Bayesian MCMC Stochastic Loss Reserve Models for Paid Loss Triangles

Glenn Meyers

Casualty Loss Reserve Seminar

September 17, 2020

Background - Bayesians vs. Frequentists

- Given the model $X \sim f(X|\theta)$
- Given the set of observations x.
  - Frequentists test the hypothesis $\theta = \theta_0$.
  - Bayesians calculate the posterior distribution $f(\theta|x)$.

$$f(\theta|x) = \frac{f(x|\theta) \cdot \pi(\theta)}{\int\limits_{\vartheta} f(x|\vartheta) \cdot \pi(\vartheta) \cdot d\vartheta}$$

- The issue — What is the prior distribution, $\pi(\theta)$?

The Philosophical Issue

- Bayesians select π "subjectively" according to prior opinion.
- Frequentists respond by saying that conclusions should be dictated solely by looking at "the data."
- Bayesians respond with "noninformative" priors.
  - Is there such a thing?

The Practical Issue — Can we do the calculations?

- For most of the 20th century, the frequentists were winning.
  - Calculations were easy with quadradic forms needed for the normal distributions.
  - The General Linear Model (PROC GLM in SAS).
  - As computers and numerical analysis progressed we got the **Generalized** Linear Model (PROC GENMOD in SAS).
- Now the Bayesians are winning - with MCMC and good software to implement it.

The Problem with Bayesian Analysis

- Let θ be an *n*-parameter vector — e.g. development factors.
- Let X be a set of observations — e.g. a loss triangle.

$$f(\theta|x) = \frac{f(x|\theta)\cdot\pi(\theta)}{\int\limits_{\vartheta_1}\cdots\int\limits_{\vartheta_n} f(x|\vartheta)\cdot\pi(\vartheta)\cdot d\vartheta}$$

- $f(X|\theta)$ is the likelihood of X given θ.
- $\pi(\theta)$ is the prior distribution of θ.
- $f(\theta|X)$ is the posterior distribution of θ.

The Problem with Bayesian Analysis

- Let θ be an n-parameter vector — e.g. development factors.
- Let X be a set of observations — e.g. a loss triangle.

$$f(\theta|x) = \frac{f(x|\theta) \cdot \pi(\theta)}{\int\limits_{\vartheta_1} \cdots \int\limits_{\vartheta_n} f(x|\vartheta) \cdot \pi(\vartheta) \cdot d\vartheta}$$

- $f(X|\theta)$ is the likelihood of X given θ.
- $\pi(\theta)$ is the prior distribution of θ.
- $f(\theta|X)$ is the posterior distribution of θ.
- Calculating the n-dimensional integral is intractable.

A New World Order

- This impasse came to an end in 1990 when a simulation-based approach to estimating posterior probabilities was introduced.
- *Sampling Based Approach to Calculating Marginal Densities*
  - Alan E. Gelfand and Adrian F.M. Smith
  - Journal of the American Statistical Association, June 1990

Markov Chains

- Let Ω be a finite state with random events

$$X_1, X_2, \ldots, X_t, \ldots$$

- A Markov chain P satisfies

$$Pr(X_t = y | X_{t-1} = x_{t-1}, \ldots, X_1 = x_1) = Pr(X_t = y | X_{t-1} = x_{t-1})$$

- The probability of an event in the chain depends only on the immediate previous event.

The Markov Convergence Theorem

- There is a branch of probability theory, called Ergodic Theory, that gives conditions for which there exists a unique stationary distribution, π, such that

$$Pr(y|X_{t-1}) \longrightarrow \pi(y)$$

$$\text{as } t \longrightarrow \infty$$

- "Convergence" means that for sufficiently large $t > T$, X_t can be thought of as a random draw from the distribution function π.

The Metropolis Hastings Algorithm
A Very Important Markov Chain.

1. Time $t = 1$: select a random initial position θ_1 in parameter space.
2. Select a proposal distribution $p(\theta|\theta_{t-1})$ that we will use to select proposed random steps away from our current position in parameter space.
3. Starting at time $t = 2$: repeat the following until you get convergence:
   - At step t, generate a proposal $\theta^* \sim p(\theta|\theta_{t-1})$.
   - Generate $U \sim$ uniform(0,1)
   - Calculate

$$R = \frac{f(\theta^*|x)}{f(\theta_{t-1}|x)} \cdot \frac{p(\theta_{t-1}|\theta^*)}{p(\theta^*|\theta_{t-1})}$$

   - If $U < R$ then $\theta_t = \theta^*$. Else, $\theta_t = \theta_{t-1}$.

Dodging the Intractable Integral

$$R = \frac{f(\theta^*|x)}{f(\theta_{t-1}|x)} \cdot \frac{p(\theta_{t-1}|\theta^*)}{p(\theta^*|\theta_{t-1})}$$

$$R = \frac{\dfrac{f(x|\theta^*)\cdot\pi(\theta^*)}{\int\limits_{\vartheta_1}\cdots\int\limits_{\vartheta_n} f(x|\vartheta)\cdot\pi(\vartheta)\cdot d\vartheta}}{\dfrac{f(x|\theta_{t-1})\cdot\pi(\theta_{t-1})}{\int\limits_{\vartheta_1}\cdots\int\limits_{\vartheta_n} f(x|\vartheta)\cdot\pi(\vartheta)\cdot d\vartheta}} \cdot \frac{p(\theta_{t-1}|\theta^*)}{p(\theta^*|\theta_{t-1})}$$

The integral $\int\limits_{\vartheta_1}\cdots\int\limits_{\vartheta_n} f(x|\vartheta)\cdot\pi(\vartheta)\cdot d\vartheta$ cancels out!

The Metropolis Hastings Algorithm
A Very Important Markov Chain.

1. Time $t = 1$: select a random initial position θ_1 in parameter space.
2. Select a proposal distribution $p(\theta|\theta_{t-1})$ that we will use to select proposed random steps away from our current position in parameter space.
3. Starting at time $t = 2$: repeat the following until you get convergence:
   - At step t, generate a proposal $\theta^* \sim p(\theta|\theta_{t-1})$.
   - Generate $U \sim$ uniform(0,1)
   - Calculate
   $$R = \frac{f(x|\theta^*) \cdot \pi(\theta^*)}{f(x|\theta_{t-1}) \cdot \pi(\theta_{t-1})} \cdot \frac{p(\theta_{t-1}|\theta^*)}{p(\theta^*|\theta_{t-1})}$$
   - If $U < R$ then $\theta_t = \theta^*$. Else, $\theta_t = \theta_{t-1}$.

