methodology, we have to specify the prior distributions for the unknown *ELR* and  $Dev_{Lag}$  parameters. Table 1 provides the prior distribution of the *ELR* parameter.

**Table 1**  
**Prior Probabilities for *ELR***

<i>ELR</i>	Prior	<i>ELR</i>	Prior	<i>ELR</i>	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

The prior distribution of loss development paths,  $Dev_{Lag}$  for *Lag* running from 1 to 10 were generated from the formula for a beta distribution:

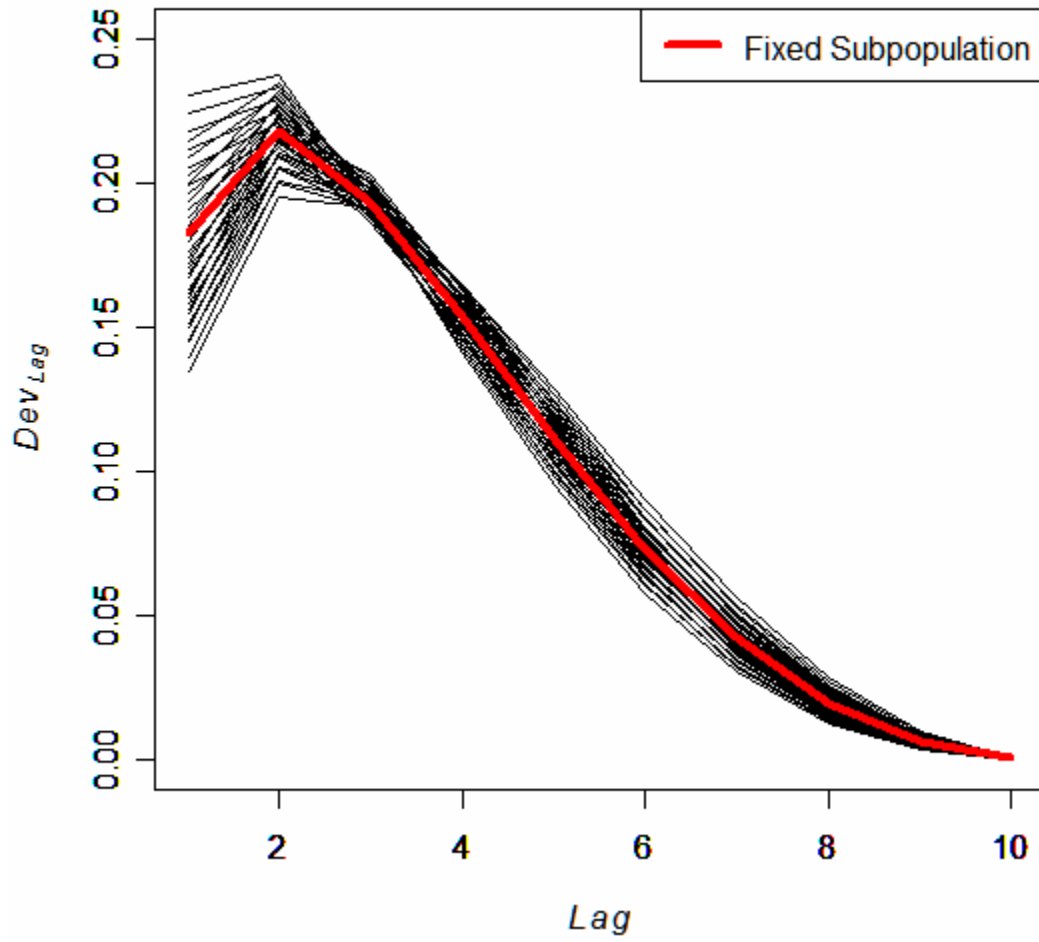
$$Dev_{Lag}(a, b) = \frac{\int_0^{Lag/10} t^{a-1} (1-t)^{b-1} dt}{\int_0^1 t^{a-1} (1-t)^{b-1} dt} \quad (4.4)$$

where *a* ran from 1.25 to 1.50 in increments of 0.05 and *b* ran from 3.25 to 3.75 in increments of 0.10. Figure 7 is a plot of all thirty-six development paths. Each path was assigned a prior probability of 1/36.

