A COPULA REGRESSION FOR MODELING MULTIVARIATE LOSS TRIANGLES AND QUANTIFYING RESERVING VARIABILITY

BY

PENG SHI

ABSTRACT

This article proposes a claims reserving model for dependent lines of business with the accommodation of association among triangles by a copula function. We show that the family of elliptical copulas is a pretty convenient choice to capture the dependencies introduced by various sources, including the common calendar year effects. To quantify the associated reserving variability, we resort to parametric bootstrapping techniques for simulating the predictive distribution of outstanding liabilities and for calculating the three components of predictive uncertainty: the model error, the process error and the estimation error. Numerical analysis is performed for a portfolio of casualty insurance from a major U.S. insurer.

KEYWORDS

Dependent loss reserving, copula, calendar year effects, predictive uncertainty.

1. INTRODUCTION

Claims reserving is a classic actuarial problem in general insurance. The development of the capital adequacy regime — Solvency II — emphasizes the role of reserving variability in the decision process, which motivates the fast-growing literature in the stochastic loss reserving models (see the excellent monograph by Wüthrich and Merz, 2008, for a review). In particular, the multivariate reserving approach has received extensive attention. Most recent examples include Merz and Wüthrich (2009a, 2009b), Zhang (2010), de Jong (2012) and Happ and Wuthrich (2013), among others.

Despite the vast studies in the multivariate claims reserving, modeling the dependency among multiple triangles is still a challenge. Because of the time-dependent evolution, correlation among payments could be introduced by various sources, among which the calendar year effect is the focus of the current literature (see, for example, Shi et al., 2012; Wüthrich, 2012; Salzmann and

Astin Bulletin 44(1), 85-102. doi: 10.1017/asb.2013.23

Wüthrich, 2012; Merz et al., 2013). Specifically, losses among triangles could be correlated due to a common calendar year effect, such as a court judgment or management decision, that could affect all open claims in the portfolio simultaneously.

To model calendar year effects, existing studies rely on the normality assumption or in more general the theory of linear models. Such practice sacrifices the flexibility in handling the skewness and heavy tails featured by loss distributions. To address this issue, we propose a multivariate claims reserving model with the accommodation of association among triangles by a copula function. Especially, we show that the family of elliptical copulas is a convenient choice where dependency could be captured by the dispersion matrix. Various forms of association are allowed in this framework, including the common calendar year effects, that could cause correlation within and across triangles. In addition, two aspects further set our analysis apart from the current studies. First, we are able to explore the usefulness of the Tweedie distribution in modeling the aggregated losses in claim triangles. Second, to quantify predictive uncertainty, the parametric nature allows to employ the modern bootstrapping for incorporating the model error along with the process and estimation errors.

The rest of this paper is organized as follows. Section 2 introduces the copula modeling framework and discusses the procedures for model estimation and selection. Section 3 focuses on the reserving variability. We adopt a parametric bootstrap that could be used to derive the predictive distribution of unpaid losses and calculate mean square error of prediction. Section 4 applies the copula model to a casualty insurance portfolio from a U.S. insurer and demonstrates the flexibility of the proposed approach. Section 5 concludes the article. Technical notes and supplementary results are provided in the Appendix.

2. Modeling

Consider an insurance portfolio of N lines of business. We use Y_{nij} to denote the incremental payment in the nth $(= 1, \dots, N)$ run-off triangle, with i $(= 1, \dots, I)$ and j $(= 0, \dots, J)$ indicating the accident year and development lag, respectively. Using above notations, I represents the most recent accident year and J represents the largest development lag, satisfying $J \leq I-1$. Without loss of generality, we consider the case $J = I - 1$. Thus, by the latest calendar year $J+1$, we only observe payments in the upper left-hand triangle. Define the vector of incremental payments in the upper and lower triangles of the insurance portfolio respectively as $\mathbf{Y}^U = \{\mathbf{Y}_{ij} : i+j \leq J+1\}$ and $\mathbf{Y}^L = \{\mathbf{Y}_{ij} : i+j > J+1\}$ with $\mathbf{Y}_{ij} = (Y_{1ij}, \dots, Y_{Nij})'$. From reserving perspective, our goal is to make inference on the unpaid losses in the lower right-hand triangle $\mathbf{Y}^L$, based on the observed payments in the upper left-hand triangle $\mathbf{Y}^U$.

2.1. Copula model

In loss reserving context, claims in each run-off typically represent aggregated losses, i.e. the sum of claims from all policyholders insured by the insurer, motivating the employment of the Tweedie (1984) distribution in modeling the incremental payments in triangles. In actuarial science, the Tweedie's compound Poisson model has been used for ratemaking (see Jørgensen and De Souza, 1994; Smyth and Jørgensen, 2002) and claims reserving analysis (see Wüthrich, 2003; Peters et al., 2009; Meyers and Shi, 2011). Along the group of reserving models, we consider a version of Tweedie's compound Poisson distribution with the constant dispersion parameter ϕ and the power parameter $1 < p < 2$, i.e. $f(y_{nij}; \mu_{nij}, \phi, p)$, for the incremental claim Y_{nij}. A log link function is further specified for the mean parameter $\log(\mu_{nij}) = \alpha_{ni} + \beta_{nj}$, where α_{ni} and β_{nj} capture the fixed effects along accident year and development year, respectively. As usual, identification constraints apply here.

Now, we introduce a copula modeling framework for jointly examining claims from multiple triangles. The proposed model is built on and extends that of Shi and Frees (2011), where like most existing studies, the authors relied on the independence assumption across accident years and focused on the pairwise association, i.e. each cell in one triangle relates to the equivalent cell in another triangle. We propose to use elliptical copulas to model dependencies among multiple loss triangles, relaxing the independence assumption across accident years. As we will show, the dispersion matrix in elliptical copulas offers great flexibility in accommodating various types of association.

