The Report of the Research Working Party on Correlations and Dependencies Among All Risk Sources
Part 3
The Common Shock Model for Correlated Insurance Losses

Glenn Meyers, FCAS, MAAA, Ph.D.

Abstract

This chapter discusses an approach to the correlation problem where losses in different lines of insurance are linked by a common variation (or shock) in the parameters of each line's loss model. The chapter begins with a simple common shock model and graphically illustrates the effect of the magnitude of the shocks on correlation. Next it describes some more general common shock models that involve common shocks to both the claim count and claim severity distributions. It derives formulas for the correlation between lines of insurance in terms of the magnitude of the common shocks and the parameters of the underlying claim count and claim severity distributions. Finally, it shows how to estimate the magnitude of the common shocks. A feature of this estimation is that it uses the data from several insurers.

1. Introduction

In the study of insurer enterprise risk management, "correlation" has been an important, but elusive phenomenon. Those who have tried to model insurer risk assuming independence have almost always understated the variability that is observed in publicly available data. Most actuaries would agree that "correlation" is the major missing link to the realistic modeling of insurance losses.

This chapter discusses an approach to the correlation problem where losses in different lines of insurance are linked by a common variation (or shock) in the parameters of each line's loss model. Here is an outline of what is to follow.

- I will begin with a simple common shock model and graphically illustrate the effect of the magnitude of the shocks on correlation.

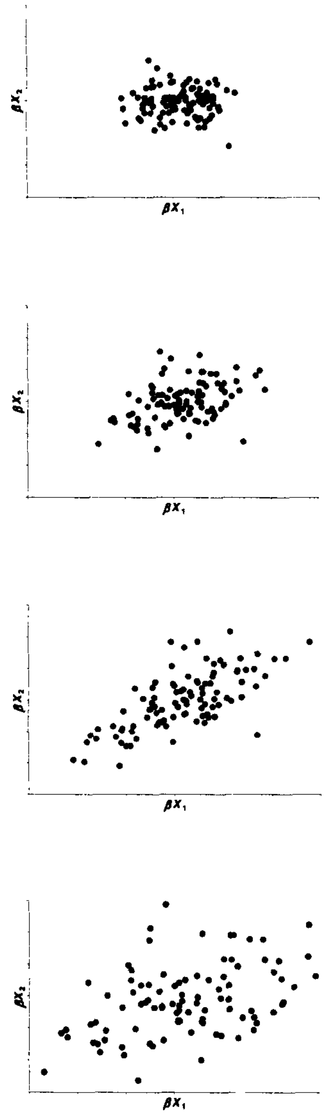
- Next I will describe some more general common shock models that involve common shocks to both the claim count and claim severity distributions. I will derive formulas for the correlation between lines of insurance in terms of the magnitude of the common shocks and the parameters of the underlying claim count and claim severity distributions.

- Finally, I will show how to estimate the magnitude of the common shocks. A feature of this estimation is that it uses the data from several insurers.

2. A Simple Common Shock Model

Let X_1 and X_2 be independent positive random variables. Also let β be a positive random variable with mean 1 and variance b. If $b > 0$, the random variables βX_1 and βX_2 tend to be larger when β is large, and tend to be smaller when β is small. Hence the random variables βX_1 and βX_2 are correlated. Figures 1-4 below illustrate this graphically.

I will refer to the β as the "common shock" and refer to the b as the magnitude of the common shocks. Figures 1-4 illustrate graphically that coefficient of correlation depend upon b and the volatility of the random variables X_1 and X_2.



βX2
βX1
βX2
βX1
βX2
βX1
βX2
βX1

I will now derive formulas for the coefficient of correlation between the random variables βX_1 and βX_2. This derivation will be detailed and I believe that it is worth the reader's time to master these details in order to appreciate much of what is to follow.

Let's begin with the derivation of two general equations from which I will derive much of what follows. These equations calculate the global covariance (or variance) in terms of the covariances (or variances) that are given conditionally on a parameter θ.

$$\begin{aligned} Cov[X,Y] &= E[X \cdot Y] - E[X] \cdot E[Y] \\ &= E_\theta\left[E[X \cdot Y \mid \theta]\right] - E_\theta\left[E[X \mid \theta]\right] \cdot E_\theta\left[E[Y \mid \theta]\right] \\ &= E_\theta\left[E[X \cdot Y \mid \theta] - E[X \mid \theta] \cdot E[Y \mid \theta]\right] \\ &\quad + E_\theta\left[E[X \mid \theta] \cdot E[Y \mid \theta]\right] - E_\theta\left[E[X \mid \theta]\right] \cdot E_\theta\left[E[Y \mid \theta]\right] \\ &= E_\theta\left[Cov[X,Y \mid \theta]\right] + Cov_\theta\left[E[X \mid \theta], E[Y \mid \theta]\right] \end{aligned} \tag{1}$$

An important special case of this equation occurs when $X = Y$.

$$Var[X] = E_\theta\left[Var[X \mid \theta]\right] + Var_\theta\left[E[X \mid \theta]\right] \tag{2}$$

Now let's apply Equations 1 and 2 to the common shock model given at the beginning of this section.

$$\begin{aligned} Cov[\beta X_1, \beta X_2] &= E_\beta\left[Cov[\beta X_1, \beta X_2 \mid \beta]\right] + Cov_\beta\left[E[\beta X_1 \mid \beta], E[\beta X_2 \mid \beta]\right] \\ &= E_\beta\left[\beta^2 Cov[X_1, X_2]\right] + Cov_\beta\left[\beta E[X_1], \beta E[X_2]\right] \\ &= E_\beta\left[\beta^2 \cdot 0\right] + E[X_1] \cdot E[X_2] \cdot Cov_\beta[\beta, \beta] \\ &= E[X_1] \cdot E[X_2] \cdot b \end{aligned} \tag{3}$$

$$\begin{aligned} Var[\beta X_1] &= E_\beta\left[Var[\beta X_1 \mid \beta]\right] + Var_\beta\left[E[\beta X_1 \mid \beta]\right] \\ &= E_\beta\left[\beta^2 \cdot Var[X_1]\right] + Var_\beta\left[\beta \cdot E[X_1]\right] \\ &= Var[X_1] \cdot E_\beta\left[\beta^2\right] + E[X_1]^2 \cdot Var_\beta[\beta] \\ &= Var[X_1] \cdot (1+b) + E[X_1]^2 \cdot b \end{aligned} \tag{4}$$

