Casualty Actuarial Society
E-Forum, Winter 2017

The CAS *E-Forum*, Winter 2017

The Winter 2017 edition of the CAS *E-Forum* is a cooperative effort between the CAS *E-Forum* Committee and various other CAS committees, task forces, or working parties. This *E-Forum* contains two independent research papers.

CAS *E-Forum*, Winter 2017

Table of Contents

E-Forum Committee

For information on submitting a paper to the *E-Forum*, visit http://www.casact.org/pubs/forum/.

Minimum Bias, GLMs and Credibility in the Context of Predictive Modeling

Christopher Gross and Jonanthan Evans

Abstract:

When predictive performance testing, rather than testing model assumptions, is used for validation, the needs for detailed model specification are greatly reduced. Minimum bias models trade some degree of statistical independence in data points in exchange for statistically much more tame distributions underlying individual data points. A combination of multiplicative minimum bias and credibility methods for predictively modeling losses (pure premiums, claim counts, and/or average severity, etc.) based on explanatory risk characteristics is defined. Advantages of this model include grounding in longstanding and conceptually lucid methods with minimal assumptions. An empirical case study is presented with comparisons between multiplicative minimum bias and a typical generalized linear model (GLM). Comparison is also made with methods of incorporating credibility into GLM.

Keywords: predictive modeling, minimum bias, credibility, ratemaking, generalized linear models

1. INTRODUCTION

As predictive models that relate losses (pure premiums, claim counts, and/or average severity, etc.) to explanatory risk characteristics become ever more commonplace, some of the practical problems that frequently emerge include:

- Models often use complex techniques that are effectively "black boxes" without a lucid conceptual basis.
- Models may require very detailed parametric or distributional assumptions. Invalid assumptions may result in biased parameters.
- A highly Frequentist approach, usually involving Maximum Likelihood Estimation (MLE), can lead to overfitting sparsely populated data bins.

Some longstanding methods can be combined to overcome these problems:

- Minimum Bias Iterative fitting of parameters is simple, longstanding in practice, and non-parametric in specification.
- Credibility methods are similarly simple and longstanding. Credibility directly solves the sparse bin problem.

Most importantly, properly done predictive testing, in contrast with testing model assumptions, makes highly detailed model specification generally unnecessary.

1.1 Research Context

The minimum bias criteria and iterative solution methodology were introduced by Bailey and Simon in [2] and [3]. Brown in [5] substituted the minimum bias criteria with MLE of Generalized Linear Models (GLM), an approach further explored by Mildenhall in [10]. Venter in [13] further discusses credibility issues related to minimum bias methods. The basic contemporary reference on credibility methods is Klugman, S., et al. [9]. Nelder and Verrall in [11] and Klinker in [8] discuss incorporating random effects into GLM to implement credibility adjustments. Brosius and Feldblum provide a modern practical guide to Minimum Bias Methods in [4]. A similar practical guide to GLM is provided by Anderson, et al. in [1]. A demonstration of predictive model fitting and testing can be found in Evans and Dean [6], particularly the predictive testing methods that will be used in this paper. "Gibbs Sampling" is a term we will use for Markov Chain Monte Carlo (MCMC) methods, as these are implemented using Gibbs Sampling software, such BUGS, WinBUGS, or JAGS. Scollnik in [12] introduces MCMC. Particularly relevant to this paper is the recent book on predictive modeling for actuaries Frees, E., et al. [7]. This book contains very detailed information on GLM, particularly incorporating credibility through Gibbs Sampling. This paper represents in a certain sense an opposite perspective from [7] and [12], by emphasizing very simple models combined with rigorous predictive testing as described in [6].

1.2 Outline

The remaining sections of this paper are:

2. Predictive Performance as the Modeling Objective

3. Multiplicative Minimum Bias Iteration

4. Incorporating Credibility

5. Anchoring And Iteration Blending For Practical Iterative Convergence

6. Testing of Individual Explanatory Variables

7. Empirical Case Study

8. Summary Discussion

Appendix A. Details of Empirical Case Study

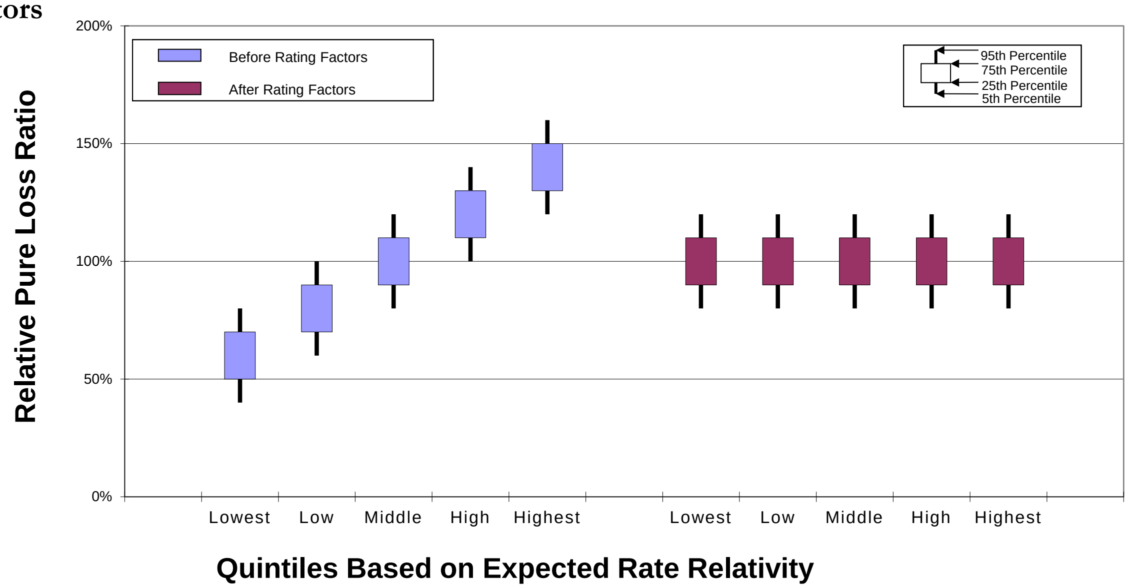
Appendix B. Gibbs Sampling Model Code

2. PREDICTIVE PERFORMANCE AS THE MODELING OBJECTIVE

Traditionally, statistical models tend to use the same data for both fitting and validation. Validation tends to involve testing of model assumptions. For example, a linear regression of the form Y = m X + b + ξ, where ξ ~ Normal(0, σ^2), might be fit, using least squares, to a set of data points (x_i, y_i), i =1,…,n. Validation tests would check to verify that the residuals ξ_i are normally distributed with constant variance and are independent of x_i, y_i, and each other. Hypothesis tests would then be performed to confirm that the probability is sufficiently remote that the actual data set would result if m = 0 or b = 0 (null hypotheses). This framework relies on detailed assumptions, without which validation testing would not be possible.

Modern predictive models split available data into multiple sets for separate fitting and validation. In the previous example, the parameters m and b might be fit to the points (x_i, y_i), i =1 ,…, k , using any method, and then tested on the points (x_i, y_i), i =k+1 ,…, n. The test would only be concerned with how well $\hat{y}_i = \hat{m}\, x_i + \hat{b}$ predicts y_i for the test set. A bootstrap quintile test might be used, where the validation points are sorted by the value $\hat{y}_i$ into 5 equal-sized groups. The average value of y_i should ascend with the quintile groups and for each group the average value of y_i should be close to the average value of $\hat{y}_i$. Figure 1 is a hypothetical example of a quintile test, with bootstrap confidence intervals added, as described by Evans and Dean in [6], validation of rating factors. Note, the assumption ξ ~ Normal(0, σ^2) and other implicit assumptions of linear regression are unnecessary here.

Figure 1. Hypothetical Example of Bootstrap Quintile Test Predictive Validation of Rating Factors



In practice, predictive modelers often split data into three or more sets (i.e., training, testing, and validation), but only the distinction between two separate data sets for fitting and validation will be covered in this paper.

In the predictive framework, detailed model assumptions are not necessary. A model, even if its assumptions seem unjustified or erroneous, is valid as long as it performs well at predicting outcomes for data that were not used to fit its parameters. This comes with the caveat that care must be taken that both the fitting and validation data should be representative of – effectively random samples of – the loss process. For example, predictive testing might be misleading if both the fitting and validation data occurred in a single year influenced by a somewhat rare catastrophe, such as a hurricane.

3. MULTIPLICATIVE MINIMUM BIAS ITERATION

Suppose the basic data available consists of actual losses $L_{i_1,\ldots,i_n} \geq 0$ and exposures $P_{i_1,\ldots,i_n} \geq 0$,($P_{i_1,\ldots,i_n} = 0 \quad \Rightarrow \quad L_{i_1,\ldots,i_n} = 0$) where $i_j = 1,\ldots,n_j$ indexes the individual classes within the classification dimension *j* and $i_1,..,i_n$ denotes the cell corresponding to the intersection of a single class in each classification dimension. Also the total exposure in any class is positive, $\sum_{i_j=k} P_{i_1,\ldots,i_n} > 0$, otherwise it would make sense to exclude the class entirely from estimating rating parameters. A multiplicative minimum bias model assumes that $L_{i_1,\ldots,i_n} = B_{i_1,\ldots,i_n} + P_{i_1,\ldots,i_n} \prod_{j=1,\ldots,n_j} X_{j,i_j}$`. The parameters X_{j,i_j} are fit with the goal of minimizing some bias function, or functions, of the residual errors $B_{i_1,\ldots,i_n}$.

The minimum bias goal is that the sum of the residual errors for each class $\sum_{i_j=k} B_{i_1,\ldots,i_n}$ should be 0. A corresponding iterative sequence of parameter estimates can be formed whose convergence corresponds to convergence to the goal:

$$X_{j,k,1} = 1$$

$$X_{j,k,t+1} = \frac{\sum_{i_j=k} L_{i_1,\ldots,i_n}}{\sum_{i_j=k} P_{i_1,\ldots,i_n} \prod_{l \neq j} X_{l,i_l,t}} \tag{3.1}$$

The effective sample is now $\sum_{j=1,\ldots,n} n_j$ data points with values $\sum_{i_j=k} L_{i_1,\ldots,i_n}$, which reduces to $\sum_{j=1,\ldots,n} n_j - (n-1)$ linearly independent numbers. There is a corresponding $(n-1)$ dimensional degeneracy in the parameters. If the parameters X_{k,i_k} are multiplied by a constant $c > 0$ and the parameters X_{l,i_l} are divided by c, where $0 \leq k < l \leq n$, then $\prod_{j=1,\ldots,n_j} X_{j,i_j}$ will be unchanged.

The Central Limit Theorem implies that the distribution of $\sum_{i_j=k} L_{i_1,\ldots,i_n}$ can be expected to more closely resemble a Normal distribution, with a generally lower coefficient of variation than the individual cell values $L_{i_1,\ldots,i_n}$. However, whereas the cellular values $L_{i_1,\ldots,i_n}$ can reasonably be assumed to be statistically independent of each other, the aggregated values $\sum_{i_j=k} L_{i_1,\ldots,i_n}$ include many statistical dependencies since there is an overlap of cells between classes in different dimensions. So, a tradeoff is made for a minimum bias iteration model. Statistical independence of sample data points, a desirable property, is partially sacrificed in exchange for the benefit of a more Normal distribution, generally having a lower coefficient of variation than the distributions underlying each sample data point. This taming of the distribution of data points means that it becomes less necessary to specify the distribution of the individual cellular loss values, or as may be the case the distributions of individual loss observations within the cells, as would be necessary for a GLM.

Example 1

Suppose there are three classification dimensions, each with 10 classes, resulting in 1,000 individual cells. We can expect about 100 times as much data underlying each class as for each cell, and correspondingly an average coefficient of variation by class that is only about 10% as much as by cell. Two classes in different dimensions overlap in 10 cells and thus actual losses between them will have a correlation coefficient of about 10%.

Multiplicative minimum bias effectively aims toward the same parameters estimates as a GLM with a logarithmic link function and Poisson likelihood function. The logarithmic link converts the sum of linear explanatory factors into a multiplicative product of their exponentials. The Poisson likelihood leads to equations for MLE that correspond to a fixed limit point of the minimum bias iteration, as pointed out by Brown in [5].

However, the Poisson distributional assumption is usually unrealistic and not a part of the minimum bias model. Data are generally not restricted to integer values. The Poisson coefficient of variation (CV) is not scale independent (it is 10 times greater when applied to dollar amounts versus when applied to the same amounts measured as pennies) and implodes for large nominal means (mean of 1,000,000 implies a CV of 0.1%). So, the Poisson assumption is important only in the optimization equations it implies for MLE.

4. INCORPORATING CREDIBILITY

Credibility adjustments $0 \le Z_{j,i_j} \le 1$ can be easily and directly incorporated into the iteration equations:

$$X_{j,k,1} = 1$$

$$X_{j,k,t+1} = (1 - Z_{j,k}) + Z_{j,k} \frac{\sum_{i_j = k} L_{i_1,\ldots,i_n}}{\sum_{i_j = k} P_{i_1,\ldots,i_n} \prod_{l \ne j} X_{l,i_l,t}} \tag{4.1}$$

Note, other than the constraint of the interval [0, 1], nothing has been specified about the determination of $Z_{j,i}$. There are many possibilities for Z_{j,i_j}, including functions of the sum of exposure $P_{j,k} = \sum_{i_j = k} P_{i_1,\ldots,i_n}$. The ultimate test will be the predictive performance of the final model regardless of whether $Z_{j,i}$ itself satisfies any traditional goals of credibility theory, such as limiting fluctuation or greatest accuracy.

For GLM, the basic and common protection against fitting parameters to data that is not credible is to throw away explanatory variables whose parameters are not statistically distinct from 0, those variables with high p-values.

To add a true credibility, or "shrinkage", adjustment is complicated. The two main approaches are:

1. General Linear Mixed Models. At least some rating factors are assumed to be random rather than fixed effects, but an MLE-like fitting method is still used. Numerical solution is rather difficult and, in practice, functions in R or procedures in SAS are used, very much as black boxes. See [7], [8] and [11] for background.
2. Bayesian Networks and Gibbs Sampling. Rating factors in each class dimension follow a prior distribution. The parameters of the prior distributions follow distributions that are very diffuse. Numerical solution is performed using a Gibbs Sampling program, such as JAGS or WinBUGS. The model itself is elaborately specified and lucid to an audience sophisticated enough read the specification. See [7] and [12] for background.

In Section 7, we will demonstrate an example of the second approach.

5. ANCHORING AND ITERATION BLENDING FOR PRACTICAL ITERATIVE CONVERGENCE

In practice the convergence of the iterative algorithms can be a problem even after the application of credibility. For one thing there is still the problem of $(n-1)$ dimensional degeneracy previously mentioned. Also, highly correlated dimensions can also contribute to non-convergence or slow convergence in practice. Other than the automatic degeneracy we will not attempt to deal with the more general convergence issue in a precise mathematical way, which appears to be an open problem for multiplicative minimum bias. From a practical point of view *anchoring* and *iteration blending* can effectively provide timely convergence.

Anchoring directly eliminates the degeneracy. One approach is to fix one of the parameters in each of $(n-1)$ dimensions to the value of 1.0; or to fix such a parameter in each of n dimensions and add a single overall base rate parameter. Another approach is to use a single overall base rate and rescale the parameters in each dimension to a weighted average of 1.0 at the end of each iteration.

Example 2

If $P=\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}$ and $L=\begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix}$ then parameter iterations will oscillate back and forth between the values $X=\begin{pmatrix}1.5 & 3.5\\ 2.0 & 3.0\end{pmatrix}$ and $X=\begin{pmatrix}0.6 & 1.4\\ 0.8 & 1.2\end{pmatrix}$. However, if we "anchor" one parameter at 1.00

the iterations will converge to $X = \begin{pmatrix} 1.000 & 2.333 \\ 1.200 & 1.800 \end{pmatrix}$.

Iteration blending can be implemented to accelerate convergence by modifying the iterative equations to be:

$$X_{j,k,t+1} = \alpha\left[(1-Z_{j,k}) + Z_{j,k}\frac{\sum_{i_j=k} L_{i_1,\ldots,i_n}}{\sum_{i_j=k} P_{i_1,\ldots,i_n}\prod_{l\neq j} X_{l,i_l,t}}\right] + (1-\alpha)X_{j,k,t+1} \tag{5.1}$$

where $0 < \alpha < 1$ is a selected constant blending parameter.

As an extreme illustration of correlation, let one classification dimension be replicated or made once redundant. Setting $\alpha = 0.5$ will allow the model to converge. Each one of the replicated dimensions will end up sharing equally in the observed predictive relationship, combining together to provide the appropriate prediction. In the case of full credibility, they will exactly reproduce the result from not replicating the dimension. With less than full credibility, the result will not be exactly the same from not replicating the dimension, but will be similar.

6. TESTING OF INDIVIDUAL EXPLANATORY VARIABLES

Sometimes predictive modeling techniques are used specifically to determine whether or not individual explanatory variables, or equivalently classification dimensions, are statistically significant. As mentioned earlier, when using GLM techniques, it is common to consider the p-values of the estimated parameters. These p-values are calculated under the distributional and other assumptions, such as independence of the GLM model being used.

Whether distributional assumptions are made (as with GLM) or not (as with minimum bias), tests of predictive performance can be performed and compared with and without a given classification dimension. In cases where the improvement is insignificant the dimension should be removed for the sake of parsimony.

7. EMPIRICAL CASE STUDY

The empirical data used in this case study consists of 371,123 records of medical malpractice payments obtained from the National Practitioner Data Bank. Three explanatory variables will be used for modeling payment amounts: *Original Year*, *Allegation Group* and *License Field*. The records will be randomly split into two sets for model fitting and validation, respectively. Further details are included in Appendix A.

7.1 GLM Model Specifications

For our GLM model we will consider:

1. The logarithmic link function, which causes the fit factors to act multiplicatively.
2. Several likelihood functions: Gaussian, Poisson, Gamma, and Inverse Gaussian. These correspond to assumptions that variance σ^2 is related to mean μ as σ^2 = constant, $\sigma^2 \propto \mu$, $\sigma^2 \propto \mu^2$, and $\sigma^2 \propto \mu^3$, respectively.
3. Initially we will ignore credibility considerations, aside from reviewing p-values, and later we will use Gibbs Sampling to incorporate credibility.

7.2 Comparison of GLM and Minimum Bias Model Results

Figures 2 and 3 and Table 1 show the bootstrap quantile testing results of fitting and performance testing models. Optimal noise-to-signal estimates along the lines described in [6] suggested using 20 quantiles. Also, see [6] for details on the definitions of the test statistics. The "old statistic" test measure is the ratio of the variance of the relative average payments after rating factors are applied to the same variance before rating factors are applied, lower being better. The "new statistic" test measure is essentially the square root of the difference between these two variances, higher being better.