Because common calendar year effects introduce association within and across triangles, a copula function is specified for the vector of claims occurred in each calendar year. To be more specific, collect all claims in calendar year $t = i + j$ in the nth triangle into the vector $\mathbf{Y}_{nt} = \{Y_{nij} : i + j = t\}$, then stack vectors $\mathbf{Y}_{nt}$, $n = 1, \ldots, N$, into the vector $\mathbf{Y}_t = (\mathbf{Y}_{1t}', \ldots, \mathbf{Y}_{Nt}')'$. For $t = 1, \ldots, I + J$, the claims in the vector $\mathbf{Y}_t$ are joined through an elliptical copula:

$$c_t(u_1, \ldots, u_{t_D}) = h_{t_D}(H^{-1}\left(u_1\right), \ldots, H^{-1}(u_{t_D})) \prod_{d=1}^{t_D} \frac{1}{h\left(H^{-1}(u_d)\right)}. \quad (1)$$

Here, $h(\cdot)$ and $H(\cdot)$ indicate the density function and distribution function of a univariate elliptical distribution, respectively, and h_{t_D} denotes the density function of a t_D-variate elliptical distribution with

$$h_{t_D}(\boldsymbol{x}) = \frac{c}{\sqrt{|\boldsymbol{\Sigma}|}} g\left(\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}\right),$$

where c is a normalizing constant and $g(\cdot)$ is the density generator function (see, for example, Fang et al., 2003). Without loss of generality, the elliptical distribution is assumed to have zero mean and unit variance, since elliptical copulas

https://doi.org/10.1017/asb.2013.23 Published online by Cambridge University Press

remain invariant under an increasing transformation of components. It is worth mentioning the dimension of $\mathbf{Y}_t$ and thus of the corresponding copula function t_D:

$$t_D = \begin{cases} tN, & \text{if } 1 \le t \le I, \\ (I+J+1-t)N, & \text{if } I+1 \le t \le I+J. \end{cases}$$

The critical part in the copula specification is the dispersion matrix $\mathbf{\Sigma}_t$ corresponding to the vector $\mathbf{Y}_t$. We adopt an approach similar to de Jong (2012) and motivate the specification of $\mathbf{\Sigma}_t$ with a Gaussian copula, a special case of the elliptical family. Under a Gaussian copula, one could assume

$$(\Phi \circ F)(y_{nij}) = \eta_{n,t=i+j} + \varepsilon_{nij}, \tag{2}$$

where $F(\cdot)$ is the Tweedie distribution function. The variable η_{nt} indicates the calendar year effect in year t for the nth triangle, and ε_{nij}, independent of η_{nt}, indicates the cell-specific effect for the nth triangle corresponding to accident year i and development lag j. We introduce correlations among loss triangles by further specifying the distributions of $\boldsymbol{\eta}_t = \{\eta_{nt}; n = 1, \ldots, N\}$ and $\boldsymbol{\varepsilon}_{ij} = \{\varepsilon_{nij} : n = 1, \ldots, N\}$ as

$$\boldsymbol{\eta}_t = \begin{pmatrix} \eta_{1t} \\ \vdots \\ \eta_{Nt} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{\Lambda}) \text{ and } \boldsymbol{\varepsilon}_{ij} = \begin{pmatrix} \varepsilon_{1ij} \\ \vdots \\ \varepsilon_{Nij} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{\Omega}),$$

with

$$\mathbf{\Lambda} = Y \begin{pmatrix} c_1^2 & c_1c_2\lambda_{12} & \cdots & c_1c_N\lambda_{1N} \\ c_2c_1\lambda_{21} & c_2^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ c_Nc_1\lambda_{N1} & c_Nc_2\lambda_{N2} & \cdots & c_N^2 \end{pmatrix} \text{ and}$$

$$\mathbf{\Omega} = \begin{pmatrix} s_1^2 & s_1s_2\omega_{12} & \cdots & s_1s_N\omega_{1N} \\ s_2s_1\omega_{21} & s_2^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ s_Ns_1\omega_{N1} & s_Ns_2\omega_{N2} & \cdots & s_N^2 \end{pmatrix}.$$

Following the above formulation, correlations among multiple triangles could be caused by both calendar year effect and cell-specific effect. To be more specific, the covariance matrix $\mathbf{\Lambda}$ captures the association within and across run-offs due to accounting year effects, and $\mathbf{\Omega}$ accommodates the pair-wise association among different triangles. To guarantee a standard normal marginal in the Gaussian copula, one requires $c_n^2 + s_n^2 = 1$ for $n = 1, \ldots, N$.

Now, we can derive the dispersion matrix in the Gaussian copula for the vector of claims in calendar year t, i.e. $\mathbf{Y}_t$. Collecting (2) for all claims in calendar year t from the insurance portfolio, we have

$$(\Phi \circ F)(\mathbf{y}_t) = (I_N \otimes 1_t)\, \boldsymbol{\eta}_t + (I_N \otimes I_t)\, \boldsymbol{\varepsilon}_t,$$

where 1_t indicates the vector of 1s of length t_D/N, and I_t and I_N indicate a t_D/N- and N-dimensional identity matrix, respectively. Note that $\boldsymbol{\varepsilon}_t$ is defined in the same order as $\mathbf{Y}_t$. Then, the dispersion matrix corresponding to $\mathbf{Y}_t$ could be shown as

$$\boldsymbol{\Sigma}_t = (I_N \otimes 1_t)\, \boldsymbol{\Lambda}\, (I_N \otimes 1_t)' + (I_N \otimes I_t)\, \boldsymbol{\Omega}\, (I_N \otimes I_t)' = \boldsymbol{\Lambda} \otimes 1_t 1_t^{'} + \boldsymbol{\Omega} \otimes I_t, \quad (3)$$

which could be easily adapted to the entire family of elliptical copulas.

2.2. Model estimation

One advantage of the model-based approach is that the likelihood-based method could be easily implemented for inference purposes. Let $\boldsymbol{\theta}$ denote the vector of parameters, including both the Tweedie model and the elliptical copula. Based on the above specification, the log-likelihood function for the claims in calendar year t in the insurance portfolio is

$$L_t(\boldsymbol{\theta}) = \ln c_t \left(F(y_{1,1,t-1}), \ldots, F(y_{1,t-1,1}), \ldots\ldots, F(y_{N,1,t-1}), \ldots, F(y_{N,t-1,1}) \right) + \sum_{n=1}^{N} \sum_{i+j=t} \ln f(y_{nij}; \mu_{nij}, \phi, p).$$

Assuming independence across calendar years, the log likelihood for all claims in the insurance portfolio is the sum of the log likelihood over all calendar years, which could be expressed as

$$L(\boldsymbol{\theta}) = \sum_{t=1}^{I} \sum_{n=1}^{N} \sum_{i+j=t} \ln f(y_{nij}; \mu_{nij}, \phi, p) + \sum_{t=1}^{I} \left\{ \ln g\left(\frac{1}{2} \boldsymbol{x}_t^{'} \boldsymbol{\Sigma}_t^{-1} \boldsymbol{x}_t \right) - \ln \sqrt{|\boldsymbol{\Sigma}_t|} - \sum_{d=1}^{t_D} \ln h(x_d) \right\}.$$

Here, we define the vector $\mathbf{x}_t = (x_{1,1,t-1}, \ldots, x_{1,t-1,1}, \ldots\ldots, x_{N,1,t-1}, \ldots, x_{N,t-1,1})^{'}$, with the element $x_{nij} = H^{-1}(F(y_{nij}; \mu_{nij}, \phi, p))$ and x_d indicating the dth component of $\mathbf{x}_t$. The dispersion matrix $\boldsymbol{\Sigma}_t$ is defined according to (3).

