

**CAS MONOGRAPH SERIES
NUMBER 1**

STOCHASTIC LOSS RESERVING USING BAYESIAN MCMC MODELS

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



CASUALTY ACTUARIAL SOCIETY

The emergence of Bayesian Markov Chain Monte-Carlo (MCMC) models has provided actuaries with an unprecedented flexibility in stochastic model development. Another recent development has been the posting of a database on the CAS website that consists of hundreds of loss development triangles with outcomes. This monograph begins by first testing the performance of the Mack model on incurred data, and the Bootstrap Overdispersed Poisson model on paid data. It then will identify features of some Bayesian MCMC models that improve the performance over the above models. The features examined include (1) recognizing correlation between accident years; (2) introducing a skewed distribution defined over the entire real line to deal with negative incremental paid data; (3) allowing for a payment year trend on paid data; and (4) allowing for a change in the claim settlement rate. While the specific conclusions of this monograph pertain only to the data in the CAS Loss Reserve Database, the breadth of this study suggests that the currently popular models might similarly understate the range of outcomes for other loss triangles. This monograph then suggests features of models that actuaries might consider implementing in their stochastic loss reserve models to improve their estimates of the expected range of outcomes.

Glenn Meyers, FCAS, MAAA, CERA, and Ph.D., retired from ISO in 2011 after a 37-year career as an actuary. He holds a B.S. in Mathematics and Physics from Alma College, an M.A. in Mathematics from Oakland University, and a Ph.D. in Mathematics from SUNY at Albany. A frequent speaker at Casualty Actuarial Society (CAS) meetings, he has served, and continues to serve, the CAS and the International Actuarial Association on various research and education committees. He has also served on the CAS Board of Directors. Over the years, he has published articles in the *Proceedings* of the Casualty Actuarial Society, *Variance*, and the *Actuarial Review*. His research and other service contributions have been recognized by the CAS on numerous occasions. He has received the Woodward-Fondiller Prize on three separate occasions, the Dorweiller Prize twice, the DFA Prize, the Reserves Prize, the Matthew Rodermund Service Award, and the Michelbacher Significant Achievement Award. In retirement he still devotes a good portion of his time in pursuit of his passion for actuarial research.

STOCHASTIC LOSS RESERVING USING BAYESIAN MCMC MODELS

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



Casualty Actuarial Society
4350 North Fairfax Drive, Suite 250
Arlington, Virginia 22203
www.casact.org
(703) 276-3100

Stochastic Loss Reserving Using Bayesian MCMC Models
By Glenn Meyers

Copyright 2015 by the Casualty Actuarial Society.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher. For information on obtaining permission for use of material in this work, please submit a written request to the Casualty Actuarial Society.

Library of Congress Cataloging-in-Publication Data

Meyers, Glenn

Stochastic Loss Reserving Using Bayesian MCMC Models / Glenn Meyers

ISBN 978-0-9624762-7-3 (print edition)

ISBN 978-0-9624762-6-6 (electronic edition)

1. Actuarial science. 2. Loss reserving. 3. Insurance—Mathematical models.

I. Meyers, Glenn.

Copyright 2015, Casualty Actuarial Society

Contents

1. Introduction	1
2. The CAS Loss Reserve Database	4
3. Validating the Mack Model	6
4. Validating the Bootstrap ODP Model.....	11
5. Bayesian Models for Incurred Loss Data	14
6. Bayesian Models for Paid Loss Data.....	24
7. Process Risk, Parameter Risk, and Model Risk	37
8. Summary and Conclusions	39
References	41
Appendix A: The Data Selection Process.....	43
Appendix B: Introduction to Bayesian MCMC Models.....	45
Appendix C: Bayesian MCMC Model Implementation	54

2015 CAS Monograph Editorial Board

C. K. Stan Khury, Editor in Chief

Emmanuel Bardis

Craig Davis

Richard Fein

Jesse Groman

Ali Ishaq

Leslie Marlo

Sholom Feldblum, consultant

Glenn Meyers, consultant

Katya Prell, consultant

Foreword

This is the inaugural volume of the new CAS Monograph Series. A CAS monograph is an authoritative, peer reviewed, in-depth work on an important topic broadly within property and casualty actuarial practice.

In this monograph Glenn Meyers introduces a novel way of testing the predictive power of two loss reserving methodologies. He first demonstrates that the method commonly used for incurred losses tends to understate the range of possible outcomes. For paid losses, both methods tend to overstate the range of expected outcomes. Then he proceeds to apply Bayesian Markov Chain Monte-Carlo models (Bayesian MCMC) to improve the predictive power by recognizing three different elements implicit in the data histories. He is careful to note that the results are based on the histories contained in the CAS Database (of loss development triangles), which prevents one from making broad unqualified statements about the conclusions drawn in this work.

This monograph lays a solid foundation for future development and research in the area of testing the predictive power of loss reserving methods generally and in the use of Bayesian MCMC models to improve confidence in the selection of appropriate loss reserving methods. Glenn Meyers manages to show the way for raising the performance standard of what constitutes a reliable loss reserving methodology in any given situation.

C. K. “Stan” Khury
Chairperson
Monograph Editorial Board

1. Introduction

The recent attempts to apply enterprise risk management principles to insurance has placed a high degree of importance on quantifying the uncertainty in the various necessary estimates with stochastic models. For general insurers, the most important liability is the reserve for unpaid losses. Over the years a number of stochastic models have been developed to address this problem. Two of the more prominent nonproprietary models are those of Mack (1993, 1994) and England and Verrall (2002).

While these, and other, models provide predictive distributions¹ of the outcomes, very little work has been done to retrospectively test, or validate,² the performance of these models in an organized fashion on a large number of insurers. Recently with the permission of the National Association of Insurance Commissioners (NAIC), Peng Shi and I, in Meyers and Shi (2011), were able to assemble a database consisting of a large number of Schedule P triangles for six lines of insurance. These triangles came from insurer NAIC Annual Statements reported in 1997. Using subsequent annual statements we “completed the triangle” so that we could examine the outcomes and validate, the predictive distribution for any proposed model.

Sections 3 and 4 attempt to validate the models of Mack (1993, 1994) and England and Verrall (2002). As it turns out, these models do not accurately predict the distribution of outcomes for the data included in the subject database. Explanation for these results include the following.

- The insurance loss environment is too dynamic to be captured in a single stochastic loss reserve model. I.e., there could be different “black swan” events that invalidate any attempt to model loss reserves.³
- There could be other models that better fit the existing data.
- The data used to calibrate the model is missing crucial information needed to make a reliable prediction. Examples of such changes could include changes in the way the underlying business is conducted, such as changes in claim processes or changes in the direct/ceded/assumed reinsurance composition of the claim values in triangles.

¹ In this monograph, the term “predictive distribution” will mean the distribution of a random variable, X , given observed data x . By this definition the range of outcomes, X , could be quite wide. This, in contrast to the common usage of the term “predict,” connotes an ability to foresee the future and, in the context of the subject matter of this monograph, implies a fairly narrow range of expected outcomes.

² An explanation of “validate” will be given in Section 3.

³ The term “black swan,” as popularized by Taleb [2007], has come to be an oft-used term representing a rare high-impact event.

Possible ways to rule out the first item above are to (1) find a better model; and/or (2) find better data. This monograph examines a number of different models and data sources that are available in Schedule P. The data in Schedule P includes net paid losses, net incurred losses, and net premiums.

A characteristic of loss reserve models is that they are complex in the sense that they have a relatively large number of parameters. A major difficulty in quantifying the uncertainty in the parameters of a complex model has been that it takes a fair amount of effort to derive a formula for the predictive distribution of outcomes. See Mack's (1993, 1994) papers and Bardis, Majidi and Murphy's (2012) paper for examples of analytic solutions. Taking advantage of the ever-increasing computer speed, England and Verrall (2002) pass the work on to computers using a bootstrapping methodology with the overdispersed Poisson distribution (ODP). Not too long ago, the Bayesian models⁴ were not practical for models of any complexity. But with the relatively recent introduction of Bayesian Markov Chain Monte Carlo (MCMC) models, complex Bayesian stochastic loss reserve models are now practical in the current computing environment.

Although Markov chains have long been studied by probability theorists, it took a while for their application to Bayesian statistics to be recognized. Starting in the 1930s, physicists began using statistical sampling from Markov chains to solve some of the more complex problems in nuclear physics. The names associated with these efforts include Enrico Fermi, John von Neumann, Stanislaw Ulam and Nicolas Metropolis. This led to the Metropolis algorithm for generating Markov chains. Later on, W. Keith Hastings (1970) recognized the importance of Markov chains for mainstream statistics and published a generalization of the Metropolis algorithm. That paper was largely ignored by statisticians at the time as they were not accustomed to using simulations for statistical inference. Gelfand and Smith (1990) provided the "aha" moment for Bayesian statisticians. They pulled together a relevant set of existing ideas at a time when access to fast computing was becoming widely available. In the words of McGrayne (2011, Part V): "Almost instantaneously MCMC and Gibbs sampling changed statisticians' entire method of attacking problems. In the words of Thomas Kuhn, it was a paradigm shift. MCMC solved real problems, used computer algorithms instead of theorems, and led statisticians and scientists into a world where 'exact' meant 'simulated' and repetitive computer simulations replaced mathematical equations. It was a quantum shift in statistics" (p. 225).

As was the case for the other social sciences, Bayesian MCMC should eventually have a profound effect on actuarial science. And in fact, its effect has already begun. Scollnik (2001) introduced actuaries to Bayesian MCMC models. De Alba (2002) along with Ntzoufras and Dellaportas (2002) quickly followed by applying these models to the loss reserving problem. Verrall (2007) applied them to the chain ladder model. In the time since these papers were written, the algorithms implementing

⁴ By a "Bayesian model" I mean a model with its parameters having a prior distribution specified by the user. By "Bayesian estimation" I mean the process of predicting the distribution of a "statistic of interest" from the posterior distribution of a Bayesian model.

Bayesian MCMC models have gotten more efficient, and the associated software has gotten more user friendly.

Here is the situation we now face. First, we are able to construct a wide variety of proposed models and predict their distribution of outcomes with the Bayesian MCMC methodology. Second, we are able to validate a proposed stochastic loss reserve model using a large number of insurers on the CAS Loss Reserve Database. If the insurance loss environment is not dominated by a series of unique “black swan” events, it should be possible to systematically search for models and data that successfully validate. This monograph describes the results I have obtained to date in my pursuit of this goal.

While I believe I have made significant progress in identifying models that do successfully validate on the data I selected from the CAS Loss Reserve Database, it should be stressed that more work needs to be done to confirm or reject these results for different data taken from different time periods.

The intended audience for this monograph consists of general insurance actuaries who are familiar with the Mack (1993, 1994) and the England and Verrall (2002) models. While I hope that most sections will be readable by a “generalist” actuary, those desiring a deeper understanding should work with the companion scripts to this monograph.⁵

The computer scripts used to implement these models is written in the R programming language. To implement the MCMC calculations the R script contains another script that is written in JAGS. Like R, JAGS is an open source programming language one can download for free. For readers who are not familiar with R and JAGS, here are some links to help the reader get started.

- <http://r-project.org> The home page of the R-Project.
- <http://mcmc-jags.sourceforge.net/> A link to download JAGS.
- <http://www.rstudio.com/> A currently popular editor for R and JAGS script.

⁵ Scripts are available at https://www.casact.org/sites/default/files/2021-03/Monograph_Tables_and_Scripts.xlsx

2. The CAS Loss Reserve Database

In order to validate a model, one need not only the data used to build the model, but also the data with outcomes that the model was built to predict. Schedule P of the NAIC Annual Statement contains insurer-level run-off triangles of aggregated losses by line of insurance. Triangles for both paid and incurred losses (net of reinsurance) are reported in Schedule P.⁶ To get the outcomes, one must look at subsequent Annual Statements.

To illustrate the calculations in this monograph, I selected incurred and paid loss triangles from a single insurer in the database, whose data are in Tables 1, 2 and 3. The data in the loss triangles above the diagonal lines are available in the 1997 Annual Statement. These data are used to build the models discussed below. The outcome data below the diagonal lines were extracted, by row, from the Annual Statements listed in the “Source” column. These data are used to validate the models.

The database, along with a complete description of how it was constructed and how the insurers were selected, is available on the CAS website at <http://www.casact.org/publications-research/research/research-resources/loss-reserving-data-pulled-naic-schedule-p>.

This monograph will fit various loss reserve models, and test the predictive distributions, to a set of 200 insurer loss triangles taken from four Schedule P (50 from each of Commercial Auto, Personal Auto, Workers Compensation and Other Liability) lines of insurance. An underlying assumption of these models is that there have not been any substantial changes in the insurer’s operation. In our real world, insurers are always tinkering with their operations. Schedule P provides two hints of possible insurer operational changes:

- Changes in the net premium from year-to-year
- Changes in the ratio of net to direct premium from year to year

The criteria for selecting the 200 insurer loss triangles rests mainly on controlling for changes in the above two items. Appendix A gives the group codes for the selected insurers by line of insurance and gives a detailed description of the selection algorithm.

⁶ Paid losses are reported in Part 3 of Schedule P. Incurred losses are the losses reported in Part 2 minus those reported in Part 4 of Schedule P.

Table 1. Illustrative Insurer Net Written Premium

AY	1	2	3	4	5	6	7	8	9	10
Premium	5812	4908	5454	5165	5214	5230	4992	5466	5226	4962

Table 2. Illustrative Insurer Incurred Losses Net of Reinsurance

AY/Lag	1	2	3	4	5	6	7	8	9	10	Source
1988	1722	3830	3603	3835	3873	3895	3918	3918	3917	3917	1997
1989	1581	2192	2528	2533	2528	2530	2534	2541	2538	2532	1998
1990	1834	3009	3488	4000	4105	4087	4112	4170	4271	4279	1999
1991	2305	3473	3713	4018	4295	4334	4343	4340	4342	4341	2000
1992	1832	2625	3086	3493	3521	3563	3542	3541	3541	3587	2001
1993	2289	3160	3154	3204	3190	3206	3351	3289	3267	3268	2002
1994	2881	4254	4841	5176	5551	5689	5683	5688	5684	5684	2003
1995	2489	2956	3382	3755	4148	4123	4126	4127	4128	4128	2004
1996	2541	3307	3789	3973	4031	4157	4143	4142	4144	4144	2005
1997	2203	2934	3608	3977	4040	4121	4147	4155	4183	4181	2006

Table 3. Illustrative Insurer Paid Losses Net of Reinsurance

AY/Lag	1	2	3	4	5	6	7	8	9	10	Source
1988	952	1529	2813	3647	3724	3832	3899	3907	3911	3912	1997
1989	849	1564	2202	2432	2468	2487	2513	2526	2531	2527	1998
1990	983	2211	2830	3832	4039	4065	4102	4155	4268	4274	1999
1991	1657	2685	3169	3600	3900	4320	4332	4338	4341	4341	2000
1992	932	1940	2626	3332	3368	3491	3531	3540	3540	3583	2001
1993	1162	2402	2799	2996	3034	3042	3230	3238	3241	3268	2002
1994	1478	2980	3945	4714	5462	5680	5682	5683	5684	5684	2003
1995	1240	2080	2607	3080	3678	2004	4117	4125	4128	4128	1997
1996	1326	2412	3367	3843	3965	4127	4133	4141	4142	4144	2005
1997	1413	2683	3173	3674	3805	4005	4020	4095	4132	4139	2006

3. Validating the Mack Model

Probably the two most popular nonproprietary stochastic loss reserve models are the Mack (1993, 1994) chain-ladder model and the England and Verrall (2002) bootstrap ODP model. This section describes an attempt to validate the Mack model on the incurred loss data from several insurers that are included in the CAS database. Validating the bootstrap ODP model will be addressed in the following section.

Let's begin with the classic chain-ladder model. Let $C_{w,d}$ denote the accumulated loss amount, either incurred or paid, for accident year, w , and development lag, d , for $1 \leq w \leq K$ and $1 \leq d \leq K$. $C_{w,d}$ is known for the "triangle" of data specified by $w + d \leq K + 1$. The goal of this model is to estimate the loss amounts in the last column of data, $C_{w,K}$ for $w = 2, \dots, K$. To use the chain-ladder model, one first calculates the age to age factors given by

$$f_d = \frac{\sum_{w=1}^{K-d} C_{w,d+1}}{\sum_{w=1}^{K-d} C_{w,d}} \quad \text{for } d = 1, \dots, K - 1.$$

The chain-ladder estimate of $C_{w,K}$ is the product of the latest reported loss, $C_{w,K+1-w}$, and the subsequent age-to-age factors $f_{K+1-w} \cdots f_{K-1}$. Putting this together, we have

$$\hat{C}_{w,K} = C_{w,K+1-w} \cdot f_{K+1-w} \cdots f_{K-1}$$

Taylor (1986, p. 40) discusses the origin of the chain-ladder model and concludes that "It appears that it probably originated in the accounting literature, and was subsequently absorbed in to, or rediscovered in, the actuarial." He goes on to say that "Of course, one must bear in mind that both the chain-ladder model and estimation method are fairly obvious and might have been derived several times in past literature." Taylor believes that the rather whimsical name of the model was first used by Professor R. E. Beard as he championed the method in the early 1970s while working as a consultant to the U.K. Department of Trade.