Similarly:

$$Var[\beta X_2] = Var[X_2] \cdot (1+b) + E[X_2]^2 \cdot b. \tag{5}$$

Equations 3-5 can now be plugged into the following equation for the coefficient of correlation.

$$\rho[\beta X_1, \beta X_2] = \frac{Cov[\beta X_1, \beta X_2]}{\sqrt{Var[\beta X_1] \cdot Var[\beta X_2]}} \tag{6}$$

Plugging Equations 3-5 into Equation 6 yields a simple expression if we give X_1 and X_2 identical distributions with a common coefficient of variation, CV.

$$\rho[\beta X_1, \beta X_2] = \frac{b}{(CV)^2 \cdot (1+b) + b} \tag{7}$$

The coefficients of correlation given in Figures 1-4 were calculated using Equation 7.

At this point we can observe that the common shock model, as formulated above, implies that the coefficient of correlation depends not only the magnitude of the shocks, but also the volatility of the distributions that receive the effect of the random shocks.

3. The Collective Risk Model

The collective risk model describes the distribution of total losses arising from a two-step process where: (1) the number of claims is random; and (2) for each claim, the claim severity is random. In this section I will specify a particular version of the collective risk model. In the next section I will subject both the claim count and claim severity distributions to common shocks across different lines of insurance and calculate the correlations implied by this model.

Let's begin by considering a Poisson distribution with mean λ and variance λ for the claim count random variable, N. Let χ be a random variable with mean 1 and variance c. The claim count distribution[1] for our version of the collective risk model will be defined by the two-step process where; (1) χ is selected at random; and (2) the claim count is selected at random from a Poisson distribution with mean $\chi\lambda$. The mean of this distribution is λ. I will refer to the parameter c as the contagion parameter.

[1] If χ has a gamma distribution, it is well known that this claim count distribution is the negative binomial distribution. None of the results derived in this paper will make use of this fact.

Using Equation 2, one calculates the variance of N as:

$$\begin{aligned} Var[N] &= E_\chi\left[Var[N \mid \chi]\right] + Var_\chi\left[E[N \mid \chi]\right] \\ &= E_\chi[\chi\lambda] + Var_\chi[\chi\lambda] \\ &= \lambda + c \cdot \lambda^2 \end{aligned} \tag{8}$$

Let Z_i be a random variable for claim severity for the i^{th} claim. We will assume that each Z_i is identically distributed with mean μ and variance σ^2. For random claim count, N, let:

$$X = Z_1 + \ldots + Z_N.$$

The mean of X is $\lambda\mu$. Using Equation 2 we calculate the variance of X as:

$$\begin{aligned} Var[X] &= E_N\left[Var[X \mid N]\right] + Var_N\left[E[X \mid N]\right] \\ &= E_N\left[N \cdot \sigma^2\right] + Var_N[N \cdot \mu] \\ &= \lambda \cdot \sigma^2 + \mu^2 \cdot \left(\lambda + c \cdot \lambda^2\right) \\ &= \lambda \cdot \left(\sigma^2 + \mu^2\right) + c \cdot \lambda^2 \cdot \mu^2 \end{aligned} \tag{9}$$

At this point, I would like to introduce a notion of risk size and specify my assumptions on how the parameters of this model change with risk size.

1. The size of the risk is proportional to the expected claim count, λ.
2. The parameters of the claim severity distribution, μ and σ, are the same for all risk sizes.
3. The contagion parameter, c, is the same for all risk sizes.

I do not claim that these assumptions are applicable to all situations. For example, increasing the size of an insured building will expose an insurer to a potentially larger property insurance claim.

I do believe these assumptions are applicable in the context of this chapter, enterprise risk management. As an insurer increases the number of risks that it insureds, its total expected claim count, λ, increases. If each risk that it adds on is similar to its existing risks, it is reasonable to expect μ and σ to be the same. One way to think of the contagion parameter, c, is as a measure of the uncertainty in the claim frequency. I believe it is reasonable to think this uncertainty applies to all risks simultaneously.

While a set of assumptions may sound reasonable, ultimately one should empirically test the predictions of such a model. I will do so below after I complete the description of my proposed model.

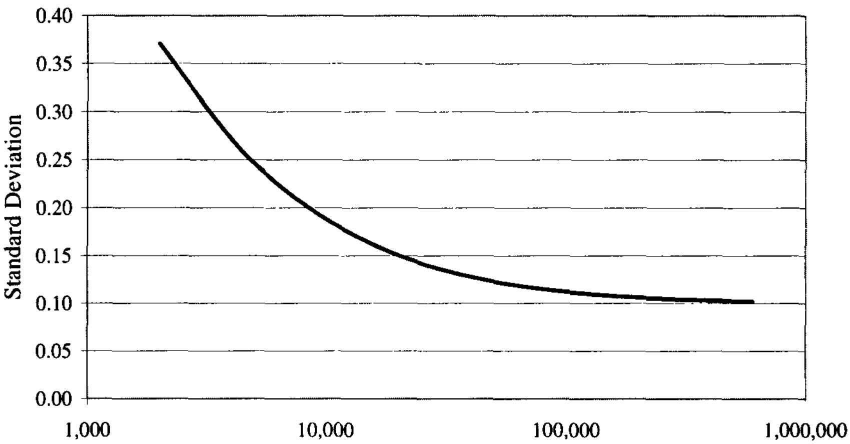
If the risk size is proportional to the expected claim count, λ, under the above assumptions it is also proportional to the expected loss $\lambda \cdot \mu$. In this chapter let's define the loss ratio as the ratio of the random loss X to its expected loss $E[X] = \lambda \cdot \mu$.

