

The Common Shock Model for Correlated Insurance Losses

by Glenn G. Meyers

ABSTRACT

This paper discusses an approach to the correlation problem in which losses from different lines of insurance are linked by a common variation (or shock) in the parameters of each line's loss model. The paper 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.

KEYWORDS

Collective risk model, internal models, hazard risk

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 in the realistic modeling of insurance losses.

This paper discusses an approach to the correlation problem in which losses from 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.

- A simple common shock model will graphically illustrate the effect of the magnitude of the shocks on correlation.
- The paper will describe some more general common shock models that involve common shocks to both the claim count and claim severity distributions. The paper 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, we will see how to estimate the magnitude of the common shocks. A feature of this estimation is that it uses the data from several insurers.

One should note that the common shock model is a kind of a “dependency” that causes correlation. It is possible for highly dependent random variables to be uncorrelated.

2. A simple common shock model

Let X_1 and X_2 be independent positive random variables. Also let β , which is independent of X_1 and X_2 , be a positive random variable with mean 1 and variance b . If $b > 0$, the random variables

Figure 1. X_1 and X_2 are independently drawn random variables with a coefficient of variation (CV) equal to 0.1. β is drawn from a distribution with $b = \text{Var}[\beta] = 0$. Thus $\rho = 0.00$

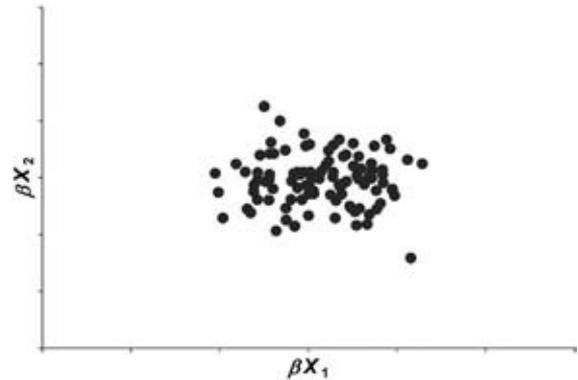
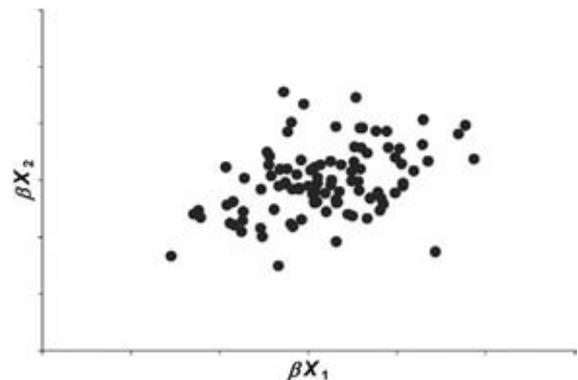


Figure 2. X_1 and X_2 are independently drawn random variables with $CV = 0.1$. β was drawn from a distribution with $b = \text{Var}[\beta] = 0.005$. Thus $\rho = 0.33$

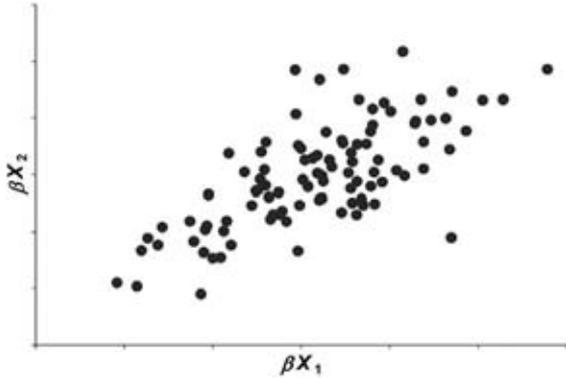


βX_1 and βX_2 tend to be larger when β is large, and tend to be smaller when β is small. Hence the random variables βX_1 and βX_2 are correlated. Figures 1 through 4 illustrate this graphically for the case in which X_1 and X_2 follow a gamma distribution.

Let us refer to β as the “common shock” and refer to b as the magnitude of the common shock. Figures 1 through 4 illustrate graphically that the coefficient of correlation depends upon b and the volatility of the gamma distributed random variables X_1 and X_2 .

We now turn to deriving formulas for the coefficient of correlation between the random vari-

Figure 3. X_1 and X_2 are independently drawn random variables with $CV = 0.1$. β was drawn from a distribution with $b = \text{Var}[\beta] = 0.020$. Thus $\rho = 0.66$



ables βX_1 and βX_2 . This derivation will be detailed. It is worth the reader's time to master these details in order to appreciate much of what is to follow.

Let us begin with the derivation of two general equations from which to 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 .

