Dependent Loss Reserving Using Copulas

Peng Shi
Northern Illinois University

Edward W. Frees
University of Wisconsin - Madison

July 29, 2010

Abstract

Modeling the dependence among multiple loss triangles is critical to loss reserving, risk management and capital allocation for property-casualty insurers. In this article, we propose a copula regression model for the prediction of unpaid losses for dependent lines of business. The proposed method, relating the payments in different run-off triangles through a copula function, allows us to use flexible parametric families for the loss distribution and to understand the associations among lines of business. Based on the copula model, a parametric bootstrap procedure is developed to incorporate the uncertainty in parameter estimates. In the actuarial applications, we consider an insurance portfolio consisting of personal and commercial automobile lines. When applied to the data of a major US property-casualty insurer, our method provides comparable point prediction of unpaid losses with the industry's standard practice. Moreover, our flexible structure renders the predictive distribution of unpaid losses, from which, the accident year reserves, calendar year reserves, as well as the aggregated reserves for the portfolio can be handily determined. One important implication of the dependence modeling is the diversification effect in the risk capital analysis. We demonstrate this effect by calculating the commonly used risk measures, including value at risk and conditional tail expectation, for the insurer's portfolio.

Keywords: Run-off triangle, Association, Copula Regression, Bootstrap

1 Introduction

Loss reserving is a classic actuarial reserving problem encountered extensively in property and casualty as well as health insurance. Typically, losses are arranged in a triangular fashion as they develop over time and as different obligations are incurred from year to year. This triangular format emphasizes the longitudinal and censored nature of the data. The primary goal of loss reserving is to set an adequate reserve to fund losses that have been incurred but not yet developed. For a single line of business written by an insurance company, there is an extensive actuarial literature describing alternative approaches for determining loss reserves. See, for example, Taylor (2000), England and Verrall (2002), and Wüthrich and Merz (2008).

However, almost every major insurer has more than one line of business. One can view losses from a line of business as a financial risk; it is intuitively appealing to think about these risks as being related to one another. It is well-known that if risks are associated, then the distribution of their sum depends on the association. For example, if two risks are positively correlated, then the variability of the sum of risks exceeds the sum of variabilities from each risk. Should an insurer use the loss reserve from the sum of two lines of business or the sum of loss reserves, each determined from a line of business? This problem of "additivity" was put forth by Ajne (1994) who notes that the most common approach in actuarial practice is the "silo" method. Here, an insurer divides its portfolio into several subportfolios (silos). A subportfolio can be a single line of business or can consist of several lines with homogeneous development pattern. The claim reserve and risk capital are then calculated for each silo and added up for the portfolio. The most important critique of this method is that the simple aggregation ignores the dependencies among the subportfolios.

In loss reserving, complicating the determination of dependencies among lines of business is the evolution of losses over time. As emphasized by Holmberg (1994) and Schmidt (2006), correlations may appear among losses as they develop over time (within an incurral year) or among losses in different incurral years (within a single development period). Other authors have focussed on correlations over calendar years, thinking of inflationary trends as a common unknown factor inducing correlation.

Much of the work on multivariate stochastic reserving methods to date has involved extending the distribution-free method of Mack (1993). Braun (2004) proposed to estimate the prediction error for a portfolio of correlated loss triangles based on a multivariate chain-ladder method. Similarly, Merz and Wüthrich (2008) considered the prediction error of another version of the multivariate chain-ladder model by Schmidt (2006), where the dependence structure was incorporated into parameter estimates. Within the theory of linear models, Hess et al. (2006) and Merz and Wüthrich (2009b) provided the optimal predictor and the prediction error for the multivariate additive loss reserving method, respectively. Motivated by the fact that not all subportfolios satisfy the same homogeneity assumption, Merz and Wüthrich (2009a) combined chain-ladder and additive loss reserving methods into one single framework. Zhang (2010) proposed a general multivariate chain-ladder model that introduces correlations among triangles using the seemingly unrelated regression technique.

These procedures have desirable properties, focusing on the mean square error of predictions. In this paper, we focus instead on tails of the distribution. Because of this, and the small sample size typically encountered in loss reserving problems, we look more to parametric methods based on distributional families. For example, in a multivariate context, Brehm (2002) employed a log-normal model for the unpaid losses of each line and a normal copula for the generation of the joint distribution. The dispersion matrix in the copula was estimated using correlations among calendar year inflation in different lines of business. Kirschner et al. (2002, 2008) presented two approaches to calculate reserve indications for correlated lines, among which, a synchronized bootstrap was suggested to resample the variability parameters for multiple triangles. Following and generalizing this result, Taylor and McGuire (2007) examined the similar bootstrap method under the generalized linear model framework, and demonstrated the calculations of loss reserves and their prediction errors. de Jong (2010) employed factor analytic techniques to handle several sources of time dependencies (by incurral year, development year, calendar year) as well as correlations among lines of business in a flexible manner.

An alternative parametric approach involving Bayesian methods has found applications when studying loss reserves for single lines of business. Some recent work include de Alba (2006), de Alba and Nieto-Barajas (2008) and Meyers (2009). The Bayesian methods for multivariate loss reserving problems have rarely been found in the literature. Merz and Wüthrich (2010) is one example, where the authors considered a bivariate Bayesian model for combining data from the paid and incurred triangles to achieve better prediction.

We employ a copula method to associate the claims from multiple run-off triangles. Despite the application of copulas in Brehm (2002) and de Jong (2010), both are focused on correlations in a model based on normal distributions. In contrast, the focus of this paper is to show how one can use a wide range of parametric families for the loss distribution to understand associations among lines of business.

Our reliance on a parametric approach has both strengths and limitations. A strength of the parametric approach is that is has historically been used for small data sets such as is typical in the loss reserve setting. With the parametric approach, we can use diagnostic methods to check model assumptions. To illustrate, in our example of data from a major US insurer, we show that the lognormal distribution is appropriate for the personal automobile line whereas the gamma distribution is appropriate for the commercial auto line. Because of our reliance on parametric families, we are able to provide an entire predictive distribution for "silo" (single line of business) loss reserves as well as for the entire portfolio. Traditionally, parametric approaches have been limited because they do not incorporate parameter uncertainty into statistical inference. However, we are able to use modern parametric bootstrapping to overcome this limitation.

A limitation of the approach presented in this paper is that we focus on the cross-sectional dependence among lines of business. We incorporate time patterns through deterministic parameters, similarly to that historically done in a loss reserve setting. Our goal is to provide a simple alternative way to view the dependence among multiple loss triangles. We show that the dependency is critical in determining an insurer's reserve ranges and risk capital. To show that our loss reserve forecasts are not ad hoc, we compare our results with the industry's standard practice, the chain-ladder prediction, as well as other alternative multivariate loss reserving methods.

The outline of this article is as follows: Section 2 introduces the copula regression model to associate losses from multiple triangles. Section 3 presents the run-off triangle data and model fit results. Section 4 discusses the predictive distribution of unpaid losses and shows its implications in determining the accident year reserves, calendar year reserves, as well as aggregate reserves. Section 5 illustrates the diversification effect of dependent loss triangles in a risk capital analysis. Section 6 concludes the article.

2 Modeling

In a loss reserving context, each element of a run-off triangle may represent incremental payments or cumulative payments, depending on the situation. Our approach applies to the incremental paid losses. Assume that an insurance portfolio consists of N subportfolios (triangles). Let i indicate the year in which an accident is incurred, and j indicate the development lag, that is the number of years from the occurrence to the time when the payment is made. Define $X_{ij}^{(n)}$ as the incremental claims in the ith accident year and the jth development year. The superscript (n), $n \in \{1, \cdots, N\}$, indicates the nth run-off triangle. Thus, the random vector of multivariate incremental claims can be expressed by

$$\mathbf{X}_{ij} = (X_{ij}^{(1)}, \cdots, X_{ij}^{(N_{ij})}),\ i \in \{0, \cdots, I\} \text{ and } j \in \{0, \cdots, J\},$$

where I denotes the most recent accident year and J denotes the latest development year. Typically, we have $I \geq J$. Note that we allow the imbalance in the multivariate triangles. With N_{ij} being the dimension of the incremental claim vector for accident year i and development lag j, $N_{ij} < N$ implies the lack of balance and $N_{ij} = N$ the complete design. The imbalance could be due to missing values in run-off triangles or the different size of each portfolio.

With above notations, the claims reserves for accident year i and calendar year k, at time I, can be shown as

$$\sum_{j=I+1-i}^{J} \mathbf{X}_{ij} \text{ for } i \in \{I+1-J, \cdots, I\},\ \text{and} \sum_{i+j=k} \mathbf{X}_{ij} \text{ for } k \in \{I+1, \cdots, I+J\},$$

respectively. Our interest is to forecast the unpaid losses in the lower-right-hand triangle, based on the observed payments in the upper-left-hand triangle. At the same time, we take into account of the dependencies among multiple run-offs in the parameter estimation and loss reserve indication. Here, we assume that all the claims will be closed in J years, that is, all payments are made within the next J years after the occurrence of an accident.