The Relevance of the Metropolis Hastings Algorithm

- Defined in terms of the conditional distribution

$$f(X|\theta)$$

and the prior distribution

$$\pi(\theta)$$

- The limiting distribution is the **posterior distribution!**

The Relevance of the Metropolis Hastings Algorithm

- Defined in terms of the conditional distribution

$$f(X|\theta)$$

and the prior distribution

$$\pi(\theta)$$

- The limiting distribution is the **posterior distribution!**
- Code $f(X|\theta)$ and $\pi(\theta)$ into a Markov chain and let it run for a while, and you have a large sample from the posterior distribution.

The Relevance of the Metropolis Hastings Algorithm

- The theoretical limiting distribution is the same, no matter what proposal distribution, $p(\theta|\theta_{t-1})$, is used.
- The choice of the proposal distribution does affect the speed of convergence. The latest software is pretty fast.
- There is no fundamental limit on the number of parameters in you model!
- The practical limit is within range of stochastic loss reserve models.

A Short History of MCMC

- Originated with the study of nuclear fission.
  - Enrico Fermi, John von Neumann, Nicolas Metropolis and Stanislaw Ulam.
  - Developed the Metropolis algorithm.
- Keith Hastings (1970) recognized the potential of the Metropolis algorithm to solve statistical problems.
- Simulations were not readily accepted by the statistical community at that time.

A Short History of MCMC

- Gelfand and Smith (1990) pulled together the relevant ideas at a time when simulation was deemed OK.
  - Seized upon by scientists in other fields.
  - Used the Gibbs sampler (A one parameter at a time special case of Metropolis Hastings algorithm).
- Statisticians in the UK started the BUGS project to produce software for MCMC.
  - **B**ayesian inference **U**sing the **G**ibbs **S**ampler

Evolution of MCMC Software

- WinBUGS (Original — now discontinued)
- OpenBUGS (Continuation of WinBUGS)
  - Designed mainly for the Windows operating system.
- JAGS — **J**ust **A**nother **G**ibbs **S**ampler
  - Originated by Martyn Plummer.
  - Runs on multiple operating systems.
  - Callable from R ("runjags" package.
- Stan (in honor of Stanislaw Ulam)
  - Stan team led by Andrew Gelman at Columbia University.
  - Runs on multiple operation systems.
  - Callable from R ("rstan" package) and other languages, e.g. Python and Matlab.

Evolution of MCMC Software

- WinBUGS (Original — now discontinued)
- OpenBUGS (Continuation of WinBUGS)
  - Designed mainly for the Windows operating system.
- JAGS — **J**ust **A**nother **G**ibbs **S**ampler
  - Originated by Martyn Plummer.
  - Runs on multiple operating systems.
  - Callable from R ("runjags" package.
- Stan (in honor of Stanislaw Ulam)
  - Stan team led by Andrew Gelman at Columbia University.
  - Runs on multiple operation systems.
  - Callable from R ("rstan" package) and other languages, e.g. Python and Matlab.
- Features of the latest software: (1) good convergence diagnostics, and (2) fast convergence.

Outline of The Main Event — Reserving

- Data taken from the CAS Loss Reserve Database. 50 "well behaved" loss triangles from each of the CA, PA, WC and OL lines of business.
- Stochastic loss reserve modeling with Bayesian MCMC
  - Focus on paid loss triangles.
  - Three models using paid triangles, one using incurred triangles and one model using combined paid and incurred triangles.
- Diagnostics on those models
  - Real-time diagnostics useful for current loss reserve analyses.
  - Long-term "Reputation" diagnostics based on how well a model performs on other "similar" loss triangles taken from the CAS Loss Reserve Database

References — CAS Monograph Series

- Stochastic Loss Reserving Using Bayesian MCMC Models
- Stochastic Loss Reserving Using Bayesian MCMC Models - 2nd Edition
- The difference between the two editions
  - 1st edition provides a gentler introduction to Bayesian MCMC.
  - 2nd edition uses up-to-date software and more diagnostics. It also includes dependencies between lines of business and risk margins.
  - Today's webinar draws on material from the second edition.
- Today's talk will focus on paid data. The monographs provide more coverage for incurred data.

Common Features of the Models in this Talk

- Let C_{wd} equal the cumulative loss (paid or incurred) for accident year, w, and development lag, d.
- All models discussed in these talk will be of the form:

$$C_{wd} \sim \text{lognormal}(\mu_{wd}, \sigma_d)$$

- The examples that follow will be 10x10 loss triangles from US Schedule P.
- As an accident year matures, an increasing proportion of claims are settled. Thus we impose the condition.

$$\sigma_1^2 > \ldots > \sigma_{10}^2$$

- The models discussed in this talk explore alternative ways to model μ_{wd}.

Schedule P Data for Paid Losses

Table: 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: 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

The CRoss Classified (CRC) Model

— Parameters

- $logelr \sim$ normal(-0.4,$\sqrt{10}$).
- $\alpha_w \sim \text{normal}(0, \sqrt{10})$ for $w = 2, \ldots, 10$. $\alpha_1 = 0$.
- $\beta_d \sim \text{normal}(0, \sqrt{10})$ for $d = 1, \ldots, 9$. $\beta_{10} = 0$.
- $a_i \sim \text{uniform}(0, 1)$ for $i = 1, \ldots, 10$

The CRoss Classified (CRC) Model

— Parameters

- $logelr \sim$ normal(-0.4,$\sqrt{10}$).
- $\alpha_w \sim \text{normal}(0, \sqrt{10})$ for $w = 2, \ldots, 10$. $\alpha_1 = 0$.
- $\beta_d \sim \text{normal}(0, \sqrt{10})$ for $d = 1, \ldots, 9$. $\beta_{10} = 0$.
- $a_i \sim \text{uniform}(0, 1)$ for $i = 1, \ldots, 10$

— Transformed Parameters

- $\sigma_d^2 = \sum_{i=d}^{10} a_i$ for $d = 1, \ldots, 10$.
  Note that this forces $\sigma_1^2 > \ldots > \sigma_{10}^2$
- $\mu_{wd} = \log(\text{Premium}_w) + logelr + \alpha_w + \beta_d$

The CRoss Classified (CRC) Model

— Parameters

- $logelr \sim$ normal(-0.4,$\sqrt{10}$).
- $\alpha_w \sim \text{normal}(0, \sqrt{10})$ for $w = 2, \ldots, 10$. $\alpha_1 = 0$.
- $\beta_d \sim \text{normal}(0, \sqrt{10})$ for $d = 1, \ldots, 9$. $\beta_{10} = 0$.
- $a_i \sim \text{uniform}(0, 1)$ for $i = 1, \ldots, 10$

— Transformed Parameters

- $\sigma_d^2 = \sum_{i=d}^{10} a_i$ for $d = 1, \ldots, 10$.
  Note that this forces $\sigma_1^2 > \ldots > \sigma_{10}^2$
- $\mu_{wd} = \log(\text{Premium}_w) + logelr + \alpha_w + \beta_d$

— Model

- Then $C_{wd} \sim \text{lognormal}(\mu_{wd}, \sigma_d)$

Run "CRC.R" to get 10,000 {.}

- The scripts are from the monograph. They can be downloaded from

  https://www.casact.org/pubs/monographs/meyers/Appendix.zip.