A full-information maximum likelihood estimation (MLE) could be implemented to calibrate the model, i.e. $\widehat{\boldsymbol{\theta}}^{\text{MLE}} = \text{argmax } L(\boldsymbol{\theta})$. However, maximum likelihood is computationally difficult for multivariate copula-based models. As an alternative, one could resort to a two-stage estimation approach, known as the method of inference function for marginals (IMF), where the first stage maximizes the likelihood from marginals, and the second stage maximizes the likelihood of dependence parameters with parameters in marginals held fixed from the first stage. Accordingly, one could split the likelihood function into two pieces:

$$L(\boldsymbol{\theta}) = L_1(\boldsymbol{\theta}^{Tweedie}, \boldsymbol{\theta}^{Copula}) + L_2(\boldsymbol{\theta}^{Tweedie}, \boldsymbol{\theta}^{Copula}),$$

where

$$L_1(\boldsymbol{\theta}^{Tweedie}, \boldsymbol{\theta}^{Copula}) = \sum_{t=1}^{I}\sum_{n=1}^{N}\sum_{i+j=t}\left\{\ln a(y_{nij};\phi,p) + \frac{1}{\phi}\left[y_{nij}b_{nij} - c(b_{nij};p)\right]\right\},$$

and

$$L_2(\boldsymbol{\theta}^{Tweedie}, \boldsymbol{\theta}^{Copula}) = \sum_{t=1}^{I}\left\{g\left(\frac{1}{2}\boldsymbol{x}_t'(\boldsymbol{\Lambda}\otimes 1_t1_t' + \boldsymbol{\Omega}\otimes I_t)^{-1}\boldsymbol{x}_t\right) - \ln\sqrt{|\boldsymbol{\Lambda}\otimes 1_t1_t' + \boldsymbol{\Omega}\otimes I_t|} - \sum_{d=1}^{t_D}\ln h(x_d)\right\}.$$

Then, the two-stage estimation could be implemented as follows:

Step I: $\widehat{\boldsymbol{\theta}}^{Tweedie} = \text{argmax } L_1(\boldsymbol{\theta}^{Tweedie}, \boldsymbol{\theta}^{Copula})$, s.j.t., $\boldsymbol{\theta}^{Copula} = \widehat{\boldsymbol{\theta}}^{Copula}$.

Step II: $\widehat{\boldsymbol{\theta}}^{Copula} = \text{argmax } L_2(\boldsymbol{\theta}^{Tweedie}, \boldsymbol{\theta}^{Copula})$, s.j.t., $\boldsymbol{\theta}^{Tweedie} = \widehat{\boldsymbol{\theta}}^{Tweedie}$.

To achieve the asymptotic efficiency of MLE, one needs to reiterate the two-step estimation r times so that the IMF will converge in probability to the MLE, i.e. $(\widehat{\boldsymbol{\theta}}^{Tweedie}, \widehat{\boldsymbol{\theta}}^{Copula}) \xrightarrow{r=+\infty} \widehat{\boldsymbol{\theta}}^{MLE}$ (see Joe, 2005).

2.3. Model selection

Because reserving problems typically involve small samples, the usual criteria AIC and SIC could lead to overfitting. Thus, we consider their bias-corrected versions, AICc (see Sugiura, 1978; Hurvich and Tsai, 1989) and SICc (see McQuarrie, 1999). To incorporate the model error and demonstrate its effect on reserving variability, we consider various specifications.

$\mathcal{M}1$: Assume $\alpha_{ni} = 0$ for $i = 1, \ldots, I$. In this model, the evolution of claims depends solely on the development lag, regardless of the year when losses are incurred. Note that one does not necessarily assume this for all lines of business simultaneously. We will consider different combinations in the insurance portfolio.

$\mathcal{M}2$: Assume $\boldsymbol{\Omega} = \text{diag}\{s_1^2, \ldots, s_N^2\}$. This model presumes that all dependencies among multiple triangles are caused by common calendar year effects. This type of model has been examined by de Jong (2012).

$\mathcal{M}3$: Assume $c_n = 0$ for $n = 1, \ldots, N$. It is easy to see that no calendar year correlation is considered in this case. As in Shi and Frees (2011), the model focuses on pair-wise association, assuming that the calendar year correlation has already been captured by the effects of accident year and development lag.

$\mathcal{M}4$: Assume that $\mathbf{\Omega} = \text{diag}\{s_1^2, \ldots, s_N^2\}$, and $c_n = 0$ for $n = 1, \ldots, N$. Suppose that the claims in different lines of business are independent of each other, and this model does not incorporate any dependency among loss triangles.

$\mathcal{M}5$: The full model specified in Section 2.1.

Note that model selection focuses on the specifications of the mean of loss distributions and the dependence structure among payments rather than the parametric forms. Specifically, the choice of the Tweedie distribution is motivated by the aggregated claims, and the employment of elliptical copulas is due to the flexible specification of the pair-wise dependence through the dispersion matrix. However, statistical inference allows for the selection of functional forms though it is not the focus of this study. For instance, to select the generator g in the elliptical family, one could refer to the t-statistic proposed by Sun et al. (2008) and Shi (2012).

3. Predictive uncertainty

Define reserve R as a linear function of $\mathbf{Y}^L$, i.e. $R = g(\mathbf{Y}^L)$, where $g(\mathbf{x}) = \boldsymbol{\eta}'\mathbf{x}$ and $\boldsymbol{\eta}$ is a known vector of the same length as $\mathbf{x}$. Our interest is to derive the predictive distribution of R conditional on the realized losses. We resort to modern parametric bootstrapping to derive the predictive distribution of reserves while incorporating both the parameter uncertainty and model error. In addition, we present the (conditional) mean-squared error of prediction (MSEP) for the reserve $R = g(\mathbf{Y}^L)$ when conditioning on a specific model and averaging over different models. The detailed proofs are provided in Section A.1 of the Appendix.