Equation 10 shows that the standard deviation of the loss ratio, $R = X/E[X]$ decreases asymptotically to $\sqrt{c}$ as we increase the size of the risk. Figure 5 below illustrates this graphically.

$$\text{Standard Deviation}[R] = \frac{\sqrt{\lambda \cdot \left(\sigma^2 + \mu^2\right) + c \cdot \lambda^2 \cdot \mu^2}}{\lambda \cdot \mu} \xrightarrow[\lambda \to \infty]{} \sqrt{c} \tag{10}$$

Figure 5

Loss Ratios for the Collective Risk Model
μ = 15,000 σ = 60,000 c = 0.01



4. Common Shocks in the Collective Risk Model

I will now apply the ideas underlying the common shock model described in Section 2, to the collective risk model described in Section 3. I will start with the claim count distributions.

Let N_1 and N_2 be two claim count random variables with $E[N_i] = \lambda_i$ and $Var[N_i] = \lambda_i + c_i \cdot \lambda_i^2$ for i = 1 and 2.

Let α be a random variable with $E[\alpha] = 1$ and $Var[\alpha] = g$.

I now introduce common shocks into the joint distribution of N_1 and N_2 by selecting N_1 and N_2 from claim count distributions with means $\alpha \cdot \lambda_1$ and $\alpha \cdot \lambda_2$ respectively and variances $\alpha \cdot \lambda_1 + c_1 \cdot (\alpha \cdot \lambda_1)^2$ and $\alpha \cdot \lambda_2 + c_2 \cdot (\alpha \cdot \lambda_2)^2$. Let's calculate the covariance matrix for N_1 and N_2.

Using Equation 2 to calculate the diagonal elements yields:

$$\begin{aligned} Var[N_i] &= E_\alpha\left[Var[N_i \mid \alpha]\right] + Var_\alpha\left[E[N_i \mid \alpha]\right] \\ &= E_\alpha\left[\alpha \cdot \lambda_i + c_i \cdot \alpha^2 \cdot \lambda_i^2\right] + Var_\alpha\left[\alpha \cdot \lambda_i\right] \\ &= \lambda_i + c_i \cdot \lambda_i^2 \cdot (1+g) + \lambda_i^2 \cdot g \\ &= \lambda_i + \lambda_i^2 \cdot (c_i + g + c_i \cdot g) \end{aligned} \tag{11}$$

Using Equation 1 to calculate the off-diagonal elements yields:

$$\begin{aligned} Cov[N_1, N_2] &= E_\alpha\left[Cov[N_1, N_2 \mid \alpha]\right] + Cov_\alpha\left[E[N_1 \mid \alpha], E[N_2 \mid \alpha]\right] \\ &= E_\alpha[0] + Cov_\alpha\left[\alpha \cdot \lambda_1, \alpha \cdot \lambda_2\right] \\ &= g \cdot \lambda_1 \cdot \lambda_2 \end{aligned} \tag{12}$$

Now let's add independent random claim severities, Z_1 and Z_2 to our common shock model. Here are the calculations for the elements of the covariance matrix for the total loss random variables X_1 and X_2.

$$\begin{aligned} Var[X_i] &= E_{N_i}\left[Var[X_i \mid N_i]\right] + Var_{N_i}\left[E[X_i \mid N_i]\right] \\ &= E_{N_i}\left[N_i \cdot \sigma_i^2\right] + Var_{N_i}\left[N_i \cdot \mu_i\right] \\ &= \lambda_i \cdot \sigma_i^2 + \mu_i^2 \cdot \left(\lambda_i + \lambda_i^2 \cdot (c_i + g + c_i \cdot g)\right) \\ &= \lambda_i \cdot \left(\sigma_i^2 + \mu_i^2\right) + \lambda_i^2 \cdot \mu_i^2 \cdot (c_i + g + c_i \cdot g) \end{aligned} \tag{13}$$

$$\begin{aligned} Cov[X_1, X_2] &= E_\alpha\left[Cov[X_1, X_2 \mid \alpha]\right] + Cov_\alpha\left[E[X_1 \mid \alpha], E[X_2 \mid \alpha]\right] \\ &= E_\alpha[0] + Cov_\alpha\left[\alpha \cdot \lambda_1 \cdot \mu_1, \alpha \cdot \lambda_2 \cdot \mu_2\right] \\ &= g \cdot \lambda_1 \cdot \mu_1 \cdot \lambda_2 \cdot \mu_2 \end{aligned} \tag{14}$$

Finally, let's multiply the claim severity random variables, Z_1 and Z_2, by a random variable β with $E[\beta] = 1$ and $\mathrm{Var}[\beta] = b$. Here are the calculations for the elements of the covariance matrix for the total loss random variables X_1 and X_2.

