

## Combining Paid and Incurred Data in a Bayesian MCMC Model

One way to include more data into a loss reserve model is to use both paid and incurred data. Over the years, a number of authors have explored this idea. See, for example, Quarg and Mack (2004), Posthuma et al. (2008) and Venter (2008). Until recently, I have avoided that path because I have been using different models for paid and incurred data. All this changed when Ned Tyrrell, FCAS, showed me how to use the Bayesian MCMC language, Stan, to combine different models for paid and incurred data. Tyrrell's motivation for doing this was that using more data will reduce the range of possible outcomes. This article shows what happens when we combine the CSR and CCL models that are in my previous work, Meyers (2015).

Let's start by specifying my current versions of these models.

### The Changing Settlement Rate (CSR) Model for Paid Losses

1. Let  $C_{wd}^p$  be the cumulative paid loss for a 10 x 10 triangle for accident year  $w$  and development year  $d$ .
2. Let  $P_w$  be the earned premium for accident year  $d$ .
3. Let  $\alpha_w \sim Normal(0, \sqrt{10})$  for  $w = 2, \dots, 10$ . Set  $\alpha_1 \equiv 0$ .
4. Let  $\beta_d^p \sim Normal(0, \sqrt{10})$  for  $d = 1, \dots, 9$ . Set  $\beta_{10}^p \equiv 0$ .
5. Let  $logelr \sim Normal(0, \sqrt{10})$ .
6. Let  $\gamma \sim Normal(0, 0.05)$ .
7. Let  $a_i^p \sim Uniform(0, 1)$  for  $d = 1, \dots, 10$ . Then set  $(\sigma_d^p)^2 = \sum_{i=d}^{10} a_i^p$ . This forces  $\sigma_1^p < \sigma_2^p < \dots < \sigma_{10}^p$ .

8. Set  $\mu_{wd}^p = log(P_w) + logelr + \alpha_w + \beta_d^p \cdot (1-\gamma)^{w-1}$  for  $w = 1, \dots, 10$  and  $d = 1, \dots, 11-w$ .

9. Then  $C_{wd}^p \sim lognormal(\mu_{wd}^p, \sigma_d^p)$   
The expected loss ratio for accident year,  $w$ , after 10 years is given by  $exp(logelr + \alpha_w + (\sigma_{10}^p)^2/2) \approx exp(logelr + \alpha_w)$  since  $(\sigma_{10}^p)^2$  is generally very small.

While there are any number of equivalent ways to specify this model, I chose to formulate the model with a *logelr* parameter since many actuaries have access to prior information about the expected loss ratio for their business. They also expect market forces to change the expected loss ratio from year to year, and the  $\alpha_w$  parameters allow for these changes.

Once the model is coded, the Stan software will draw a sample from the posterior distribution of the parameters  $logelr, \{\alpha_w\}_{w=1}^{10}, \{\beta_d^p\}_{d=1}^{10}, \gamma$  and  $\{\sigma_d^p\}_{d=1}^{10}$ . Let's refer to this collection of parameters as  $\theta_p$ . With these parameters, one can calculate any statistic of interest to the actuary, such as the expected outcome and the standard deviation of the outcomes.

### The Correlated Chain Ladder (CCL) Model for Incurred Losses

1. Let  $C_{wd}^l$  be the cumulative incurred loss for a 10 x 10 triangle for accident year  $w$  and development year  $d$ .
2. Let  $P_w$  be the earned premium for accident year  $w$ .
3. Let  $\alpha_w \sim Normal(0, \sqrt{10})$  for  $w = 2, \dots, 10$ . Set  $\alpha_1 \equiv 0$ .

4. Let  $\beta_d^l \sim Normal(0, \sqrt{10})$  for  $d = 1, \dots, 9$ . Set  $\beta_{10}^l \equiv 0$ .
5. Let  $logelr \sim Normal(0, \sqrt{10})$ .
6. Let  $\rho \sim \beta(2, 2)$  scaled to go between -1 and 1, where  $\beta(\cdot, \cdot)$  denotes the  $\beta$  distribution.
7. Let  $a_i^l \sim Uniform(0, 1)$  for  $d = 1, \dots, 10$ . Then set  $(\sigma_d^l)^2 = \sum_{i=d}^{10} a_i^l$ . This forces  $\sigma_1^l < \sigma_2^l < \dots < \sigma_{10}^l$ .
8. Set  $\mu_{1d} = log(P_1) + logelr + \beta_d^l$  for  $d = 1, \dots, 10$ .
9. Set  $\mu_{wd}^l = log(P_w) + logelr + \alpha_w + \beta_d^l + \rho \cdot (log(C_{w-1,d}^l) - \mu_{w-1,d}^l)$  for  $w = 2, \dots, 10$  and  $d = 1, \dots, 11-w$ .
10. Then  $C_{wd}^l \sim lognormal(\mu_{wd}^l, \sigma_d^l)$

Again, once the model is coded, the Stan software will draw a sample from the posterior distribution of the parameters  $logelr, \{\alpha_w\}_{w=2}^{10}, \{\beta_d^l\}_{d=1}^{10}, \rho$  and  $\{\sigma_d^l\}_{d=1}^{10}$ . As above, let's refer to this collection of parameters as  $\theta_l$ .