Outline of "CRC.R"

- Get 10,000 parameter vectors by running Stan within R. $\{logelr\}, \{\alpha_{1:10}\}, \{\beta_{1:10}\},$ and $\{\sigma_{1:10}\}$
- Calculate $\{\mu_{2:10,10}\} = \log(\text{Premium}_w) + \{logelr\} + \{\alpha_{2:10}\}$ Note $\beta_{10} = 0$
- Simulate "ultimate" outcomes

$$\{C_{2:10,10}\} \sim \{\text{lognormal}(\mu_{2:10,10}, \sigma_{10})\}$$

- Calculate the totals $\{\sum_{w=1}^{10} C_{w,10}\}$
- Calculate "statistics of interest" of $\{\sum_{w=1}^{10} C_{w,10}\}$
  - Mean$[\{\sum_{w=1}^{10} C_{w,10}\}]$
  - Standard Deviation$[\{\sum_{w=1}^{10} C_{w,10}\}]$
  - Percentile of actual "ultimates" from holdout lower triangle data.

CRC Output for the Paid Illustrative Loss Triangle

w	Premium	Estimate	SE	CV	Outcome	Percentile
1	5812	3912	0	0.0000	3912	
2	4908	2564	114	0.0445	2527	
3	5454	4149	193	0.0465	4274	
4	5165	4315	223	0.0517	4341	
5	5214	3566	203	0.0569	3583	
6	5230	3410	249	0.0730	3268	
7	4992	5208	445	0.0854	5684	
8	5466	3630	442	0.1218	4128	
9	5226	4392	817	0.1860	4144	
10	4962	4976	1762	0.3541	4139	
Total	52429	40121	2487	0.0620	40000	51.88

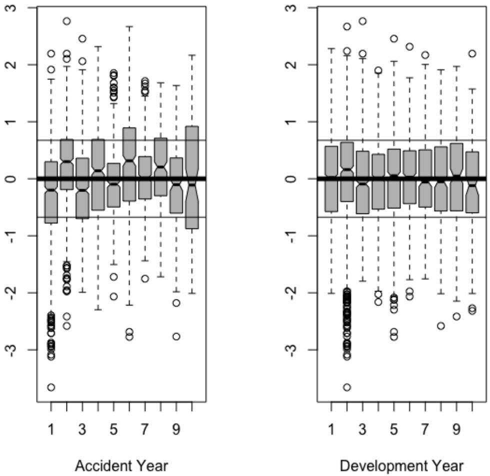
Standardized Residual Boxplots

- We have a large sample of parameter vectors, rather than a single parameter vector that arises from, say, a maximum likelihood estimate.
- Take a random sample of 100 js and use the corresponding parameter vectors, μ_{wd}^{j} and σ_{d}^{j} from the posterior distribution and calculate the standardized residuals, r_{wd}^{j} , for the log of all losses in the training (upper) loss triangle.

$$r_{wd}^{j} = \frac{\log(C_{wd}) - \mu_{wd}^{j}}{\sigma_{d}^{j}}$$

  for $w = 1 \ldots, 10, d = 1, \cdots, 11 - w$ and for each j.
- Do Boxplots by accident year and development year.

Standardized Residual Boxplots — CRC Model



The Stochastic Cape Cod (SCC) Model

— Parameters

- $logelr \sim$ normal(-0.4,$\sqrt{10}$).
- α_w is dropped.
- $\beta_d \sim$ normal$(0, \sqrt{10})$ for $d = 1, \ldots, 9$. $\beta_{10} = 0$.
- $a_i \sim$ uniform$(0, 1)$ for $i = 1, \ldots, 10$

— Transformed Parameters

- $\sigma_d^2 = \sum_{i=d}^{10} a_i$ for $d = 1, \ldots, 10$.
  Note that this forces $\sigma_1^2 > \ldots > \sigma_{10}^2$
- $\mu_{wd} = \log(\text{Premium}_w) + logelr + \beta_d$

— Model

- Then $C_{wd} \sim$ lognormal(μ_{wd}, σ_d)

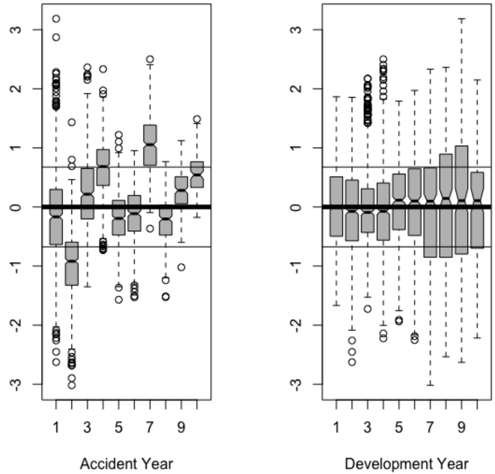
SCC Output for the Paid Illustrative Loss Triangle

w	Premium	Estimate	SE	CV	Outcome	Percentile
1	5812	3912	0	0.0000	3912	
2	4908	2905	640	0.2203	2527	
3	5454	4265	724	0.1698	4274	
4	5165	4199	703	0.1674	4341	
5	5214	3376	694	0.2056	3583	
6	5230	3097	690	0.2228	3268	
7	4992	4645	665	0.1432	5684	
8	5466	3180	700	0.2201	4128	
9	5226	3639	657	0.1805	4144	
10	4962	3506	591	0.1686	4139	
Total	52429	36725	3950	0.1075	40000	83.38