3.1. Conditioning on a model

The expectation and variance of reserve R given the model $\mathcal{M}$ and associated parameters $\boldsymbol{\theta}$ could be expressed as

$$\widetilde{R}_{\mathcal{M}} = \mathrm{E}[R|\boldsymbol{\theta}, \mathcal{M}]$$

and

$$\widetilde{V}_{\mathcal{M}} = \mathrm{Var}[R|\boldsymbol{\theta}, \mathcal{M}].$$

Simple calculations show that the best estimate of reserve that minimizes the conditional MSEP given model $\mathcal{M}$ is

$$\widehat{R}^{B_{\mathcal{M}}} = \mathrm{E}\left[\widetilde{R}_{\mathcal{M}}|\mathbf{Y}^U, \mathcal{M}\right] = \mathrm{E}\left[R|\mathbf{Y}^U, \mathcal{M}\right]. \tag{4}$$

Here, $\widehat{R}^{B_{\mathcal{M}}}$ is a consistent estimator and $\mathbf{Y}^U$-measurable. Thus, the (conditional) mean-squared prediction error for a given model $\mathcal{M}$ could be decomposed as

$$\mathrm{MSEP}_{R|\mathbf{Y}^U,\mathcal{M}}\left(\widehat{R}^{B_{\mathcal{M}}}\right) = \mathrm{E}\left[\widetilde{V}_{\mathcal{M}}|\mathbf{Y}^U,\mathcal{M}\right] + \mathrm{Var}\left[\widetilde{R}_{\mathcal{M}}|\mathbf{Y}^U,\mathcal{M}\right]. \tag{5}$$

Note that in this context the unknown parameter vector $\boldsymbol{\theta}$ is modeled stochastically. Hence, the above decomposition of MSEP is similar to the Bayesian perspective, the first term quantifying the process error and the second term quantifying the parameter error.

3.2. Mixing over models

One step further, one could incorporate the model error into the predictive uncertainty by mixing over different models. Specifically, the best consistent estimate of reserve incorporating the model error could be derived as

$$\widehat{R}^B = \mathrm{E}\left[\widetilde{R}|\mathbf{Y}^U\right] = \mathrm{E}[R|\mathbf{Y}^U],$$

where

$$\widetilde{R} = \mathrm{E}\left[\widetilde{R}_{\mathcal{M}}|\boldsymbol{\theta}\right] = \mathrm{E}[R|\boldsymbol{\theta}]. \tag{6}$$

Since $\widehat{R}^B$ is $\mathbf{Y}^U$-measurable, we show that the prediction uncertainty could be decomposed into three terms as follows:

$$\begin{aligned}\mathrm{MSEP}_{R|\mathbf{Y}^U}\left(\widehat{R}^B\right) = &\,\mathrm{E}\left[\mathrm{E}\left[\widetilde{V}_{\mathcal{M}}|\boldsymbol{\theta}\right]|\mathbf{Y}^U\right] + \mathrm{E}\left[\mathrm{Var}\left[\widetilde{R}_{\mathcal{M}}|\boldsymbol{\theta}\right]|\mathbf{Y}^U\right]\\ &+ \mathrm{Var}\left[\mathrm{E}\left[\widetilde{R}_{\mathcal{M}}|\boldsymbol{\theta}\right]|\mathbf{Y}^U\right].\end{aligned} \tag{7}$$

The first and second terms in (7) correspond to the process error and the parameter error, respectively, when averaging over various models. The last term explicitly quantifies the model error in the reserving variability. The calculations of the three components are implemented based on parametric bootstrapping.

4. Application

4.1. Data characteristics

We consider an insurance portfolio consisting of two casualty lines of business, workers' compensation and commercial auto liability. The triangle data are from the Schedule P of the National Association of Insurance Commissioners (NAIC) database. Our analysis focuses on the run-offs with incremental paid losses. The triangles of the two casualty lines of business are summarized in Section A.2 for reference purposes.

We normalize the payment by dividing by the net premiums earned in the corresponding accident year. The multiple time-series plots of loss ratios are displayed in the first two panels in Figure 1. Each curve corresponds to an accident

https://doi.org/10.1017/asb.2013.23 Published online by Cambridge University Press

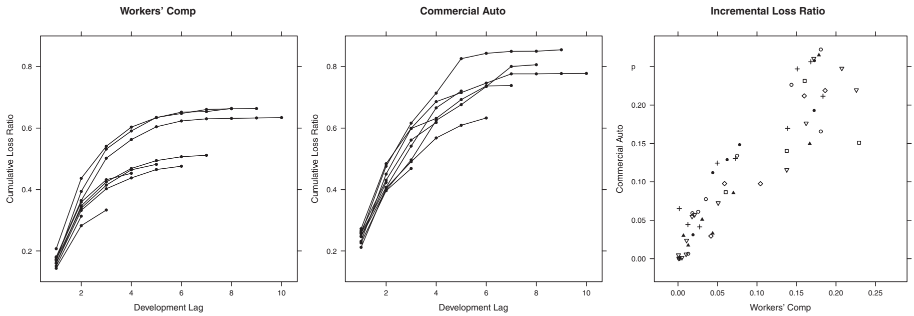

FIGURE 1: Loss development characteristics. The first two figures show the multiple time-series plots of cumulative loss ratios for workers' compensation and commercial auto liability. The third figure shows the scatter plot of incremental loss ratios.

year, showing the development pattern of losses over time. A comparison of the two panels suggests a strong accident year effect in workers' compensation and a higher volatility in commercial auto liability. We also observe a larger ultimate loss ratio in commercial auto liability line, presumably because of the relatively lower legal costs.

To obtain some knowledge on the correlation between the two lines of business, the third panel in Figure 1 presents the scatter plot of incremental loss ratios. This plot suggests a strong positive, although nonlinear, relationship with the corresponding Pearson correlation 0.911. Note that the scatter plot describes a pair-wise correlation. However, a common calendar year effect could introduce correlation between different lines as well as within a single line. Thus, the loss ratios in each calendar year are differentiated by a unique symbol. We observe the heterogeneity in correlation cross calendar years, indicating a mixed effect of calendar year correlation and pair-wise correlation.