2.1 Distributional Models

Due to the typical small sample size of run-off triangles, we focus on the modeling of incremental payments based on parametric distributional families. In this context, claims are usually normalized by an exposure variable that measures the volume of the business. We define the normalized incremental claims as $Y_{ij}^{(n)} = X_{ij}^{(n)}/\omega_i^{(n)}$, where $\omega_i^{(n)}$ denotes the exposure for the ith accident year in the nth triangle. Assume that $Y_{ij}^{(n)}$ is from a parametric distribution:

$$F_{ij}^{(n)} = \text{Prob}(Y_{ij}^{(n)} \leq y_{ij}^{(n)}) = F^{(n)}(y_{ij}^{(n)}; \eta_{ij}^{(n)}, \boldsymbol{\gamma}^{(n)}),\ n = 1, \cdots, N. \tag{1}$$

Here, we allow different distributional families $F^{(n)}(\cdot)$ for incremental losses of lines with different characteristics or development patterns. In claims reserving problems, the systematic component $\eta_{ij}^{(n)}$, which determines the location, is often a linear function of explanatory variables (covariates), that is $\eta_{ij}^{(n)} = \mathbf{x}_{ij}^{(n)\prime}\boldsymbol{\beta}^{(n)}$. The covariate vector $\mathbf{x}_{ij}^{(n)}$ includes the predictive variables that are used for forecasting the unpaid losses in the nth triangle, and $\boldsymbol{\beta}^{(n)}$ represents the corresponding coefficients to be estimated. The vector $\boldsymbol{\gamma}^{(n)}$, summarizing additional parameters in the distribution of $Y_{ij}^{(n)}$, determines the shape and scale. Except for the location parameter, we assume that all other parameters are the same for incremental claims within each individual run-off triangle.

Among parametric distributional families, the log-normal and gamma distributions have been extensively studied for incremental claims in the loss reserving literature. The log-normal model, introduced by Kremer (1982), examined the logarithm of incremental losses and used the multiplicative structure for the mean. Another approach based on a log-normal distribution is the Hoerl curve (see England and Verrall (2002)). Using a log link function, the Hoerl curve replaced the chain-ladder type systematic component in the log-normal model with one that is linear in development lag and and its logarithm. With the same linear predictor as in the chain-ladder method, Mack (1991) proposed using a gamma distribution for claim amounts. Other parametric approaches, including the Wright's model (see Wright (1990)) and the generalized linear model (GLM) framework (see Renshaw and Verrall (1998)), also considered the gamma distribution for incremental claims.

In our applications, we follow the idea behind the chain-ladder model and use two factors, accident year and development lag, for covariates. Thus, the systematic component for the nth subportfolio can be expressed as:

$$\eta_{ij}^{(n)} = \zeta^{(n)} + \alpha_i^{(n)} + \tau_j^{(n)},\ n = 1, \cdots, N, \tag{2}$$

where constraints $\alpha_0^{(n)} = 0$ and $\tau_0^{(n)} = 0$ are used in the model development. Specifically, we consider the form $\eta_{ij}^{(n)} = \mu_{ij}^{(n)}$ for a log-normal distribution with location parameter μ and scale parameter σ. For a gamma distribution with shape parameter κ and scale parameter θ, one could apply the canonical inverse link $\eta_{ij}^{(n)} = (\kappa^{(n)}\theta_{ij}^{(n)})^{-1}$ in the GLM framework. Alternatively, as pointed out by Wüthrich and Merz (2008), a log-link $\eta_{ij}^{(n)} = \log(\kappa^{(n)}\theta_{ij}^{(n)})$ is typically a natural choice in the insurance reserving context.

2.2 Copula Regression

For large property-casualty insurers, different lines of business are very often related and reserve indications must reflect the dependencies among the corresponding multiple loss triangles. In a regression context, a natural choice to accommodate the dependence among lines of business is the seemingly unrelated regression (SUR) introduced by Zellner (1962). The SUR extends the linear model and allows correlated errors between equations. However, due to the long-tailed nature, insurance data are often phrased in a non-linear regression framework, such as GLMs (see de Jong and Heller (2008)). Within a GLM, one can introduce correlations via latent variables. Both approaches to dependency modeling are limited to the concept of linear correlation. Furthermore, assuming a common distributional family for all triangles might not be appropriate, since subportfolios often present heterogeneous development patterns. Merz and Wüthrich (2009a) addressed this problem by combining the multivariate chain-ladder and the multivariate additive loss reserving method and thus allowing the chain-ladder for one triangle and the additive loss reserving method for the other in a portfolio. However, no work has appeared to date to address the same problem in a parametric setup.

In this work, we employ parametric copulas to understand the dependencies among run-off triangles. In stead of linear correlation, we examine a more general concept of dependence - association. A copula is a multivariate distribution with all marginals following uniform distribution on $[0, 1]$. It is a useful tool for understanding relationships (both linear and nonlinear) among multiple responses (see Joe (1997)). For statistical inference and prediction purposes, it is more interesting to place a copula in a multivariate regression context. The application of copula regression in actuarial science is recent. Frees and Wang (2005, 2006) developed a copula-based credibility estimates for longitudinal insurance claims. Sun et al. (2008) employed copulas in a similar manner to forecast nursing home utilization. Frees and Valdez (2008) and Frees et al. (2009) used copulas to accommodate the dependencies among claims from various types of coverage in auto insurance. Shi and Frees (2010) introduced a longitudinal quantile regression model using copulas to examine insurance company expenses.

Consider a simple case where an insurance portfolio consists of two lines of business (N=2). According to Sklar's theorem (see Nelsen (2006)), the joint distribution of normalized incremental claims $(Y_{ij}^{(1)}, Y_{ij}^{(2)})$ can be uniquely represented by a copula function as

$$F_{ij}(y_{ij}^{(1)}, y_{ij}^{(2)}) = \text{Prob}(Y_{ij}^{(1)} \leq y_{ij}^{(1)}, Y_{ij}^{(2)} \leq y_{ij}^{(2)}) = C(F_{ij}^{(1)}, F_{ij}^{(2)}; \boldsymbol{\phi}), \tag{3}$$

where $C(\cdot; \boldsymbol{\phi})$ denotes the copula function with parameter vector $\boldsymbol{\phi}$, and marginal distribution functions $F_{ij}^{(1)}$ and $F_{ij}^{(2)}$ follow equation (1). This specification renders the flexibility of modeling claims of the two subportfolios with different distributional families.

In model (3), the dependence between two run-off triangles is captured by the association parameter $\boldsymbol{\phi}$. In addition to the linear correlation, it also measures non-linear relationships. Some non-linear association measures include Spearman's rho ρ_s and Kandall's tau ρ_τ:

$$\rho_s(Y_{ij}^{(1)}, Y_{ij}^{(2)}) = 12 \int\int_{[0,1]^2} C(u, v)\mathrm{d}u\mathrm{d}v - 3, \tag{4}$$

$$\rho_\tau(Y_{ij}^{(1)}, Y_{ij}^{(2)}) = 4 \int\int_{[0,1]^2} C(u, v)\frac{\partial^2 C}{\partial u \partial v}(u, v)\mathrm{d}u\mathrm{d}v - 1. \tag{5}$$

Another type of non-linear association is the tail dependence. Based on the copula C, the upper and lower tail dependence can be derived by:

$$\rho_{Upper}(Y_{ij}^{(1)}, Y_{ij}^{(2)}) = \lim_{u \to 1^-} \frac{\bar{C}(u, u)}{1 - u}, \text{ and } \rho_{Lower}(Y_{ij}^{(1)}, Y_{ij}^{(2)}) = \lim_{u \to 0^+} \frac{C(u, u)}{u}, \tag{6}$$

where $\bar{C}(u, v)$ denotes the associated survival copula $\bar{C}(u, v) = 1 - u - v + C(u, v)$. Note all above non-linear dependence measures only depend on the association parameter $\boldsymbol{\phi}$. For example, a bivariate frank copula, which captures both positive and negative association, is defined as

$$C(u, v) = \frac{1}{\phi} \log\left(1 + \frac{(e^{-\phi u} - 1)(e^{-\phi v} - 1)}{e^{-\phi} - 1}\right).$$

It is straight forward to show that the corresponding Spearman's rho and Kandall's tau are:

$$\rho_s(Y_{ij}^{(1)}, Y_{ij}^{(2)}) = 1 - \frac{4}{\phi}[1 - D_1(\phi)],$$

$$\rho_\tau(Y_{ij}^{(1)}, Y_{ij}^{(2)}) = 1 - \frac{12}{\phi}[D_1(\phi) - D_2(\phi)],$$

where $D_k(\cdot),\ k = 1$ or 2, denotes the Debye function.

As a parametric approach, model (3) can be easily estimated using a likelihood based estimation method. Let $c(\cdot)$ denote the probability density function corresponding to the copula distribution function $C(\cdot)$. The log-likelihood function for the insurance portfolio is:

$$L = \sum_{i=0}^{I}\sum_{j=0}^{I-i} \ln c(F_{ij}^{(1)}, F_{ij}^{(2)}; \boldsymbol{\phi}) + \sum_{i=0}^{I}\sum_{j=0}^{I-i} \ln(f_{ij}^{(1)} + f_{ij}^{(2)}), \tag{7}$$

where $f_{ij}^{(n)}$ denotes the density of marginal distribution $F_{ij}^{(n)}$, that is $f_{ij}^{(n)} = f^{(n)}(y_{ij}^{(n)}; \eta_{ij}^{(n)}, \boldsymbol{\gamma}^{(n)})$ for $n = 1,\ 2$. The model is estimated using observed paid losses $y_{ij}^{(n)}$, for $(i, j) \in \{(i, j) : i + j \leq I\}$, and a reserve is set up to cover future payments $y_{ij}^{(n)}$, for $(i, j) \in \{(i, j) : i + j > I\}$.