A key assumption that we can make to combine these models is that the *logelr* and the  $\{\alpha_w\}_{w=1}^{10}$  parameters are the same for both the paid and incurred loss models. An additional modification, suggested to me by Ned Tyrrell, is to drop the assumption that  $\beta_{10}^l \equiv 0$ . This modification accounts for the fact that the case incurred losses recognize the further adjustments that could happen after the 10th development year.

The Stan software combines the paid and incurred models by adding the log-likelihoods,  $ll(\theta_p | \{C_{wd}^p\})$  and  $ll(\theta_l | \{C_{wd}^l\})$ . Stan then provides a sample from the posterior distribution of  $\theta_p$  and  $\theta_l$  in which the parameters *logelr* and  $\{\alpha_w\}_{w=1}^{10}$  are the same in both parameter sets

$\theta_p$  and  $\theta_l$ . One can then calculate the statistics of interest for both the paid and incurred triangles.

I ran each model for Commercial Auto Insurer #353 (The Illustrative Insurer in Meyers (2015)). Table 1 contains the loss estimates and the standard deviations for standalone CSR and CCL models. Table 2 contains the loss estimates and standard deviations for the combined CSR and CCL model.

Note that the standard deviations of the estimates for the combined model in Table 2 are smaller than the standard deviations standalone estimates in Table 1. To see how often this happens, I ran the combined and standalone models on 50 loss triangles in each of the Commercial Auto (CA), Personal Auto (PA), Workers' Compensation (WC) and Other Liability (OL) lines of insurance<sup>1</sup> Figure 1 shows a histogram of the standard deviation ratios for both the CSR and CCL models. The results show that the standard deviation is reduced in a clear majority of the cases, showing the positive effect of the additional data in reducing the uncertainty in the estimates.

While reducing the predictive standard deviation of the outcomes is desirable, it is not the goal of a stochastic loss reserve model. The goal is to correctly predict the distribution of outcomes. Following the methodology I proposed in Meyers (2015), Figures 2 and 3 test the predictive distribution on the observed outcomes by comparing the pp-Plots of the CSR and CCL models derived from the combined model with the corresponding plots from the standalone models. The CA and OL lines pass the test. PA just barely misses, but the

**Table 1. Standalone CSR and CCL Models.**

AV	Premium	CSR Estimate	CSR Std. Dev.	CCL Estimate	CCL Std. Dev.
1	5,812	3,912	0	3,917	0
2	4,908	2,562	107	2,549	65
3	5,454	4,134	181	4,110	125
4	5,165	4,283	208	4,312	143
5	5,214	3,514	193	3,549	131
6	5,230	3,319	222	3,332	155
7	4,992	4,965	414	5,291	304
8	5,466	3,311	406	3,803	337
9	5,226	3,739	746	4,180	619
10	4,962	3,747	1,407	4,116	1,261
<b>Table</b>	<b>52,429</b>	<b>37,485</b>	<b>2,338</b>	<b>39,159</b>	<b>1,774</b>

**Table 2. Combined Model for CSR and CCL.**

AV	Premium	CSR Estimate	CSR Std.Dev.	CCL Estimate	CCL Std.Dev.
1	5,812	3,912	0	3,912	0
2	4,908	2,544	65	2,549	53
3	5,454	4,123	112	4,132	92
4	5,165	4,304	123	4,314	106
5	5,214	3,557	110	3,563	97
6	5,230	3,389	131	3,395	122
7	4,992	5,209	260	5,218	248
8	5,466	3,790	297	3,797	293
9	5,226	4,094	474	4,101	469
10	4,962	3,802	794	3,809	796
<b>Total</b>	<b>52,429</b>	<b>38,722</b>	<b>1,343</b>	<b>38,796</b>	<b>1,294</b>

combined models perform better than the standalone models.

For WC, the combined model performs noticeably worse. I am not sure why this is the case, but it is worth noting that the difference between the paid and incurred losses is noticeably larger for WC than for the other lines.

While there are some questions that remain to be answered, I believe the combined models are worthy of further consideration.