Mack (1993, 1994) turns the deterministic chain ladder model into a stochastic model by first treating $\tilde{C}_{w,d}$ as a random variable that represents the accumulated loss amount in the (w, d) cell. He then makes three assumptions.⁷

⁷ Depending on the context, various quantities, such as $C_{w,d}$, will represent observations, estimates or random variables. In situations where it might not be clear, let's adopt the convention that for a quantity X , \tilde{X} will indicate that X is being treated as a random, or simulated, variable, \hat{X} will denote an estimate of X , and a bare X will be treated as a fixed observation or parameter.

1. $E[\tilde{C}_{w,d+1} | C_{w,1}, \dots, C_{w,d}] = C_{w,d} \cdot f_d$
2. For any given d , the random variables $\tilde{C}_{v,d}$ and $\tilde{C}_{w,d}$ are independent for $v \neq w$.
3. $\text{Var}[\tilde{C}_{w,d+1} | C_{w,1}, \dots, C_{w,d}] = C_{w,d} \cdot \alpha_d^2$

The Mack estimate for $E[\tilde{C}_{w,K}]$ for $w = 2, \dots, K$ is given by

$$\hat{C}_{w,K} = C_{w,K+1-w} \cdot \hat{f}_{K+1-w} \cdots \hat{f}_{K-1}$$

where

$$\hat{f}_d = \frac{\sum_{w=1}^{K-d} C_{w,d+1}}{\sum_{w=1}^{K-d} C_{w,d}}$$

Given his assumptions above, Mack then derives expressions for the standard deviations $SD[\tilde{C}_{w,K}]$ and $SD[\sum_{w=2}^K \tilde{C}_{w,K}]$. Table 4 applies Mack’s expressions to the illustrative insured data in Table 2 using the R “ChainLadder” package.

In addition to the loss statistics calculated by the Mack expressions, Table 4 contains the outcomes $\{C_{w,10}\}$ from Table 2. Following Mack’s suggestion, I calculated the percentile of $\sum_{w=1}^{10} C_{w,10}$ assuming a lognormal distribution with matching the mean and the standard deviation.

Taken by itself, an outcome falling in the 86th percentile gives us little information, as that percentile is not unusually high. If the percentile was, say, above the 99.5th percentile, suspicion might be warranted. My intent here is to test the general applicability of the Mack model on incurred loss triangles. To do this, I selected 200 incurred loss

Table 4. Mack Model Output for the Illustrative Insurer Incurred Losses

w	$\hat{C}_{w,10}$	SD	CV	$C_{w,10}$	Percentile
1	3917	0	0.000	3917	
2	2538	0	0.000	2532	
3	4167	3	0.001	4279	
4	4367	37	0.009	4341	
5	3597	34	0.010	3587	
6	3236	40	0.012	3268	
7	5358	146	0.027	5684	
8	3765	225	0.060	4128	
9	4013	412	0.103	4144	
10	3955	878	0.222	4181	
Total	38914	1057	0.027	40061	86.03

triangles, 50 each from four different lines of insurance, and calculated the percentile of the $\sum_{w=1}^{10} C_{w,10}$ outcome for each triangle. My criteria for “general applicability of the model” is that these percentiles should be uniformly distributed. And for a sufficiently large sample, uniformity is testable! Klugman, Panjer, and Willmot (2012, Section 16.3) describe a variety of tests that can be applied in this case.

Probably the most visual test for uniformity is a plot of a histogram. If the percentiles are uniformly distributed, we should expect the height of the bars to be equal. Unless the sample size is very large, this will rarely be the case because of random fluctuations. A visual test of uniformity that allows one to test for statistical significance is the p - p plot combined with the Kolmogorov–Smirnov (K–S) test. Here is how it works. Suppose one has a sample of n predicted percentiles ranging from 0 to 100 and sort them into increasing order. The expected value of these percentiles is given by $\{e_i\} = 100 \cdot \{1/(n+1), 2/(n+1), \dots, n/(n+1)\}$. One then plots the expected percentiles on the horizontal axis against the sorted predicted percentiles on the vertical axis. If the predicted percentiles are uniformly distributed, we expect this plot to lie along a 45° line. According to the K–S test as described by Klugman, Panjer, and Willmot (2012, p. 331), one can reject the hypothesis that a set of percentiles $\{p_i\}$ is uniform at the 5% level if $D \equiv \max |p_i - f_i|$ is greater than its critical value, $136/\sqrt{n}$ where $\{f_i\} = 100 \cdot \{1/n, 2/n, \dots, n/n\}$. This is represented visually on a p - p plot by drawing lines at a distance $136/\sqrt{n}$ above and below the 45° line.⁸ We reject the hypothesis of uniformity if the p - p plot lies outside the band defined by those lines. For the purposes of this monograph, a model will be deemed “validated” if it passes the K–S test at the 5% level.

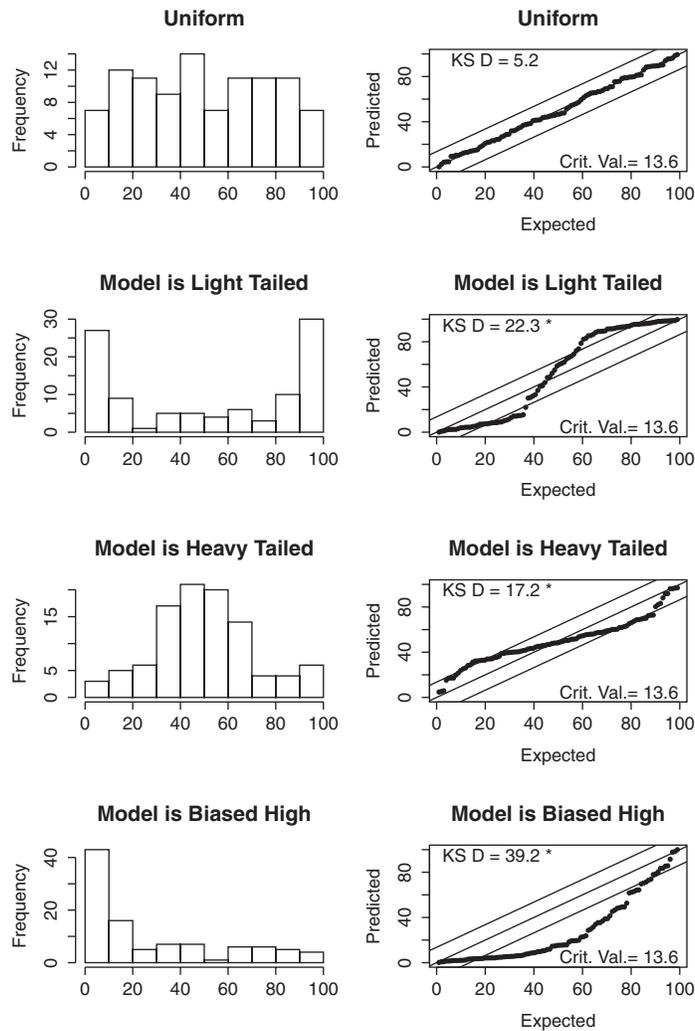
Klugman, Panjer, and Willmot (2012, p. 332) also discusses a second test of uniformity that is applicable in this situation. The Anderson–Darling (A–D) test is similar to the Kolmogorov–Smirnov test, but it is more sensitive to the fit in the extreme values (near the 0th and the 100th percentile) of the distribution. I applied the A–D test along with the K–S test on the models described in this monograph with the result that almost all A–D tests failed. If in the future someone develops a more refined model, we can raise the bar to the more stringent A–D test. Until that happens, I think the K–S test is the best tool to differentiate between models.

Figure 1 shows both histograms and p - p plots for simulated data with $n = 100$. The plots labeled “Uniform” illustrate the expected result. The K–S D statistic accompanies each p - p plot. The “*” indicates that the D statistic is above its critical value.

Figure 1 also shows p - p plots for various departures from uniformity. For example, if the predicted distribution is too light in the tails, there are more than expected high and low percentiles in the predicted outcomes and we see a p - p plot that looks like a slanted “S” curve. If the predicted distribution is too heavy in the tails, there are more than expected middle percentiles in the predicted outcomes and we see a p - p plot that looks like a slanted backward “S” curve. If the model predicts results that are in general too high, predicted outcomes in the low percentiles will be more frequent.

⁸ This is an approximation as $f_i \approx e_i$.

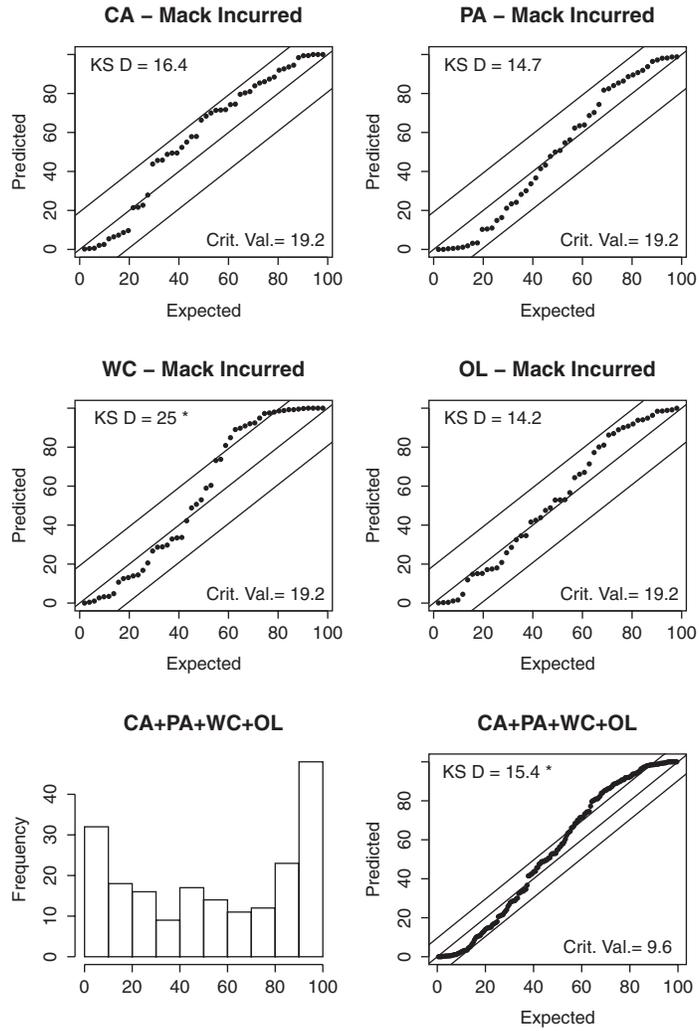
Figure 1. p - p Plots Test for Uniformity



To validate the Mack model, I repeated the calculations for the 200 selected incurred loss reserve triangles.

Figure 2 shows the p - p plots for the Mack model. The plots were first done separately for the outcome percentiles in each line of insurance. Although the plots fall inside the K-S band for three of the four lines, the plots for all four of the lines resemble the slanted “S” curve that is characteristic of a light tailed predicted distribution. When we combine the outcome percentiles of all four lines, the p - p plot lies outside the K-S band and we conclude that the distribution predicted by the Mack model is too light in the tails for these data. In all the validation plots below the K-S critical values are 19.2 and 9.6 for the individual lines and all lines combined respectively.

Figure 2. $p-p$ Plots for the Mack Model on Incurred Loss Triangles



4. Validating the Bootstrap ODP Model

This section does an analysis similar to that done in the last section for the bootstrap ODP model as described by England and Verrall (2002) and implemented by the R “ChainLadder” package. This model was designed to work with incremental losses, $I_{w,d}$, rather than the cumulative losses $C_{w,d}$, where $I_{w,1} = C_{w,1}$ and $I_{w,d} = C_{w,d} - C_{w,d-1}$ for $d > 1$.

A key assumption made by this model is that the incremental losses are described by the overdispersed Poisson distribution with

$$E[\tilde{I}_{w,d}] = \alpha_w \cdot \beta_d \quad \text{and} \quad \text{Var}[\tilde{I}_{w,d}] = \phi \cdot \alpha_w \cdot \beta_d$$

The parameters of the model can be estimated by a standard generalized linear model (GLM) package.⁹ They then use a bootstrap resampling procedure to quantify the volatility of the estimate.

England and Verrall point out that the using the ODP model on incremental losses almost all but requires one to use paid, rather than incurred, losses since the overdispersed Poisson model is defined only for nonnegative losses. Incurred losses include estimates by claims adjusters that can (and frequently do) get adjusted downward. Negative incremental paid losses occasionally occur because of salvage and subrogation, but a feature of the GLM estimation procedure allows for negative incremental losses as long as all column sums of the loss triangle remain positive.

Table 5 gives the estimates of the mean, the standard deviation for both the ODP (with 10,000 bootstrap simulations) and Mack models on the data in Table 3. The predicted percentiles of the 10,000 outcomes are also given for each model.

The validation p - p plots, similar to those done in the previous section, for both the ODP and the Mack models on paid data, are in Figures 2 and 3. The results for both models are quite similar. Neither model validates on the paid triangles. A comparison of the p - p plots in Figures 3 and 4 with the illustrative plots in Figure 1 suggests that the expected loss estimates of both models tend to be too high for these data.

Let’s now consider the results of this and the prior section. These sections show that two popular models do not validate on outcomes of the 200 Schedule P triangles drawn from the CAS Loss Reserve Database. These models do not validate in different ways when we examine paid and incurred triangles. For incurred triangles, the distribution

⁹ England and Verrall (2002) use a log link function in their GLM. They also note that the GLM for the ODP maximizes the quasi-likelihood, allowing the model to work with continuous (non-integer) losses.

Table 5. ODP and Mack Model Output for the Illustrative Insurer Paid Losses

w	ODP			Mack			Outcome $C_{w,10}$
	$\hat{C}_{w,10}$	SD	CV	$\hat{C}_{w,10}$	SD	CV	
1	3912	0	0	3912	0	0.0000	3912
2	2532	21	0.0083	2532	0	0.0000	2527
3	4163	51	0.0123	4162	3	0.0007	4274
4	4369	85	0.0195	4370	28	0.0064	4341
5	3554	96	0.027	3555	35	0.0098	3583
6	3211	148	0.0461	3213	157	0.0489	3268
7	5161	240	0.0465	5167	251	0.0486	5684
8	3437	332	0.0966	3442	385	0.1119	4128
9	4220	572	0.1355	4210	750	0.1781	4144
10	4635	1048	0.2261	4616	957	0.2073	4139
Total	39193	1389	0.0354	39177	1442	0.0368	40000
Percentile		73.91			72.02		

Figure 3. $p-p$ Plots for the Bootstrap ODP Model on Paid Loss Triangles

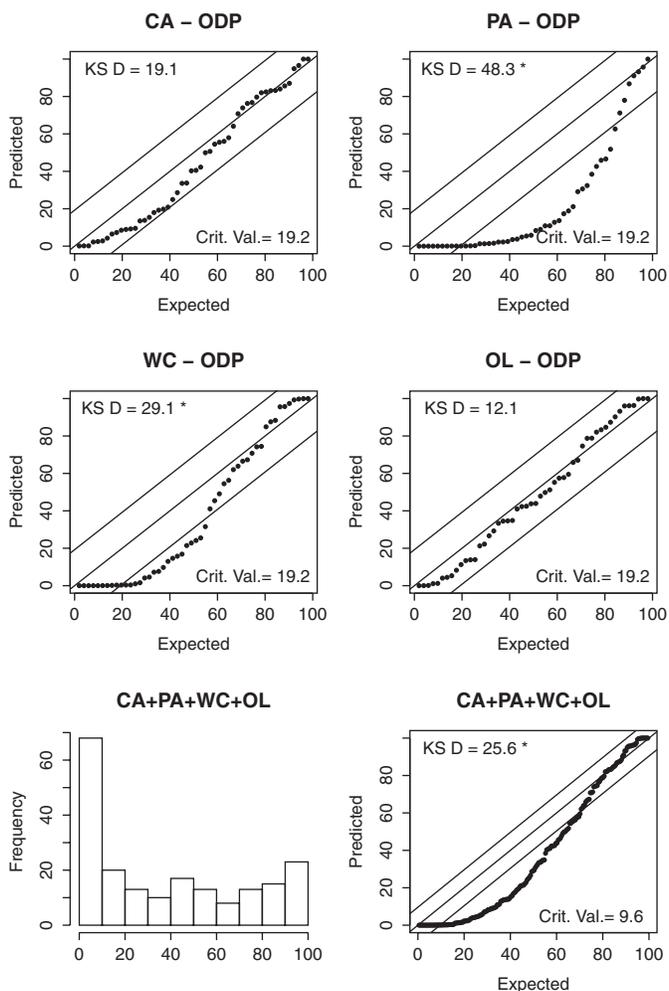
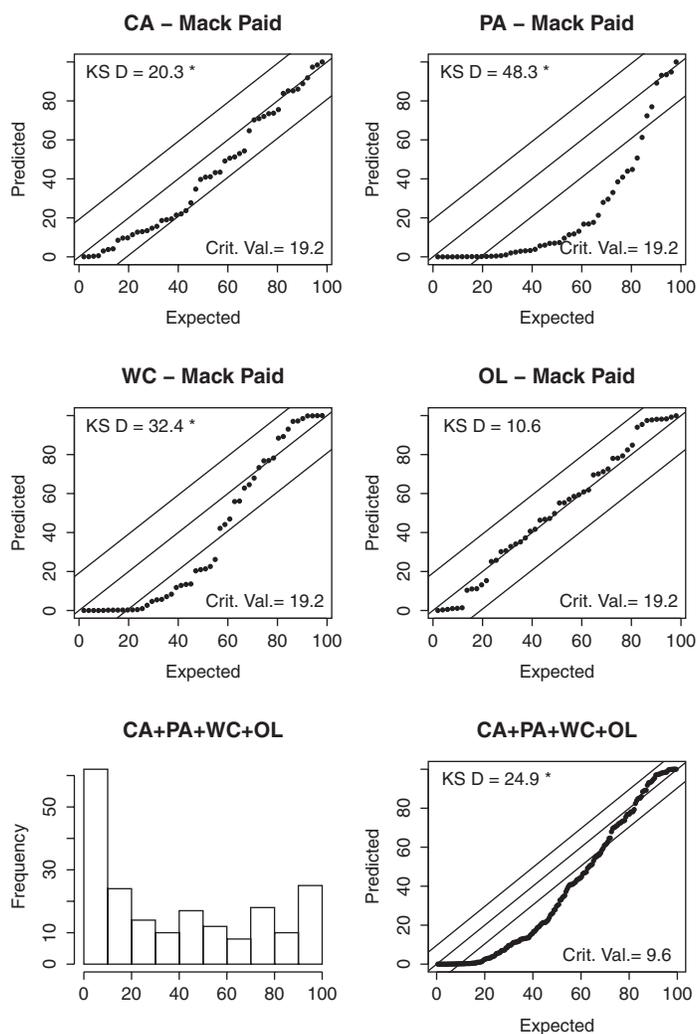


Figure 4. p - p Plots for the Mack Model on Paid Loss Triangles



predicted by the Mack model has a light tail. For paid triangles, the distributions predicted by both the Mack and the bootstrap ODP models tend to produce expected loss estimates that are too high. There are two plausible explanations for these observations:

1. The insurance loss environment has experienced changes that are not observable at the current time.
2. There are other models that can be validated.