One benefit of dependence modeling using copulas is that a copula preserves the shapes of marginals. Thus, one can take advantage of standard statistical inference procedures in choosing marginal distributions, that is the distributional family for each claims triangle. Regarding the association between triangles, various approaches have been proposed for the choice of copulas (see

Genest et al. (2009) for a comprehensive review). We exploit this specification by examining the Akaike's Information Criterion (AIC). In addition, due to the parametric setup, the entire predictive distribution for the loss reserves of each line of business as well as for the portfolio can be derived using Monte Carlos simulation techniques.

A limitation of parametric approaches is that the parameter uncertainty is not incorporated into statistical inference. To tackle this issue, one can consider a Bayesian framework, where the data are used to improve the prior and hence the new posterior distribution are used together with the sampling distribution to compute the predictive distribution for unpaid losses. However, we take a frequentist's perspective and choose to overcome this limitation by using modern bootstrapping. The detailed simulation and bootstrapping procedures are summarized in Appendix A.1.

2.3 Model Extension

This section discusses the potential extensions to the copula regression model by relaxing the model assumptions. We intend to provide a more general framework for dependent loss reserving, though the empirical analysis will focus on the basic setup. The first generalization is to adapt model (3) to the case of multivariate ($N > 2$) run-off triangles. Similar to the bivariate case, the model enjoys the computational advantage. Rewriting model (3), the joint density of $(Y_{ij}^{(1)}, \cdots, Y_{ij}^{(N_{ij})})$ can be expressed by:

$$f_{ij}(y_{ij}^{(1)}, \cdots, y_{ij}^{(N_{ij})}) = c(F_{ij}^{(1)}, \cdots, F_{ij}^{(N_{ij})}; \boldsymbol{\phi}) \prod_{n=1}^{N_{ij}} f_{ij}^{(n)}. \tag{8}$$

For the case of unbalanced data $N_{ij} < N$, the copula density in (8) will be replaced with the corresponding sub-copula. Thus, the parameters can be estimated by maximizing the log-likelihood function:

$$L = \sum_{i=0}^{I} \sum_{j=0}^{I-i} \ln c(F_{ij}^{(1)}, \cdots, F_{ij}^{(N_{ij})}; \boldsymbol{\phi}) + \sum_{i=0}^{I} \sum_{j=0}^{I-i} \sum_{n=1}^{N_{ij}} \ln f_{ij}^{(n)}. \tag{9}$$

To accommodate the pairwise association among N triangles, one could employ the family of elliptical copulas. The definition of elliptical copulas is given in Appendix A.2. A natural way to introduce dependency is through the association matrix $\boldsymbol{\Sigma}$ of an elliptical copula:

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1N} \\ \rho_{21} & 1 & \cdots & \rho_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{N1} & \rho_{N2} & \cdots & 1 \end{pmatrix}. \tag{10}$$

Here, $\rho_{ij} = \rho_{ji}$ captures the pairwise association between the ith and jth triangles. The sub-copula for model (8) is the elliptical copula generated by the corresponding sub-matrix of $\boldsymbol{\Sigma}$.

In both models (3) and (8), we assume an identical association for all claims in the triangle, regardless of the accident year and development lag. This assumption could be relaxed by specifying different copulas for claims with regard to the accident year or development lag. For example, the association among run-off triangles could vary over accident years, then the model (8)-(9) becomes (11)-(12), respectively:

$$f_{ij}(y_{ij}^{(1)},\cdots,y_{ij}^{(N_{ij})}) = c_i(F_{ij}^{(1)},\cdots,F_{ij}^{(N_{ij})};\boldsymbol{\phi})\prod_{n=1}^{N_{ij}} f_{ij}^{(n)},\ \ i=0,\cdots,I \tag{11}$$

$$L = \sum_{i=0}^{I}\sum_{j=0}^{I-i}\ln c_i(F_{ij}^{(1)},\cdots,F_{ij}^{(N_{ij})};\boldsymbol{\phi}) + \sum_{i=0}^{I}\sum_{j=0}^{I-i}\sum_{n=1}^{N_{ij}}\ln f_{ij}^{(n)}. \tag{12}$$

In the above specification, copula functions c_i for $i \in 1,\cdots,I$ could be from the same distribution with different association matrix $\mathbf{\Sigma}_i$. Or they might have the same association structure but are based on different distributions, for example, the normal copula is used for one accident year, while the t-copula for the other. Following the same rationale, we could allow the association among triangles to vary over development years or calendar years.

The more general and also more complicated case is when the independence assumption for claims in each triangle is relaxed. Within a single triangle, the incremental payments may present dependency over development lags or calendar years. To introduce such type of dependence, one might refer to multivariate longitudinal modeling techniques. In the copula regression framework, to capture the association within and between triangles simultaneously, we choose to replace matrix (10) with:

$$\mathbf{\Sigma} = \begin{pmatrix} P_1 & \sigma_{12}P_{12} & \cdots & \sigma_{1N}P_{1N} \\ \sigma_{21}P_{21} & P_2 & \cdots & \sigma_{2N}P_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1}P_{N1} & \sigma_{N2}P_{N2} & \cdots & P_N \end{pmatrix}. \tag{13}$$

In the formulation (13), P_n, $n = 1,\cdots,N$ is a correlation matrix that describes the association for claims within the nth triangle. $\sigma_{ij} = \sigma_{ji}$ measures the concurrent association between the ith and jth triangles. P_{ij} implies the lag correlation that is straightforward to be derived from P_i and P_j. As mentioned before, our goal is to provide a general modeling framework for dependent loss reserving. We leave the detailed discussion of this complicated case to the future study.

3 Empirical Analysis

The copula regression model is applied to the claims triangles of a major US property-casualty insurer. We pay more attention to the data analysis and select models that are closely fit by the data. Also, we carefully interpret the association between lines of business.

3.1 Data

The run-off triangle data are from the Schedule P of the National Association of Insurance Commissioners (NAIC) database. The NAIC is an organization of insurance regulators that provides a forum to promote uniformity in the regulation among different states. It maintains one of the world's largest insurance regulatory databases, including the statutory accounting report for all insurance companies in the United States. The Schedule P includes firm level run-off triangles of aggregated claims for major personal and commercial lines of business for property-casualty insurers. And the triangles are available for both incurred and paid losses.

We consider the triangles of paid losses in Schedule P of year 1997. Each triangle contains losses for accident years 1988-1997 and at most ten development years. The preliminary analysis shows that the dependencies among lines of business vary across firms. As a result, our analysis will focus on one single major insurance company. Recall that we assume that all claims will be closed in I (=10 in our case) years. This assumption is not reasonable for long-tail lines of business. Thus, we limit our application to an insurance portfolio that consists of two lines of business with relative short tails, personal auto and commercial auto.

Table 1 and Table 2 display the cumulative paid losses for personal and commercial auto lines, respectively. We observe that the portfolio is not evenly distributed in the two lines of business, with personal auto much larger than the commercial auto. In loss reserving literature, payments are typically normalized by an exposure variable that measures the volume of the business, such as number of policies or premiums. We normalize the payment by dividing by the net premiums earned in the corresponding accident year, and we focus on the normalized incremental payments in the following analysis. The exposure variable is also exhibited in the above tables.

To examine the development pattern of each triangle, we present the multiple time series plot of loss ratios for personal and commercial auto lines in Figure 1. Each line corresponds to an accident year. The decreasing trend confirms the assumption that all claims will be closed within ten years. A comparison of the two panels shows that the development of the personal automobile line is less volatile than that of the commercial automobile line.

Table 1. Cumulative Paid Losses for Personal Auto Line (in thousand of dollars)

		Development Lag									
Accident Year	Premiums	0	1	2	3	4	5	6	7	8	9
1988	4,711,333	1,376,384	2,587,552	3,123,435	3,437,225	3,605,367	3,685,339	3,724,574	3,739,604	3,750,469	3,754,555
1989	5,335,525	1,576,278	3,013,428	3,665,873	4,008,567	4,197,366	4,274,322	4,309,364	4,326,453	4,338,960	
1990	5,947,504	1,763,277	3,303,508	3,982,467	4,346,666	4,523,774	4,601,943	4,649,334	4,674,622		
1991	6,354,197	1,779,698	3,278,229	3,939,630	4,261,064	4,423,642	4,508,223	4,561,672			
1992	6,738,172	1,843,224	3,416,828	4,029,923	4,329,396	4,506,238	4,612,534				
1993	7,079,444	1,962,385	3,482,683	4,064,615	4,412,049	4,650,424					
1994	7,254,832	2,033,371	3,463,912	4,097,412	4,529,669						
1995	7,739,379	2,072,061	3,530,602	4,257,700							
1996	8,154,065	2,210,754	3,728,255								
1997	8,435,918	2,206,886									

Table 2. Cumulative Paid Losses for Commercial Auto Line (in thousand of dollars)

		Development Lag									
Accident Year	Premiums	0	1	2	3	4	5	6	7	8	9
1988	267,666	33,810	79,128	125,677	160,883	184,243	196,745	203,347	206,720	209,093	209,871
1989	274,526	37,663	89,434	130,432	159,928	172,597	183,801	189,586	193,806	195,716	
1990	268,161	40,630	96,948	153,130	185,603	201,431	209,840	216,960	218,085		
1991	276,821	40,475	90,172	129,485	153,529	166,685	179,280	182,188			
1992	270,214	37,127	88,110	122,264	147,719	167,140	172,868				
1993	280,568	41,125	94,427	134,716	174,628	181,278					
1994	344,915	57,515	125,396	212,130	230,239						
1995	371,139	61,553	193,761	214,684							
1996	323,753	112,103	145,353								
1997	221,448	37,554									

