The Technical Provisions in Solvency II
What EU Insurers Could Do if They Had Schedule P

Glenn Meyers, FCAS, MAAA, Ph.D.

Abstract: The goal of this paper is to demonstrate how publicly available data can be used to calculate the technical provisions in Solvency II. This is a purely hypothetical exercise, since the publicly available data is in America, and Solvency II applies to the European Union. Using American Schedule P data, this paper:

Develops "prior information" to be used in an empirical Bayesian loss reserving method.

Uses the Metropolis-Hastings algorithm to develop a posterior distribution of parameters for a Bayesian Analysis.

Develops a series of diagnostics to assess the applicability of the Bayesian model.

Uses the results to calculate the best estimate and the risk margin in accordance with the principles underlying Solvency II.

Develops an ongoing process to regularly compare projected results against experience.

The paper includes analyses of the Schedule P data for four American Insurers based on its methodology.

Keywords: Solvency II, reserving methods, reserve variability, uncertainty and ranges, Bayesian estimation

---

1. INTRODUCTION

In 2009 the European Parliament passed a new act for regulating insurers known as Solvency II. Its objectives include:

- Increased focus on effective risk management, control, and governance,
- Market consistent valuation of assets and liabilities,
- Increased disclosure and transparency.

This act will become effective on October 31, 2012. Because of the growing international nature of the business of insurance, the development of the provisions in this act has been watched and debated by interested parties worldwide.

This paper focuses on calculating the "technical provisions" specified in this act[1]. The "technical provisions" refer to the insurer's liability for unpaid losses. Specifically:

- "The value of the technical provisions shall be equal to the sum of a best estimate and a risk margin."[2]

---

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

[2] Article 77

- "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."[5]

With regard to technical provisions, the act also requires insurers to have "processes and procedures in place to insure that best estimates, and the assumptions underlying the calculation of the best estimates, are regularly compared against experience.

When the comparison identifies systematic deviations between the experience and the best estimate, the insurer shall make appropriate adjustments to the actuarial methods and/or the assumptions being made."[6]

These provisions of the act implicitly, if not explicitly, call for a stochastic model of the loss development process. Details such as the particular models and the data being used are not specified.

In America, insurers are required to report very detailed data to regulators. Relevant to the topic of technical provisions is Schedule P of the National Association of Insurance Commissioners (NAIC) Annual Statement.[7] This data contains net premiums, along with paid and incurred loss triangles spanning a period of ten accident years. The data is organized into 36 specific lines of insurance such as Personal Auto, Commercial Auto, Homeowners, and Workers' Compensation. Note that all dollar amounts are in thousands.

This paper describes how to use data provided by the NAIC to develop a stochastic model for the loss development process. A feature of this model will be that it draws on the information provided by several insurers to provide "prior information" for use in the Bayesian estimation of the model parameters. The Bayesian methodology will also quantify the uncertainty in the parameters.

[3] Article 77
[4] Article 77
[5] Article 80
[6] Article 83
[7] One can purchase an electronic copy of the Annual Statements for all American insurers for a nominal price from the NAIC. http://www.naic.org/store_financial_home.htm.

This paper will then show how to use this model to carry out the calculations required for the technical provisions of Solvency II. In watching parts of the debate that led to Solvency II, I have seen reasonable alternatives to its methodology. This paper will explore some of those alternatives.

The data in Schedule P is available to the public for all American insurers and thus the calculations described in this paper can be done by external interested parties. The intent of this paper is not to replace the more detailed analysis that insurers can do internally. Instead its intent is to do a credible analysis with publicly available data.

2. A STOCHASTIC MODEL OF THE LOSS DEVELOPMENT PROCESS

The stochastic model in this paper describes the random incremental paid loss, $X_{AY,Lag}$, for accident year AY, and settlement lag, Lag. The data used to fit the model will consist of a loss triangle of ten accident years of incremental paid net losses and the net earned premium for each accident year. The model can be used to predict the distribution of losses paid in future settlement lags through the tenth year. It can also be used to predict the distribution of sums of losses for any given combination of future settlement lags in the given accident years.

For a given accident year, AY, and settlement lag, Lag, the expected loss is equal to

$$\mu_{AY,Lag} = Premium_{AY} \cdot ELR_{AY} \cdot Dev_{Lag} \cdot t^{AY+Lag-1}, \tag{1}$$

where:

- $Premium_{AY}$ is the accident year premium obtained from the data,
- ELR_{AY} is a parameter representing the expected loss ratio for the accident year,
- Dev_{Lag} is a parameter representing the incremental paid loss development factor for the settlement lag,
- t is a parameter representing the calendar year trend for the claim frequency.

The claim severity, Z, in this model is a random variable with a gamma distribution,

$$f(z) = \frac{z^{\alpha-1} \cdot e^{-z/\theta}}{\Gamma(\alpha) \cdot \theta^{\alpha}}. \tag{2}$$

The claim severity distribution will vary by settlement lag with its mean given by the parameter $\tau_{Lag} = \alpha \cdot \theta_{Lag}$ and a fixed shape parameter, $\alpha = 1/2$. In accordance with the general observation that claim severity increases with the settlement lag, this model sets

$$\tau_{Lag} = sev \cdot \left(1 - \left(1 - \frac{Lag}{10}\right)^3\right) \text{ for } Lag = 1,2 \ldots,10. \tag{3}$$

As was done in Meyers (2007a), the claim count, N, in this model has a distribution with its mean given by $\lambda_{AY,Lag} = \mu_{AY,Lag}/\tau_{Lag}$, and its variance given by

$$Var[N] = \lambda_{AY,Lag} + c \cdot \lambda^2_{AY,Lag}. \quad (4)$$

The model, as described by Equations 1-4, depends upon the unknown parameters

- ELR_{AY}, for AY = 1,2,…,10.
- Dev_{Lag}, for Lag = 1,2,…,10.
- *sev* (the claim severity for the 10th settlement lag).
- *t* (the calendar year frequency trend factor).
- *c* (the contagion parameter).

My selection of the fixed parameters in the model (i.e. the τ_{Lag} parameters used to describe variation by settlement lag and the α parameter in the gamma claim severity distribution) was based on a combination of prior experience and sensitivity testing.

The expected loss in each (AY,Lag) cell is given by Equation (1). The variance of the loss in each cell is given by:

$$Var\left[X_{AY,Lag}\right] = \mu_{AY,Lag} \cdot \tau_{Lag} \cdot \left(1 + 1/\alpha\right) + c \cdot \mu^2_{AY,Lag}. \quad (5)$$

For each (AY,Lag) cell, the model will be approximated by a Tweedie distribution with the same mean and variance[8]. The mean and variance of the Tweedie distribution are given by μ and $\phi \cdot \mu^p$, respectively, with p = $(\alpha+2)/(\alpha+1)$. Using the value of p that is implied by the value of α and solving for the ϕ that forces the variances to be equal yields:

$$\phi_{AY,Lag} = \frac{\mu^{1-p}_{AY,Lag} \cdot \tau_{Lag}}{2-p} + c \cdot \mu^{2-p}_{AY,Lag}. \quad (6)$$

Note that the approximation is exact if N has a Poisson distribution with (implied) c = 0.

3. BAYESIAN ESTIMATION OF THE MODEL PARAMETERS

It is generally regarded as good statistical practice to use models with as few parameters as possible. As illustrated by Meyers (2008), too many parameters can lead to overfitting problems when estimating the parameters by maximum likelihood. Attempts such as Clark (2006) and Meyers (2009) to formulate models for loss reserving, with a small number of parameters have not found

[8] See Meyers (2009) and/or Smyth and Jørgensen (2002) for an introduction to the Tweedie distribution.

general use in the actuarial community.[9]

In the same paper, Meyers (2008) suggests, by way of example, that a Bayesian analysis can overcome the problems associated with overfitting. The paper recommends using a mixture of models over the posterior distribution of parameters. This paper takes a similar Bayesian approach.

Let *Parm* denote the set of unknown parameters $\left\{\left\{ELR_{AY}\right\},\left\{Dev_{Lag}\right\},sev,t,c\right\}$. Let $\mathbf{X}=\left\{x_{AY,Lag},AY=1,\ldots,10,Lag=1,\ldots,11-AY\right\}$ denote the observed incremental paid losses from a 10x10 Schedule P loss development triangle. According to Bayes' Theorem:

$$f\left(Parm|\,\mathbf{X}\right)\propto \ell\left(\mathbf{X}\,|\,Parm\right)\cdot f\left(Parm\right) \tag{7}$$

where:

- $f\left(Parm|\,\mathbf{X}\right)$ is the posterior distribution of *Parm*.
- $\ell\left(\mathbf{X}\,|\,Parm\right)$ is the likelihood function of **X.**
- $f\left(Parm\right)$ is the prior distribution of Parm.

The likelihood function is given by

$$\ell\left(\mathbf{X}\,|\,Parm\right)=\prod_{AY=1}^{10}\prod_{Lag=1}^{11-AY} dtweedie\left(x_{AY,Lag}\,|\,p,\mu_{AY,Lag},\phi_{AY,Lag}\right), \tag{8}$$

where:

- *dtweedie* is the probability density function for the Tweedie distribution.
- p is the power parameter. $p = (\alpha+2)/(\alpha+1) = 1.67$.
- $\mu_{AY,Lag}$ and $\phi_{AY,Lag}$ are calculated from *Parm* and Equations 1 and 6.

Following Meyers (2009) this paper uses the Metropolis-Hastings algorithm to generate a sample of 500 parameter sets that represent the posterior distribution. Appendix A describes how that algorithm was implemented in this paper. That appendix also provides the code (written in the R programming language) used for this paper.

This paper uses a gamma distribution (Equation 2) to represent its prior distributions. Table 1 gives the α and θ parameters of the prior distribution for each parameter in *Parm*.

[9] I intend no disparagement here. I consider Clark's paper to be a very good introduction to the use of maximum likelihood methods for fitting loss reserve models.

Table 1

			Implied	
Parameter	α	θ	Mean	Std. Dev.
sev	1.3676	136.2478	186.3386	159.34
t	1290.2307	0.0008	0.9931	0.0276
c	0.0740	0.1391	0.0103	0.0379
ELR_1	29.8506	0.0237	0.7073	0.1295
ELR_2	33.8347	0.0227	0.7674	0.1319
ELR_3	35.3338	0.0214	0.7545	0.1269
ELR_4	24.4908	0.0285	0.6981	0.1411
ELR_5	28.6618	0.0254	0.7272	0.1358
ELR_6	25.6341	0.0304	0.7790	0.1539
ELR_7	16.8043	0.0501	0.8417	0.2053
ELR_8	14.3680	0.0602	0.8650	0.2282
ELR_9	9.3053	0.1017	0.9465	0.3103
ELR_{10}	6.3667	0.1609	1.0246	0.4061
Dev_1	15.8100	0.0135	0.2137	0.0537
Dev_2	42.8538	0.0059	0.2517	0.0385
Dev_3	56.4944	0.0036	0.2028	0.0270
Dev_4	30.4528	0.0046	0.1403	0.0254
Dev_5	10.2309	0.0085	0.0870	0.0272
Dev_6	5.8094	0.0083	0.0480	0.0199
Dev_7	3.6954	0.0068	0.0250	0.0130
Dev_8	2.3934	0.0057	0.0135	0.0087
Dev_9	1.3559	0.0066	0.0090	0.0077
Dev_{10}	0.4552	0.0200	0.0091	0.0135

These prior distributions were obtained by the following steps.

1. Obtain the maximum likelihood estimates (MLEs) of the parameters for 50 large active insurers using Schedule P data.
2. Using the MLEs obtained in Step 1 as prior means, run the Metropolis-Hastings algorithm to get a sample of 100 parameter sets.
3. Using the 5,000 parameter sets obtained from the Steps 1 and 2 above, fit the gamma distributions by matching the mean and standard deviation of the gamma distribution with the sample mean and standard deviation for each parameter in the set.

Loss reserving is considered by many to be an art that depends on the data and actuarial judgment. The experience gained from many reserving analyses often forms the basis of such judgments. These steps taken to derive the prior distribution are an attempt to capture the experience needed for such judgments in a repeatable and transparent way. The Bayesian approach taken by this paper merges the data with the "judgment" supplied by the prior distribution.

For a given insurer, the iterations generated by the Metropolis-Hastings algorithm can be thought of as a sample of equally likely parameter sets describing the posterior distribution of their loss development process. Denote the n^{th} parameter set by:

$$Parm_n = \left\{ sev_n, t_n, c_n, \left\{ ELR_{n,AY} \right\}, \left\{ Dev_{n,Lag} \right\} \right\}. \tag{9}$$

Each $Parm_n$ can be used to construct "statistics of interest" that can be either used to describe parameter risk, or be averaged to get an overall expected value. The sections below provide several examples of statistics of interest that involve model diagnostics, prediction intervals, and items in a financial statement, such as a best estimate and a risk margin.

4. EXAMPLES WITH FOUR ILLUSTRATIVE INSURERS

This paper has illustrative analyses with data from four real insurers. The paid loss triangles were taken from the 1997 Schedule P each insurer reported to the NAIC for the commercial auto line of insurance. The data are reported in the form of cumulative paid losses for each accident year. Incremental paid losses were obtained by taking the difference of the cumulative paid losses by settlement lag. Occasionally, the cumulative paid losses decreased with subsequent settlement lag. My understanding of the reporting instructions is that this should not happen, but when it did happen, I removed the negative incremental paid loss from the data, and fit the models without that data point. The data used for fitting the model consisted of the earned net premium, the incremental paid losses indexed by accident year and settlement lag. These data are tabulated in Appendix B. Table 2 gives an indication of the size of each insurer.

Table 2

Insurer	1997 Net Premium
1	73,359
2	24,030
3	99,940
4	241,228

Before selecting the particular insurers to put in this paper, I fit the model to the data from several insurers. I selected these insurers to illustrate the variety of stories that these kinds of data can tell. I would discourage any attempts to draw conclusions about the Commercial Auto line of business or about other insurers not analyzed in this paper.

Let us start by looking at the variability of each parameter in the model. Exhibits 1-3 plot histograms of the s*ev*, *t* and *c* parameters. The top of each exhibit has a histogram of a sample of parameters taken from the prior distribution. This shows how much of the initial uncertainty in the parameters is reduced by each insurer's data. Here are some casual observations about the *sev*, *t,* and *c* parameters

The width of the histograms indicates uncertainty in the parameters. An inspection of the exhibits indicates that there is no apparent relationship between the parameter uncertainty and the size of the insurer.

Exhibit 2 confirms a general industry trend of a slight decrease in claim frequency over time for commercial auto. Given that the trend factor of 1.00 is close to the center of the histograms, one might be tempted to drop the trend parameter but, in light of the industry trend, I chose to keep it in.

As indicated in Meyers (2007a), a positive c parameter indicates that there is a random external factor that affects all claims at once. The c parameter is the coefficient of variation squared of the external factor. For insurers 2 and 3, the minimal size of the c parameter indicates that the external factor is something usual, such as changing inflation rates. The c parameter for Insurer 4 is enormous. Something is systematically affecting large blocks of claims.

Exhibit 4 shows the $\{ELR\}$ and $\{Dev\}$ parameters expressed as paths over time for both the prior and posterior distributions. One general observation is that the uncertainty in the $\{ELR\}$ parameters decreases as we gain information over time. In other words, we have better information about the loss ratio for earlier years.

It might seem natural to define the "parameter estimates" as the mean of the parameter sets $Parm_n$. But the analyses below do not make any use of such a parameter estimate. Instead they create "statistics of interest" as functions of each parameter set. They then combine them by either:

1. taking an average "statistic of interest" over all the $Parm_n$s;
2. plotting related statistics of interest; or
3. simulating predicted losses derived from a random selection of $Parm_n$s.

Exhibit 1a

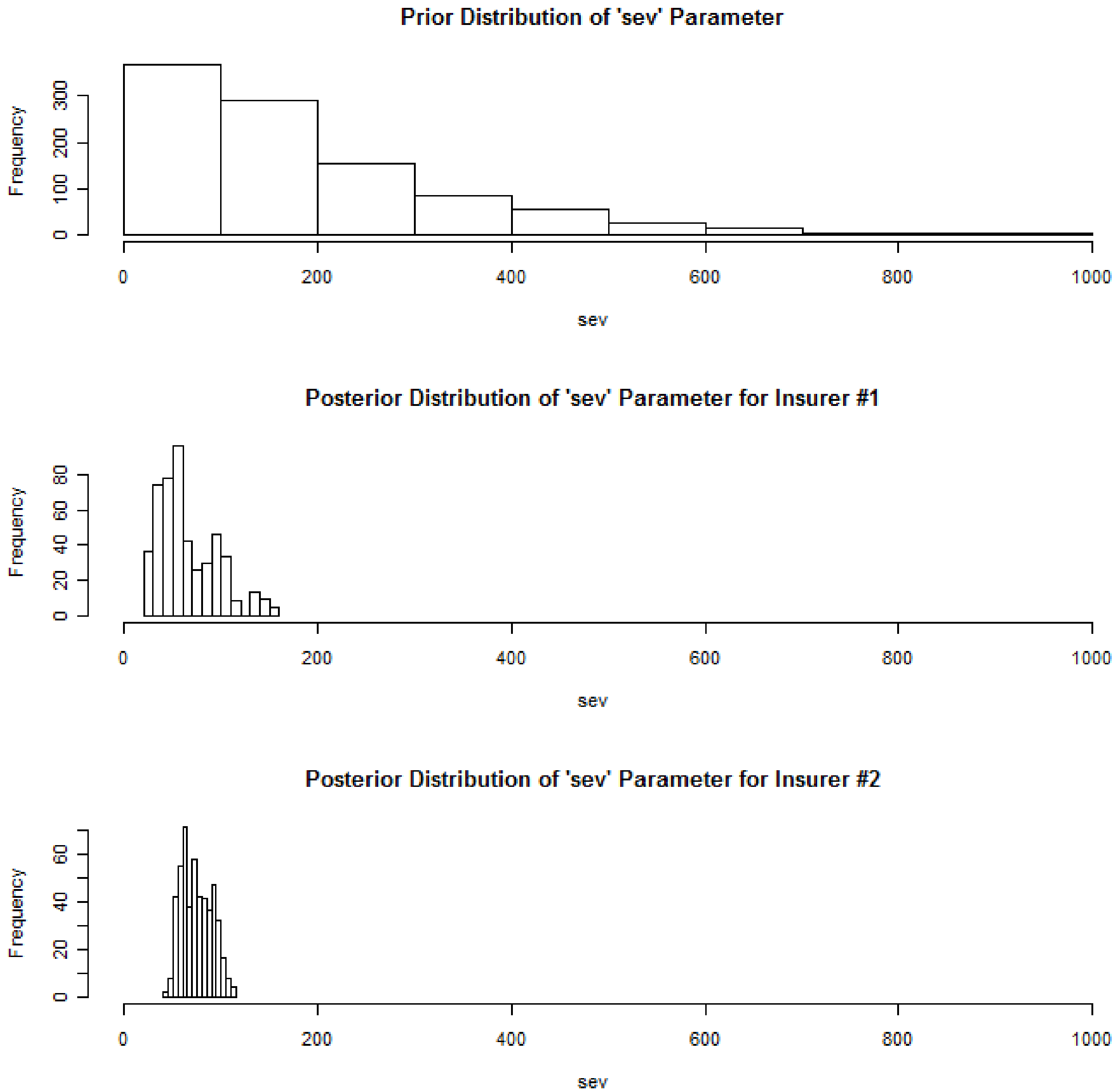

Prior Distribution of 'sev' Parameter
Frequency
0
100
200
300
0
200
400
600
800
1000
sev
Posterior Distribution of 'sev' Parameter for Insurer #1
Frequency
0
20
40
60
80
0
200
400
600
800
1000
sev
Posterior Distribution of 'sev' Parameter for Insurer #2
Frequency
0
20
40
60
0
200
400
600
800
1000
sev