To disprove the first explanation, one can develop models that do validate. Failing to develop a model that validates may give credence to, but does not necessarily confirm, that the first explanation is true. This monograph now turns to describing some efforts to find models that do validate.

5. Bayesian Models for Incurred Loss Data

I will begin this section on Bayesian MCMC models by quoting the advice of Verrall (2007). “For the readers for whom this is the first time they have encountered MCMC methods, it is suggested that they simply accept that they are a neat way to get the posterior distributions for Bayesian models and continue reading the paper. If they like the ideas and would like to find out more . . .” they should read the introduction in Appendix B. Keep in mind that the state of the art (e.g., faster multi-core personal computers, more efficient algorithms and more user-friendly software) is still rapidly advancing. Appendix C explains what I did with the current state of the art, as I perceived it, at the time I was writing this monograph.

Now let’s get to the loss reserve models. As pointed out in Section 3, the Mack model did not validate on the insurers listed in Appendix A using the loss data that are in the CAS Loss Reserve Database. This section presents two Bayesian MCMC models that were proposed in an attempt to find a model that does validate on these data.

The way the Mack model did not validate, i.e., it underestimated the variability of the ultimate loss estimates, suggested a direction to go in order to fix it. Here are two ways to improve the recognition of the inherent variability of the predictive distribution.

1. The Mack model multiplies the age-to-age factors by the last observed loss, $C_{w,11-w}$. One can think of the $C_{w,11-w}$ s as fixed level parameters. A model that treats the level of the accident year as random will predict more risk.
2. The Mack model assumes that the loss amounts for different accident years are independent. A model that allows for correlation between accident years could increase the standard deviation of $\sum_{w=1}^{10} \tilde{C}_{w,10}$.

I propose two different models to address the underestimation of the variability of the ultimate loss. The first model replaces the fixed level parameters, given by the last observed accident year, in the Mack model with random level parameters. As we shall see, this model improves the estimation of the variability, but does not go far enough. The second, and more complicated model, considers correlation between the accident years.

The Leveled Chain Ladder (LCL) Model

Let:

1. $\mu_{w,d} = \alpha_w + \beta_d$.
2. $\tilde{C}_{w,d}$ has a lognormal distribution with log mean $\mu_{w,d}$ and log standard deviation σ_d subject to the constraint that $\sigma_1 > \sigma_2 > \dots > \sigma_{10}$.

To prevent overdetermining the model, set $\beta_{10} = 0$. The parameters $\{\alpha_w\}$, $\{\sigma_d\}$ and the remaining $\{\beta_d\}$ are assigned relatively wide prior distributions as follows:

1. Each $\alpha_w \sim \text{normal}(\log(\text{Premium}_w) + \text{logelr}, \sqrt{10})$ where the parameter $\text{logelr} \sim \text{uniform}(-1, 0.5)$.¹⁰
2. Each $\beta_d \sim \text{uniform}(-5, 5)$ for $d < 10$.
3. Each $\sigma_d = \sum_{i=d}^{10} a_i$ where $a_i \sim \text{uniform}(0, 1)$.

The hierarchical structure of the priors in (3) above assures that $\sigma_1 > \sigma_2 > \dots > \sigma_{10}$. The rationale behind this structure is that as d increases, there are fewer claims that are open and subject to random outcomes.

The next model adds a between-year correlation feature.¹¹

The Correlated Chain-Ladder (CCL) Model

Let:

1. Each $\alpha_w \sim \text{normal}(\log(\text{Premium}_w) + \text{logelr}, \sqrt{10})$ where the parameter $\text{logelr} \sim \text{uniform}(-1, 0.5)$.
2. $\mu_{1,d} = \alpha_1 + \beta_d$.
3. $\mu_{w,d} = \alpha_w + \beta_d + \rho \cdot (\log(C_{w-1,d}) - \mu_{w-1,d})$ for $w > 1$.
4. $\tilde{C}_{w,d}$ has a lognormal distribution with log mean $\mu_{w,d}$ and log standard deviation σ_d subject to the constraint that $\sigma_1 > \sigma_2 > \dots > \sigma_{10}$.

Note that the CCL model reduces to the LCL model when $\rho = 0$.

If the parameters $\{\alpha_w\}$, $\{\beta_d\}$, and ρ are given, the parameter ρ is equal to the coefficient of correlation between $\log(\tilde{C}_{w-1,d})$ and $\log(\tilde{C}_{w,d})$. To see this we first note that unconditionally:

$$\begin{aligned} E(\log(\tilde{C}_{w,d})) &= \mu_{w,d} \\ &= \alpha_w + \beta_d + \rho \cdot (\log(\tilde{C}_{w-1,d}) - \mu_{w-1,d}) \\ &= \alpha_w + \beta_d \end{aligned}$$

Given $C_{w-1,d}$ we have that:

$$\begin{aligned} E((\log(\tilde{C}_{w,d}) - (\alpha_w + \beta_d)) \cdot (\log(C_{w-1,d}) - \mu_{w-1,d})) \\ &= (\mu_{w,d} - (\alpha_w + \beta_d)) \cdot (\log(C_{w-1,d}) - \mu_{w-1,d}) \\ &= \rho \cdot (\log(C_{w-1,d}) - \mu_{w-1,d})^2 \end{aligned}$$

¹⁰ The JAGS expression for a normal distribution uses what it calls a “precision” parameter equal to the reciprocal of the variance. The standard deviation, $\sqrt{10}$, corresponds to the rather low precision of 0.1.

¹¹ Some of the models I tried before getting to this one are described in my working paper Meyers (2012). Note that what I call the LCL model in that paper is different from the LCL model above.

Then the coefficient of correlation between $\tilde{C}_{w,d}$ and $\tilde{C}_{w-1,d}$ is given by:

$$\begin{aligned} & E_{C_{w-1,d}} \left(E_{C_{w,d}} \left(\frac{\log(\tilde{C}_{w,d}) - (\alpha_w + \beta_d)}{\sigma_d} \cdot \frac{\log(C_{w-1,d}) - \mu_{w-1,d}}{\sigma_d} \mid C_{w-1,d} \right) \right) \\ &= E_{C_{w-1,d}} \left(\frac{\rho \cdot (\log(\tilde{C}_{w-1,d}) - \tilde{\mu}_{w-1,d})^2}{\sigma_d^2} \right) = \rho \end{aligned}$$

To prevent overdetermining the model, set $\beta_{10} = 0$. The parameters $\{\alpha_w\}$, $\{\sigma_d\}$, ρ and the remaining $\{\beta_d\}$ are assigned relatively wide prior distributions as follows:

1. Each $\alpha_w \sim \text{normal}(\log(\text{Premium}_w) + \text{logelr}, \sqrt{10})$ where the parameter $\text{logelr} \sim \text{uniform}(-1, 0.5)$.¹²
2. Each $\beta_d \sim \text{uniform}(-5, 5)$ for $d < 10$.
3. $\rho \sim \text{uniform}(-1, 1)$ —The full permissible range for ρ .
4. Each $\sigma_d = \sum_{i=d}^K a_i$ where $a_i \sim \text{uniform}(0,1)$.

I deliberately chose rather diffuse¹³ prior distributions since I had no direct knowledge of the claims environment other than the data that are reported in Schedule P. While preparing annual statements, actuaries with more direct knowledge of the claims environment normally attempt to reflect this knowledge in their unpaid loss estimates. Bornhuetter and Ferguson (1972) describe a very popular method where one can reflect knowledge of an insurer’s expected loss ratio in their estimates. With minor modifications of the JAGS script, one can reflect this knowledge by specifying more restrictive priors for $\{\alpha_w\}$ parameters and the logelr parameter.

The predictive distribution of outcomes is a mixed distribution where the mixing is specified by the posterior distribution of parameters. Here is what the script for the CCL model does.

The predictive distribution for $\sum_{w=1}^{10} C_{w,10}$ is generated by a simulation. For each parameter set $\{\alpha_w\}$, $\{\beta_d\}$, $\{\sigma_d\}$ and $\{\rho\}$, start with the given $C_{1,10}$ and calculate the mean, $\mu_{2,10}$. Then simulate $\tilde{C}_{2,10}$ from a lognormal distribution with log mean, $\mu_{2,10}$, and log standard deviation, σ_{10} . Similarly, use the result of this simulation to simulate $\tilde{C}_{2,10}, \dots, \tilde{C}_{10,10}$. Then form the sum $C_{1,10} + \sum_{w=2}^{10} \tilde{C}_{w,10}$. The script generates 10,000 simulations that make up a sample from the predictive distribution from which one can calculate various statistics such as the mean, standard deviation and the percentile of the outcome. Here is a more detailed explanation of this process.

Given the group code for an insurer in the CAS Loss Reserve Database, the R script for the CCL Model performs the following steps:

1. Reads in the data triangle $\{C_{w,d}\}$ for the insurer identified by the group code.
2. Runs the JAGS script and gets a sample of 10,000 parameter sets, $\{\alpha_w\}$, $\{\beta_d\}$, $\{\sigma_d\}$ and ρ from the posterior distribution of the CCL model.

¹² The JAGS expression for a normal distribution uses what it calls a “precision” parameter equal to the reciprocal of the variance. The standard deviation, corresponds to the rather low precision of 0.1.

¹³ One might also use a “noninformative” prior distribution. Noninformative prior distributions are usually attached to a specific mathematical objective. See, for example, Section 3.3 of Berger (1985).

3. Simulates 10,000 copies, one for each parameter set in (2) above, of $\{\tilde{C}_{w,10}\}_{d=2}^{10}$. The simulation proceeds as follows.
 - Set $\mu_{1,10} = \alpha_1 + \beta_{10}$. Recall that $C_{1,10}$ is given in the original data.
 - Set $\tilde{\mu}_{2,10} = \alpha_2 + \beta_{10} + \rho \cdot (\log(C_{1,10}) - \mu_{1,10})$. Simulate $\tilde{C}_{2,10}$ from a lognormal distribution with log mean $\mu_{2,10}$ and log standard deviation σ_{10} .
 - Set $\tilde{\mu}_{3,10} = \alpha_3 + \beta_{10} + \rho \cdot (\log(\tilde{C}_{2,10}) - \mu_{2,10})$. Simulate $\tilde{C}_{3,10}$ from a lognormal distribution with log mean $\mu_{2,10}$ and log standard deviation σ_{10} .
 - ...
 - Set $\tilde{\mu}_{10,10} = \alpha_{10} + \beta_{10} + \rho \cdot (\log(\tilde{C}_{9,10}) - \tilde{\mu}_{9,10})$. Simulate $\tilde{C}_{10,10}$ from a lognormal distribution with log mean $\mu_{10,10}$ and log standard deviation σ_{10} .
4. For each w , calculate summary statistics $\hat{C}_{w,10} = \text{mean}(\tilde{C}_{w,10})$ and SD = standard deviation ($\tilde{C}_{w,10}$). Calculate similar statistics for the total $C_{1,10} + \sum_{w=2}^{10} \tilde{C}_{w,10}$.
5. Calculate the percentile of the outcome by counting how many of the 10,000 instances of $\sum_{w=2}^{10} \tilde{C}_{w,10}$ are \leq the actual outcomes $\sum_{w=2}^{10} C_{w,10}$.

Table 6 gives the results from the first five MCMC samples produced by the script for the CCL model applied to the losses for the illustrative insurer in Table 2. The top 31 rows of that table were generated in Step 2 of the simulation above. The remaining rows were generated in Step 3.

Table 6. Illustrative MCMC Simulations

	MCMC Sample Number				
	1	2	3	4	5
α_1	8.2763	8.2452	8.2390	8.2591	8.2295
α_2	7.8226	7.7812	7.8008	7.8048	7.7810
α_3	8.2625	8.3200	8.2929	8.2883	8.2642
α_4	8.3409	8.3286	8.3539	8.3622	8.3159
α_5	8.2326	8.1166	8.1093	8.1855	8.1523
α_6	8.1673	8.0307	8.0491	8.1727	8.0470
α_7	8.6403	8.4776	8.4113	8.5815	8.4871
α_8	8.2177	8.2488	8.2708	8.0752	8.1763
α_9	8.3174	8.2007	8.2589	8.3744	8.2653
α_{10}	7.4101	8.0036	8.7584	8.4241	8.8420
β_1	-0.5125	-0.5180	-0.6504	-0.4947	-0.7384
β_2	-0.2756	-0.1014	-0.1231	-0.2138	-0.0844
β_3	-0.1271	-0.0313	-0.0622	-0.0758	-0.0498
β_4	-0.1013	-0.0090	0.0165	0.0439	0.0479
β_5	0.0518	-0.0109	0.0060	0.0034	0.0610
β_6	0.0180	0.0885	0.0139	0.0175	0.0709
β_7	0.0105	0.0583	0.0205	0.0427	0.0362

(continued on next page)

Table 6. Illustrative MCMC Simulations(continued)

	MCMC Sample Number				
	1	2	3	4	5
β_8	0.0400	-0.0090	0.0612	0.0444	0.0338
β_9	0.0005	0.0287	0.0419	0.0116	0.0333
β_{10}	0.0000	0.0000	0.0000	0.0000	0.0000
σ_1	0.3152	0.2954	0.3164	0.1895	0.2791
σ_2	0.2428	0.1982	0.2440	0.1858	0.1711
σ_3	0.1607	0.1632	0.2078	0.1419	0.1089
σ_4	0.1245	0.1133	0.0920	0.0842	0.0800
σ_5	0.0871	0.0830	0.0694	0.0747	0.0794
σ_6	0.0733	0.0649	0.0626	0.0508	0.0463
σ_7	0.0324	0.0281	0.0294	0.0368	0.0352
σ_8	0.0279	0.0247	0.0172	0.0270	0.0330
σ_9	0.0171	0.0239	0.0130	0.0267	0.0329
σ_{10}	0.0170	0.0237	0.0105	0.0241	0.0244
ρ	0.1828	0.4659	0.4817	0.1901	0.2155
$\mu_{1,10}$	8.2763	8.2452	8.2390	8.2591	8.2295
$\tilde{C}_{1,10}$	3917	3917	3917	3917	3917
$\tilde{\mu}_{2,10}$	7.8221	7.7942	7.8172	7.8074	7.7904
$\tilde{C}_{2,10}$	2520	2468	2480	2432	2453
$\tilde{\mu}_{3,10}$	8.2643	8.3278	8.2924	8.2862	8.2674
$\tilde{C}_{3,10}$	3893	4190	3939	4090	3802
$\tilde{\mu}_{4,10}$	8.3414	8.3345	8.3474	8.3679	8.3107
$\tilde{C}_{4,10}$	4229	4212	4233	4346	4075
$\tilde{\mu}_{5,10}$	8.2341	8.1219	8.1109	8.1873	8.1527
$\tilde{C}_{5,10}$	3761	3285	3269	3597	3676
$\tilde{\mu}_{6,10}$	8.1670	8.0192	8.0400	8.1728	8.0593
$\tilde{C}_{6,10}$	3450	3127	3120	3552	3196
$\tilde{\mu}_{7,10}$	8.6365	8.4910	8.4140	8.5819	8.4893
$\tilde{C}_{7,10}$	5488	4719	4441	5299	4765
$\tilde{\mu}_{8,10}$	8.2129	8.2340	8.2634	8.0739	8.1720
$\tilde{C}_{8,10}$	3652	3847	3933	3295	3708
$\tilde{\mu}_{9,10}$	8.3156	8.2106	8.2655	8.3794	8.2752
$\tilde{C}_{9,10}$	4112	3538	3949	4426	3914
$\tilde{\mu}_{10,10}$	7.4112	7.9853	8.7659	8.4271	8.8414
$\tilde{C}_{10,10}$	1613	3001	6511	4507	6763

Table 7. CCL, LCL, and Mack Models on Illustrative Insurer Incurred Data

w	CCL			LCL			Mack			Outcome
	$C_{w,10}$	SD	CV	$C_{w,10}$	SD	CV	$C_{w,10}$	SD	CV	$C_{w,10}$
1	3917	0	0.000	3917	0	0.000	3,917	0	0.000	3,917
2	2545	57	0.022	2544	59	0.023	2,538	0	0.000	2,532
3	4110	113	0.028	4110	106	0.026	4,167	3	0.001	4,279
4	4314	130	0.030	4307	122	0.028	4,367	37	0.009	4,341
5	3549	123	0.035	3545	115	0.032	3,597	34	0.010	3,587
6	3319	146	0.044	3317	132	0.040	3,236	40	0.012	3,268
7	5277	292	0.055	5315	265	0.050	5,358	146	0.027	5,684
8	3796	331	0.087	3775	301	0.080	3,765	225	0.060	4,128
9	4180	622	0.149	4203	561	0.134	4,013	412	0.103	4,144
10	4155	1471	0.354	4084	1157	0.283	3,955	878	0.222	4,181
Total	39161	1901	0.049	39116	1551	0.040	38,914	1,057	0.027	40,061
Percentile		73.72			76.38			86.03		

Table 7 gives the estimates of the mean and standard deviation, by accident year and in total, for the LCL, the CCL, and the Mack Models for the illustrative insurer. The predicted percentiles of the 40,061 outcome are also given for each model. Note that the standard deviations of the predicted outcomes were significantly higher for the CCL and the LCL models than they were for the Mack Model. This is generally the case, as can be seen in Figure 5. This figure plots the standard deviations (in the log scale) of the CCL and LCL models against those of the Mack Model for the 200 loss triangles listed in Appendix A. The higher standard deviations of the CCL model over the LCL model can be attributed to the generally positive correlation parameters that are shown in Figure 6 for the illustrative insurer. Generally this is the case for other insurers as can be seen in Figure 7.