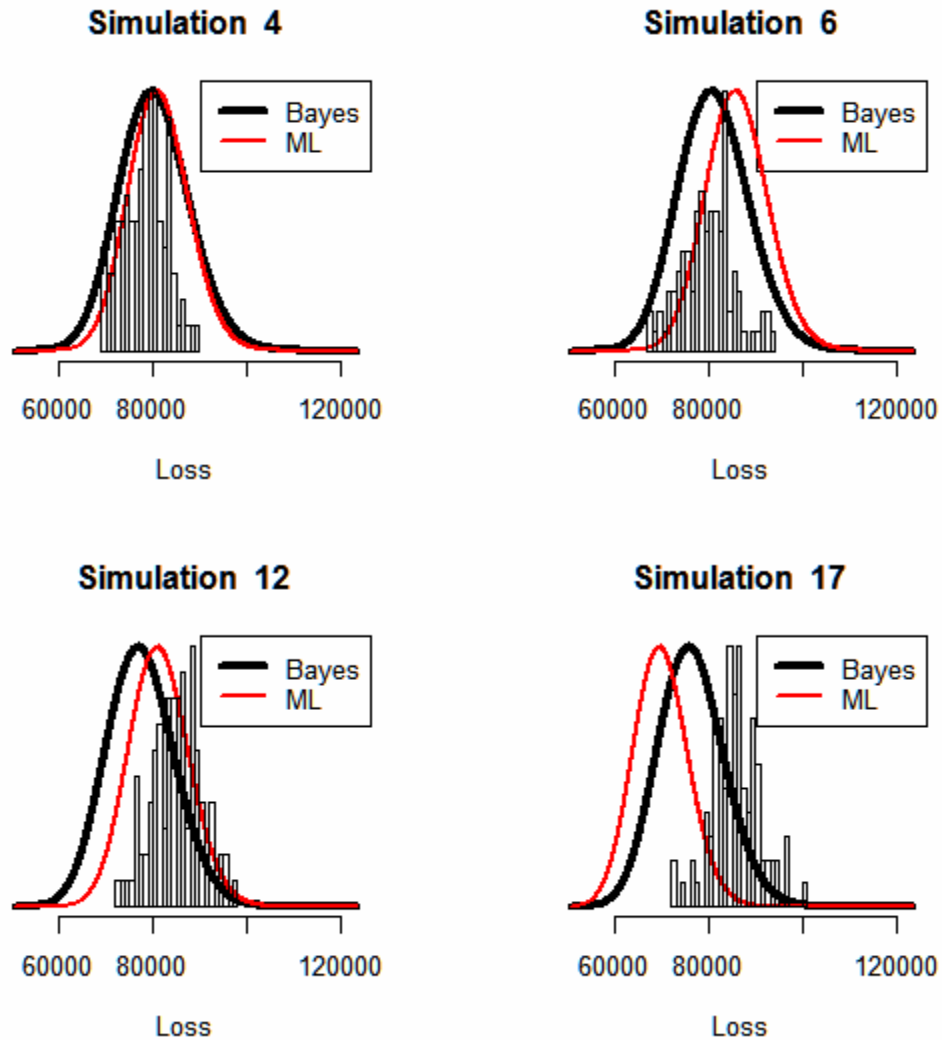
As was done in the previous section, Bayesian fits will be done for a fixed subpopulation ( $Dev_{Lag}$  calculated with  $a = 1.45$ ,  $b = 3.45$ , and  $ELR = 0.700$ ) and the fits will also be done for randomly selected subpopulations. The annual premium used in these examples was set equal to 50,000 (in units of 1,000's.)

Figures 8-12 show the results of simulations of the maximum likelihood and Bayesian fitting methodologies on this model.