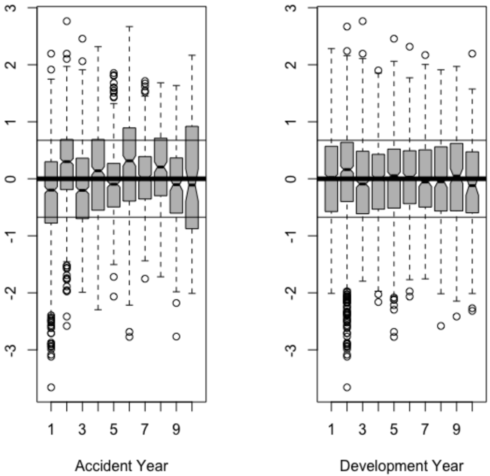
CRC Output for the Paid Illustrative Loss Triangle

w	Premium	Estimate	SE	CV	Outcome	Percentile
1	5812	3912	0	0.0000	3912	
2	4908	2564	114	0.0445	2527	
3	5454	4149	193	0.0465	4274	
4	5165	4315	223	0.0517	4341	
5	5214	3566	203	0.0569	3583	
6	5230	3410	249	0.0730	3268	
7	4992	5208	445	0.0854	5684	
8	5466	3630	442	0.1218	4128	
9	5226	4392	817	0.1860	4144	
10	4962	4976	1762	0.3541	4139	
Total	52429	40121	2487	0.0620	40000	51.88

Standardized Residual Boxplots — SCC Model



Standardized Residual Boxplots — CRC Model



Model Selection With the “loo” Package”

- Given two different models for the same data, how do you select the “better” model?

Model Selection With the "loo" Package"

- Given two different models for the same data, how do you select the "better" model?
- "loo" stands for **L**eave **O**ne **O**ut.
- Maintained by members of the stan development team.
- Vehtari, A., Gelman, A., and Gabry, J. (2015). "Efficient implementation of leave-one-out cross validation and WAIC for evaluating fitted Bayesian models."
- See the documentation of the "loo" package for the latest version of the paper.

Selecting Models Fit By Maximum Likelihood

- If we fit a model, $f(x|\theta)$, by maximum likelihood, define

$$AIC = 2 \cdot p - 2 \cdot L(x|\hat{\theta})$$

- Where:
  - p is the number of parameters in the model.
  - $L(x|\hat{\theta})$ is the maximum log-likelihood of the model specified by f.
- Lower AIC indicates a better fit.
  - Encourages a larger log-likelihood.
  - Penalizes an increase in the number of parameters.

AIC in a Bayesian Environment?

- When doing maximum likelihood estimation, we have a single parameter vector.
- With MCMC we have 10,000 parameter vectors.

AIC in a Bayesian Environment?

- When doing maximum likelihood estimation, we have a single parameter vector.
- With MCMC we have 10,000 parameter vectors.
- In a Bayesian environment — Should the penalty for a parameter be as great when there is strong prior information?

Selecting Bayesian MCMC Models with the LOOIC Statistic

- Given an MCMC model with parameters $\{\theta^i\}_{i=1}^{10,000}$, define

$$LOOIC = 2 \cdot \hat{p}_{loo} - 2 \cdot \overline{\{L(x|\theta^i)\}_{i=1}^{10,000}}$$

- Where
  - $\hat{p}_{loo}$ is the ***effective*** number of parameters.

$$\hat{p}_{loo} = \overline{\{L(x|\theta^i)\}_{i=1}^{10,000}} - \sum_{n=1}^{N} \overline{\{L(x_n|x_{(-n)}, \theta^i)\}_{i=1}^{10,000}}$$

  - $x_{(-n)} = x_1, \ldots, x_{n-1}, x_{n+1}, \ldots, x_N$
  - $L(x_n|x_{(-n)}, \theta^i)$ is the log-likelihood of x_n from a model fit using all data except x_n.

Selecting Bayesian MCMC Models with the LOOIC Statistic

- After some algebra we can see that

$$LOOIC = -2 \cdot \sum_{n=1}^{N} \overline{\{L(x_n|x_{(-n)}, \theta^i)\}_{i=1}^{10,000}}$$

which we like as it favors the model with the largest likelihood, and the smallest LOOIC, on the "holdout" data.

Selecting Bayesian MCMC Models with the LOOIC Statistic

- After some algebra we can see that

$$LOOIC = -2 \cdot \sum_{n=1}^{N} \overline{\{L(x_n|x_{(-n)}, \theta^i)\}_{i=1}^{10,000}}$$

which we like as it favors the model with the largest likelihood, and the smallest LOOIC, on the "holdout" data.

- Some "loo" package features.
  - $\sum_{n=1}^{N} \overline{\{L(x_n|x_{(-n)}, \theta^i)\}_{i=1}^{10,000}} \equiv ELPD_{loo}$
  - "loo" does not calculate each summand in $ELPD_{loo}$ by MCMC. Instead it approximates the sum using a 10,000 x N matrix of log-likelihoods.

Model Comparison (loo) Statistics

Model	$\widehat{elpd}_{loo}$	p_{loo}	LOOIC
CRC-Paid	47.80	14.97	-95.60
SCC-Paid	-5.14	8.75	10.28

- In my monograph, I calculated the loo statistics for 200 loss triangles, and the CRC fit better than the SCC model for all 200 triangles!

Model Comparison (loo) Statistics

Model	$\widehat{elpd}_{loo}$	p_{loo}	LOOIC
CRC-Paid	47.80	14.97	-95.60
SCC-Paid	-5.14	8.75	10.28

- In my monograph, I calculated the loo statistics for 200 loss triangles, and the CRC fit better than the SCC model for all 200 triangles!
- At this point I want to introduce what I call "**Reputation**" statistics. That is I want to draw conclusions about a model based on looking at a lot of other triangles.

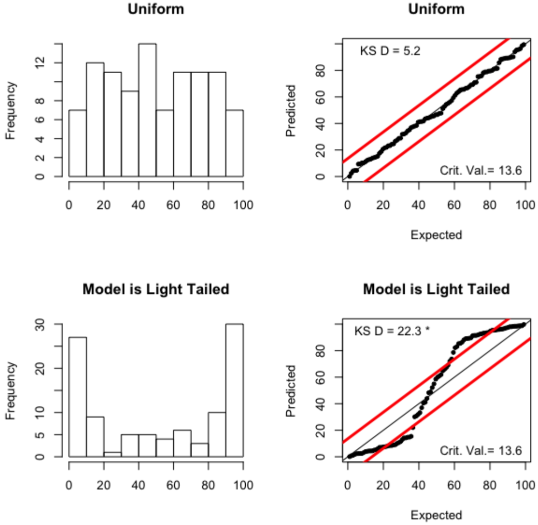
Testing Percentiles of Outcomes with the Lower Triangle Holdout Data.

- Use the model to calculate the percentile of the predictive distribution of the actual outcome.
- We should expect the percentiles of the outcomes to be uniformly distributed.
- Uniformity is testable with model fits and outcomes of several insurers.