Although Figures 2 and 3 only correspond to the Minimum Bias fits, Table 1 demonstrates that the Log-Poisson GLM was identical to the Minimum Bias approach, and the best fitting model. In fact, we checked the individual predicted values and verified that they were numerically identical. Log-Gaussian and Log-Gamma were almost as good. The MLE for our run of Log-Inverse Gaussian failed to converge, almost certainly driven by its unrealistic variance assumption.

Figures 4 and 5 correspond to "Traditional" univariate rate relativities for the three explanatory variables. Rating factors are calculated separately and independently in each classification dimension. The Traditional method clearly performs much worse than Minimum Bias and the convergent GLMs, but is still a great improvement over no adjustment.

Figure 2. Bootstrap 20 Quantiles Test Validation of Minimum Bias Rating Factors

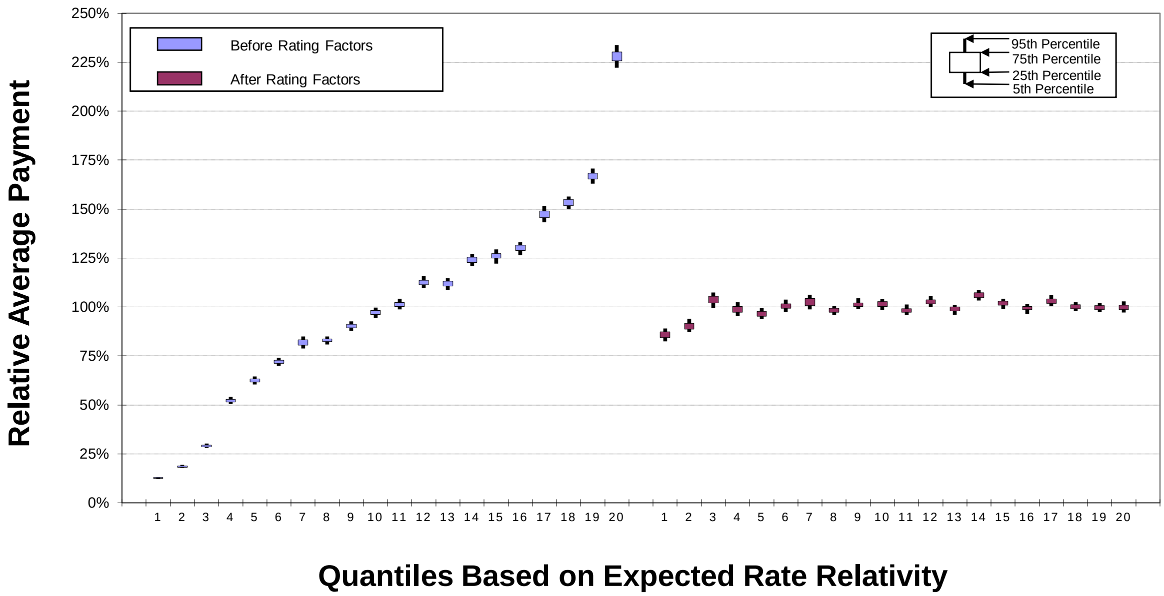


Figure 3. Allegation Nature - Bootstrap Test Validation of Minimum Bias Rating Factors

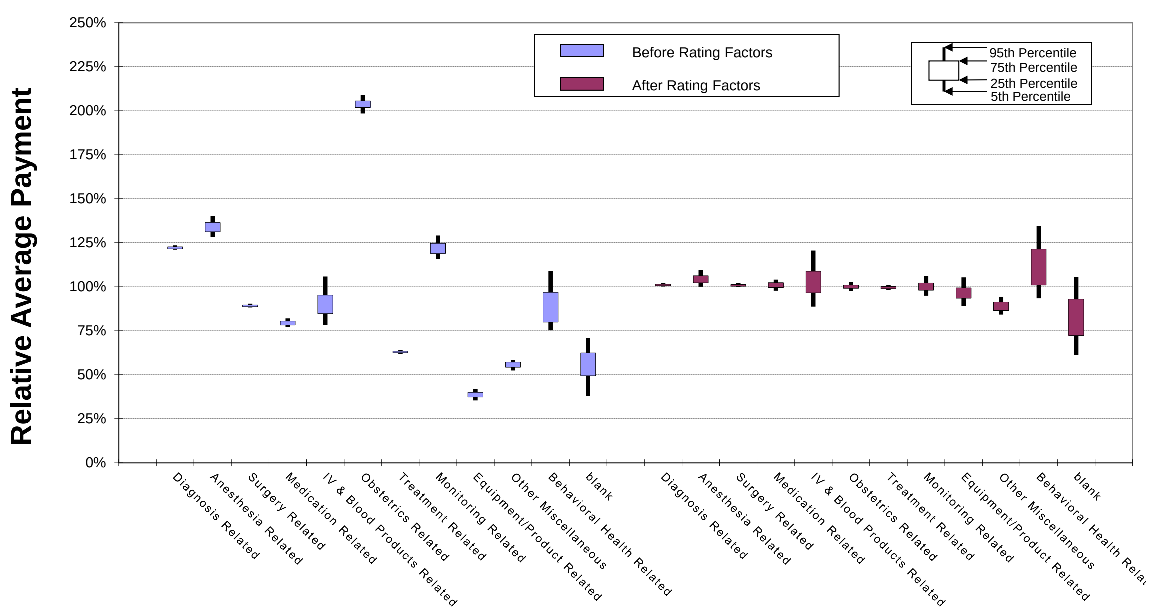


Figure 4. Bootstrap 20 Quantiles Test Validation of Traditional Rating Factors

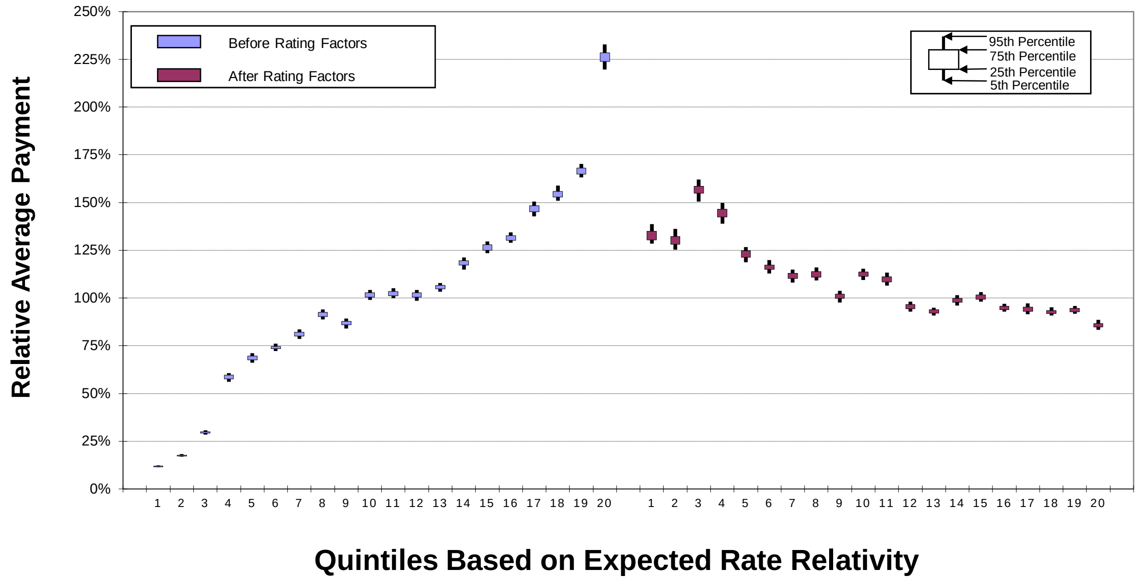


Figure 5. Allegation Nature - Bootstrap Test Validation of Traditional Rating Factors

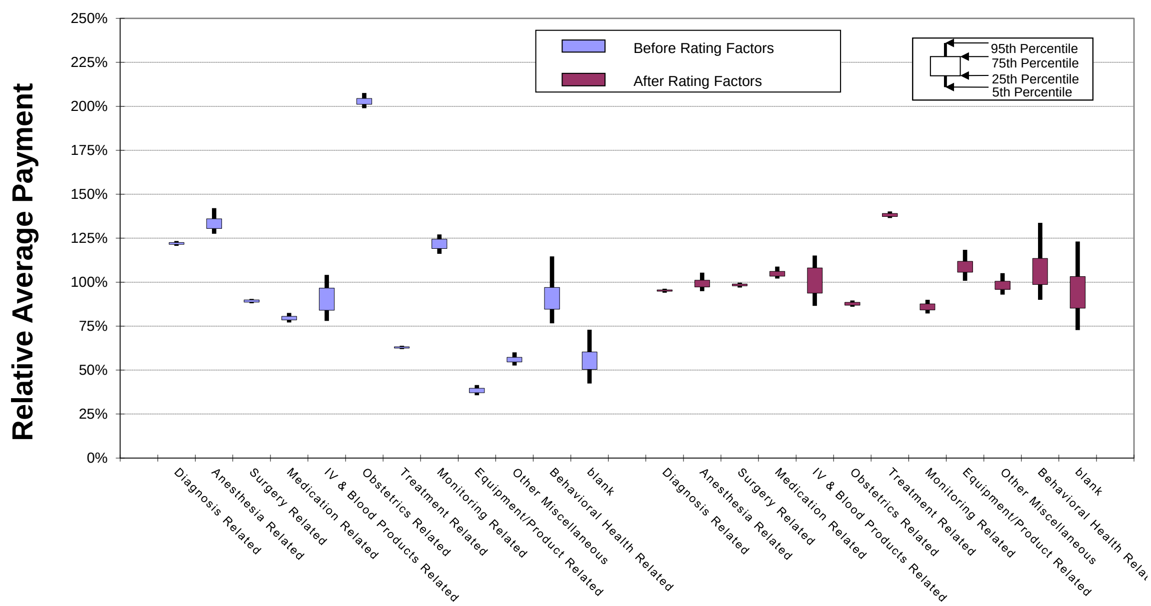


Table 1. Predictive Performance Statistics for Various Models

	20 Quantiles		Allegation Nature	
	Old Statistic	New Statistic	Old Statistic	New Statistic
Mult. Minimum Bias	0.007	0.512	0.023	0.425
GLMs				
Log-Gaussian	0.010	0.511	0.041	0.422
Log-Poisson	0.007	0.512	0.023	0.425
Log-Gamma	0.009	0.511	0.033	0.422
Log-InverseGaussian	Failed to Converge		Failed to Converge	
Traditional	0.135	0.470	0.089	0.408

At this point we have a clear picture of the relative predictive performance of the different models. However, we have not specifically tested the validity of any of the model assumptions, such as likelihoods, independence assumptions, etc. The optimal performance of Minimum Bias/Log-Poisson is likely due to the general validity of its implicit connection to the Central Limit Theorem as discussed earlier.

The GLM assumption that all risks are identically distributed is potentially problematic taken together with the log-link function.

Figures 6 through 8 illustrate the lack of distributional consistency for this dataset. We have broken the observations in the training data into 20 quantiles weighted by modeled values, sorted by actual/modeled result. Using the same breakpoints, determined from the entire training dataset, we then calculated the amount of summed modeled values for each allegation group. If the errors were identically distributed for each allegation group there should be only a random fluctuation around the 5% of total expected for each bin.

Figure 6 shows all allegation natures and naturally each bin demonstrates no differences in the weighted proportion. Figure 7 shows that the anesthesia related allegation group has a much higher percentage in the lowest bin than what would have been expected from the overall population,

Figure 6. All Allegation Nature 20 Value Weighted Quantile Bins

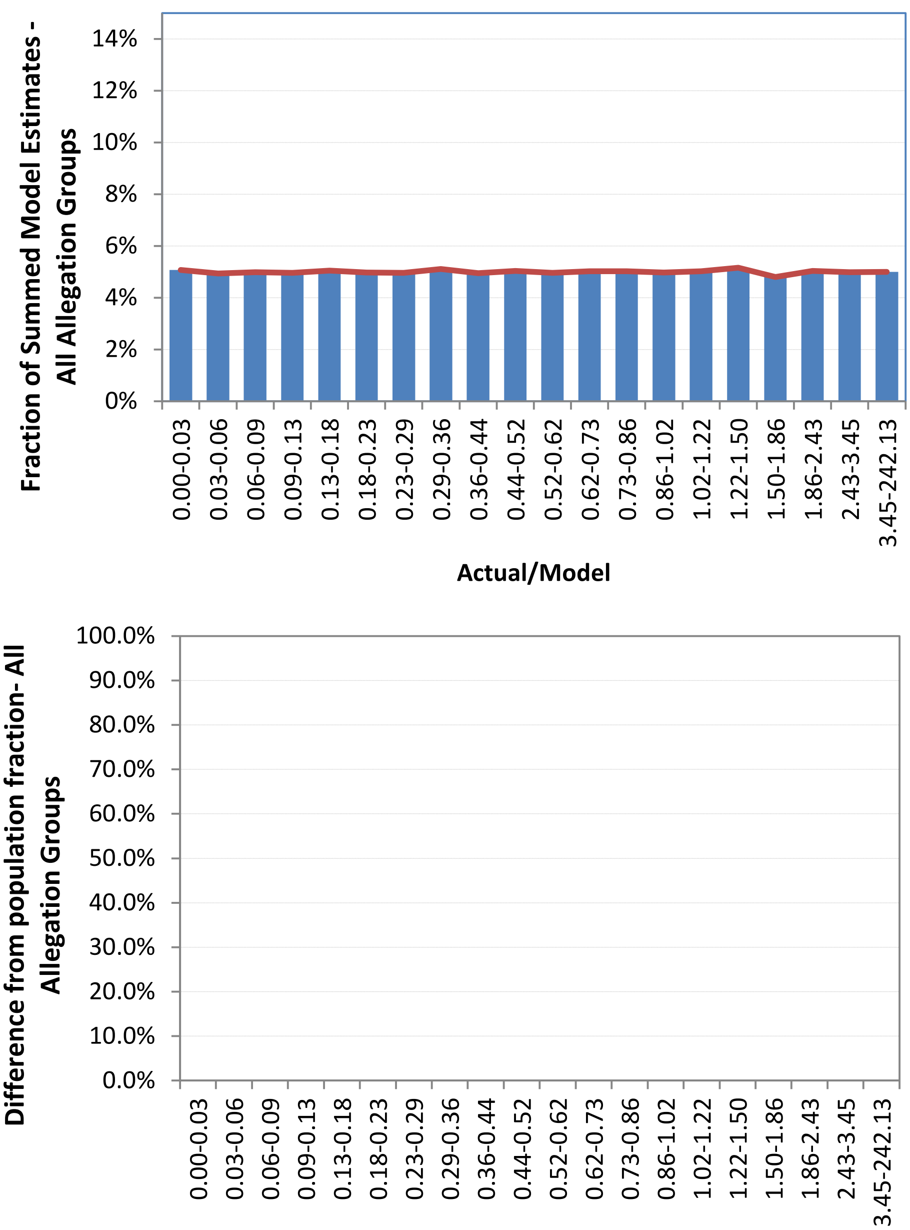


Figure 7. Anesthesia Allegation 20 Value Weighted Quantile Bins

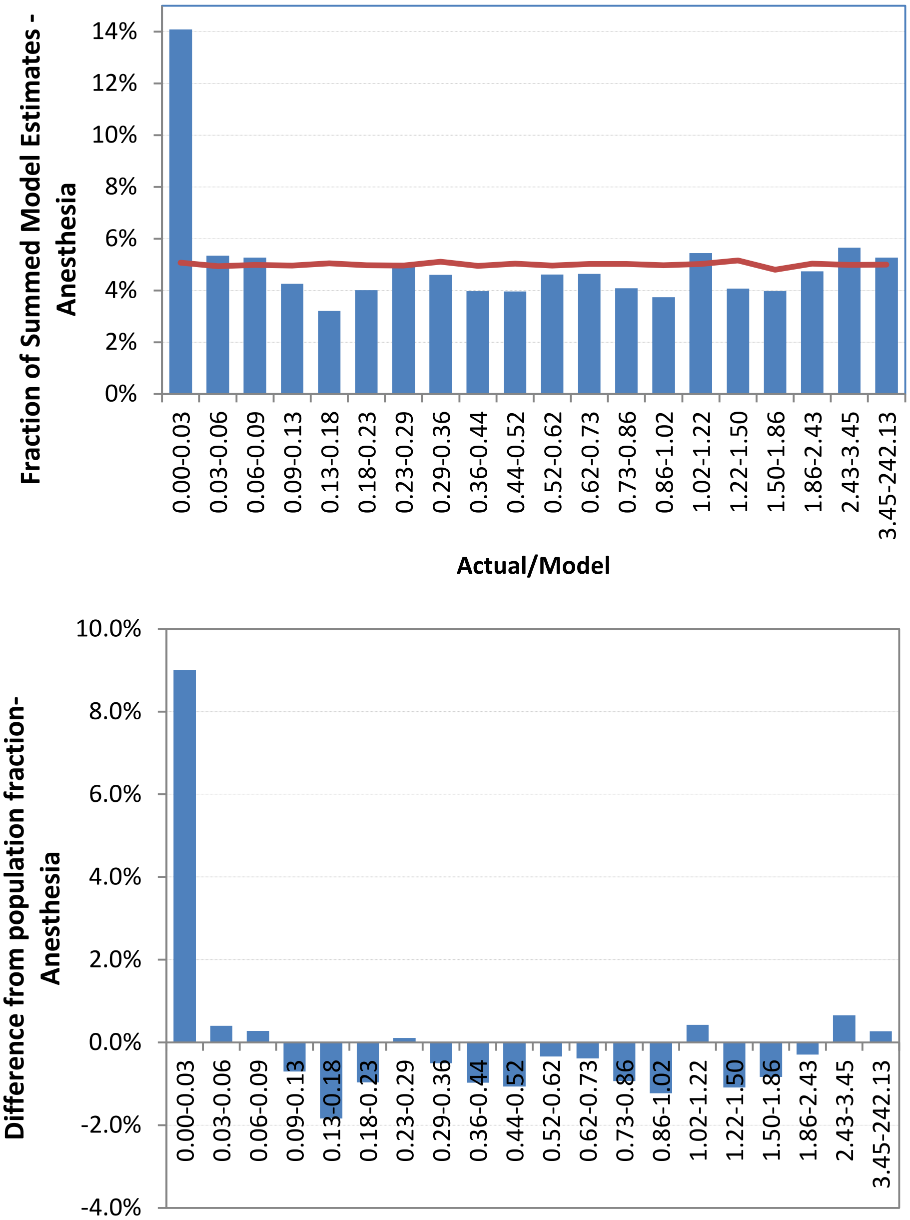
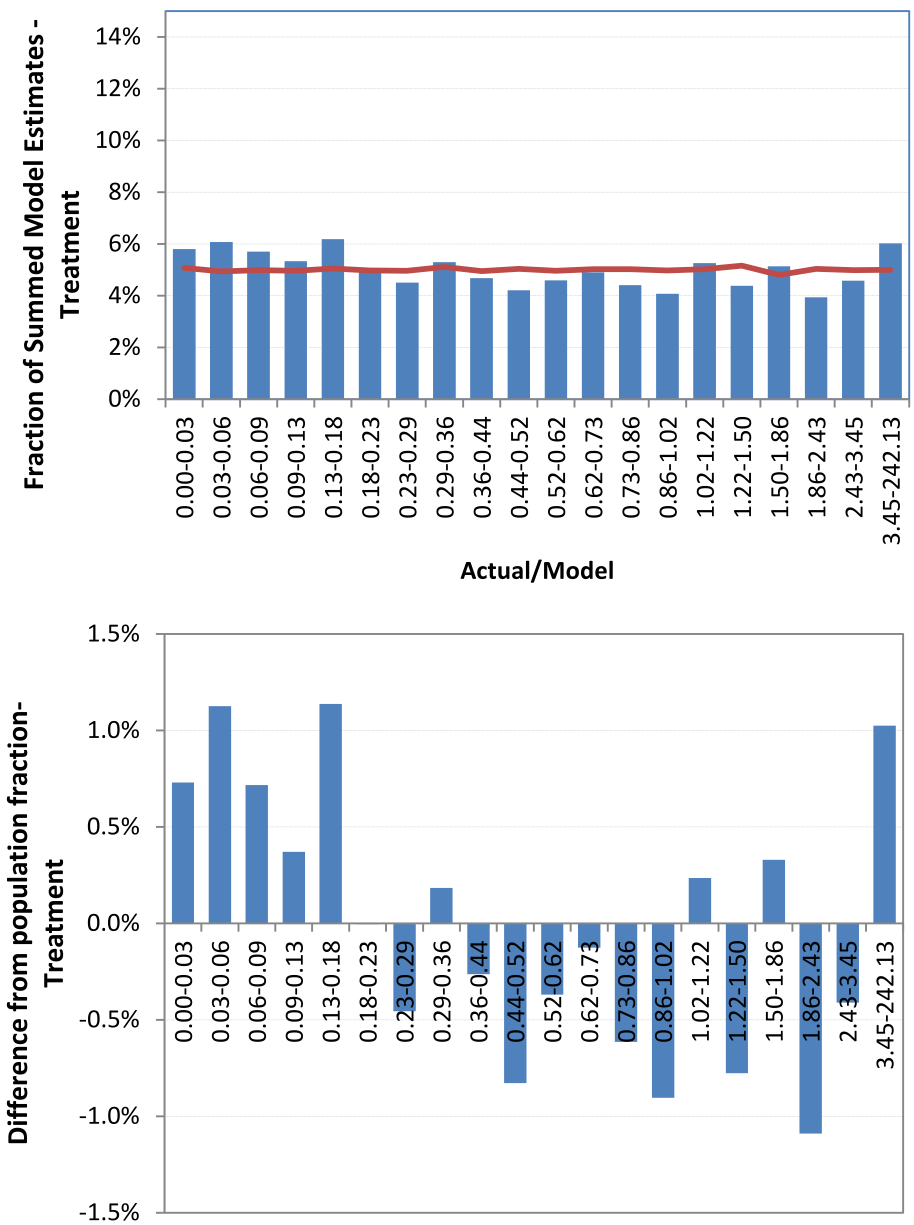