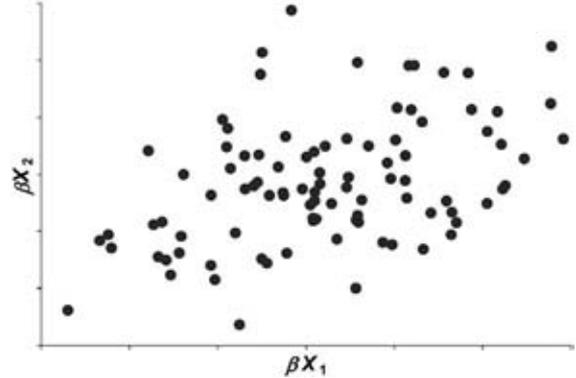
$$\begin{aligned}
 \text{Cov}[X, Y] &= E[X \cdot Y] - E[X] \cdot E[Y] \\
 &= E_{\theta}[E[X \cdot Y | \theta]] - E_{\theta}[E[X | \theta]] \cdot E_{\theta}[E[Y | \theta]] \\
 &= E_{\theta}[E[X \cdot Y | \theta]] - E_{\theta}[E[X | \theta] \cdot E[Y | \theta]] \\
 &\quad + E_{\theta}[E[X | \theta] \cdot E[Y | \theta]] \\
 &\quad - E_{\theta}[E[X | \theta]] \cdot E_{\theta}[E[Y | \theta]] \\
 &= E_{\theta}[\text{Cov}[X, Y | \theta]] + \text{Cov}_{\theta}[E[X | \theta], E[Y | \theta]].
 \end{aligned}
 \tag{2.1}$$

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

$$\text{Var}[X] = E_{\theta}[\text{Var}[X | \theta]] + \text{Var}_{\theta}[E[X | \theta]].
 \tag{2.2}$$

Now let us apply Equations (2.1) and (2.2) to the common shock model given at the beginning of this section.

Figure 4. X_1 and X_2 are independently drawn random variables with $CV = 0.2$. β was drawn from a distribution with $b = \text{Var}[\beta] = 0.020$. Thus $\rho = 0.33$



$$\begin{aligned}
 \text{Cov}[\beta X_1, \beta X_2] &= E_{\beta}[\text{Cov}[\beta X_1, \beta X_2 | \beta]] \\
 &\quad + \text{Cov}_{\beta}[E[\beta X_1 | \beta], E[\beta X_2 | \beta]] \\
 &= E_{\beta}[\beta^2 \text{Cov}[X_1, X_2 | \beta]] \\
 &\quad + \text{Cov}_{\beta}[\beta E[X_1], \beta E[X_2]] \\
 &= E_{\beta}[\beta^2 \cdot 0] + E[X_1] \cdot E[X_2] \\
 &\quad \cdot \text{Cov}_{\beta}[\beta, \beta] \\
 &= E[X_1] \cdot E[X_2] \cdot b.
 \end{aligned}
 \tag{2.3}$$

$$\begin{aligned}
 \text{Var}[\beta X_1] &= E_{\beta}[\text{Var}[\beta X_1 | \beta]] + \text{Var}_{\beta}[E[\beta X_1 | \beta]] \\
 &= E_{\beta}[\beta^2 \cdot \text{Var}[X_1]] + \text{Var}_{\beta}[\beta \cdot E[X_1]] \\
 &= \text{Var}[X_1] \cdot E_{\beta}[\beta^2] + E[X_1]^2 \cdot \text{Var}_{\beta}[\beta] \\
 &= \text{Var}[X_1] \cdot (1 + b) + E[X_1]^2 \cdot b.
 \end{aligned}
 \tag{2.4}$$

Similarly:

$$\text{Var}[\beta X_2] = \text{Var}[X_2] \cdot (1 + b) + E[X_2]^2 \cdot b.
 \tag{2.5}$$

The following equation defines the coefficient of correlation of two random variables, βX_1 and βX_2 .

$$\rho[\beta X_1, \beta X_2] = \frac{\text{Cov}[\beta X_1, \beta X_2]}{\sqrt{\text{Var}[\beta X_1] \cdot \text{Var}[\beta X_2]}}.
 \tag{2.6}$$

Combining Equations (2.3) through (2.5) with Equation (2.6) yields a simple expression for the correlation coefficient of βX_1 and βX_2 if we give X_1 and X_2 identical distributions with a common coefficient of variation, CV .

$$\rho[\beta X_1, \beta X_2] = \frac{b}{(CV)^2 \cdot (1 + b) + b}. \quad (2.7)$$

The coefficients of correlation given in Figures 1 through 4 were calculated using Equation (2.7).

At this point we can observe that the common shock model, as formulated above, implies that the coefficient of correlation depends not only on the magnitude of the shocks, but also on 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. This section specifies a particular version of the collective risk model. The next section subjects both the claim count and claim severity distributions to common shocks across different lines of insurance and calculates the correlations implied by this model.

It will be helpful to think of the “common shocks” as either inflation or judicial trends that affect two or more lines of insurance simultaneously. The more severe shocks such as those caused by natural catastrophes are not the focus of this paper.

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 distribution¹ for this version of the collective risk model will be defined by the two-step

¹If 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.

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 . Let’s refer to the parameter c as the contagion parameter.

Using Equation (2.2), one calculates the variance of N as:

$$\begin{aligned} \text{Var}[N] &= E_\chi[\text{Var}[N | \chi]] + \text{Var}_\chi[E[N | \chi]] \\ &= E_\chi[\chi\lambda] + \text{Var}_\chi[\chi\lambda] \\ &= \lambda + c \cdot \lambda^2. \end{aligned} \quad (3.1)$$

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:

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

The mean of X is $\lambda\mu$. Using Equation (2.2) we calculate the variance of X as:

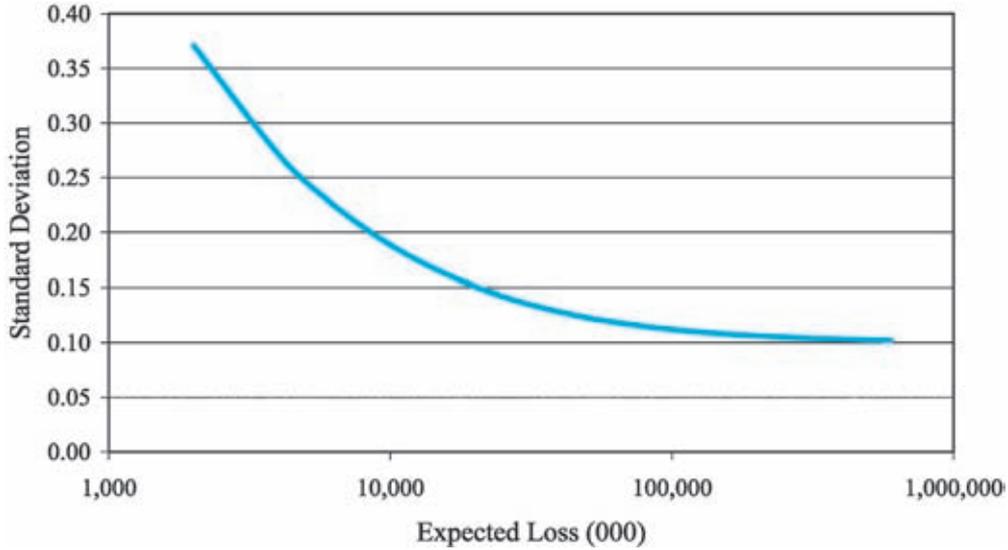
$$\begin{aligned} \text{Var}[X] &= E_N[\text{Var}[X | N]] + \text{Var}_N[E[X | N]] \\ &= E_N[N \cdot \sigma^2] + \text{Var}_N[N \cdot \mu] \\ &= \lambda \cdot \sigma^2 + \mu^2 \cdot (\lambda + c \cdot \lambda^2) \\ &= \lambda \cdot (\sigma^2 + \mu^2) + c \cdot \lambda^2 \cdot \mu^2. \end{aligned} \quad (3.2)$$

In this paper, let us define the size of risk as the expected loss of the risk. Let us also define the following 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. Note that this implies that the size of risk will be proportional to the expected claim count.
3. The contagion parameter, c , is the same for all risk sizes.

I believe that this is appropriate in the context of an insurer considering its volume in a particular line of business. It would not be appropriate in other contexts, such as property insurance where

Figure 5. Standard deviation of normalized loss ratios for the collective risk model $\mu = 15,000$ $\sigma = 60,000$
 $c = 0.01$



increasing the size of an insured building will expose the insurer to a potentially larger insurance claim.

I believe these assumptions are applicable in the context of this paper, which is enterprise risk management. As an insurer increases the number of risks that it insures, its total expected claim count, λ , increases. If each additional risk that it insures is similar to its existing risks, it is reasonable to expect that μ and σ will not change. 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. Such a test will be documented later in this paper.

Let us define the normalized loss ratio as the ratio of the random loss X to its expected loss $E[X] = \lambda \cdot \mu$. Defining it this way allows us to remove the cyclic effects of premium in the usual loss ratio. It is also worth noting that I did not normalize the loss ratio with respect to standard deviation.

Equation (3.3) shows that the standard deviation of the normalized loss ratio, $R = X/E[X]$ decreases asymptotically to \sqrt{c} as we increase the size of the risk. Figure 5 illustrates this graphically.

Standard Deviation[R] =

$$\frac{\sqrt{\lambda \cdot (\sigma^2 + \mu^2) + c \cdot \lambda^2 \cdot \mu^2}}{\lambda \cdot \mu} \xrightarrow{\lambda \rightarrow \infty} \sqrt{c}. \quad (3.3)$$

4. Common shocks in the collective risk model

This section applies the ideas underlying the common shock model described in Section 2 to the collective risk model described in Section 3. Let us 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 $\text{Var}[N_i] = \lambda_i + c_i \cdot \lambda_i^2$ for $i = 1$ and 2 .

Let α be a random variable, which is independent of N_1 and N_2 with $E[\alpha] = 1$ and $\text{Var}[\alpha] = g$.

Let us 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 $\alpha \cdot$

λ_1 and $\alpha \cdot \lambda_2$ and variances $\alpha \cdot \lambda_1 + c_1 \cdot (\alpha \cdot \lambda_1)^2$ and $\alpha \cdot \lambda_2 + c_2 \cdot (\alpha \cdot \lambda_2)^2$, respectively Let's calculate the covariance matrix for N_1 and N_2 .

Using Equation (2.2) to calculate the diagonal elements yields:

$$\begin{aligned} \text{Var}[N_i] &= E_\alpha[\text{Var}[N_i | \alpha]] + \text{Var}_\alpha[E[N_i | \alpha]] \\ &= E_\alpha[\alpha \cdot \lambda_i + c_i \cdot \alpha^2 \cdot \lambda_i^2] + \text{Var}_\alpha[\alpha \cdot \lambda_i] \\ &= \lambda_i + c_i \cdot \lambda_i^2 \cdot (1 + g) + \lambda_i^2 \cdot g \\ &= \lambda_i + \lambda_i^2 \cdot (c_i + g + c_i \cdot g). \end{aligned} \quad (4.1)$$

Using Equation (2.1) to calculate the off-diagonal elements yields:

$$\begin{aligned} \text{Cov}[N_1, N_2] &= E_\alpha[\text{Cov}[N_1, N_2 | \alpha]] \\ &\quad + \text{Cov}_\alpha[E[N_1 | \alpha], E[N_2 | \alpha]] \\ &= E_\alpha[0] + \text{Cov}_\alpha[\alpha \cdot \lambda_1, \alpha \cdot \lambda_2] \\ &= g \cdot \lambda_1 \cdot \lambda_2. \end{aligned} \quad (4.2)$$