4.2. Model inference

The five classes of models described in Section 2.3 are estimated using the likelihood-based method. The model selection criteria suggest that the best model assumes the Tweedie's compound Poisson distribution for the two lines of business in the insurance portfolio, with the fixed effects of both accident year and development year for the workers' compensation and the fixed effects of only development year for the commercial auto liability. A Gaussian copula is used to accommodate the association among different lines due to common calendar year effect as well as line-specific effect. Table 1 summarizes the estimation results for the best model with associated confidence interval derived using the bootstrap method.

Consistent with Figure 1, we observe strong effects of both accident year and development year for the workers' compensation and the effects of only development year for the commercial auto liability. The higher power parameter of the Tweedie model for the commercial auto liability suggests the higher

TABLE 1

PARAMETER ESTIMATES FOR THE COPULA MODEL.

	Workers' Compensation						Commercial Auto Liability	
	Estimate	95% CI		Estimate	95% CI		Estimate	95% CI
int	−1.606	[−1.715, −1.499]				int	−1.399	[−1.514, −1.288]
α_2	0.045	[−0.069, 0.158]	β_2	0.077	[0.003, 0.150]	α_2	−0.298	[−0.469, −0.129]
α_3	0.040	[−0.079, 0.159]	β_3	−0.642	[−0.736, −0.550]	α_3	−0.764	[−0.961, −0.571]
α_4	−0.219	[−0.349, −0.089]	β_4	−1.329	[−1.457, −1.206]	α_4	−1.056	[−1.281, −0.841]
α_5	−0.282	[−0.417, −0.146]	β_5	−1.770	[−1.933, −1.614]	α_5	−1.456	[−1.726, −1.199]
α_6	−0.242	[−0.382, −0.103]	β_6	−2.552	[−2.801, −2.321]	α_6	−1.967	[−2.316, −1.651]
α_7	−0.256	[−0.403, −0.109]	β_7	−3.415	[−3.820, −3.062]	α_7	−2.290	[−2.719, −1.910]
α_8	−0.459	[−0.625, −0.297]	β_8	−3.833	[−4.401, −3.361]	α_8	−5.570	[−6.708, −4.005]
α_9	−0.290	[−0.463, −0.119]	β_9	−6.897	[−24.592, −4.444]	α_9	−5.958	[−7.516, −3.649]
α_{10}	−0.161	[−0.383, 0.053]	β_{10}	−10.502	[−28.983, −4.184]	α_{10}	−22.268	[−34.156, −4.739]
p	1.165	[1.013, 1.480]				p	1.387	[1.166, 1.715]
ϕ	0.002	[0.001, 0.005]				ϕ	0.019	[0.008, 0.039]
c_1	0.272	[0.021, 0.442]						
c_2	0.511	[0.427, 0.566]						
c_{12}	0.226	[−0.061, 0.577]						
s_{12}	0.161	[0.017, 0.325]						

volatility in loss development. The association parameters in the copula function indicate the significant dependency between the two lines after controlling for the effects in both accident years and development years. The results also support the model assumption that the correlation between lines comes from two sources: parameters c_{12} and s_{12} measure the correlation due to common calendar year effects and line-specific effects, respectively, and parameters c_i or s_i $(i = 1, 2)$ determine the weight of the two sources when arriving at the pair-wise association among paid losses. For this particular insurance portfolio, 14% of the association could be explained by calendar year effects and 82% could be attributed to the pair-wise correlation. The implied correlation between loss payments of the two lines represents a combined result. The correlation between lines has important implications on the reserve indication. To emphasize this point, a simulation study is provided in Section A.3 to show the combined effect of c_{12} and s_{12} on the implied correlation between incremental payments and thus on the predictive distribution of outstanding liabilities of the insurance portfolio.

To incorporate modeling error into the prediction of unpaid losses, one needs the distribution of $\mathcal{M}$. Assuming that $\mathcal{M}$ contains all possible specifications in the five classes in Section 2.3, we derive its probabilistic distribution using parametric bootstrap. Specifically, the pseudo-payments in bootstrap are generated from the best model and then are used to calibrate the candidates. Then, $\mathcal{M}$ takes the model chosen by the information-based model selection criteria. The empirical distribution of $\mathcal{M}$ is estimated by repeating the model selection process a large number of times.

4.3. Reserve indication

With the estimation results in the previous section, one could derive the predictive distribution of unpaid losses. To illustrate, the resulting average predictive distribution of loss reserves that takes account for the model error is displayed in Figure 2. The model averaging is implemented using the selection distribution from SICc because it provides the most favorable results for this particular insurance portfolio. Though not reported here, our analysis shows that other model selection criteria could lead to significantly different predictions, emphasizing the predictive variability due to mis-specification error. Agreeing with a previous analysis, we observe higher uncertainty in workers' compensation than commercial auto liability. In fact, the loss reserves are 18,839 [15,923, 22,006] and 17,722 [15,840, 19,677] for the worker's compensation and the commercial auto liability, respectively, where the numbers in brackets represent the 95% confidence band. When aggregating reserves from individual lines, one obtains the reserve for the insurance portfolio. Our estimate is 36,561 with confidence interval [32,893, 40,437]. The narrower confidence band than the silo method simply demonstrates the diversification effect due to the positive association among different lines in the portfolio.

https://doi.org/10.1017/asb.2013.23 Published online by Cambridge University Press

TABLE 2

COMPARISON OF RESERVING VARIABILITY.

	Workers' Comp	Commercial Auto	Insurance Portfolio
		Conditioning on a model	
$\sqrt{PE}$	755.303	1271.347	1478.786
$\sqrt{EE}$	1795.316	1110.040	2193.791
$\sqrt{MSEP}$	1947.728	1687.754	2645.661
		Averaging over models	
$\sqrt{PE}$	1040.853	1198.596	1587.453
$\sqrt{EE}$	1575.248	1974.456	2678.812
$\sqrt{ME}$	1311.468	99.751	1275.185
$\sqrt{MSEP}$	2298.854	2311.939	3364.839