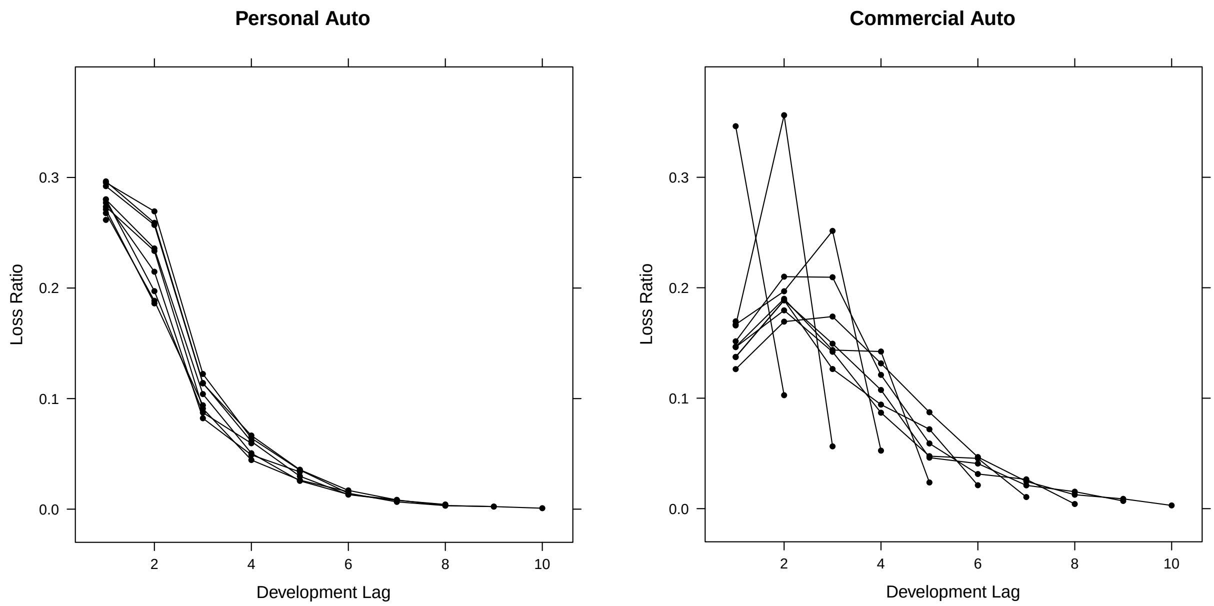

Figure 1: Multiple time series plots of loss ratios for personal auto and commercial auto lines.

The scatter plot of loss ratios is exhibited in Figure 2. This plot suggests a strong positive, although nonlinear, relationship between commercial and personal auto lines. In fact, the corresponding Pearson correlation is 0.725. Since we are comparing the payments from two triangles of the same accident year and development lag, the strong correlation reflects the effects of the potential distortions that might affect all open claims. Such distortion could be a calendar year inflation, for example, a decision to accelerate the payments in all business lines.

3.2 Model Inference

This section fits the copula regression model. Since a copula splits the modeling of marginals and dependence structure, one could evaluate the goodness-of-fit for the marginal and joint distributions separately. As for marginals, preliminary analysis suggests that a lognormal regression is appropriate for the personal auto line and a gamma regression is appropriate for the commercial auto line. To show the reasonable model fits for the two triangles, we exhibit the qq-plots of marginals for personal and commercial auto lines in Figure 3. Note that the analysis is performed on residuals from each regression model, because one wants to take out the effects of covariates (the accident year and development year effects for our case). For the lognormal regression, the residual is defined as $\hat{\varepsilon}_{ij} = (\ln y_{ij} - \hat{\mu}_{ij})/\hat{\sigma}$, and for the gamma regression, the residual is defined as $\hat{\varepsilon}_{ij} = y_{ij}/\hat{\theta}_{ij}$. These plots show that the marginal distributions for personal and commercial auto lines seem to be well-specified. There is some concern that the lower tail of the commercial auto distribution could be improved.

The results of formal statistical tests are reported in Table 3. The three goodness-of-fit statistics assess the relationship between the empirical distribution and the estimated parametric distribution.

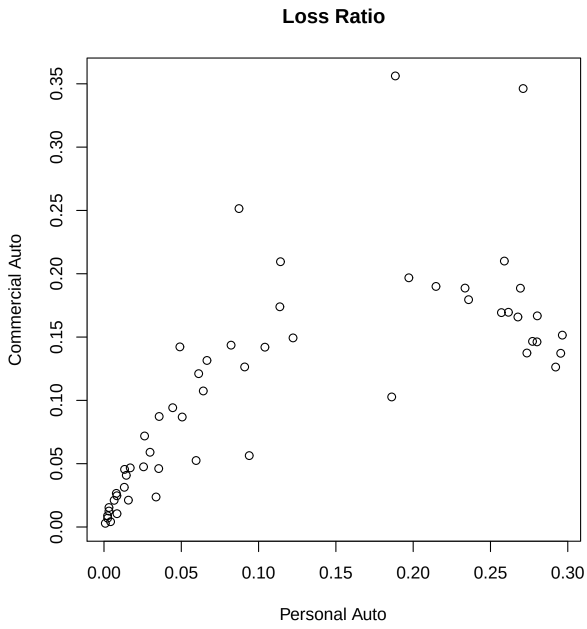

Figure 2: Loss ratios of personal auto line versus commercial auto line.

A large p-value indicates a nonsignificant difference between the two. Both probability plots and hypothesis tests suggest that the lognormal regression and gamma regression fit well for personal and commercial auto lines, respectively.

Table 3. p-values of Goodness-of-Fit

	Personal Auto (Lognormal)	Commercial Auto (Gamma)
Kolmogorov-Smirnov	>0.150	0.081
Cramer-von Mises	>0.250	0.135
Anderson-Darling	>0.250	0.125

With the specifications of marginals, we reexamine the dependence between the two lines. Recall that there is a strong positive correlation between the loss ratios of personal and commercial auto lines. This correlation might reflect, to some extent, the accident year and/or development year effects. To isolate these effects, we look at the relationship between the percentile ranks of residuals from the two lines, as shown in Figure 4. The percentile rank is calculated by $P(\hat{\varepsilon}_{ij}) = \hat{G}(\hat{\varepsilon}_{ij})$, where $\hat{G}$ denotes the estimated distribution function of the residual. In our calculation, $\hat{G}$ represents a standard normal distribution for the personal auto line and a gamma distribution with shape parameter $\hat{\kappa}$ and scale parameter 1 for the commercial auto line.

The scatter plot in Figure 4 implies a negative relationship between residuals of the two triangles. The correlation coefficient is -0.2. It is noteworthy that the loss ratios from personal and commercial auto lines become negatively correlated, after purging off the effects of accident year

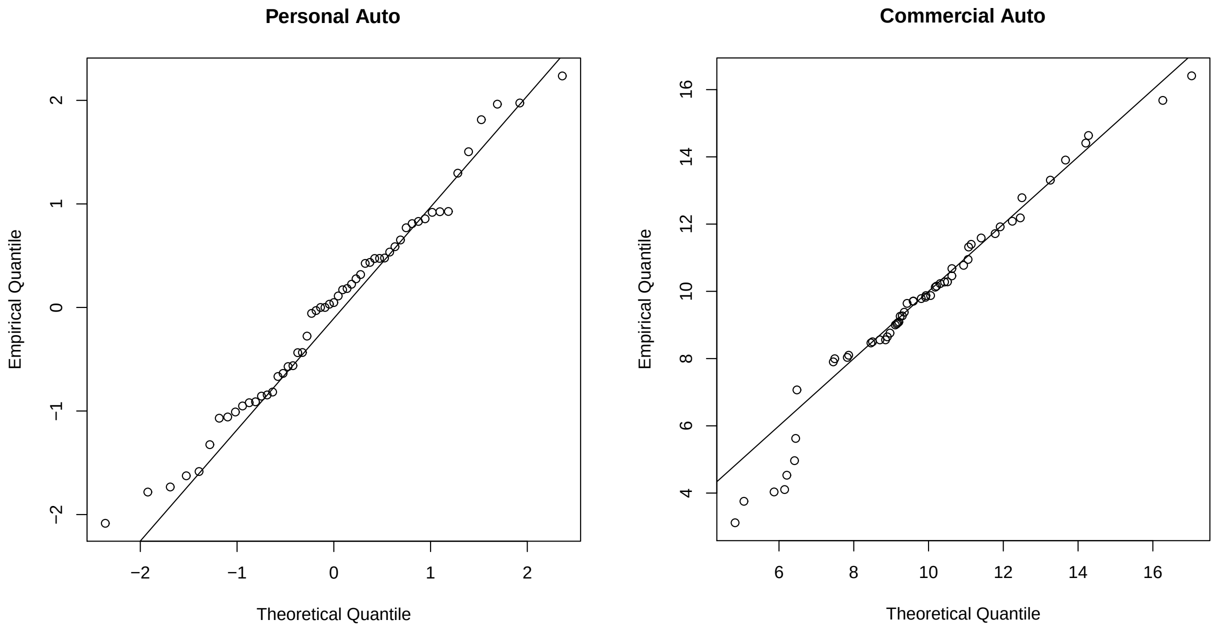

Figure 3: QQ plots of marginals for personal and commercial auto lines.

and development lag. This is an important implication from the risk management perspective, as will be shown in Section 5.