Figure 8. Treatment Allegation 20 Value Weighted Quantile Bins



of the error distribution. Figure 8 shows that, while not as dramatic, the treatment related allegation group shows greater variation than the overall error distribution, with more of the highest and lowest values.

This is far from uncommon with highly-skewed insurance data. The problem is compounded by the multiple dimensions of data. Error distributions could be, and likely are, differently distributed across many of the dimensions, if not every dimension being analyzed. Without adjustment, the basic assumption in a GLM is that the errors are identically distributed. The use of the log-link function, in conjunction with maximum likelihood estimation, puts a great deal of faith in the distributional assumption, inferring conclusions about results in the tail, based on the more voluminous observations at the lower parts of the distribution. But it is the tail itself that is of primary interest in most insurance questions, with the majority of the aggregate losses being caused by the minority of claims. Despite the unreasonable implied assumption of a log-Poisson GLM, because it happens to have effectively the same parameter estimation formulas as the multiplicative minimum bias approach, which has the associated Central Limit Theorem advantages previously described, it is less vulnerable to these distributional differences.

Table 2 shows a comparison of the model biases by allegation group on the validation data using multiplicative minimum bias with full credibility vs. GLM with a log-Gaussian assumption, by comparing actual aggregated results by allegation group to aggregated modeled results over a number of bootstrapped test sets. Despite the log-Gaussian assumption better characterizing the distribution of the data than does the log-Poisson assumption, it ultimately produces estimates that are more vulnerable to distributional differences. The only allegation group with a worse log-Gaussian mean bias was that for Equipment/Product related payments, and in that group, both sets of bootstrapped ranges contained zero, suggesting that the bias measure was inconclusive.

Table 2. Bootstrapped (Actual – Modeled)/Modeled By Allegation Nature

	Multiplicative Minimum Bias			Log-Gaussian		
	Mean	5th %	95th %	Mean	5th %	95th %
Diagnosis	1.0%	0.1%	2.0%	1.3%	0.4%	2.3%
Anesthesia	4.3%	0.0%	9.5%	7.1%	2.5%	11.9%
Surgery	0.8%	-0.3%	2.1%	1.1%	-0.2%	2.5%
Medication	0.9%	-2.2%	4.0%	2.2%	-0.6%	5.4%
IV & Blood Products	3.0%	-11.3%	20.5%	3.6%	-6.8%	15.9%
Obstetrics	0.1%	-2.4%	2.8%	-0.4%	-2.3%	1.8%
Treatment	-0.5%	-2.0%	1.1%	-2.5%	-4.0%	-1.0%
Monitoriing	0.2%	-5.1%	6.2%	0.9%	-4.3%	5.7%
Equipment/Product	-3.4%	-11.0%	5.4%	0.0%	-9.3%	8.7%
Other	-11.1%	-15.8%	-5.7%	-14.3%	-19.6%	-8.9%
Behavioral Health	11.9%	-6.5%	34.4%	13.2%	-10.4%	40.9%
Blank	-17.0%	-38.8%	5.5%	-20.7%	-37.7%	-0.6%

7.3 Incorporating Credibility into Minimum Bias

Although the overall predictive performance without any credibility adjustments was very good, there are reasons to explore credibility. In some sparsely populated classes for License Field, rating variables might be so unreliable as to lead to adverse selection problems in real world applications.

In the previous example, the p-values for the rating factors in the Log-Poisson were all infinitesimally low (the largest p-value ~ 10^{-204}). This is likely due to the problematic general phenomenon that p-values tend to always implode with very large volumes of data, such as the volume in the example. In stark contrast, most of the p-values for the Log-Gaussian and Log-Gamma models were high, from 1% to approaching 100%. Whether these p-value results indicate any of the likelihood selections are valid, or not, they demonstrate the generally awkward nature of trying to use p-values and class consolidation to handle the lack of credibility in sparsely populated classes.

Rather than attempt a p-value based class consolidation, we will explore the impact of a very simple credibility adjustment for Minimum Bias. We select the very simple form $Z_{j,i_j} = \frac{P_{j,i_j}}{P_{j,i_j} + K}$ where P_{j,i_j} is the number of records where the i_j class for classification dimension j and $K \geq 0$ is a judgmental selection. Table 3 shows that this simple credibility adjustment only tends to erode overall predictive value for this large dataset with only truly predictive variables included.

Table 3. Predictive Performance Statistics for Credibility Adjusted Multiplicative Minimum Bias

	20 Quantiles		Allegation Nature	
	Old Statistic	New Statistic	Old Statistic	New Statistic
Mult. Minimum Bias				
K = 0	0.007	0.512	0.023	0.425
K = 1	0.009	0.511	0.032	0.425
K = 10	0.010	0.511	0.030	0.423
K = 25	0.009	0.510	0.029	0.425
K = 50	0.010	0.511	0.022	0.424
K = 100	0.011	0.511	0.028	0.425
K = 200	0.013	0.509	0.031	0.423
K = 700	0.023	0.505	0.082	0.414

To construct a smaller example where credibility is more relevant, we will use a random set of only 5,000 records for fitting and another random set of 5,000 records for testing, shown in Tables 4 and 5 and Figures 9 through 12. We will also do a full test using all the remaining 366,123 records not used for fitting, shown in Tables 6 and 7 and Figures 13 and 14.

Table 4. Smaller Sample Predictive Performance Statistics for Various Models

	6 Quantiles		Allegation Nature	
	Old Statistic	New Statistic	Old Statistic	New Statistic
Mult. Minimum Bias	0.021	0.463	2.216	-0.683
GLMs				
Log-Gaussian	0.041	0.448	3.252	-0.785
Log-Poisson	0.021	0.463	2.216	-0.683
Log-Gamma	0.052	0.445	2.245	-0.704
Log-InverseGaussian	Failed to Converge		Failed to Converge	
Traditional	0.524	0.302	2.419	-0.751

Table 5. Smaller Sample Predictive Performance Statistics for Credibility Adjusted Multiplicative Minimum Bias

	6 Quantiles		Allegation Nature	
	Old Statistic	New Statistic	Old Statistic	New Statistic
Mult. Minimum Bias				
K = 0	0.021	0.463	2.216	-0.683
K = 1	0.016	0.457	1.138	-0.419
K = 10	0.012	0.461	0.454	0.246
K = 25	0.022	0.458	0.394	0.316
K = 50	0.043	0.450	0.376	0.338
K = 100	0.068	0.449	0.373	0.345
K = 200	0.093	0.432	0.384	0.345
K = 700	0.255	0.387	0.479	0.319

As Tables 4 through 7 and Figures 9 through 14 show, the incorporation of credibility was particularly important when distinguishing differences between the allegation groups. Actuaries are regularly asked to provide estimates of the impact of rating variables despite having less than fully

Figure 9. Smaller Sample Bootstrap 6 Quantiles Test Validation of Minimum Bias Rating Factors

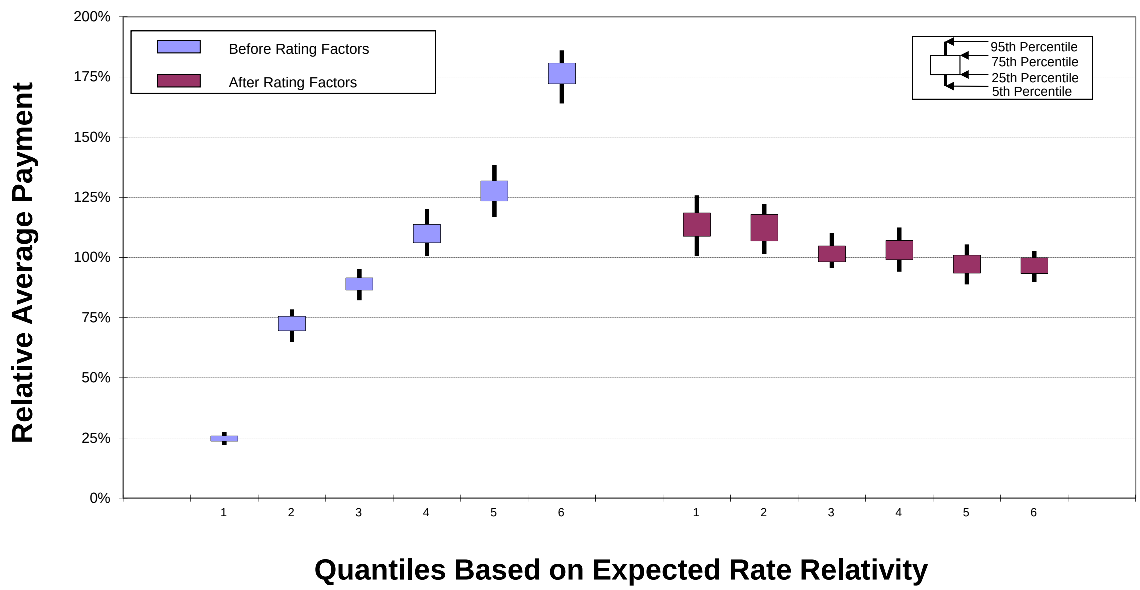


Figure 10. Smaller Sample Allegation Nature - Bootstrap Test Validation of Minimum Bias Rating Factors

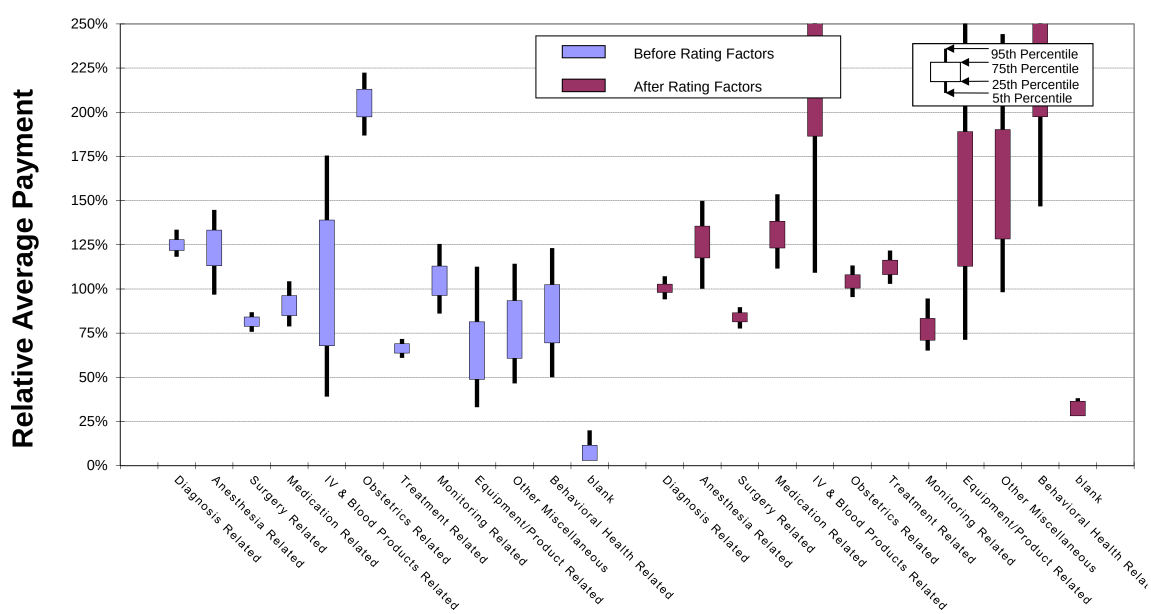


Figure 11. Smaller Sample Bootstrap 6 Quantiles Test Validation of Minimum Bias (Credibility K = 10) Rating Factors

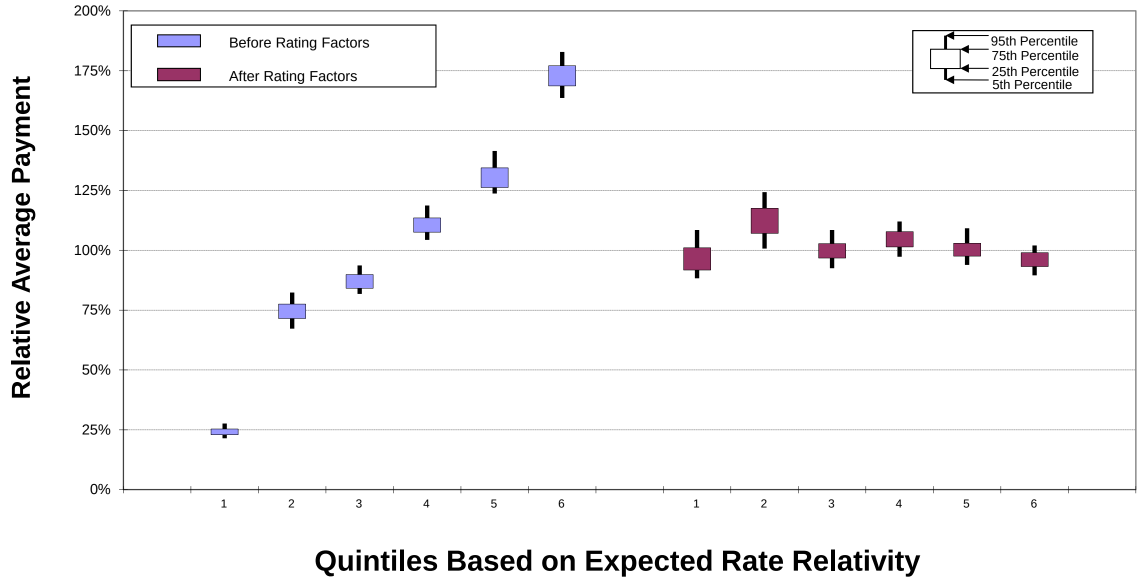


Figure 12. Smaller Sample Allegation Nature - Bootstrap Test Validation of Minimum Bias (Credibility K = 10) Rating Factors

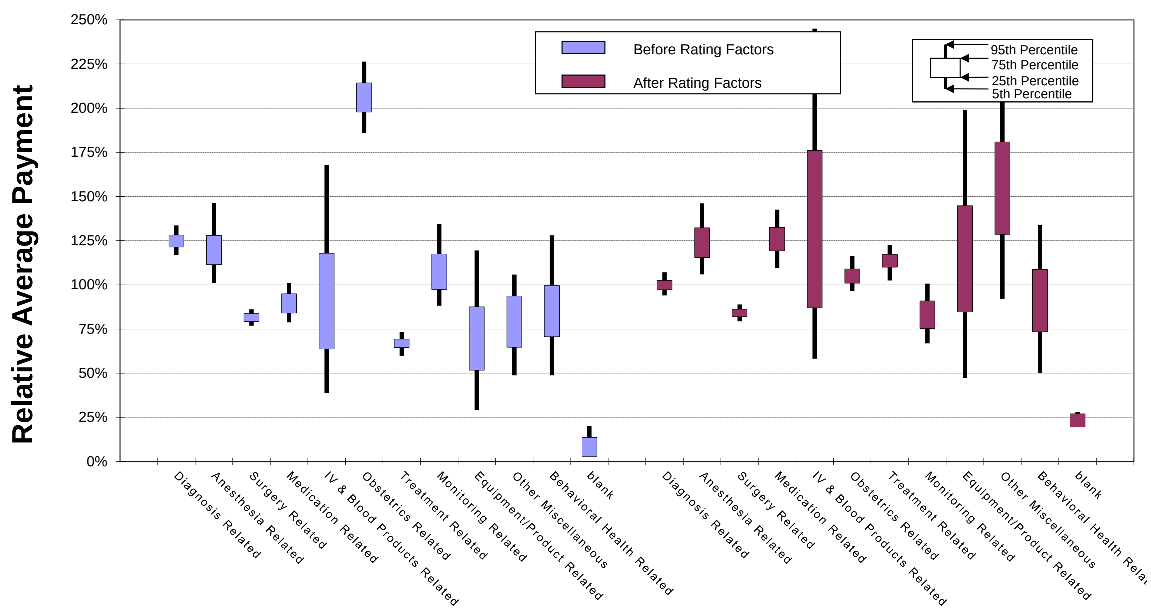


Table 6. Full Test of Smaller Sample Predictive Performance Statistics for Various Models

