EXPLORATIONS BY GLENN MEYERS

Bayesian Model Selection

A common complaint I hear from classically trained statisticians when I discuss loss reserve models is that we should be careful of overfitting. As I have been writing about fitting models with over 30 parameters to a 10 x 10 loss triangle (with 55 observations), I must admit that, at least on the surface, this sounds pretty bad. My response has always been that if there were a loss reserve model with a small number of parameters "out there," someone would have found it by now. We need to deal with models with a large number of parameters.

I was drawn to Bayesian MCMC modeling because it is well equipped to handle these situations. Given a "sensible" model, it is possible to get a statistically valid predictive distribution of outcomes for any number of parameters. In fact, that is what I have done in my monograph *Stochastic Loss Reserving with Bayesian MCMC Models*[1] where I successfully validated stochastic loss reserve models on the holdout lower triangle data.

While a model's successful validation on 10-year-old data should be a consideration in deciding which model to use, I have been hearing from actuaries who are considering Bayesian MCMC models with fewer parameters on current data. This article discusses how to compare the performance of alternative Bayesian MCMC models on current data while taking the number of parameters into account.

Let's start the discussion with a review of the Akaike Information Criteria (AIC).

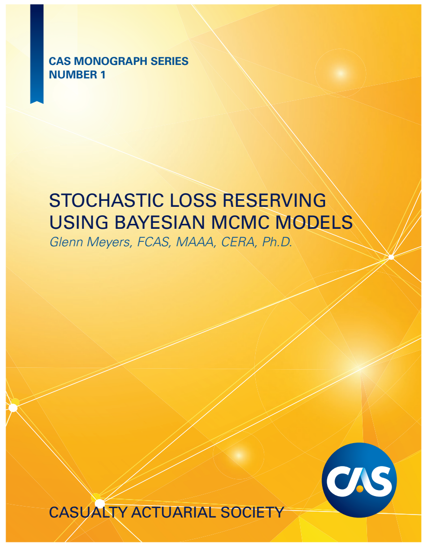
Suppose that we have a model with a data vector, $\mathrm{x} = \{x_n\}_{n-1}^{N}$ and a parameter vector θ with p parameters. Let $\hat{\theta}$ be the parameter value that maximizes the likelihood, L, of the data x. Then the AIC is defined as

$$\mathrm{AIC} = 2 \cdot \mathrm{p} - 2 \cdot \sum_{n-1}^{N} \log(\mathrm{L}(x_n|\hat{\theta})).$$

Given a choice of models, the model with the lowest AIC is usually 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 MCMC environment. Instead of a single maximum likelihood estimate of the parameter vector, there is an entire sample, $\{\theta_s\}_{s-1}^{S}$ 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 the parameters' informative priors or hierarchical structures or both.

To address these concerns, Gelman et al. describe a statistic, called the Watanabe-Akaike Information Criterion (WAIC) that generalizes the AIC in a way that is appropriate for Bayesian MCMC models.[2]



Stochastic Loss Reserving with Bayesian MCMC Models by Glenn Meyers is the first in the CAS series of Monographs. It can be downloaded at http://www.casact.org/pubs/index.cfm?fa=monographs.

First define the computed log pointwise predictive density as

$$L_{\mathrm{WAIC}} = \sum_{n-1}^{N} \log\left(\frac{1}{S}\sum_{s-1}^{S} L(x_n|\theta_s)\right).$$

The L_{WAIC} statistic replaces $\sum_{n-1}^{N} \log(\mathrm{L}(x_n|\hat{\theta}))$ in the expression for the AIC with the log of the average likelihood taken over the sample from the posterior distribution.

Next, define the effective number of parameters p_{WAIC} as

$$p_{\mathrm{WAIC}} = \sum_{n-1}^{N} Var_n[\log(L(x_n|\theta_s))].$$

p_{WAIC} has the property that it decreases with the tightness of the prior distribution. For a normal linear model with large sample size, known variance

[1] http://www.casact.org/pubs/monographs/index.cfm?fa=meyers-monograph01

[2] Gelman, Carlin, Stern, Denson, Vehtari and Rubin. Bayesian Data Analysis – Third Edition. CRC Press, Ch. 7.