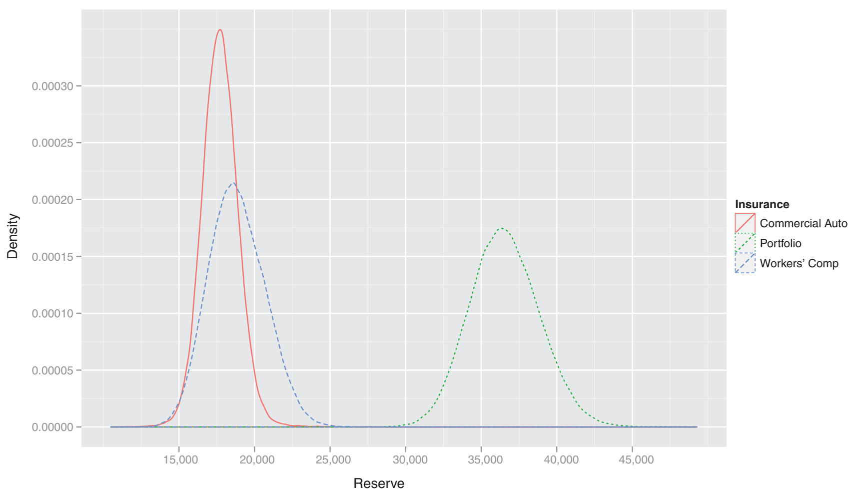

FIGURE 2: Predictive distribution of reserves for individual lines as well as the portfolio.

We conclude the analysis with a comparison between the two-component MESP using (5) and the three-component MSEP using (7). The results for each individual line as well as the insurance portfolio are exhibited in Table 2. We condition on the model with the best fit in (5) and average over models with the SICc-based distribution in (7). The comparison shows that, first, though the model error accounts for a small percentage of predictive variability, it indeed introduces extra uncertainty in the prediction of reserves. Second, the uncertainty of model mis-specification is lower in commercial auto liability than in workers' compensation, presumably due to the non-significant effect along accident years. Finally, the subadditivity of MSEP demonstrates

https://doi.org/10.1017/asb.2013.23 Published online by Cambridge University Press

the diversification effects when aggregating losses from dependent lines of business.

5. Conclusion

This article considered joint modeling run-off triangles from dependent lines of business. We proposed a model-based reserving approach and showed its flexibility in accommodating the unique features of the triangle data when compared with existing studies. We focused on the family of elliptical copulas because its dispersion matrix is a convenient vehicle to introduce various types of association. In particular, the dependency structure was specified in a way that both calendar year effects and pair-wise association could cause dependency among triangles. To quantify the associated reserving variability, we took advantage of the powerful parametric bootstrapping that allows for the incorporation of parameter uncertainty as well as model error into the reserving variability. We also provided guidance on the calculation of the mean-squared error of prediction accordingly.

The proposed method was applied to a casualty insurance portfolio consisting of workers' compensation and commercial auto liability from a U.S. property-casualty insurer. We showed that the association along calendar years was non-ignorable and the dependencies due to both calendar year effect and pair-wise correlation were well captured by the elliptical copula. When quantifying the prediction uncertainty, we calibrated various models with different specifications. Our analysis suggested that the model mis-specification error should be treated carefully, though its effect could be mitigated when an appropriate model selection criterion is used.

Acknowledgement

I thank Editor Mario V. Wüthrich as well as anonymous reviewers for comments that helped improve the paper.

References

DE JONG, P. (2012) Modeling dependence between loss triangles using copula. *North American Actuarial Journal*, **16**(1), 74–86.

FANG, K., KOTZ, S. and NG, K. (2003) *Symmetric Multivariate and Related Distributions*. London, England: Chapman & Hall.

HAPP, S. and WUTHRICH, M. (2013) Paid-incurred chain reserving method with dependence modeling. *ASTIN Bulletin,* **43**(1), 1–20.

HURVICH, C. and TSAI, C. (1989) Regression and time series model selection in small samples. *Biometrika*, **76**(2), 297–307.

JOE, H. (2005) Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, **94**(2), 401–419.

JØRGENSEN, B. and DE SOUZA, M. (1994) Fitting Tweedie's compound Poisson model to insurance claims data. *Scandinavian Actuarial Journal*, **1**(1), 69–93.
MCQUARRIE, A. (1999) A small-sample correction for the Schwarz SIC model selection criterion. *Statistics & Probability Letters*, **44**(1), 79–86.
MERZ, M. and WÜTHRICH, M. (2009a) Combining chain-ladder and additive loss reserving methods for dependent lines of business. *Variance*, **3**(2), 270–291.
MERZ, M. and WÜTHRICH, M. (2009b) Prediction error of the multivariate additive loss reserving method for dependent lines of business. *Variance*, **3**(1), 131–151.
MERZ, M., WÜTHRICH, M. and HASHORVA, E. (2013) Dependence modeling in multivariate claims run-off triangles. *Annals of Actuarial Science*, **7**(1), 3–25.
MEYERS, G. and SHI, P. (2011) *The retrospective testing of stochastic loss reserve models*. Casualty Actuarial Society E-Forum, Summer 2011, http://www.casact.org/pubs/forum/11sumforum/.
PETERS, G., SHEVCHENKO, P. and WÜTHRICH, M. (2009) Model uncertainty in claims reserving within Tweedie's compound Poisson models. *ASTIN Bulletin*, **39**(1), 1–33.
SALZMANN, R. and WÜTHRICH, M. (2012) Modeling accounting year dependence in runoff triangles. *European Actuarial Journal*, **2**(2), 227–242.
SHI, P. (2012) Multivariate longitudinal modeling of insurance company expenses. *Insurance: Mathematics and Economics*, **51**(1), 204–215.
SHI, P., BASU, S. and MEYERS, G. (2012) A Bayesian log-normal model for multivariate loss reserving. *North American Actuarial Journal*, **16**(1), 29–51.
SHI, P. and FREES, E. (2011) Dependent loss reserving using copulas. *ASTIN Bulletin*, **41**(2), 449–486.
SMYTH, G. and JØRGENSEN, B. (2002) Fitting Tweedie's compound Poisson model to insurance claims data: Dispersion modelling. *ASTIN Bulletin*, **32**(1), 143–157.
SUGIURA, N. (1978) Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics: Theory and Methods*, **7**(1), 13–26.
SUN, J., FREES, E. and ROSENBERG, M. (2008) Heavy-tailed longitudinal data modeling using copulas. *Insurance Mathematics and Economics*, **42**(2), 817–830.
TWEEDIE, M. (1984) An index which distinguishes between some important exponential families. In *Statistics: Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference* (eds. J. Ghosh and J. Roy), pp. 579–604. Calcutta, India: Indian Statistical Institute.
WÜTHRICH, M. (2003) Claims reserving using Tweedie's compound Poisson model. *ASTIN Bulletin*, **33**(2), 331–346.
WÜTHRICH, M. (2012) Discussion of "A Bayesian log-normal model for multivariate loss reserving". *North America Actuarial Journal*, **16**(3), 398–401.
WÜTHRICH, M. and MERZ, M. (2008) *Stochastic Claims Reserving Methods in Insurance*. Chichester, England: John Wiley & Sons.
ZHANG, Y. (2010) A general multivariate chain ladder model. *Insurance: Mathematics and Economics*, **46**(3), 588–599.

