EXPLORATIONS By GLENN MEYERS

Revisiting Cape Cod

The first time I visited Cape Cod was in the summer of 1970. We stayed near the Cape Cod National Seashore, enjoyed the beaches and wandered through the shops in Provincetown. Given that we were on a graduate student budget, we stayed at a campground in a classic 1960s-style umbrella tent, cooking our meals in an old Coleman stove fueled by white gasoline — quite Spartan by today's standards, but it worked well for us at the time.

Over a decade later, I became familiar with the actuarial Cape Cod. A prominent loss reserve formula dating from the early 1970s was the Bornhuetter-Ferguson method. This method, for paid losses, estimates the unpaid losses for a given accident year *w* by

$$\textit{Earned Premium}_w \bullet \textit{ELR} \bullet \textit{Expected Unpaid Losses}$$

where:

1. The expected unpaid loss is estimated by a standard loss reserve method, such as the chain ladder method.
2. The expected loss ratio (*ELR*) is to be judgmentally selected by the actuary.

When using the Bornhuetter-Ferguson method, many actuaries have felt the need to back up their judgment with a data-driven estimate of the expected loss ratio. In response, Hans Bühlmann and James Stanard developed a method to estimate this expected loss ratio in the early 1980s. (See Stanard 1985.) The method was so named as it sprang out of an actuarial conference held on Cape Cod. Over the years, it too has become a prominent loss reserve formula in the P&C actuary's toolkit.

Starting in 1990, statisticians began developing the statistical model-building methodology now known as Bayesian Markov Chain Monte Carlo (MCMC). Actuaries began looking at it in the early 2000s, and by 2005, the CAS recognized the potential of this methodology for building stochastic loss reserve models.[1]

It was not long after this that I got involved with stochastic loss reserve modeling with Bayesian MCMC. At first, the going was slow as I had a day job. But when I retired at the end of 2011, I was able to concentrate on it without a lot of other responsibilities.

One benefit of my late entry into the fray was that the MCMC methodology had begun to mature, and there was some terrific software that made it fairly easy to build new MCMC models. As I started modeling the Schedule P loss triangles in the CAS Loss Reserve Database, I quickly found myself building models that were different from the usual models that actuaries were using. This brings up the question: "How do we select which model to use?" The purpose of this article is to show how to select between alternative Bayesian MCMC models. We will look at two of the simpler models in the second edition of my monograph, "Stochastic Loss Reserving Using Bayesian MCMC Models." These models will describe the cumulative loss, C_{wd}, for accident year w and development year d. The first will be a CRoss-classified model by accident year and development year. The second will be a stochastic version of the actuarial Cape Cod model.

The CRoss-Classified (CRC) Model

1. $logelr \sim \text{Normal}(-0.4,\sqrt{10})$.
2. $\alpha_w \sim \text{Normal}(0,\sqrt{10})$ for $w = 2,\ldots,10$. Set $\alpha_1 = 0$.
3. $\beta_d \sim \text{Normal}(0,\sqrt{10})$ for $d = 1,\ldots,9$. Set $\beta_{10} = 0$.
4. $a_i \sim \text{Uniform}(0,1)$ for $i = 1,\ldots,10$.
5. Set $\sigma_d^2 = \Sigma_{i=d}^{10} a_i$ for $d = 1,\ldots,10$. Note that this forces $\sigma_1^2 > \ldots > \sigma_{10}^2$.
6. Set $\mu_{wd} = \log(\text{Premium}_w) + logelr + \alpha_w + \beta_d$.
7. Then $C_{wd} \sim \text{lognormal}(\mu_{wd},\sigma_d)$.