Table 1 - Group 620 - Commercial Auto

AY	Premium	DY1	DY2	DY3	DY4	DY5	DY6	DY7	DY8	DY9	DY10
1	30,224	4,381	9,502	15,155	18,892	20,945	21,350	21,721	21,934	21,959	21,960
2	35,778	5,456	9,887	13,338	17,505	20,180	20,977	21,855	21,877	21,912	
3	42,257	7,083	15,211	21,091	27,688	28,725	29,394	29,541	29,580		
4	47,171	9,800	17,607	23,399	29,918	32,131	33,483	33,686			
5	53,546	8,793	19,188	26,738	31,572	34,218	35,170				
6	58,004	9,586	18,297	25,998	31,635	33,760					
7	64,119	11,618	22,293	33,535	39,252						
8	68,613	12,402	27,913	39,139							
9	74,552	15,095	27,810								
10	78,855	16,361									

Table 2 - Summary - Predictive Distributions of the Outcomes

Model	Estimate	Std. Dev.	L_{WAIC}	p_{WAIC}	WAIC
CSR	383,355	19,706	94.6	13.3	-162.61
ZSR	413,667	17,606	90.2	12.6	-155.24
SCC	402,803	22,629	40.8	8.0	-65.65

and uniform prior distribution of the coefficients, p_{WAIC} is approximately equal to p.

The final expression for the WAIC is analogous to that of the AIC and is given by

$$WAIC = 2 \cdot p_{WAIC} - 2 \cdot L_{WAIC}.$$

As with the AIC, the model with the lower WAIC is preferred.

Let's now show this calculation on the Changing Settlement Rate (CSR) model using the loss triangle in Table 1.

The CSR model is defined as follows:

1. $logelr \sim$ Uniform(-1, 0.5).
2. $\beta_d \sim$ Uniform(-5, 5) for $d = 1, \dots, 9$. $\beta_{10} = 0$.
3. $\gamma \sim$ Normal(0, 0.05).
4. $\alpha_w \sim$ Normal(log(Premium$_w$) + $logelr$, $\sqrt{10}$) for w = 1, …, 10.
5. $\sigma_d^2 = \sum_{i=d}^{10} a_i$, $a_i \sim$ Uniform(0, 1).
6. $\log(C_{wd}) \sim$ Normal($a_w + \beta_d(1-\gamma)^{w-1}$, σ_d)

Let's consider two simplifications to the model. The first simplification is to fix the settlement rate, γ=0. The second simplification is the set α_w = log(Premium$_w$) + $logelr$. Let's call the model with only the first simplification the Zero Settlement Rate (ZSR) model, and model with both simplifications the Stochastic Cape Cod (SCC) model as it forces the expected loss ratio to be the same for all accident years. The nominal number of parameters for the three models is 31, 30 and 20, respectively.

I then took a sample of size 10,000 from the posterior distribution of parameters for each of the models using Bayesian MCMC. Table 2 shows some summary statistics for the predictive distributions of the outcomes. The p_{WAIC}, L_{WAIC} and WAIC statistics are also given.

Subject to simulation error, we expect to see lower values of the log of the average likelihood, L_{WAIC}, for simpler models. We should also expect to see lower values of the effective number of parameters, p_{WAIC}, for simpler models. The model that is preferred depends upon the difference between the two statistics.

For this example, the CSR model (with a posterior mean $\gamma = 0.03$) is the preferred model. Behind it is the ZSR model, and way behind it is the SCC model. I have run these models on other insurers and found that, on some occasions, the ZSR is the preferred model.

The R scripts that produced these results are posted on the CAS website. Model changes were implemented by short modifications of the JAGS script that can be activated or removed by using comments. ●