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 :

$$\begin{aligned} \text{Var}[X_i] &= E_{N_i}[\text{Var}[X_i | N_i]] + \text{Var}_{N_i}[E[X_i | N_i]] \\ &= E_{N_i}[N_i \cdot \sigma_i^2] + \text{Var}_{N_i}[N_i \cdot \mu_i] \\ &= \lambda_i \cdot \sigma_i^2 + \mu_i^2 \cdot (\lambda_i + \lambda_i^2 \cdot (c_i + g + c_i \cdot g)) \\ &= \lambda_i \cdot (\sigma_i^2 + \mu_i^2) + \lambda_i^2 \cdot \mu_i^2 \cdot (c_i + g + c_i \cdot g). \end{aligned} \quad (4.3)$$

$$\begin{aligned} \text{Cov}[X_1, X_2] &= E_\alpha[\text{Cov}[X_1, X_2 | \alpha]] \\ &\quad + \text{Cov}_\alpha[E[X_1 | \alpha], E[X_2 | \alpha]] \\ &= E_\alpha[0] + \text{Cov}_\alpha[\alpha \cdot \lambda_1 \cdot \mu_1, \alpha \cdot \lambda_2 \cdot \mu_2] \\ &= g \cdot \lambda_1 \cdot \mu_1 \cdot \lambda_2 \cdot \mu_2. \end{aligned} \quad (4.4)$$

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

X_2 :

$$\begin{aligned} \text{Var}[X_i] &= E_\beta[\text{Var}[X_i | \beta]] + \text{Var}_\beta[E[X_i | \beta]] \\ &= E_\beta[\lambda_i \cdot \beta^2 \cdot (\sigma_i^2 + \mu_i^2) + \lambda_i^2 \cdot \beta^2 \cdot \mu_i^2 \\ &\quad \cdot (c_i + g + c_i \cdot g)] + \text{Var}_\beta[\lambda_i \cdot \beta \cdot \mu_i] \\ &= (\lambda_i \cdot (\sigma_i^2 + \mu_i^2) + \lambda_i^2 \cdot \mu_i^2 \cdot (c_i + g + c_i \cdot g)) \\ &\quad \cdot E[\beta^2] + \lambda_i^2 \cdot \mu_i^2 \cdot \text{Var}[\beta] \\ &= \lambda_i \cdot (\mu_i^2 + \sigma_i^2) \cdot (1 + b) + \lambda_i^2 \cdot \mu_i^2 \\ &\quad \cdot (c_i + g + b + c_i \cdot g + c_i \cdot b + g \cdot b \\ &\quad + c_i \cdot g \cdot b). \end{aligned} \quad (4.5)$$

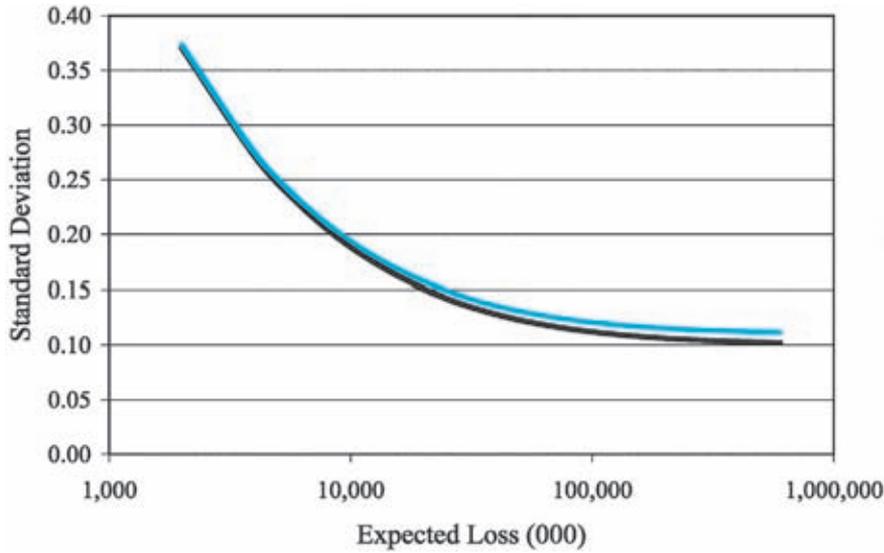
$$\begin{aligned} \text{Cov}[X_1, X_2] &= E_\beta[\text{Cov}[X_1, X_2 | \beta]] \\ &\quad + \text{Cov}_\beta[E[X_1 | \beta], E[X_2 | \beta]] \\ &= E_\beta[g \cdot \lambda_1 \cdot \beta \cdot \mu_1 \cdot \lambda_2 \cdot \beta \cdot \mu_2] \\ &\quad + \text{Cov}_\beta[\lambda_1 \cdot \beta \cdot \mu_1, \lambda_2 \cdot \beta \cdot \mu_2] \\ &= g \cdot \lambda_1 \cdot \mu_1 \cdot \lambda_2 \cdot \mu_2 \cdot E[\beta^2] \\ &\quad + \lambda_1 \cdot \mu_1 \cdot \lambda_2 \cdot \mu_2 \cdot \text{Var}[\beta] \\ &= \lambda_1 \cdot \mu_1 \cdot \lambda_2 \cdot \mu_2 \cdot (b + g + b \cdot g). \end{aligned} \quad (4.6)$$

The description of this version of the collective risk model can be completed 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. The second assumption keeps the math simple without sacrificing the main themes of this paper. Admittedly, there are many other assumptions one could make that are equally plausible. For example, the b s and g s could vary by line of insurance. In practice I have allowed g to vary by line of insurance. It is left as an exercise to the reader to show that you can replace g in Equations (4.4) and (4.6) with $\sqrt{g_1 \cdot g_2}$ when the coefficient of correlation between α_1 and α_2 is equal to one.