The validation p - p plots for the LCL and CCL models run on the selected 200 triangles are given in Figures 8 and 9. For the LCL model:

- The p - p plots combined lines of insurance lie within the Kolmogorov–Smirnov bounds for Commercial Auto, Personal Auto and Workers Comp.
- All four lines have the slanted S pattern that characterizes models that are too thin in the tails. This pattern is reinforced in the combined plot, and the resulting plot does not lie within the Kolmogorov–Smirnov bounds. But the combined plot is an improvement over the corresponding Mack p - p plot.

For the CCL Model:

- The p - p plots for all four lines lie within the Kolmogorov–Smirnov bounds, but just barely so for the Other Liability line.
- While the combined p - p plot lies within the Kolmogorov–Smirnov bounds, the slanted S pattern indicates a mildly thin tail predicted by the model.

Figure 5. Compare Standard Deviation for CCL and LCL with Mack

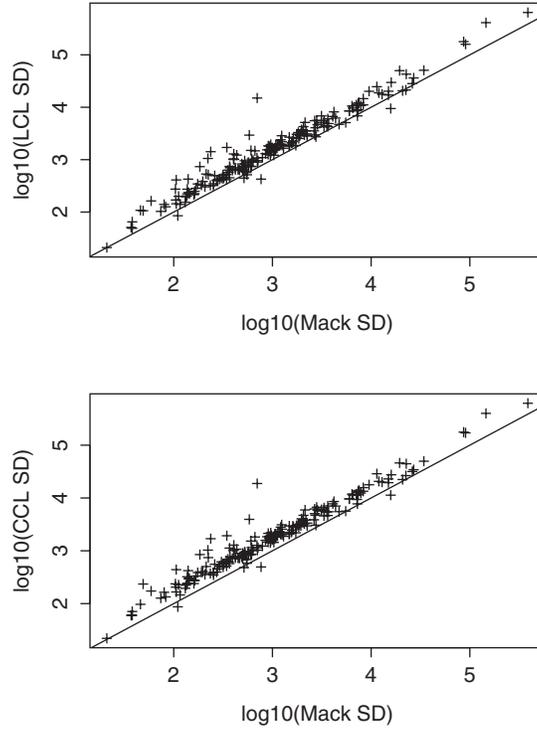


Figure 6. Posterior Distribution of ρ for the Illustrative Insurer

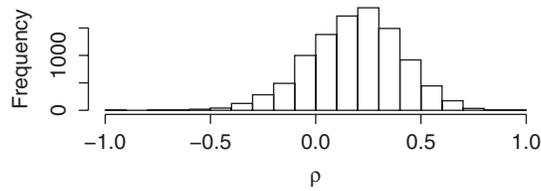


Figure 7. Posterior Mean of ρ for the 200 Incurred Loss Triangles

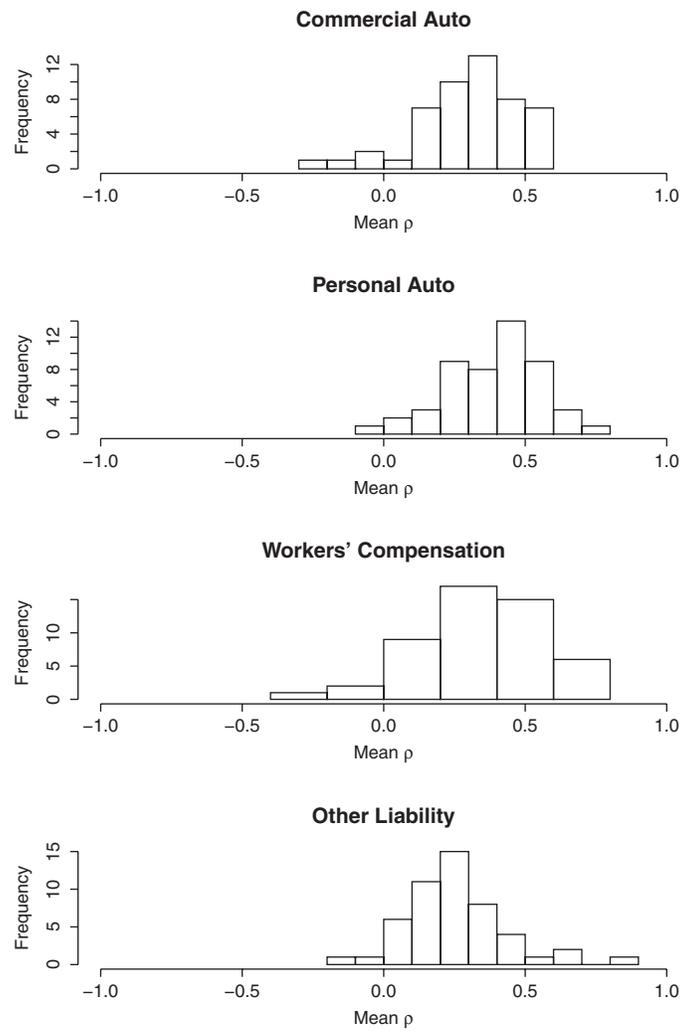


Figure 8. p - p Plots for the LCL Model on the Incurred Loss Triangles

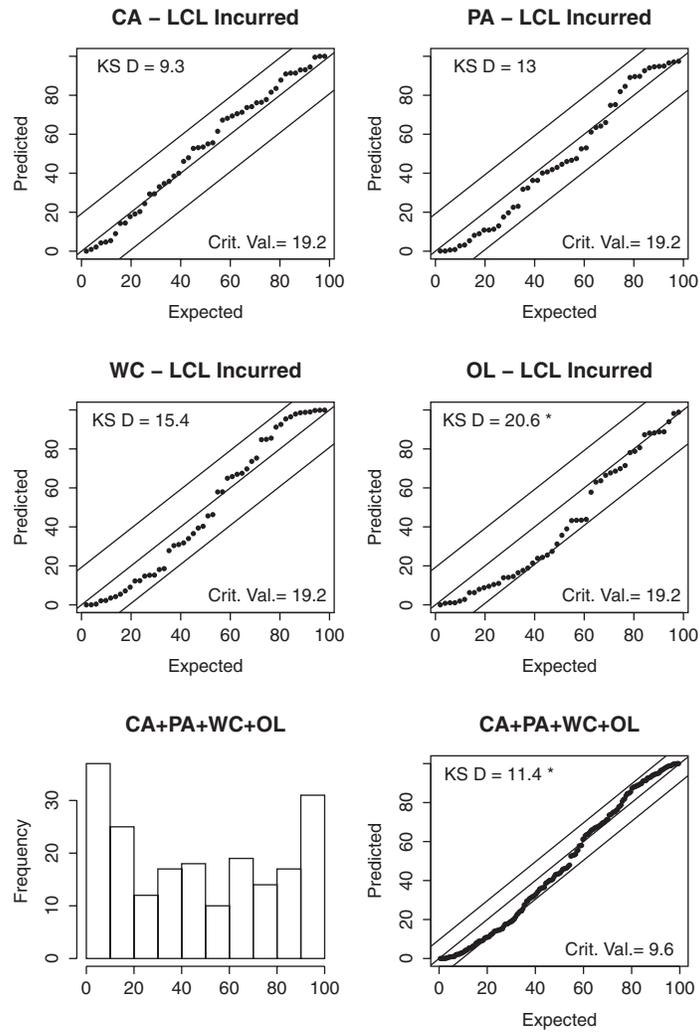
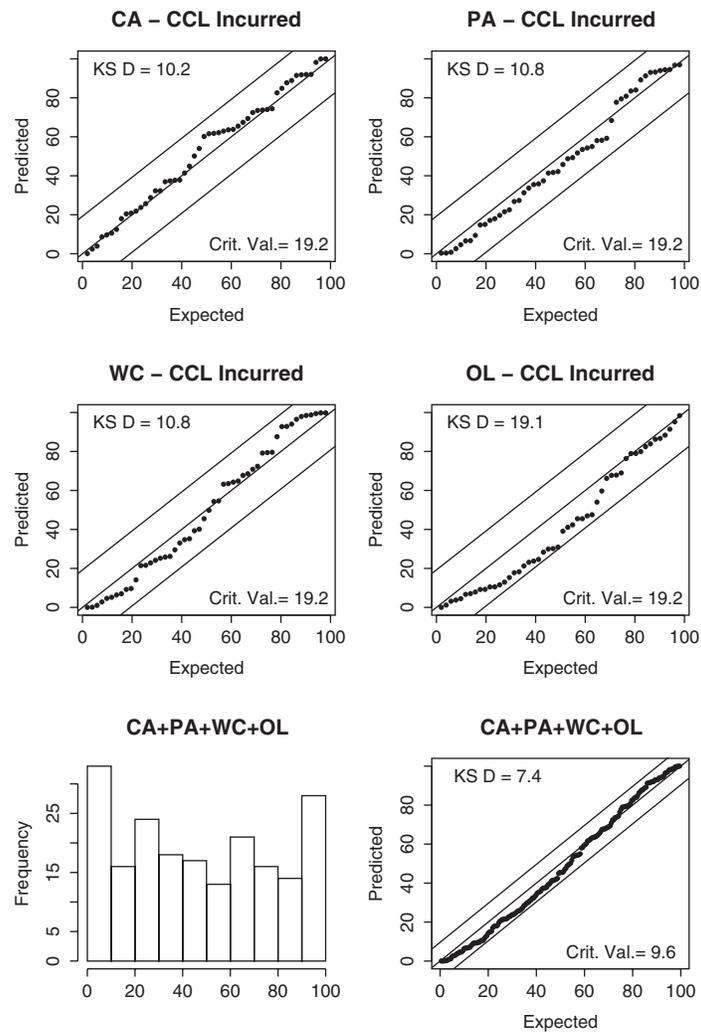


Figure 9. p - p Plots for the CCL Model on the Incurred Loss Triangles



6. Bayesian Models for Paid Loss Data

Given the improved validation of the CCL model on incurred loss data, it seems appropriate to try it out on paid loss data. Table 8 shows the CCL and ODP estimates. As should be expected given the results in Section 5, the standard deviation of the outcomes produced by the CCL model are noticeably higher than those produced by the ODP model.

The validation $p-p$ plots for the CCL model applied to paid data are in Figure 10. When comparing this plot with the validation $p-p$ plots for the ODP model (Figure 3) and the Mack model (Figure 4), we see that all three models show tend to produced estimates that are too high for these loss triangles.

Given the improved validation of the CCL model with incurred loss data, it is tempting to conclude that the incurred loss data contains crucial information that is not present in the paid loss data. However, there is also the possibility that a model other than the ODP or the CCL may be appropriate. A feature of such a model might be that it has a trend along the payment year ($= w + d - 1$). Models with a payment year trend have been proposed in the writings of Ben Zehnwirth over the years. See, for example, Barnett and Zehnwirth (2000). The inclusion of a payment year trend in a model has two important consequences.

1. The model should be based on incremental paid loss amounts rather than cumulative paid loss amounts. Cumulative losses include settled claims which do not change with time.
2. Incremental paid loss amounts tend to be skewed to the right and are occasionally negative. We need a loss distribution that allows for these features.

One distribution that has these properties is the skew normal distribution. This distribution is starting to be applied in actuarial settings. See, for example, Pigeon, Antonio and Denuit (2013) Here is a description of this distribution taken from Frühwirth-Schnatter and Pyne (2010). This distribution has three parameters.

1. μ —the location parameter.
2. ω —the scale parameter, with $\omega > 0$.
3. δ —the shape parameter, with $\delta \in (-1, 1)$.¹⁴

¹⁴ The reference calls the shape parameter α and then define $\delta = \alpha / \sqrt{1 + \alpha^2}$. The parameter designation, α , was already taken in this monograph.

Table 8. CCL and ODP Models on Illustrative Insurer Paid Data

w	CCL			ODP			Outcome $C_{w,10}$
	$\hat{C}_{w,10}$	SD	CV	$\hat{C}_{w,10}$	SD	CV	
1	3912	0	0	3912	0	0.0000	3912
2	2568	114	0.0444	2532	21	0.0083	2527
3	4157	199	0.0479	4163	51	0.0123	4274
4	4330	234	0.0540	4369	85	0.0195	4341
5	3574	212	0.0593	3554	96	0.0270	3583
6	3417	259	0.0758	3211	148	0.0461	3268
7	5235	465	0.0888	5161	240	0.0465	5684
8	3664	463	0.1264	3437	332	0.0966	4128
9	4444	870	0.1958	4220	572	0.1355	4144
10	5036	1961	0.3894	4635	1048	0.2261	4139
Total	40337	2692	0.0667	39193	1389	0.0354	40000
Percentile		49.18			73.91		

The skew normal distribution is defined as the sum of two random variables

$$X \sim \mu + \omega \cdot \delta \cdot Z + \omega \cdot \sqrt{1 - \delta^2} \cdot \varepsilon$$

where $Z \sim \text{truncated normal}_{(0,\infty)}(0,1)$ and $\varepsilon \sim \text{normal}(0,1)$. This distribution can also be expressed as a mixed truncated normal-normal distribution by setting

$$X \sim \text{normal}\left(\mu + \omega \cdot \delta \cdot Z, \omega \cdot \sqrt{1 - \delta^2}\right).$$

In looking at either expression for the skew normal distribution one can see that when $\delta = 0$, the skew normal becomes a normal distribution. As δ approaches one, the distribution becomes more skewed and becomes a truncated normal distribution when $\delta = 1$. Figure 11 plots¹⁵ the density functions for $\mu = 0$, $\omega = 15$ and δ close to one.¹⁶

It should be apparent that the coefficient of skewness can never exceed the coefficient of skewness of the truncated normal distribution, which is equal to 0.995. As it turns out, this constraint is important. I have fit models with the skew normal distribution that otherwise are similar to what will be described below and found that for most triangles, δ is very close to its theoretical limit. This suggests that a distribution with a higher coefficient of skewness is needed.

¹⁵ Using the R “sn” package.

¹⁶ The parameters in Figures 11 and 12 are representative of what one could expect in the later settlement lags where negative incremental losses frequently occur.