Exhibit 1b

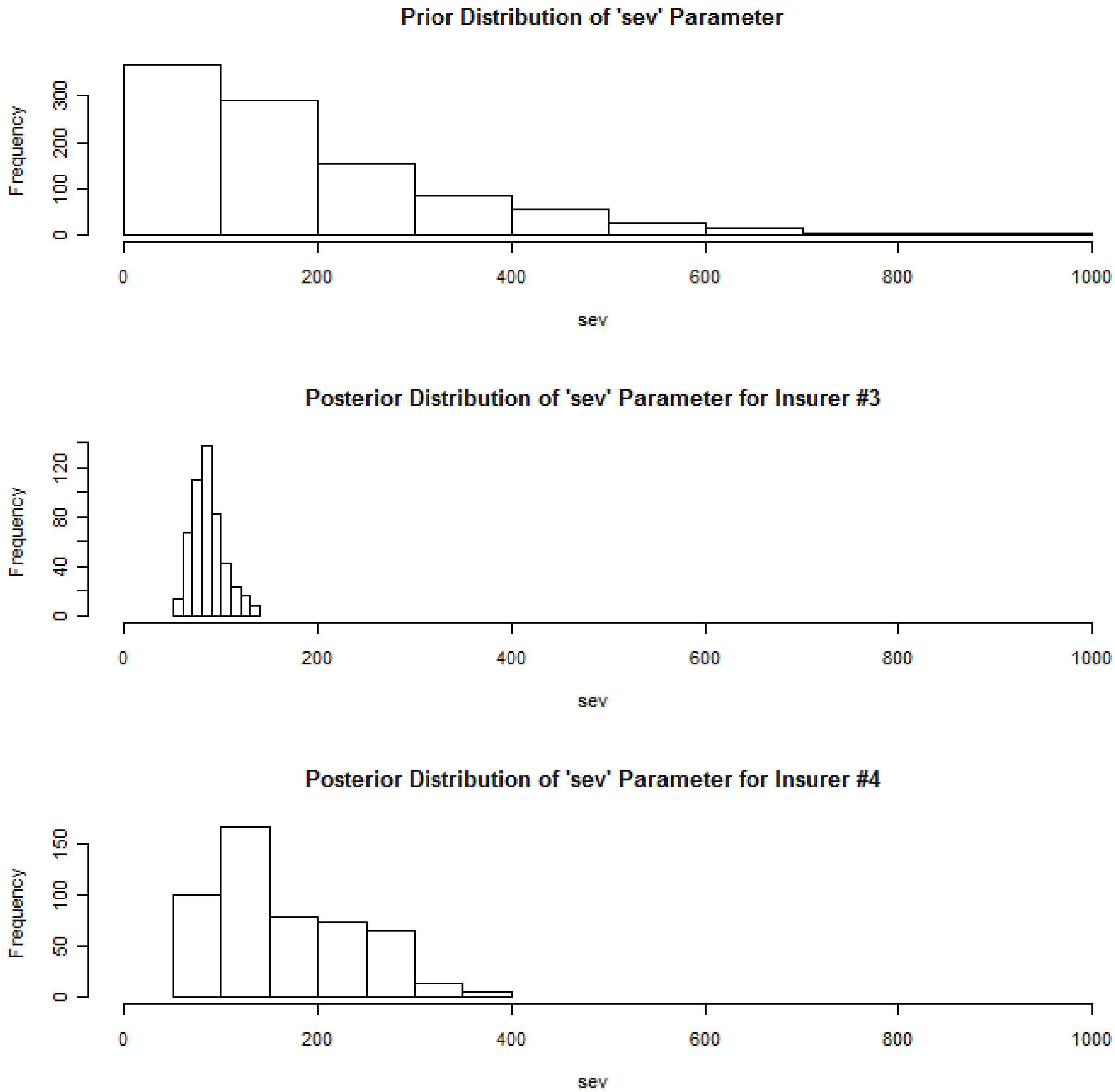

Prior Distribution of 'sev' Parameter
Frequency
0
100
200
300
0
200
400
600
800
1000
sev
Posterior Distribution of 'sev' Parameter for Insurer #3
Frequency
0
40
80
120
sev
Posterior Distribution of 'sev' Parameter for Insurer #4
Frequency
0
50
100
150
sev

Exhibit 2a

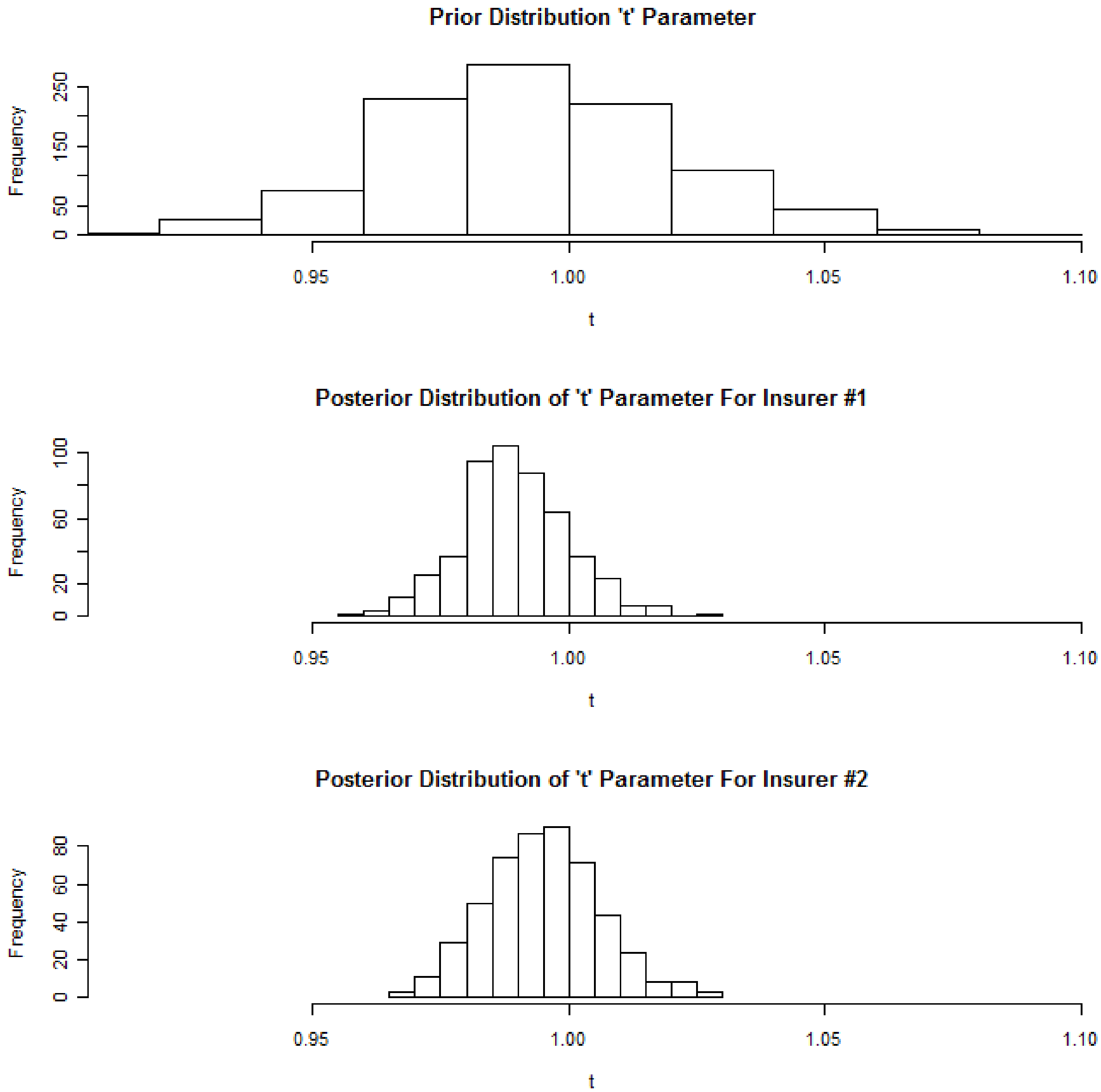
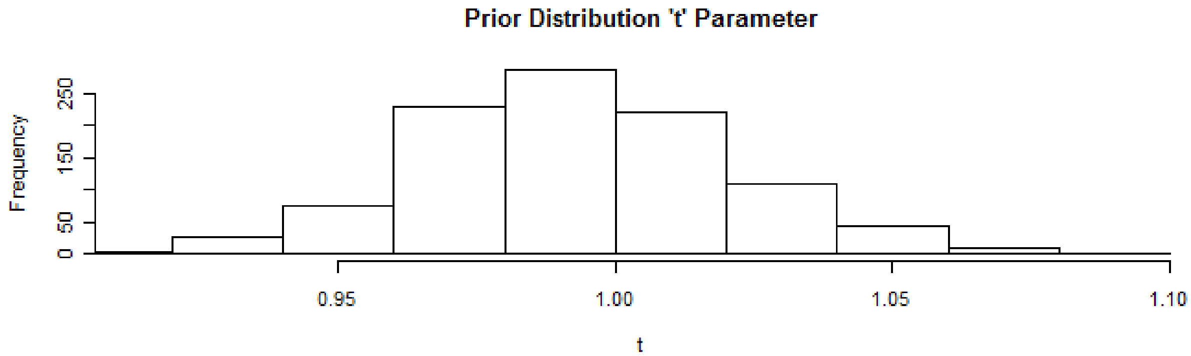
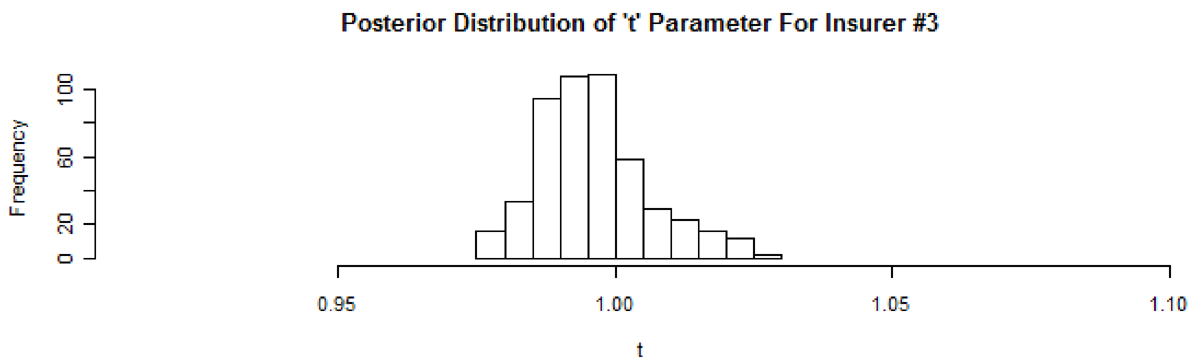

Prior Distribution 't' Parameter
Frequency
0
50
150
250
0.95
1.00
1.05
1.10
t
Posterior Distribution of 't' Parameter For Insurer #1
Frequency
0
20
60
100
0.95
1.00
1.05
1.10
t
Posterior Distribution of 't' Parameter For Insurer #2
Frequency
0
20
40
60
80
0.95
1.00
1.05
1.10
t

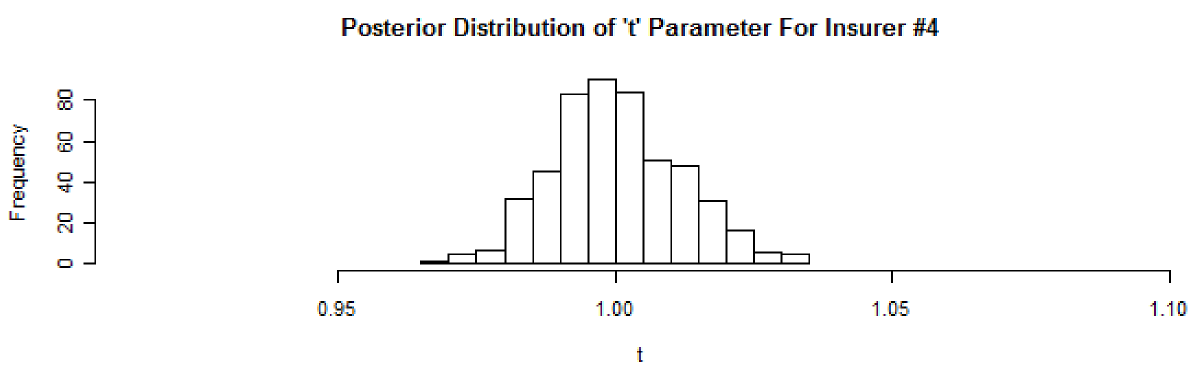
Exhibit 2b



Prior Distribution 't' Parameter
Frequency
0
50
150
250
0.95
1.00
1.05
1.10
t



Posterior Distribution of 't' Parameter For Insurer #3
Frequency
0
20
60
100
0.95
1.00
1.05
1.10
t



Posterior Distribution of 't' Parameter For Insurer #4
Frequency
0
20
40
60
80
0.95
1.00
1.05
1.10
t

Exhibit 3a

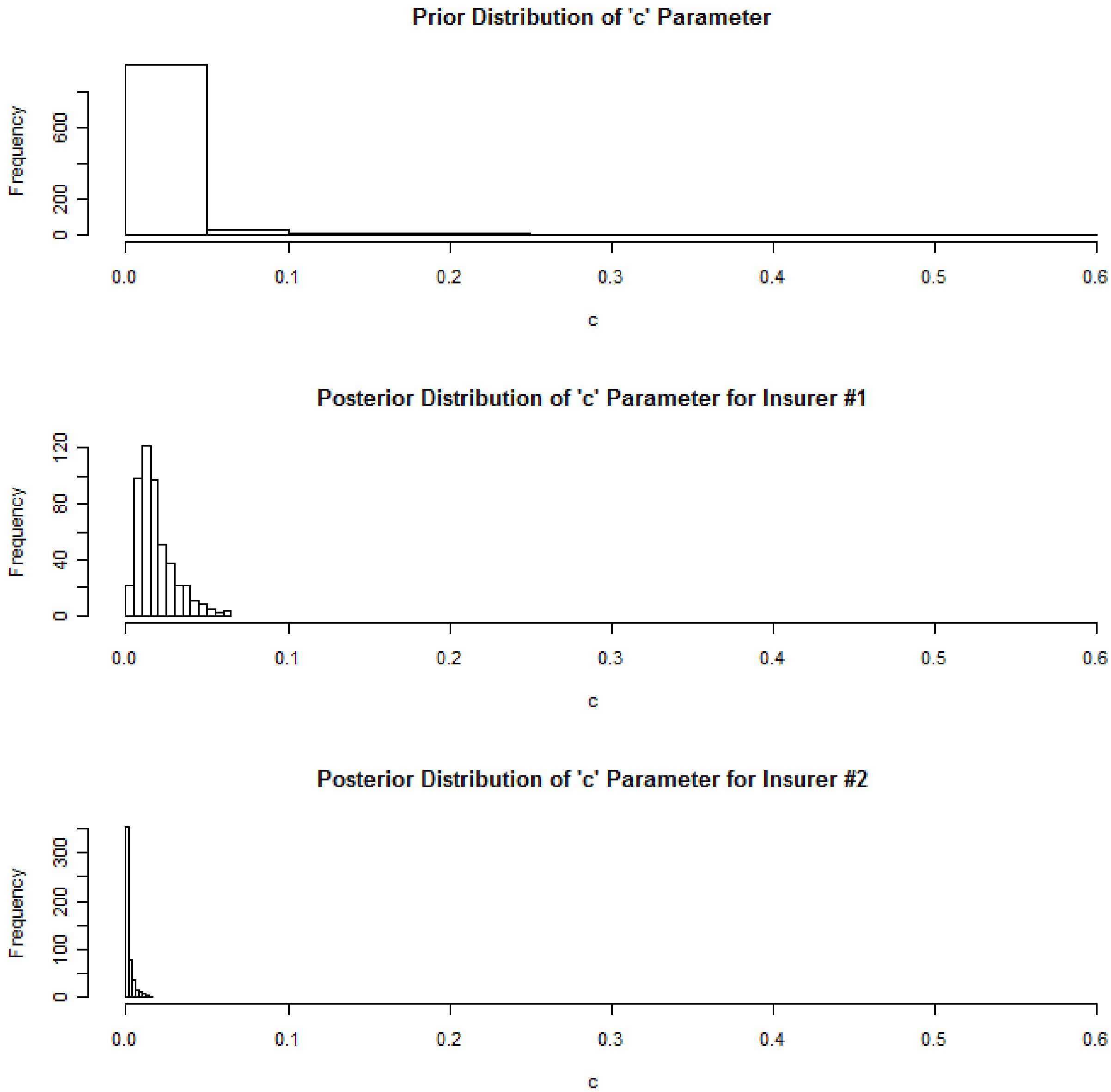
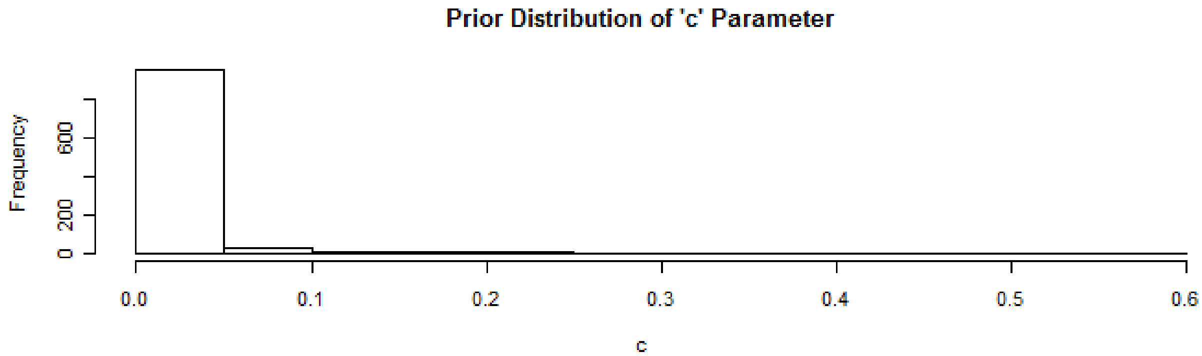
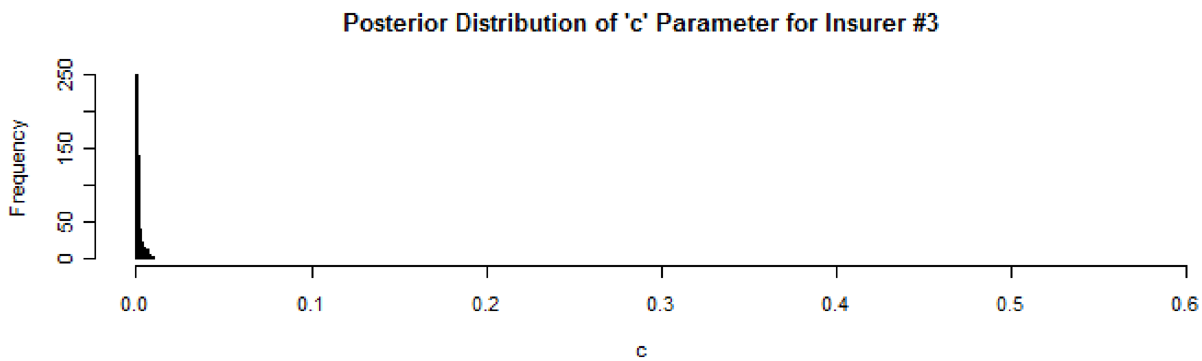