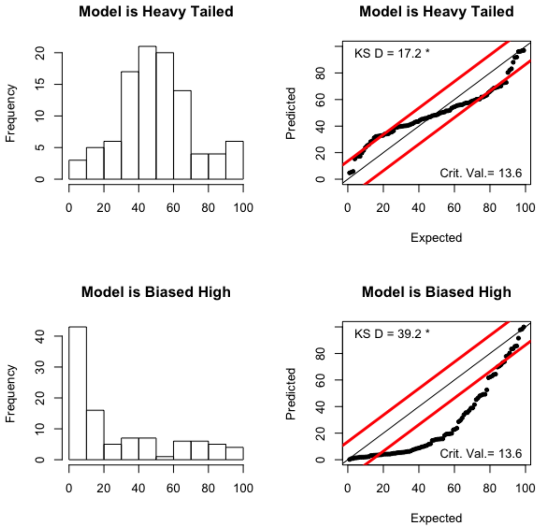
Testing Percentiles of Outcomes with the Lower Triangle Holdout Data.

- Use the model to calculate the percentile of the predictive distribution of the actual outcome.
- We should expect the percentiles of the outcomes to be uniformly distributed.
- Uniformity is testable with model fits and outcomes of several insurers.
- PP Plots - Plot the sorted values of a uniformly distributed set of numbers (Expected) against the sorted percentiles of the outcomes predicted by the model (Predicted).
  - We expect the plot to lie along a 45^o line.
- Kolmogorov-Smirnov test puts bounds around how far the difference between the predicted and expected can be.

PP Plot Characteristics

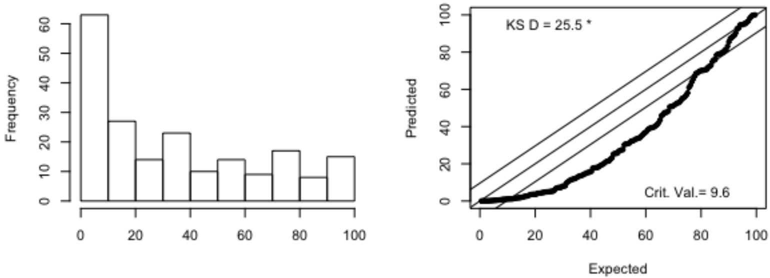


PP Plot Characteristics



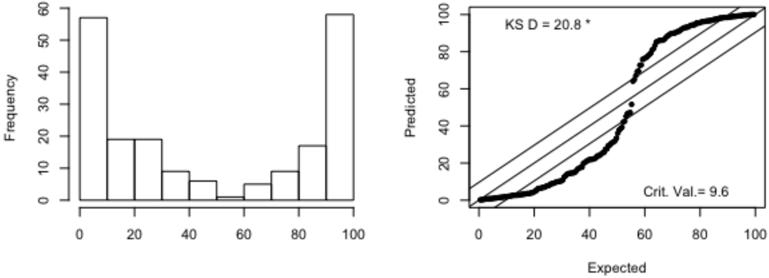
CRC PP Plot on 200 Paid Loss Triangles

CRC model is biased high



SCC PP Plot on 200 Paid Loss Triangles

SCC model has tails that are too light!



So Where Are We?

- We have two models
  - The **CR**oss-**C**lassified (CRC) model
  - The **S**tochastic **C**ape **C**od (SCC) model

So Where Are We?

- We have two models
  - The **CR**oss-**C**lassified (CRC) model
  - The **S**tochastic **C**ape **C**od (SCC) model
- The SCC model has fewer parameters than the CRC model. — Good!
- But the additional parameters in the CRC model add information.

So Where Are We?

- We have two models
  - The **CR**oss-**C**lassified (CRC) model
  - The **S**tochastic **C**ape **C**od (SCC) model
- The SCC model has fewer parameters than the CRC model. — Good!
- But the additional parameters in the CRC model add information.
- The SCC models tend to have tails that are too thin.
- The CRC models tend to have an upward bias.
  - An anonymous referee to my first monograph suggested that claim settlement was speeding up.
- This led to the **C**hanging **S**ettlement **R**ate model.

The Changing Settlement Rate (CSR) Model

— Parameters

- $logelr \sim$ normal(-0.4, $\sqrt{10}$).
- $\alpha_w \sim \text{normal}(0, \sqrt{10})$ for $w = 2, \ldots, 10$. $\alpha_1 = 0$.
- $\beta_d \sim \text{normal}(0, \sqrt{10})$ for $d = 1, \ldots, 9$. $\beta_{10} = 0$.
- $\gamma \sim \text{normal}(0, 0.05)$.
- $a_i \sim \text{uniform}(0, 1)$ for $i = 1, \ldots, 10$

— Transformed Parameters

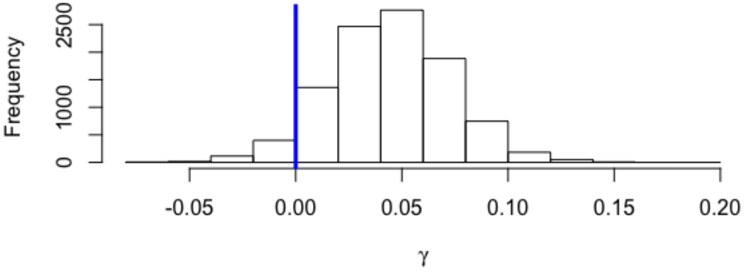
- $\sigma_d^2 = \sum_{i=d}^{10} a_i$ for $d = 1, \ldots, 10$.
  Note that this forces $\sigma_1^2 > \ldots > \sigma_{10}^2$
- $\mu_{wd} = \log(\text{Premium}_w) + logelr + \alpha_w + \beta_d \cdot (1 - \gamma)^{w-1}$

— Model

- Then $C_{wd} \sim \text{lognormal}(\mu_{wd}, \sigma_d)$

The Settlement Rate Parameter — γ

Figure: CSR Posterior Distribution of γ



CSR Output for the Paid Illustrative Loss Triangle

w	Premium	Estimate	SE	CV	Outcome	Percentile
1	5812	3912	0	0.0000	3912	
2	4908	2566	113	0.0440	2527	
3	5454	4139	189	0.0457	4274	
4	5165	4292	215	0.0501	4341	
5	5214	3516	192	0.0546	3583	
6	5230	3332	235	0.0705	3268	
7	4992	4971	426	0.0857	5684	
8	5466	3323	407	0.1225	4128	
9	5226	3756	742	0.1976	4144	
10	4962	3790	1416	0.3736	4139	
Total	52429	37597	2401	0.0639	40000	86.26