Figure 6. Standard deviation of normalized loss ratios for the collective risk model $\mu = 15,000$ $\sigma = 60,000$ $c = 0.01$



The following example illustrates the implications of this model for normalized loss ratios as we vary the size of risk. The example will assume that $\mu = 16,000$, $\sigma = 60,000$ and $c = 0.010$ for each line of insurance. The additional parameters will be $b = g = 0.001$. The final sections will show that these are reasonable choices of the parameters.

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

However, the coefficient of correlation between normalized loss ratios R_1 and R_2 , as defined by

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

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

Actuaries often find that their expectations of the coefficients 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 the following shall now illustrate.

Let us consider the covariance matrix for an insurer writing n lines of business:

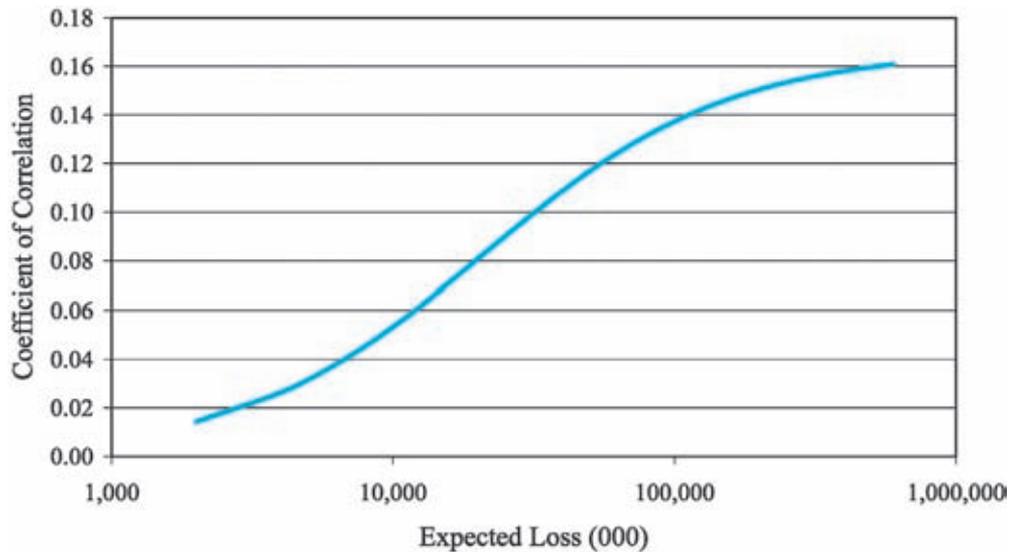
$$\begin{pmatrix} \text{Var}[X_1] & \text{Cov}[X_1, X_2] & \cdots & \text{Cov}[X_1, X_n] \\ \text{Cov}[X_2, X_1] & \text{Var}[X_2] & \cdots & \text{Cov}[X_2, X_n] \\ \cdots & \cdots & \cdots & \cdots \\ \text{Cov}[X_n, X_1] & \text{Cov}[X_n, X_2] & \cdots & \text{Var}[X_n] \end{pmatrix}$$

The standard deviation of the insurer’s total losses, $X_1 + \cdots + X_n$, 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 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.

5. An empirical test of the model

The collective risk model, as defined above, makes predictions about how the volatility and

Figure 7. Coefficient of correlation between normalized loss ratios of two equally sized risks for the collective risk model $\mu = 15,000$ $\sigma = 60,000$ $c = 0.01$ $b = g = 0.001$



correlation statistics of normalized 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. This section will demonstrate that data that is publicly available from 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. With Schedule P data for several insurers, various statistics such as standard deviations and coefficients of correlation between lines of insurance were calculated. Testing the model consisted of comparing these statistics with available information about each insurer.

Schedule P data presents several difficulties. The discussion that follows describes ways that Meyers, Klinker and Lalonde [3] dealt with these issues.

Schedule P premiums and reserves vary in largely predictable ways due to conditions that are present in the insurance market. These con-

ditions 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 the additional volatility caused by the underwriting cycle by using paid, rather than incurred, losses and estimating the ultimate incurred losses with industrywide paid loss development factors. We also attempted to smooth out differences in loss ratios that we deemed “predictable.” Appendix A in Meyers et al. [3] describes this process in greater detail.