Figure 10. $p-p$ Plots for the CCL Model on Paid Loss Triangles

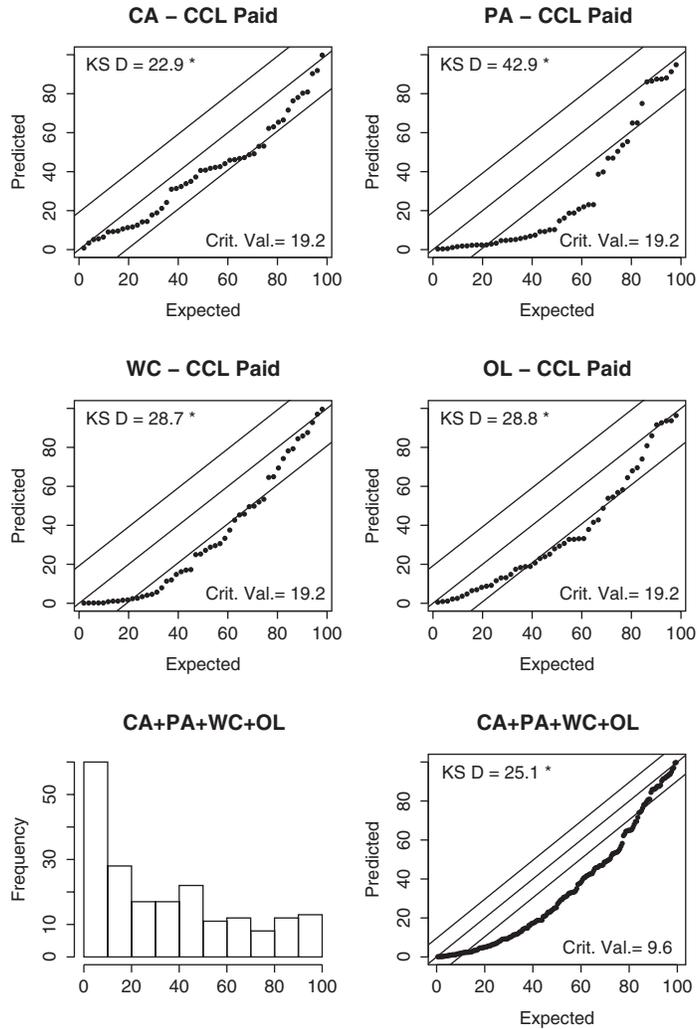


Figure 11. The Skew Normal Distribution

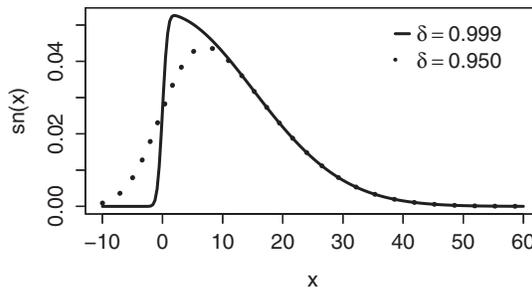
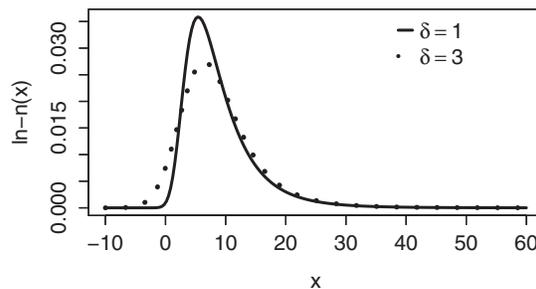


Figure 12. The Mixed Lognormal-Normal Distribution



The formulation of the skew normal distribution described by Frühwirth-Schnatter and Pyne (2010) suggests an alternative. Simply replace the truncated normal distribution with another skewed distribution, such as the lognormal distribution. Here is one way to do that. Define

$$X \sim \text{normal}(Z, \delta), \text{ where } Z \sim \text{lognormal}(\mu, \sigma).$$

Let's call this distribution the mixed lognormal-normal (ln-n) distribution with parameters given by δ , μ and σ . Figure 12 plots the density functions for $\mu = 2$, $\sigma = 0.6$, and two different values of δ .

Now that we have a loss distribution with the desired features of skewness and a domain that includes negative numbers, let's describe a model for incremental paid losses with a calendar-year trend.

The Correlated Incremental Trend (CIT) Model

Let:

1. $\mu_{w,d} = \alpha_w + \beta_d + \tau \cdot (w + d - 1)$.
2. $Z_{w,d} \sim \text{lognormal}(\mu_{w,d}, \sigma_d)$ subject to the constraint that $\sigma_1 < \sigma_2 < \dots < \sigma_{10}$.
3. $\tilde{I}_{1,d} \sim \text{normal}(Z_{1,d}, \delta)$.
4. $\tilde{I}_{w,d} \sim \text{normal}(Z_{w,d} + \rho \cdot (\tilde{I}_{w-1,d} - Z_{w-1,d}) \cdot e^\tau, \delta)$ for $w > 1$.

When comparing the CIT model with the CCL model (as it might be applied to incremental losses) there are some differences to note.

- The CCL model was applied to cumulative losses. One should expect σ_d to decrease as d increases as a greater proportion of claims are settled. In the CIT model, one should expect that the smaller less volatile claims to be settled earlier. Consequently, σ_d should increase as d increases.
- In the CCL model, the autocorrelation feature was applied to the logarithm of the cumulative losses. Since there is the possibility of negative incremental losses, it was necessary to apply the autocorrelation feature in Step 4 above after leaving the "log" space. The hierarchical feature of the mixed lognormal-normal distribution

provides the opportunity to do this. For a given set of parameters, ρ is the coefficient of correlation between $\tilde{I}_{w-1,d}$ and $\tilde{I}_{w,d}$.

- The trend factor, τ , is applied additively in the “log” space in Step 1 above. As the autocorrelation feature in Step 4 above is applied outside of the “log” space, it is necessary to trend the prior payment year’s difference by multiplying that difference by e^τ .

To prevent overdetermining the model, set $\beta_{10} = 0$. The parameters $\{\alpha_w\}, \{\sigma_d\}, \rho$, and the remaining $\{\beta_d\}$ are assigned prior distributions as follows:

1. Each $\alpha_w \sim \text{normal}(\log(\text{Premium}_w) + \text{logelr}, \sqrt{10})$ where $\text{logelr} \sim \text{uniform}(-5, 1)$.
2. Each $\beta_d \sim \text{uniform}(0, 10)$ for $d = 1$ to 4 and $\beta_d \sim \text{uniform}(0, \beta_{d-1})$ for $d > 4$. This assures that β_d decreases for $d > 4$.
3. $\rho \sim \text{uniform}(-1, 1)$ —The full permissible range for ρ .
4. $\tau \sim \text{normal}(0, 0.0316)$ —corresponding to a precision parameter used by JAGS of 1000.
5. $\sigma_1^2 \sim \text{uniform}(0, 0.5)$, $\sigma_d^2 \sim \text{uniform}(\sigma_{d-1}^2, \sigma_{d-1}^2 + 0.1)$.
6. $\delta \sim \text{uniform}(0, \text{Average Premium})$

There are two deviations from the selection of diffuse prior distributions that are in the CCL model.

- I first tried a wider prior for τ . In examining the MCMC output I noticed that quite often, the value of τ was less than -0.1 , which I took to be unreasonably low. This low value was usually compensated for by offsetting high values for the α and/or β parameters. This could have a noticeable effect on the final result, so I decided to restrict the volatility of τ to what I considered to be a reasonable range of payment year changes.
- In examining the MCMC output, I noticed that, occasionally, high values of σ_d would occur. This led to unreasonably high simulated losses in the output, so I decided to limit how fast σ_d could increase with d .

The predictive distributions of the sum, $\sum_{d=1}^{10} \tilde{I}_{w,d}$ for each w , and the overall sum, $\sum_{w=1}^{10} \sum_{d=1}^{10} \tilde{I}_{w,d}$ are simulated 10,000 times with a Bayesian MCMC model. The details are very similar to those described in Section 5 and will not be given here.

By setting the prior distribution of ρ equal to zero, we eliminate the between accident year correlation. Following the naming convention of the last section, let’s call this model the Leveled Incremental Trend (LIT) model.

Table 9 shows the estimates of for the illustrative insurer with the CIT and the LIT model on paid data.

Before producing these distributions, I had no particular expectation of how ρ would be distributed for paid data. However, I did expect τ to be predominantly negative since the p - p plots in Figures 3, 4 and 10 indicated that the all the other models predicted results that were too high.

Let’s first examine the effects of between-year correlation in the CIT model. Figure 13 gives the posterior distributions for ρ for the illustrative insurer. Figure 14 gives the histograms of the posterior means ρ for each insurer by line of business.

Table 9. CIT and LIT Models on Illustrative Insurer Paid Data

w	CIT			LIT			Outcome $C_{w,10}$
	$\hat{C}_{w,10}$	SD	CV	$\hat{C}_{w,10}$	SD	CV	
2	2539	9	0.0035	2538	9	0.0035	2527
3	4183	21	0.0050	4185	20	0.0048	4274
4	4395	40	0.0091	4393	32	0.0073	4341
5	3553	42	0.0118	3566	32	0.0090	3583
6	3063	101	0.0330	3151	40	0.0127	3268
7	5062	123	0.0243	5065	111	0.0219	5684
8	3512	514	0.1464	3355	234	0.0697	4128
9	4025	707	0.1757	4138	594	0.1435	4144
10	4698	1482	0.3155	4703	1489	0.3166	4139
Total	38942	1803	0.0463	39006	1723	0.0442	40000
Percentile		79.04			79.69		

As seen in Figure 14, the posterior means of ρ for the paid data were not as overwhelmingly positive as we saw in the incurred data shown in Figure 7. Figure 15 shows a small but noticeable difference between the standard deviations of the CIT and LIT models.

My efforts to rein in the correlation between the $\{\alpha_w\}$, the $\{\beta_d\}$, and the τ parameters were, at best, only partially successful, as Figure 16 indicates. The analogous plot for the LIT model is very similar. With the given data, it is hard for the CIT and the LIT models to sort out the effects of the level plus the development and the trend.

As seen in Figure 17, the posterior means of τ were predominantly negative. But as pointed out above, a negative might be offset by higher $\{\alpha_w\}$ s and $\{\beta_d\}$ s. Figure 18 shows only a handful of triangles where there was a noticeable decrease in the final expected loss estimates. And most of those differences appeared in the Other Liability line of business.

Figures 19 and 20 show the validation $p-p$ plots for the CIT and the LIT models. As do the Mack, ODP and CCL models on paid data indicate, the predictive distributions for the CIT and LIT models tend to overstate the estimates of the expected loss.

Figure 13. Posterior Distribution of ρ and for the Illustrative Insurer

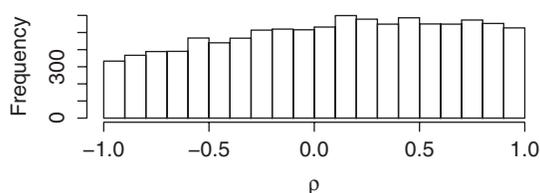


Figure 14. Posterior Mean of ρ by Line and Insurer for Paid Loss Data

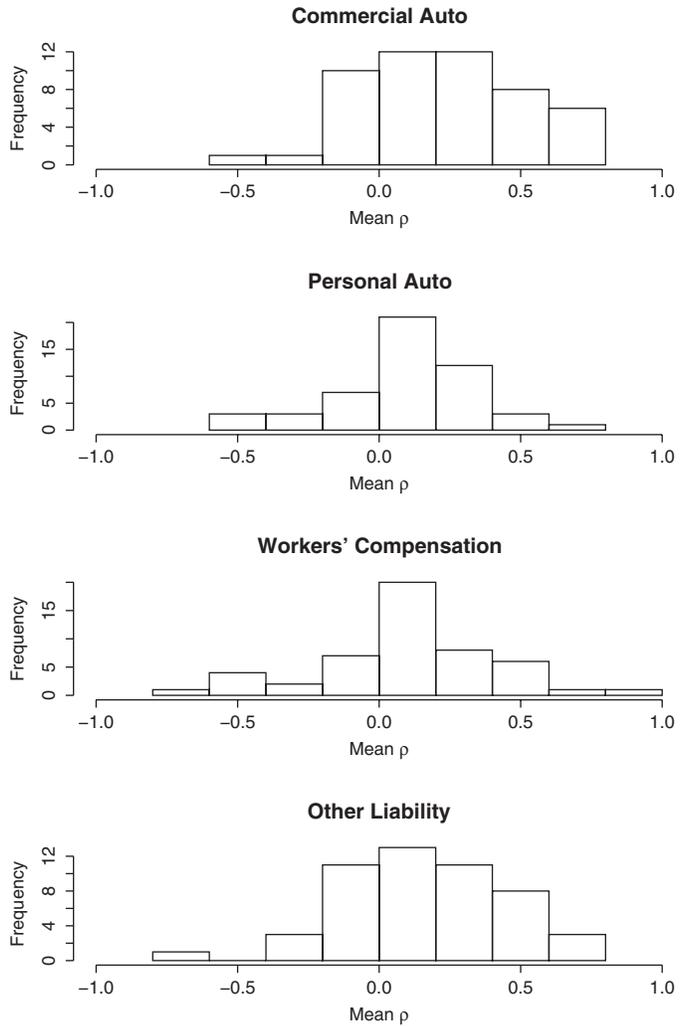


Figure 15. Compare the Standard Deviations of CIT and LIT Models for Paid Loss Data

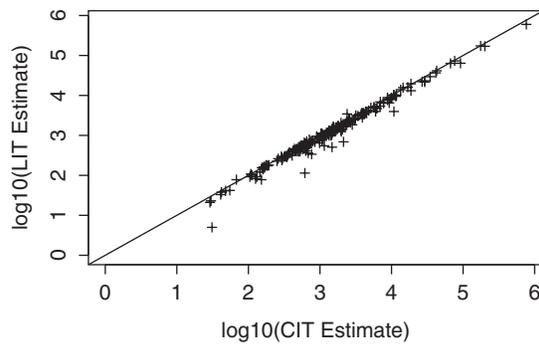


Figure 16. Correlations Between Parameters in the CIT Model for the Illustrative Insurer

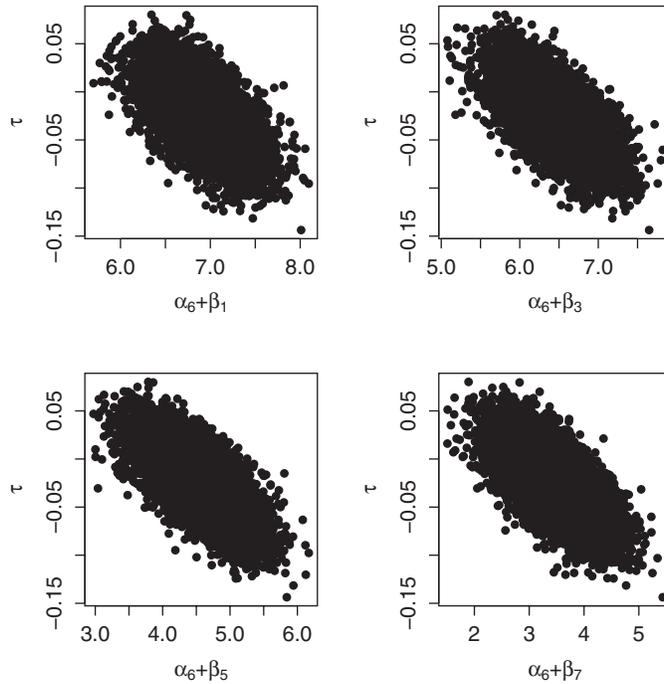


Figure 17. Posterior Mean of τ by Line and Insurer for Paid Loss Data

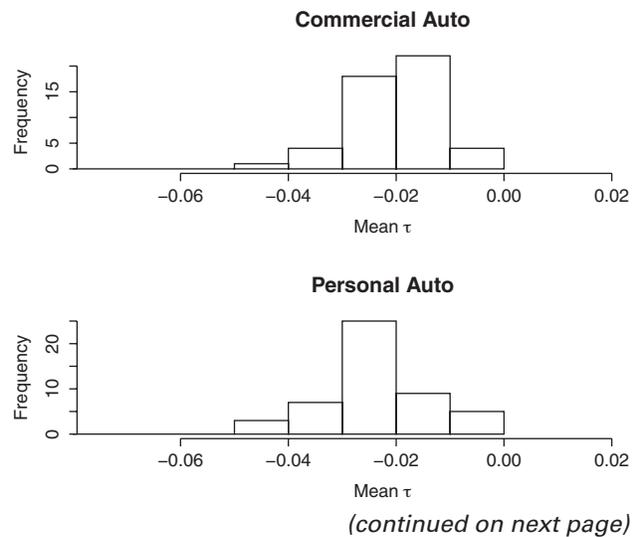


Figure 17. Posterior Mean of τ by Line and Insurer for Paid Loss Data (continued)