$$\begin{aligned}
Var[X_i] &= E_\beta\left[Var[X_i \mid \beta]\right] + Var_\beta\left[E[X_i \mid \beta]\right] \\
&= E_\beta\left[\lambda_i \cdot \beta^2 \cdot \left(\sigma_i^2 + \mu_i^2\right) + \lambda_i^2 \cdot \beta^2 \cdot \mu_i^2 \cdot \left(c_i + g + c_i \cdot g\right)\right] + Var_\beta\left[\lambda_i \cdot \beta \cdot \mu_i\right] \\
&= \left(\lambda_i \cdot \left(\sigma_i^2 + \mu_i^2\right) + \lambda_i^2 \cdot \mu_i^2 \cdot \left(c_i + g + c_i \cdot g\right)\right) \cdot E\left[\beta^2\right] + \lambda_i^2 \cdot \mu_i^2 \cdot Var[\beta] \\
&= \lambda_i \cdot \left(\mu_i^2 + \sigma_i^2\right) \cdot (1+b) + \lambda_i^2 \cdot \mu_i^2 \cdot \left(c_i + g + b + c_i \cdot g + c_i \cdot b + g \cdot b + c_i \cdot g \cdot b\right)
\end{aligned} \tag{15}$$

$$\begin{aligned}
Cov[X_1, X_2] &= E_\beta\left[Cov[X_1, X_2 \mid \beta]\right] + Cov_\beta\left[E[X_1 \mid \beta], E[X_2 \mid \beta]\right] \\
&= E_\beta\left[g \cdot \lambda_1 \cdot \beta \cdot \mu_1 \cdot \lambda_2 \cdot \beta \cdot \mu_2\right] + Cov_\beta\left[\lambda_1 \cdot \beta \cdot \mu_1, \lambda_2 \cdot \beta \cdot \mu_2\right] \\
&= g \cdot \lambda_1 \cdot \mu_1 \cdot \lambda_2 \cdot \mu_2 \cdot E\left[\beta^2\right] + \lambda_1 \cdot \mu_1 \cdot \lambda_2 \cdot \mu_2 \cdot Var[\beta] \\
&= \lambda_1 \cdot \mu_1 \cdot \lambda_2 \cdot \mu_2 \cdot (b + g + b \cdot g)
\end{aligned} \tag{16}$$

I now complete my description of this version of the collective risk model with the following two assumptions.

1. b and g are the same for all risk sizes.
2. b and g are the same for all lines of insurance.

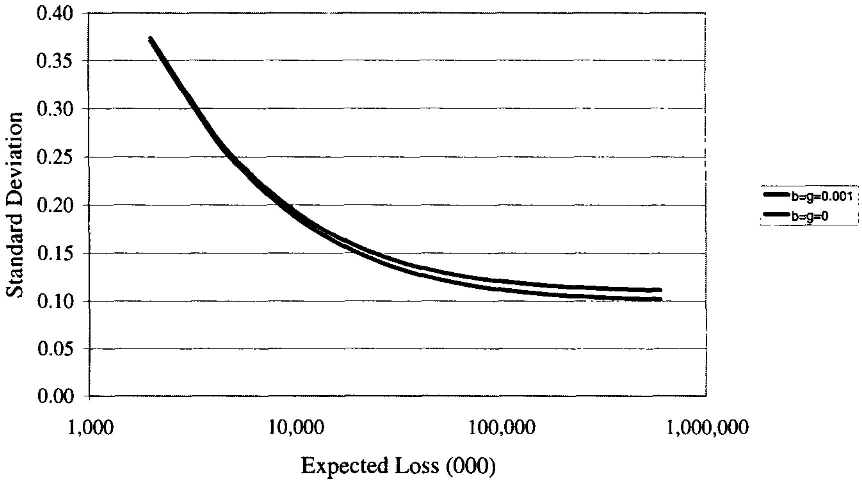
The parameters b and g represent parameter uncertainty that applies across lines of insurance and it seems reasonable to assume that this uncertainty is independent of the size of risk. I made the second assumption to keep the math simple without sacrificing the main themes of this chapter. In practice I have allowed g to vary by line of insurance. I will leave it as an exercise to the reader to show that you can replace g in Equations 14 and 16 with $\sqrt{g_1 \cdot g_2}$ when the coefficient of correlation between α_1 and α_2 is equal to one.

Now I will illustrate the implications of this model for loss ratios as we vary the size of risk. My example will assume that μ = 16,000, σ = 60,000 and c = 0.010 for each line of insurance. The additional parameters will be b = g = 0.001. In the final sections, I will show that these are reasonable choices of the parameters.

First let's note that since b and g are small compared to c, introducing b and g into the model has little effect on the standard deviation of the loss ratio, although what effect there is, increases with the size of the risk. This is illustrated by Figure 6.

Figure 6

Loss Ratios for the Collective Risk Model
μ = 15,000 σ = 60,000 c = 0.01

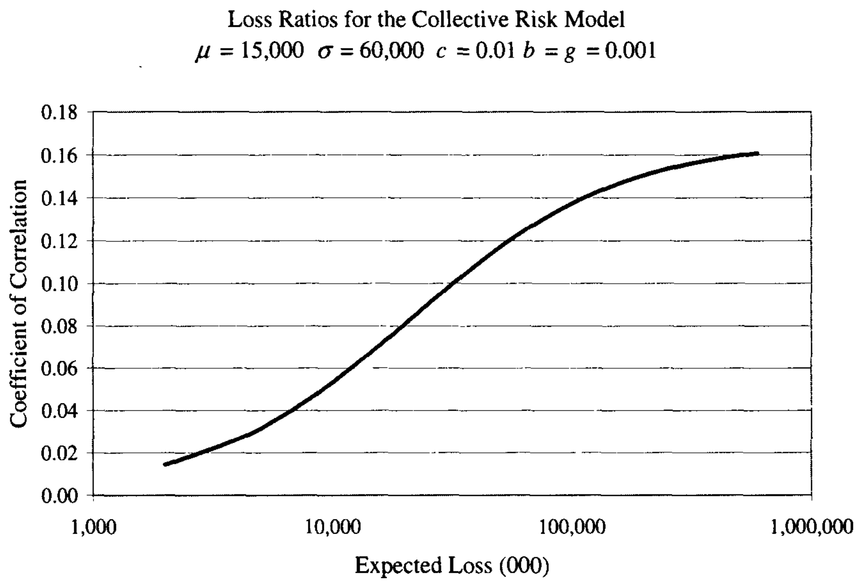


However, the coefficient of correlation, as defined by:

$$\rho[R_1, R_2] = \frac{Cov[R_1, R_2]}{\sqrt{Var[R_1] \cdot Var[R_2]}};$$

increases significantly as you increase the size of the risk. In Figure 7 below, it is almost negligible for small risks.