The above analysis suggests that an appropriate copula should be able to accommodate negative correlation. We consider the Frank copula and the Gaussian copula in this study. For comparison purposes, we also examine the product copula that assumes independence and is a special case of the other two. The likelihood-based method is used to estimate the copula regression model, and the estimation results are summarized in Table 4.

We report the parameter estimates, the corresponding t statistics, as well as the the value of the log likelihood function for each model. The result suggests that the negative association between personal and commercial auto lines are not negligible. First, the t-statistics for the dependence parameter in both Frank and Gaussian copula models indicate significant association. In the Frank copula, a dependence parameter of -2.60 corresponds to a Spearmans rho of -0.39. In the Gaussian copula, a dependence parameter of -0.36 corresponds to a Spearmans rho of -0.34. Second, since both models nest the product copula as a special case, we can perform a likelihood ratio test to examine the model fit. Compared with the independence case, the Frank copula model gives a χ^2 statistics of 5.12, and the gaussian copula model gives a χ^2 statistics of 6.84. Consistently, the model is of better fit when incorporating the dependence between the two lines of business.

The model selection is based on a likelihood-based goodness-of-fit measure. According to the AIC, we choose the Gaussian copula as our final model for the determination of reserves. As a step of model validation, one needs to examine how well the Gaussian copula fits the data. We adopt the t-plot method introduced by Sun et al. (2008). The t-plot employs the properties of the elliptical

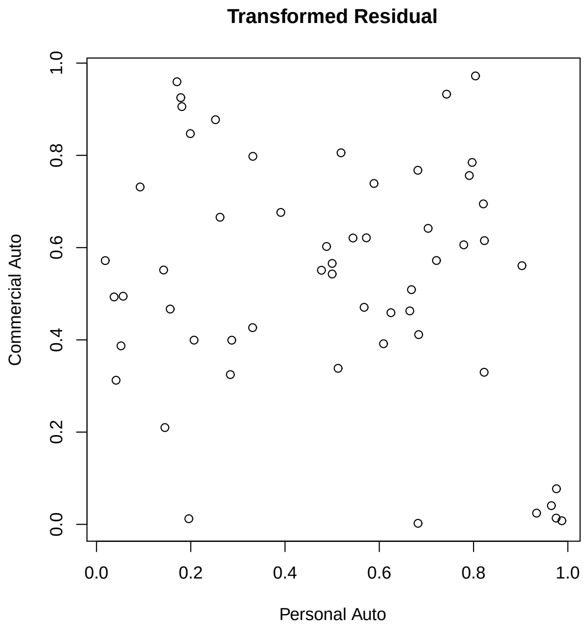

Figure 4: Scatter plot of residual percentiles from commercial and personal auto lines.

distribution and is designed to evaluate the goodness-of-fit for the family of elliptical copulas. We display the plot in Figure 5. The linear trend along the 45 degree line provides evidence that the Gaussian copula is a suitable model for the dependency. A statistical test is performed for sample correlation. The correlation coefficient between the sample and theoretical quantiles is 0.914. Based on 5,000 simulation, the p-value for the correlation is 0.634, indicating the nonsignificant difference between the empirical and theoretical distributions.

Table 4. Estimates for the Copula Regression Model with Different Copula Specifications

	Product				Frank				Gaussian			
	Personal Auto		Commercial Auto		Personal Auto		Commercial Auto		Personal Auto		Commercial Auto	
	Estimates	t-stat	Estimates	t-stat	Estimates	t-stat	Estimates	t-stat	Estimates	t-stat	Estimates	t-stat
Intercept	-1.1367	-26.55	5.8029	5.26	-1.1182	-22.49	6.1506	4.87	-1.1185	-26.11	5.8557	5.42
AY=1989	-0.0327	-0.78	0.6650	0.47	-0.0547	-1.23	-0.3025	-0.21	-0.0441	-1.07	0.1851	0.13
AY=1990	-0.0284	-0.65	-0.3299	-0.25	-0.0470	-0.98	-0.6312	-0.45	-0.0493	-1.13	-0.4864	-0.38
AY=1991	-0.1309	-2.86	1.0813	0.73	-0.1552	-3.01	0.5782	0.37	-0.1517	-3.30	0.8870	0.61
AY=1992	-0.1747	-3.62	1.0568	0.72	-0.1962	-3.66	0.8063	0.50	-0.1923	-3.99	0.9771	0.67
AY=1993	-0.1745	-3.39	0.5652	0.40	-0.1879	-3.49	0.6472	0.44	-0.1956	-3.79	0.6354	0.45
AY=1994	-0.1729	-3.10	-0.1869	-0.14	-0.1921	-3.34	-0.1393	-0.10	-0.1950	-3.48	-0.1622	-0.12
AY=1995	-0.2234	-3.60	-0.3964	-0.29	-0.2533	-4.13	-0.7472	-0.52	-0.2564	-4.05	-0.5061	-0.37
AY=1996	-0.2444	-3.35	-0.8879	-0.63	-0.2670	-2.40	-1.1801	-0.74	-0.2725	-3.69	-1.1194	-0.80
AY=1997	-0.2042	-2.07	0.0935	0.04	-0.2164	-2.06	-0.2457	-0.10	-0.2190	-2.22	0.0406	0.02
Dev=2	-0.2244	-5.37	-0.8418	-1.02	-0.2182	-4.65	-1.0517	-1.16	-0.2212	-5.27	-0.7457	-0.90
Dev=3	-1.0469	-23.95	0.3268	0.33	-1.0421	-22.10	0.1202	0.12	-1.0480	-23.97	0.3609	0.37
Dev=4	-1.6441	-35.90	3.3330	2.49	-1.6331	-32.81	3.3388	2.34	-1.6455	-35.90	3.3571	2.52
Dev=5	-2.2540	-46.68	11.5877	4.73	-2.2418	-43.49	12.6428	4.95	-2.2582	-46.74	11.7195	4.77
Dev=6	-3.0130	-58.57	20.6612	5.23	-2.9953	-54.34	21.7878	5.20	-3.0158	-58.60	20.8855	5.28
Dev=7	-3.6713	-65.89	42.1216	5.39	-3.6508	-64.08	44.1762	5.56	-3.6760	-65.94	42.6381	5.42
Dev=8	-4.4935	-72.39	87.3407	5.03	-4.4844	-73.26	94.3617	5.74	-4.5040	-72.55	89.5681	5.07
Dev=9	-4.9109	-67.29	120.2184	4.18	-4.9136	-64.05	120.3563	4.14	-4.9205	-67.31	120.3904	4.18
Dev=10	-5.9134	-60.06	344.7043	3.05	-5.9227	-56.48	344.5201	2.93	-5.9264	-60.03	344.4750	3.05
Scale/Shape	0.0887	10.49	9.6425	5.33	0.0913	10.07	9.2518	5.29	0.0890	10.40	9.6002	5.31
Dependence					-2.6021	-2.2723			-0.3586	-2.8961		
LogLik	345.3001				347.8606				348.7210			

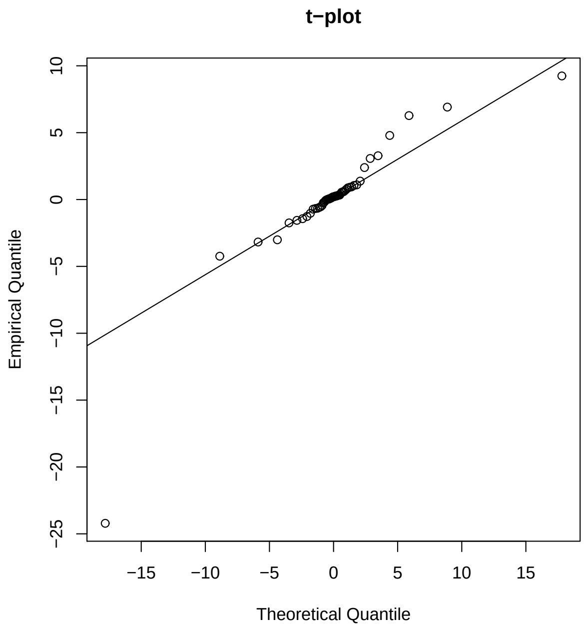

Figure 5: t-plot of residual percentiles of the Gaussian copula regression.

3.3 Comparison with Chain-Ladder Method

To show that our loss reserve forecasts are not ad hoc, this subsection compares the performance of the copula model with one of the industry's benchmark, the chain-ladder method. The chain-ladder method is implemented via a over-dispersed poisson model. We examine both fitted values and point predictions from the copula model and the chain-ladder fit. The results are presented in Figure 6 and Figure 7.

Figure 6 compares the fitted loss ratio $\hat{y}_{ij}$, for $i + j \leq I$, from the two methods. The fitted values from the copula model are calculated as $\exp(\hat{\mu}_{ij} + 1/2\hat{\sigma}^2)$ for the personal auto line, and $(\hat{\kappa}\hat{\theta}_{ij})^{-1}$ for the commercial auto line. Figure 7 demonstrates the relationship for point predictions of unpaid losses, that is $\hat{y}_{ij}$ when $i + j > I$. We use the predictive mean as the best estimate for unpaid losses. The predictive mean is derived based on the simulation procedure described in Appendix A.1. These panels show that both fitted values and point predictions from the copula model are closely related to those from the chain ladder fit. Thus, a point estimate of aggregated reserves for the insurance portfolio should be close to the chain-ladder forecast.