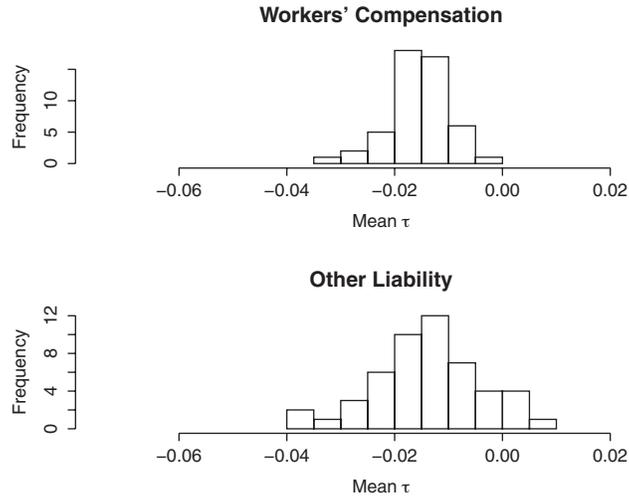


Figure 18. Comparing Estimates for the CCL and the CIT Models for Paid Data

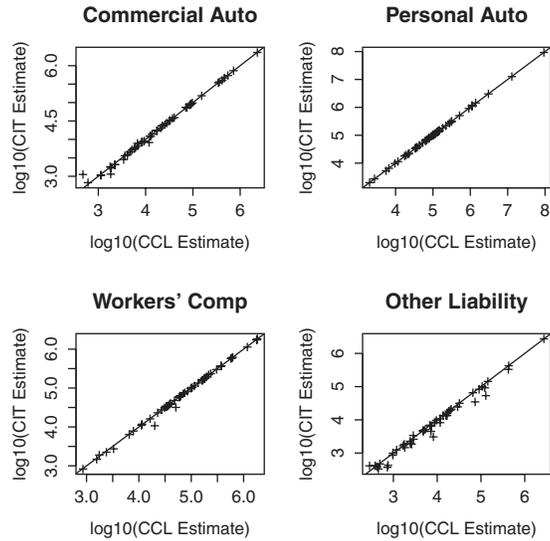


Figure 19. p - p Plots for the CIT Model

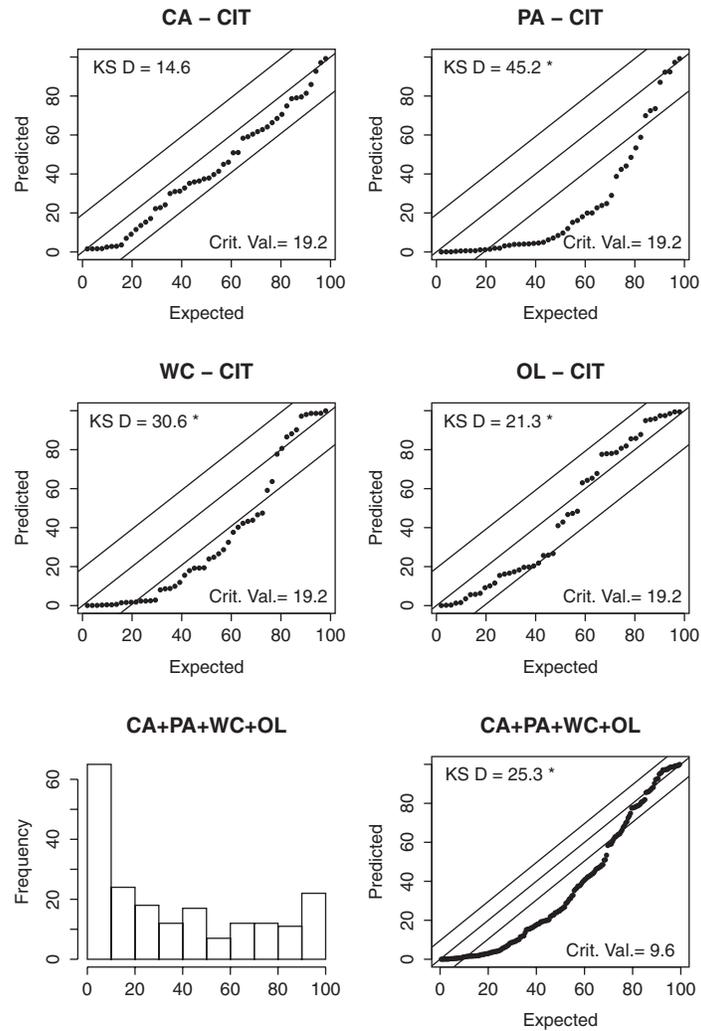
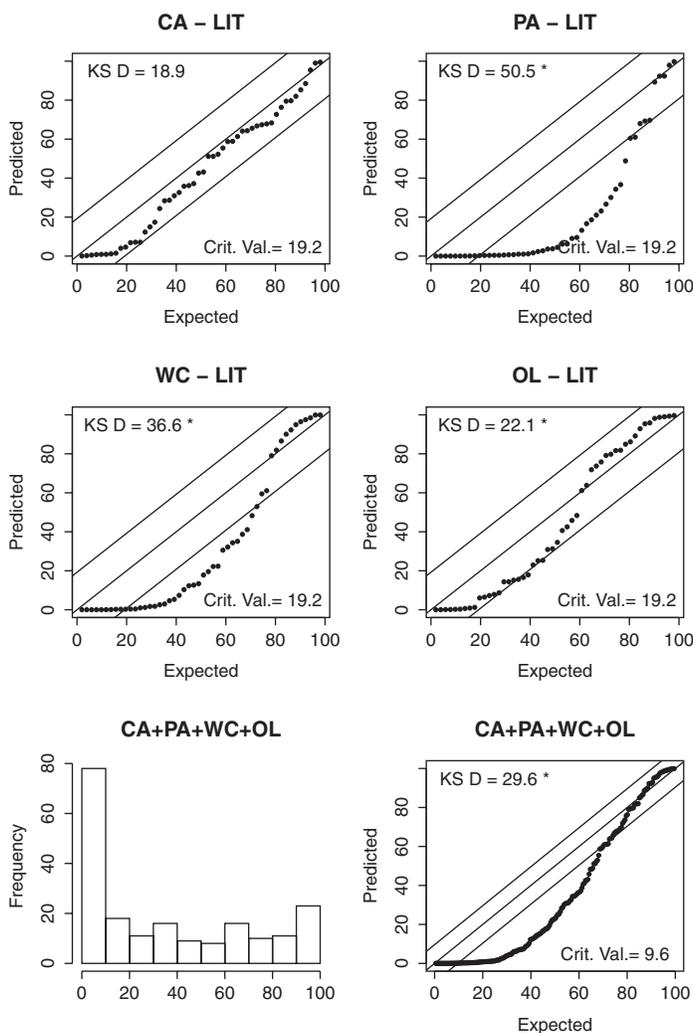


Figure 20. p - p Plots for the LIT Model



So, in spite of a serious attempt to improve on the results produced by the earlier models on paid data, the CIT and LIT models did not achieve the desired improvement. This result tends to support the idea that is generally accepted, that the incurred data reflects real information that is not in the paid data.

A reviewer of this monograph checked with some colleagues and found that claims are “reported and settled faster today due to technology,” and suggested that the CIT model might not fully reflect this change. A model that addresses the possibility of a speedup of claim settlement is the following.

The Changing Settlement Rate (CSR) Model

Let:

1. Each $\alpha_w \sim \text{normal}(\log(\text{Premium}_w) + \text{logelr}, \sqrt{10})$ where the parameter $\text{logelr} \sim \text{uniform}(-1, 0.5)$.
2. $\beta_d \sim \text{uniform}(-5, 5)$ for $d = 1, \dots, 9$, $\beta_{10} = 0$.

3. $\mu_{w,d} = \alpha_w + \beta_d \cdot (1 - \gamma)^{(w-1)}$ $\gamma \sim \text{normal}(0, 0.025)$.
4. Each $\sigma_d = \sum_{i=d}^{10} a_i$ where $a_i \sim \text{uniform}(0, 1)$.
5. $\tilde{C}_{w,d}$ has a lognormal distribution with log mean $\mu_{w,d}$ and log standard deviation σ_d subject to the constraint that $\sigma_1 > \sigma_2 > \dots > \sigma_{10}$.

Since $\beta_{10} = 0$ and cumulative paid losses generally increase with the development year, d , β_d for $d < 10$ is usually negative. Then for each $d < 10$, a positive value of γ will cause $\beta_d \cdot (1 - \gamma)^{(w-1)}$ to increase with w and thus indicate a speedup in claim settlement. Similarly, a negative value of γ will indicate a slowdown in claim settlement.

Table 10 shows the results for the CSR model on the illustrative insurer.

Figure 21 shows that the posterior distribution of γ is predominantly positive. This confirms the reviewer’s contention that the claim settlement rate is, in general, increasing.

The validation $p-p$ plots in Figure 22 shows that for three of the four lines of insurance, the CSR model corrects the bias found in the earlier models. This model also correctly predicts the spread of the predicted percentile of the outcomes for those lines. While the CSR model still exhibits bias for the personal auto line of business, the bias is significantly smaller than that of the other models.

It appears that the incurred loss data recognized the speedup in claim settlements.

Table 10. CIT and CSR Models on Illustrative Insurer Paid Data

w	CIT			CSR			Outcome $C_{w,10}$
	$\hat{C}_{w,10}$	SD	CV	$\hat{C}_{w,10}$	SD	CV	
1	3912	0	0	3912	0	0	3912
2	2539	9	0.0035	2559	103	0.0403	2527
3	4183	21	0.0050	4135	173	0.0418	4274
4	4395	40	0.0091	4285	198	0.0462	4341
5	3553	42	0.0118	3513	180	0.0512	3583
6	3063	101	0.0330	3317	216	0.0651	3268
7	5062	123	0.0243	4967	404	0.0813	5684
8	3512	514	0.1464	3314	402	0.1213	4128
9	4025	707	0.1757	3750	734	0.1957	4144
10	4698	1482	0.3155	3753	1363	0.3632	4139
Total	38942	1803	0.0463	37506	2247	0.0599	40000
Percentile		79.04			87.62		

Figure 21. Posterior Mean of γ by Line and Insurer for Paid Loss Data

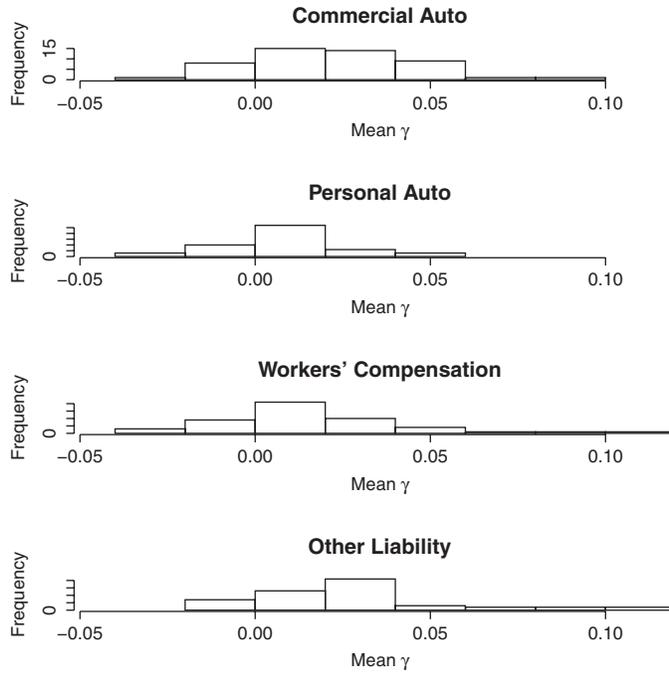
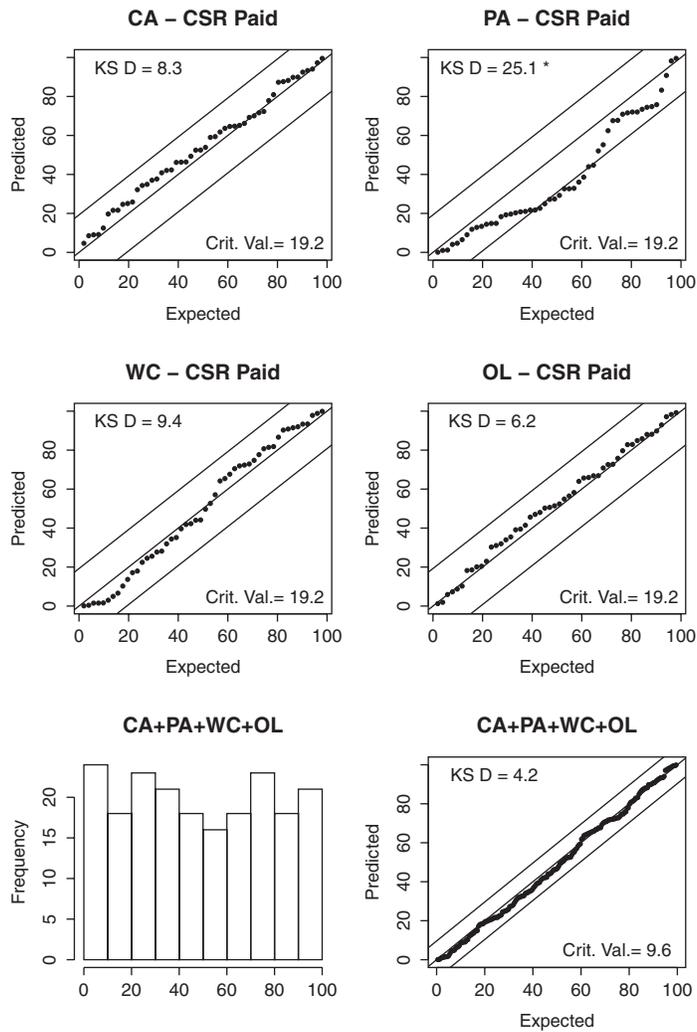


Figure 22. $p-p$ Plots for the CSR Model



7. Process Risk, Parameter Risk and Model Risk

Let us now address a topic that frequently comes up in stochastic modeling discussions – process risk, parameter risk and model risk. One way to describe process and parameter risk is to consider the relationship for a random variable X conditioned on a parameter θ .

$$\text{Var}[X] = E_{\theta}[\text{Var}[X|\theta]] + \text{Var}_{\theta}[E[X|\theta]].$$

Let's call the left side of the above equation the "Total Risk." Let's call the first term of the right side the "Process Risk" as it represents the average variance of the outcomes from the expected result. Finally, let's call the second term the "Parameter Risk" as it represents the variance due to the many possible parameters in the posterior distribution. Another often-used term that overlaps with parameter risk is the "range of reasonable estimates."

For the CCL model, the parameter θ is represented by the vector

$$(\alpha_1, \dots, \alpha_{10}, \beta_1, \dots, \beta_9, \sigma_1, \dots, \sigma_{10}, \rho).$$

The MCMC sample simulates 10,000 parameters denoted by θ_i . We then have the illustrative insurer:

$$\text{Total Risk} = \text{Var} \left[\sum_{w=1}^{10} \tilde{C}_{w,10} \right] = 1901^2.$$

The random variables $\mu_{w,10}$ are derived from the posterior distribution of the α_w . One can then use the formula for the mean of a lognormal distribution to calculate:

$$\text{Parameter Risk} = \text{Var}_{\theta} \left[E \left[\sum_{w=1}^{10} \tilde{C}_{w,10} | \theta \right] \right] = \text{Var} \left[\sum_{w=1}^{10} e^{\mu_{w,10} + \frac{\sigma_{10}^2}{2}} \right] = 1893^2.$$

For this example, the parameter risk is very close to the total risk, and hence there is minimal process risk. I have repeated this calculation on several (including some very large) insurers and I obtained the same result that process risk is minimal.

Model risk is the risk that one did not select the right model. If the possible models fall into the class of "known unknowns" one can view model risk as parameter risk. It is possible to formulate a model as a weighted average of the candidate models, with the weights as parameters. If the posterior distribution of the weights assigned to each

model has significant variability, this is an indication of model risk. Viewed in this light, model risk is a special case of parameter risk.

As a thought experiment, one can consider what happens if we were to run this model on a very large dataset. The parameter risk will shrink towards zero and any remaining risk, such as model risk, will be interpreted as process risk.

This thought experiment is of largely academic interest since all aggregated loss triangles one finds in practice are small datasets. But it does serve to illustrate some of the theoretical difficulties that occur when one tries to work with the parameter/process/model classifications of risk. My own preference is to focus on total risk, as that it is the only risk that we can test by looking at actual outcomes.

8. Summary and Conclusions

The central thrust of this monograph is twofold.

- It implements the idea of large-scale retrospective testing of stochastic loss reserve models on real data. The goal is not to comment on the reserves of individual insurers. Instead the goal is to test the predictive accuracy of specific models.
- As shortcomings in existing models are identified, it demonstrates that Bayesian MCMC models can be developed to overcome some of these shortcomings.

The principle behind the retrospective testing is that a specific model is built with data that we customarily observe. The model is used to predict a distribution of outcomes that we will observe in the future. When we do observe outcomes for a large number of predictions, we expect the percentiles of the outcomes to be uniformly distributed. If they are not uniformly distributed, we look for a better model. We may or may not find one.

The data used in this study comes from the CAS Loss Reserve Database. It consists of hundreds of paid and incurred loss triangles that Peng Shi and I obtained from a proprietary database maintained by the NAIC. We are grateful that the NAIC allowed us to make these data available to the public. The data I used to build the models came from the 1997 NAIC Annual Statements. The outcomes came from subsequent statements.

Here is a high-level summary of the results obtained with these data.

- For incurred data, the variability predicted by Mack model is understated. One of its key assumptions is that the losses from different accident years are independent. This monograph proposes the correlated chain ladder (CCL) model as an alternative. This model allows for a particular form of dependency between accident years. It finds that the CCL model predicts the distribution of outcome correctly within a specified confidence level.
- For paid data, the bootstrap ODP model, the Mack model and the CCL model tend to give estimates of the expected ultimate loss that are high. This suggests that there is a change in the loss environment that is not being captured in these models. This monograph proposes three models, the Leveled Incremental Trend (LIT), the Correlated Incremental Trend (CIT) model, and the Changing Settlement Rate (CSR) as alternatives. The first two models allow for payment year trends. While the introduction of a payment year trend seems plausible given the bias identified in the earlier models, the performance of the LIT and CIT models are similar to the earlier models in the validation $p-p$ plots. The CSR model corrects the bias identified in the previous models for three of the four lines of insurance, and has significantly less bias on the fourth line of insurance.

- Note that for the “Other Liability” line of insurance, the Mack and ODP models validate better than any of the new models proposed in this monograph. While it might be a small sample problem, the sample is not all that small. This suggests that more study is needed. Note that these results are for a specific annual statement year—1997. Studies such as this should be repeated on other annual statement years to see if the above conclusions still hold.

In preparing this monograph I have made every effort to adhere to the “open source” philosophy. The data is publicly available. The software is publicly available for free. The R and JAGS scripts used in creating these models are to be made publicly available. I have purposely restricted my methods to widely used software (R, JAGS and RStudio) in order to make it easy for others to duplicate and improve on these results.