Figure 7



When I show similar exhibits to other actuaries, I often find that their expectations of the coefficient of correlations are much higher. My best rationale for these expectations is that most expect a positive number between 0 and 1, and 0.5 seems like a good choice.

Even so, these (perhaps) seemingly small correlations can have a significant effect for a multiline insurer seeking to manage its risk as I shall now illustrate.

Let's consider the covariance matrix for an insurer writing *n* lines of business.

$$\begin{pmatrix} Var[X_1] & Cov[X_1,X_2] & \dots & Cov[X_1,X_n] \\ Cov[X_2,X_1] & Var[X_2] & \dots & Cov[X_2,X_n] \\ \dots & \dots & \dots & \dots \\ Cov[X_n,X_2] & Cov[X_n,X_2] & \dots & Var[X_n] \end{pmatrix}$$

The standard deviation of the insurer's total losses, $X_1 + \dots + X_n$, is the square root of the sum of the elements of the covariance matrix. If $b = g = 0$, this sum consists of the n variances along the diagonal. If b and/or g $\neq 0$, then there are $n^2 - n$ off-diagonal covariances included in the sum. As n increases, so does the effect of even a "small" correlation. This is illustrated in Figures 8 and 9.

Figure 8

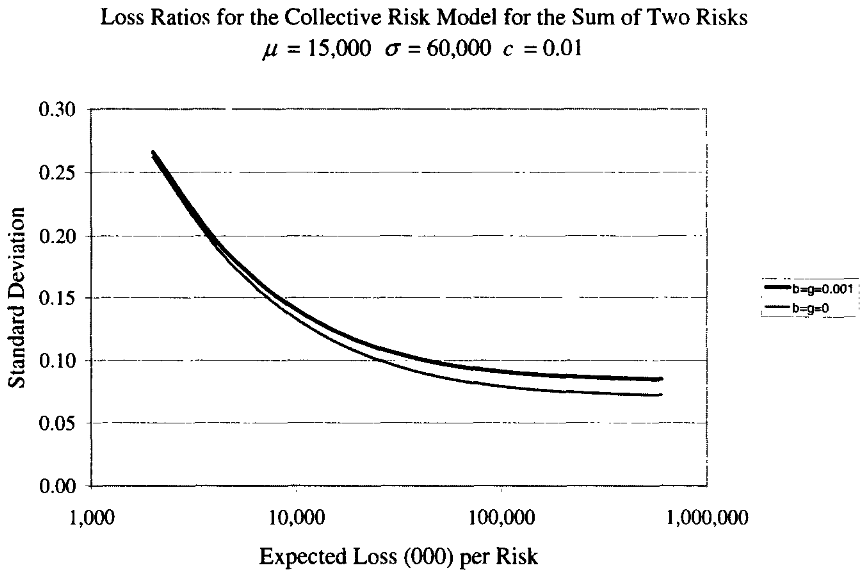
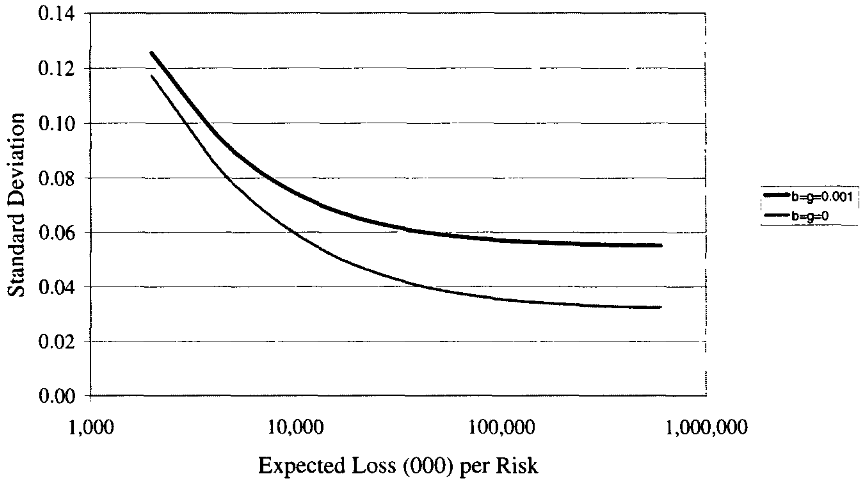


Figure 9

Loss Ratios for the Collective Risk Model for the Sum of Ten Risks
$\mu = 15{,}000 \quad \sigma = 60{,}000 \quad c = 0.01$



5. An Empirical Test of the Model

The collective risk model, as defined above, makes predictions about how the volatility and correlation statistics of loss ratios vary with insurer characteristics. These predictions should, at least in principle, be observable when one looks at a sizeable collection of insurance companies. In this section I will demonstrate that data that is publicly available on Schedule P is consistent with the major predictions of this model.

Data in Schedule P includes net losses, reported to date, and net premium by major line of insurance over a 10-year period of time. With Schedule P data for several insurers I calculated various statistics such as standard deviations and coefficients of correlation between lines of insurance for several insurers. Testing the model consisted of comparing these statistics with available information about each insurer.

But first I will discuss some of the difficulties with Schedule P data and discuss how, in work done jointly with Fred Klinker (see Meyers, Klinker and Lalonde [3] for details), we dealt with these difficulties.