We focus on point estimates in this subsection, though a reasonable reserve range is more informative to a reserving actuary. As mentioned in Section 1, various methods have been proposed to estimate the chain-ladder prediction error for correlated run-off triangles. By contrast, our parametric setup allows to provide not only the prediction error, but also a predictive distribution of reserves. Another implication of this comparison is that for this particular insurer, the dependence among triangles does not play an important role in determining the point estimate of reserves.

However, the dependencies are critical, as we will show in the following sections, to the predictive distribution, and thus the reserve range.

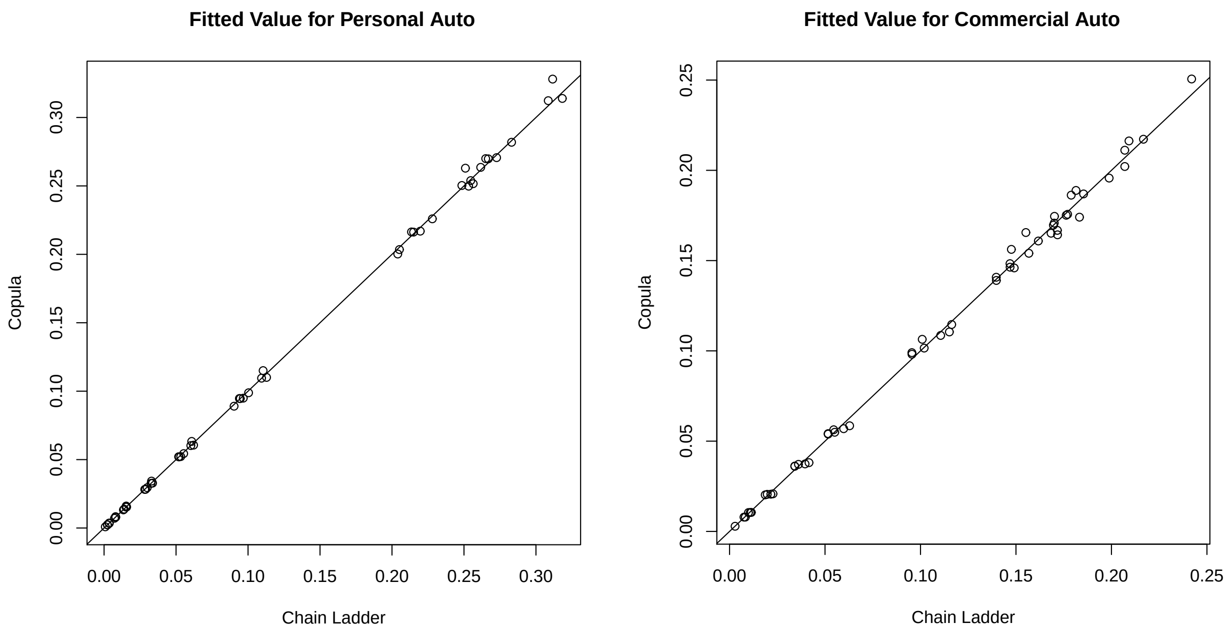

Figure 6: Scatter plots of fitted value between the chain-ladder method and copula model for the personal auto and commercial auto lines.

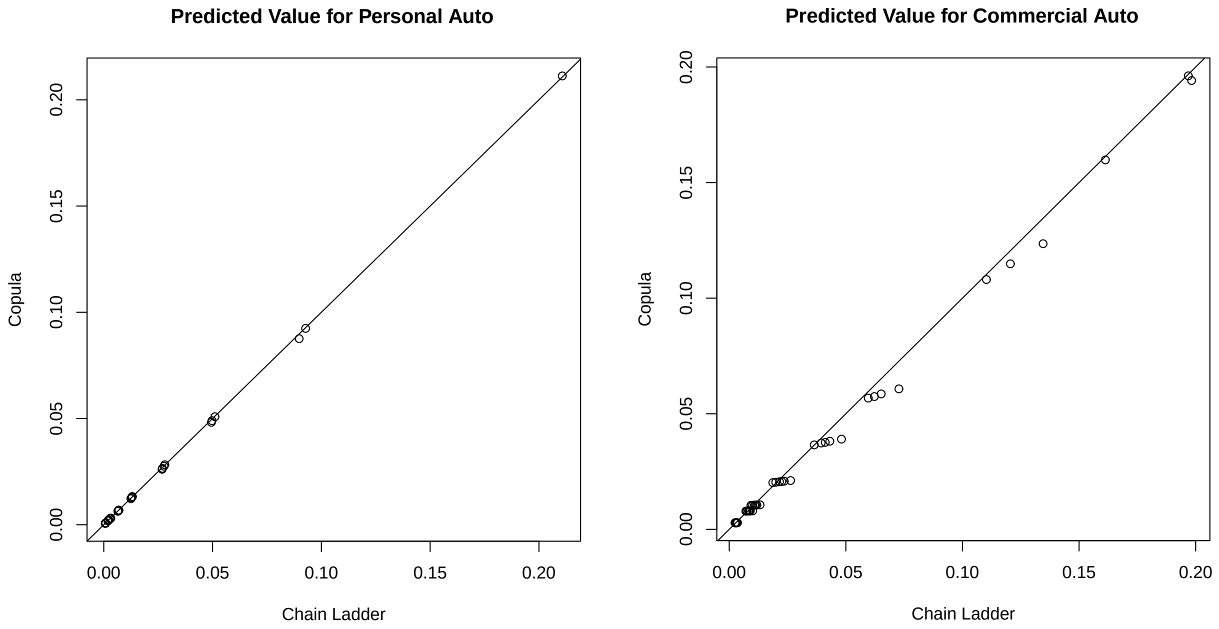

Figure 7: Scatter plots of predicted value between the chain-ladder method and copula model for the personal auto and commercial auto lines.

4 Loss Reserving Indications

In practice, reserving actuaries are more interested in reserve ranges rather than point estimates. This section demonstrates the role of dependencies in the aggregation of claims from multiple run-off triangles. Also, a bootstrap analysis is performed to show the effects of the uncertainty in parameter estimates on the predictive distribution of reserves.

4.1 Prediction of Total Unpaid Losses

Based on the copula regression model, a predictive distribution could be generated for unpaid losses using the Monte Carlo simulation techniques in Appendix A.1. We display the predictive distribution of aggregated reserves for the portfolio in Figure 8. The left panel exhibits the kernel density and the right panel exhibits the empirical CDF. To demonstrate the effect of dependency, the distributions derived from both product and Gaussian copula models are reported. The simulations are based on the parameter estimates in Table 4.

The first panel shows that the Gaussian copula produces a tighter distribution than the product copula. The second panel shows close agreement between the two distributions. The tighter distribution indicates the diversification effect of the correlated subportfolios. Recall that our data show a negative association between the personable auto and commercial auto lines. On the contrary, if two subportfolios are positively associated, one expects to see a predictive distribution that spreads out more than the product copula. In fact, such cases are identified in the preliminary analysis for other insurers, and we include one example in the case studies in Appendix A.3. We need to point out, a fatter distribution does not mean that there is no diversification effect in the insurance portfolio, because the diversification occurs when subportfolios are not perfectly correlated.

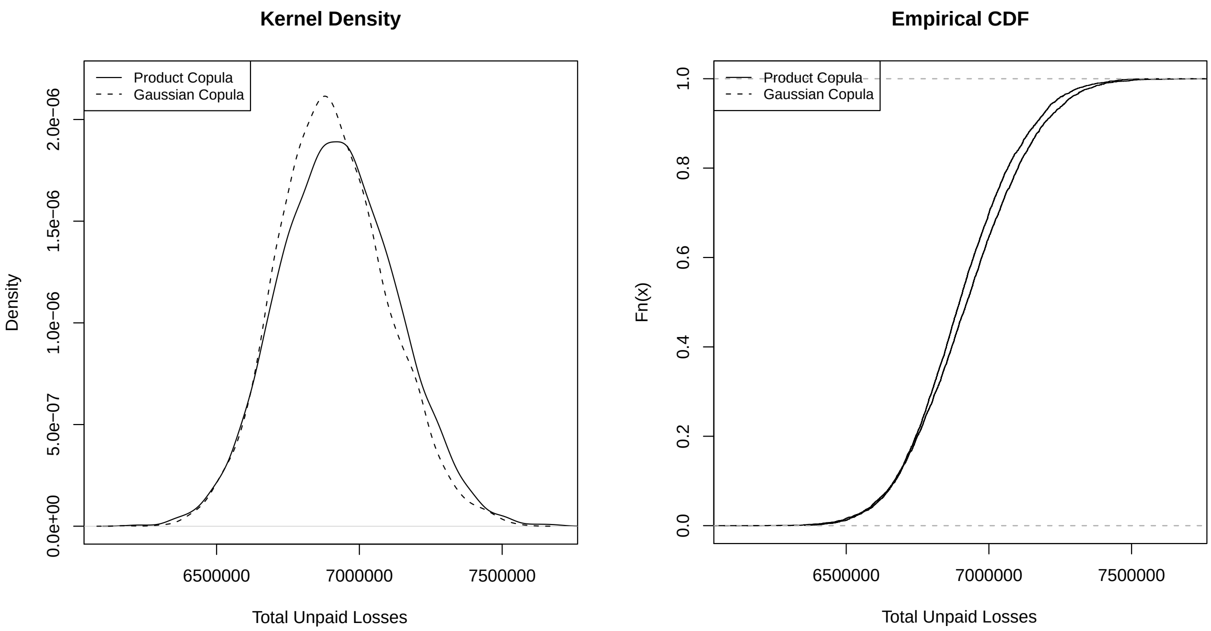

Figure 8: Simulated predictive distributions of total unpaid losses from the product and Gaussian copulas models.