In building the Bayesian models I used prior distributions that were as diffuse as I could make them. The restrictions I did make (for example, the restriction that $\sigma_1 > \sigma_2 > \dots > \sigma_{10}$ in the CCL model) reflected my experience over several years of general model building. I did not have intimate knowledge of each insurer’s business operations. Those with knowledge of an insurer’s business operation should be able to incorporate this knowledge to obtain better results. As all probabilities are conditional, the Bayesian methodology allows for one to incorporate additional information by adjusting the prior distributions. I made every effort to code the models transparently so that such adjustments are easy to make.

The models proposed in this monograph are offered as demonstrated improvements over current models. I expect to see further improvements over time. The Bayesian MCMC methodology offers a flexible framework with which one can make these improvements.

References

- Bardis, Emmanuel Theodore, Ali Majidi, and Daniel M. Murphy. 2012. "A Family of Chain-Ladder Factor Models for Selected Link Ratios." *Variance* 6:143–160.
- Barnett, Glen, and Ben Zehnwirth. 2000. "Best Estimates for Reserves." *Proceedings of the Casualty Actuarial Society* 87:245–321.
- Berger, James O. 1985. *Statistical Decision Theory and Bayesian Analysis*. 2nd ed. New York: Springer-Verlag.
- Bornhuetter, Ronald L., and Ronald E. Ferguson. 1972. "The Actuary and IBNR." *Proceedings of the Casualty Actuarial Society* 59:181–195.
- Brooks, Steve, Andrew Gelman, Galin L. Jones, and Xiao-Li Meng, eds. 2011. *Handbook of Markov Chain Monte Carlo*. London: Chapman and Hall/CRC.
- De Alba, Enrique. 2002. "Bayesian Estimation of Outstanding Claims Reserves." *North American Actuarial Journal* 6(4):1–20.
- England, P. D., and R. J. Verrall. 2002. "Stochastic Claims Reserving in General Insurance." Meeting paper. Presented to the Institute and Faculty of Actuaries, 28 January. <http://www.actuaries.org.uk/research-and-resources/documents/stochastic-claims-reserving-general-insurance>.
- Gelfand, A. E., and A. F. M. Smith. 1990. "Sampling-Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association* 85(410):398–409.
- Frühwirth-Schnatter, Sylvia, and Saumyadipta Pyne. 2010. "Bayesian Inference for Mixtures of Univariate and Multivariate Skew-Normal and Skew-t Distributions." *Biostat* 11(2):317–336.
- Hartman, Brian. 2014. "Bayesian Computational Methods." In *Predictive Modeling Applications in Actuarial Science*, vol. 1, *Predictive Modeling Techniques*, eds. Edward W. Frees, Richard A. Derrig and Glenn Meyers, 315–333. New York: Cambridge University Press.
- Hastings, W. K. 1970. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications." *Biometrika* 57(1):97–109.
- Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences*. Hoboken, N. J.: Wiley.
- Klugman, Stuart A., Harry H. Panjer, and Gordon E. Willmot. 2012. *Loss Models, From Data to Decisions*. 4th ed., Hoboken, N. J.: Wiley.
- Mack, Thomas. 1993. "Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates." *ASTIN Bulletin* 23(2):213–225.
- Mack, Thomas. 1994. "Measuring the Variability of Chain Ladder Reserve Estimates." *Casualty Actuarial Society Forum* (Spring):101–182.

- McGrayne, Sharon Bertsch. 2011. *The Theory That Would Not Die*. New Haven, Conn.: Yale University Press.
- Metropolis, N., and S. Ulam. "The Monte Carlo Method." *Journal of the American Statistical Association* 44(247):335–341.
- Meyers, Glenn G. 2009. "Pure Premium Regression with the Tweedie Model." *The Actuarial Review* 36(2):12.
- Meyers, Glenn G., and Peng Shi. 2011. "The Retrospective Testing of Stochastic Loss Reserve Models." *Casualty Actuarial Society E-Forum* (Summer):1–37.
- Meyers, Glenn G. 2012. "The Leveled Chain Ladder Model for Stochastic Loss Reserving." *Casualty Actuarial Society E-Forum* (Summer):1–33.
- Ntzoufras, I., and P. Dellaportas. 2002. "Bayesian Modelling of Outstanding Liabilities Incorporating Claim Count Uncertainty." *North American Actuarial Journal* 6(1): 113–128.
- Pigeon, Mathiew, Katrien Antonio, and Michel Denuit. 2013. "Individual Loss Reserving with the Multivariate Skew Normal Framework." *ASTIN Bulletin* 43(3): 399–428.
- Scollnik, D.P.M. 2001. "Actuarial Modeling with MCMC and BUGS." *North American Actuarial Journal* 5(2):96–124.
- Taleb, N. N. 2010. *The Black Swan: The Impact of the Highly Improbable*. 2nd ed. New York: Random House.
- Taylor, G. C. 1986. *Claims Reserving In Non-Life Insurance*. New York: Elsevier.
- Verrall, Richard. 2007. "Obtaining Predictive Distributions for Reserves Which Incorporate Expert Opinion." *Variance* 1:53–80.

Appendix A. The Data Selection Process

When selecting the loss triangles to use in this monograph my overriding consideration was that the process should be mechanical and well defined. There are two potential mistakes one can make in selecting the insurers to analyze.

- If one were to take all the insurers in the database, or randomly select the insurers, there could be some insurers who made significant changes in their business operations that could violate the assumptions underlying the models.
- If one is too selective, one runs the risk of selecting only those data that best fit a chosen model. For example, let's suppose that I wanted the CCL model to fit the incurred data even better than it does. As an extreme case, noting that CCL model still appears to be a bit light in the tails, I could have replaced some of the insurers that have outcomes in the tail with other insurers that have outcomes in the middle.

While I did not have inside information on any changes in the business operations, Schedule P provides some hints in their reporting of both net and direct earned premium by accident year. Both of these data elements are in the CAS Loss Reserve Database.

- If an insurer makes significant changes in its volume of business over the 10-year period covered by Schedule P, a change in business operation could be inferred.
- If an insurer makes significant changes in its net to direct premium ratio over the 10-year period, a change in its reinsurance strategy could be inferred.

To carry out an analysis of this sort, I needed a large number of insurers. After looking at the quality and consistency of the data available in the CAS Loss Reserve Database, I decided to use 50 insurers in each of four major lines of insurance—Commercial Auto, Personal Auto, Workers Compensation, and Other Liability. Early on I concluded that there were an insufficient number of insurers in the Products Liability and the Medical Malpractices lines to obtain an adequately sized selection.

To implement these considerations, I calculated the coefficients of variation for the net earned premiums and the net to direct premium ratios over the ten available years. By trial and error, I then set up limits on these coefficients (CV) of variation that obtained the desired number of insurers. This procedure should have eliminated some of the insurers that changed their business operations.

After some provisional testing, I eliminated insurer group 38997 from the Personal Auto and Workers Comp lines, and insurer groups 16373, 44598 and 14885 from the

Other Liability line because the R “ChainLadder” package produced “NA” results for the Mack calculation of the standard deviation. I also eliminated insurer group 14451 from the Other Liability line because the MCMC algorithm took very long to converge for paid losses. After eliminating these insurer groups I adjusted the CV limits to give 50 insurers for each line. The final CV limits are given in Table 11. The final list of the selected insurer groups are in Table 12.

Table 11. CV Limits for Insurer Triangles

	Commercial Auto	Personal Auto	Workers’ Comp	Other Liability
CV(Premium)	<0.399	<0.450	<0.950	<0.390
CV(Net/Direct)	<0.125	<0.125	<0.200	<0.125

Table 12. Group Codes for Selected Insurers

	Commercial Auto		Personal Auto		Workers’ Comp		Other Liability	
353	13420	353	13641	86	13528	620	14370	
388	13439	388	13889	337	14176	669	14915	
620	13641	620	14044	353	14320	671	15113	
833	13889	692	14176	388	14508	683	15148	
1066	14044	715	14257	671	14974	715	15210	
1090	14176	1066	14311	715	15148	833	15571	
1538	14257	1090	14443	1252	15199	1538	17043	
1767	14320	1538	15199	1538	15334	1767	17450	
2135	14974	1767	15393	1767	18309	2003	17493	
2208	15199	2003	15660	2135	18538	2135	18163	
2623	18163	2143	16373	2712	18767	2143	18686	
2712	18767	3240	16799	3034	18791	2208	24830	
3240	19020	4839	18163	3240	21172	3240	26797	
3492	21270	5185	18791	5185	23108	5185	27065	
4839	25275	6947	23574	6408	26433	5320	28550	
5185	27022	7080	25275	6807	27529	6459	30449	
6408	27065	8427	27022	7080	30589	6947	30651	
6459	29440	8559	27065	8559	32875	7625	32301	
6947	31550	10022	27499	8672	33499	10657	33049	
7080	32301	13420	27766	9466	34576	13501	36315	
8427	34606	13439	29440	10385	35408	13919	38733	
10022	35483	13501	31550	10699	37370	13994	41068	
10308	37036	13528	34509	11347	38687	14044	41580	
11037	38733	13587	34592	11703	38733	14176	42439	
11118	44598	13595	34606	13439	41300	14257	43354	

Appendix B. Introduction to Bayesian MCMC Models

Since the recognition of Markov Chain Monte Carlo as a powerful tool for doing Bayesian analyses in 1990, there have been many efforts to create software to aid in these analyses. Progress in making the available software faster and more user friendly is still being made. In spite of this progress, I believe that it is necessary for an actuary to have a picture of what is happening inside the black box. The purpose of this appendix is to provide a brief description of what is inside the black box.

A Markov chain is a random process where the transition to the next state depends only on its current state, and not on prior states. Formally, a Markov chain, X_t , for $t = 1, 2, \dots$ is a sequence of vectors satisfying the property that

$$\Pr(X_{t+1} = x | X_1 = x_1, X_2 = x_2, \dots, X_t = x_t) = \Pr(X_{t+1} = x | X_t = x_t).$$

The properties of Markov chains have been well studied by scholars. Those interested in these studies can start with Chapter 4 of Jackman (2009). What actuaries need to know about Markov chains in Bayesian MCMC analyses can be summarized as follows.

- There is a certain class of Markov chains, generally called “ergodic,” for which the vectors, $\{X_t\}$, approaches a limiting distribution. That is to say that as T increases, the distribution of $\{X_t\}$ for all $t > T$ approaches a unique limiting distribution.
- The Markov chains used in Bayesian MCMC analyses, such as the Metropolis Hastings algorithm, are members of this class.
- Let x be a vector of observations and let y be a vector of parameters in a model. In Bayesian MCMC analyses, the Markov chain is defined in terms of the prior distribution, $p(y)$, and the conditional distribution, $f(x|y)$. The limiting distribution is the posterior distribution, $f(y|x)$. That is to say, if we let the chain run long enough, the chain will randomly visit all states with a frequency that is proportional to their posterior probabilities.

The operative phrase in the above is “long enough.” In practice we want to: (1) develop an algorithm for obtaining a chain that is “long enough” as quickly as possible; and (2) develop criteria for being “long enough.”

Here is how Bayesian MCMC analyses work in practice.

1. The user specifies the prior distribution, $p(y)$, and the conditional distribution, $f(x|y)$.

2. The user selects a starting vector, x_1 , and then, using a computer simulation, runs the Markov chain through a sufficiently large number, t_1 , of iterations. This first phase of the simulation is called the “adaptive” phase, where the algorithm is automatically modified to increase its efficiency.
3. The user then runs an additional t_2 iterations. This phase is called the “burn-in” phase. t_2 is selected to be high enough so that a sample taken from subsequent t_3 periods represents the posterior distribution.
4. The user then runs an additional t_3 iterations and then takes a sample, $\{x_i\}$, from the $(t_2 + 1)^{th}$ step to the $(t_2 + t_3)^{th}$ step to represent the posterior distribution $f(y|x)$.
5. From the sample, one then constructs various “statistics of interest” that are relevant to the problem addressed by the analysis.

The most common algorithms for generating Bayesian Markov chains are variants of the Metropolis-Hastings algorithm.

Given a prior distribution, $p(y)$, and a conditional distribution, $f(x|y)$, the Metropolis-Hastings algorithm introduces a third distribution, $J(y_t|y_{t-1})$, called the “proposal” or “jumping” distribution. Given a parameter vector, y_{t-1} , the algorithm generates a Markov chain by the following steps.

1. Select a candidate value, y^* , at random from the proposal distribution, $J(y_t|y_{t-1})$.
2. Compute the ratio

$$R \equiv R_1 \times R_2 = \frac{f(x|y^*) \cdot p(y^*)}{f(x|y_{t-1}) \cdot p(y_{t-1})} \times \frac{J(y_{t-1}|y^*)}{J(y^*|y_{t-1})}$$

3. Select U at random from a uniform(0,1) distribution.
4. If $U < R$ then set $y_t = y^*$. Otherwise set $y_t = y_{t-1}$.

The first part of the ratio, R_1 , represents the ratio of the posterior probability of the proposal, y^* , to the posterior probability of y_{t-1} . The higher the value of R_1 , the more likely will be accepted into the chain. Regardless of how the proposal density distribution is chosen, the distribution of y_t can be regarded as a sample from the posterior distribution, after a suitable burn-in period.

To see the issues that can arise when implementing the Metropolis-Hastings algorithm, let us examine the following made-up example.

Sample Claim Data				
484	1407	2262	5015	6500
603	1565	2654	5354	6747
631	1894	2672	5464	9143
1189	2140	4019	5598	12782
1229	2244	4318	6060	18349

We want to model the losses using a lognormal distribution with unknown parameter μ and known parameter $\sigma = 1$.

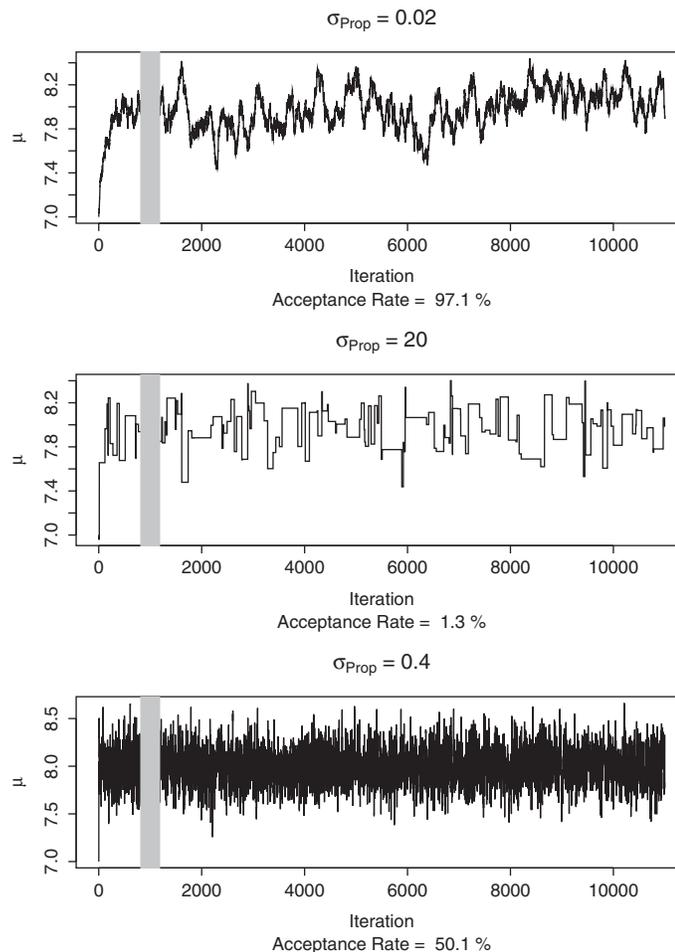
The prior distribution of μ is a normal distribution with mean 8 and standard deviation 1. For the proposal distribution of $(\mu_t|\mu_{t-1})$, I chose a normal distribution with mean μ_{t-1} and standard deviation σ_{Prop} . The starting value, μ_1 , was set equal to 7.00. For this example, there is no adaptive phase and the burn-in phase was 1,000 iterations.

To illustrate the effect of the choice of the proposal distribution, I ran the Metropolis-Hastings algorithm using the normal proposal distributions with $\sigma_{Prop} = 0.02$ (low volatility), $\sigma_{Prop} = 20$ (high volatility) and $\sigma_{Prop} = 0.4$ (volatility just about right). Figure 23 shows plots of the value of μ_t as the chain progresses for each choice of σ_{Prop} . These plots are generally called trace plots in the MCMC literature.

Note that while the starting value $\mu_1 = 7$ was outside of the high density region of the posterior distribution of μ , as t increases μ_t moves rather quickly into the high density region for $\sigma_{Prop} = 20$ and $\sigma_{Prop} = 0.4$. It takes a bit longer for $\sigma_{Prop} = 0.02$, as the differences between μ^* and μ_{t-1} tend to be small.

If $\sigma_{Prop} = 0.02$, μ^* will be close to μ_{t-1} and the ratio in Step 2 of the Metropolis-Hastings algorithm will be relatively high and thus μ_t will be close to μ_{t-1} . In the first

Figure 23. Trace Plot 1: Metropolis—Hastings Example



trace plot of Figure 23 we see a high degree of autocorrelation between successive iterations. If $\sigma_{prop} = 20$, μ^* could be quite far from μ_{t-1} and the ratio in Step 2 could be relatively low and thus μ_t will equal μ_{t-1} . In the second trace plot of Figure 23 we still see a high degree of autocorrelation. If $\sigma_{prop} = 0.4$, μ^* can be far enough away from μ_{t-1} to reduce the autocorrelation, and close enough to avoid rejection and the setting of $\mu_t = \mu_{t-1}$. Getting a good value for σ_{prop} is balancing act. The third trace plot in Figure 23 shows a relatively low degree of autocorrelation and suggests that μ_t for $t = 1001, \dots, 11000$ is a representative sample from the posterior distribution.