Schedule P premiums and reserves vary in largely predictable ways due to conditions that are present in the insurance market. These conditions are often referred to as the underwriting cycle. The underwriting cycle contributes an artificial volatility to underwriting results that lies outside the statistical realm of insurance risk. The measures insurance managers take to deal with the statistical realm of insurance risk, i.e. reinsurance and diversification, are different than those measures they take to deal with the underwriting cycle.

We dealt with these difficulties by first using paid, rather than incurred, losses and estimating the ultimate incurred losses with industrywide paid loss development factors. Next we attempted to smooth out differences in loss ratios that we deemed "predictable." Appendix A in the Meyers *et. al.* paper referenced above describes this process in greater detail.

After making the above adjustments, two other difficulties should be discussed. First, the use of industrywide loss development factors removes the volatility that takes place after the report date of the loss. As such, we should expect the volatilities we measure to understate the ultimate volatility.

Second, Schedule P losses are reported net of reinsurance. In addition, policy limits are not reported. Rather than incorporate this information directly into our estimation, we did sensitivity tests of our model varying limits and reinsurance provisions over realistic scenarios.

Here I present results for commercial automobile liability insurance. I feel this is a good choice because: (1) it is a shorter tailed line than general liability and the underestimation of volatility will not be as great; (2) the use of reinsurance is not as great as it is in the general liability lines of insurance; and (3) commercial auto is not as prone to catastrophes as the property lines of insurance.

5.1 Standard Deviation of Loss Ratios vs. Size of Insurer

As illustrated in Figure 5, the collective risk model predicts that the standard deviation of insurer loss ratios should decrease as the size of the insurer increases. In Figure 10 we can see that this prediction is consistent with the observed standard deviations calculated from the Schedule P data described above. In this figure we plotted the empirical standard deviation of 55 commercial auto insurers against the average (over the 10 years of reported data) expected loss for the insurer[2].

Figure 10 also includes the standard deviations predicted by the collective risk model. The series denoted by "LowLim" used claim severity distribution parameters taken from a countrywide ISO claim severity distributions evaluated at the $300,000 occurrence limit. In this series I set $c = 0.007, g = 0.0005$ and $b = 0$. See Section 6 below for my commentary on selecting b and g.

Now we (at ISO) know from data reported to us that, depending on the subline (e.g. light and medium trucks or long-haul trucks), typically 65% to 90% of all commercial auto insurance policies are written at the $1 million policy limit. But since I also believed that the Schedule P data understates the true volatility of the loss ratios, I selected the $300,000 policy limit for the test.

For the sake of comparison, the series "HiLim" represents a judgmental adjustment that one might use to account for problems with the Schedule P data. I used claim severity distribution parameters taken from a countrywide ISO claim severity distributions evaluated at the $1,000,000 occurrence limit. In this series I set $c = 0.010, g = 0.0010$ and $b = 0$.

Figure 11 provides a comparable plot of loss ratios simulated from a collective risk model using the same parameters I used for the "LowLim" series.

The two plots both suggest that the Schedule P data is well represented by the collective risk mode — for an individual line of insurance.

---

[2] Since the expected loss varies by each observation of annual losses, the annual loss ratios are not identically distributed according to the collective risk model. I don't think this is a serious problem here since the volume of business is fairly consistent from year to year.

Figure 10

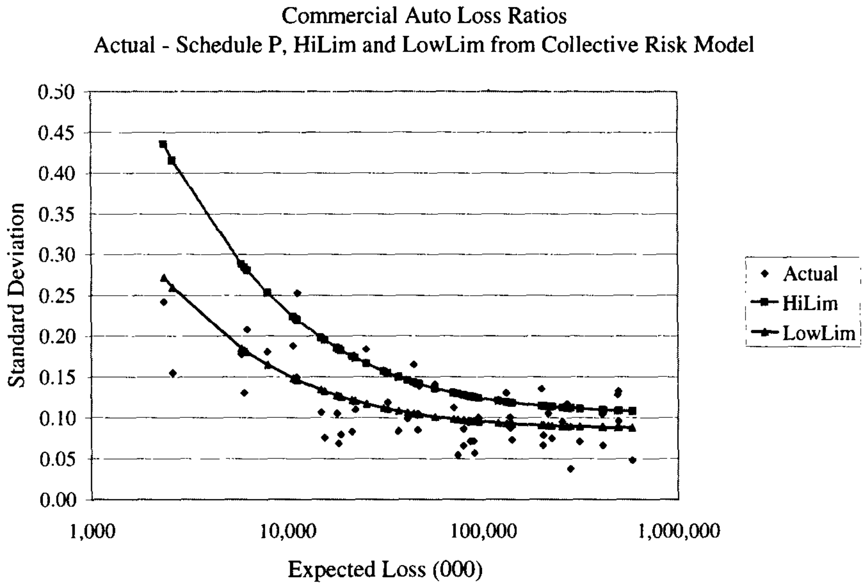
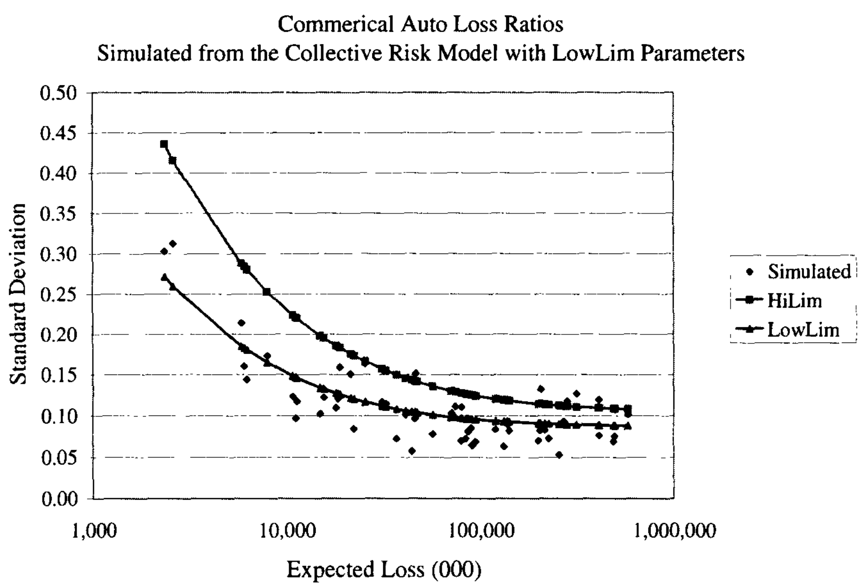