Prior Distribution of 'c' Parameter
Frequency
0
200
600
0.0
0.1
0.2
0.3
0.4
0.5
0.6
c
Posterior Distribution of 'c' Parameter for Insurer #1
Frequency
0
40
80
120
0.0
0.1
0.2
0.3
0.4
0.5
0.6
c
Posterior Distribution of 'c' Parameter for Insurer #2
Frequency
0
100
200
300
0.0
0.1
0.2
0.3
0.4
0.5
0.6
c

Exhibit 3b


Prior Distribution of 'c' Parameter
Frequency
0
200
600
0.0
0.1
0.2
0.3
0.4
0.5
0.6
c


Posterior Distribution of 'c' Parameter for Insurer #3
Frequency
0
50
150
250
0.0
0.1
0.2
0.3
0.4
0.5
0.6
c

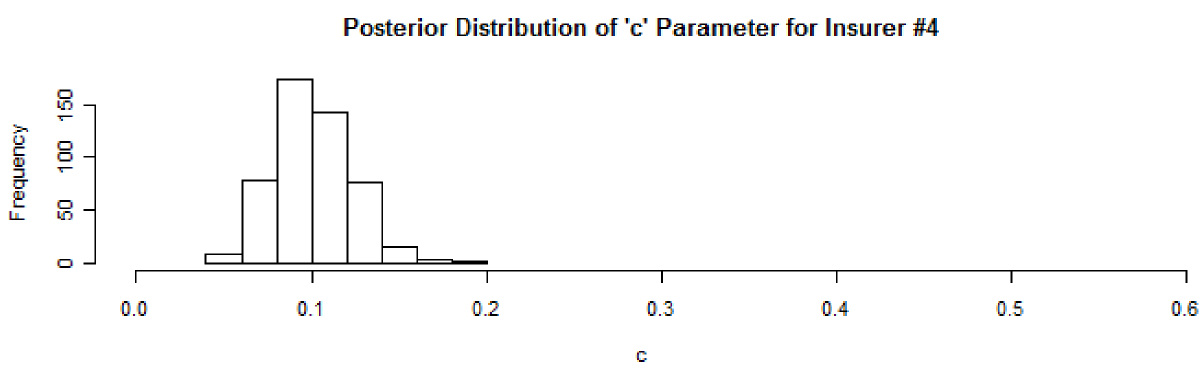
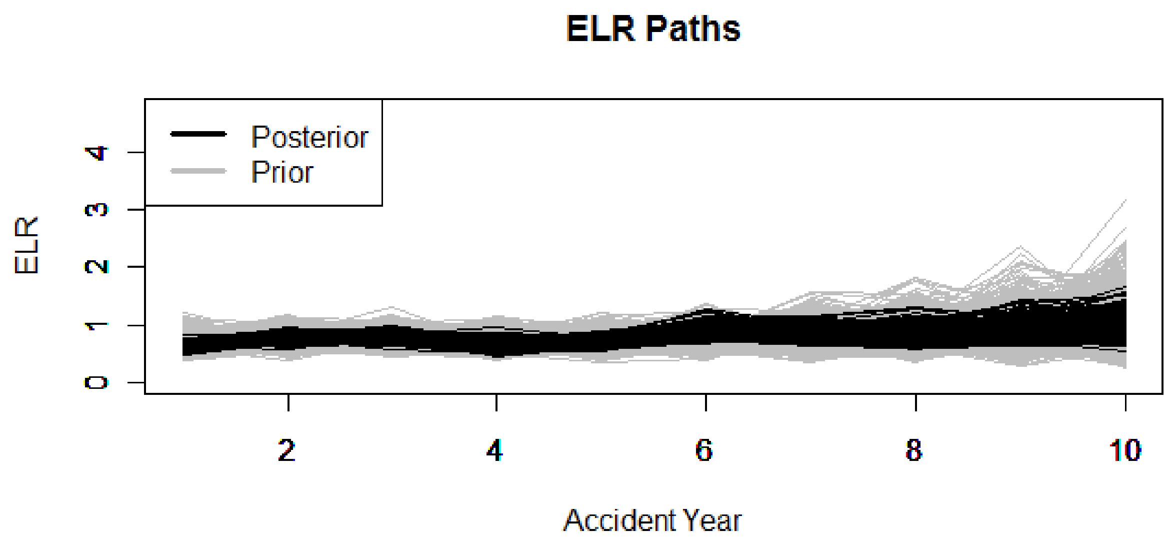

Posterior Distribution of 'c' Parameter for Insurer #4
Frequency
0
50
100
150
0.0
0.1
0.2
0.3
0.4
0.5
0.6
c

Exhibit 4a
ELR and Dev Paths for Insurer #1



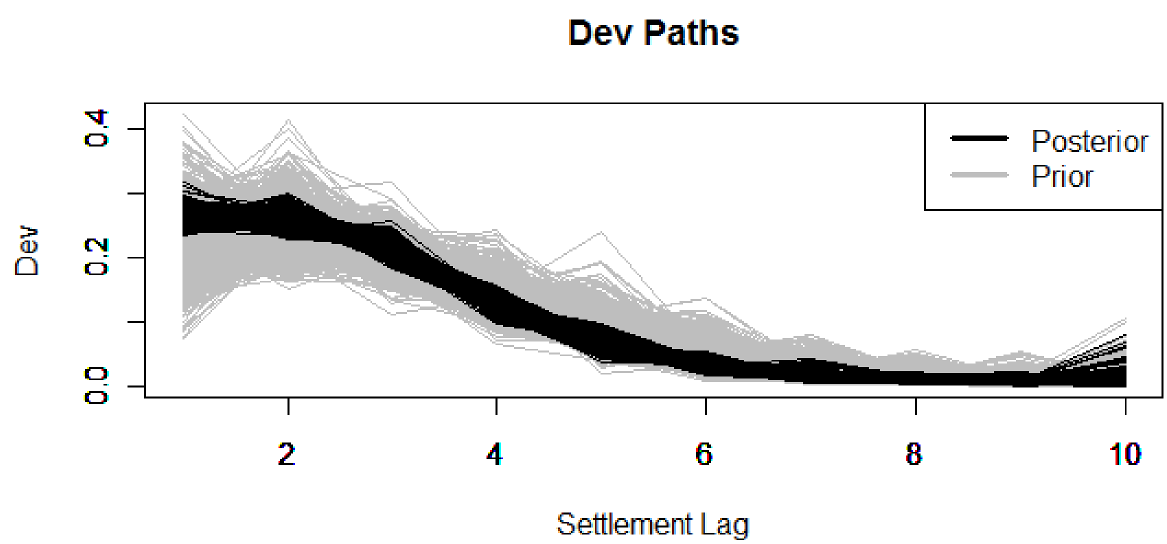
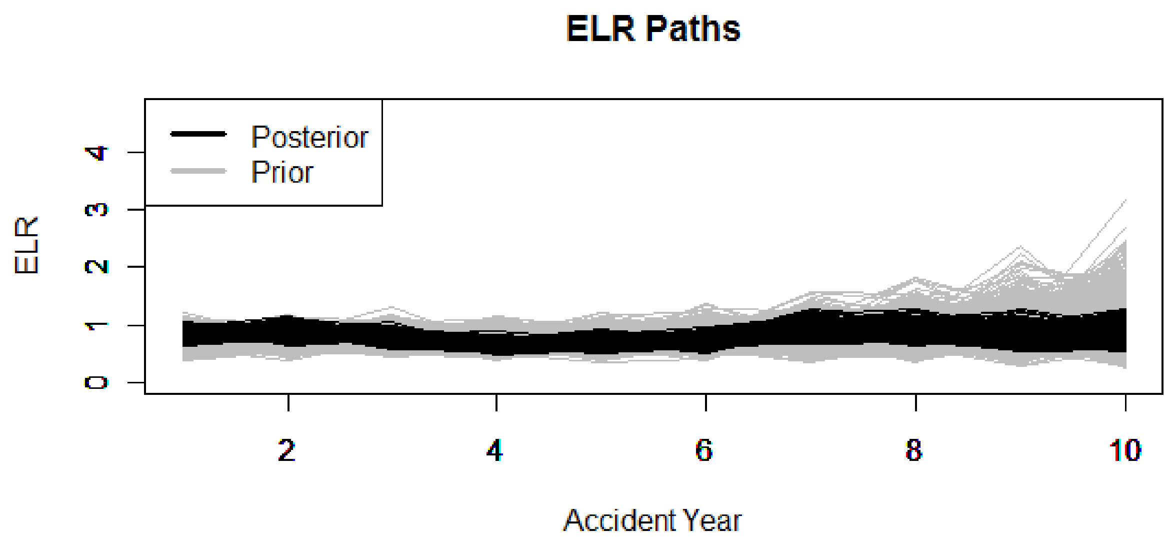


Exhibit 4b
ELR and Dev Paths for Insurer #2



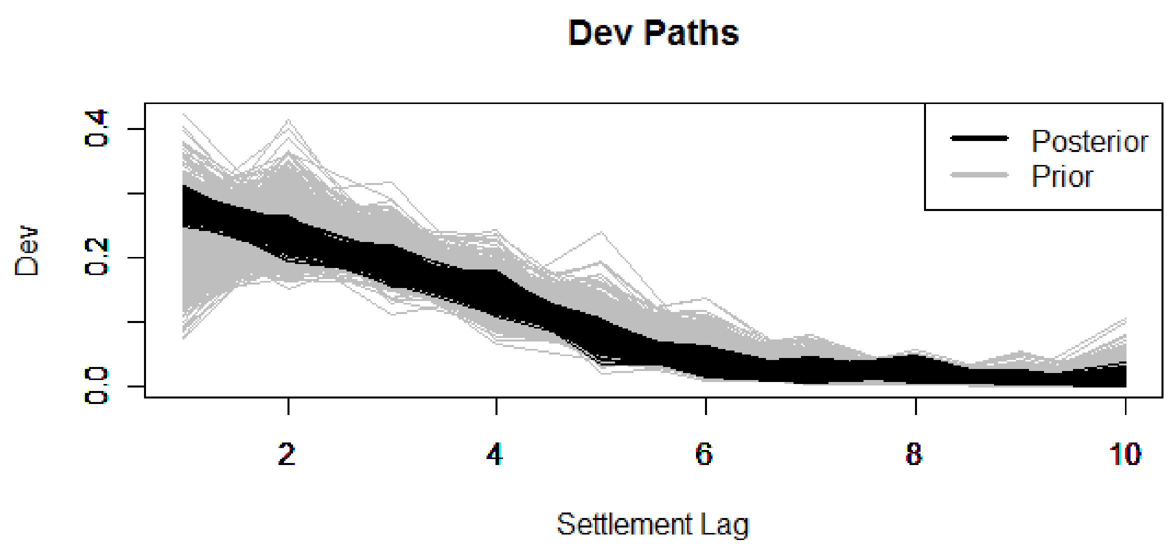


Exhibit 4c
ELR and Dev Paths for Insurer #3

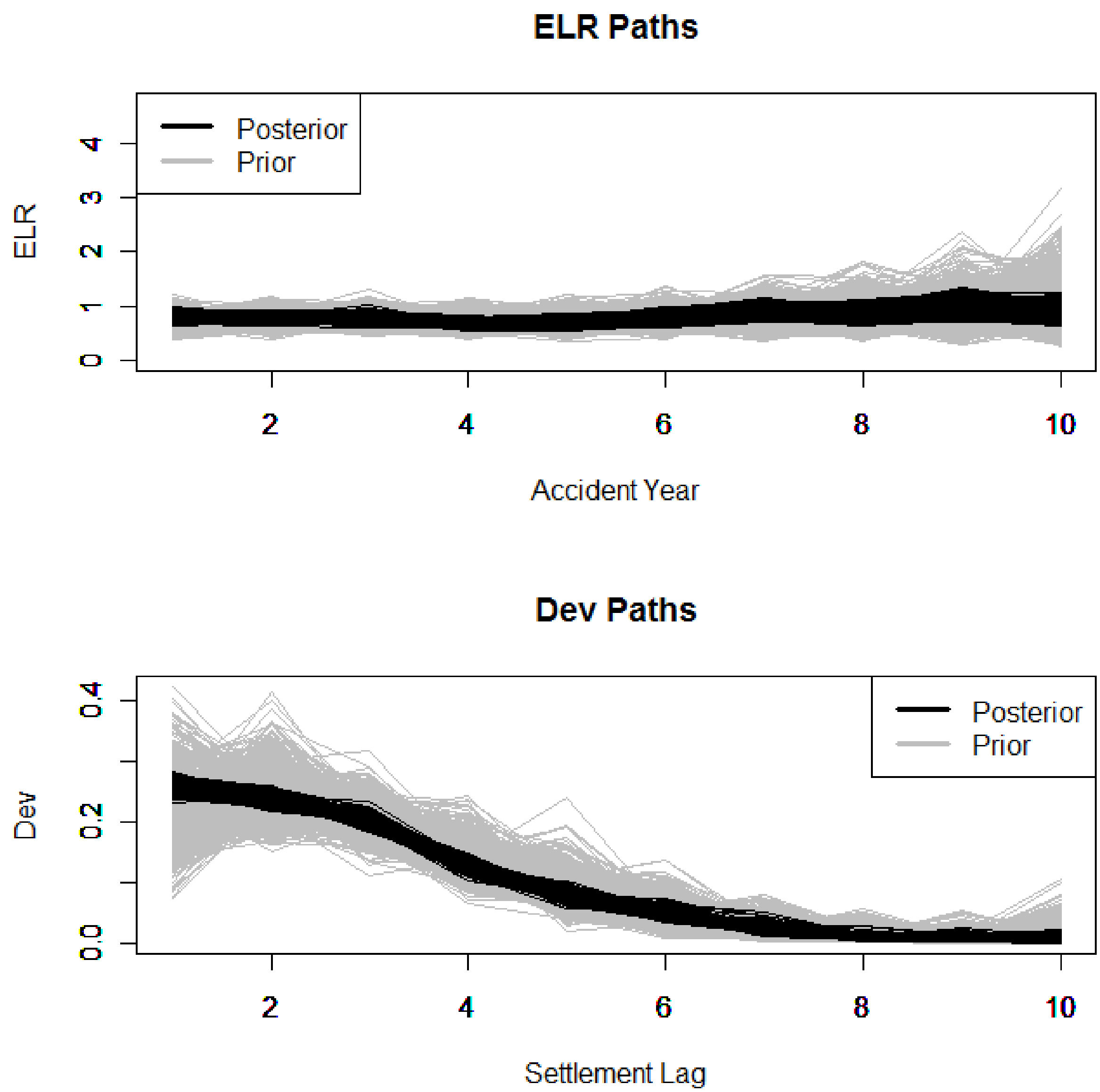
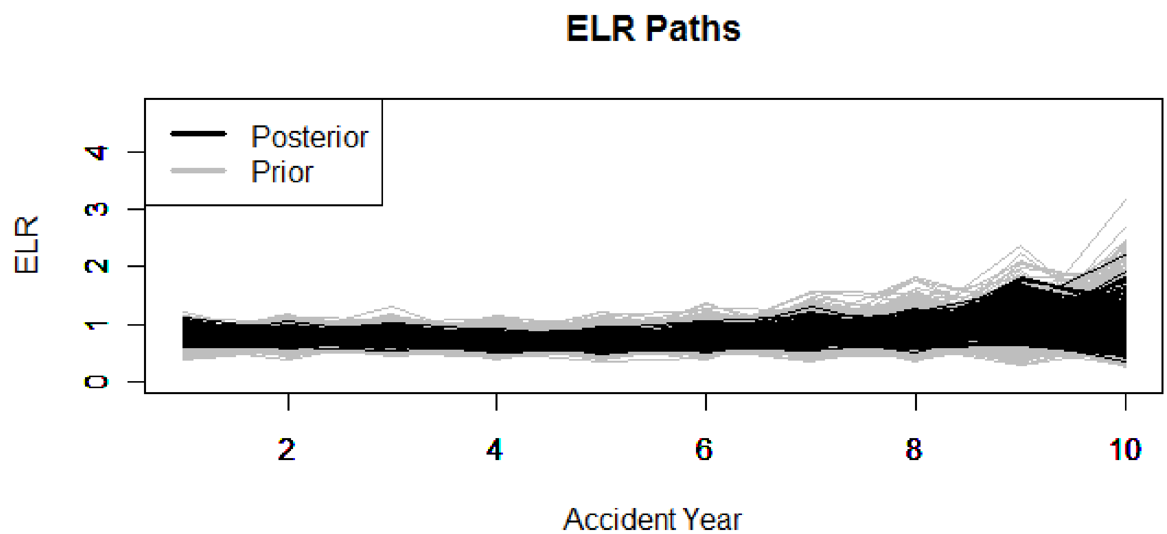
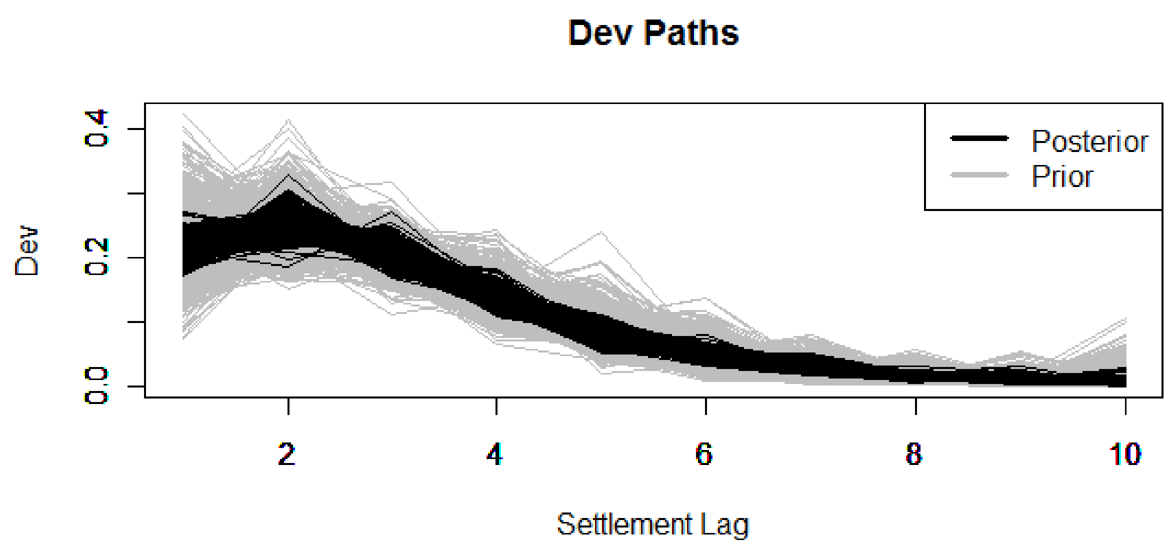


Exhibit 4d
ELR and Dev Paths for Insurer #3





5. MODEL DIAGNOSTICS

The model specified in Sections 2 and 3 predicts that the losses in each (*AY, Lag*) cell are a mixture of 500 Tweedie distributions. For a given value x in an (*AY*, *Lag*) cell, the cumulative probability is given by:

$$F_{AY,Lag}(x) = \frac{1}{500}\sum_{n=1}^{500} ptweedie\left(x \mid p, \mu_{n,AY,Lag}, \phi_{n,AY,Lag}\right), \tag{10}$$

and the mean loss for each (*AY*, *Lag*) cell is given by:

$$\mu_{AY,Lag} = \frac{1}{500}\sum_{n=1}^{500} \mu_{n,AY,Lag}, \tag{11}$$

where $\mu_{n,AY,Lag}$ and $\phi_{n,AY,Lag}$ are given by Equations 1 and 6 for each $Parm_n$, and *ptweedie* is the cumulative distribution function for the Tweedie distribution.