After making the above adjustments, two other difficulties remain. 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. To see this mathematically, consider Equation (2.2)

Figure 8. Standard deviation of normalized loss ratios for the collective risk model for the sum of two equally sized risks $\mu = 15,000$ $\sigma = 60,000$ $c = 0.01$

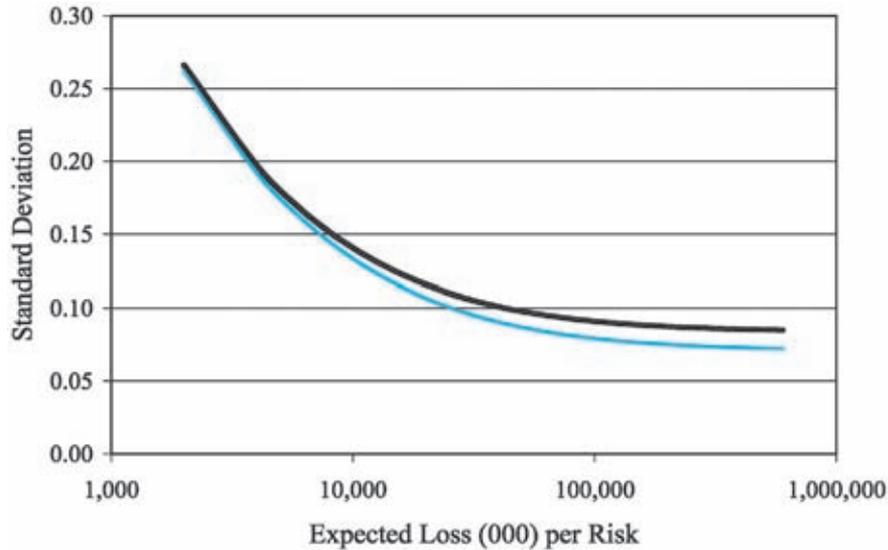
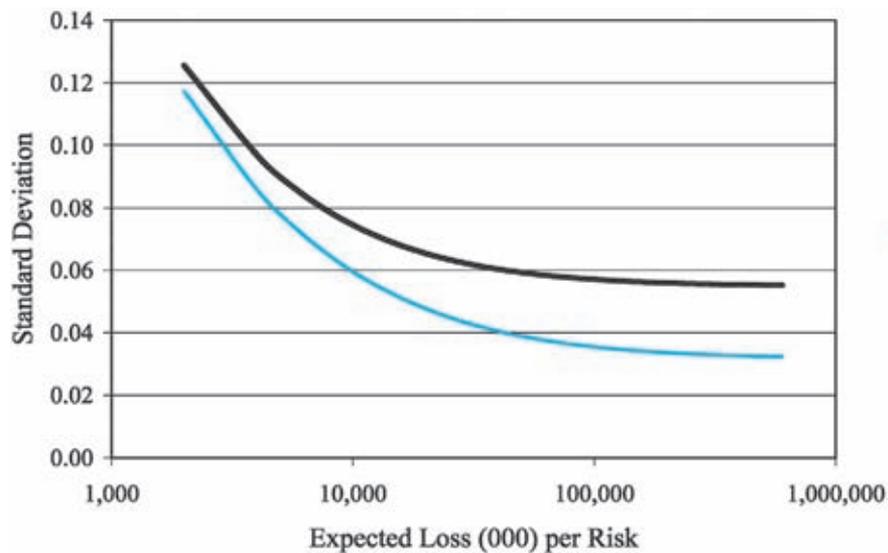


Figure 9. Standard deviation of normalized loss ratios for the collective risk model for the sum of ten equally sized risks $\mu = 15,000$ $\sigma = 60,000$ $c = 0.01$



with θ as a random variable representing future development. If we assume that the current loss estimates equal the expected ultimate loss, the second term in Equation (2.2) is the current variance, and the sum of the two terms is what we expect to be the ultimate variance.

Second, Schedule P losses are reported net of reinsurance. In addition, policy limits are not re-

ported. Rather than incorporate this information directly into our estimation, we did sensitivity tests of our model, varying limits and reinsurance provisions within realistic scenarios. See Figures 10 and 11.

This example presents results for commercial automobile liability insurance. I feel this is a good choice because: (1) it is a shorter

Figure 10. Commercial auto normalized loss ratios actual—Schedule P HiLim and LowLim from the collective risk model