Figure 7  
Prior Payout Patterns



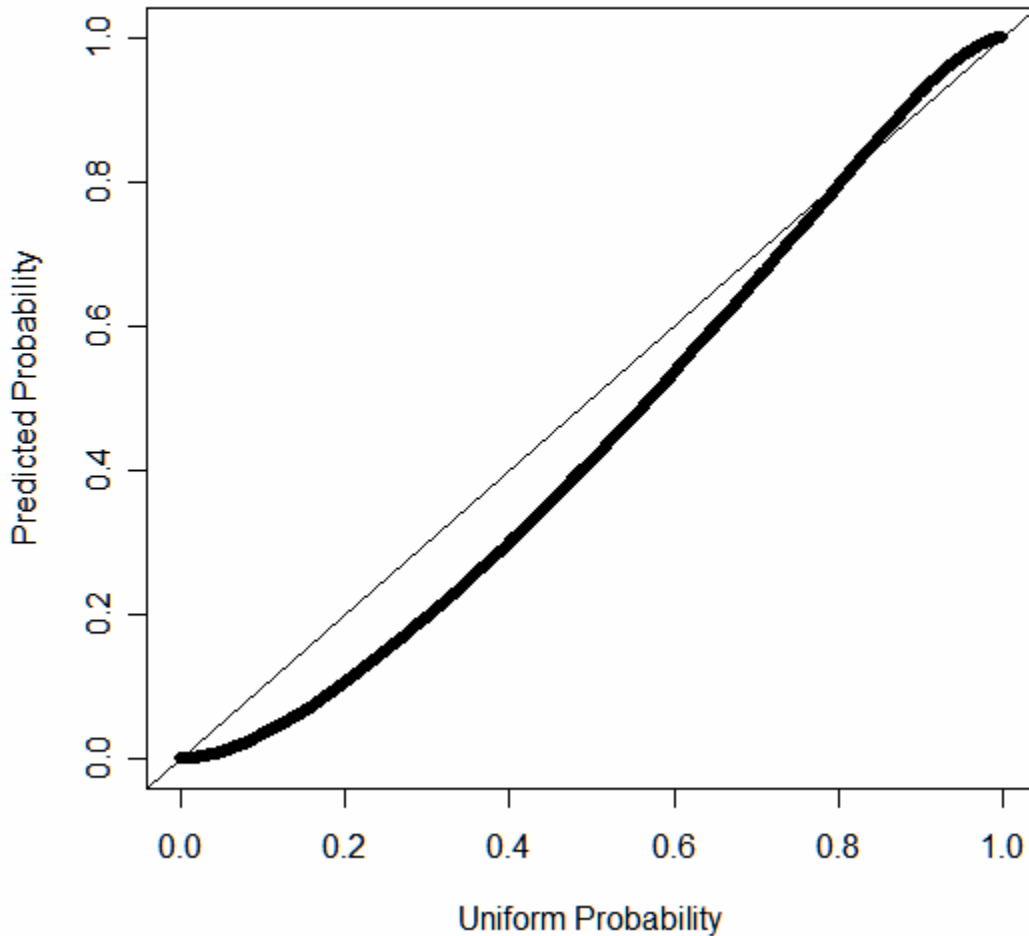
**Figure 8**  
**Simulation of Bayesian and Maximum Likelihood Fits**  
**For Fixed Subpopulation**



- These plots were selected from a set of 150 fits of simulated data from the fixed subpopulation. Both the Bayesian fits and the maximum likelihood fits were based on the same training sample consisting of a single triangle of 55 observations. These fits were used to predict the total outstanding losses.
- The histograms were based on the total outstanding losses of 100 testing samples simulated from the same model.
- The Bayesian fits generally indicated more volatility than the maximum likelihood fits as they recognize uncertainty in the parameters of the model.

**Figure 9**

**Combined P-P Plot of the Maximum Likelihood Fitting Methodology  
With Training Samples Selected From the Fixed Subpopulation**