Denote the cumulative probabilities of each observed data point $x_{AY,Lag}$ by $p_{AY,Lag} = F_{AY,Lag}(x_{AY,Lag})$. Both the $\mu_{AY,Lag}$s and the $p_{AY,Lag}$s are given in Appendix B. Table 3 shows that the sum of the actual losses and the predicted losses are in excellent agreement.

Table 3

Insurer	Actual $\sum_{AY=1}^{10}\sum_{Lag=11-AY}^{10} x_{AY,Lag}$	Expected $\sum_{AY=1}^{10}\sum_{Lag=11-AY}^{10} \mu_{AY,Lag}$	Ratio $\frac{\text{Actual}}{\text{Expected}}$
1	269,804	269,916	0.9996
2	114,873	114,202	1.0059
3	394,629	394,854	0.9994
4	1,793,604	1,822,626	0.9841

For a well-fitting model one should expect that the collection of probabilities $\{p_{AY,Lag}\}$ will be uniformly distributed on the interval from zero to one. Following Meyers (2007b) this can be checked graphically with P-P plots. These plots compare the sorted probabilities, $\{p_{AY,Lag}\}$, with the expected probabilities. If the sorted probabilities are indeed uniform, the points in these plots will lie on a 45^{o} line.

Exhibits 5a-5d provide P-P plots for each of the four insurers. One should expect random variation from the 45^{o} line, and so the P-P plots also include confidence bands at the 99% and the 95% level based on the Kolmogorov-Smirnov test.

If the probabilities, $\{p_{AY,Lag}\}$, are truly random, one should also expect these probabilities to be independent of accident year, settlement lag, and calendar year (i.e., $AY+Lag-1$). Exhibits 5a-5d also contain plots of the probabilities against these variables. These plots are analogous to those described by Barnett and Zehnwirth (2000).

Here are some casual observations about the diagnostics.

- The P-P plots for all four insurers lie within the 99% confidence bands. The plots for Insurers 1, 2 and 3 all lie within the 95% confidence band, although the plot for Insurer 1 is just barely inside that band. The plot for Insurer 4 lies outside the 95% band.
- For Insurer 1, the set $\{p_{AY,Lag}\}$ for the first two accident years appears to be less spread out than expected.
- For Insurer 3, the small amount of overlap in the $p_{AY,Lag}$s in the later calendar years shows evidence of instability in the calendar year trend.
- For Insurer 4, the clearly nonrandom pattern in the calendar year plot leads to rather strange-looking patterns in the accident year and settlement lag plots.

In spite of the excellent agreement between the sum of the actual and the expected losses as identified in Table 3, the statistical diagnostics identify some potential problems with the model fits.

Exhibit 5a
Diagnostic Plots for Insurer #1

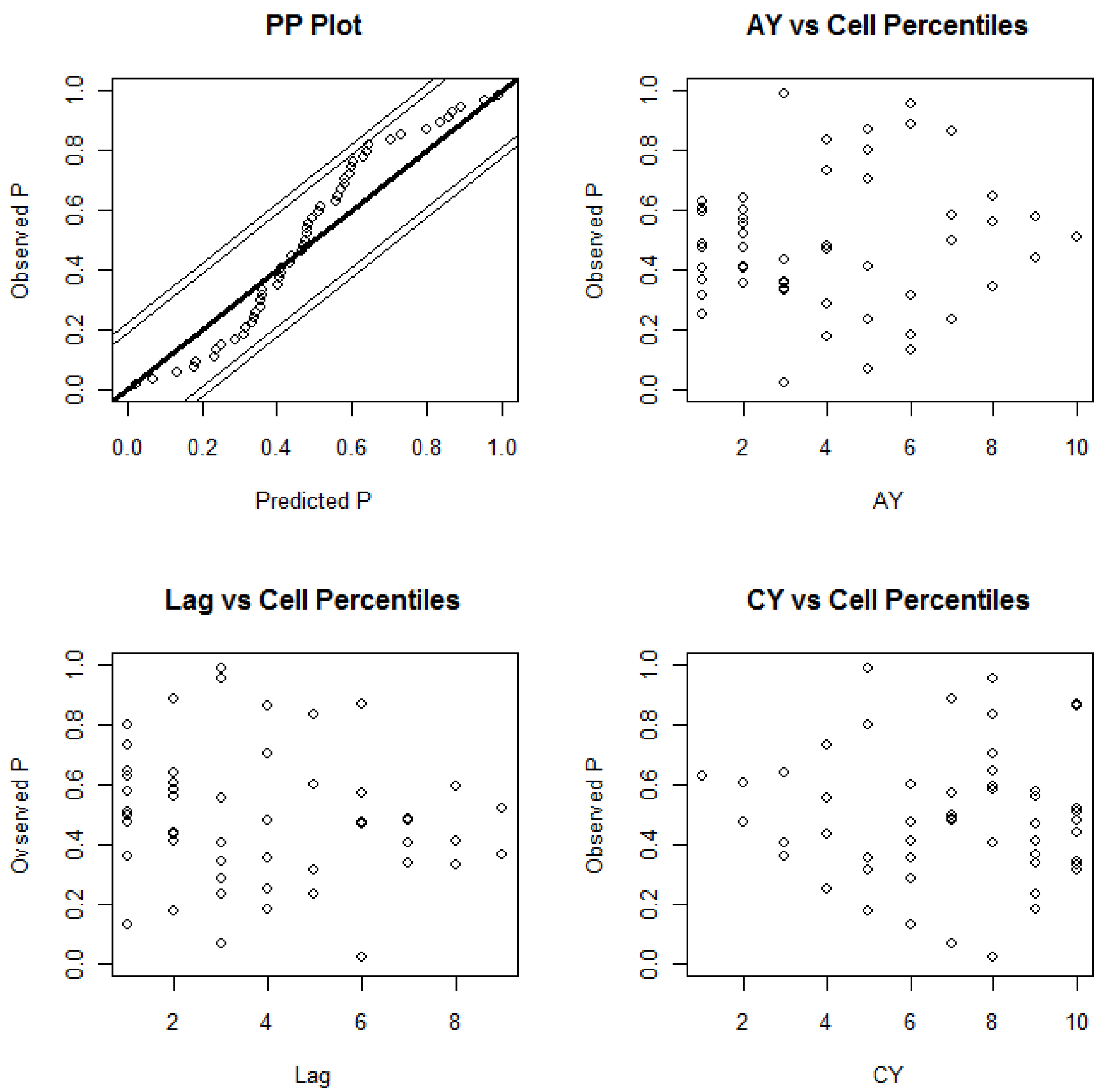


Exhibit 5b

Diagnostic Plots for Insurer #2

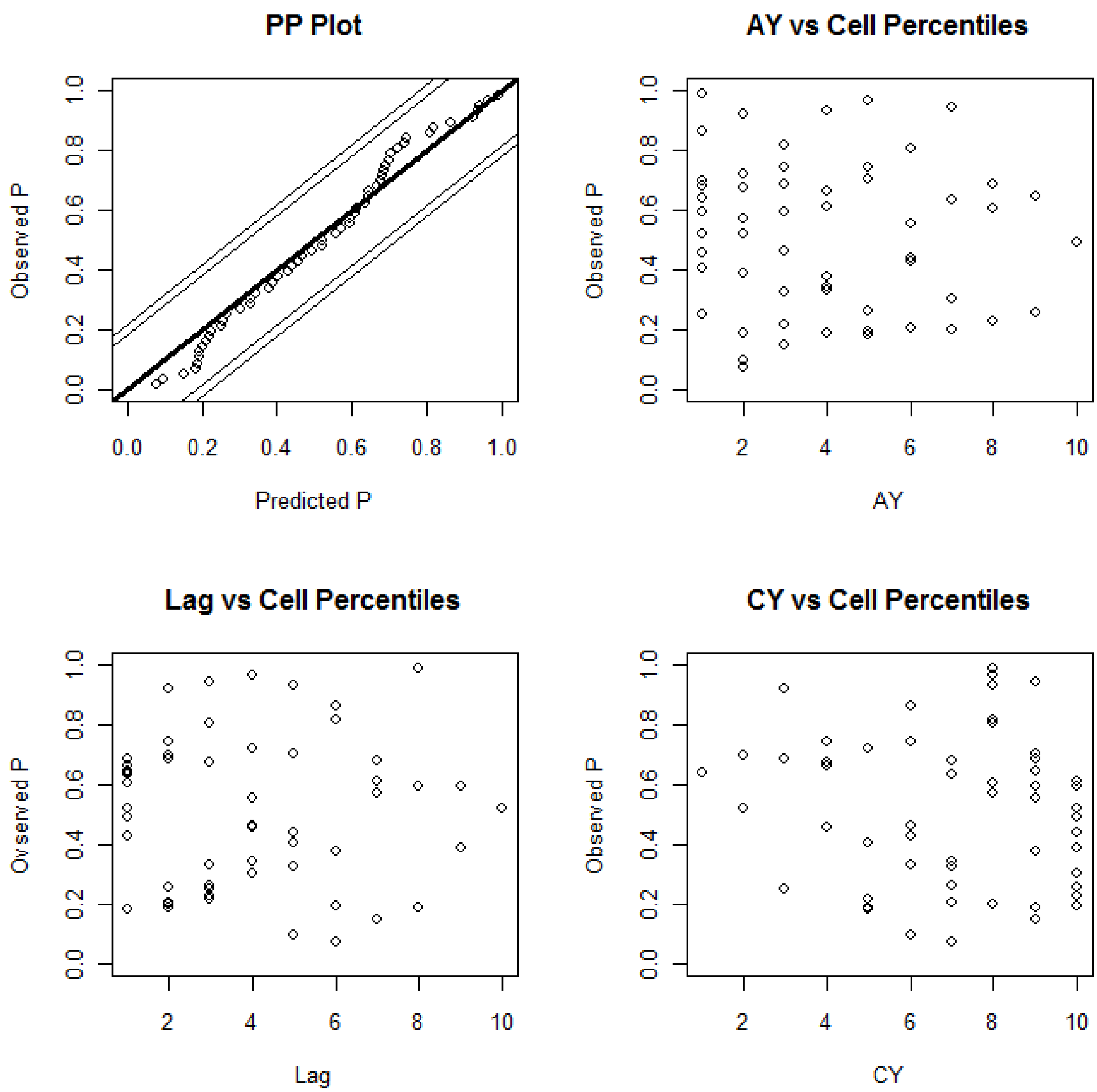


Exhibit 5c
Diagnostic Plots for Insurer #3

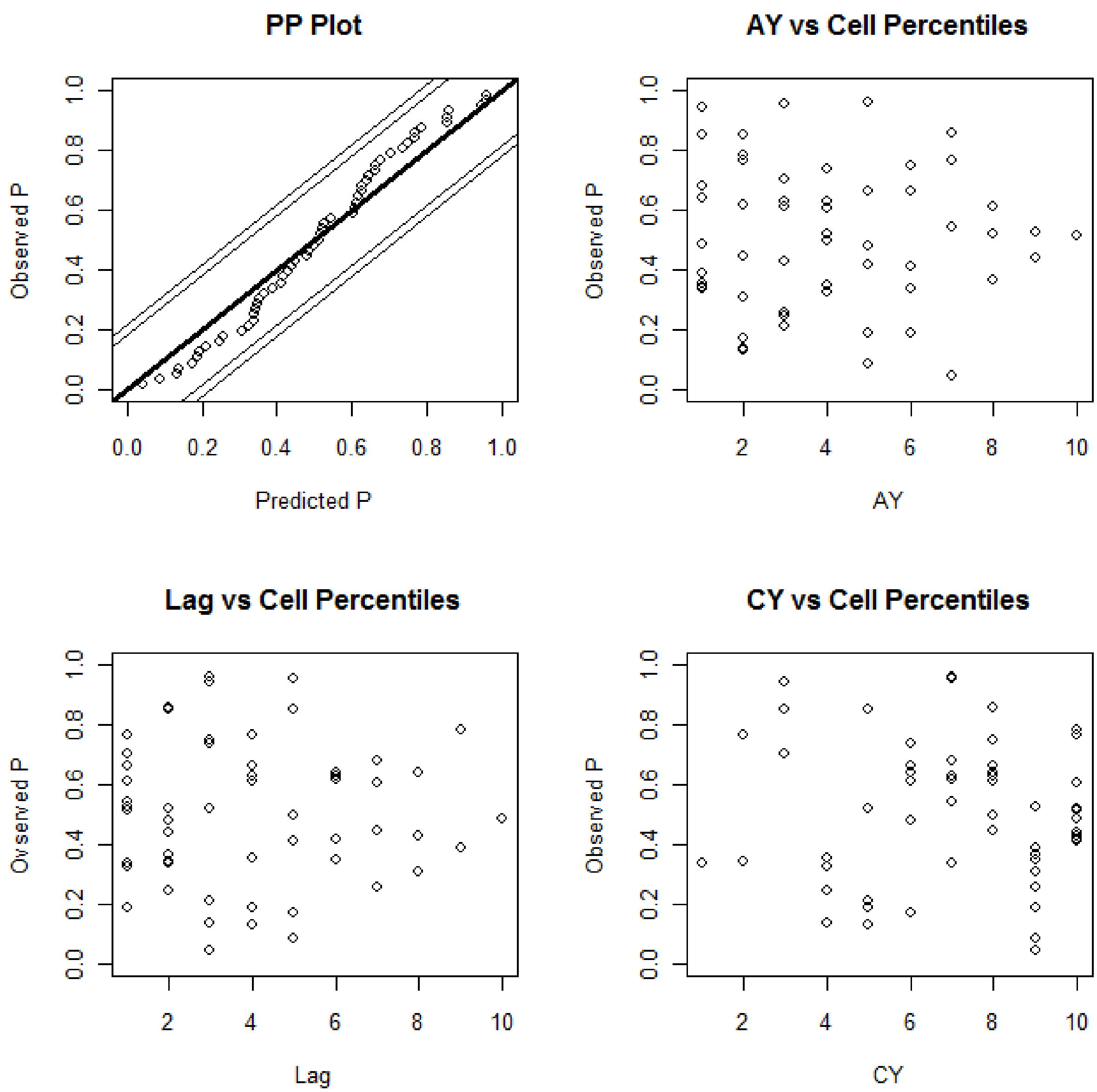
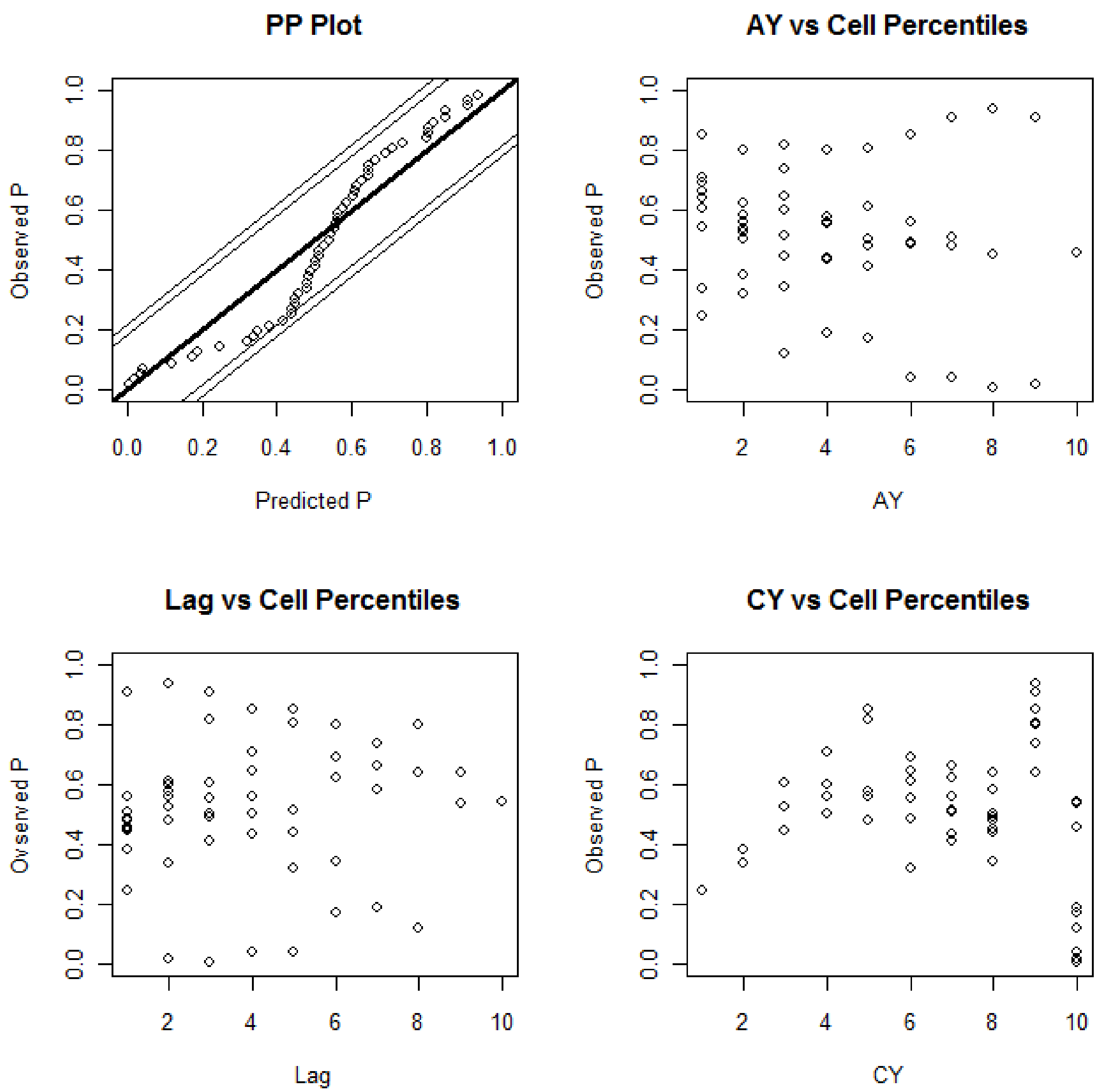


Exhibit 5d
Diagnostic Plots for Insurer #4



6. RETROSPECTIVE TESTS

As stated in the introduction, Solvency II requires insurers to have "processes and procedures in place to insure that best estimates, and the assumptions underlying the calculation of the best estimates, are regularly compared against experience." This section shows how to use the model to predict the distribution of paid loss outcomes for the next calendar year. Observing the next calendar year's total paid loss, $\sum_{AY=2}^{10} x_{AY,12-AY}$, one can check to see if the cumulative probability of that sum, as determined by its predictive distribution, lies within a normal range, say 0.05 to 0.95.