The Stochastic Cape Cod (SCC) Model

1. $logelr \sim \text{Normal}(-0.4,\sqrt{10})$.
2. $\beta_d \sim \text{Normal}(1,\sqrt{10})$ for $d = 1,\ldots,9$. Set $\beta_{10} = 0$.
3. $a_i \sim \text{Uniform}(0,1)$ for $i = 1,\ldots,10$.
4. Set $\sigma_d^2 = \Sigma_{i=d}^{10} a_i$ for $d = 1,\ldots,10$. Note that this forces $\sigma_1^2 > \ldots > \sigma_{10}^2$.
5. Set $\mu_{wd} = \log(\text{Premium}_w) + logelr + \beta_d$.
6. Then $C_{wd} \sim \text{Lognormal}(\mu_{wd},\sigma_d)$.

The difference between the two models is that while the SCC model forces a common expected loss ratio on all accident years, the CRC model allows the expected loss ratio to vary by accident year. As the prior distributions are fairly wide, the expected loss ratios are governed mainly by the data for both models.

The numerical examples in Table 1 in this article are identical to the numerical examples in my monograph, Meyers (2019), using the illustrative paid loss triangle from the commercial auto line of business.[2] The posterior means of the parameters for each model are in Table 1.

Table 1. Posterior Means of Parameters

Parameter	CRC	SCC
logelr	-0.3965	-0.4033
α_1	0.0000	
α_2	-0.2541	
α_3	0.1217	
α_4	0.2152	
α_5	0.0149	
α_6	-0.0343	
α_7	0.4354	
α_8	-0.0199	
α_9	0.2060	
α_{10}	0.3435	
β_1	-1.1999	-1.0897
β_2	-0.5751	-0.4926
β_3	-0.2825	-0.2155
β_4	-0.0954	-0.0170
β_5	-0.0628	-0.0439
β_6	-0.0170	0.0109
β_7	-0.0060	0.0214
β_8	-0.0038	-0.0418
β_9	-0.0056	-0.1251
β_{10}	0.0000	0.0000
σ_1	0.2965	0.4608
σ_2	0.2073	0.3691
σ_3	0.1334	0.3183
σ_4	0.0946	0.2853
σ_5	0.0730	0.2579
σ_6	0.0576	0.2351
σ_7	0.0472	0.2132
σ_8	0.0384	0.1887
σ_9	0.0300	0.1572
σ_{10}	0.0202	0.1051

Some observations on the parameters:

- There are jumps in the $\{\alpha_w\}$ parameters for the CRC model. This indicates that loss ratios are varying significantly by accident year parameters.
- The $\{\beta_d\}$ parameters for the SCC model do not gradually increase toward zero as the accident year matures. For a line of business like commercial auto, one would expect the upward development of the paid losses to gradually approach the ultimate loss.
- The $\{\sigma_d\}$ parameters are noticeably

[1] See Section 3.2.4 of the report of the CAS Working Party on Quantifying Variability in Reserve Estimates (2005).

[2] Additional model outputs that are not germane to this article are in the monograph.

larger for the SCC model. There remains a fair amount of uncertainty in the parameter estimates for the later development years in that model.

These observations highlight the fact that the SCC model is not simply a Bayesian MCMC version of the actuarial Cape Cod model. The principle difference is that the actuarial Cape Cod model first estimates the loss development factors (which are subject to the actuary's sense of being "reasonable.") The model then estimates the expected loss ratio. This is in contrast to the SCC which estimates all the parameters simultaneously. But as both models have a single parameter for the expected loss and the same number of development year parameters, one should expect the less constrained SCC model to have a better "fit."

So now let's consider our measure of fit. To shorten our notation, let $\{\theta^j\} = \{logelr^j, \alpha^j_{2:10}, \beta^j_{1:9}, \sigma^j_{1:10}\}$ denote the parameter set from the sample of size J from the posterior distribution of the CRC model. For the SCC model, drop the $\{\alpha^j_{2:10}\}$ from the $\{\theta^j\}$.

Given that we now have two models, we now discuss how we compare models using only the upper triangle data. Let's start the discussion with a review of the Akaike Information Criteria (AIC).