The R/Stan scripts for the combined and standalone models are on the CAS

website along with summary statistics for all 200 loss triangles.

**References**

1. Meyers, Glenn G., *Stochastic Loss Reserving Using Bayesian MCMC Models*, CAS Monograph No. 1, 2015. <http://www.casact.org/pubs/monographs/index.cfm?fa=meyers-monograph01>
2. B. Posthuma, E.A. Cator, W. Veerkamp, and E.W. van Zwet, "Combined Analysis of Paid and Incurred Losses," *Casualty Actuarial*

<sup>1</sup> There have been a handful of changes from the set of insurers I used in Meyers(2015). The original set had a small number of "questionable" triangles. When the pp-Plots in Figures 2 and 3 are compared with the corresponding plots in Meyers (2015), the conclusions still hold.

*Society E-Forum*, Fall 2009. <http://www.casact.org/pubs/forum/08fforum/12Posthuma.pdf>

3. Quarg, Gerhard and Thomas Mack. 2004. "Munich chain-ladder - A reserving method that reduces

the gap between IBNR projections based on paid losses and IBNR projections based on incurred losses." *Blätter der DGVM*, Heft 4, 597-630.

4. Venter, Gary, "Distribution and

Value of Reserves Using Paid and Incurred Triangles," *Casualty Actuarial Society E-Forum*, Fall 2008. <http://www.casact.org/pubs/forum/08fforum/15Venter.pdf> ●