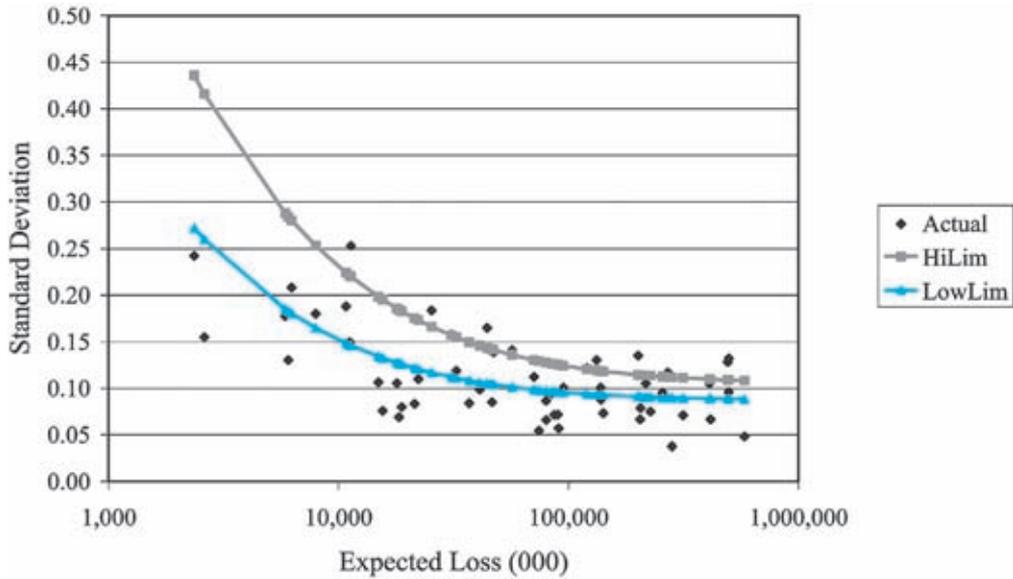
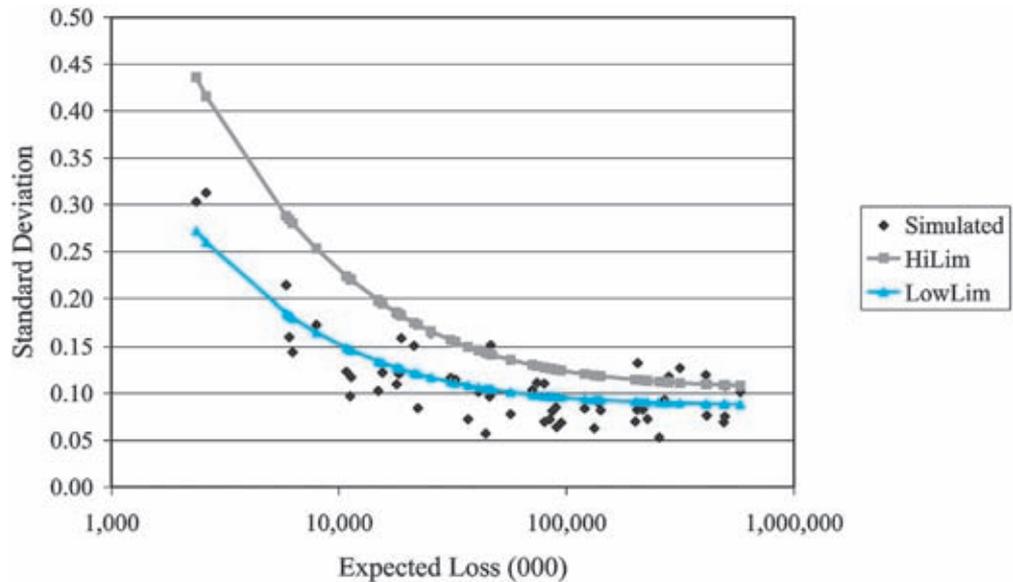


Figure 11. Commercial auto normalized loss ratios simulated from collective risk model

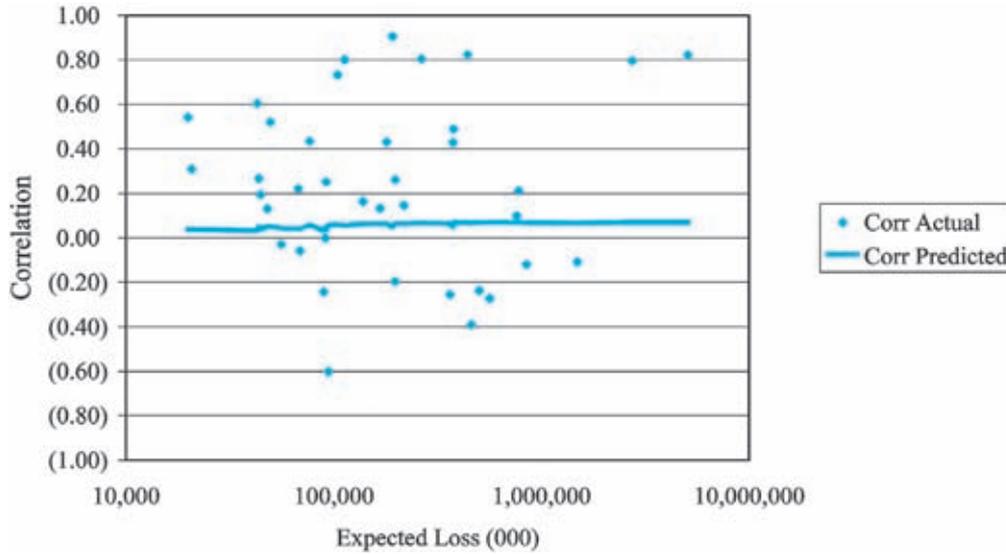


tailed line than general liability and the underestimation of volatility will not be as great; (2) the use of reinsurance is not as great as it is in the general liability lines of insurance; and (3) commercial auto is not as prone to catastrophes as the property lines of insurance.

5.1. Standard deviation of normalized loss ratios vs. size of insurer

As illustrated in Figure 5, the collective risk model predicts that the standard deviation of insurer normalized loss ratios should decrease as the size of the insurer increases. In Figure 10 we

Figure 12. Commercial auto and personal auto normalized loss ratios actual—Schedule P, predicted—collective risk model



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 “LowLim” used claim severity distribution parameters taken from a countrywide ISO claim severity distributions evaluated at the \$300,000 occurrence limit. In this series, $c = 0.007$, $g = 0.0005$, and $b = 0$, were selected judgmentally. See Section 6 for 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 normalized 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. Claim severity distribution parameters were taken from countrywide ISO claim severity distributions evaluated at the \$1 million occurrence limit. In this series $c = 0.010$, $g = 0.0010$, and $b = 0$ were selected judgmentally.