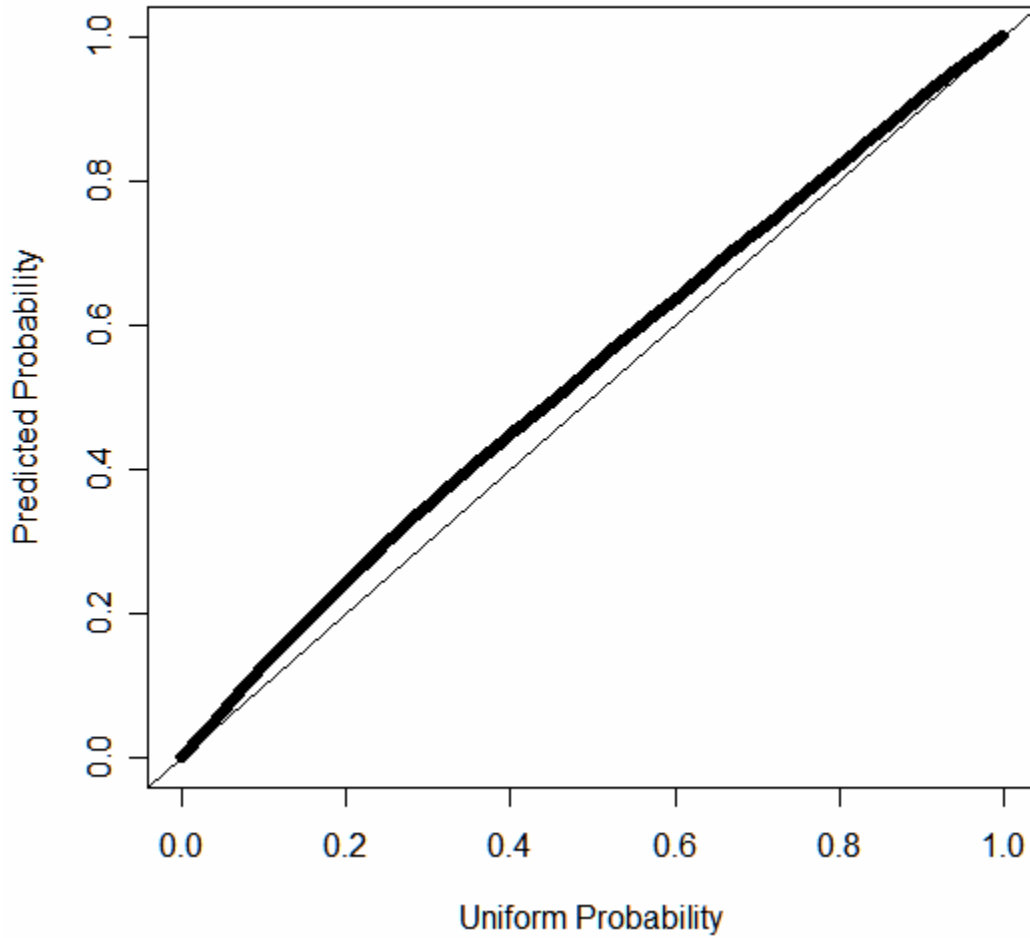


- This p-p plot has the tilted S curve that is characteristic of overfitting.
- There are fewer than expected observations predicted at the lower percentiles. Thus the maximum likelihood fitting methodology has an upward bias in these simulated results.
- The mean absolute error in the reserve estimate for the maximum likelihood fitting methodology was 1,860. This is significantly higher than the mean absolute error for the Bayesian fitting methodology of 1,389.