Figure11



5.2 Coefficients of Correlation vs. the Size of the Insurer

As Figure 7 illustrates, a second prediction of the collective risk model is that the coefficients of correlation will increase with the size of the insurer. In Figure 12 below, we plotted the empirical coefficient of correlation between commercial auto and personal auto for 38 insurers of both lines, against the average (over 10 years of experience reported for the two lines of insurance) expected loss. A comparable plot based on simulated data from the model underlying the URM is in Figure 13[3].

We observe that the coefficient of correlation is a very volatile statistic for both the empirical data and the simulated data which has a built-in assumption consistent with our hypothesis. This serves to illustrate the difficulty in measuring the effect of correlation in insurance data.

To provide a deeper analysis of the correlation problem I will make the assumption that the common shock random variables α and β operate on all insurers simultaneously. For random loss ratios R_1 and R_2:

$$E\left[(R_1-1)\cdot(R_2-1)\right]=\frac{Cov\left[X_1,X_2\right]}{\lambda_1\cdot\mu_1\cdot\lambda_2\cdot\mu_2}=b+g+b\cdot g; \tag{17}$$

which I derived from Equation 16.

Now we have already established that the standard deviation of loss ratios decreases with the size of the insurer. Thus the denominator of:

$$\rho\left[R_1,R_2\right]=\frac{E\left[(R_1-1)\cdot(R_2-1)\right]}{Std\left[R_1\right]\cdot Std\left[R_2\right]}$$

should decrease. If we can demonstrate with the Schedule P data, that the numerator does not also decrease, we can conclude that the prediction that coefficients of correlation will increase is consistent with the Schedule P data. It is to this we now turn.

[3] It may seem odd that the predicted correlation curve is not smooth. It is not smooth because the horizontal axis is the average of the commercial auto and the personal auto expected loss, while the actual split between the two expected losses varies significantly between insurers.

Figure 12

Commercial Auto and Personal Auto Loss Ratios
Actual – Schedule P, Predicted – Collective Risk Model

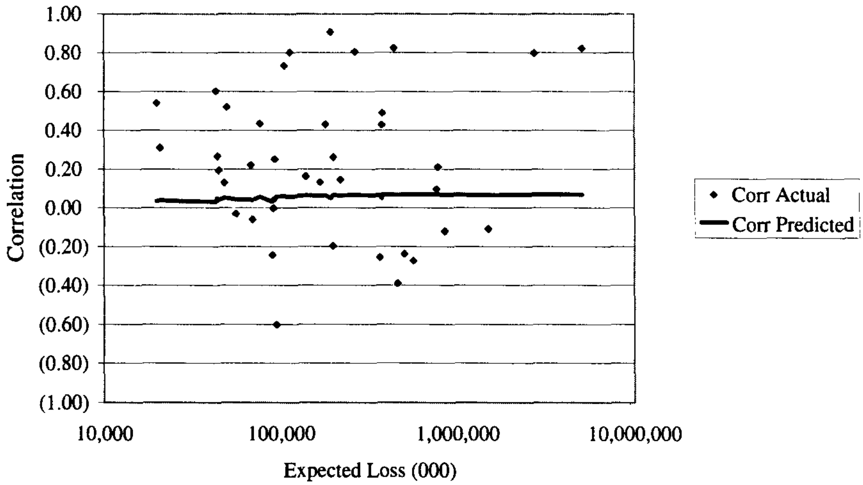
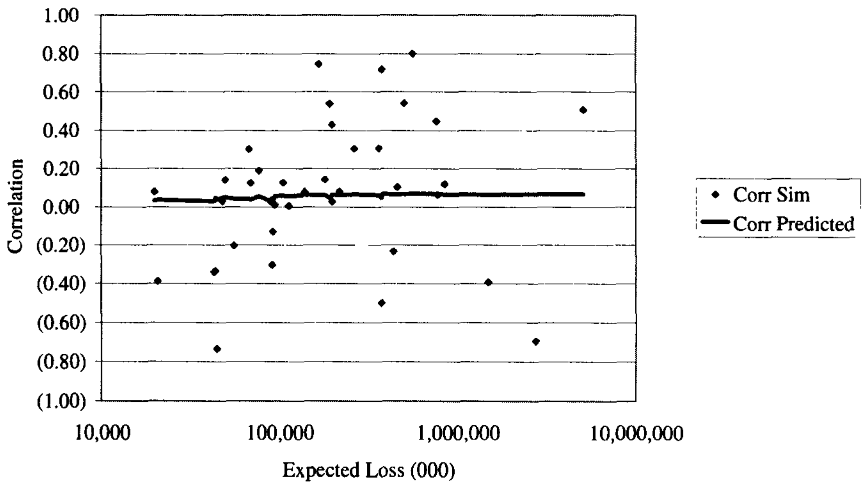


Figure 13

Commercial Auto and Personal Auto Loss Ratios
Simulated and Predicted from the Collective Risk Model



The data used in the test that $(R_1-1)\cdot(R_2-1)$ was independent of insurer size consisted of all possible pairs (15,790 in all) of r_1 and r_2, and the associated expected losses, taken from the same year and different insurers. I fit a line[4] to the ordered pairs

$$(\textit{Average Size of the Insurer}, (r_1-1)\cdot(r_2-1))$$

and obtained a slope of $+1.95\times10^{-10}$. This slightly positive slope means that an increasing coefficient of correlation is consistent with the Schedule P data.

