EXPLORATIONS BY GLENN MEYERS

Dependencies in Stochastic Loss Reserve Models

My approach to the problem of correlations, or more generally, "dependencies," has usually appealed to some kind of causal modeling such as a common shock model. That changed at last November's CAS Centennial meeting when I attended a presentation by Yanwei (Wayne) Zhang about his paper, "Predicting Multivariate Insurance Loss Payments Using a Bayesian Copula Framework,"[1] written jointly with Vanja Dukic. The CAS awarded this paper the ARIA Prize. This annual prize, established in 1997 by the American Risk and Insurance Association, is made to the author or authors of a paper published by the *Journal of Risk and Insurance* that provides the most valuable contribution to casualty actuarial science.

The idea behind the Zhang-Dukic paper has a very simple high-level description. Suppose we have two Bayesian Markov chain Monte Carlo (MCMC) models, say Model 1 and Model 2. We can then use Bayesian MCMC to fit a joint (Model 1, Model 2) distribution. Zhang and Dukic expressed their bivariate distribution in its most general formulation as a copula, but any kind of bivariate distribution will work. Having just written a monograph for the CAS on Bayesian MCMC stochastic loss reserve models,[2] I thought I would give their approach a try on one of the models in my monograph.

I chose a changing settlement rate (CSR) model and applied it to paid 10 x 10 triangles taken from the commercial auto, and personal auto lines in the CAS Loss Reserve Database.[3] Following is a high-level description of this model.

Let w and d be subscripts for the accident year and development year, respectively. Let C_{wd}^{X} denote the cumulative paid loss for line X; $X = 1$ for commercial auto and $X = 2$ for personal auto. The univariate version of the model takes the following form:

$$\log(C_{wd}^{X}) \sim \text{normal}(\mu_{wd}^{X}, \sigma_{wd}^{X}),$$

where each μ_{wd}^{X} is a function of accident year and development year parameters. In all, there are 30 parameters in this model. There is a link to the full description of this model, along with the R/JAGS script, in the web version of this article.

The bivariate version of this model takes the following form.

$$\begin{pmatrix}\log(C_{wd}^{1})\\ \log(C_{wd}^{2})\end{pmatrix} \sim \text{normal}\left(\begin{pmatrix}\mu_{wd}^{1}\\ \mu_{wd}^{2}\end{pmatrix}, \begin{pmatrix}(\sigma_{d}^{1})^2 & \rho\sigma_{d}^{1}\sigma_{d}^{2}\\ \rho\sigma_{d}^{1}\sigma_{d}^{2} & (\sigma_{d}^{2})^2\end{pmatrix}\right).$$

This model has 30 parameters for each line of insurance, plus a correlation parameter ρ, for a total of 61 parameters. Fitting a Bayesian MCMC model yields a sample of 10,000 parameter sets from the posterior distribution. I ran this model on two insurers in the CAS Loss Reserve Database. Of particular interest is the correlation parameter, ρ. Figures 1 and 2 describe the posterior distribution of ρ for each insurer. Table 1 below gives some summary statistics of the predictive distribution of outcomes for the marginal distributions and the sum of losses in each line of insurance.

Table 1

Insurer #353	Net Premium	Expected Loss	S.D. Loss
Line 1 Marginal	52,429	37,845	1,824
Line 2 Marginal	155,061	126,439	2,018
Line 1 + Line 2	207,490	164,285	2,523
Insurer #388			
Line 1 Marginal	1,086,150	777,078	133,916
Line 2 Marginal	1,270,861	1,040,930	75,989
Line 1 + Line 2	2,357,011	1,818,009	155,108

I have run this model on several other insurers and found that these two insurers represent fairly well what happens with other insurers.

Based on my work to date on this topic, here are some general observations.

- My biggest surprise is that it is not uncommon for insurers to have negatively correlated logarithms of losses. As a quick reality check, I calculated the standard deviations of

[1] http://www.marsinsights.com/publication/bayesianCopulaOneComp.pdf
[2] http://www.casact.org/pubs/monographs/papers/01-Meyers.PDF
[3] http://www.casact.org/research/index.cfm?fa=loss_reserves_data