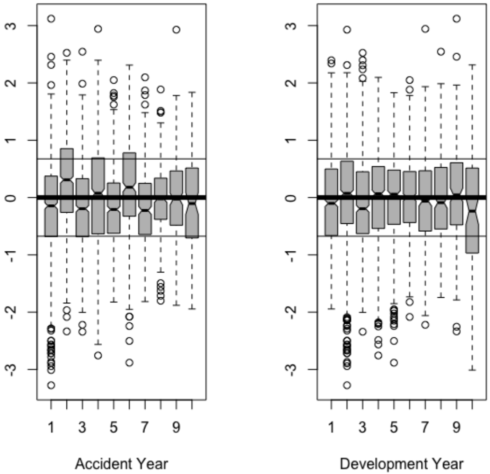
Model	$\widehat{elpd}_{loo}$	p_{loo}	LOOIC
CSR-Paid	49.76	15.09	-99.53

CRC Output for the Paid Illustrative Loss Triangle

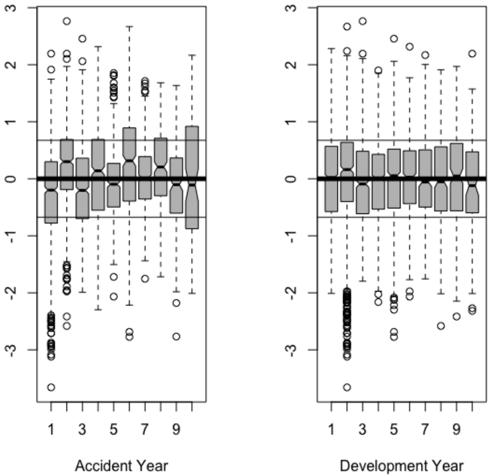
w	Premium	Estimate	SE	CV	Outcome	Percentile
1	5812	3912	0	0.0000	3912	
2	4908	2564	114	0.0445	2527	
3	5454	4149	193	0.0465	4274	
4	5165	4315	223	0.0517	4341	
5	5214	3566	203	0.0569	3583	
6	5230	3410	249	0.0730	3268	
7	4992	5208	445	0.0854	5684	
8	5466	3630	442	0.1218	4128	
9	5226	4392	817	0.1860	4144	
10	4962	4976	1762	0.3541	4139	
Total	52429	40121	2487	0.0620	40000	51.88

Model	$\widehat{elpd}_{loo}$	p_{loo}	LOOIC
CRC-Paid	47.80	14.97	-95.60

Standardized Residual Boxplots — CSR Model

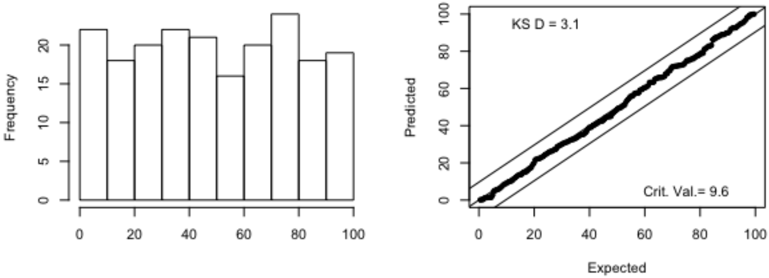


Standardized Residual Boxplots — CRC Model



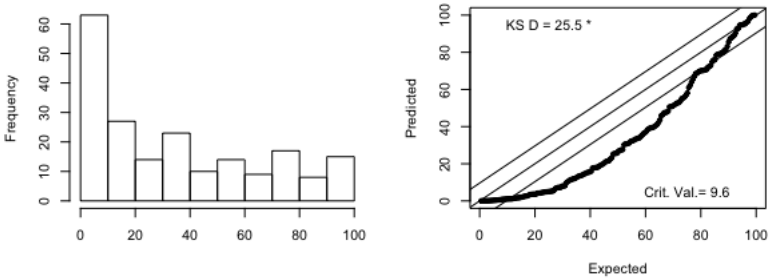
CSR PP Plot on 200 Paid Loss Triangles

Clearly a uniform distribution of percentiles



CRC PP Plot on 200 Paid Loss Triangles

CRC model is biased high



Using Incurred Data to Better Estimate Paid Losses

Table: 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: 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

The Correlated Accident Year (CAY) Model

— Parameters

- $logelr \sim$ normal(-0.4,$\sqrt{10}$).
- $\alpha_w \sim \text{normal}(0, \sqrt{10})$ for $w = 2, \ldots, 10$. $\alpha_1 = 0$.
- $\beta_d \sim \text{normal}(0, \sqrt{10})$ for $d = 1, \ldots, 9$. $\beta_{10} = 0$.
- $\rho_{pos} \sim \text{beta}(2, 2)$
- $a_i \sim \text{uniform}(0, 1)$ for $i = 1, \ldots, 10$.

— Transformed Parameters

- $\rho = 2 \cdot \rho_{pos} - 1$, This allows ρ to take on any value in the interval (-1,1).
- $\sigma_d^2 = \sum_{i=d}^{10} a_i$ for $d = 1, \ldots, 10$. This forces $\sigma_1^2 > \ldots > \sigma_{10}^2$.

The Correlated Accident Year (CAY) Model

— Transformed Parameters - continued

- $\mu_{1,d} = \log(\text{Premium}_1) + logelr + \beta_d$.
- $\mu_{wd} = \log(\text{Premium}_w) + logelr + \alpha_w + \beta_d$
  $+\rho \cdot (\log(C_{w-1,d}) - \mu_{w-1,d})$ for $w > 1$.

— Comment

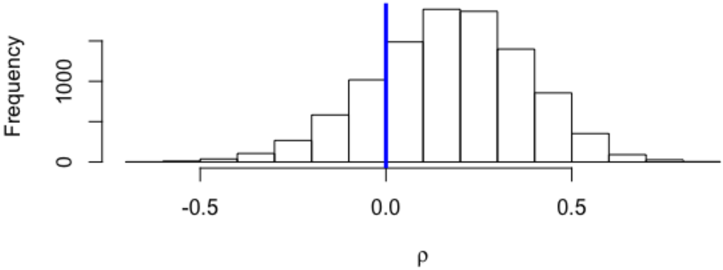
- This step generates a correlation between the accident years.

— Model

- Then $C_{wd} \sim \text{lognormal}(\mu_{wd}, \sigma_d)$.