	20 Quantiles		Allegation Nature	
	Old Statistic	New Statistic	Old Statistic	New Statistic
Mult. Minimum Bias	0.031	0.488	1.906	-0.403
GLMs				
Log-Gaussian	0.038	0.482	2.673	-0.556
Log-Poisson	0.031	0.488	1.906	-0.403
Log-Gamma	0.072	0.474	3.256	-0.653
Log-InverseGaussian	Failed to Converge		Failed to Converge	
Traditional	0.489	0.350	2.158	-0.471

Table 7. Full Test of Smaller Sample Predictive Performance Statistics for Credibility Adjusted Multiplicative Minimum Bias

	20 Quantiles		Allegation Nature	
	Old Statistic	New Statistic	Old Statistic	New Statistic
Mult. Minimum Bias				
K = 0	0.031	0.488	1.906	-0.403
K = 1	0.020	0.492	0.835	0.139
K = 10	0.012	0.494	0.169	0.380
K = 25	0.013	0.493	0.187	0.379
K = 50	0.026	0.489	0.215	0.372
K = 100	0.063	0.479	0.246	0.364
K = 200	0.117	0.460	0.289	0.355
K = 700	0.300	0.399	0.427	0.317

Figure 13. Full Test of Smaller Sample Bootstrap 6 Quantiles Test Validation of Minimum Bias (Credibility K = 10) Rating Factors

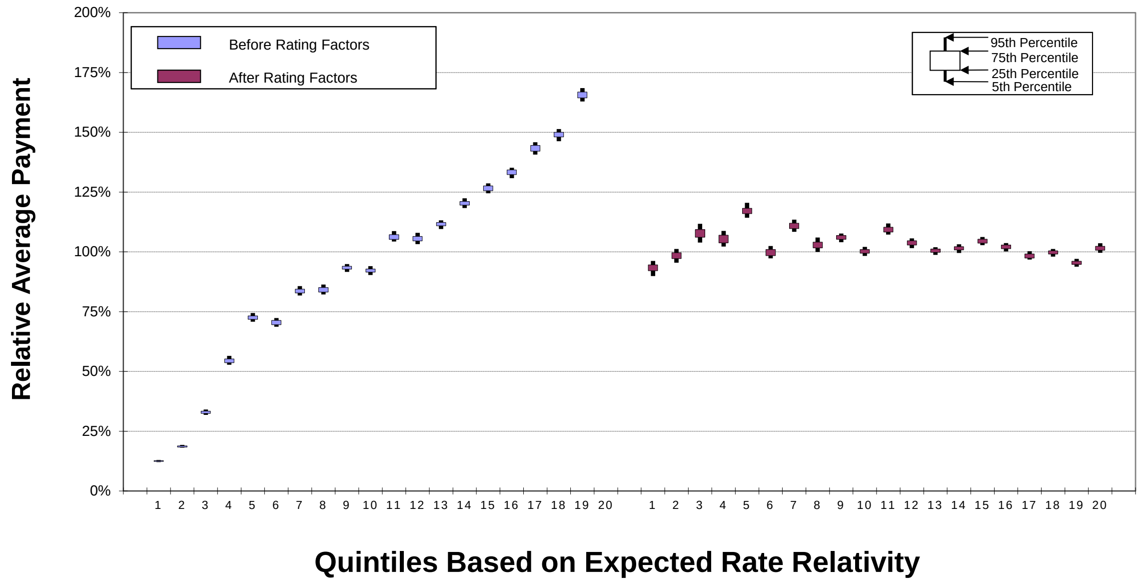
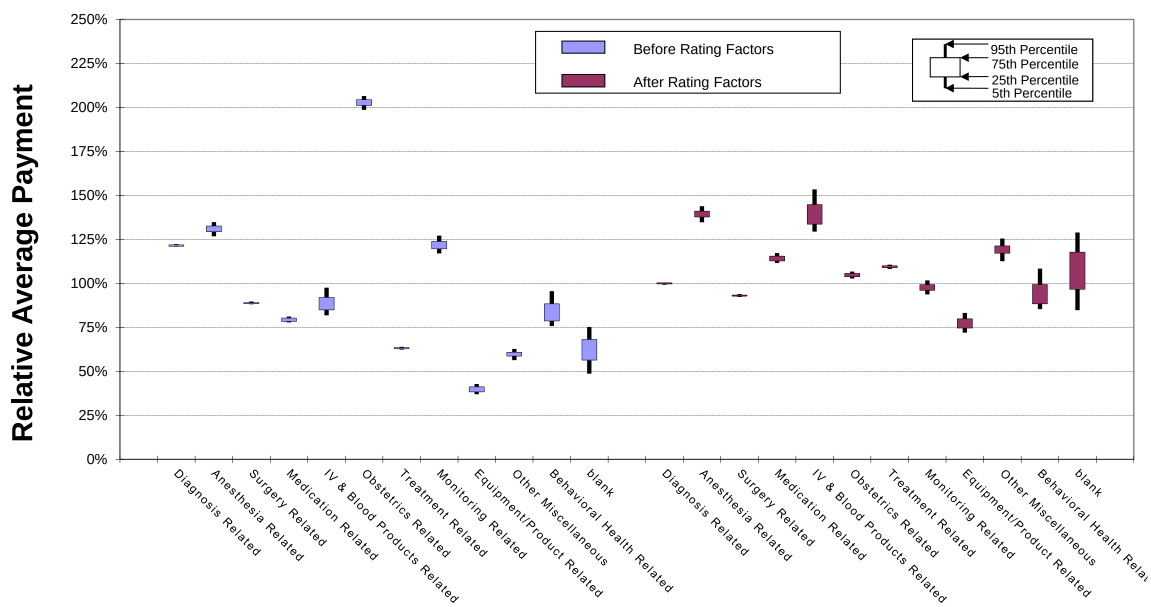


Figure 14. Full Test of Smaller Sample Allegation Nature - Bootstrap Test Validation of Minimum Bias (Credibility K = 10) Rating Factors



credible data. While the overall result may appear to be relatively unaffected by increasing the credibility standard, the ability to more robustly differentiate between them is illustrated.

8.4 Incorporating Credibility Into GLM

We can incorporate credibility, or "shrinkage" of parameter estimates, into a GLM model by defining a hierarchical Bayesian Network of random variables:

$$U_{1,j} = 0 \quad j = 1,2,3$$

$$U_{1,4} = \text{Uniform}(0, 20)$$

$$U_{i,1} \sim \text{Normal}(-\sigma_1{}^2 / 2,\ \sigma_1{}^2) \quad i = 2,\ldots,83$$

$$U_{i,2} \sim \text{Normal}(-\sigma_1{}^2 / 2,\ \sigma_1{}^2) \quad i = 2,\ldots,12$$

$$U_{i,3} \sim \text{Normal}(-\sigma_1{}^2 / 2,\ \sigma_1{}^2) \quad i = 2,\ldots,9$$

$$\sigma_1{}^2 \sim \text{Lognormal}(0,10)$$

$$\sigma_2{}^2 \sim \text{Lognormal}(0,10)$$

$$\delta_k \sim \text{Normal}(-\sigma_2{}^2 / 2,\ \sigma_2{}^2) \quad k = 1,\ldots,n$$

$$Y_k \sim \text{Poisson}(\text{Exp}(\delta_k + U_{1,4} + U_{i_{1,k},1} + U_{i_{2,k},2} + U_{i_{3,k},1})) \quad k = 1,\ldots,n$$

Y_k are the individual actual claim amounts to be fit. $U_{i,j}$ are parameters in log space, with $U_{1,4}$ being a constant and the other j= 1,2, or 3 corresponding to *License Field, Allegation Group,* and *Original Year*, respectively. $i_{j,k}$ is an index of which class the Y_k observation falls into in each classification dimension. δ_k is a random over-dispersion for each observation which itself has variance $\sigma_2{}^2$. $\sigma_1{}^2$ is the parameter variance for each class parameter. Since $U_{1,4}$, $\sigma_1{}^2$, and $\sigma_2{}^2$ follow highly diffuse distributions they will effectively be "fitted" parameters when Gibbs Sampling is performed. $\sigma_1^{\ 2}$, and $\sigma_2^{\ 2}$ conceptually correspond to parameter and process variances in credibility.

We will also defined a simpler form of this model eliminating the over-dispersion arising from $\sigma_1^{\ 2}$ and $\sigma_2^{\ 2}$. Running this simpler model numerically produced the same parameters as the MLE Log-Poisson/Minimum Bias with no credibility adjustment, confirming that our Gibbs Sampling model is constructed and coded on the right track up to to the point of adding credibility adjustments.

When the model including the δ_k and $\sigma_2^{\ 2}$ was run numerically there was a shrinkage effect observed in the set of parameters. Table 8 shows that the range of the $U_{i,1}$ contracted significantly with over-dispersion. There was a slight broadening of the ranges for $U_{i,2}$ and $U_{i,3}$, which is not unreasonable as none of the corresponding classes in these dimensions are sparsely populated.

Table 8. Shrinkage Effect in Range of Gibbs Sampled Parameter Fits

	$U_{i,1}$		$U_{i,2}$		$U_{i,3}$	
	Min	Max	Min	Max	Min	Max
Large Split						
w/o overdispersion	-4.103	0.775	-0.920	0.473	0.000	0.691
w overdispersion	-2.173	0.550	-0.975	0.742	-0.040	0.494
Smaller Sample						
w/o overdispersion	-6.570	2.234	-1.405	0.432	0.000	0.742
w overdispersion	-2.033	0.963	-1.992	0.318	-0.069	0.691

Unfortunately, although there was a credibility-like shrinkage affect, the predictive performance actually deteriorated. Figures 15 and 16 show the deteriorating situation when the Gibbs Sampling with over-dispersion is included in the large split of the data. Table 9 shows the deterioration in test statistics for both the large split and smaller sample.

There are potential criticisms of the Bayesian network model as we have defined it. For example, the anchoring of the parameters for the first classes $U_{1,j} = 0 \quad j = 1,2,3$; offsetting the prior distributions on parameters so as to have mean 1 after exponentiation $U_{i,1} \sim \text{Normal}(-\sigma_1^{\ 2}/2,\ \sigma_1^{\ 2}) \quad i = 2,\ldots,83$; the same parameter variance $\sigma_1^{\ 2}$ was used for all three classification dimensions; etc. However, the authors experimented with a myriad of alterations to the model definition, even going so far as to convert the likelihood function into a Negative Binomial distribution to capture the impact of over-dispersion of the Poisson more directly. In all cases tried performance deteriorated further or did not improve. The earlier presented multiplicative minimum bias model with incorporated credibility would be vulnerable to similar or more extensive potential criticisms. Yet implementing it went quickly and easily produced desirable results.

This failed modeling experience in no way proves that a well performing Gibbs Sampled Bayesian model cannot be defined in this context. Obviously, well performing examples for much simpler situations, such as one classification dimension and an identity link function, are well known and easy to construct. Nor is the point that the theory behind these models does not provide deep insights into understanding modeling and statistical estimation. However, in this case, orders of magnitude more input of resources both in time and sophistication in effort than was used for minimum bias produced inferior predictive performance. Though neither author of this paper is a specialist in Gibbs Sampling methods, one author (Evans) has used them occasionally for over 10 years and informally consulted several more experienced specialists (in Acknowledgements). As of this writing, we have not been able to diagnose why the model as defined performs so much more poorly than a regular MLE GLM with no shrinkage effect. Whether the model is in some way poorly designed or, much less likely, one of the many technical choices made in running the Gibbs Sampling software should be tuned differently, does not alter the key conclusion. Namely, that the tremendous additional resource and intellectual burdens of such detailed and sophisticated models may offer no advantage, or may even be disadvantageous, in many practical situations of predictive modeling.

Figure 15. Full Test of Smaller Sample Bootstrap 20 Quantiles Test Validation of Gibbs Sampled Rating Factors with Shrinkage

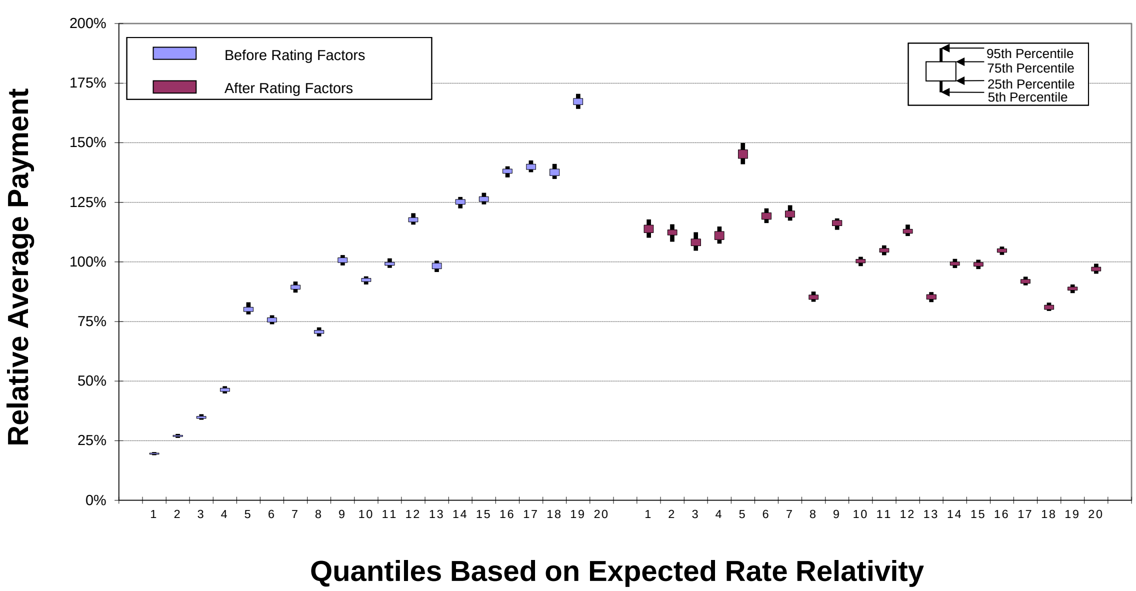


Figure 16. Full Test of Smaller Sample - Allegation Nature - Bootstrap Test Validation of Gibbs Sampled Rating Factors with Shrinkage

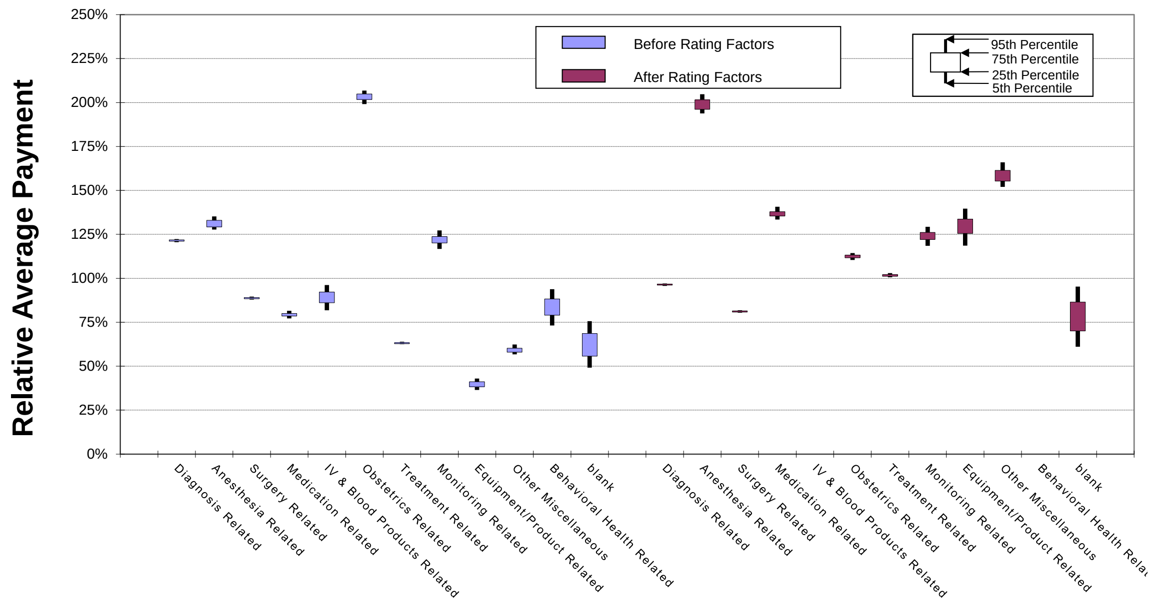


Table 9. Test Statistics for Gibbs Sampled Rating Factors

	Quantiles		Allegation Nature	
	Old Statistic	New Statistic	Old Statistic	New Statistic
Large Split (20 Quantiles)				
w/o overdispersion	0.007	0.512	0.023	0.425
w overdispersion	0.102	0.463	0.219	0.376
Smaller Sample (6 Quantiles)				
w/o overdispersion	0.021	0.463	2.216	-0.683
w overdispersion	0.101	0.403	3.616	-0.943
Full Test Smaller Sample (20 Quantiles)				
w/o overdispersion	0.031	0.488	1.906	-0.403
w overdispersion	0.098	0.448	4.723	-0.818

8. SUMMARY DISCUSSION

The predictive modeling framework greatly reduces the burdens of model specification, because models are validated based on their predictive performance rather than hypothesis testing of model assumptions. Minimum bias models transform basic data in such a way as to partially sacrifice sample independence in exchange for much tamer distributions of individual data points that are much less needy of detailed distributional specification. The combination of multiplicative minimum bias iteration with a generic incorporation of credibility as presented in this paper demonstrates that a very simple model, without complete distributional specification, in practice can provide comparable or better predictive value than a far more complex model, such as a typical GLM.

GLM models are fit to individual data points and require specification of the distributions underlying each data point. Consequently, GLM models can be significantly vulnerable to inaccurate specifications and their fundamental complexity makes the practical incorporation of credibility adjustments, such as including random effects or fitting parameters through Gibbs sampling, very complex.

Philosophically, simpler modeling is desirable. In practice, simpler models are beneficial in many ways, such as lower skill requirements for operational personnel and greater lucidity to a much wider audience. Some previous papers, such as Brown in [5] and Mildenhall in [10], have highlighted the sense in which minimum bias iteration is a special case of GLM and encouraged – at least implicitly – minimum bias practitioners to switch to GLM as a richer framework. There is some irony that with the advent of the predictive framework minimum bias may often be somewhat more advantageous, in principle and practice.

While GLM models are powerful and belong in the set of tools applied by actuaries, consideration should also be given to multiplicative minimum bias models and the traditional actuarial concept of partial credibility. Ultimately the test of any predictive model should be how it performs on out-of-sample data.