Figure 1

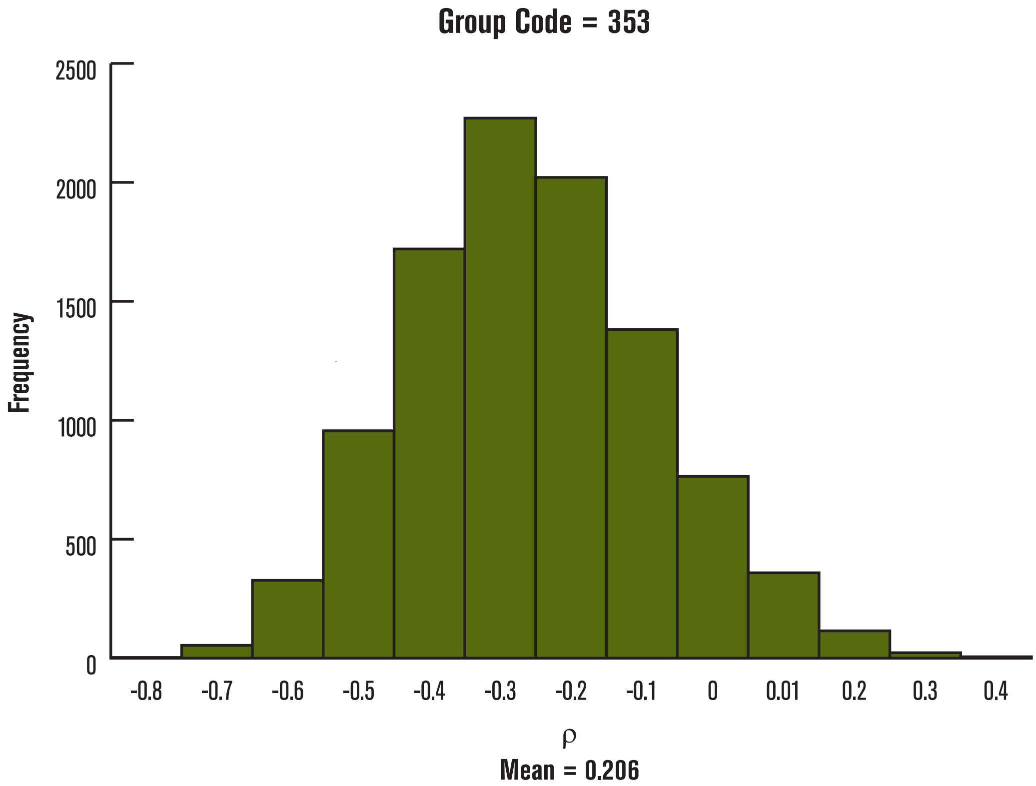
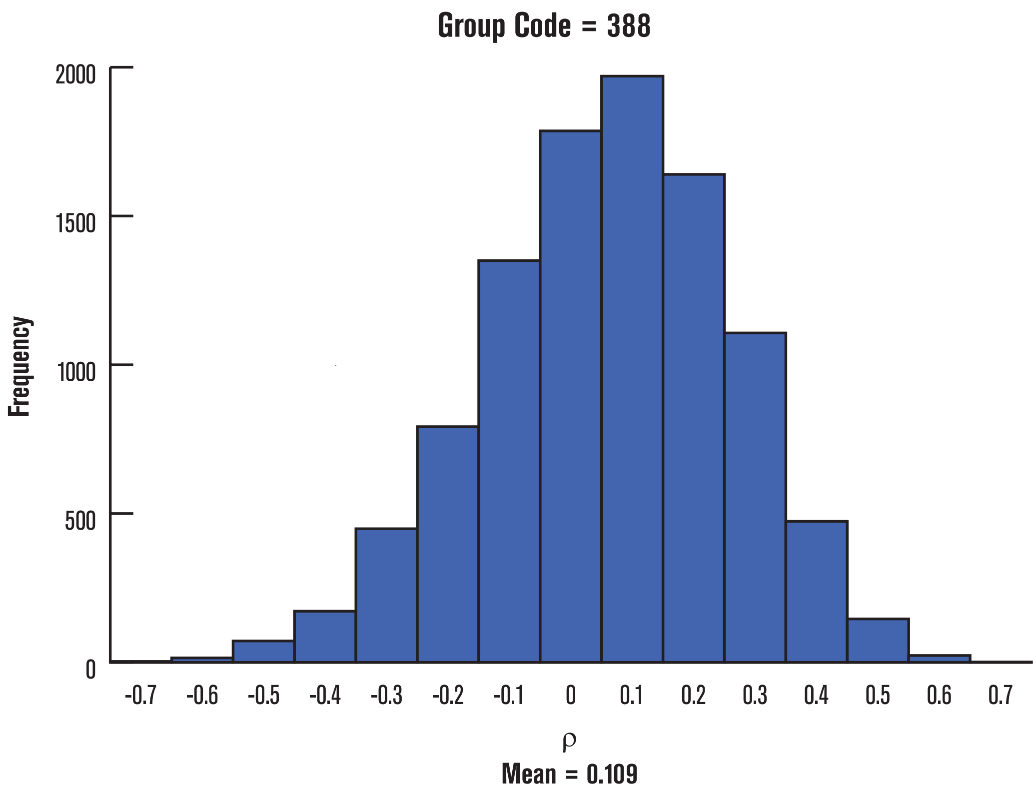


Figure 2



the total loss under the assumption of independence between lines and got following:

- 2,720 for Insurer #353. This is greater than the standard deviation obtained with the bivariate model, which is to be expected as the posterior mean coefficient of correlation is negative.
- 153,973 for insurer #388. This is less than the standard deviation obtained from the bivariate model, which is to be expected as the posterior mean coefficient of correlation is positive.
- If these results hold up under further scrutiny, it could imply that there is a sizeable diversification benefit for multiline insurers in various risk-based capital and liability risk margin regimes. For example, the liability risk margins under Solvency II are additive by line of insurance, which is tantamount to assuming that the lines of insurance are perfectly correlated.
- The prior distribution I chose for this model for each σ_d has a lighter tail than the prior distribution I used for the CSR model in my monograph. The results were unstable for the prior distribution I used there. I would very much like to have a model that worked well for all prior distributions. I am still thinking about how to handle prior distributions of σ_d with heavier tails.

As we can see, there is work that remains to be done. But nonetheless, we should all thank Zhang and Dukic for a fine piece of basic research that could have far reaching effects on solvency management for the insurance industry.