PENG SHI
Actuarial Science, Risk Management, and Insurance Department
Wisconsin School of Business
University of Wisconsin–Madison
975 University Avenue
Madison, Wisconsin 53706, USA
E-Mail: pshi@bus.wisc.edu

https://doi.org/10.1017/asb.2013.23 Published online by Cambridge University Press

APPENDIX

A.1 Technical notes

Proof of (4):

$$\begin{aligned}\widehat{R}^{B_{\mathcal{M}}} &= \mathrm{E}\left[\widetilde{R}_{\mathcal{M}}|\mathbf{Y}^U, \mathcal{M}\right] = \int \mathrm{E}\left[R|\boldsymbol{\theta}, \mathcal{M}\right] f(\boldsymbol{\theta}|\mathbf{Y}^U, \mathcal{M})d\boldsymbol{\theta}\\ &= \int R\left[\int f(R|\boldsymbol{\theta}, \mathcal{M}) f(\boldsymbol{\theta}|\mathbf{Y}^U, \mathcal{M})d\boldsymbol{\theta}\right] dR = \int R\left[\int f(R, \boldsymbol{\theta}|\mathbf{Y}^U, \mathcal{M})d\boldsymbol{\theta}\right] dR\\ &= \int Rf(R|\mathbf{Y}^U, \mathcal{M})dR = \mathrm{E}\left[R|\mathbf{Y}^U, \mathcal{M}\right].\end{aligned}$$

Proof of (5):

$$\begin{aligned}\mathrm{MSEP}_{R|\mathbf{Y}^U,\mathcal{M}}\left(\widehat{R}^{B_{\mathcal{M}}}\right) &= \mathrm{E}\left[\left(R - \widehat{R}^{B_{\mathcal{M}}}\right)^2|\mathbf{Y}^U, \mathcal{M}\right]\\ &= \mathrm{Var}\left[R|\mathbf{Y}^U, \mathcal{M}\right] + \mathrm{E}\left[\left(\mathrm{E}\left[R|\mathbf{Y}^U, \mathcal{M}\right] - \widehat{R}^{B_{\mathcal{M}}}\right)^2|\mathbf{Y}^U, \mathcal{M}\right]\\ &= \mathrm{E}\left[\mathrm{Var}(R|\boldsymbol{\theta}, \mathcal{M})|\mathbf{Y}^U, \mathcal{M}\right] + \mathrm{Var}\left[\mathrm{E}(R|\boldsymbol{\theta}, \mathcal{M})|\mathbf{Y}^U, \mathcal{M}\right]\\ &= \mathrm{E}\left[\widetilde{V}_{\mathcal{M}}|\mathbf{Y}^U, \mathcal{M}\right] + \mathrm{Var}\left[\widetilde{R}_{\mathcal{M}}|\mathbf{Y}^U, \mathcal{M}\right].\end{aligned}$$

Proof of (6):

$$\begin{aligned}\widehat{R}^{B} &= \mathrm{E}\left[\widetilde{R}|\mathbf{Y}^U\right] = \mathrm{E}\left[\mathrm{E}\left[\mathrm{E}\left[R|\boldsymbol{\theta}, \mathcal{M}\right]|\boldsymbol{\theta}\right]|\mathbf{Y}^U\right]\\ &= \int\int\int Rf(R|\boldsymbol{\theta}, \mathcal{M}) f(\mathcal{M}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\mathbf{Y}^U)d\mathcal{M}dRd\boldsymbol{\theta}\\ &= \int\int Rf(R|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\mathbf{Y}^U)dRd\boldsymbol{\theta} = \int \mathrm{E}\left[R|\boldsymbol{\theta}\right] f(\boldsymbol{\theta}|\mathbf{Y}^U)d\boldsymbol{\theta}\\ &= \int Rf(R|\mathbf{Y}^U)dR = \mathrm{E}\left[R|\mathbf{Y}^U\right].\end{aligned}$$

Proof of (7):

$$\begin{aligned}&\mathrm{MSEP}_{R|\mathbf{Y}^U}\left(\widehat{R}^{B}\right)\\ &= \mathrm{E}\left[\left(R - \widehat{R}^{B}\right)^2|\mathbf{Y}^U\right] = \mathrm{Var}\left[R|\mathbf{Y}^U\right] + \mathrm{E}\left[\left(\mathrm{E}\left[R|\mathbf{Y}^U\right] - \widehat{R}^{B}\right)^2|\mathbf{Y}^U\right]\\ &= \mathrm{E}\left[\mathrm{Var}(R|\mathcal{M})|\mathbf{Y}^U\right] + \mathrm{Var}\left[\mathrm{E}(R|\mathcal{M})|\mathbf{Y}^U\right]\\ &= \mathrm{E}\left[\mathrm{E}\left[\mathrm{Var}[R|\boldsymbol{\theta}, \mathcal{M}]|\mathcal{M}\right]|\mathbf{Y}^U\right] + \mathrm{E}\left[\mathrm{Var}\left[\mathrm{E}[R|\boldsymbol{\theta}, \mathcal{M}]|\mathcal{M}\right]|\mathbf{Y}^U\right] + \mathrm{Var}\left[\mathrm{E}\left[\mathrm{E}[R|\boldsymbol{\theta}, \mathcal{M}]|\mathcal{M}\right]|\mathbf{Y}^U\right]\\ &= \mathrm{E}\left[\mathrm{E}\left[\widetilde{V}_{\mathcal{M}}|\mathcal{M}\right]|\mathbf{Y}^U\right] + \mathrm{E}\left[\mathrm{Var}\left[\widetilde{R}_{\mathcal{M}}|\mathcal{M}\right]|\mathbf{Y}^U\right] + \mathrm{Var}\left[\mathrm{E}\left[\widetilde{R}_{\mathcal{M}}|\mathcal{M}\right]|\mathbf{Y}^U\right].\end{aligned}$$

A.2 Run-off triangles

See Tables A.1 and A.2.