One way to determine this predictive distribution is to take a large sample, say 10,000 or so, of random *X*s from the following simulation algorithm.

Simulation Algorithm 1

1. Select a random parameter set from the list, $\mathrm{P}_n = \left\{sev_n, t_n, c_n, \left\{ELR_{n,AY}\right\}, \left\{Dev_{n,Lag}\right\}\right\}$.
2. For each (*AY,Lag*) cell in next calendar year (*AY* = 2,…,10, *Lag* = 12 – *AY*):
    a. Calculate $\mu_{AY,Lag}$ from Equation 1.
    b. Calculate $\phi_{AY,Lag}$ from Equation 6.
    c. Select a random loss $X_{AY,Lag}$ from a Tweedie distribution with parameters p = 1.67, $\mu_{AY,Lag}$, and $\phi_{AY,Lag}$.
3. Set $\mathrm{X} = \sum_{AY=2}^{10} \mathrm{X}_{AY,12-AY}$

Following Meyers (2009), this paper uses the fast Fourier transform (FFT) to calculate the predictive distributions. It is faster and more numerically precise, but admittedly harder to implement. The R code for doing this is included in Appendix A.

When comparing the predictive distributions of this paper with predictive distributions derived from formulas in other papers, e.g., Mack (1993), one should be careful to distinguish between the predictive distribution of estimates, $\left\{\sum_{AT=2}^{10} \mu_{AY,12-AY}\right\}$, and the predictive distribution of outcomes, $\left\{\sum_{AY=2}^{10} X_{AY,12-AY}\right\}$. For retrospective tests we need the latter. Exhibits 6a-6d below provide the predictive distributions for both random variables.

After fitting the model to the 1997 paid loss triangle, I then obtained test data consisting of incremental paid loss data from the 1998 Schedule P and calculated the implied *p*-value for

$\left\{ \sum_{AY=2}^{10} x_{AY,12-AY} \right\}$. That and other summary statistics are in Table 4. P-values for individual cell losses in the test data are given in Appendix B.

Table 4

Insurer	Actual $\sum_{AY=2}^{10} x_{AY,12-AY}$	Expected $\sum_{AY=2}^{10} \mu_{AY,12-AY}$	Ratio $\frac{\text{Actual}}{\text{Expected}}$	*p*-value
1	41,403	40,240	102.89%	0.6408
2	11,082	13,089	84.67%	0.1080
3	46,735	57,389	81.44%	0.0019
4	102,257	212,926	48.02%	0.0000

Here are some casual observations about the results.

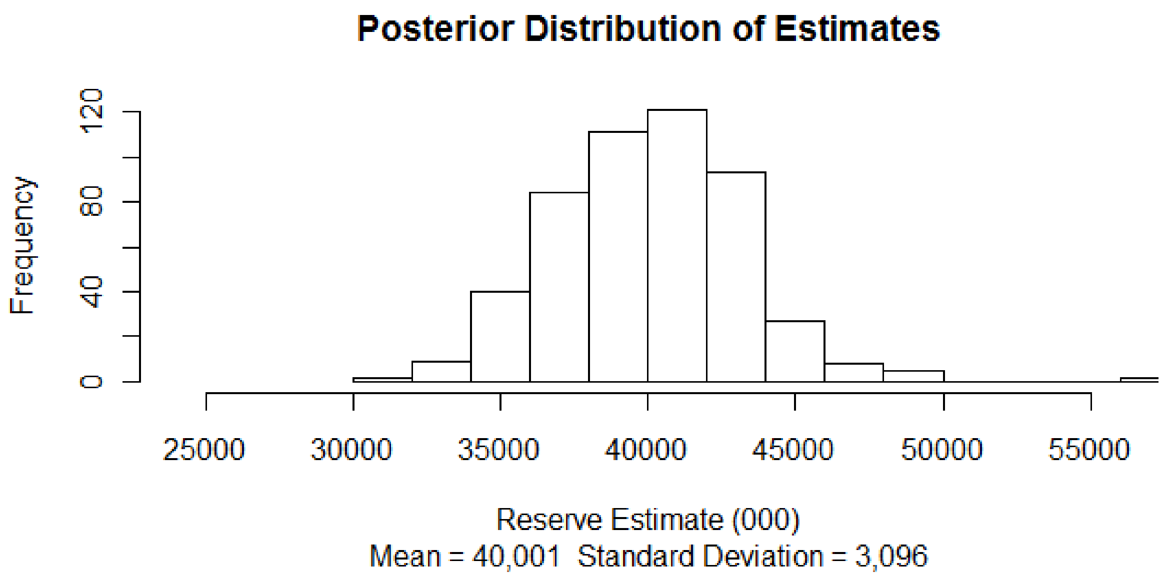
- The agreement between actual and expected results is not as good as obtained when fitting the data. Taken by itself, that is not necessarily a bad result. The test data contained only a single calendar year of data, while the data used for fitting contained 10 calendar years of data. The law of large numbers does not have a large enough number to work its magic.
- The *p*-values for Insurers 1 and 2 appear to be in the normal range. Thus, no change in assumptions seems necessary at this time.
- The *p*-value for Insurer 3 appears to be out of the normal range. If we examine the cell p-values for the test data in Appendix B, we see that all except the (*AY,Lag*) = (9,3) appear to be normal. The abnormality for the total calendar year loss appears to be caused by one bad cell. To test this, I calculated the predictive distribution for that same calendar year without the (9,3) cell. The results of this calculation are in Table 5 below. With that adjustment, the p-value moves into the normal range. An investigation into the (9,3) cell is called for. It may be a simple miscode, or some unusual event that caused the outlier.

Table 5

Insurer	Actual $\sum_{AY=2,\neq 9}^{10} x_{AY,12-AY}$	Expected $\sum_{AY=2,\neq 9}^{10} \mu_{AY,12-AY}$	Ratio $\frac{\text{Actual}}{\text{Expected}}$	p-value
3	35,861	39,063	91.80%	0.1646

- The extraordinarily low p-value for Insurer 4 cannot be explained by a single outlier. In looking at the cell p-values for the test data in Appendix B, one can see several cells with low *p*-values. This indicates there is something wrong with the structure of the model. This was apparent in the diagnostics, Exhibit 5d, of the previous section. The extraordinarily high *c* parameter and the very noticeable swings in the cell p-values by calendar provide an early indication of the problems with the model when applied to Insurer 4.

Exhibit 6a
Predictive Distributions for Insurer #1



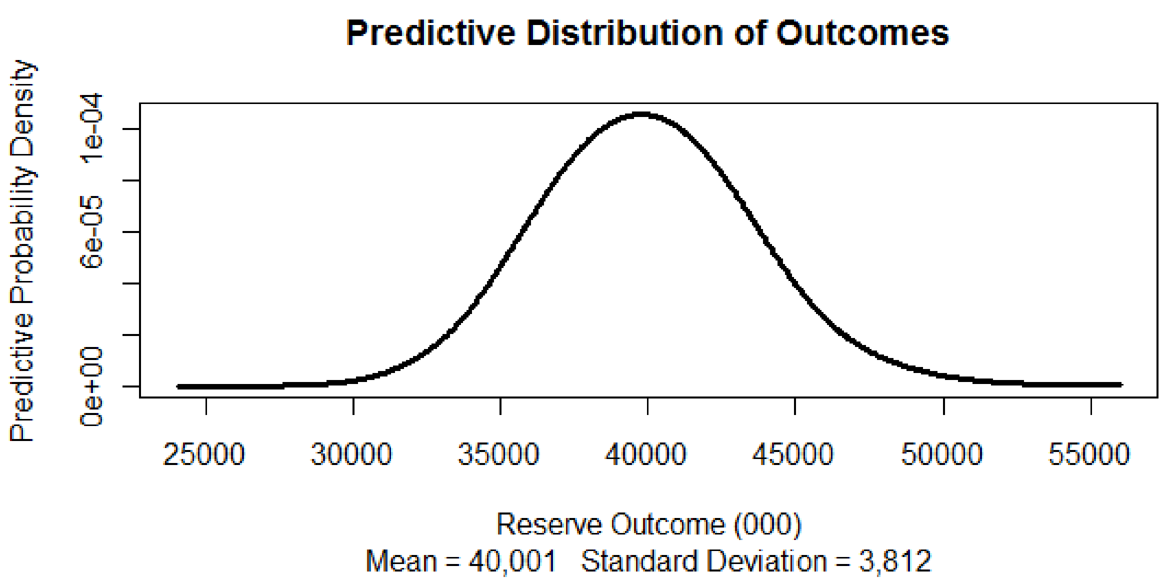
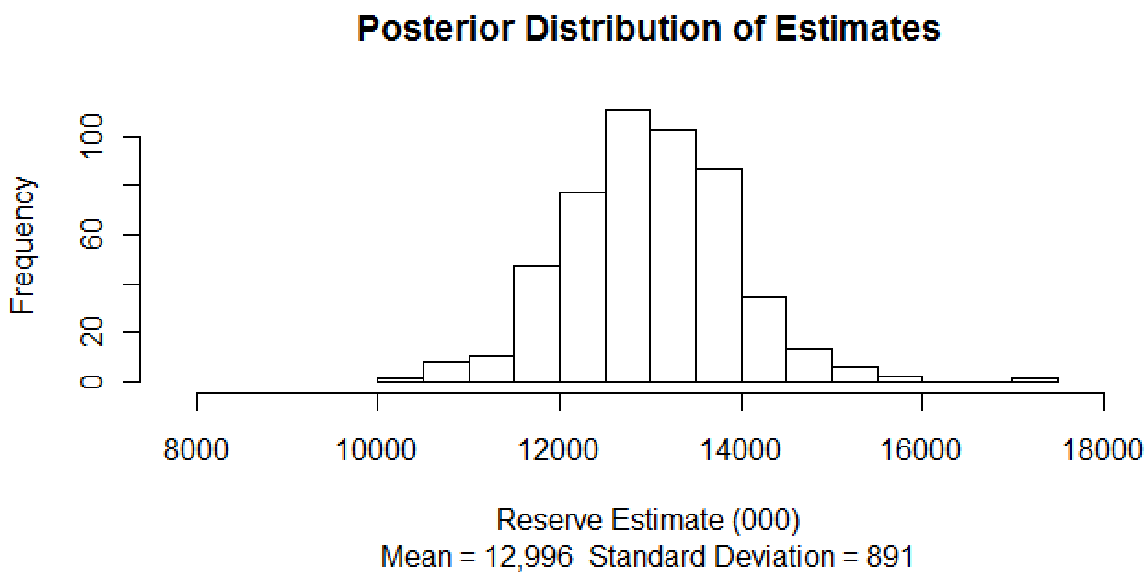


Exhibit 6b
Predictive Distributions for Insurer #2



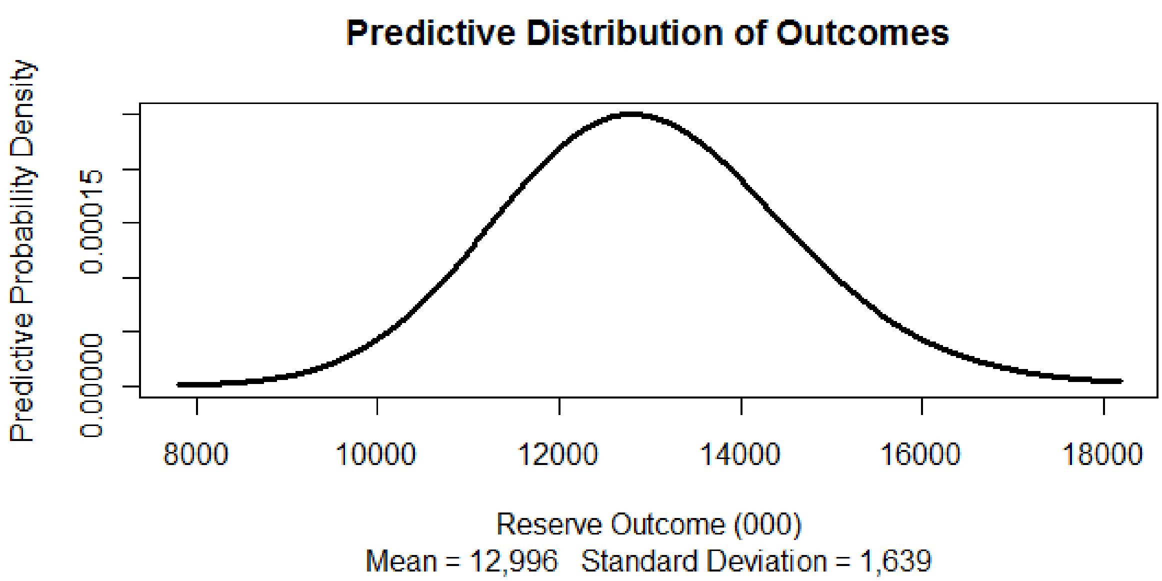
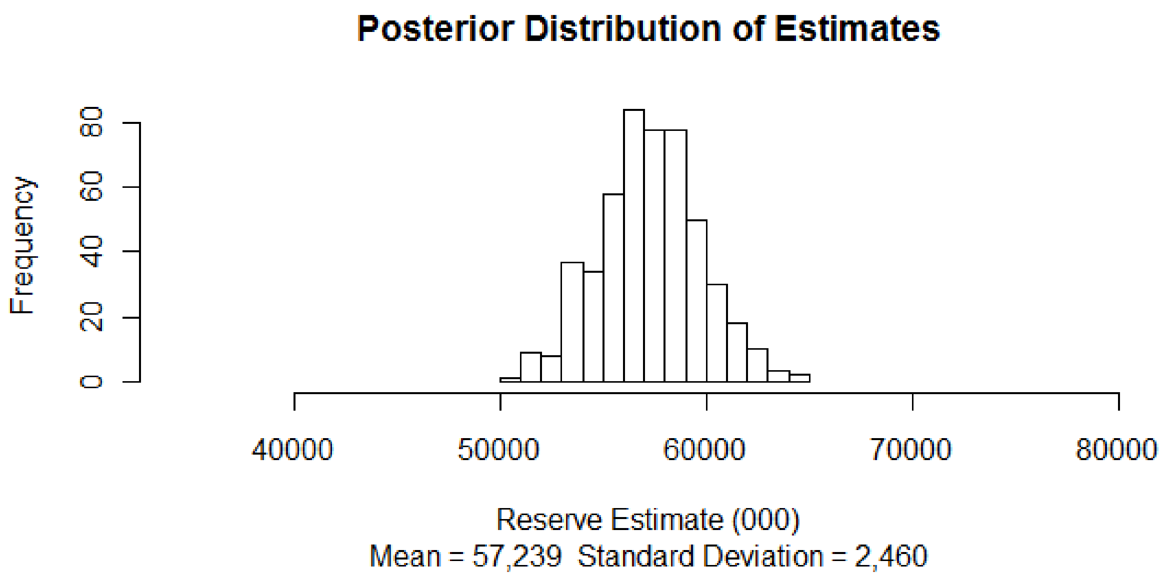