- 

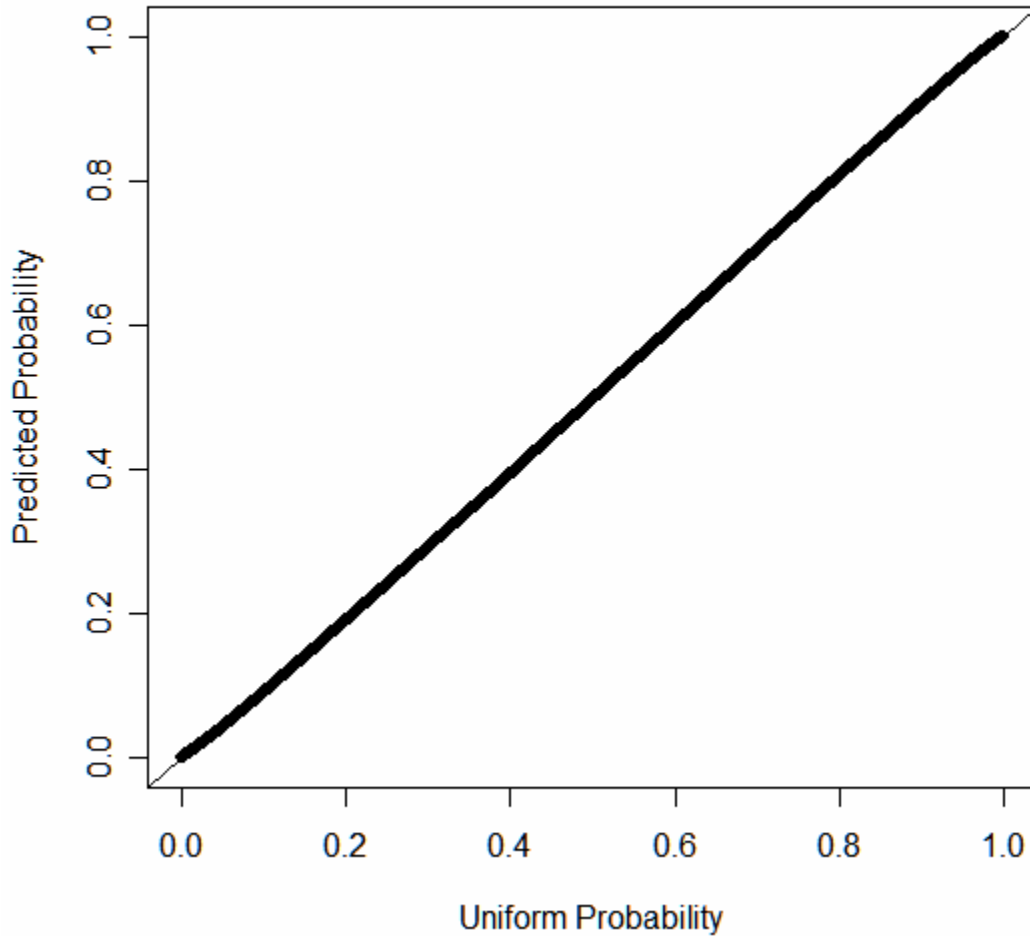
**Figure 10**

**Combined P-P Plot of the Bayesian Fitting Methodology  
With Training Samples Selected From the Fixed Subpopulation**



- Most observations had a higher than expected predicted percentile. This means that this Bayesian fitting methodology has a slight downward bias for this simulation.
- The downward bias could be attributed to the fact that the subpopulation's expected loss ratio is 0.70 while the lower expected loss ratio received greater weight in the prior distributions specified in Table 1 above. Bayesian statisticians would argue that this is the correct result if you really believe your prior distribution.
- The mean absolute error in the reserve estimate for the Bayesian fitting methodology was 1,389. This is significantly lower than the mean absolute error for the maximum likelihood fitting methodology of 1,860.

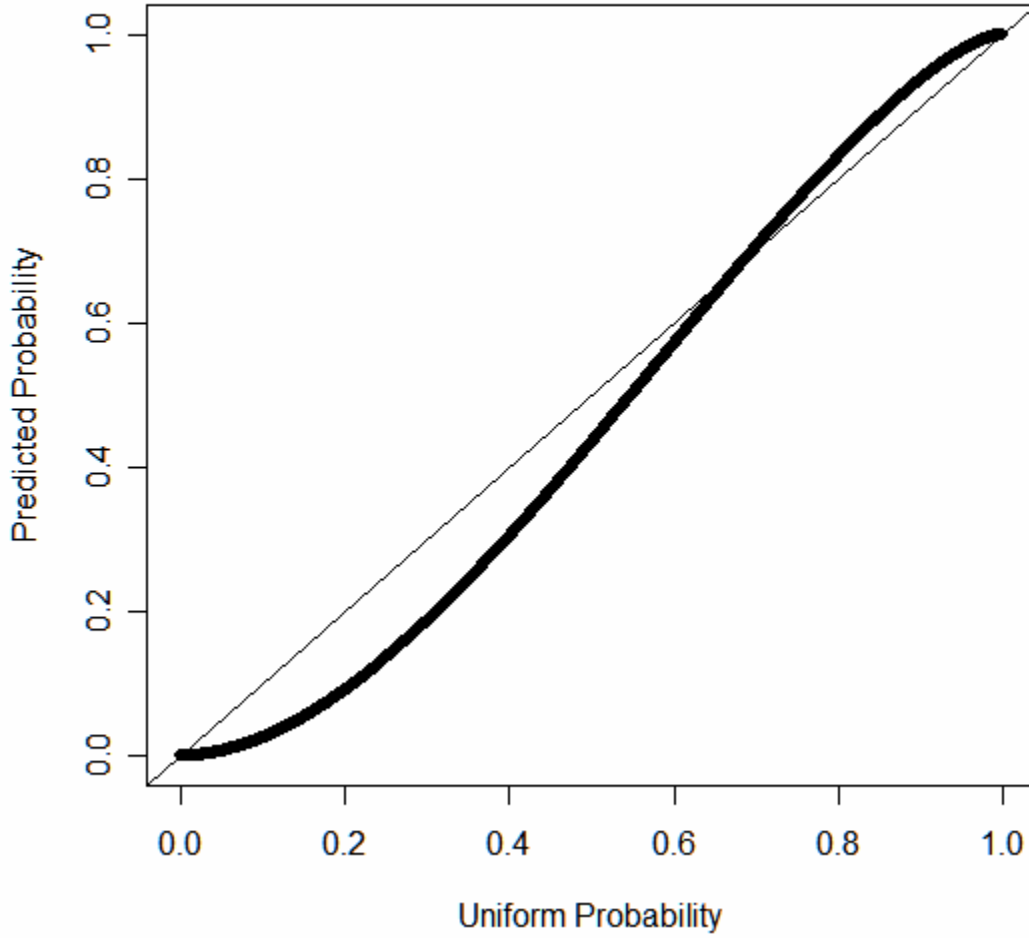
**Figure 11**  
**Combined P-P Plot of the Bayesian Fitting Methodology**  
**With Training Samples Selected at Random From All Subpopulations**