Acknowledgment

The authors acknowledge are thankful to Jose Couret, Louise Francis, Chris Laws and Frank Schmid for answering some questions that arose in the course of writing this paper.

Appendix A Details of Empirical Case Study

The empirical data used in this case study consists of 371,123 records of medical malpractice payments obtained from the National Practitioner Data Bank. Three explanatory variables were used for modeling payment amounts: *Original Year*, *Allegation Group* and *License Field*. The following Tables A.1 through A.3 display record counts by each of the explanatory variables overall and for the individual predictive modeling splits.

Table A.1 Counts of Records by License Field

		Large Split		Smaller Sample		
License Field	Total	Fit	Test	5,000 Fit	5,000 Test	Full Test
Allopathic Physician (MD)	271,443	135,514	135,929	3644	3661	267,799
Phys. Intern/Resident (MD)	2,113	1,063	1,050	34	28	2,079
Osteopathic Physician (DO)	17,612	8,829	8,783	237	244	17,375
Osteo. Phys. Intern/Resident (DO)	324	161	163	8	6	316
Dentist	46,516	23,425	23,091	623	596	45,893
Dental Resident	145	64	81	4	3	141
Pharmacist	1,890	952	938	24	20	1,866
Pharmacy Intern [available 9/9/2002]	2	1	1	0	0	2
Pharmacist, Nuclear	6	4	2	0	0	6
Pharmacy Assistant	19	12	7	0	0	19
Pharmacy Technician [available 9/9/2002]	12	7	5	0	1	12
Registered (RN) Nurse	5,715	2,885	2,830	91	80	5,624
Nurse Anesthetist	1,568	777	791	19	19	1,549
Nurse Midwife	873	431	442	18	8	855
Nurse Practitioner	1,288	598	690	19	24	1,269
Doctor of Nursing Practice [available 11/8/2010]	1	-	1	0	0	1
Advanced Nurse Practitioner [3/5/02 - 9/9/02]	4	3	1	0	0	4
LPN or Vocational Nurse	692	345	347	9	9	683
Clinical Nurse Specialist [available 9/9/02]	18	12	6	1	0	17
Certified Nurse Aide/Nursing Assistant [available 10/17/05]	36	18	18	0	1	36
Nurses Aide	78	39	39	2	2	76
Home Health Aide (Homemaker)	22	10	12	0	0	22
Health Care Aide/Direct Care Worker [available 10/17/05]	3	1	2	0	0	3
Psychiatric Technician	15	10	5	0	0	15
Dietician	22	11	11	0	1	22
Nutritionist	1	1	-	0	0	1
EMT, Basic	200	106	94	3	2	197
EMT, Cardiac/Critical Care	28	17	11	0	0	28
EMT, Intermediate	26	13	13	1	2	25
EMT, Paramedic	59	32	27	0	1	59
Clinical Social Worker	206	107	99	2	0	204
Podiatrist	7,654	3,809	3,845	92	113	7,562
Clinical Psychologist [last use 9/9/02]	875	436	439	15	15	860
Psychologist [available 9/9/02]	352	174	178	2	5	350
School Psychologist [available 9/9/02]	1	-	1	0	0	1
Audiologist	39	23	16	2	1	37
Art/Recreation Therapist	2	1	1	0	0	2
Massage Therapist	82	54	28	3	1	79
Occupational Therapist	85	43	42	0	0	85
Occup. Therapy Assistant	11	7	4	0	0	11
Physical Therapist	1,094	545	549	14	14	1,080

Table A.1 Counts of Records by License Field (continued)

		Large Split		Smaller Sample		
License Field	Total	Fit	Test	5,000 Fit	5,000 Test	Full Test
Phys. Therapy Assistant	94	48	46	0	3	94
Rehabilitation Therapist	9	3	6	0	0	9
Speech/Language Pathologist	14	9	5	0	0	14
Hearing Aid/Instrument Specialist [available 10/17/05]	2	1	1	0	0	2
Medical Technologist [changed to 501(6/15/09)]	64	28	36	0	0	64
Medical/Clinical Lab Technologist [available 6/15/09]	1	1	-	0	0	1
Medical/Clinical Lab Technician [available 6/15/09]	2	-	2	0	0	2
Surgical Technologist [available 6/15/09]	7	4	3	0	0	7
Surgical Assistant [available 6/15/09]	1	-	1	0	0	1
Cytotechnologist [available 11/22/99]	11	7	4	0	0	11
Nuclear Med. Technologist	14	5	9	0	0	14
Rad. Therapy Technologist	12	5	7	0	0	12
Radiologic Technologist	169	89	80	1	0	168
X-Ray Technician or Operator [available 6/15/09]	5	2	3	0	0	5
Acupuncturist	58	22	36	0	0	58
Athletic Trainer [available 11/22/99]	6	3	3	1	0	5
Chiropractor	5,834	2,928	2,906	78	87	5,756
Dental Assistant	15	8	7	1	1	14
Dental Hygienist	41	22	19	1	2	40
Denturist	27	8	19	0	0	27
Homeopath	6	5	1	1	0	5
Medical Assistant	33	14	19	1	0	32
Counselor, Mental Health	167	84	83	1	2	166
Midwife, Lay (Non-Nurse)	22	14	8	0	0	22
Naturopath	17	9	8	0	0	17
Ocularist	25	12	13	0	1	25
Optician	17	10	7	0	0	17
Optometrist	715	367	348	6	11	709
Orthotics/Prosthetics Fitter	9	5	4	1	0	8
Phys. Asst., Allopathic	1,713	847	866	26	22	1,687
Phys. Asst., Osteopathic	137	71	66	3	3	134
Perfusionist [available 11/22/99]	8	2	6	1	0	7
Podiatric Assistant	14	9	5	0	0	14
Prof. Counselor	209	109	100	4	3	205
Prof. Cnslr., Alcohol	9	2	7	0	1	9
Prof. Cnslr., Family/Marriage	177	96	81	4	5	173
Prof. Cnslr, Substance Abuse	23	13	10	0	0	23
Marriage and Family Therapist [available 9/9/02]	27	15	12	1	0	26
Respiratory Therapist	48	24	24	1	0	47
Resp. Therapy Technician	14	4	10	0	0	14
Other Health Care Pract, Not Classified [available 11/22/99]	45	31	14	0	0	45
Unspecified or Unknown	170	86	84	1	2	169
Total	371,123	185,562	185,561	5,000	5,000	366,123

Table A.2 Counts of Records by Allegation Nature

		Large Split		Smaller Sample		
Allegation Nature	Total	Fit	Test	5,000 Fit	5,000 Test	Full Test
Diagnosis Related	105,674	52,516	53,158	1,409	1,388	104,265
Anesthesia Related	10,974	5,421	5,553	127	153	10,847
Surgery Related	88,763	44,538	44,225	1,176	1,211	87,587
Medication Related	20,197	10,047	10,150	259	268	19,938
IV & Blood Products Related	1,259	625	634	14	16	1,245
Obstetrics Related	25,988	13,081	12,907	384	345	25,604
Treatment Related	100,666	50,517	50,149	1,380	1,372	99,286
Monitoring Related	7,313	3,594	3,719	103	106	7,210
Equipment/Product Related	2,037	989	1,048	32	24	2,005
Other Miscellaneous	7,404	3,791	3,613	106	106	7,298
Behavioral Health Related	677	361	316	7	9	670
blank	171	82	89	3	2	168
Total	371,123	185,562	185,561	5,000	5,000	366,123

Table A.3 Counts of Records by Origination Year Group

		Large Split		Smaller Sample		
Origination Year	Total	Fit	Test	5,000 Fit	5,000 Test	Full Test
1990-1992	40,574	20,306	20,268	568	515	40,006
1993-1994	39,016	19,480	19,536	570	529	38,446
1995-1996	37,048	18,557	18,491	516	509	36,532
1997-1998	35,689	17,838	17,851	490	493	35,199
1999-2000	38,036	19,045	18,991	469	516	37,567
2001-2002	39,277	19,650	19,627	491	533	38,786
2003-2004	36,565	18,256	18,309	472	508	36,093
2005-2007	47,519	23,756	23,763	659	646	46,860
2008-2012	57,399	28,674	28,725	765	751	56,634
Total	371,123	185,562	185,561	5,000	5,000	366,123

Appendix B Gibbs Sampling Model Code

<u>With Poisson Over-dispersion</u>

```
model
{
        U[1,4]~dunif(0,20)
        U[1,1]<-0
        U[1,2]<-0
        U[1,3]<-0
        Tau[1]  ~ dlnorm(0,0.1)
        Mu<-  -pow(Tau[1],-1)/2
        Tau[2] ~ dlnorm(0,0.1)
        Mu2<-  -pow(Tau[2],-1)/2
        Tau[3]<-Tau[1]/Tau[2]
        for(i in 2:N1) { U[i,1]~dnorm(Mu,Tau[1]) }
        for(i in 2:N2) { U[i,2]~dnorm(Mu,Tau[1]) }
        for(i in 2:N3) { U[i,3]~dnorm(Mu,Tau[1]) }
        for(i in 1:N) {
                ProcError[i]~dnorm(Mu2,Tau[2])
                lambda1[i]<-exp(min(20,ProcError[i]+U[1,4]+U[X[i,1],1]+U[X[i,2],2]+U[X[i,3],3]))
                Y[i]~dpois(lambda1[i])
                }
        }
```

<u>Without Poisson Over-dispersion</u>

```
model
{
        U[1,4]~dunif(0,20)
        U[1,1]<-0
        U[1,2]<-0
        U[1,3]<-0
        Tau[1]  ~ dlnorm(0,0.1)
        Mu<-  -pow(Tau[1],-1)/2

        for(i in 2:N1) { U[i,1]~dnorm(Mu,Tau[1]) }
        for(i in 2:N2) { U[i,2]~dnorm(Mu,Tau[1]) }
        for(i in 2:N3) { U[i,3]~dnorm(Mu,Tau[1]) }
        for(i in 1:N) {
                lambda1[i]<-exp(min(20,U[1,4]+U[X[i,1],1]+U[X[i,2],2]+U[X[i,3],3]))
                Y[i]~dpois(lambda1[i])
                }
        }
```

REFERENCES

[1] Anderson, D., et al., "A Practitioner's Guide to Generalized Linear Models," *CAS Exam Study Note* **2007**, 3rd ed.

[2] Bailey, Robert A., "Insurance Rates with Minimum Bias", *PCAS* **1963**, Vol. L, 4–13.

[3] Bailey, Robert A. and LeRoy J. Simon, "Two Studies in Automobile Insurance Ratemaking", *PCAS* **1960**, Vol. XLVII, 1–19.

[4] Brosius, E. and Feldblum, S., "The Minimum Bias Procedure: A Practitioner's Guide," *PCAS* **2003**, Vol. XC, 196-273.

[5] Brown, Robert L., "Minimum Bias with Generalized Linear Models", PCAS **1988**, Vol. LXXV, 187–217.

[6] Evans, J. and Dean, C., "The Optimal Number of Quantiles for Predictive Performance Testing of the NCCI Experience Rating Plan," *Variance* **2014**, Vol. 8 No. 2, 89-104.

[7] Frees, E., et al., *Predictive Modeling Applications in Actuarial Science: Volume 1, Predictive Modeling Techniques*, Cambridge University Press, **2014**.

[8] Klinker, F., "Generalized Linear Mixed Models for Ratemaking: A Means of Introducing Credibility into a Generalized Linear Model Setting," *CAS Forum* **2001**, Winter Vol. 2, 1-25.

[9] Klugman, S., et al., *Loss Models: From Data to Decisions 4th Edition*, Wiley, **2012**.

[10] Mildenhall, S., "Minimum Bias and Generalized Linear Models," *PCAS* **1999**, Vol. LXXVI, 393-487.

[11] Nelder, J. and Verrall, R., "Credibility Theory and Generalized Linear Models," *ASTIN Bulletin* **1997**, Vol. 27:1, 71-82.

[12] Scollnik, D., "An Introduction to Markov Chain Monte Carlo Methods and their Actuarial Applications," *PCAS* **1996**, Vol. LXXXIII, 114–165.

[13] Venter, Gary G., "Discussion of Minimum Bias with Generalized Linear Models," *PCAS* **1990**, Vol. LXXVII, 337–349.

Biographies of the Authors

Christopher Gross, ACAS, MAAA is a property and casualty consulting actuary with 20 years of experience. He is the founder and president of Gross Consulting, which provides actuarial services and the Cognalysis™ line of actuarial software: Claim Life Cycle Model (CLCM™), MultiRate™, and Cognalysis Reserving System™ .

Jonathan Evans, FCAS, FSA, FCA, CERA, MAAA, WCP is a property and casualty consulting actuary with 20 years of experience. He is currently President of Convergent Actuarial Services, Inc.

A Cost of Capital Risk Margin Formula For Non-Life Insurance Liabilities

Glenn Meyers, FCAS, MAAA, CERA, Ph.D.

January 5, 2017

Abstract

A Bayesian MCMC stochastic loss reserve model provides an arbitrarily large number of equally likely parameter sets that enable one to simulate future cash flows of the liability. Using these parameter sets to represent all future outcomes, it is possible to describe any future state in the model's time horizon including those states necessary to calculate a cost of capital risk margin. This paper shows how to use the MCMC output to: (1) Calculate the risk margin for an "ultimate" time horizon; (2) Calculate the risk margin for a one-year time horizon; and (3) Analyze the effect of diversification in a risk margin calculation for multiple lines of insurance.

Key Words

Stochastic Loss Reserving, Bayesian MCMC, Capital Requirements, Risk Margins

1 Introduction

With the growing influence of Bayesian MCMC models in stochastic loss reserving such as Meyers (2015) this paper will illustrate one way to use such a model to calculate a cost of capital risk margin for non-life insurance liabilities. The need for such a calculation is found in the "technical provisions" specified in the European Union's Solvency II act.[1]

These technical provisions refer to the insurer's liability for unpaid losses. Specifically:

1. "The value of the technical provisions shall be equal to the sum of a best estimate and a risk margin."

2. "The best estimate shall correspond to the probability-weighted average of future cash flows, taking account of the time value of money using the relevant risk-free interest rate term structure."

3. "The risk margin shall be calculated by determining the cost of providing an amount of eligible own funds equal to the Solvency Capital Requirement necessary to support the insurance obligations over the lifetime thereof."

4. "Insurance undertakings shall segment their insurance obligations into homogeneous risk groups, and as a minimum by lines of business, when calculating the technical provisions."

A Bayesian MCMC stochastic loss reserve model provides an arbitrarily large number (say 10,000) of equally likely parameter sets that enable one to simulate future cash flows of the liability. From these parameter sets, it is possible to describe any future state in the model's time horizon including those states necessary to calculate the technical provisions. That is what this paper will do.

Here is a high-level description of that cash flow.

1. At the end of the current calendar year (call this time $t = 0$), the insurer posts its best estimate of the liability. The insurer also posts the amount of capital, C_0, needed to contain the uncertainty in this estimate. It invests C_0 in a fund that earns income at the risk-free interest rate i.

[1] The provisions quoted here are stated in Section 2, Article 77 and Article 80, of Chapter VI of the act, p 222. http://register.consilium.europa.eu/pdf/en/09/st03/st03643-re01.en09.pdf.

2. At the end of the next calendar year, at time $t = 1$, the insurer uses its next year of loss experience to reevaluate its liability. It then posts its updated estimate of the liability and the capital, C_1, needed to contain the uncertainty in this estimate. The difference between $C_0 \cdot (1 + i)$ and C_1 is returned to the investor. If that difference is negative, as it occasionally will be, the investor is expected to contribute an amount to make up that difference.

3. The process continues for future calendar years, t, with the amount,

$$C_{t-1} \cdot (1 + i) - C_t,$$

being returned to (or being contributed by) the investor.

4. At some time $t = u$, the loss is deemed to at ultimate, i.e. no significant changes in the loss is anticipated and so we set $C_t = 0$ for $t > u$. For the examples in this paper, $u = 9$.

The present value, discounted at the risky rate r, of the amount returned is equal to

$$\sum_{t=1}^{u} \frac{C_{t-1} \cdot (1 + i) - C_t}{(1 + r)^t}.$$

Since $r > i$, this present value will be less than the initial capital investment of C_0. To adequately compensate the investor for taking on the risk of insuring policyholder losses, the difference can be made up at time $t = 0$ by what we now define as the cost of capital risk margin, R_{COC}.

$$R_{COC} \equiv C_0 - \sum_{t=1}^{u} \frac{C_{t-1} \cdot (1 + i) - C_t}{(1 + r)^t} = (r - i) \cdot \sum_{t=0}^{u} \frac{C_t}{(1 + r)^t} \tag{1}$$

with the second equality coming after some algebraic manipulations.

Note that R_{COC} is similar to, but not identical to, the Solvency II risk margin.

$$R_{SII} \equiv (r - i) \cdot \sum_{t=0}^{u} \frac{C_t}{(1 + i)^t} \tag{2}$$

The problem that now needs to be addressed is the calculation of the C_ts. A straightforward way to project a future cash flow for this process would be to take a fitted Bayesian MCMC model and simulate an additional calendar year of losses for $t = 1$. Then fit another

Bayesian MCMC model to the original data and the simulated data to get the loss estimate and capital requirements for $t = 1$. Then continue this process for $t = 2, \ldots, u$.

While the execution speed of Bayesian MCMC software has significantly increased in recent years, repeating this for 10,000 simulated future cash flows would undoubtedly strain the patience of most practicing actuaries. This paper will propose a faster, but conceptually identical, way to calculate the capital requirements for this process.

Now that we have defined the cost of capital risk margin, here is the route this paper will take to address the problems that need to be solved to calculate the risk margin.

- First we show how to use the Bayesian MCMC machinery to calculate the cash flows and corresponding loss estimates implied by the model.
- Then we show how to calculate the best estimate and the risk margins from the cash flows.
- Then we will investigate the effect of insurer size and line of business on risk margins.
- Then we will address the effect of diversification by line of business.

While the examples this paper focus on an "ultimate" time horizon, jurisdictions such as the European Union require insurers to calculate their capital requirements and their risk margin based on a one-year time horizon. The final section will show, with an example, how to adjust the models so that the one-year time horizon can be incorporated within the framework of this paper.

The data for the examples in this paper are taken from the CAS Loss Reserve Database. The data consist of 50 loss triangles in the Commercial Auto (CA), Personal Auto (PA), Workers' Compensation (WC) and Other Liability (OL) lines of insurance. The loss triangles used in this paper were selected from the list given in Appendix A of Meyers (2015).

The algorithms described in this paper are computationally intensive. As one reads this paper, they might question if the computations can be done in a reasonable time. The answer is yes. The scripts that are included with the paper were run on my standard issue high-end laptop. The run times for the calculations are about two minutes per loss triangle for the model in Section 3 and about seventeen minutes per triangle for the model in Section 5.