Exhibit 6c
Predictive Distributions for Insurer #3



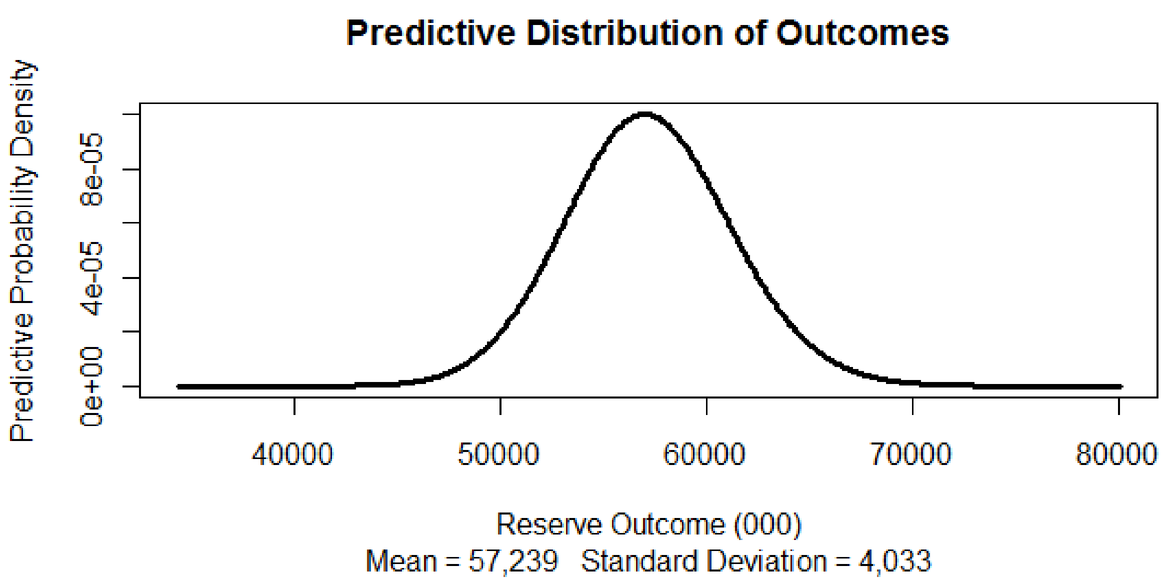
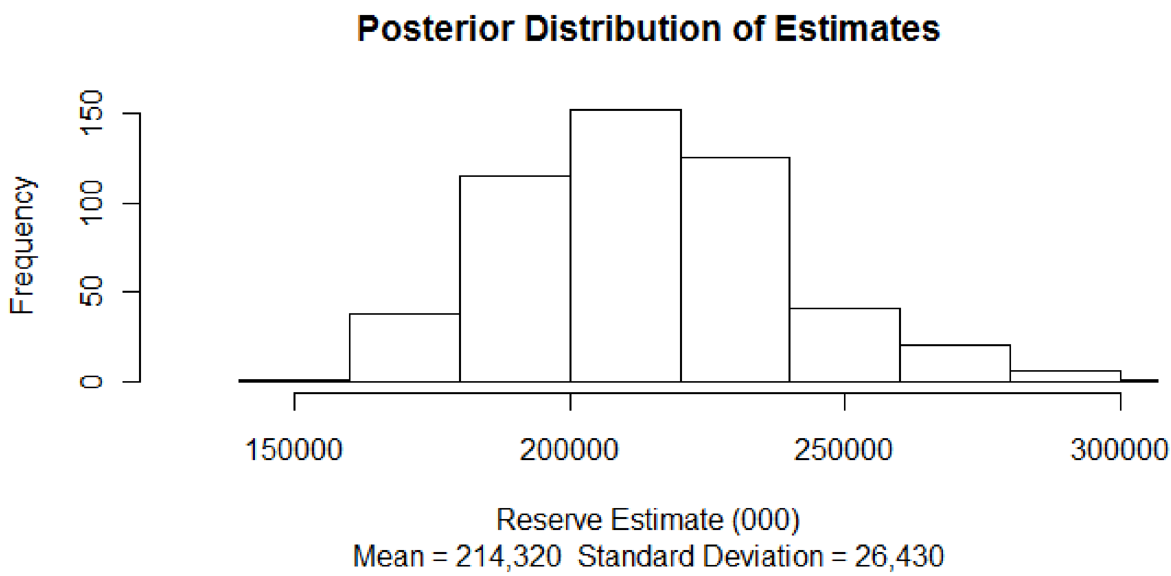
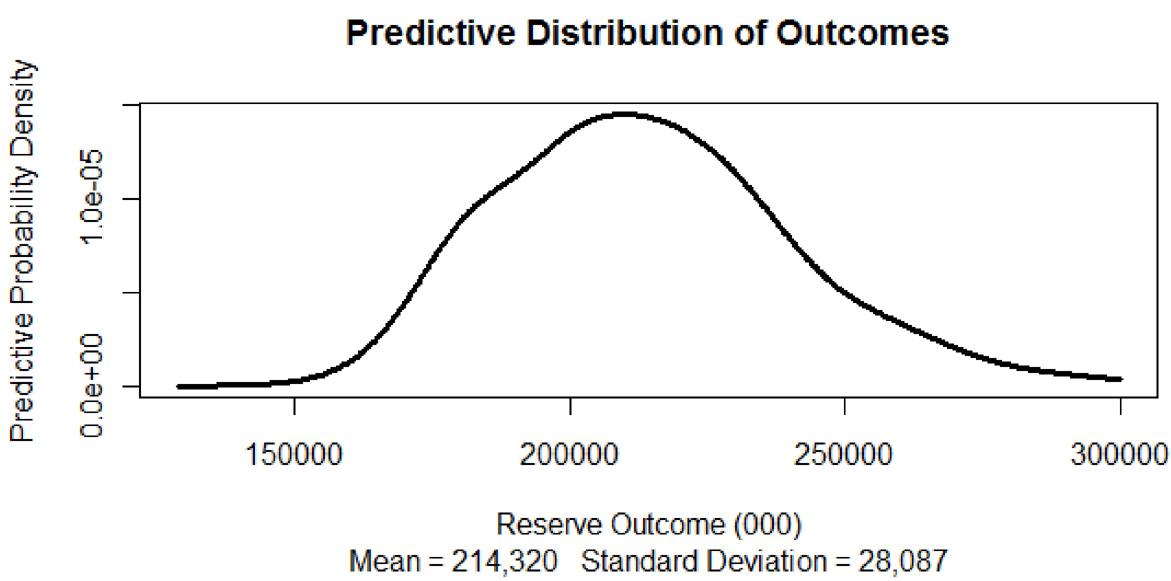


Exhibit 6d
Predictive Distributions for Insurer #4





7. BEST ESTIMATES AND RISK MARGINS

As stated in the introduction, according to the Solvency II Framework Directive:

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

"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."

"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."

This section shows how to use the model developed above to calculate the current estimate and the risk margin.

Let's start with the best estimate. Given that the future cash flows generated by the Metropolis-Hastings algorithm are equally likely, the formula for the best estimate becomes.

$$\sum_{AY=2}^{10}\sum_{Lag=12-AY}^{10}\left(\frac{1}{500}\sum_{n=1}^{500}\mu_{n,AY,Lag}\right)\cdot\frac{1}{\left(1+i\right)^{AY+Lag-11.5}}, \tag{12}$$

where i is the "relevant risk-free interest rate." This formula assumes that the liabilities expire mid-year.

Articles 104 and 105 of the Framework Directive call for the Solvency Capital Requirement to have sufficient capital to cover losses over the next 12 months with a probability (Value-at-Risk or VaR) of 99.5%. Both the time horizon of one year and the VaR standard are controversial among actuaries.

Instead of the VaR requirement, many actuaries prefer the Conditional Tail Expectation (CTE), which is the average of all outcomes above a given percentile (say 99%) of the outcomes. Another common name for the CTE is the Tail Value at Risk (TVaR). My speculation on why the EU chose the VaR requirement is that many feel uncomfortable calculating tail probabilities at the high end of the distribution of outcomes. I believe that when one calculates the distribution of outcomes as described above, the VaR and TVaR calculations are equally reliable. So the examples in this paper will use the TVaR at 99% to calculate the Solvency Capital Requirement.

A rationale for the one-year time horizon is that it will provide regulators sufficient time to take corrective action if necessary. Not everybody agrees. As we shall see below, the choice of the time horizon can make a significant difference in the risk margin. This paper will calculate the risk margin assuming both a single year and a 100-year time horizon.

The first risk margin formula discussed here is called the Capital Cash Flow (CCF) risk margin. In words, this formula assumes that the insurer's investors need to put up capital to take on the loss reserve risk. As claims are settled, the insurer expects to release capital over time. The CCF risk margin is the profit that the insurer's investors would need to be persuaded to take on this risky venture.

We will now discuss the details. Let:

- i = Risk-free rate of return on investments.
- r = Total rate of return demanded by the reinsurer for taking additional insurance risk.
- t = Time the loss reserve liability is set.
- C_t = Amount of capital required to support an insurance portfolio at time t.

First look at the cash flow of the insurance transaction.

- At time $t = 0$, investors contribute a sum of C_0 to the insurer, which earns a risk-free rate of return, i, over the next year.
- At time $t = 0$, the investors collects a (market value) risk margin, MVM_{CCF}. Equivalently, one could say that the investor contributes $C_0 - MVM_{CCF}$ to the insurer.
- At time $t = 1$, the investors expect to keep C_1 invested in the insurer, and they expect to receive a cash flow $C_0(1+i) - C_1$ at the end of year 1. Since the loss the insurers are required to pay and C_1 is uncertain, the investors discount the value of the amount returned at the risky rate of return $r > i$.
- Continuing on to time t, the investors expect to keep C_t invested in the insurer, and they expect a cash flow of $C_{t-1}(1+i) - C_t$ at the end of year t.

Since the cash flows are uncertain, it is appropriate to discount the cash flow at the risky rate of return, r. This leads to the following expression,

$$C_0 = MVM_{CCF} + \sum_{t=1}^{\infty} \frac{C_{t-1}\left(1+i\right) - C_t}{\left(1+r\right)^t}. \qquad (13)$$

This equation implies

$$\begin{aligned} MVM_{CCF} &= C_0 - \sum_{t=1}^{\infty} \frac{C_{t-1}\left(1+i\right) - C_t}{\left(1+r\right)^t} \\ &= \frac{C_0\left(1+r-1-i\right)}{1+r} + \frac{C_1\left(1+r-1-i\right)}{\left(1+r\right)^2} + \frac{C_2\left(1+r-1-i\right)}{\left(1+r\right)^3} + \ldots \\ &= \left(r-i\right)\sum_{t=0}^{\infty} \frac{C_t}{\left(1+r\right)^{t+1}}. \end{aligned} \tag{14}$$

There are two other risk margin formulas that involve slightly similar calculations. Let's call the next formula MVMSST because of its similarity to that used in the Swiss Solvency Test

$$MVM_{SST} = \left(r-i\right)\cdot\sum_{t=1}^{\infty} \frac{C_t}{\left(1+i\right)^{t+1}}. \tag{15}$$

MVM_{SST} differs from MVM_{CCF} in two ways. First it discounts the C_ts at the risk-free rate i, rather than the risky rate r. Second, it starts at time $t = 1$ rather than at time $t = 0$.

Let's call the last risk margin formula MVM_{QIS4} because of its resemblance to that used by some in their response to the CEIOPS Quantitative Impact Survey #4,

$$MVM_{QIS4} = \left(r-i\right)\cdot\sum_{t=0}^{\infty} \frac{C_t}{\left(1+i\right)^{t+1}}. \tag{16}$$

MVM_{QIS4} differs from MVM_{SST} in that it starts at time $t = 0$.

I used the term "resemblance" in the description of MVM_{SST} and MVM_{QIS4} because we now use a different calculation of C_t..

For a one-year time horizon, C_0 depends upon the distribution of the sum of outcomes in calendar year 11, i.e., $\sum_{AY-2}^{10} X_{AY,12-AY}$. Simulation Algorithm 1 describes the distribution of these losses.

Other calendar years and other time horizon involve random sums over different sets of (*AY*,*Lag*) cells, and Simulation Algorithm 1 can be modified to accommodate any given set of cells. As in the previous section, this paper uses the FFT methodology to calculate the predictive distribution of outcomes and the TVaR statistics.

Tables 7 and 8 below describe the calculation of the $C_r s$ for the one year and the 10 year time horizons for Insurer 1. The calculation accounts for the time value of money. Table 6 shows the result of the best estimate and risk margin calculations for Insurer 1 for two time horizons and the three risk margin formulas above.

Table 6

Insurer 1						
r = 10%	i =4%		Best Estimate = 91,220			
Time Horizon	MVM_{CCF}	%	MVM_{SST}	%	MVM_{QIS4}	%
1	1,994	2.2%	1,854	2.0%	2,411	2.6%
10	5,082	5.6%	4,736	5.2%	6,129	6.7%

Table 7

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
t	L_t^{Nom}	ΔL_t^{Nom}	L_t^{Disc}	TVaR_t^{Nom}	$\Delta\text{TVaR}_t^{Nom}$	TVaR_t^{Disc}	C_t
0	40,375	13,882	37,526	52,875	15,933	48,415	10,889
1	26,493	12,004	24,870	36,942	15,641	34,103	9,233
2	14,490	6,867	13,624	21,301	8,603	19,516	5,893
3	7,622	3,661	7,165	12,698	4,741	11,524	4,358
4	3,962	1,919	3,719	7,957	2,606	7,150	3,432
5	2,042	766	1,910	5,352	834	4,779	2,869
6	1,276	484	1,205	4,517	230	4,119	2,914
7	792	341	760	4,287	190	4,050	3,290
8	451	451	442	4,097	4,097	4,017	3,575

(1) The time, *t*, after the liability is set.

(2) The expected value of payments in the next calendar year, $L_t^{Nom} = \sum_{AY=2+t}^{10} \mu_{AY,12+t-AY}$.

(3) $\Delta L_t^{Nom} = L_t^{Nom} - L_{t+1}^{Nom}$.

(4) The discounted liability, $L_t^{Disc} = \sum_{k=t}^{8} \frac{\Delta L_k^{Nom}}{(1+i)^{k-t+0.5}}$.

(5) The Tail-Value-at-Risk, i.e., the conditional expected value of the random loss, $\sum_{AY=2+t}^{10} X_{AY,12+t-AY}$, given that the loss exceeds the 99th percentile.

(6) $\Delta\text{TVaR}_t^{Nom} = \text{TVaR}_t^{Nom} - \text{TVaR}_{t+1}^{Nom}$.

(7) The discounted TVaR $_t^{Disc} = \sum_{k=t}^{8} \frac{\Delta TVaR_k^{Nom}}{(1+i)^{k-t+0.5}}$.

The needed capital at time *t* is expected to be $C_t = \text{TVaR}_t^{Disc} - L_t^{Disc}$.

Table 8

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
t	L_t^{Nom}	ΔL_t^{Nom}	L_t^{Disc}	TVaR_t^{Nom}	$\Delta\mathrm{TVaR}_t^{Nom}$	TVaR_t^{Disc}	C_t
0	97,503	40,375	91,220	128,894	48,491	118,529	27,309
1	57,128	26,493	53,695	80,403	31,742	73,819	20,124
2	30,635	14,490	28,824	48,661	17,133	44,401	15,576
3	16,145	7,622	15,201	31,528	9,412	28,705	13,504
4	8,523	3,962	8,035	22,116	6,225	20,255	12,219
5	4,561	2,042	4,317	15,891	4,321	14,717	10,400
6	2,519	1,276	2,407	11,570	3,673	10,899	8,493
7	1,243	792	1,202	7,898	3,801	7,590	6,388
8	451	451	442	4,097	4,097	4,017	3,575

(1) The time, t, after the liability is set.

(2) The expected value of all future payments, $L_t^{Nom} = \sum_{AY=2+t}^{10} \sum_{Lag=12+t-AY}^{10} \mu_{AY,Lag}$.

(3) $\Delta L_t^{Nom} = L_t^{Nom} - L_{t+1}^{Nom}$.

(4) The discounted liability, $L_t^{Disc} = \sum_{k=t}^{8} \frac{\Delta L_k^{Nom}}{(1+i)^{k-t+0.5}}$.

(5) The Tail-Value-at-Risk, i.e., the conditional expected value of the random loss,

$\sum_{AY=2+t}^{10} \sum_{Lag=12+t-AY}^{10} X_{AY,Lag}$, given that the loss exceeds the 99th percentile.

(6) $\Delta\mathrm{TVaR}_t^{Nom} = \mathrm{TVaR}_t^{Nom} - \mathrm{TVaR}_{t+1}^{Nom}$.

(7) The discounted $\mathrm{TVaR}_t^{Disc} = \sum_{k=t}^{8} \frac{\Delta\mathrm{TVaR}_k^{Nom}}{(1+i)^{k-t+0.5}}$.

(8) The needed capital at time t is expected to be $C_t = \mathrm{TVaR}_t^{Disc} - L_t^{Disc}$.

8. NEXT STEPS

The goal of this paper was to demonstrate how publicly available data can be used to calculate the technical provisions in Solvency II. This is a purely hypothetical exercise, since the publicly available data is in America, and Solvency II applies to the European Union.

Even if the Americans were to adopt something like Solvency II, or the Europeans were to adopt reporting requirements similar to the American Schedule P, there is more work to be done. The 10 years of paid data reported in Schedule P are reasonably close to final for commercial auto. But losses in other lines of insurance can take longer than 10 years to settle. Schedule P does have incurred data that can be useful in getting estimates of outstanding losses beyond the 10-year maturity reporting limit of Schedule P. There are loss reserving methods now available that integrate both paid and incurred data. See, for example, Quarg and Mack (2008) or Posthuma, Cator, Veerkamp, and van Zwet (2008). One thing that could be done is to integrate Schedule P incurred losses into the empirical Bayesian framework developed in this paper.

REFERENCES

[1.] Barnett, Glen and Ben Zehnwirth, "Best Estimates for Reserves," *Proceedings of the Casualty Actuarial Society* LXXXVII , 2000, 245-321.

[2.] Clark, David R., "LDF Curve Fitting and Stochastic Loss Reserving: A Maximum Likelihood Approach," *Casualty Actuarial Society Forum,* Fall 2003, pp. 41-92, http://www.casact.org/pubs/forum/03fforum/03ff041.pdf.

[3.] Mack, T., "Distribution-Free Calculation of the Standard Error of Chain-Ladder Reserve Estimates," *ASTIN Bulletin* 23, 1993, pp. 213-225, http://www.casact.org/library/astin/vol23no2/213.pdf.

[4.] Meyers, Glenn G., "The Common Shock Model for Correlated Insurance Losses," *Variance* 1:1, 2007a, pp. 40-52, http://www.variancejournal.org/issues/01-01/040.pdf.

[5.] Meyers, Glenn G., "Estimating Predictive Distributions for Loss Reserve Models," *Variance* 1:2, 2007b, pp. 248-272, http://www.variancejournal.org/issues/01-02/248.pdf.

[6.] Meyers, Glenn G., "Thinking Outside the Triangle," ASTIN Colloquium, 2008. http://www.actuaries.org/ASTIN/Colloquia/Orlando/Papers/Meyers.pdf.

[7.] Meyers, Glenn G., "Stochastic Loss Reserving with the Collective Risk Model," *Variance* 3:2, 2009, pp. 239-269, http://www.variancejournal.org/issues/03-02/239.pdf..

[8.] Posthuma, B., E.A. Cator, W. Veerkamp, and E.W. van Zwet, "Combined Analysis of Paid and Incurred Losses," *CAS E-Forum*, Fall 2008, http://www.casact.org/pubs/forum/08fforum/12Posthuma.pdf.

[9.] Quarg, Gerhard, and Thomas Mack, "Munich Chain Ladder: A Reserving Method that Reduces the Gap between IBNR Projections Based on Paid Losses and IBNR Projections Based on Incurred Losses," *Variance* 2:2, 2008, pp. 266-299, http://www.variancejournal.org/issues/02-02/266.pdf.

[10.] Smyth, Gordon K. and Bent Jørgensen, "Fitting Tweedie's Compound Poisson Model to Insurance Claims Data: Dispersion Modeling," *ASTIN Bulletin*, Vol. 32, No. 1, 2002, pp. 143-157, http://www.casact.org/library/astin/vol32no1/143.pdf.

APPENDIX A – ANNOTATED R CODE

The methodology in this paper follows that of Meyers (2009). This appendix assumes that the reader is familiar with the methodology of that paper. I think the methodology needs further development before it can be considered to be mature. This paper makes a few evolutionary steps along that path.

This paper makes two improvements over the code in Meyers (2009).

First it adds the *sev*, *t*, and *c* parameters to the model. Note that Simulation Algorithm 4 or Meyers (2009) introduces the {*ELR*} and {*Dev*} parameters into the Metropolis-Hastings algorithm in two separate steps. This paper introduces the *sev* and *t* parameters into the algorithm as an additional step, and then introduces the *c* parameter as a second additional step.

Next it revises the "speedy Tweedie" approximation of Appendix B of Meyers (2009). The function "dtweedie" in R's Tweedie package is relatively slow compared to other density functions available in R. Appendix B makes use of the fact that the dtweeedie works nearly as fast on vectors as it does on single numbers. So it calculates the function dtweedie($\boldsymbol{y}$,p,$\boldsymbol{y}$,ϕ) over a vector $\boldsymbol{y}$ that spans the range needed. It then approximates the function by a single cubic polynomial. This paper attains a more accurate approximation with a piecewise cubic interpolation that is just as fast.

To run the program, you input the name of a comma separated value file containing the first four columns of the data in Appendix B. You then specify the names of the various output files (identified with various tables in the paper. Finally you have to provide a list of cells whose random sum you want to predict. It generally consists of cells that make up one or more calendar years. When testing against holdout data, you must take care to match the cells in the holdout data with the list of cells the go into the predictive distribution.