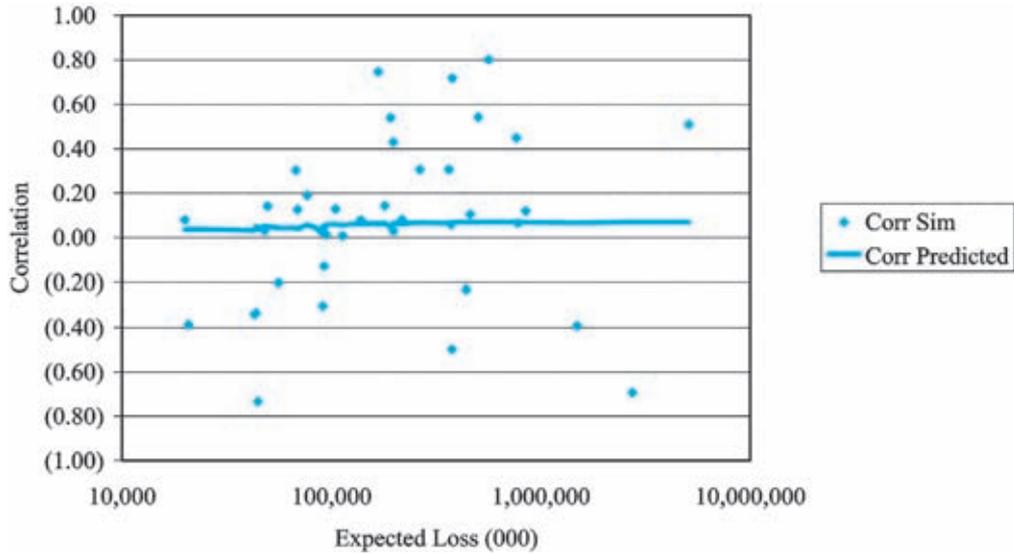
Figures 10 and 11 also provide a comparable plot of normalized loss ratios simulated from a collective risk model, one per insurer, using the same parameters used for the “LowLim” and “HiLim” series. The two plots suggest that the Schedule P data is well represented by the collective risk model—for an individual line of insurance.

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

²Since the expected loss varies by each observation of annual losses, the annual normalized loss ratios are not identically distributed according to the collective risk model. I do not think this is a serious problem here since the volume of business is fairly consistent from year to year for the insurers selected in this sample.

Figure 13. Commercial auto and personal auto normalized loss ratios simulated and predicted from the collective risk model



of correlation will increase with the size of the insurer. In Figure 12, we plotted the empirical coefficients 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 underlying model 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, assume that the common shock random variables α and β operate on all insurers simultaneously. For random normalized loss ratios R_1 and R_2 the covariance is calculated

³It 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.

as

$$E[(R_1 - 1) \cdot (R_2 - 1)] = \frac{\text{Cov}[X_1, X_2]}{\lambda_1 \cdot \mu_1 \cdot \lambda_2 \cdot \mu_2} = b + g + b \cdot g; \tag{5.1}$$

which is derived from Equation (4.5).

Now as already established, the standard deviation of the normalized loss ratios decreases with the size of the insurer. Thus the denominator of:

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

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

In the test that $(R_1 - 1) \cdot (R_2 - 1)$ was independent of insurer size the data consisted of all 15,790 possible pairs of r_1 and r_2 , and the associated expected losses, taken from the same year and different insurers. The line⁴ fit to the ordered

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

pairs

(Average Size of the Insurer, $(r_1 - 1) \cdot (r_2 - 1)$) produced a slope of $+1.95 \times 10^{-10}$. This slightly positive slope means that an increasing coefficient of correlation is consistent with the Schedule P data.

Equation (5.1) also provides us with a way to estimate the quantity $b + g + b \cdot g$. One simply has to calculate the weighted average of the 15,790 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, 200 weighted averages were simulated using the “LowLim” parameters (except that $b = g = 0$) with the result that the highest weighted average was 0.000318. Since the final weighted average of 0.00054 is greater than all the simulations generated under the null hypothesis that $b + g + b \cdot g = 0$, the result is statistically significant.

One final simulation with the “LowLim” parameters (except that $b = 0$ and $g = 0.00054$) calculated 200 slopes results in a slope of 1.95×10^{-10} , which was just below the 49th highest. Thus the fitted 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 paper was written in the hope of supplying some objectivity to this quantification. ISO has worked on quantifying this correlation. ISO has 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 some judgments must be made. Here are some of the considerations made in selecting the final models. Comments are always welcome.

- First of all, the selection of the model itself is subject to judgment. As noted above, there are other ways of introducing the “common shock idea” that are very likely to be consistent with the data.
- 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 $b + g + b \cdot g$, 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 in these trends when selecting the final parameters.

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 paper demonstrates that we do need to consider correlation between lines of insurance.

References

- [1] Meyers, Glenn G., Discussion of Wang: “Aggregation of Correlated Risk Portfolios: Models & Algorithms,” *Proceedings of the Casualty Actuarial Society* 86, 1999, pp. 781–805. <http://www.casact.org/pubs/proceed/proceed99/99705.pdf>
- [2] Meyers, Glenn G., “Estimating Between Line Correlations Generated by Parameter Uncertainty,” *Casualty Actuarial Society Forum*, Summer 1999, pp. 197–222. (<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,” *Casualty Actuarial Society Forum*, Summer 2003, pp. 15–82 (<http://www.casact.org/pubs/forum/03sforum/03sf015.pdf>).
- [4] Wang, Shaun S., “Aggregation of Correlated Risk Portfolios: Models and Algorithms,” *Proceedings of the Casualty Actuarial Society* 85, 1998, pp. 848–939 (<http://www.casact.org/pubs/proceed/proceed98/980848.pdf>).