Equation 17 also provides us with a way to estimate the quantity $b + g + b\,g$. One simply has to calculate the weighted average of the 15,970 products of $(r_1-1)\cdot(r_2-1)$, 0.00054. Since the 15,790 observations are not independent, the usual tests of statistical significance do not apply. To test the statistical significance of this result, I simulated 200 weighted averages using the "LowLim" parameters (except that $b = g = 0$) with the result that the highest weighted average was 0.000318. Thus we can reject the hypothesis that $b + g + b\,g = 0$.

I did one final simulation with the "LowLim" parameters (except that b = 0 and g = 0.00054) and calculated 200 slopes, with the result that the slope of 1.95×10^{-10} was just below the 49th highest. Thus this slope would not be unusual if the collective risk model is the correct model.

6. The Role of Judgment in Selecting Final Parameters

Historically, most actuaries have resorted to judgment in the quantification of correlation. This chapter was written in the hope of supplying some objectivity to this quantification. My employer, Insurance Services Office (ISO), has worked on quantifying this correlation. We have conducted analyses similar to the one described above for several lines of business using both Schedule P data and individual insurer data reported to ISO. In the end, no data set is perfect for the job and we end up making some judgments. Here are some of the considerations we made in selecting our final models. Comments are always welcome.

[4] I used a weighted least squares fit, using the inverse of the product of the predicted standard deviations of the loss ratio as the weights. This gives the higher volume, and hence more stable, observations more weight.

- We have reason to believe that the data we observe understates the ultimate variability since there are some claims that have yet to be settled. As a result we judgmentally increased the c, b and g parameters in the final model.

- Since the estimation procedure described provides an estimate of $b + g + b\,g$, it is impossible to distinguish between the claim frequency common shocks, as quantified by g, and the claim severity common shocks as quantified by b. A lot of work has been done with claim severity and claim frequency trend and one can look to uncertainties is these trends when selecting the final parameters.

- While one might argue that the distinction between claim frequency common shocks and claim severity common shocks is unimportant, the way we apply them does make a difference. For claim frequency we group the various lines of insurance judgmentally, with some support from the data. For example, the same common shock for claim frequency applies to personal and commercial auto, but different common shocks apply to the commercial liability lines. We apply claim severity shocks across all lines. Meyers, Klinker and Lalonde [2003] describe this model more fully.

Accounting data such as Schedule P may not be the best source for such analyses, but if we cannot see the effect of correlation in the accounting data, I would ask, do we need to worry about correlation? I believe that the analysis in this chapter demonstrates that we do need to consider correlation between lines of insurance.

7. Acknowledgements

This chapter is largely an exposition of work that appeared in a series of prior papers that I will now describe. A significant advance in the correlation literature was made by Shaun Wang [4] with the publication of his work on a project that was sponsored by the CAS. It is in this paper that I first heard the term "common shock model." I rather quickly followed up with two related papers. In Meyers[1], I originally developed the model that is described in Section 4 of this paper, and in Meyers[2] I developed methodology to parameterize the model with data that was "theoretically" available. A few years later we –

Meyers, Klinker and Lalonde [3] — followed up with another methodology to parameterize the model with data that was actually available. The original version of this methodology is described in Appendix A and Fredrick Klinker deserves the lion's share of the credit for developing it. I would described Section 5 as a minor improvement to this methodology.

References

[1] Meyers, Glenn G., 1999, "A Discussion of Aggregation of Correlated Risk Portfolios: Models & Algorithms, by Shaun S. Wang," 1999, *PCAS,* Vol. LXXXVI, 781-805. http://www.casact.org/pubs/proceed/proceed99/99705.pdf

[2] Meyers, Glenn G., "Estimating Between Line Correlations Generated by Parameter Uncertainty," *CAS Forum*, Summer 1999, 16-82. http://www.casact.org/pubs/forum/99sforum/99sf197.pdf

[3] Meyers, Glenn G., Fredrick L. Klinker and David A. Lalonde, "The Aggregation and Correlation of Insurance Risk," *CAS Forum*, Summer 2003, 16-82. http://www.casact.org/pubs/forum/03sforum/03sf015.pdf

[4] Wang, Shaun S., "Aggregation of Correlated Risk Portfolios: Models & Algorithms," *PCAS*, 1998, Vol. LXXXV, 848-939. http://www.casact.org/pubs/proceed/proceed98/980848.pdf

Biography

Glenn Meyers is the Chief Actuary for ISO Innovative Analytics. He holds a bachelor's degree in mathematics and physics from Alma College in Alma, Mich., a master's degree in mathematics from Oakland University, and a Ph.D. in mathematics from the State University of New York at Albany. Glenn is a Fellow of the Casualty Actuarial Society and a member of the American Academy of Actuaries. Before joining ISO in 1988, Glenn worked at CNA Insurance Companies and the University of Iowa.

Glenn's current responsibilities at ISO include the development of scoring products. Prior responsibilities have included working on ISO Capital Management products, increased limits and catastrophe ratemaking, ISO's, and Property Size-of-Loss Database (PSOLD), ISO's model for commercial property size of loss distributions.

Glenn's work has been published in *Proceedings of the Casualty Actuarial Society (CAS).* He is a three-time winner of the Woodward-Fondiller Prize, a two-time winner of the Dorweiller Prize and a winner of the Dynamic Financial Analysis Prize. He is a frequent speaker at CAS meetings and seminars.

His service to the CAS includes membership on various education and research committees. He currently serves on the International Actuarial Association Solvency Committee and the CAS Board of Directors.