2 Cash Flows and Statistics of Interest

This paper will use the Changing Settlement Rate (CSR) model described in Meyers (2015) as modified in Meyers (2016). As shown in these papers, this model has been successfully validated on the lower triangle holdout data for a set of 200 loss triangles, 50 from each of four lines of business. The model is fit to a cumulative paid loss triangle, $T_0 \equiv \{X_{w,d}\}$ where the accident year, $w = 1, \ldots, 10$ and the development year, $d = 1, \ldots, 11 - w$. This model allows for accident year effects, development year effects and a variable claim settlement rate. The details of the model are in the references above. What is relevant for this paper is that given the loss triangle, T_0, the model uses Bayesian MCMC to obtain a sample of 10,000 equally likely lognormal,$\{\mu_{w,d}^j, \sigma_d^j\}_{j=1}^{10{,}000}$, parameter sets from the posterior distribution, $\{\mu_{w,d}, \sigma_d | T_0\}$. This paper assumes that these parameter sets can be used to describe the possible future cash flows by a simulation.

With these parameter sets we can calculate the best estimate as the probability weighted average of the present value of expected future cash flows. This will be equal to the expected value of the differences in the cumulative payments between development years, i.e.

$$E_{Best} = \frac{\sum_{j=1}^{10{,}000} \sum_{w=2}^{10} \sum_{d=12-w}^{10} exp\left(\mu_{w,d}^j + (\sigma_d^j)^2/2\right) - exp\left(\mu_{w,d-1}^j + (\sigma_{d-1}^j)^2/2\right)}{10{,}000 \cdot (1+i)^{w+d-11.5}} \tag{3}$$

This calculation assumes that the losses are paid one half year before the end of future calendar year $t = w + d - 11$.

For the scope of this paper, let's also select the ultimate loss, U_j, associated with the j^{th} parameter set to be the sum of the expected values of the losses for $d = 10$ over all the accident years. i.e.,

$$U_j = \sum_{w=1}^{10} exp\left(\mu_{w,10}^j + (\sigma_{10}^j)^2/2\right) \tag{4}$$

For those wishing to allow for loss development after $d = 10$, I suggest that a Bayesian MCMC version of Clark (2003) would be a good place to begin.

For the lower triangle of $\{X_{w,d}^j\}_{j=1}^{10{,}000}$, define the simulated loss trapezoid for future calendar year t that includes the upper loss triangle, T_0, and the first t diagonals of from the lower loss triangle, i.e.

$$T_t^j \equiv \begin{cases} X_{w,d} \text{ for } w = 1, \ldots, 10 \text{ and } d = 11 - w, \ldots, 10 \\ X_{w,d}^j \text{ for } w = t+1, \ldots, 10 \text{ and } d = 12 - w, \ldots, \min(11 - w + t, 10) \end{cases} \tag{5}$$

where $X^j_{w.d}$ is simulated from a lognormal distribution with parameters $\mu^j_{w,d}$ and σ^j_d.

Let's temporarily drop the assumption that we know the parameter set index j. All we have is an observed loss trapezoid, T_t. Then using Bayes' Theorem and the fact that initially, all j are equally likely, the probability that the parameter set index is equal to j given T_t is given by

$$\Pr\left[J = j|T_t\right] = \frac{\prod\limits_{X_{w,d}\in T_t} \phi\left(\log(X_{w,d})|\mu^j_{w,d}\sigma^j_d\right)}{\sum\limits_{k=1}^{10{,}000} \prod\limits_{X_{w,d}\in T_t} \phi\left(\log(X_{w,d})|\mu^k_{w,d}\sigma^k_d\right)} \tag{6}$$

where ϕ is the probability density function for the normal distribution.

At this point, there are a number of options one can choose to calculate the various statistics that are of interest to insurer risk managers. For example, given T_t, one could calculate the ultimate loss estimate, E_t as

$$E_t \equiv E\left[\sum_{w=1}^{10} X_{w,10}|T_t\right] = \sum_{j=1}^{10{,}000} \Pr\left[J = j|T_t\right] \cdot U_j. \tag{7}$$

If one accepts that the Bayesian MCMC output as representative of all future scenarios, Equation 7 is exactly the right calculation for the loss estimate given T_t. But let's consider what one should do to calculate, say, the 99.5th percentile. First one should sort the scenarios in order of increasing U_j. It is not uncommon to find a case where there is a scenario, j, with $\Pr[J \leq j|T_9] = 0.9900$ and $\Pr[J \leq j+1|T_9] = 0.9960$. For cases such as this, I tried a linear interpolation that occasionally yielded small discretization errors that gave theoretically impossible results[2].

To avoid these annoying cases, I decided to calculate the statistics of interest by first taking a random sample of size 10,000 (with replacement), $\{S_t\}$, of the U_js with sampling probabilities $\Pr[J = j|T_t]$. This is subject to an additional simulation error, but it should be small.

The "statistics of interest" for risk margin are, for $t = 0, \ldots, 9$:

1. The mean, E_t, which is equal to the arithmetic average of $\{S_t\}$.
2. The Tail Value-at-Risk at the α level (TVaR@α), which is equal to the arithmetic average of the $(1-\alpha)\cdot$10,000 highest values of $\{S_t\}$[3].

[2] Such as a negative capital when the required assets were determined by the TVaR measure of risk.

[3] While this paper does not use the Value-at-Risk (VaR) in its examples, one could calculate the VaR@α as the $(1-\alpha)\cdot 10{,}000^{th}$ highest value of $\{S_t\}$.

Let's denote the total required capital by $C_t \equiv \text{TVaR@}\alpha - E_t$.

We summarize the above in the following algorithm.

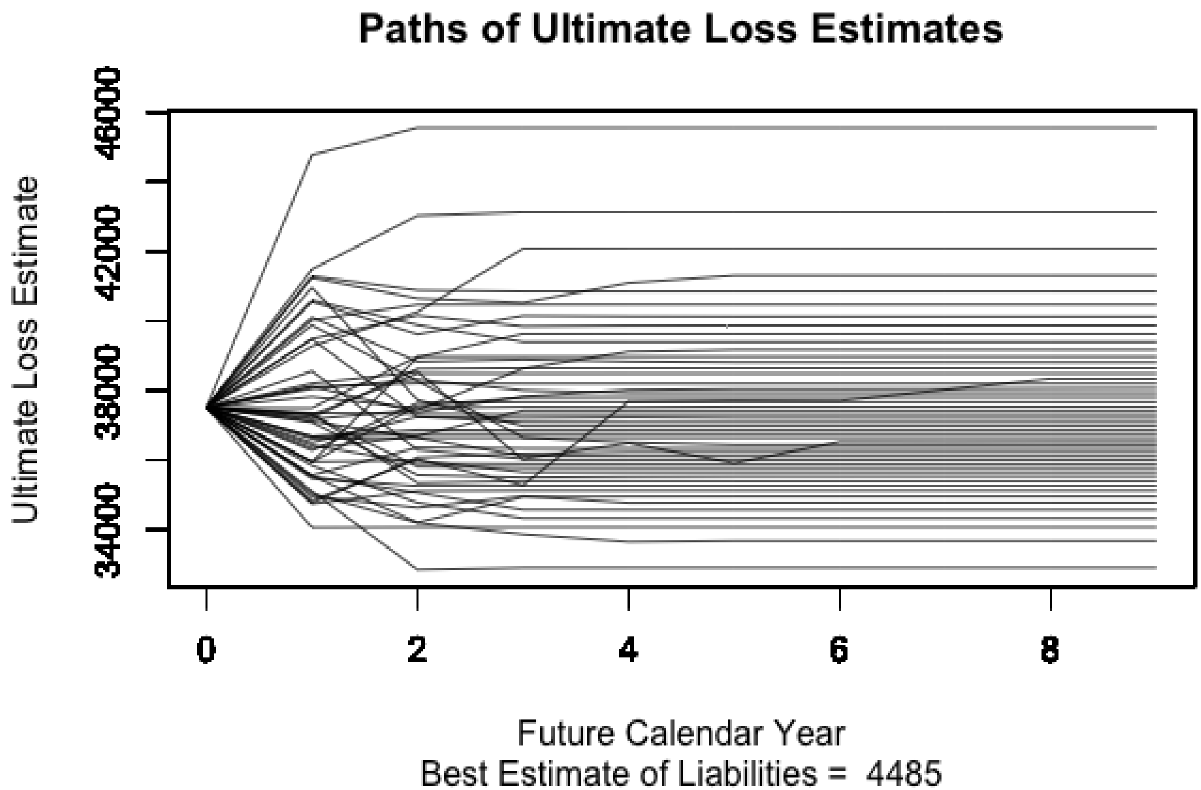
Algorithm 1 Calculate Capital Scenarios

for $k = 1, \ldots, 10,000$ **do**
  for $t = 0, \ldots, 9$ **do**
    Simulate cash flows $\{T_t^k\}$ using the parameter set $\{(\mu_{w,d}^k, \sigma_d^k)\}$
    Use Equation 6 to calculate $\Pr\left[J = j | T_t^k\right]$ for each $j = 1, \ldots, 10,000$
    Take a random sample of size 10,000 with replacement, $\{S_t^k\}$, of $\{U_j\}_{j=1}^{10{,}000}$ with sampling probabilities $\Pr\left[J = j | T_t^k\right]$.
    Set E_t^k equal to the arithmetic average of $\{S_t^k\}$.
    Set C_t^k equal to the arithmetic average of the highest $(1-\alpha)\cdot 10{,}000$ highest values of $\{S_t^k\}$, minus E_t^k.
  end for
end for

The examples in this paper use $\alpha = 97\%$. This selection is for illustrative purposes only.

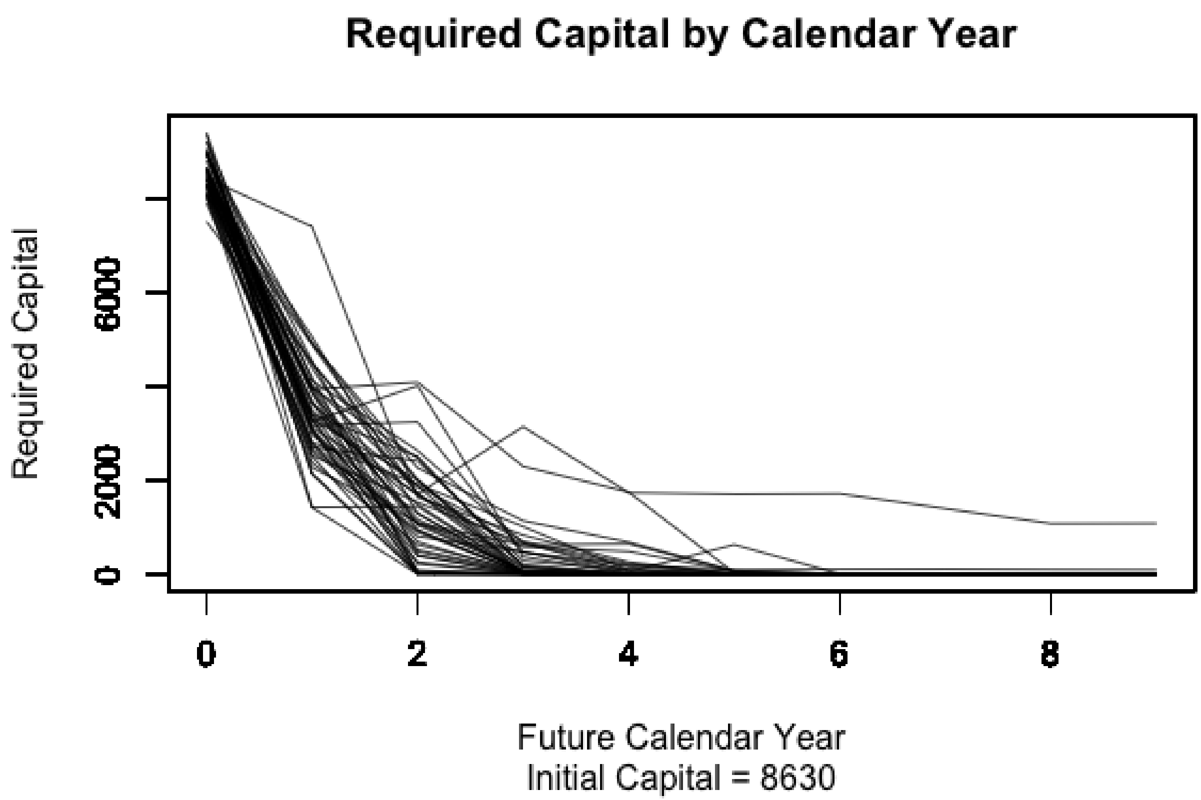
Calculating E_t^k for $t = 0, \ldots, 9$ yields the k^{th} path that the loss estimate takes as it moves toward its ultimate value. Of interest for what follows is the set of possible paths that the loss estimate can take. Figure 1 shows the paths for the paths that contain the 100th, the 300th,. . ., and the 9,900th highest E_9^ks of Insurer #353 for Commercial Auto in the CAS Loss Reserve Database. This figure illustrates that the E_t^ks tend to become more certain over time.

Figure 1


Paths of Ultimate Loss Estimates
Ultimate Loss Estimate
34000
38000
42000
46000
0
2
4
6
8
Future Calendar Year
Best Estimate of Liabilities = 4485

Also of interest is the paths of the required capital, C_t^k, for $t = 0, \ldots, 9$. Figure 2 shows the paths of C_t^k that correspond to the paths taken by E_t^k in Figure 1. This figure illustrates that as the estimates of the E_t^ks become more more certain, the required capital, C_t^k, tends to decrease over time.

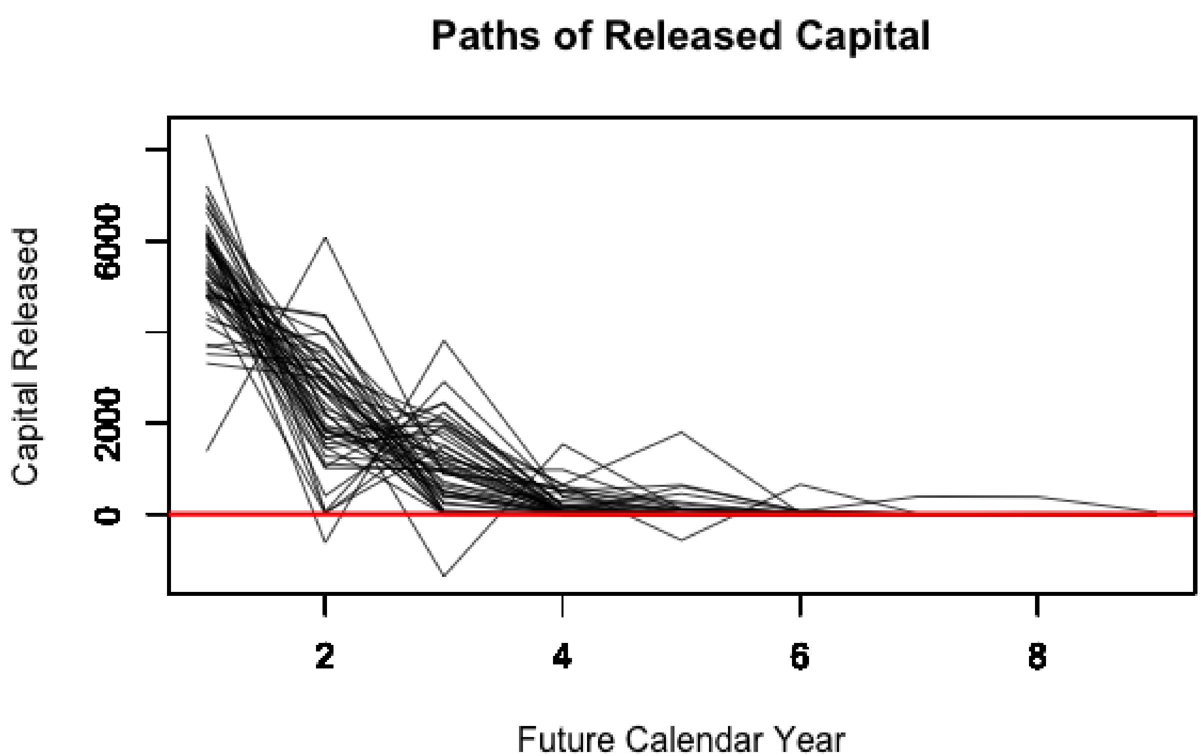
Figure 2



3 Risk Margins

This section applies the cost of capital risk margin formula, given by Equation 1, to the set of required capital paths, $\{C_0^k, \ldots, C_9^k\}_{k=1}^{10,000}$. Recall that the that formula defined the cost of capital risk margin as the present value of the capital released as the loss reserve liability becomes more certain. Figure 3 shows the paths of released capital that correspond to the paths taken by the C_t^ks in Figure 2. In general, this figure shows that most of the capital gets released early on, and that occasionally it is necessary to add capital.

Figure 3



Applying Equation 1 we get for each k

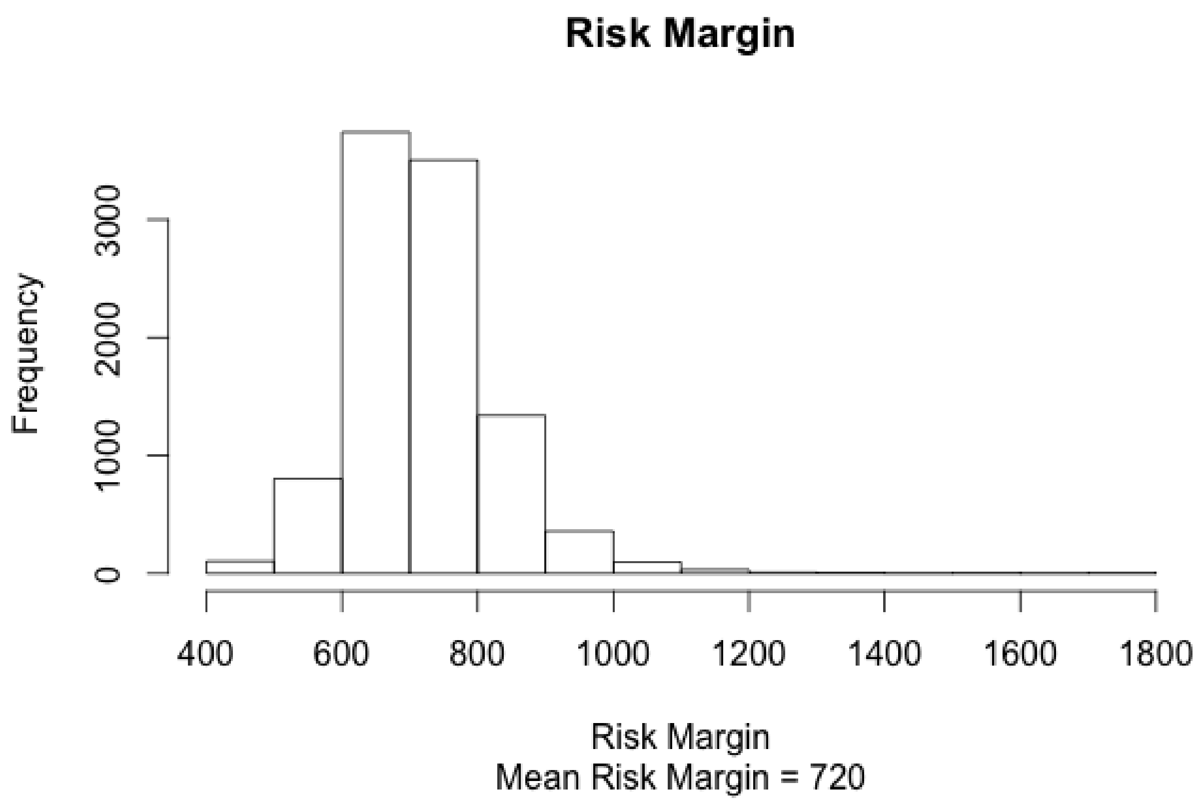
$$R_{COC}^k \equiv C_0^k - \sum_{t=1}^{u} \frac{C_{t-1}^k \cdot (1+i) - C_t^k}{(1+r)^t} \tag{8}$$

Then the risk margin is given by

$$R_{COC} = \frac{1}{10,000} \sum_{k=1}^{10,000} R^k_{COC} \tag{9}$$

Figure 4 shows a histogram of the R^k_{COC}s for our example.