Figure 1

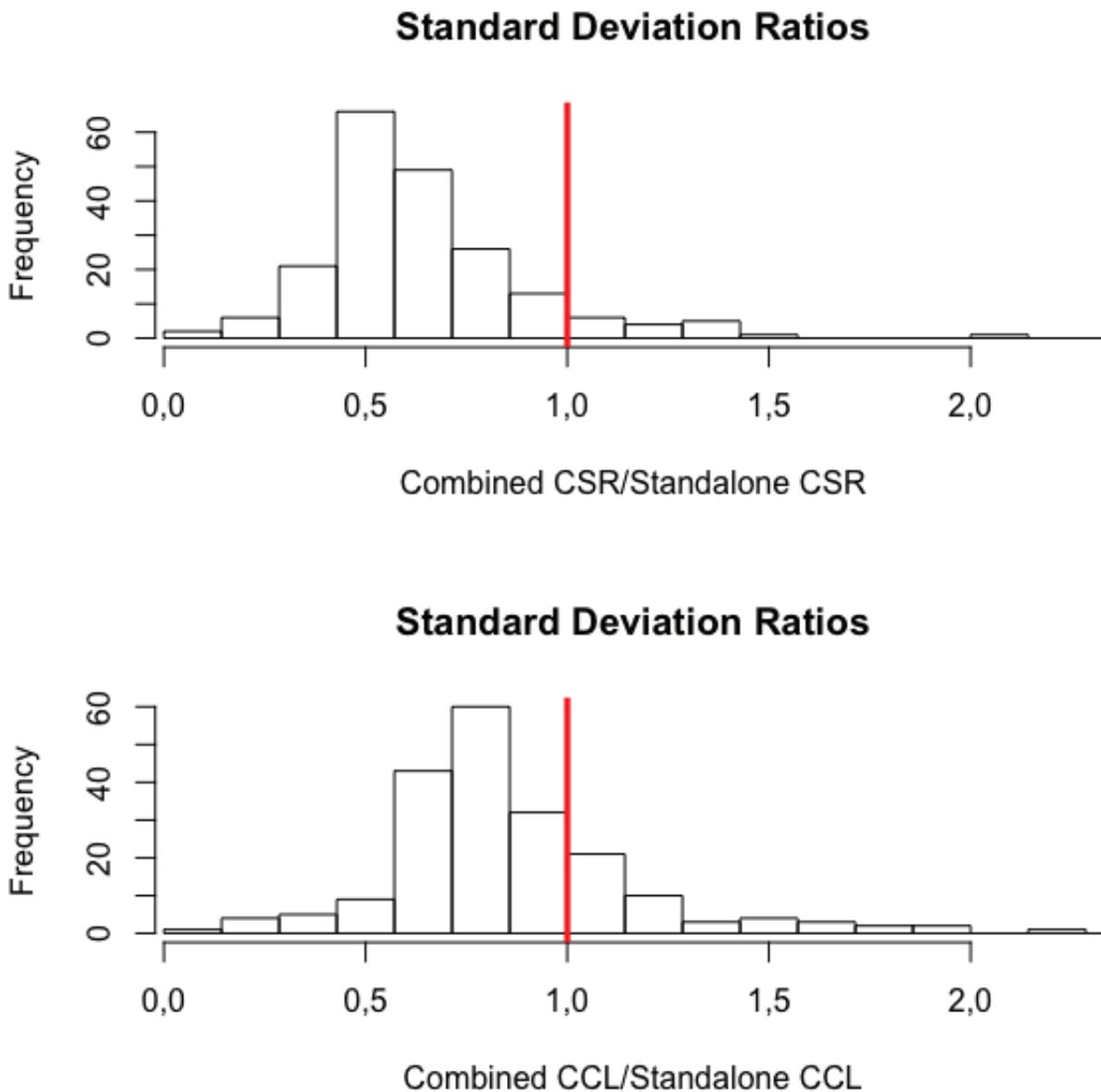


Figure 2

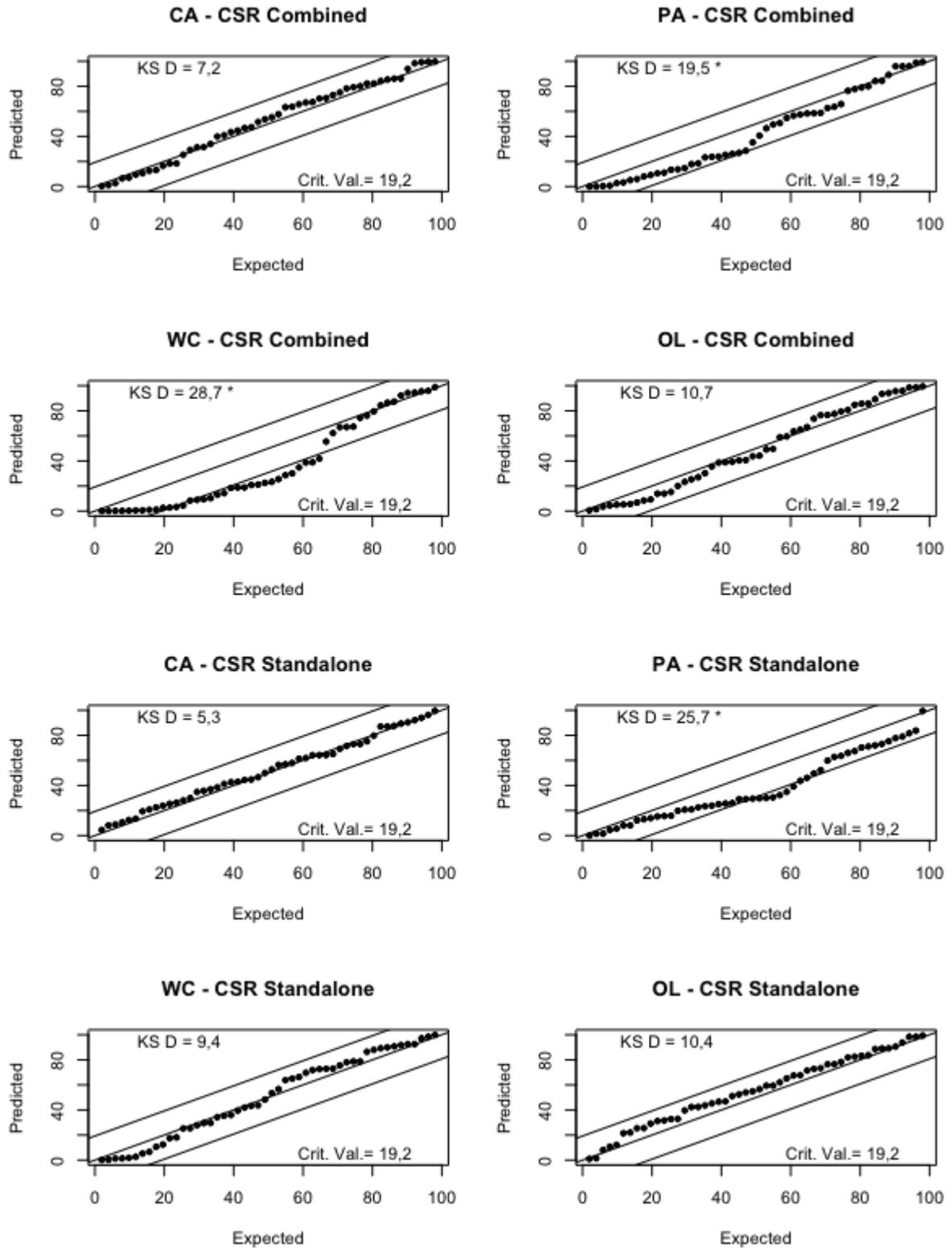


Figure 3