Hopefully the program comments make this clear.

```
#
# Input
#
insurer="Insurer 1 Data.csv"        # input file
adata=read.csv(insurer)
outname="Insurer 1 Summary"         # Table 3 and Table 6
#outname2="Insurer 1 Cells.csv"     # Appendix B  comment out if not testing
#outname3="Insurer 1 Test.csv"      # Table 4     comment out if not testing
tweedie.p=1.67
npost=500
#
# set up the (AY,Lag) pairs included in the predictive distribution
#
# in ayXX and lagXX below, the XX refers to the calendar year
#
ay11=2:10
lag11=12-ay11
ay12=3:10
lag12=13-ay12
ay13=4:10
lag13=14-ay13
ay14=5:10
lag14=15-ay14
ay15=6:10
lag15=16-ay15
ay16=7:10
lag16=17-ay16
ay17=8:10
lag17=18-ay17
ay18=9:10
lag18=19-ay18
ay19=10:10
lag19=20-ay19
#
# select which (AY,Lag) cells to include in predictive distribution
#
# examples
# use for the next calendar year
pred.ay=ay11
pred.lag=lag11
# use for all outstanding losses
#pred.ay=c(ay11,ay12,ay13,ay14,ay15,ay16,ay17,ay18,ay19)
#pred.lag=c(lag11,lag12,lag13,lag14,lag15,lag16,lag17,lag18,lag19)
# use for insurer 1 retro test (missing ay=3)
#ayins1=c(2,4,5,6,7,8,9,10)
#lagins1=12-ayins1
# use for insurer 3 retro test (missing ay=2)
#ayins3=3:10
#lagins3=12-ayins3
cys=unique(pred.lag+pred.ay-1)
#
# discretized gamma severity distribution
#
library(actuar)
```

```
discrete.gamma<-function(tau,p,h,fftn){
    alpha=(2-p)/(p-1)
    theta=tau/alpha
    m=2^fftn
    dpar<-rep(0,m)
    x<-h*0:(m-1)
    lev=levgamma(x,alpha,scale=theta)
    dpar[1]=1-lev[2]/h
    dpar[2:(m-1)]=(2*lev[2:(m-1)]-lev[1:(m-2)]-lev[3:(m)])/h
    dpar[m]=1-sum(dpar[1:(m-1)])
    return(dpar)
    } # end discrete.gamma function
#
# model with variable dev,elr,sev,con
#
fact.crm.llike1=function(dev,elr,sev,con){
  cyt=sev[2]^(rdata$ay+rdata$lag-1)
  eloss=rdata$premium*dev[rdata$lag]*elr[rdata$ay]*cyt
  phi=(eloss^(1-tweedie.p)*sev[1]*tau[rdata$lag])/(2-tweedie.p)+
              con*eloss^(2-tweedie.p)
  llike=ldtweedie.scaled(rdata$loss,eloss,phi)
  return(sum(llike))
  }
num=250
front=matrix(0,num,10)
log.y1=front
log.ybot=0
library(statmod)
library(tweedie)
ldtweedie.front=function(y,lyf,lf){
  ly=log(y)
  del=lyf[2]-lyf[1]
  low=pmax(floor((ly-lyf[1])/del),1)
  d01=(lf[low+1]-lf[low])/del
  d12=(lf[low+2]-lf[low+1])/del
  d23=(lf[low+3]-lf[low+2])/del
  d012=(d12-d01)/2/del
  d123=(d23-d12)/2/del
  d0123=(d123-d012)/3/del
  ld=lf[low]+(ly-lyf[low])*d01+(ly-lyf[low])*(ly-lyf[low+1])*d012+
           (ly-lyf[low])*(ly-lyf[low+1])*(ly-lyf[low+2])*d0123
  return(ld)
  }
#
ldtweedie.scaled=function(y,mu,phi){
  dev=y
  ll=y
  k=(1/phi)^(1/(2-tweedie.p))
  ky=k*y
  yp=ky>0
  dev[yp]=2*((k[yp]*y[yp])^(2-tweedie.p)/((1-tweedie.p)*
     (2-tweedie.p))-k[yp]*y[yp]*(k[yp]*mu[yp])^(1-tweedie.p)/
     (1-tweedie.p)+(k[yp]*mu[yp])^(2-tweedie.p)/(2-tweedie.p))
  ll[yp]=log(k[yp])+ldtweedie.front(ky[yp],log.y1,front)-dev[yp]/2
  ll[!yp]=-mu[!yp]^(2-tweedie.p)/phi[!yp]/(2-tweedie.p)
  return(ll)
  }
```

```
#
# log prior and proposal density functions
#
log.prior=function(dev,elr,sev,con){
  ld=dgamma(dev,alpha.dev,scale=theta.dev,log=T)
  le=dgamma(elr,alpha.elr,scale=theta.elr,log=T)
  ls=dgamma(sev,alpha.sev,scale=theta.sev,log=T)
  lc=dgamma(con,alpha.con,scale=theta.con,log=T)
  return(sum(ld,le,ls,lc))
  }

log.proposal.den=function(x,m,alpha){
  d=dgamma(x,alpha,scale=m/alpha,log=T)
  return(sum(d))
  }
#
# main program
#
# initialize variables for metropolis hastings
#
set.seed(12345)
nmh=11000                    # number of MH scenarios
#
# parameters for the prior distribution
#
alpha.sev=c(1.367644674,1290.230651)
theta.sev=c(136.2478465,0.00076972)
alpha.con=0.074005011
theta.con=0.139142639
alpha.elr=c(29.85060994,33.8347283, 35.33377535,24.49077508,28.66183085,
            25.63407528,16.80427236,14.36801632,9.305348568,6.366703316)
theta.elr=c(0.023695076,0.022680106,0.021353992,0.028504884,0.025371532,
            0.030388169,0.050089616,0.060203232,0.101715232,0.160927171)
alpha.dev=c(15.80995889,42.85381689,56.49438570,30.45284406,10.23093999,
            5.809417079,3.695390712,2.393367923,1.355938768,0.455240196)
theta.dev=c(0.013514659,0.005874493,0.003588986,0.004605868,0.008501860,
            0.008263645,0.006753167,0.005653256,0.006622295,0.020023956)
tau=1-(1-(1:10)/10)^3
alpha.prop.elr=500
alpha.prop.sev=500
alpha.prop.con=500
alpha.prop.dev=2000*alpha.dev*theta.dev
#
# get insurer data and set up tweedie model
#
rdata=subset(adata,adata$ay+adata$lag<12)  #separate  test  data  from  fitting
data
#
# set up the 'speedy tweedie' calculation
#
eloss.max=max(rdata$loss)
phi.min=eloss.max^(1-tweedie.p)*tau[1]/2/(2-tweedie.p)
k.max=(1/phi.min)^(1/(2-tweedie.p))
log.ytop=log(eloss.max*k.max)
log.ybot=0
```

```
del=(log.ytop-log.ybot)/num
log.y1=seq(from=log.ybot,to=log.ytop,length=num)
front=log(dtweedie(exp(log.y1),tweedie.p,exp(log.y1),1))
#
# initialize metropolic hastings arrays and select starting values
#
mh.dev=matrix(0,nmh,10)
mh.elr=mh.dev
mh.sev=matrix(0,nmh,2)
mh.con=mh.sev
mh.dev[1,]=alpha.dev*theta.dev  # use prior mean for mh starting values
mh.elr[1,]=alpha.elr*theta.elr
mh.sev[1,]=alpha.sev*theta.sev
mh.con[1]=alpha.con*theta.con
prev.log.post=fact.crm.llike1(mh.dev[1,],mh.elr[1,],mh.sev[1,],mh.con[1])+
              log.prior(mh.dev[1,],mh.elr[1,],mh.sev[1,],mh.con[1])
#
# generate samples using mh algorithm
#
  for (i in 2:nmh){
    devmh=rgamma(10,shape=alpha.prop.dev,scale=mh.dev[i-1,]/alpha.prop.dev)
    devmh=devmh/sum(devmh)
    u=log(runif(1))
    log.post=fact.crm.llike1(devmh,mh.elr[i-1,],mh.sev[i-1,],mh.con[i-1])+
               log.prior(devmh,mh.elr[i-1,],mh.sev[i-1,],mh.con[i-1])
    r=log.post-prev.log.post+
      log.proposal.den(mh.dev[i-1,],devmh,alpha.prop.dev)-
      log.proposal.den(devmh,mh.dev[i-1,],alpha.prop.dev)
    mh.dev[i,]=mh.dev[i-1,]
    if(u<r){
      mh.dev[i,]=devmh
      prev.log.post=log.post
      }
    #
    elrmh=rgamma(10,shape=alpha.prop.elr,scale=mh.elr[i-1,]/alpha.prop.elr)
    u=log(runif(1))
    log.post=fact.crm.llike1(mh.dev[i,],elrmh,mh.sev[i-1,],mh.con[i-1])+
               log.prior(mh.dev[i,],elrmh,mh.sev[i-1,],mh.con[i-1])
    r=log.post-prev.log.post+
      log.proposal.den(mh.elr[i-1,],elrmh,alpha.prop.elr)-
      log.proposal.den(elrmh,mh.elr[i-1,],alpha.prop.elr)
    mh.elr[i,]=mh.elr[i-1,]
    if(u<r){
      mh.elr[i,]=elrmh
      prev.log.post=log.post
      }
    #
    sevmh=rgamma(2,shape=alpha.prop.sev,scale=mh.sev[i-1,]/alpha.prop.sev)
    u=log(runif(1))
    log.post=fact.crm.llike1(mh.dev[i,],mh.elr[i,],sevmh,mh.con[i-1])+
               log.prior(mh.dev[i,],mh.elr[i,],sevmh,mh.con[i-1])
    r=log.post-prev.log.post+
      log.proposal.den(mh.sev[i-1,],sevmh,alpha.prop.sev)-
      log.proposal.den(sevmh,mh.sev[i-1,],alpha.prop.sev)
    mh.sev[i,]=mh.sev[i-1,]
    if(u<r){
      mh.sev[i,]=sevmh
```

```
      prev.log.post=log.post
      }
    conmh=rgamma(1,shape=alpha.prop.con,scale=mh.con[i-1]/alpha.prop.con)
    u=log(runif(1))
    log.post=fact.crm.llike1(mh.dev[i,],mh.elr[i,],mh.sev[i,],conmh)+
                   log.prior(mh.dev[i,],mh.elr[i,],mh.sev[i,],conmh)
    r=log.post-prev.log.post+
      log.proposal.den(mh.con[i-1],conmh,alpha.prop.con)-
      log.proposal.den(conmh,mh.con[i-1],alpha.prop.con)
    mh.con[i]=mh.con[i-1]
    if(u<r){
      mh.con[i]=conmh
      prev.log.post=log.post
      }
    }
#
# sample mh parameters
#
samp=sample(1001:nmh,size=npost)
#
# calculate predited percentiles of observed losses in training data
#
pctloss=rep(0,dim(rdata)[1])
meanloss=pctloss
tpct=rep(0,npost)
for (i in 1:dim(rdata)[1]){
  cyt=mh.sev[samp,2]^(rdata$ay[i]+rdata$lag[i]-1)
  mu=rdata$premium[i]*mh.elr[samp,rdata$ay[i]]*mh.dev[samp,rdata$lag[i]]*cyt
  meanloss[i]=mean(mu)
  phi=(mu^(1-tweedie.p)*mh.sev[samp]*tau[rdata$lag[i]])/(2-tweedie.p)+
       mh.con[samp]*mu^(2-tweedie.p)
  for (j in 1:npost){
    tpct[j]=ptweedie(rdata$loss[i],tweedie.p,mu[j],phi[j])
    }
  pctloss[i]=mean(tpct)
  if (rdata$loss[i]==0) pctloss[i]=pctloss[i]*runif(1)
  }
#
# plot results
#
windows(record=T)
#
# trace plot of estimates
#
nmh.pred=rep(0,nmh)
ay.prem=rep(0,10)
for (j in unique(pred.ay)){
  ay.prem[j]=mean(rdata$premium[rdata$ay==j])
  }
pred.mean=rep(0,nmh)
for (i in 1:nmh){
  for (j in unique(pred.ay)){
    ayp=(pred.ay==j)
    for (k in pred.lag[ayp]){
      cyt=mh.sev[i,2]^(j+k-1)
      nmh.pred[i]=nmh.pred[i]+ay.prem[j]*mh.elr[i,j]*mh.dev[i,k]*cyt
      }
```

```
    }
  }
plot(1:nmh,nmh.pred,type="l",main="Trace Plot for Mean Loss")
#
# plot of elr paths
#
set.seed(12345)
prior.elr=matrix(0,1000,10)
for (j in 1:1000){
  prior.elr[j,]=rgamma(10,shape=alpha.elr,scale=theta.elr)
  }
par(mfrow=c(2,1))
plot(1:10,prior.elr[1,],ylim=range(0,1.5*prior.elr),
       main="ELR Paths",
       xlab="Accident Year",ylab="ELR",type="n")
legend("topleft",legend=c("Posterior","Prior"),
       col=c("black","grey"),lwd=c(3,3))
for (j in 1:1000){
  par(new=T)
  plot(1:10,prior.elr[j,],ylim=range(0,1.5*prior.elr),main="",
         xlab="",ylab="",col="grey",type="l")
  }
for (j in samp){
  par(new=T)
  plot(1:10,mh.elr[j,],ylim=range(0,1.5*prior.elr),main="",
         xlab="",ylab="",col="black",type="l",lwd=1)
    }
#
# plot of dev paths
#
prior.dev=matrix(0,1000,10)
for (j in 1:1000){
  prior.dev[j,]=rgamma(10,shape=alpha.dev,scale=theta.dev)
  }

plot(1:10,prior.dev[1,],ylim=range(0,prior.dev),
       main="Dev Paths",
       xlab="Settlement Lag",ylab="Dev",type="n")
legend("topright",legend=c("Posterior","Prior"),
       col=c("black","grey"),lwd=c(3,3))
for (j in 1:1000){
  par(new=T)
  plot(1:10,prior.dev[j,],ylim=range(0,prior.dev),main="",
         xlab="",ylab="",col="grey",type="l")
  }
for (j in samp){
  par(new=T)
  plot(1:10,mh.dev[j,],ylim=range(0,prior.dev),main="",
         xlab="",ylab="",col="black",type="l",lwd=1)
    }
#
# plot of severity parameters
#
par(mfrow=c(2,1))
prior.sev1=rgamma(1000,alpha.sev[1],scale=theta.sev[1])
hist(prior.sev1,main="Prior Distribution of 'sev' Parameter",
  xlim=range(prior.sev1,mh.sev[,1]),xlab="sev")
```

```
hist(mh.sev[samp,1],xlim=range(prior.sev1,mh.sev[,1]),xlab="sev",
 main="Posterior Distribution of 'sev' Parameter")
#
# plot of calendar year trend parameters
#
par(mfrow=c(2,1))
prior.sev2=rgamma(1000,alpha.sev[2],scale=theta.sev[2])
hist(prior.sev2,main="Prior Distribution 't' Parameter",
  xlim=range(prior.sev2,mh.sev[,2]),xlab="t")
hist(mh.sev[samp,2],xlim=range(prior.sev2,mh.sev[,2]),xlab="t",
 main="Posterior Distribution of 't' Parameter")
#
# plot of contagion parameters
#
par(mfrow=c(2,1))
prior.con=rgamma(1000,alpha.con,scale=theta.con)
hist(prior.con,main="Prior Distribution of 'c' Parameter",
  xlim=range(prior.con,mh.con),xlab="c")
hist(mh.con[samp],xlim=range(prior.con,mh.con),xlab="c",
 main="Posterior Distribution of 'c' Parameter")
#
# pp plot of cell loss percentiles for training data
#
par(mfrow=c(2,2))
#
# pp plot of cell loss percentiles for training data
#
plot(sort(pctloss),1:length(pctloss)/
    (1+length(pctloss)),
    xlim=c(0,1),ylim=c(0,1),xlab="Predicted P",ylab="Observed P",
    main="PP Plot")
crit.val1=1.63/sqrt(length(pctloss)) # 1.36 for 5%, 1.63 for 1%
crit.val2=1.36/sqrt(length(pctloss))
abline(0,1,lwd=3)
abline(crit.val1,1)
abline(-crit.val1,1)
abline(crit.val2,1)
abline(-crit.val2,1)
#
#  plots of ay, lag and calendar year vs percentile for training data
#
plot(rdata$ay,pctloss,main="AY vs Cell Percentiles",ylim=c(0,1),
     xlab="AY",ylab="Observed P")
plot(rdata$lag,pctloss,main="Lag vs Cell Percentiles",ylim=c(0,1),
     xlab="Lag",ylab="Ovserved P")
plot(rdata$ay+rdata$lag-1,pctloss,main="CY vs Cell Percentiles",
     ylim=c(0,1),xlab="CY",ylab="Observed P")
#
# calculate predictive distributions of outcomes - takes some time
#

fftn=14
h=max(rdata$premium)*10/2^fftn
niceh=c(5,10,20,25,40,50,75,100,125,150,200,250,500,750,1000)
h=min(subset(niceh,niceh>h))
x=h*(0:(2^fftn-1))
phiz=matrix(0,2^fftn,9)
```

```
phix=complex(2^fftn,0,0)
postnum=0
eloss=matrix(0,length(samp),length(pred.ay))
for (k in 1:npost){
  i=samp[k]
  phixp=complex(2^fftn,1,0)
  for (j in 1:length(pred.ay)){
    ay=pred.ay[j]
    lag=pred.lag[j]
    premium=min(subset(rdata$premium,rdata$ay==ay))
    tau1=mh.sev[i,1]*tau[lag]*mh.sev[i,2]^(ay+lag-1)
    phiz=fft(discrete.gamma(tau1,tweedie.p,h,fftn))
    eloss[k,j]=premium*mh.elr[i,ay]*mh.dev[i,lag]*mh.sev[i,2]^(ay+lag-1)
    lam=eloss[k,j]/tau1
    phixp=phixp*exp(lam*(phiz-1))
    }
  phix=phix+phixp
  postnum=postnum+1
  print(postnum)
  }
pred=round(Re(fft(phix/npost,inverse=TRUE)),12)/2^fftn
mean.outcome=sum(x*pred)
sd.outcome=sqrt(sum(x*x*pred)-mean.outcome^2)
pred.range=(x>.6*mean.outcome)&(x<1.4*mean.outcome)
#
# plot distribution of estimates
#
par(mfrow=c(2,1))
pred.mean=rowSums(eloss)
hist(pred.mean,
   main="Posterior Distribution of Estimates",
   xlim=range(x[pred.range]),xlab="Reserve Estimate (000)",
   sub=paste("Mean =",format(round(mean(pred.mean)),big.mark=","),
   " Standard Deviation =",format(round(sd(pred.mean)),big.mark=",")))
#
# plot distribution of outcomes
#
xb=(x[cumsum(pred)>.99])
pb=pred[cumsum(pred)>.99]
tvar=sum(xb*pb)/sum(pb)
predb=pred[pred.range]
plot(x[pred.range],predb/h,type="l",col="black",lwd=3,
     ylim=c(0,max(predb/h)),
     xlim=range(x[pred.range]),
     main="Predictive Distribution of Outcomes",
     xlab="Reserve Outcome (000)",ylab="Predictive Probability Density",
     sub=paste("Mean =",format(round(mean.outcome),big.mark=","),
     "  Standard Deviation =",format(round(sd.outcome),big.mark=",")))
#
# write out summary statistics including tvar
#
outlab=c("input  data","train  sum  actual","train  sum  predicted","train  sum
ratio")
outlab=c(outlab,"pred mean","pred sd est","pred.sd out","pred tvar","cyid")
results=rep(0,9)
results[1]=insurer
results[2]=sum(rdata$loss)
```

```
results[3]=sum(meanloss)
results[4]=sum(rdata$loss)/sum(meanloss)
results[5]=mean.outcome
results[6]=sd(pred.mean)
results[7]=sd.outcome
results[8]=tvar
results[9]=sum(unique(pred.ay+pred.lag-1))
df.results=data.frame(outlab,results)
df.results
outname=paste(outname,results[9],".csv")
write.csv(df.results,file=outname,row.names=F)
#
# calculate predited percentiles of observed losses in test data
#
edata=subset(adata,adata$ay+adata$lag>11)  #separate  test  data  from  fitting
data
e.pctloss=rep(0,dim(edata)[1])
e.meanloss=e.pctloss
tpct=rep(0,npost)
for (i in 1:dim(edata)[1]){
  cyt=mh.sev[samp,2]^(edata$ay[i]+edata$lag[i]-1)
  mu=edata$premium[i]*mh.elr[samp,edata$ay[i]]*mh.dev[samp,edata$lag[i]]*cyt
  e.meanloss[i]=mean(mu)
  phi=(mu^(1-tweedie.p)*mh.sev[samp]*tau[edata$lag[i]])/(2-tweedie.p)+
       mh.con[samp]*mu^(2-tweedie.p)
  for (j in 1:npost){
    tpct[j]=ptweedie(edata$loss[i],tweedie.p,mu[j],phi[j])
    }
  e.pctloss[i]=mean(tpct)
  if (edata$loss[i]==0) e.pctloss[i]=e.pctloss[i]*runif(1)
  }
#
# calculate p-value for test data
#
actual=sum(edata$loss)
predicted=round(sum(e.meanloss))
ratio=round(100*actual/predicted,2)
b=(x<actual)
pvalue=round(max(cumsum(pred)[b]),4)
pvalue
testout=data.frame(actual,predicted,ratio,pvalue)
write.csv(testout,file=outname3,row.names=F)
#
# reproduce Insurer X Data
#
test=rep(0,dim(rdata)[1])
e.test=rep(1,dim(edata)[1])
ay=c(rdata$ay,edata$ay)
lag=c(rdata$lag,edata$lag)
premium=c(rdata$premium,edata$premium)
loss=c(rdata$loss,edata$loss)
pctloss=c(pctloss,e.pctloss)
meanloss=c(meanloss,e.meanloss)
test=c(test,e.test)
cell.results=data.frame(ay,lag,premium,loss,pctloss,meanloss,test)
cell.results
write.csv(cell.results,outname2,row.names=F)
```