Though we observe the diversification effect of dependencies, Figure 8 does not suggest a substantial difference between the two predictive distributions. This seems counterintuitive when relating to the significant dependence parameter of -0.36 in the Gaussian copula model. To explain this discrepancy, we display the simulated total unpaid losses of the personal auto line versus the commercial auto line in Figure 9. Consistently, the left panel shows that the product copula assumes no relationship (i.e., independence) between the commercial and personal auto lines. The right panel shows that the Gaussian copula permits a negative, and nonlinear, relationship. However, the sizes of the two lines of business are quite different, with the personal auto line dominating the insurance portfolio. Thus, the diversification effect is offset by the unevenly business allocation.

We confirm this with the analysis of other insurers that show negative relationship between the personal and commercial auto lines. The results are reported in Appendix A.3. An important implication of this observation is that the insurer might consider expanding the commercial auto line or shrinking the personal auto line to take best advantage of the diversification effect. Also as will be shown in the next section, such dependence analysis is crucial in determining the risk capital of the insurer.

As mentioned in Section 2.2, to overcome the issue of potential model overfitting, we implement a parametric bootstrap analysis to incorporate the uncertainty of parameter estimates into the predictive distribution. Figure 10 presents the simulated and the bootstrapping distributions for

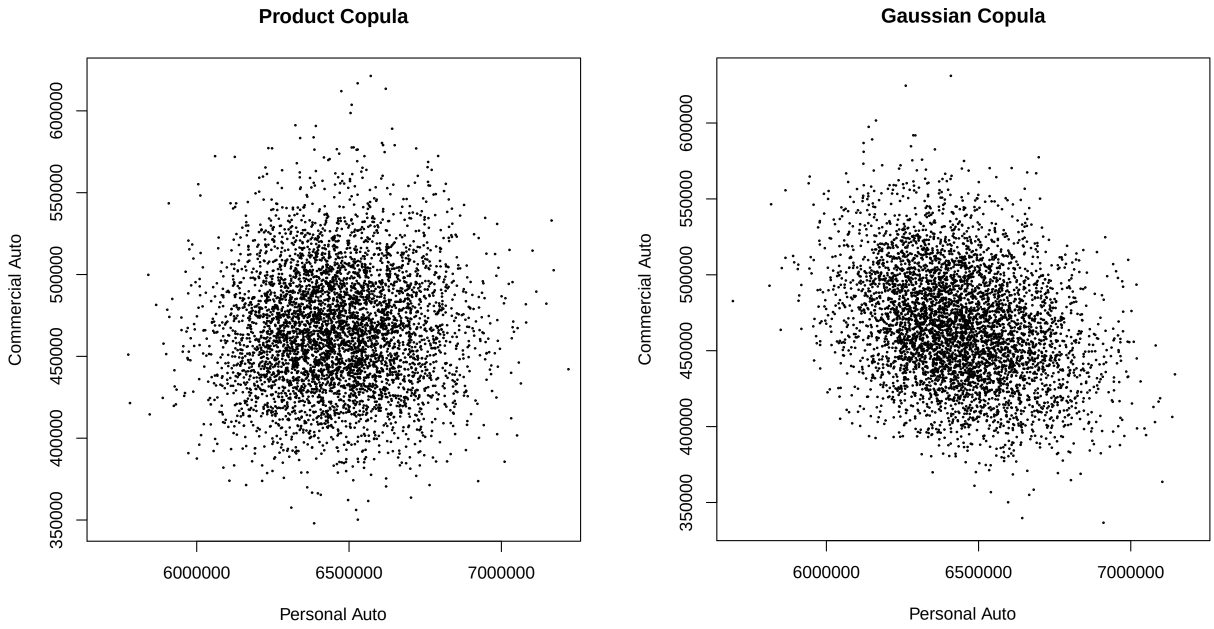

Figure 9: Plots of simulated total unpaid losses for the personal auto and commercial auto lines.

both personal and commercial auto lines. As expected, the bootstrapping distribution is fatter than the simulated distribution, because like Bayesian methods, the bootstrap technique involves various sources of uncertainty.

4.2 Prediction by Year

For accounting and risk management purposes, reserving actuaries might also be interested in the accident year and calendar year reserves. The accident year reserve represents a projection of the unpaid losses for accidents occurred in a particular year, and the calendar year reserve represents a projection of the payments for a certain calendar year. The loss reserving literature focused on the accident year reserve and total reserve (predictions and their mean square errors) for dependent lines of business. However, the extension to the calendar year reserve is not always straightforward. In this section, we demonstrate that the copula regression model is easily adapted for both accident year and calendar year reserves.

From the Gaussian copula model, we simulate the unpaid losses for each accident year i and development lag j, and thus the unpaid losses for a certain accident year or calendar year. Table 5 and Table 6 present the point estimate and a symmetric confidence interval for the accident year and calendar year reserves, respectively. We report the results for both individual lines and aggregated lines from the Gaussian copula model. For comparison purposes, we also report the results for aggregated lines from the product copula. The predicted losses are calculated using the sample mean, and the lower and upper bounds are calculated using the 5th and 95th percentile of the predictive distribution, respectively. Under the Gaussian copula model, the sum of the predicted

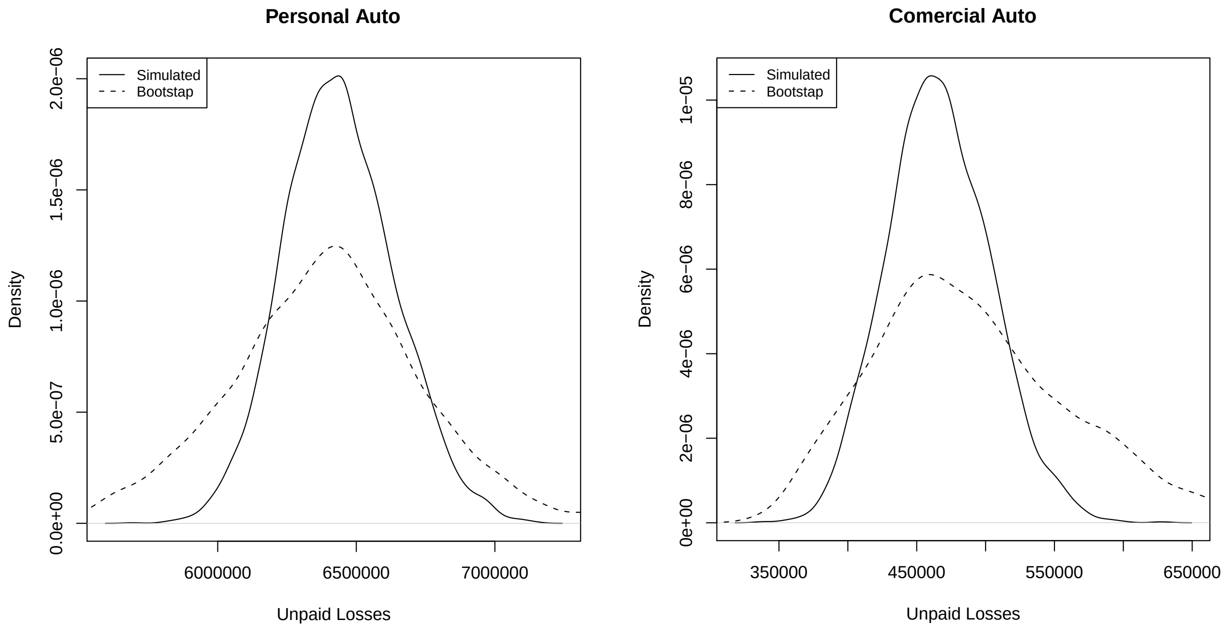

Figure 10: Predictive distributions of total unpaid losses without and with incorporation of uncertainty in parameter estimates.

losses of individual lines is equal to that of the portfolio. However, this additive relationship is not true for other estimates, such as percentiles (see Kirschner et al. (2008)). When compared with the product copula, we observe a narrower confidence interval for aggregated losses due to the negative dependence between the two subportfolios, though the point estimate is close to the independence case. This implies that the association assumed by a copula has a greater impact on the predictive distribution than the predicted mean.

In addition, to account for the uncertainty in parameter estimates, we resort to the bootstrap technique in Appendix A.1. The bootstrapping predictions of accident year and calendar year reserves are displayed in Table 7 and Table 8, respectively. We report the predictive mean and the symmetric confidence interval at 5% significance level for individual lines and combined lines. Not surprisingly, we see that the point prediction is close to and the confidence interval is wider than the corresponding observations in Table 5 and Table 6.