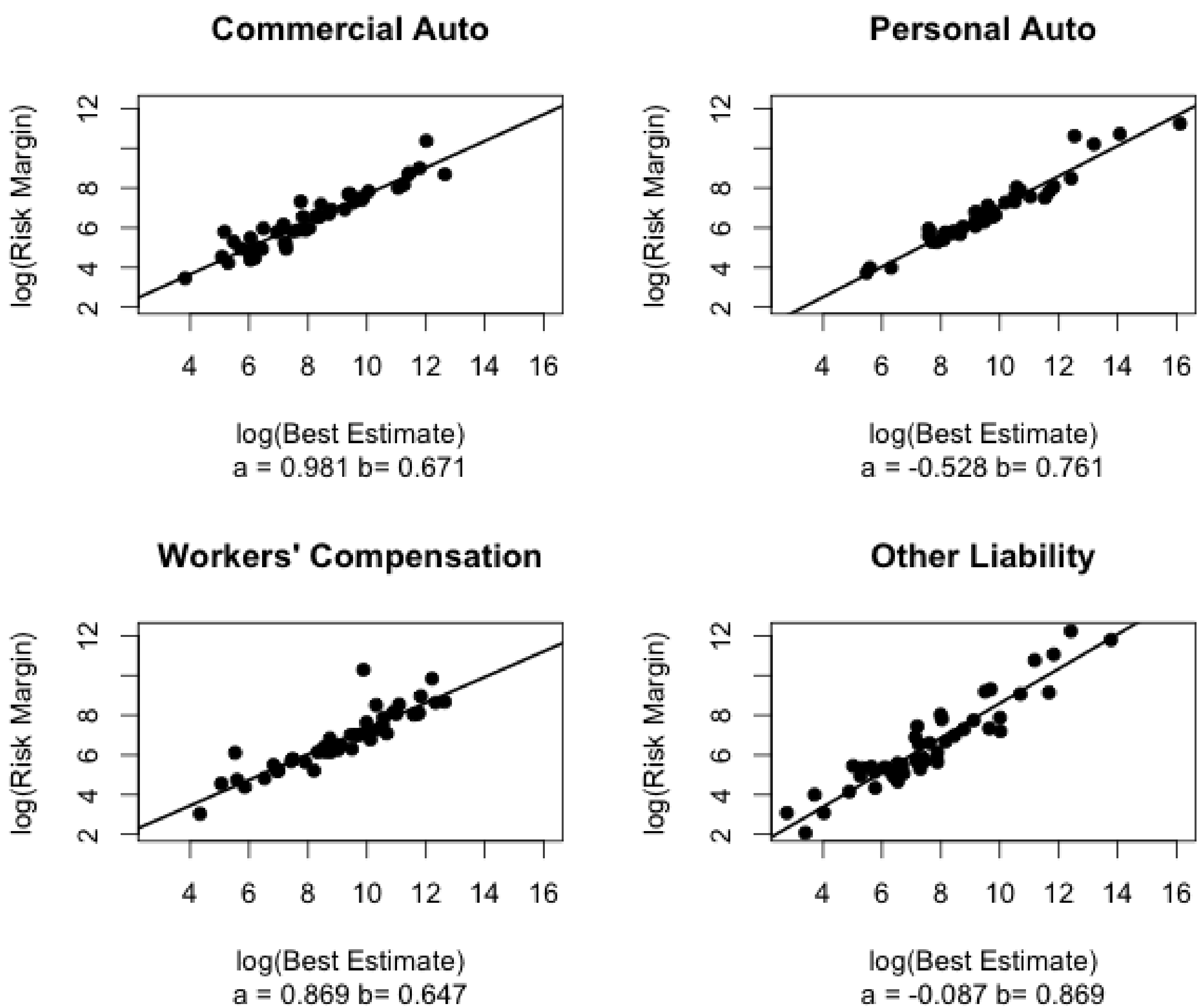
Figure 4



Of interest is the ratio of the risk margin and the size of the best estimate. To investigate, I calculated the risk margins for all 200 loss triangles in our data. After some exploratory analysis, I concluded that: (1) there are significant differences by line of business; and (2) there is an approximate linear relationship between the log of the risk margin and the log of the best estimate. Figure 5 shows the plots of the $log(R_{COC})$ against $log(E_{Best})$, along with the coefficients of an ordinary linear regression of the form

$$log(R_{COC}) = a + b \cdot log(E_{Best}) \tag{10}$$

Figure 5[4]



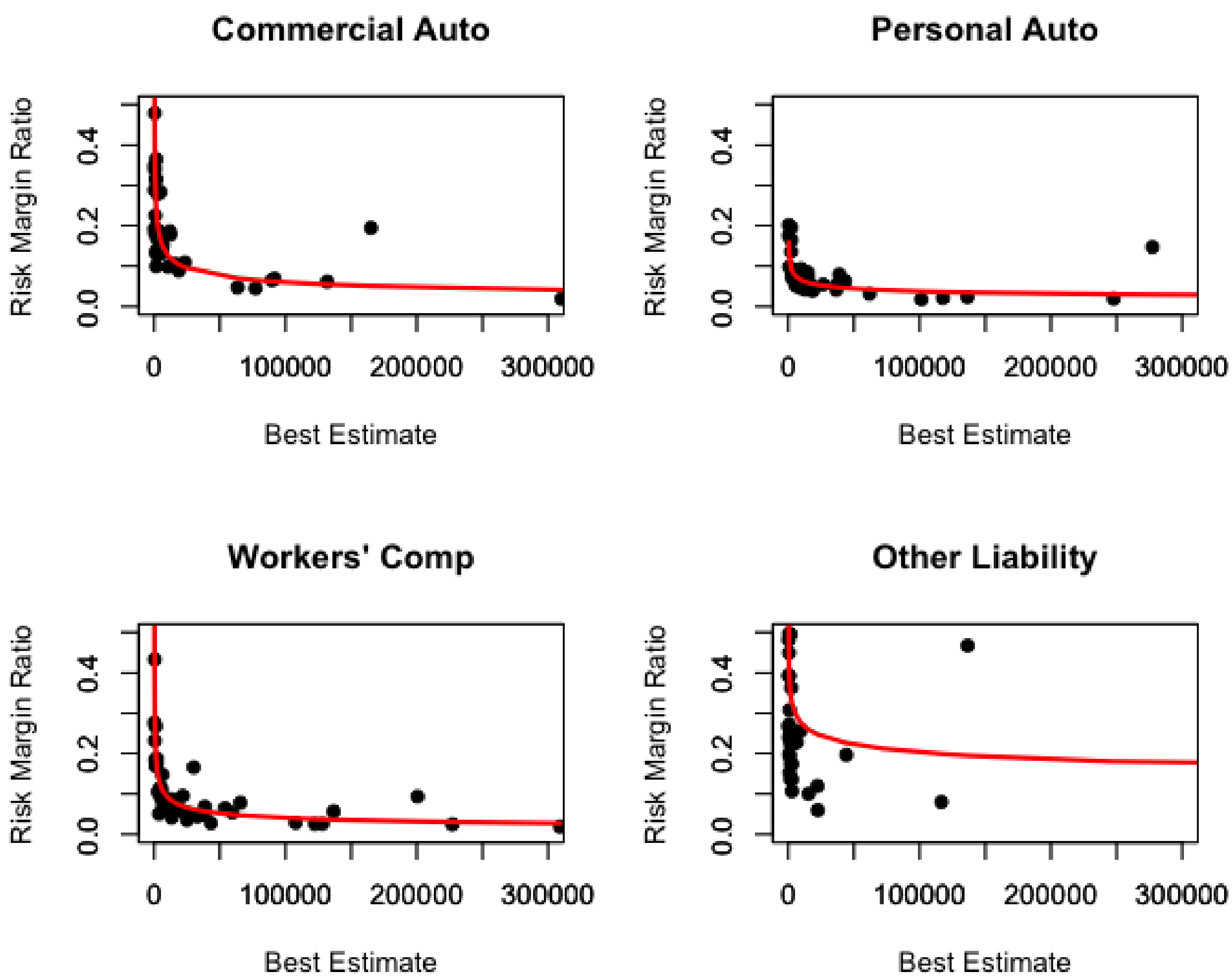
[4]Three small volatile insurers had negative best estimates and were excluded from the linear regression.

We can rewrite Equation 10 in the form

$$\frac{R_{COC}}{E_{Best}} = e^a \cdot (E_{Best})^{b-1} \tag{11}$$

Note from Figure 5 that that $b < 1$ for all four lines of insurance. This implies that the risk margin to best estimate ratio decreases as the best estimate increases. As Figure 6 shows the ratio can be quite high for insurers with small best estimates. I can see where some insurers might object, especially if the line with the high ratio is a small part of the insurer's book of business.

Figure 6[5]



[5]A small number of estimates fell outside the range of these figures.

4 Diversification

As stated in the introduction, the EU Solvency II provision states explicitly that "Insurance undertakings shall segment their insurance obligations into homogeneous risk groups, and as a minimum by lines of business, when calculating the technical provisions." This means that the total risk margin for a multiline insurer is the sum of the risk margins over its individual lines of business.

Longtime observers of the insurance business have long recognized that multiline insurers benefit from the diversification of their risk of loss. This being the case, they might well want to reflect the benefits of diversification in their risk margins. The problem with a formal recognition of diversification is that the benefits have been difficult to quantify. What many are afraid of is the possibility that significant losses from the different lines of business could happen at the same time. This possibility is often referred to as the "dependency problem."

As such, the Solvency II non-recognition of diversification may appear to some to be prudent.

Mathematical tools that can be used to describe dependency have been available for quite some time. See, for example, Frees and Valdez (1998) and Wang (1998). The main tool described in these papers is called a copula, which is a multivariate distribution on an L-dimensional unit hypercube in which the marginal distributions have a uniform(0,1) distribution. Given a copula $\mathscr{C}$ and samples $\{_lS_t^k\}$, (see Section 2) for each line l of L lines of business, one begins to calculate R_{COC} by first executing the following algorithm.

Algorithm 2 Calculate Samples for Dependent Lines

for $k = 1, \ldots, 10,000$ **do**
  for $t = 1, \ldots, 9$ **do**
    Simulate an L-tuple vector $\{P_l^k\}_{l=1}^L$ of uniform(0,1) numbers from the copula $\mathscr{C}$.
    For each line of business, l, select $_lQ_t^k$ to be the $P_l \cdot 10,000$ highest value of $\{_lS_t^k\}$.
  end for
  Set the total ultimate loss $_TS_t^k = _1Q_t^k + \cdots + _LQ_t^k$.
end for

Use the output of this algorithm to calculate, $\{_TC_t^k\}_{k=1}^{10,000}$ for $t = 1, \ldots, 9$, and Equations 8 and 9 to calculate $_TR_{COC}$.

So if one believes that the lines of business are correlated, it is possible to calculate the risk margin for the total liability that reflects whatever diversification that is warranted by one's choice of a dependency structure. As it turns out, there has been some recent empirical work on determining that structure.

Let's first look at Avanzi, Taylor and Wong (2016). The point of their paper is that correlations can arise from an inappropriate model. To quote their abstract – "We show with some real examples that, sometimes, most (if not all) of the correlation can be "explained' by an appropriate methodology. Two major conclusions stem from our analysis."

1. "In any attempt to measure cross-LoB correlations, careful modeling of the data needs to be the order of the day. The exercise will not be well served by rough modeling, such as the use of simple chain ladders, and may indeed result in the prescription of excessive risk margins and/or capital margins."

2. "Such empirical evidence as examined in the paper reveals cross-LoB correlations that vary only in the range zero to very modest. There is little evidence in favor of the high correlation assumed in some jurisdictions. The evidence suggests that these assumptions derived from either poor modeling or a misconception of the cross-LoB dependencies relevant to the purpose to which they are applied."

Meyers (2016) arrives at a similar conclusion. This paper first shows how to fit a bivariate CSR model, that allows for dependencies, to triangles for two lines of business from the same insurer. It then compares the fit of the bivariate model to a similar bivariate model that assumes independence for 102 within insurer pairs. Taking into account the additional parameter introduced by the dependent model, it concludes that the model assuming independence has a better fit for all 102 pairs of triangles.

In other words, the appropriate dependency structure is to assume that the lines of business are independent. This assumes, as demonstrated in Meyers (2016) for the CSR model used in this paper, that careful modeling has been carried out.

The independence assumption allows us to simplify the procedure described at the beginning of this section. Given the samples $\{{}_lS_t^k\}$,For each line l of L lines of business, one begins to calculate R_{COC} by first executing following algorithm.

Algorithm 3 Calculate Samples for Independent Lines

for $k = 1, \ldots, 10,000$ **do**
    for $t = 1, \ldots, 9$ **do**
        Set the total ultimate loss sample to be $\{{}_TS_t^k\} = \{{}_1S_t^k\} + \cdots + \{{}_LS_t^k\}$.
    end for
end for

Use the sample,$\{{}_TS_t^k\}$, to obtain $\{{}_TC_t^k\}_{k=1}^{10,000}$ for $t = 1, \ldots, 9$. Then use Equations 8 and 9 to calculate the combined risk margin, ${}_TR_{COC}$.

The combined risk margins in this paper were calculated using the independence assumption. This choice was *not* made for mathematical convenience. Meyers (2016) shows how to estimate the parameters of a model with dependency between the lines. The steps outlined at the beginning of this section show how to implement a dependency assumption if warranted.

From the loss triangles studied in Meyers (2015) there were five insurers with a loss triangle in all four lines. Table 1 gives the combined risk margin for these five insurers in the "Total" rows in the "Allocated Risk Margin" column. Over all five insurers, the diversification credit,

$$1 - \frac{\text{Combined Risk Margin}}{\text{Total Standalone Risk Margin}},$$

ranged from 30.3% to 48.3%.

Of interest, if not essential, is to see how this combined risk margin is allocated down to the individual lines of insurance. Allocating the cost of capital to individual lines is more important for pricing than for financial reporting as the former case requires an insurer to quote a price for an individual insurance contract. For the latter case, a risk margin need only apply to the total insurer liabilities.

Allocating the cost of capital has been debated in the actuarial profession for decades. About 15 years ago, there were a number of papers that address the issue in a pricing context. Mango and Ruhm (2003) and Meyers (1999) are two of many papers that were published around then. Forgoing the seemingly endless discussion that accompanies this topic, this paper allocates combined capital to lines of insurance in proportion to the lines marginal cost of capital.

Once one has done the coding necessary to calculate the combined risk margin, it takes only a little additional computer run time to allocate the combined risk margin to individual lines. So let's proceed.

Given the samples, $\{_lS_t^k\}$, for each line l of L lines of business, one begins to calculate marginal cost of capital for line l, $_{(l)}R_{COC}$, by first executing Algorithm 4 below. Then for each line l execute Algorithm 5.

The fourth column of Table 1 gives the marginal cost of capital, $_{(l)}R_{COC}$, by insurer for each line of insurance. Note that the sum of the marginal cost of capitals by line is less than the combined cost of capital in the "Total" column. We then allocate the cost of capital by line of insurance in proportion to the marginal capital by

$$_{(l)}R_{ACOC} \equiv _{(l)} R_{COC} \cdot \frac{_TR_{COC}}{_{(1)}R_{COC} + \cdots + _{(L)} R_{COC}} \qquad (12)$$

Note that there are many instances where the diversification credit is in excess of 80%. This occurs when a "small" line of insurance is part of the portfolio of a "large" insurer. Regardless of what one thinks of allocating the cost of capital, one cannot deny that a "small" line of insurance adds little to the risk of a large insurer. The insurer size effect illustrated in Figure 6 can be significantly reduced by taking diversification into account.

Algorithm 4 Calculate Leave-Line-Out Samples

---

for $k = 1, \ldots, 10,000$ **do**
    Set the total ultimate loss sample to be $\{_T S_t^k\} = \{_1 S_t^k\} + \cdots + \{_L S_t^k\}$.
    for $t = 1, \ldots, 9$ **do**
        Set the leave-line-out ultimate loss sample for line l to be $\{_{(-l)} S_t^k\} = \{_T S_t^k\} - \{_l S_t^k\}$.
    end for
end for

Algorithm 5 Calculate Marginal Cost of Capital

---

for $l = 1, \ldots L$ **do**
    for $t = 1, \ldots 9$ **do**
        Use the sample, $\{_{(-l)} S_t^k\}$, to calculate the leave-line-out capital, $\{_{(-l)} C_t^k\}_{k=1}^{10{,}000}$.
    end for
    Use Equations 8 and 9 to calculate the leave-line-out cost of capital risk margin, $_{(-l)}R_{COC}$.
    Calculate the marginal cost of capital risk margin, $_{(l)}R_{COC} \equiv {}_T R_{COC} - {}_{(-l)}R_{COC}$.
end for

Table 1

Grp./Line	Estimated Ult. Loss	Best Esitmate	Marginal Risk Margin	Allocated Risk Margin	Standalone Risk Margin	Diver. Credit
1528/CA	88,756	13,822	464	852	1,447	41.1%
PA	311,659	36,507	542	996	1,519	34.4%
WC	129,762	13,207	61	111	550	79.8%
OL	19,143	4,697	243	447	1,065	58.0%
Total	549,320	68,233	1,310	2,406	4,581	47.5%
1767/CA	2,205,897	310,203	108	171	5,963	97.1%
PA	90,312,996	9,921,107	20,620	32,566	76,527	57.4%
WC	1,677,179	227,010	175	276	5,637	95.1%
OL	2,443,660	956,344	76,495	120,812	132,428	8.8%
Total	96,639,732	11,414,664	97,398	153,825	220,555	30.3%
3240/CA	97,298	18,684	346	554	1,653	66.5%
PA	1,092,757	136,373	1,862	2,983	3,208	7.0%
WC	38,960	4,155	42	67	467	85.7%
OL	13,774	2,638	36	58	459	87.4%
Total	1,242,789	161,850	2,286	3,663	5,787	36.7%
5185/CA	96,071	23,262	837	1,592	2,544	37.4%
PA	268,908	43,305	952	1,811	2,719	33.4%
WC	100,322	16,768	91	173	1,100	84.3%
OL	140,606	22,440	216	410	1,346	69.5%
Total	605,907	105,775	2,095	3,985	7,709	48.3%
14176/CA	28,929	11,759	982	1,716	2,191	21.7%
PA	144,563	26,494	380	663	1,439	53.9%
WC	111,498	20,075	263	460	1,229	62.6%
OL	5,290	1,287	35	61	326	81.3%
Total	290,280	59,615	1,660	2,900	5,185	44.1%

5 One-Year Time Horizon

The risk margin calculations above assumed an "ultimate" time horizon to establish the required capital. Some regulatory jurisdictions, e.g. Solvency II, specify that the insurer should assume a one-year time horizon. This section extends the methodology of the previous sections to cover the one-year time horizon.