Suppose that we have a model with a data vector, $\mathbf{x} = \{x_i\}^N_{i=1}$, and a parameter vector θ, with p parameters. Let $\hat{\theta}$ be the parameter value that maximizes the log-likelihood, L, of the data, x. Then the AIC is defined as

$$\text{AIC} = 2 \cdot p - 2 \cdot L(\mathbf{x} \mid \hat{\theta}) \qquad (1)$$

Given a choice of models, the model with the lowest AIC is to be preferred. This statistic rewards a model for having a high log-likelihood, but it penalizes the model for having more parameters.

There are problems with the AIC in a Bayesian environment. Instead of a single maximum likelihood estimate of the parameter vector, there is an entire sample of parameter vectors taken from the model's posterior distribution. There is also the sense that the penalty for the number of parameters should not be as great in the presence of strong prior information. To address these concerns, Gelman et al. (2014, Chapter 7) describe statistics that generalize the AIC in a way that is appropriate for Bayesian MCMC models. Here is a brief overview of one of these statistics.

First, given a stochastic model, $p(\mathbf{x}|\theta)$, define the expected log predictive density as

$$elpd = \sum_{i=1}^{I} \log\left(\int p(x_i|\theta) \cdot f(\theta) d\theta\right) \qquad (2)$$

where f is the unknown density of θ.

If $\{\theta\}^J_{j=1}$ is a random sample from the posterior distribution of θ, define the computed log predicted density as

$$\widehat{lpd} = \sum_{i=1}^{I} \log\left(\frac{1}{J}\sum_{j=1}^{J} p(x_i|\theta^j)\right) \qquad (3)$$

Note that if we replace $\{\theta\}^J_{j=1}$ with the maximum likelihood estimate, $\hat{\theta}$, $\widehat{lpd}$ is equal to $L(\mathbf{x} \mid \hat{\theta})$ in Equation 1.

If the data vector, **x**, comes from a holdout sample — i.e., **x** was not used to generate the parameters, $\{\theta\}^J_{j=1}$ — then the $\widehat{lpd}$ is an unbiased estimate of *elpd*. But if the data vector, **x**, comes from the training sample, i.e., **x** was used to generate the parameters, $\{\theta\}^J_{j=1}$, then we expect $\widehat{lpd}$ to be higher than *elpd*. The amount of that bias is called the "effective number of parameters" which we denote by p.

Now let's consider what is called "leave one out cross validation" or "loo" for short. For the data point, x_i, one might obtain a sample of parameters $\{\theta_{(-i)}\}$ by an MCMC sample using all values of **x** except x_i. After doing this calculation for all observed data points in **x**, one can then use Equation 3 to calculate an unbiased estimate of the *epld*.

$$\widehat{elpd}_{loo} = \sum_{i=1}^{I} \log\left(\frac{1}{J}\sum_{j=1}^{J} p(x_i|\theta^j_{(-i)})\right) \qquad (4)$$

Methods to efficiently estimate $\widehat{elpd}_{loo}$ have been developed. Vehtari et al. (2017) provide the most up-to-date approaches that are incorporated in the R "loo" package.

When comparing two models, the model with the highest $\widehat{elpd}_{loo}$ should be preferred. For historical reasons, many prefer to state the results on the deviance scale, which similar to that of the AIC in Equation 1. This is done by writing:

$$\text{LOOIC} \equiv -2 \cdot \widehat{elpd}_{loo} = 2 \cdot p_{loo} - 2 \cdot \widehat{lpd} \qquad (5)$$

Table 2 provides the model comparison statistics for the illustrative triangle. These statistics strongly favor the CRC model. Moreover, when comparing the statistics for the models applied to the 50 commercial auto loss triangles in Meyers (2019), the CRC model is strongly favored for all 50 triangles.

Table 2: Model Comparison Statistics

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

An underlying assumption in the Bornhuetter-Ferguson and the Cape

Cod models is that the expected loss for each year is proportional to that accident year's premium. However, if that is known not to be the case, an actuary can adjust the premium to the level appropriate for that accident year. A near-perfect way to do this is to first run the CRC model and then multiply the premium for accident year w by $exp(\bar{\alpha}_w)$, where $\bar{\alpha}_w$ is the posterior mean of the $\{\alpha_w\}$ parameters obtained by fitting the CRC model.