Posterior Distribution of ρ - Illustrative Insurer

Figure: CAY Posterior Mean of ρ for the Set of 200 Paid Loss Triangles



CAY Model Output for the Incurred Illustrative Loss Triangle

w	Premium	Estimate	SE	CV	Outcome	Percentile
1	5812	3917	0	0.0000	3917	
2	4908	2547	65	0.0255	2532	
3	5454	4107	127	0.0309	4279	
4	5165	4308	144	0.0334	4341	
5	5214	3547	133	0.0375	3587	
6	5230	3329	152	0.0457	3268	
7	4992	5285	296	0.0560	5684	
8	5466	3790	323	0.0852	4128	
9	5226	4180	621	0.1486	4144	
10	4962	4183	1373	0.3282	4181	
Total	52429	39193	1859	0.0474	40061	73.24

- Plots, goodness of fit and holdout statistics for this model are in my monographs.
- Bottom line — CAY model is better fit for some, but not all triangles.

Integrated Paid and Incurred (IPI) Model

— Common Parameters

- *logelr* $\sim$ normal(-0.4,$\sqrt{10}$)
- $\alpha_w \sim$ normal$(0, \sqrt{10})$ for $w = 2, \ldots, 10$. $\alpha_1 = 0$

Integrated Paid and Incurred (IPI) Model

— Common Parameters

- $logelr \sim$ normal(-0.4,$\sqrt{10}$)
- $\alpha_w \sim \text{normal}(0, \sqrt{10})$ for $w = 2, \ldots, 10$. $\alpha_1 = 0$

— Paid Parameters

- ${}_P\beta_d \sim \text{normal}(0, \sqrt{10})$ for $d = 1, \ldots, 10$
- $\gamma \sim \text{normal}(0, 0.05)$
- ${}_Pa_i \sim \text{uniform}(0, 1)$ for $i = 1, \ldots, 10$

— Incurred Parameters

- ${}_I\beta_d \sim \text{normal}(0, \sqrt{10})$ for $d = 1, \ldots, 9$. ${}_I\beta_{10} = 0$
- $\rho_{pos} \sim$ beta(2,2)
- ${}_Ia_i \sim \text{uniform}(0, 1)$ for $i = 1, \ldots, 10$

Integrated Paid and Incurred (IPI) Model

— Transformed Parameters

- ${}_P\sigma_d^2 = \sum_{i=d}^{10} {}_P a_i$ for $d = 1, \ldots, 10$
- ${}_P\mu_{wd} = \log(\text{Premium}_w) + logelr + \alpha_w +_P \beta_d \cdot (1-\gamma)^{w-1}$
- .
- $\rho = 2 \cdot \rho_{pos} - 1$
- ${}_I\sigma_d^2 = \sum_{i=d}^{10} {}_I a_i$ for $d = 1, \ldots, 10$
- ${}_I\mu_{1,d} = \log(\text{Premium}_1) + logelr +_I \beta_d$
- ${}_I\mu_{wd} = \log(\text{Premium}_w) + logelr + \alpha_w +_I \beta_d$
  $+\rho \cdot (\log({}_I C_{w-1,d}) -_I \mu_{w-1,d})$ for $w > 1$

— Model

- ${}_I C_{wd} \sim \text{lognormal}({}_I\mu_{wd}, {}_I\sigma_d)$
- ${}_P C_{wd} \sim \text{lognormal}({}_P\mu_{wd}, {}_P\sigma_d)$

Summary of Common Parameters for IPI

	CSR Model		CAY Model		IPI Model	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
logelr	-0.3956	0.0246	-0.3945	0.0150	-0.3951	0.0109
α_1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
α_2	-0.2541	0.0272	-0.2619	0.0156	-0.2618	0.0090
α_3	0.1188	0.0308	0.1105	0.0213	0.1157	0.0119
α_4	0.2089	0.0373	0.2124	0.0253	0.2140	0.0153
α_5	-0.0002	0.0445	0.0083	0.0308	0.0091	0.0186
α_6	-0.0581	0.0617	-0.0586	0.0401	-0.0657	0.0263
α_7	0.3881	0.0787	0.4499	0.0521	0.4319	0.0383
α_8	-0.1097	0.1166	0.0248	0.0819	-0.0207	0.0619
α_9	0.0462	0.1914	0.1601	0.1453	0.1248	0.1056
α_{10}	0.0645	0.3467	0.1779	0.2984	0.1571	0.1947

Note the smaller standard deviations for the IPI model.

IPI Output for the Paid Illustrative Loss Triangle

w	Premium	Estimate	SE	CV	Outcome	Percentile
1	5812	3912	0	0.0000	3912	
2	4908	2543	58	0.0228	2527	
3	5454	4124	98	0.0238	4274	
4	5165	4309	111	0.0258	4341	
5	5214	3544	97	0.0274	3583	
6	5230	3299	109	0.0330	3268	
7	4992	5180	223	0.0431	5684	
8	5466	3612	234	0.0648	4128	
9	5226	4009	429	0.1070	4144	
10	4962	3986	796	0.1997	4139	
Total	52429	38518	1253	0.0325	40000	88.50

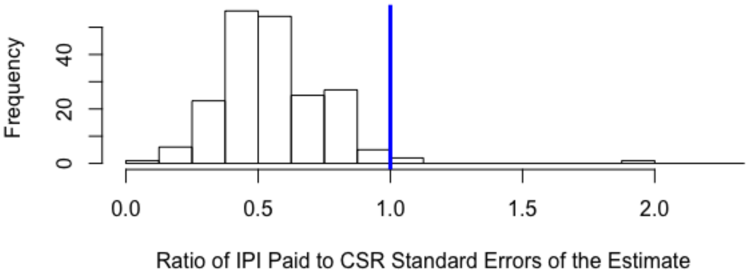
Model	$\widehat{elpd}_{loo}$	p_{loo}	LOOIC
IPI-Paid	63.54		-127.08

CSR Output for the Paid Illustrative Loss Triangle

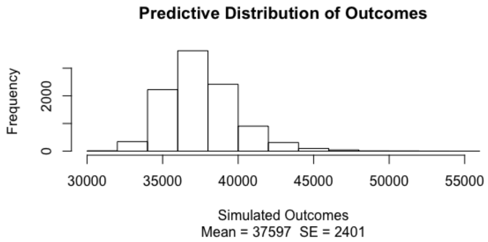
w	Premium	Estimate	SE	CV	Outcome	Percentile
1	5812	3912	0	0.0000	3912	
2	4908	2566	113	0.0440	2527	
3	5454	4139	189	0.0457	4274	
4	5165	4292	215	0.0501	4341	
5	5214	3516	192	0.0546	3583	
6	5230	3332	235	0.0705	3268	
7	4992	4971	426	0.0857	5684	
8	5466	3323	407	0.1225	4128	
9	5226	3756	742	0.1976	4144	
10	4962	3790	1416	0.3736	4139	
Total	52429	37597	2401	0.0639	40000	86.26