A high-level description of the methodology is to use a Bayesian MCMC model to obtain 10,000 equally likely scenarios that represent the future evolution of the line of business that produced the loss triangle. Then, as new losses come in, it uses Bayes' Theorem to update the probability of each scenario. From these updated probabilities, one then calculate the statistics that are needed to calculate the risk margin.

A key step in this methodology is to assign a unique ultimate loss estimate to each scenario. As Figure 1 illustrates, changes in the ultimate loss estimate for the later development periods are relatively rare, so the assignment of the ultimate loss estimate, U_j, specified in Equation 4, to the j^{th} scenario is a good approximation.

However as Figure 1 also illustrates, there is significant uncertainty in the ultimate loss after one additional year of development. So the assignment of the ultimate loss estimate of U_j to the j^{th} scenario is not a good approximation.

Under a one-year time horizon capital requirement, the capital is determined by the estimate of the ultimate losses after one more calendar year of loss experience. To calculate the risk margin we will need the distribution of ultimate loss estimates at the end or each calendar year. These estimates will depend upon the calendar year, t.

To get a good approximation, $O_{t,j}$, of the expected ultimate loss for the j^{th} scenario, one can simulate future loss experience from the parameter set of that scenario and calculate the ultimate loss estimate, M times. Then set $O_{t,j}$ equal to the average of those estimates. The details are in the Algorithm 6 below.

Both the accuracy of the estimate of $O_{t,j}$ and the computer run time increase with M. I experimented with different values of M and found that $M = 12$ obtained results that were sufficiently accurate given the intrinsic variation of the underlying MCMC simulation.

Use Algorithm 7 to calculate the risk margin for the one year time horizon. In this algorithm, one simply substitutes $O_{t+1,j}$ for U_j in the 5^{th} step of Algorithm 1. Given the output of Algorithm 7, one then calculates risk margins using Equations 8 and 9.

Algorithm 6 Calculate Scenario Estimates by Calendar Year

```
for m = 1, ..., M do
    for j = 1, ..., 10,000 do
        for t = 1, ..., 9 do
            Simulate T_t using the parameters (μ^j_{w,d}, σ^j_d).
            Use Equation 6 to calculate {Pr[N = n|T_t]}_{n=1}^{10,000}.
            Use Equation 7 to calculate the ultimate loss estimate, O^m_{t,j}.
        end for
        Set O^m_{10,j} = O^m_{9,j}
    end for
end for
for j = 1, ..., 10,000 do
    for t = 1, ..., 10 do
        Set O_{t,j} = mean(O^m_{t,j}).
    end for
end for
```

Algorithm 7 Calculate Capital Scenarios for a One-Year Time Horizon

```
for k = 1, ..., 10,000 do
    for t = 0, ..., 9 do
        Simulate cash flows {T^k_t} using the parameter set {(μ^k_{w,d}, σ^k_d)}
        Use Equation 6 to calculate Pr[J = j|T^k_t] for each j = 1, ..., 10,000
        Take a random sample of size 10,000 with replacement, {S^k_t}, of the {O_{t+1,j}}_{j=1}^{10,000}
with sampling probabilities Pr[J = j|T^k_t].
        Set E^k_t equal to the arithmetic average of {S^k_t}.
        Set C^k_t equal to the arithmetic average of the highest (1 − α)·10,000 highest values
of {S^k_t}, minus E^k_t.
    end for
end for
```

Figures 7-9 show the one-year time horizon capital paths, release paths and risk margins of Insurer #353 for Commercial Auto that correspond to Figures 2, 3 and 4, respectively for the ultimate time horizon.

Figure 7

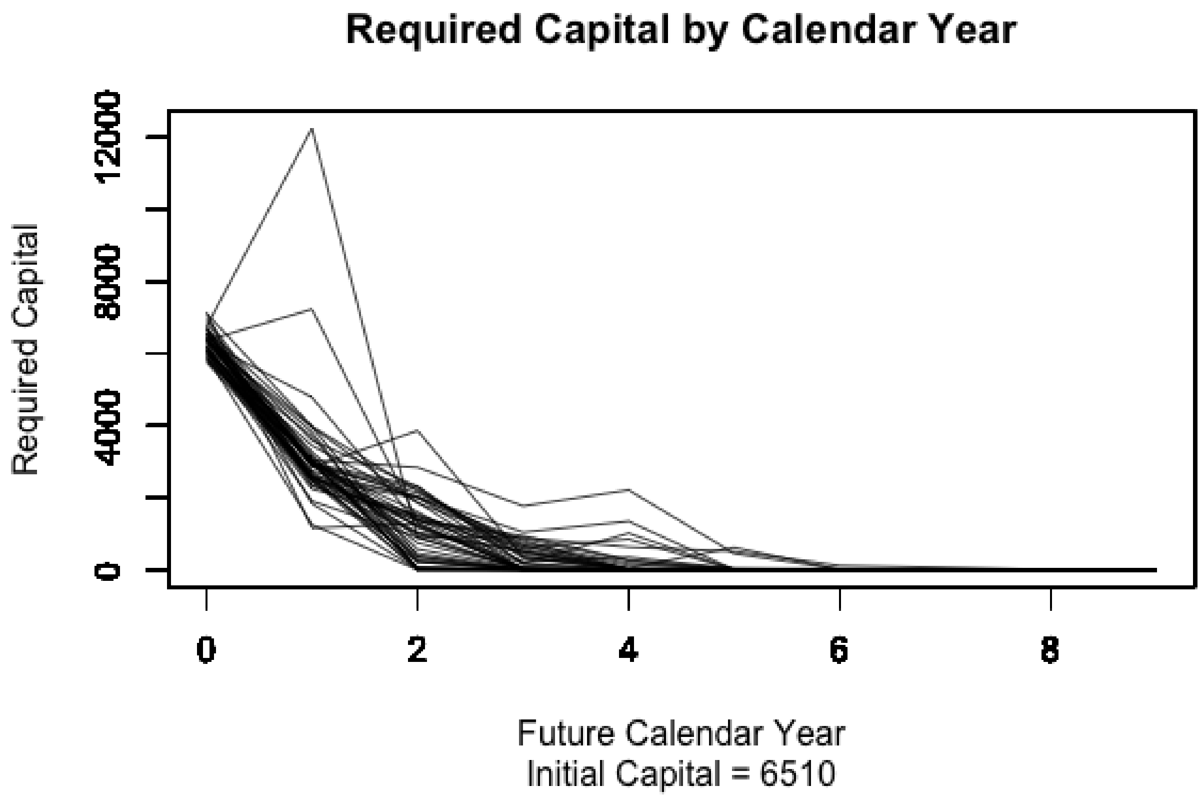


Figure 8

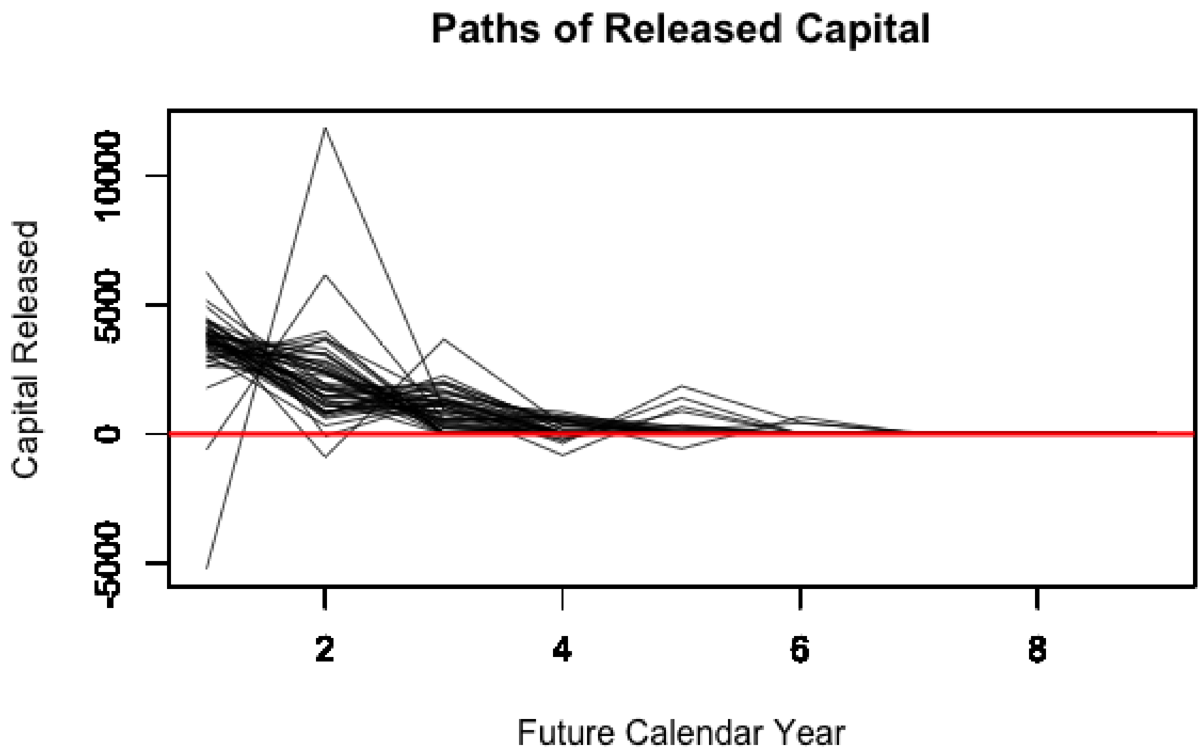
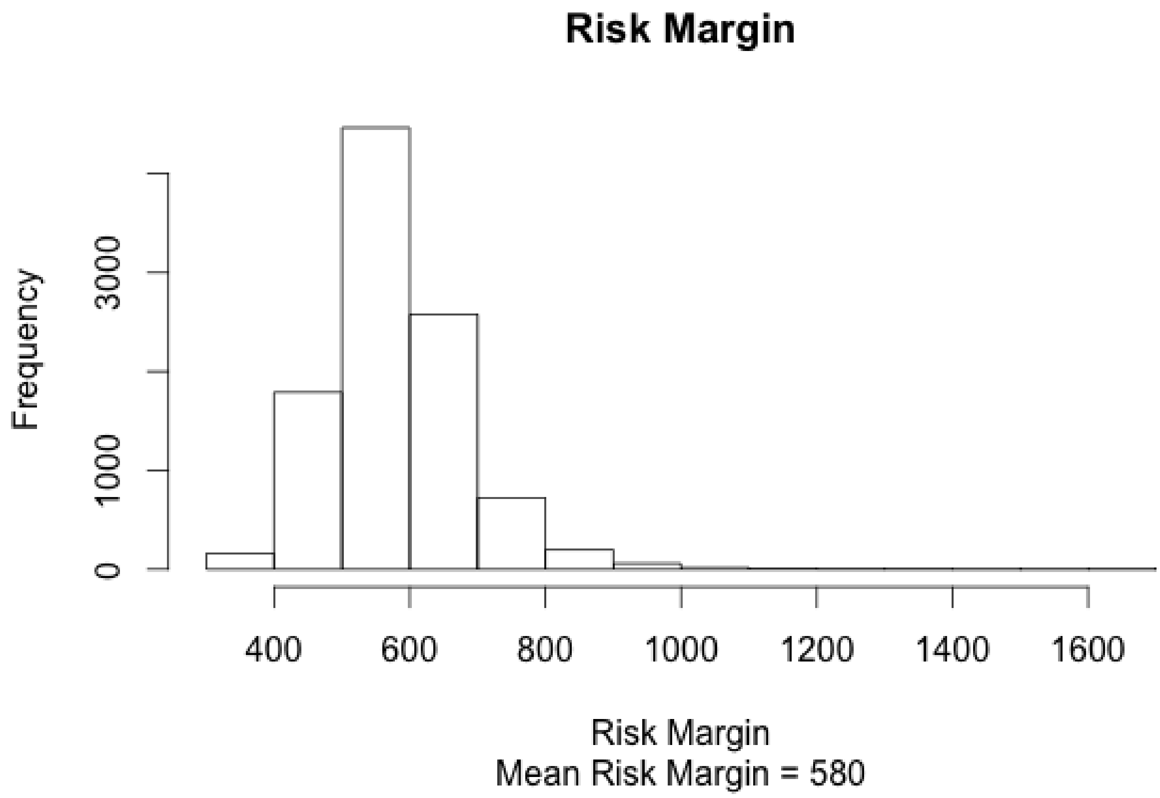

Paths of Released Capital
Capital Released
10000
5000
0
-5000
2
4
6
8
Future Calendar Year

Figure 9


Risk Margin
Frequency
0
1000
3000
400
600
800
1000
1200
1400
1600
Risk Margin
Mean Risk Margin = 580

6 Concluding Remarks

There has not been universal agreement on the assumptions underlying a cost-of-capital risk margin formula. Beyond the underlying Bayesian MCMC stochastic loss reserve model, this paper makes the the following key assumptions.

1. The required required assets for an insurer are determined by the TVaR@α measure of risk.

2. The required capital calculation assumes an "ultimate" time horizon.

3. The distribution of outcomes for the different lines of business are independent.

In numerous advisory committee meetings held at International Actuarial Association events, I heard the following argument supporting the one-year time horizon. Insolvency is usually not an instantaneous event. If the insurer finds itself under stress within a year, it will have time to make the necessary adjustments.

At the same meetings I also heard the following heuristic definition of a risk margin. The risk margin is to provide sufficient funds to transfer its liability to another insurer. "Sufficient funds" should include the cost of capital.

My approach to risk margins was governed by the following considerations.

1. The term of such a portfolio risk transfer contract is unlikely to be for a single year, with the risk reverting back to the original insurer at the end of the year.

2. For a multi-line insurer, the risk being transferred is unlikely consist of a single line of insurance.

3. Dependency between lines is model dependent. In Meyers (2016) I demonstrated that the independence assumption is warranted for the CSR model used in this paper.

4. The theoretical advantages of the TVaR@α over the Var@α have been well-documented by Artzner *et. al.* (1999). Whatever computational difficulty there may have been with the TVaR is not an issue with the methodology used in this paper.

In recognition of the fact that reasonable people may differ on their assumptions, this paper points the way to use alternative assumptions. The methodology described in this paper should be readily adopted for any Bayesian MCMC model.

7 Appendix

Included with this paper is a zip archive containing the following.

- RM 1Line.R - The script that produces the risk margin calculations in Sections 2 and 3.
- RM 4Line.R - The script that produces the risk margin calculations in Section 4.
- RM 1Line 1yr.R - The script that produces the risk margin calculations in Section 5.
- Risk Margins for 200 Triangles.xlsx - Risk margin single line calculations for all 200 triangles

The computer language for the scripts is R (https://www.r-project.org.) The computer language for the MCMC calculations is Stan (http://mc-stan.org/interfaces/rstan.html.)

References

1. Artzner, Philippe; Delbaen, Freddy; Eber, Jean-Marc; Heath, David (1999). "Coherent Measures of Risk". Mathematical Finance. 9 (3): 203?228. https://people.math.ethz.ch/~delbaen/ftp/preprints/CoherentMF.pdf

2. Avanzi, B., Taylor, G. and Wong, B. (2016) "Correlations Between Insurance Lines of Business: An Illusion or a Real Phenomenon? Some Methodological Considerations", *ASTIN Bulletin*, 46(2), pp. 225?263. doi: 10.1017/asb.2015.31. https://www.cambridge.org/core/journals/astin-bulletin-journal-of-the-iaa/article/correlations-5BBB6C52A882572B4F4D587984ED10A7

3. Clark, David R.,"LDF Curve-Fitting and Stochastic Reserving, A Maximum Likelihood Approach,"*Casualty Actuarial Society Forum*, Fall 2003. http://www.casact.org/pubs/forum/03fforum/03ff041.pdf

4. Frees, Edward W. and Valdez, Emiliano Andres, "Understanding Relationships Using Copulas," *North American Actuarial Journal* 1998: 2:1, pp. 1-25 https://www.soa.org/Library/Journals/NAAJ/1998/january/naaj9801_1.aspx

5. Kreps, Rodney E., "Reinsurer Risk Loads from Marginal Capital Requirements,"*Proceedings of the Casualty Actuarial Society Casualty Actuarial Society*,1990:LXXVII, p196. http://www.casact.org/pubs/proceed/proceed90/90196.pdf

6. Mango, Donald F., and Ruhm, David L., "A Method of Implementing Myers-Read Capital Allocation in Simulation,"*Casualty Actuarial Society Forum* Fall 2003. http://www.casact.org/pubs/forum/03fforum/03ff451.pdf

7. Meyers, Glenn G., "Underwriting Risk", *Casualty Actuarial Society Forum*,1999: Spring pp.185-220. http://www.casact.org/pubs/forum/99spforum/99spf185.pdf

8. Meyers, Glenn G., *Stochastic Loss Reserving Using Bayesian MCMC Models*, CAS Monograph No. 1, 2015. http://www.casact.org/pubs/monographs/index.cfm?fa=meyers-monograph01

9. Meyers, Glenn G. "Dependencies in Stochastic Loss Reserve Models," submitted for publication in *Variance* An earlier version of the paper is published on the Winter (2016) edition of *CAS EForum*. http://www.casact.org/pubs/forum/16wforum/Meyers.pdf

10. Wang, Shaun S.,"Aggregation of Correlated Risk Portfolios:Models and Algorithms" *Proceedings of the Casualty Actuarial Society Casualty Actuarial Society* 1998: LXXXV pp. 848-939. http://www.casact.org/pubs/proceed/proceed98/980848.pdf

Appendices

Risk Margins for 200 triangles (Excel file)
RM_Line1.R (R file)
RM_Line1_1yr.R (R file)
RM_4Lines.R (R file)