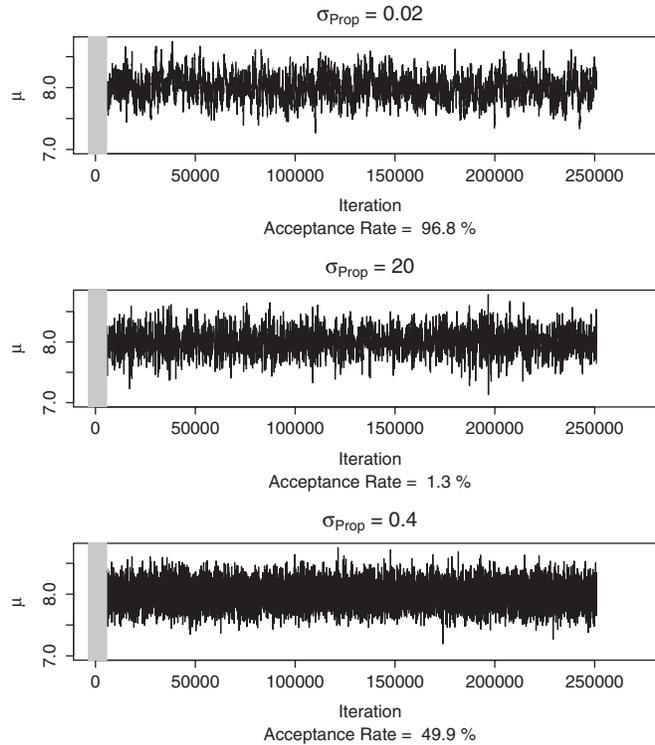
For a single parameter model, like the one in this example, it is relatively easy to scale the proposal distribution by trial and error to minimize autocorrelation. For models with many parameters, like the ones in the next section, such manual scaling is not practical. This problem has been studied extensively and here is a short description of the current state of the art.

A good statistic to look at when trying to minimize autocorrelation in the Metropolis-Hastings algorithm is the acceptance rate of y^* into the Markov chain. I have scanned a number of sources, e.g., Chapter 5 in Jackman (2009), or Chapter 4 of Brooks et al. (2011), that suggest that an acceptance rate of about 50% is near optimal for a one parameter model. The optimal acceptance rate decreases to about 25% as we increase the number of parameters in our model. Also, the researchers have developed methods to automatically adjust the proposal density function in the Metropolis-Hastings algorithm. Chapter 4 of Brooks et al. (2011) provides a recent description of the state of the art. We shall see below that all this has been mechanized in JAGS. The phase of generating the Markov chain where the proposal density function is optimized is called the “adaptive” phase.

As models become more complex, adaptive MCMC may not be good enough to eliminate the autocorrelation. While the theory on Markov chain convergence still holds, there is no guarantee on how fast it will converge. So if one observes significant autocorrelation after the best scaling effort, the next best practice is to increase t_3 until there are a sufficient number of ups and downs in the trace plot and then take a sample of the $t_1 + t_2 + 1$ to $t_1 + t_2 + t_3$ iterations. This process is known as “thinning.” Figure 24 shows what happens when we increase t_3 to 250,000 and record every 25th observation.

Before leaving this example, let us examine how one might turn the posterior distribution of μ into something of interest to actuaries. One reason actuaries fit a lognormal distribution to a set of claims is that they want to determine the cost of an excess layer. Given the parameters μ and σ of a lognormal distribution, there are formulas in Appendix A of Klugman, Panjer, and Willmot (2012) that give the cost of an excess layer of loss. The functions that calculate these formulas are included in the R “actuar” package. As the posterior distribution of μ reflects the parameter risk in our model, it is also possible to reflect the parameter risk in the expected cost of a layer by calculating the expected cost of the layer for each μ in the simulated posterior distribution. Also, it is possible to simulate an actual outcome of a loss, X , in a layer given each μ in the posterior distribution. The distribution of X calculated in this way

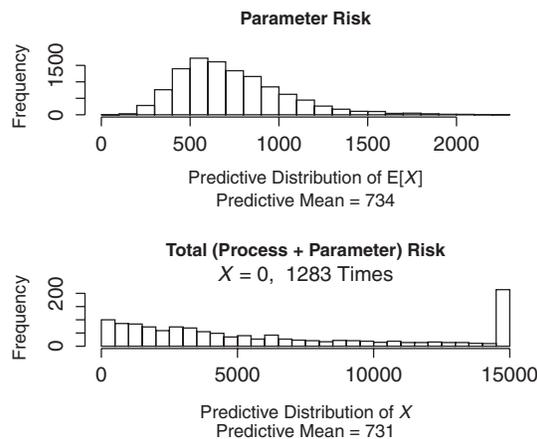
Figure 24. Trace Plot 2: Metropolis—Hastings Example with Thinning



reflects both the parameter risk and the process risk in the model. Figure 25 shows the predictive distribution of the expected cost of the layer between 10,000 and 25,000, $E[X]$, and the predicted outcome of losses X in that layer.

As statisticians and practitioners became aware of the potential for Bayesian MCMC modeling in solving real-world problems, a general software initiative to implement Bayesian MCMC analyses, called the BUGS project, began. BUGS is an

Figure 25. Predictive Distributions of the Cost of a Layer between 10,000 and 25,000



acronym for **B**ayesian inference **U**sing **G**ibbs **S**ampling.¹⁷ The project began in 1989 in the MRC Biostatistics Unit, Cambridge, and led initially to the ‘Classic’ BUGS program, and then onto the WinBUGS software developed jointly with the Imperial College School of Medicine at St Mary’s, London. The project’s web site is at <http://www.mrc-bsu.cam.ac.uk/bugs/>. The various software packages associated with the BUGS project have captured many of the good techniques involved in Bayesian MCMC modeling.

On the advice of some colleagues I chose to use the JAGS (**J**ust **A**nother **G**ibbs **S**ampler) package. It has the additional feature that it runs on a variety of platforms (Window, Mac, Linux and several varieties of Unix). Like R, it can be downloaded for free.

I use JAGS with R. My typical MCMC program begins by reading in the data, calling the JAGS script using the R package “runjags.” I then fetch the sample of the posterior back into the R program where I calculate various “statistics of interest.”

While I realize that JAGS is doing something more sophisticated, I find it helpful to “think” of JAGS as using a simple version of the Metropolis–Hasting algorithm similar to that illustrated in the example above. Once a model is specified, there are three stages in running a JAGS program:

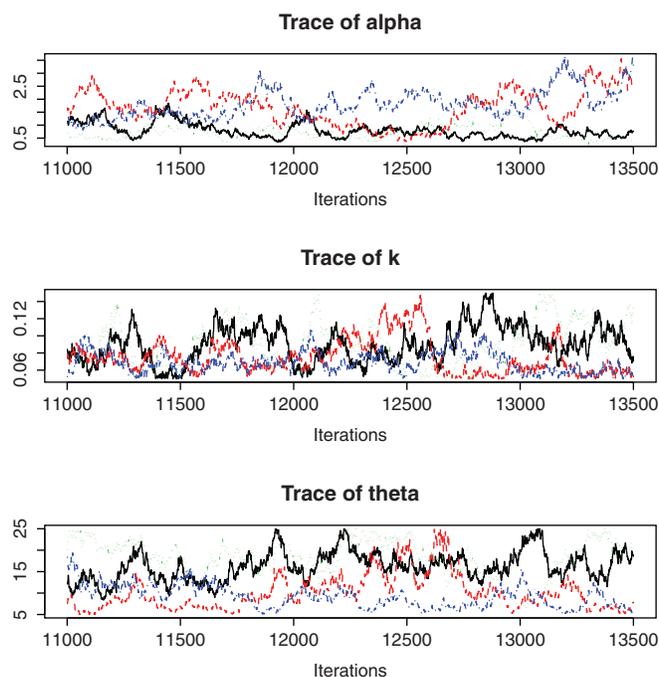
1. The adaptive stage where JAGS modifies the proposal distribution in the Metropolis–Hastings algorithm. JAGS will issue a warning if it thinks that you haven’t allowed enough iterations for adapting. Let’s denote the number of iterations for scaling by t_1 .
2. The burn-in stage runs until we have reached the limiting posterior distribution. JAGS has diagnostics (described below) that indicate convergence. The burn-in stage runs from iterations $t_1 + 1$ to $t_1 + t_2$.
3. The sampling stage that produces the sample of the posterior distribution. The sampling stage runs from iterations $t_1 + t_2 + 1$ to $t_1 + t_2 + t_3$.

JAGS has a number of convergence diagnostics that are best illustrated with an example. We are given the total losses from a set of thirty insurance policies in the following table.

Exposure	Loss	Exposure	Loss	Exposure	Loss
51	23	226	273	368	410
66	138	231	275	374	482
119	53	254	259	377	500
125	88	255	200	381	424
131	80	258	123	392	242
152	136	268	275	444	431
196	165	279	327	449	337
197	136	295	509	478	399
225	328	340	457	484	458
225	347	364	317	495	553

¹⁷ Gibbs sampling is an MCMC algorithm that is a special case of the Metropolis Hastings algorithm. This is demonstrated in Chapter 1 of Brooks *et al.* (2011).

**Figure 26. Trace Plots Without Thinning—
CRM Example**



Our task is to use these data to estimate the expected cost of losses in excess of 1000 for an insurance policy with an exposure of 800. Note that in our data, there is no insurance policy with an exposure as high as 800, and no loss over 1000.

Let's use the collective risk model with a Poisson distribution for the claim count, and a distribution for the claim severity. Here is the description of the model using the notation in Klugman, Panjer, and Willmot (2012).¹⁸

1. $\lambda = k \cdot \text{Exposure}$
2. $n \sim \text{Poisson}(\lambda)$
3. $\text{Loss} \sim \Gamma(n \cdot \alpha, \theta)$
4. $k \sim \text{Uniform}(0.05, 0.15)$
5. $\alpha \sim \text{Uniform}(0.1, 10)$
6. $\theta \sim \text{Uniform}(5, 200)$

In JAGS, the script looks pretty much like the model description above after a change in notation for the distribution parameters. Let's first consider convergence diagnostics. First of all, with JAGS one can run multiple independent chains. I first ran this model with 1,000 iterations for the adaptive stage, 10,000 iterations for the burn-in stage and then 2,500 iterations for the sampling stage. JAGS then produces trace plots for all four chains, colored differently, superimposed on each other. A visual indication of convergence is that all the chains bounce around in the same general area. Figure 26 shows the trace plots produced by JAGS for the three parameters in this example.

¹⁸ This particular version of the collective risk model is called a Tweedie distribution. See Meyers (2009).

As we can see from the trace plots, the chains are very distinct, so we should conclude that the chains have not converged.

A second diagnostic provided by JAGS is the Gelman-Rubin statistic for each parameter. Here is a heuristic description of the statistic.¹⁹ First estimate the within chain variability, \widehat{W} , and the between chain variability, \widehat{B} . Gelman and Rubin then recommend that one use the statistic

$$\sqrt{\widehat{R}} = \sqrt{\frac{\widehat{W} + \widehat{B}}{\widehat{W}}}.$$

The $\sqrt{\widehat{R}}$ is called the “potential scale reduction (or ‘shrink’) factor.” or PSRF. This statistic will approach one as the number of iterations increases, since the between chain variability will approach zero. What we need to know is how long the chains have to be before we can stop and get a representative sample of the posterior distribution. Chapter 6 of Brooks et al. (2011) recommends that we accept convergence if the PSRF is 1.1 or below for all parameters. The default for the “runjags” package is 1.05, which is what I used in for the models in this monograph. The PSRFs for this JAGS run were 1.87, 1.21 and 1.92 for the parameters α , k and θ , respectively.

Continuing the example, I reran the JAGS model with same parameters but thinned the chains to take every 25th iteration. The results are in Figure 27. The PSRFs for this JAGS run were 1.03, 1.02 and 1.01 for the parameters α , k and θ respectively. So we can accept that the run has converged.

JAGS then sent 10,000 parameter sets $\{\alpha_i, k_i, \theta_i\}$ back to the R script. R then simulated losses to the insurance policy as follows.

For $t = 1$ to 10,000.

1. Set $\lambda = k_t \cdot 800$.
2. Select n_t at random from a Poisson distribution with mean λ .
3. Select $Loss_t$ at random from a $\Gamma(n_t \cdot \alpha_t, \theta_t)$ distribution.²⁰

Figure 28 shows a histogram of the ground up losses from the above simulation and the expected cost of the layer in excess of 1,000.

The examples in this appendix illustrate the ideas behind Bayesian MCMC models, those being the adaptive phase, the burn-in phase, the sampling phase, and convergence testing. Understanding these concepts should enable one to start running these kinds of models. When running these models one should keep in mind that the state of the art is still evolving, so one should periodically check the current literature and software developments on Bayesian MCMC modeling for recent developments.

¹⁹ See Jackman (2009, Section 6.2) or Hartman (2014) for a more detailed description of this statistic.

²⁰ If each X_i has a $\Gamma(\alpha, \theta)$ distribution, then $X_1 + \dots + X_n$ has a $\Gamma(n \cdot \alpha, \theta)$ distribution.

**Figure 27. Trace Plots With Thinning—
CRM Example**

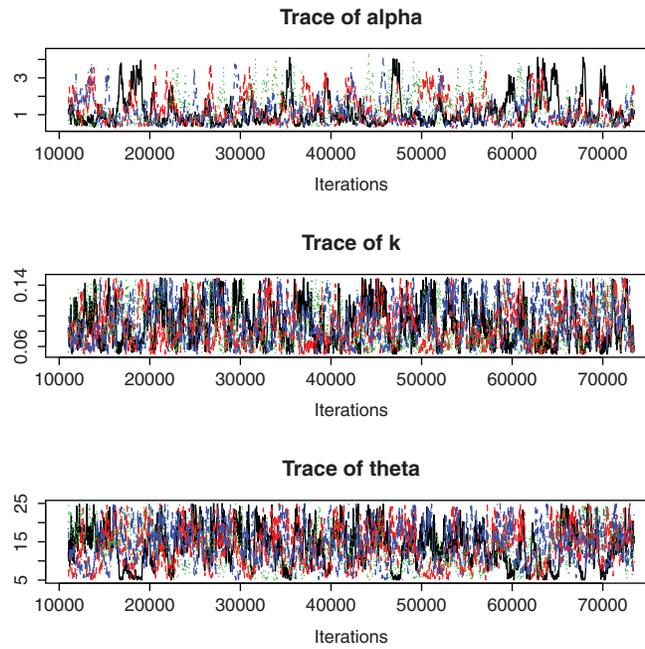
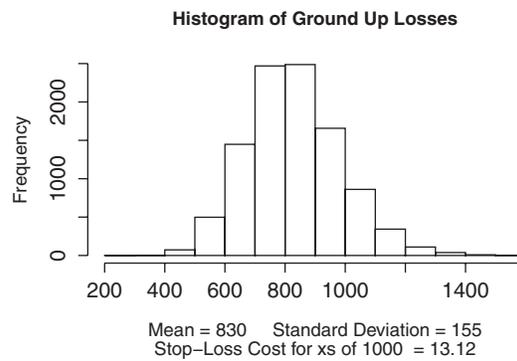


Figure 28. Output—CRM Example



Appendix C. Bayesian MCMC Model Implementation

The state of the art and the software for Bayesian MCMC modeling is still evolving. Since there may be upgrades by the time the reader sees this monograph, I think that it is important for me to describe the computing environment in which I ran the models in this monograph.

My computer was a Macbook Pro with a quad core processor. On this computer I used R version 3.0.2 and JAGS version 3.3, implementing JAGS with the “runjags” package. The main consideration in selecting the “runjags” package was that made it easy to run the four chains in parallel with my quad core computer. Running the chains in parallel made a significant improvement in the run time.

For the LCL, CCL, and CSR models I used 1,000 iterations for the adaptive phase, and 10,000 iterations for the burn-in phase. I ran the model inside a loop, with the sampling phase initially set at 10,000 iterations with a thinning parameter equal to four. If the maximum PSRF for the parameters I monitored was greater than 1.05, I doubled the number of iterations in the sampling phases and the thinning parameter and ran the simulation again—continuing until the target PSRF target was achieved.

For most of the LCL and CCL models on incurred data, the initial run achieved the PSRF target. The highest thinning parameter was 32. Convergence was somewhat slower for the CCL and CSR models on the paid data. There was one triangle that required a thinning parameter equal to 512.

For the CIT and LIT models on the paid data, I increased the burn-in to 50,000 iterations. Convergence was noticeably slower. Far fewer triangles met the PSRF target with a thinning parameter set equal to four.

The R/JAGS scripts for all models are in a spreadsheet that will be distributed with this monograph. For each model, I put these scripts inside a loop that ran all 200 triangles while I was otherwise occupied. Summary statistics for all 200 triangles are also included in the spreadsheet and because I fixed the random number seed, the scripts are able to reproduce the summary statistics for any of the triangles.

ABOUT THE SERIES:

CAS monographs are authoritative, peer-reviewed, in-depth works focusing on important topics within property and casualty actuarial practice. For more information on the CAS Monograph Series, visit the CAS website at www.casact.org.



**Expertise. Insight.
Solutions.**

www.casact.org