APPENDIX B – INSURER DATA

This appendix contains the data for the four insurers analyzed in this paper, along with selected results particular to the accident year and settlement lag. The first four columns were used to fit the model. What follows is a description of each data element.

1. Accident Year (1987=1)
2. Settlement Lag
3. Net Premium
4. Incremental Paid Net Loss
5. P-value – $F_{AY,Lag}(x_{AY,Lag})$ (Equation 10)
6. Mean of the predictive distribution –$\mu_{AY,Lag}$ (Equation 11)
7. Test Indicator (= 0 if used for fitting, =1 in used for testing)

APPENDIX B – INSURER #1

AY	Lag	Premium	Loss	*p*-value	E[Loss]	Test
1	1	29,701	5,234	0.62800	4,979.25	0
1	2	29,701	5,172	0.59792	4,978.05	0
1	3	29,701	3,708	0.40137	4,025.16	0
1	4	29,701	1,783	0.25697	2,297.84	0
1	5	29,701	923	0.31750	1,224.53	0
1	6	29,701	537	0.47430	627.50	0
1	7	29,701	175	0.45819	266.75	0
1	8	29,701	145	0.58558	175.37	0
1	9	29,701	8	0.36715	110.23	0
2	1	27,526	5,234	0.46129	5,383.23	0
2	2	27,526	5,683	0.62545	5,385.46	0
2	3	27,526	4,392	0.54172	4,355.15	0
2	4	27,526	2,134	0.35274	2,485.43	0
2	5	27,526	1,377	0.58789	1,323.74	0
2	6	27,526	673	0.56343	678.62	0
2	7	27,526	155	0.38298	287.97	0
2	8	27,526	81	0.40826	189.45	0
2	9	27,526	47	0.50786	118.20	0
3	1	30,750	5,702	0.37457	6,117.66	0
3	2	30,750	5,865	0.44487	6,111.99	0
3	3	30,750	7,966	0.98836	4,945.90	0
3	4	30,750	2,472	0.36650	2,822.51	0
3	6	30,750	143	0.03003	770.53	0
3	7	30,750	152	0.32027	326.31	0
3	8	30,750	73	0.33895	215.34	0
4	1	35,814	6,349	0.73986	5,706.63	0
4	2	35,814	4,611	0.18397	5,704.71	0
4	3	35,814	3,959	0.29187	4,617.25	0
4	4	35,814	2,522	0.47976	2,635.60	0
4	5	35,814	1,924	0.82504	1,402.37	0
4	6	35,814	622	0.46919	719.95	0
4	7	35,814	206	0.45015	303.85	0
5	1	42,277	8,377	0.80655	7,291.36	0
5	2	42,277	6,890	0.41615	7,290.34	0
5	3	42,277	4,055	0.06994	5,897.94	0
5	4	42,277	3,795	0.70103	3,363.19	0
5	5	42,277	1,292	0.24508	1,791.99	0
5	6	42,277	1,422	0.85865	918.92	0
6	1	50,088	9,291	0.13432	11,322.40	0
6	2	50,088	13,836	0.88970	11,316.67	0
6	3	50,088	12,441	0.95467	9,154.43	0
6	4	50,088	4,086	0.19012	5,224.03	0
6	5	50,088	2,293	0.31767	2,780.94	0
7	1	56,921	12,029	0.51493	12,071.18	0
7	2	56,921	12,462	0.59261	12,068.66	0
7	3	56,921	8,369	0.24543	9,759.61	0
7	4	56,921	7,034	0.85965	5,567.06	0
8	1	61,406	13,119	0.65267	12,416.19	0
8	2	61,406	12,618	0.56142	12,410.83	0
8	3	61,406	9,117	0.33922	10,037.48	0
9	1	67,983	15,860	0.56056	15,630.32	0
9	2	67,983	14,893	0.41787	15,622.43	0
10	1	73,359	16,498	0.51160	16,687.70	0
2	10	27,526	0	0.36054	143.28	1
4	8	35,814	194	0.62159	201.01	1
5	7	42,277	324	0.51047	390.57	1
6	6	50,088	1,769	0.74071	1,427.91	1
7	5	56,921	4,783	0.95281	2,972.68	1
8	4	61,406	7,954	0.92811	5,735.63	1
9	3	67,983	12,655	0.52867	12,650.30	1
10	2	73,359	13,724	0.20078	16,718.34	1

APPENDIX B – INSURER #2

AY	Lag	Premium	Loss	*p*-value	E[Loss]	Test
1	1	15,274	3,718	0.63752	3,538.62	0
1	2	15,274	3,243	0.68616	2,962.24	0
1	3	15,274	1,889	0.25142	2,350.47	0
1	4	15,274	1,697	0.45221	1,826.75	0
1	5	15,274	731	0.39982	904.88	0
1	6	15,274	770	0.85561	434.59	0
1	7	15,274	287	0.67785	242.21	0
1	8	15,274	1,086	0.98831	245.97	0
1	9	15,274	49	0.58684	97.24	0
1	10	15,274	20	0.55321	89.22	0
2	1	15,722	3,844	0.53156	3,819.48	0
2	2	15,722	4,196	0.92276	3,197.64	0
2	3	15,722	2,806	0.67816	2,536.67	0
2	4	15,722	2,310	0.72096	1,973.39	0
2	5	15,722	414	0.09853	977.83	0
2	6	15,722	78	0.08000	467.88	0
2	7	15,722	232	0.56845	261.61	0
2	8	15,722	36	0.18582	265.54	0
2	9	15,722	5	0.38171	104.54	0
3	1	16,266	3,854	0.68477	3,594.26	0
3	2	16,266	3,378	0.72771	3,010.23	0
3	3	16,266	1,860	0.21581	2,387.61	0
3	4	16,266	1,736	0.45833	1,857.23	0
3	5	16,266	662	0.32273	920.45	0
3	6	16,266	697	0.81186	439.60	0
3	7	16,266	20	0.14912	245.46	0
3	8	16,266	228	0.58356	249.83	0
4	1	17,017	3,184	0.66632	2,975.79	0
4	2	17,017	1,948	0.18750	2,493.52	0
4	3	17,017	1,670	0.32708	1,979.80	0
4	4	17,017	1,257	0.33866	1,538.59	0
4	5	17,017	1,433	0.92792	761.65	0
4	6	17,017	217	0.37523	364.88	0
4	7	17,017	190	0.60559	203.25	0
5	1	18,016	2,837	0.19607	3,335.78	0
5	2	18,016	3,180	0.74147	2,793.63	0
5	3	18,016	1,794	0.26688	2,216.25	0
5	4	18,016	2,923	0.96200	1,723.16	0
5	5	18,016	1,035	0.70247	853.66	0
5	6	18,016	136	0.19534	408.70	0
6	1	18,395	3,380	0.44921	3,482.92	0
6	2	18,395	2,394	0.21652	2,916.78	0
6	3	18,395	2,859	0.80912	2,313.58	0
6	4	18,395	1,836	0.56140	1,798.69	0
6	5	18,395	763	0.44185	891.65	0
7	1	18,932	4,948	0.64629	4,715.55	0
7	2	18,932	3,288	0.20046	3,948.47	0
7	3	18,932	4,385	0.93990	3,130.72	0
7	4	18,932	2,024	0.30727	2,437.57	0
8	1	20,857	5,116	0.60750	4,936.96	0
8	2	20,857	4,466	0.67641	4,131.59	0
8	3	20,857	2,659	0.22562	3,276.81	0
9	1	24,348	5,702	0.63089	5,466.01	0
9	2	24,348	3,953	0.24678	4,577.60	0
10	1	24,030	5,450	0.48231	5,527.01	0
2	10	15,722	0	0.32000	95.88	1
3	9	16,266	124	0.73523	99.20	1
4	8	17,017	99	0.42612	207.58	1
5	7	18,016	42	0.23582	228.69	1
6	6	18,395	497	0.66174	427.07	1
7	5	18,932	1,118	0.48390	1,208.99	1
8	4	20,857	2,558	0.53458	2,553.63	1
9	3	24,348	2,531	0.09281	3,632.89	1
10	2	24,030	4,113	0.30838	4,635.13	1

APPENDIX B – INSURER #3

AY	Lag	Premium	Loss	*p*-value	E[Loss]	Test
1	1	39,383	7,701	0.32545	8,174.66	0
1	2	39,383	7,072	0.32563	7,615.81	0
1	3	39,383	8,473	0.94002	6,589.57	0
1	4	39,383	3,549	0.34291	3,969.54	0
1	5	39,383	3,327	0.85339	2,512.85	0
1	6	39,383	1,804	0.64478	1,630.75	0
1	7	39,383	817	0.69365	669.91	0
1	8	39,383	330	0.64372	298.40	0
1	9	39,383	105	0.33264	281.69	0
1	10	39,383	63	0.48387	160.76	0
2	1	44,770	9,609	0.78559	8,816.87	0
2	2	44,770	9,540	0.86835	8,213.95	0
2	3	44,770	5,755	0.12805	7,108.68	0
2	4	44,770	3,198	0.12712	4,282.46	0
2	5	44,770	1,898	0.15681	2,711.84	0
2	6	44,770	1,912	0.63151	1,756.90	0
2	7	44,770	602	0.45221	723.08	0
2	8	44,770	122	0.29190	320.84	0
2	9	44,770	462	0.76641	303.63	0
3	1	50,914	10,780	0.70914	10,187.56	0
3	2	50,914	8,570	0.24187	9,491.07	0
3	3	50,914	7,062	0.19366	8,214.06	0
3	4	50,914	5,220	0.62343	4,946.45	0
3	5	50,914	4,849	0.95961	3,134.63	0
3	6	50,914	2,220	0.63779	2,031.78	0
3	7	50,914	488	0.24846	834.47	0
3	8	50,914	239	0.41814	371.71	0
4	1	56,904	9,098	0.32524	9,619.73	0
4	2	56,904	8,974	0.51734	8,959.88	0
4	3	56,904	8,522	0.73617	7,753.12	0
4	4	56,904	4,985	0.63832	4,672.87	0
4	5	56,904	2,864	0.48789	2,955.96	0
4	6	56,904	1,576	0.33841	1,917.89	0
4	7	56,904	857	0.62105	787.71	0
5	1	62,551	9,446	0.19963	10,441.16	0
5	2	62,551	9,620	0.48453	9,727.04	0
5	3	62,551	10,928	0.96008	8,417.18	0
5	4	62,551	5,506	0.67673	5,069.52	0
5	5	62,551	1,973	0.07126	3,209.63	0
5	6	62,551	1,858	0.41349	2,082.31	0
6	1	67,205	13,791	0.67421	13,226.53	0
6	2	67,205	11,656	0.34160	12,322.62	0
6	3	67,205	11,664	0.75153	10,661.42	0
6	4	67,205	5,323	0.18643	6,421.71	0
6	5	67,205	3,731	0.39758	4,067.67	0
7	1	74,056	16,783	0.53513	16,707.39	0
7	2	74,056	17,370	0.85083	15,560.35	0
7	3	74,056	10,413	0.03875	13,467.69	0
7	4	74,056	9,144	0.77486	8,113.31	0
8	1	81,035	17,389	0.62141	16,946.46	0
8	2	81,035	15,132	0.36836	15,783.79	0
8	3	81,035	13,653	0.51292	13,658.91	0
9	1	90,568	22,871	0.54331	22,723.13	0
9	2	90,568	20,819	0.45069	21,167.46	0
10	1	99,940	22,916	0.47903	23,057.99	0
3	9	50,914	169	0.35782	352.10	1
4	8	56,904	372	0.62074	351.26	1
5	7	62,551	607	0.34127	856.37	1
6	6	67,205	2,079	0.26920	2,638.39	1
7	5	74,056	6,121	0.79950	5,138.10	1
8	4	81,035	8,253	0.52544	8,230.16	1
9	3	90,568	10,874	0.00027	18,325.98	1
10	2	99,940	18,260	0.10827	21,496.17	1

APPENDIX B – INSURER #4

AY	Lag	Premium	Loss	*p*-value	E[Loss]	Test
1	1	267,666	33,810	0.23507	45,787.91	0
1	2	267,666	45,318	0.34147	54,471.20	0
1	3	267,666	46,549	0.60459	44,198.73	0
1	4	267,666	35,206	0.71245	30,137.93	0
1	5	267,666	23,360	0.84899	16,758.58	0
1	6	267,666	12,502	0.68697	10,818.72	0
1	7	267,666	6,602	0.65542	5,883.45	0
1	8	267,666	3,373	0.64375	3,018.67	0
1	9	267,666	2,373	0.63989	2,176.39	0
1	10	267,666	778	0.50883	1,138.65	0
2	1	274,526	37,663	0.36921	44,076.07	0
2	2	274,526	51,771	0.53039	52,532.45	0
2	3	274,526	40,998	0.50420	42,610.85	0
2	4	274,526	29,496	0.56482	29,101.82	0
2	5	274,526	12,669	0.32337	16,191.71	0
2	6	274,526	11,204	0.62130	10,418.89	0
2	7	274,526	5,785	0.58105	5,664.48	0
2	8	274,526	4,220	0.80284	2,908.03	0
2	9	274,526	1,910	0.54331	2,088.72	0
3	1	268,161	40,630	0.44228	44,604.21	0
3	2	268,161	56,318	0.61502	53,036.35	0
3	3	268,161	56,182	0.82143	43,011.27	0
3	4	268,161	32,473	0.65752	29,363.66	0
3	5	268,161	15,828	0.52041	16,328.33	0
3	6	268,161	8,409	0.35092	10,522.64	0
3	7	268,161	7,120	0.73419	5,721.56	0
3	8	268,161	1,125	0.13471	2,948.32	0
4	1	276,821	40,559	0.54905	40,480.52	0
4	2	276,821	49,755	0.58209	48,176.91	0
4	3	276,821	39,323	0.55355	39,099.08	0
4	4	276,821	24,081	0.44189	26,694.41	0
4	5	276,821	13,209	0.43920	14,806.69	0
4	6	276,821	12,655	0.79588	9,561.64	0
4	7	276,821	2,921	0.19429	5,198.07	0
5	1	270,214	37,515	0.47765	39,993.47	0
5	2	270,214	51,068	0.62719	47,562.16	0
5	3	270,214	34,410	0.42354	38,619.62	0
5	4	270,214	25,529	0.51768	26,323.12	0
5	5	270,214	19,433	0.80611	14,676.52	0
5	6	270,214	5,728	0.17976	9,452.97	0
6	1	280,568	41,454	0.47870	44,192.46	0
6	2	280,568	53,552	0.57022	52,642.06	0
6	3	280,568	40,599	0.49452	42,665.27	0
6	4	280,568	40,026	0.85032	29,100.24	0
6	5	280,568	6,750	0.04135	16,206.07	0
7	1	344,915	57,783	0.50911	60,172.70	0
7	2	344,915	68,136	0.49687	71,598.64	0
7	3	344,915	86,915	0.90859	58,013.32	0
7	4	344,915	18,328	0.04564	39,731.67	0
8	1	371,139	62,011	0.43460	69,225.13	0
8	2	371,139	132,553	0.94135	82,082.69	0
8	3	371,139	21,083	0.00703	66,759.41	0
9	1	323,753	112,592	0.91665	72,997.37	0
9	2	323,753	33,783	0.02117	87,133.82	0
10	1	221,448	38,181	0.44006	43,939.91	0
2	10	274,526	887	0.56864	1,116.16	1
3	9	268,161	1,662	0.46527	2,132.64	1
4	8	276,821	1,043	0.15090	2,686.03	1
5	7	270,214	2,898	0.19996	5,153.72	1
6	6	280,568	5,513	0.11725	10,453.71	1
7	5	344,915	11,551	0.09181	22,134.60	1
8	4	371,139	17,129	0.01972	45,665.08	1
9	3	323,753	24,089	0.01220	70,644.07	1
10	2	221,448	37,485	0.28817	52,940.26	1