- The predicted percentiles are distributed almost exactly as expected. While Figure 10 shows that there may be some bias for selected subpopulations, Figure 11 shows that the Bayesian fitting methodology is unbiased in the aggregate for this simulation.
- The mean absolute error in the reserve estimate for the Bayesian fitting methodology was 2,003. This is significantly lower than the mean absolute error for the maximum likelihood fitting methodology of 4,665.

Figure 12

**Combined P-P Plot of the Maximum Likelihood Fitting Methodology  
With Training Samples Selected at Random From All Subpopulations**



- This p-p plot has the tilted S curve that is characteristic of overfitting.
- The asymmetry of the plot indicates that there was an upward bias in the predicted distribution.
- The mean absolute error in the reserve estimate for the maximum likelihood fitting methodology was 4,665. This is significantly higher than the mean absolute error for the Bayesian fitting methodology of 2,003.

## 5. Conclusion

This paper uses simulation models to analyze the effects of overfitting. Starting with an example with the normal distribution where the effects of overfitting are generally understood, it illustrates graphical tests to identify the effects of overfitting. An overfit model will typically understate the true volatility of process under consideration. The “tilted S curve” on the p-p plot for a testing sample is an indication of overfitting.

Maximum likelihood is a classic statistical fitting methodology known to perform well on large datasets. In the typical loss reserve setting we have a case where we fit a model with many parameters to a relatively small dataset. This paper fits a model with eight parameters to a dataset with 55 observations. Simulations on test samples obtain the tilted S curve on the p-p plots. In addition, the simulations show an upward bias for the maximum likelihood fitting methodology for the model considered in the paper.

This paper then proposed a Bayesian fitting methodology as an alternative to maximum likelihood. This methodology depends on external information on the range of plausible models. One example shows that depending on where a particular sample is drawn from the range of plausible models, the predictive distribution could be biased, but if one is truly uncertain on where a particular sample belongs in the range, one should expect to obtain a predictive distribution that fits well on a testing sample.

The Bayesian methodology depends upon getting the right prior information. Meyers [2006] illustrates one way to get this external information on live data. He studied the commercial auto Schedule P data that is insurers report on NAIC Annual Statements. He first noted the following.

1. Maximum likelihood estimates of the  $Dev_{Lag}$  parameters on large insurers were noticeably more stable than maximum likelihood estimate on small insurers.
2. Aggregated empirical payment patterns were similar for large and small insurers.

The maximum likelihood estimates of the  $Dev_{Lag}$  and  $ELR$  parameters of 40 large insurers were used to form the set of plausible models in a Bayesian analysis on a set of 109 insurers for which there were six subsequent years available as testing samples. The combined p-p plot on the testing samples followed, within the bounds of sampling errors, a

straight line. This situation appears to be similar to that underlying the construction of Figure 11 in which each insurer (or training sample) was selected at random from the population with  $Dev_{Lag}$  and  $ELR$  parameters represented by the 40 large insurers.

The calculations and the exhibits in this paper were coded in R. The code and the R workspaces are available from the author at [gmeyers@iso.com](mailto:gmeyers@iso.com).

## References

1. David R. Clark, "LDF Curve Fitting and Stochastic Loss Reserving: A Maximum Likelihood Approach," *CAS Forum*, Fall 2003, pp 41-91.  
<http://www.casact.org/pubs/forum/03fforum/03ff041.pdf>
2. Stuart A. Klugman, Harry H. Panjer and Gordon E. Willmot, *Loss Models: From Data to Decisions*, 2004, Wiley Series in Probability and Statistics.  
<http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471215775.html>
3. Glenn G. Meyers, "Estimating Predictive Distributions for Loss Reserve Models," *CAS Forum*, Fall 2006, pp 159-203.  
<http://www.casact.org/pubs/forum/06fforum/163.pdf>