TABLE A.1

INCREMENTAL PAID LOSSES FOR WORKERS' COMPENSATION.

	Development Lag									
Accident Year	0	1	2	3	4	5	6	7	8	9
1988	3,178	3,225	2,420	1,063	733	334	121	23	27	19
1989	3,708	4,982	3,039	1,300	970	285	285	54	13	
1990	5,220	5,771	2,628	1,566	770	458	42	260		
1991	4,198	4,874	2,040	1,148	669	328	128			
1992	3,597	3,878	1,578	794	616	244				
1993	4,281	4,134	1,855	1,233	442					
1994	5,329	5,401	2,171	626						
1995	4,631	4,475	1,641							
1996	4,217	4,530								
1997	4,169									

TABLE A.2

INCREMENTAL PAID LOSSES FOR COMMERCIAL AUTO LIABILITY.

	Development Lag									
Accident Year	0	1	2	3	4	5	6	7	8	9
1988	2,491	1,935	1,058	791	269	284	275	0	8	4
1989	2,526	2,164	1,383	963	1,103	171	63	5	46	
1990	2,760	1,682	1,087	1,436	568	659	727	64		
1991	2,633	2,491	1,688	375	695	506	15			
1992	2,502	2,278	1,010	915	488	280				
1993	3,226	1,872	1,678	1,557	686					
1994	3,567	2,229	1,757	769						
1995	3,215	2,410	1,030							
1996	3,516	2,509								
1997	3,628									

TABLE A.3

CORRELATION OF UNPAID LOSSES AMONG LINES OF BUSINESS.

		ω_{12}						
		−0.75	−0.50	−0.25	0.00	0.25	0.50	0.75
λ_{12}	−0.75	−0.676	−0.499	−0.396	−0.159	−0.107	0.098	0.181
	−0.50	−0.549	−0.368	−0.317	−0.115	−0.082	0.101	0.261
	−0.25	−0.527	−0.419	−0.194	−0.110	0.031	0.116	0.309
	0.00	−0.434	−0.320	−0.072	0.103	0.120	0.221	0.381
	0.25	−0.329	−0.183	−0.031	0.041	0.278	0.368	0.538
	0.50	−0.237	−0.134	−0.019	0.174	0.300	0.457	0.541
	0.75	−0.118	−0.078	0.065	0.316	0.384	0.497	0.632

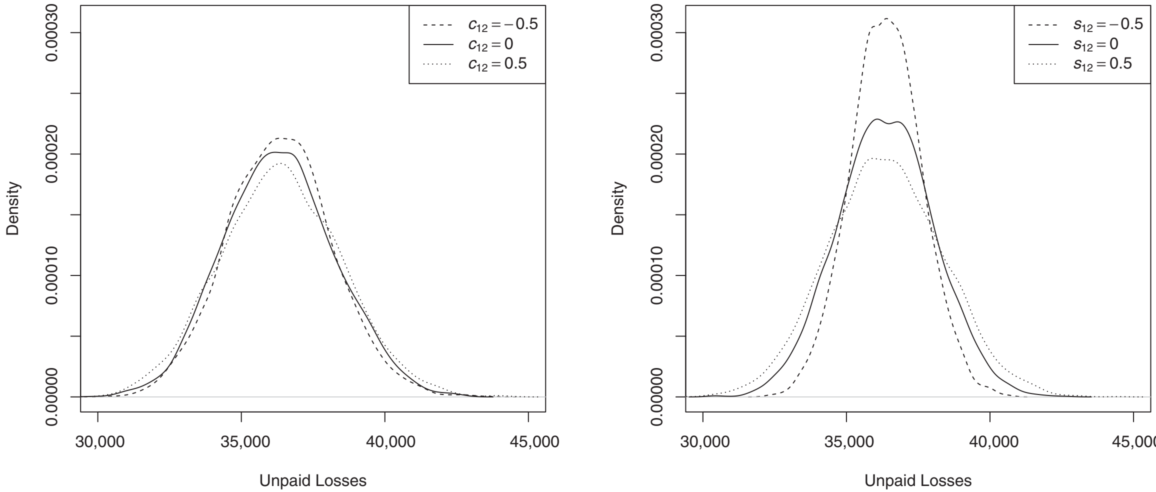

FIGURE 3: Simulated distribution of the outstanding liability of the insurance portfolio.

https://doi.org/10.1017/asb.2013.23 Published online by Cambridge University Press

A.3 Simulated correlation effects

It is well known that the dependencies among losses from different lines of business have a great implication on the determination of reserves and the associated predictive variability for the portfolio that aggregates losses from individual lines. This section provides a simple simulation study to show that the proposed method offers such flexibility to capture the complicated dependence structure within and between individual lines.

In this work, the dependency between workers' compensation and commercial auto liability is accommodated by $\mathbf{\Lambda}$ and $\mathbf{\Omega}$, where λ_{12} introduces the correlation due to calendar year effect and ω_{12} introduces the pair-wise correlation. To illustrate the way that such specification affects the correlation between payments in the two triangles, we present in Table A.3 the implied correlation coefficients based on Monte Carlo simulations under different combinations of $(\lambda_{12}, \omega_{12})$. In the simulation, we fix other association parameters at their estimated value, i.e. $c_1 = 0.272$ and $c_2 = 0.511$. Apparently, the correlation among payments is determined jointly by λ_{12} and ω_{12}: when fixing either one of the two parameters, increasing the other one increases the correlations. Thus extra flexibility is gained when allowing the two parameters to determine the dependence structure among triangles.

To further demonstrate the implications of dependency on reserve prediction, we show the predictive distribution of the outstanding liabilities for the insurance portfolio in Figure 3. The first panel shows the prediction under different c_{12} while holding s_{12} at its estimates 0.161. Similarly, the second panel shows the prediction under different s_{12} when fixing c_{12} at its estimates 0.226. First, we see that in both panels a larger association parameter lead to a wider predictive distribution, since a larger association parameter results in a higher implied association among losses and thus a weaker diversification effect. The implication for risk management practice is that the risk capital could be reduced when taking advantage of the dependencies among different lines of business. Second, the effect of s_{12} on diversification effects is larger than that of c_{12}. This is because that implied correlation among losses is also affected by the parameters c_i or s_i $(i = 1, 2)$ that determine the weight of calendar year effect and pair-wise correlation. Recall that for our insurance portfolio, only 14% of the dependency is attributed to the common calendar year effect.