The model comparison statistics for the illustrative triangle with this adjustment are in the first row of Table 3. They indicate that the adjustment leads to a strongly better fit. This is also true for the other 49 commercial auto triangles in the monograph. However, note that the adjustment came from the same data that we are fitting. A one-word description of this practice is "cheating!" What this exercise shows is that it is theoretically possible to adjust the premium so that the SCC obtains a better fit.

So what about in practice? A common rationale for adjusting the premium is the so-called underwriting cycle. There are 50 commercial auto triangles in the data used in my monograph. For each commerical auto auto-loss triangle, I adjusted the premium using an average $\bar{\alpha}_w$ where the average was taken from the remaining 49 loss triangles in our data. The model comparison statistics for the illustrative triangle with this second adjustment are in the second row of Table 3. They indicate that the unadjusted SCC model provides a better fit. When applied to the other commerical auto triangles in my monograph, I found that the unadjusted SCC fit better than the second adjusted SCC for 32 of the 50 loss triangles.

Table 3: Model Comparison Statistics

Model	$\widehat{elpd}_{loo}$	p_{loo}	LOOIC
SCC-Adj-1	67.09	9.59	-134.99
SCC-Adj-2	-7.73	9.25	15.45

My takeaway from this exercise is that while it is theoretically possible that some premium adjustment may save the stochastic Cape Cod, in practice it is going to be difficult. To any actuary considering a SCC-like model, I suggest also considering a CRC-like model. And, as I show in my monograph, there may be even better models.

When revisiting the actuarial Cape Cod model, I brought with me some very powerful tools that allowed this significantly improved fit. My laptop computer is, by several orders of magnitude, more powerful than the computers available to those who developed the original Cape Cod model. This computer power led to the development of the Bayesian MCMC technology, which produces the 29 x 10,000 array of parameters that computes our predictive distribution of ultimate losses. It also needs a 55 x 10,000 array of log-likelihoods that we use to evaluate the fit of the Bayesian model. With all this technology, one would hope we could improve our loss reserving methodology, and I think we have done so.

In spite of the powerful technology I used above, I found myself wondering why, with such an overwhelming difference, the actuarial profession had not noticed this problem before. Well, I have just attended the CLRS and found out that the problem has been noticed. See Spencer Gluck (1997). This paper allowed the estimated expected loss ratio to vary by accident year as a weighted average of the loss ratios for nearby accident years. The "generalized Cape Cod" model put forth in that paper formed the basis of a session by Jon Sappington and Enbo Jiang. This session showed how to use bootstrapping to calculate the variability of the estimates for the generalized Cape Cod model.

I have not been back to the geographic Cape Cod since our original visit 50 years ago. If I do revisit Cape Cod, I will insist on today's modern conveniences. It would be a nice hotel with a swimming pool, cable TV, fine dining at restaurants and, of course, free Wi-Fi.

References

1. CAS Working Party on Quantifying Variability in Reserve Estimates, "The Analysis and Estimation of Loss & ALAE Variability: A Summary Report," Casualty Actuarial Society *Forum*, Fall 2005.
2. Gluck, Spencer M., Balancing Development and Trend In Loss Reserve Analysis, Proceedings of the Casualty Actuarial Society Casualty Actuarial Society, 1997: LXXXIV, 482-532.
3. Meyers, Glenn G. 2019. "Stochastic Loss Reserving Using Bayesian MCMC Models" 2nd Edition, *CAS Monograph Series*, Number 8.
4. Stanard, James N. 1985. "A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques," *Proceedings of the Casualty Actuarial Society LXXII*, p. 124.
5. Vehtari, Aki, Andrew Gelman and Jonah Gabry, "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC," *Statistics and Computing*, 2017, 27:5, pp 1413-1432. (Check for updates.)