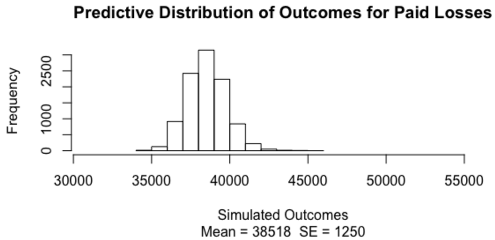
Model	$\widehat{elpd}_{loo}$	p_{loo}	LOOIC
CSR-Paid	49.76	15.09	-99.53

Standard Error Reductions by the IPI Model

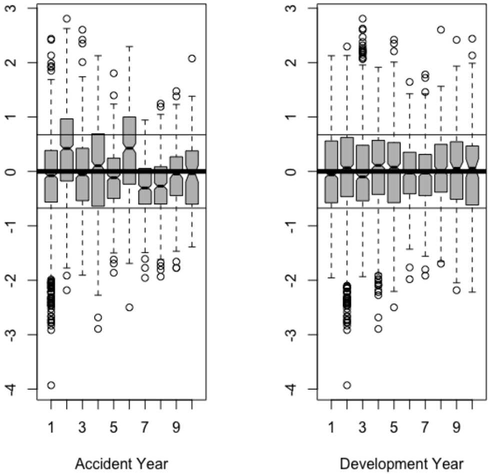


Predictive Distributions of Paid Outcomes CSR(top) and IPI(bottom)

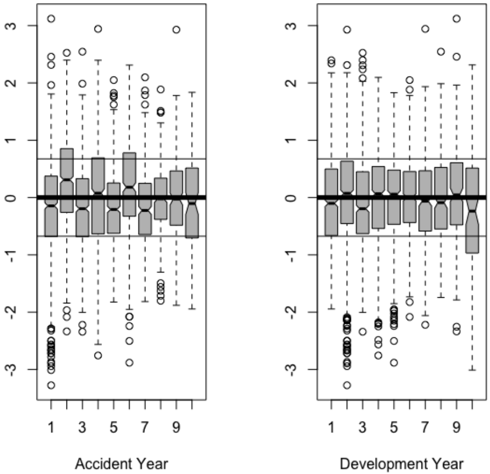




Standardized Residual Boxplots — IPI Paid Model

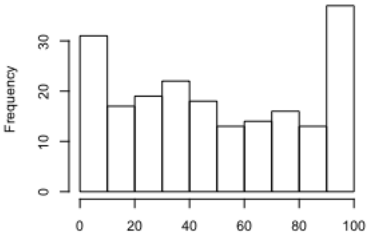


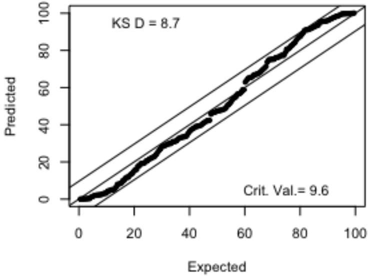
Standardized Residual Boxplots — CSR Model



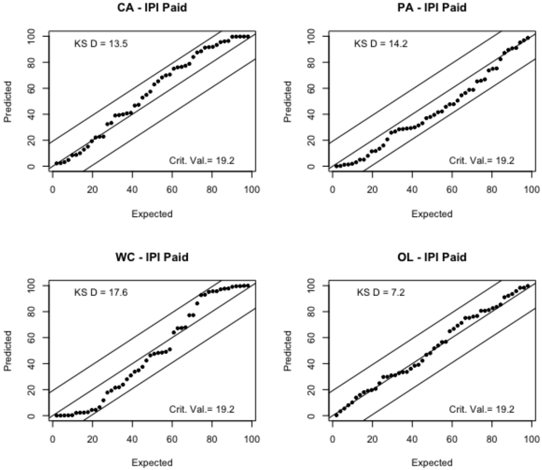
IPI PP Plot on 200 Paid Loss Triangles

Close to a uniform distribution of percentiles





IPI PP Plot on 4x50 Paid Loss Triangles



For WC, paid and incurred losses can be different.

$\widehat{elpd}_{loo}$ Paid Model Pairwise Comparisons

Line	IPI>CSR	IPI>CRC	CSR>CRC	CRC>SCC
CA	46	45	26	50
PA	41	42	27	50
WC	18	22	25	50
OL	41	40	23	50
Total	146	149	100	200

$\widehat{elpd}_{test}$ Paid Model Pairwise Comparisons

Line	IPI>CSR	IPI>CRC	CSR>CRC	CRC>SCC
CA	43	44	27	49
PA	42	44	30	47
WC	32	40	30	48
OL	39	42	32	47
Total	156	170	119	191

"Test" data are the lower triangle holdout data.

Summary

- Focused on paid loss triangles
- Four different models on paid triangles — IPI, CSR, CRC and SCC
- Provided diagnostics to help choose models
  - Real-time diagnostics — "loo" statistics and Standardized Residual Boxplots.
  - "Reputation" diagnostics on holdout data for several loss triangles — $\widehat{elpd}_{test}$ statistics and-PP Plots.
- We can rule out the "unadjusted" SCC.
- Reputation statistics tend to favor the IPI and CSR models. But there are enough counterexamples to suggest that real-time testing should be done.

On Prior Distributions

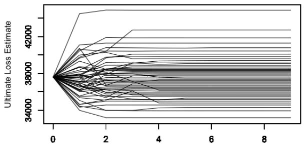
- In this work I tend to use wide or "weakly informative" priors. I like to leave room for surprises, but rule out "ridiculous" parameters. Ridiculous parameters can lead to numerical problems. See, for example, John Major's article on "Bayesian Dragons.
- Section 5 in my 2019 monograph discusses how I choose priors.
- Recommended reading —Prior Choice Recommendations by Andrew Gelman.

Prior Recommendations

- I recommend starting with wide proper priors checking for surprises.
- Then narrow the priors reflecting additional information.
- Be prepared to show and defend your initial run and runs with your priors to management, auditors and regulators.

Why Use a Stochastic Model?

- The parameters $\{\mu_{wd}, \sigma_d\}$ contain the information needed to plot alternative development paths.



- For each path calculate Capital→Released Capital→PV of Released Capital
- Cost of capital risk margin =

$$\text{Original Capital} - E[\text{PV of Released Capital}]$$

- Details in the 2nd edition of my monograph.