Table 5. Prediction by Accident Year (in thousand dollars)

	Gaussian Copula									Product Copula		
	Personal Auto			Commercial Auto			Combined Lines			Combined Lines		
Accident Year	Predicted Loss	Lower Bound	Upper Bound	Predicted Loss	Lower Bound	Upper Bound	Predicted Loss	Lower Bound	Upper Bound	Predicted Loss	Lower Bound	Upper Bound
1989	4,473	3,842	5,171	783	421	1,223	5,256	4,628	5,940	5,288	4,534	6,127
1990	18,484	16,490	20,670	2,916	1,826	4,222	21,400	19,421	23,490	21,611	19,226	24,156
1991	37,639	34,190	41,346	5,840	4,044	7,948	43,478	40,146	47,037	44,036	39,997	48,174
1992	84,562	77,266	92,298	11,121	7,958	14,798	95,683	88,720	102,894	96,387	88,266	105,125
1993	182,163	166,773	198,545	21,842	15,897	28,600	204,005	189,424	219,314	205,758	188,978	223,723
1994	391,121	358,021	426,498	47,217	34,596	61,204	438,338	407,428	471,871	442,869	406,557	480,514
1995	772,141	711,015	838,099	93,906	69,348	122,360	866,046	806,726	927,262	880,703	810,962	956,953
1996	1,514,921	1,393,988	1,642,700	149,850	111,191	196,601	1,664,772	1,552,369	1,785,327	1,679,953	1,552,086	1,813,197
1997	3,432,930	3,150,531	3,741,914	132,516	100,849	167,200	3,565,446	3,292,717	3,861,489	3,554,116	3,269,488	3,862,129

Table 6. Prediction by Calendar Year (in thousand dollars)

	Gaussian Copula									Product Copula		
	Personal Auto			Commercial Auto			Combined Lines			Combined Lines		
Calendar Year	Predicted Loss	Lower Bound	Upper Bound	Predicted Loss	Lower Bound	Upper Bound	Predicted Loss	Lower Bound	Upper Bound	Predicted Loss	Lower Bound	Upper Bound
1998	3,255,936	2,978,060	3,548,064	190,402	145,847	242,317	3,446,338	3,176,345	3,730,997	3,452,885	3,172,230	3,757,344
1999	1,561,233	1,434,947	1,696,765	121,567	92,319	154,769	1,682,799	1,561,253	1,812,121	1,689,263	1,560,637	1,824,589
2000	825,221	755,510	898,476	113,180	94,615	133,869	938,401	872,274	1,008,447	945,011	870,431	1,020,790
2001	422,804	385,756	461,565	39,294	29,771	50,428	462,098	426,762	499,477	465,695	426,344	507,046
2002	201,444	183,898	220,163	22,638	16,741	29,414	224,082	207,668	241,304	226,047	207,235	245,605
2003	98,470	89,432	108,367	11,916	8,857	15,482	110,387	101,920	119,934	111,476	102,096	121,646
2004	44,723	40,632	49,160	5,963	4,205	8,039	50,686	46,744	54,983	51,322	46,748	56,307
2005	21,694	19,265	24,314	2,678	1,709	3,855	24,372	22,089	26,822	24,591	21,991	27,439
2006	5,923	5,117	6,826	635	338	1,005	6,558	5,786	7,413	6,618	5,735	7,598

Table 7. Bootstrapping Prediction of Accident Year Reserves (in thousand dollars)									
	Personal Auto			Commercial Auto			Portfolio		
Accident Year	Predicted Loss	Lower Bound	Upper Bound	Predicted Loss	Lower Bound	Upper Bound	Predicted Loss	Lower Bound	Upper Bound
1989	4,610	3,733	5,775	783	319	1,412	5,393	4,573	6,501
1990	18,812	16,229	21,727	2,911	1,701	4,404	21,723	18,988	24,666
1991	38,138	33,830	42,602	5,835	3,989	8,158	43,972	39,804	48,258
1992	84,907	75,380	94,959	11,368	8,140	15,332	96,276	87,064	106,282
1993	181,772	163,710	202,274	22,256	16,432	29,098	204,028	187,510	224,238
1994	388,959	346,314	437,985	48,325	36,823	61,966	437,284	396,243	485,498
1995	771,358	684,100	864,330	97,450	70,485	127,058	868,807	791,252	955,942
1996	1,522,266	1,293,289	1,791,688	152,169	108,548	207,014	1,674,436	1,457,959	1,939,494
1997	3,437,295	2,813,539	4,097,722	134,833	81,096	211,087	3,572,128	2,965,710	4,215,735

Table 8. Bootstrapping Prediction of Calendar Year Reserves(in thousand dollars)									
	Personal Auto			Commercial Auto			Portfolio		
Calendar Year	Predicted Loss	Lower Bound	Upper Bound	Predicted Loss	Lower Bound	Upper Bound	Predicted Loss	Lower Bound	Upper Bound
1998	3,334,058	2,947,623	3,805,128	193,229	140,587	262,391	3,527,288	3,160,827	3,972,060
1999	1,590,484	1,418,392	1,813,571	123,195	89,422	168,223	1,713,679	1,555,939	1,922,046
2000	847,739	747,624	979,943	112,444	94,174	132,571	960,182	863,347	1,084,457
2001	432,974	378,047	498,869	39,851	30,219	50,314	472,825	421,499	534,634
2002	206,221	179,547	238,058	22,856	16,885	29,758	229,076	204,577	259,035
2003	101,526	87,396	117,845	11,876	8,420	15,869	113,402	100,251	127,885
2004	46,509	39,710	54,229	5,779	4,021	7,936	52,288	45,794	59,546
2005	22,704	18,482	27,689	2,596	1,627	3,811	25,300	21,332	30,006
2006	6,321	4,803	8,026	603	287	1,028	6,924	5,471	8,593

4.3 Comparison with Existing Methods

This section compares the prediction of unpaid losses of the insurance portfolio from the copula model with various existing approaches. We consider both parametric and non-parametric methods in the literature. Using a non-Bayesian framework, only a few methods provide the entire predictive distribution of aggregated reserves for the portfolio. Among them, Brehm (2002) approximated the silo loss reserve with a log-normal distribution and aggregated different lines of business through a normal copula. The author estimated the dispersion matrix in the copula through the calendar year inflation parameters in the Zehnwirth's model. An alternative approach is presented by Kirschner et al.(2002, 2008), where a synchronous bootstrapping technique was used to generate the unpaid losses from multiple triangles based on an overdispersed poisson model. This resampling technique was examined under the GLM framework by Taylor and McGuire (2007). When modeling the association among multiple run-off triangles, these two methods share a common assumption with our copula regression model, i.e., an identical dependence for all claims in the triangle.

The first comparison is performed with the above two parametric approaches. Figure 11 displays the predictive distributions of the total unpaid losses of the insurance portfolio from various parametric methods. For the synchronous bootstrap, we follow Taylor and McGuire (2007) and assume a gamma distribution for both personal and commercial auto lines. Without taking parameter uncertainty into consideration, the copula model and the log-normal model in Brehm (2002) provide tighter distributions. Note that for this particular insurer, the log-normal distribution is not a good approximation for the total unpaid losses of the portfolio. Applying the parametric bootstrap in Appendix A.1, the predictive distribution from the copula model moves toward the result of Taylor and McGuire (2007). This is explained by the fact that both methods incorporate the parameter uncertainty by resampling run-off triangles.

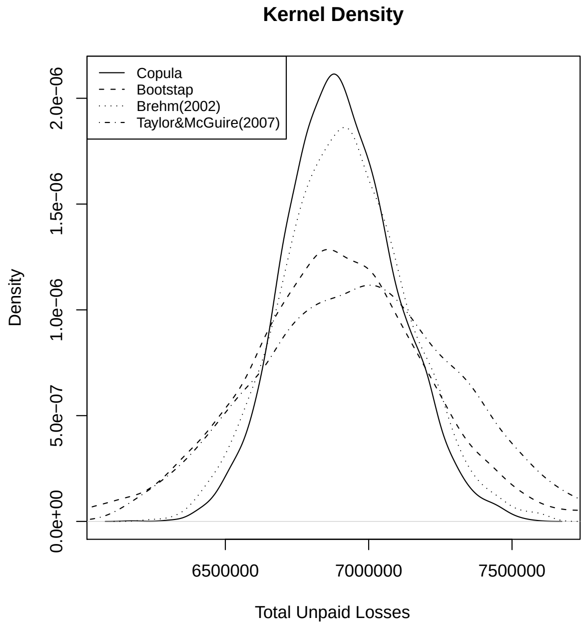

Figure 11: Predictive distributions of total unpaid losses of the insurance portfolio from various parametric methods.

The second comparison is made between parametric and non-parametric methods. Three non-parametric approaches are applied to the insurance portfolio: the multivariate chain-ladder method in Merz and Wüthrich (2008), the multivariate additive loss reserving method in Merz and Wüthrich (2009b), and the combined multivariate chain-ladder and additive loss reserving method in Merz and Wüthrich (2009a). The estimated aggregated reserves and the corresponding prediction error from various approaches are summarized in Table 9. Note that the reported prediction standard error represents the standard deviation of the predictive distribution for the parametric models, and the mean square error of prediction for the non-parametric models. The predictions from the parametric approaches agree with the observations in Figure 11: the copula model and log-normal approximation provides tighter predictions, while the bootstrapped copula and GLM involve more predictive variability. The comparisons with non-parametric methods are rather interesting. In general, we observe the consistency between the two bootstrap models and the four non-parametric approaches. Recall that the main difference between the multivariate chain-ladder and multivariate additive loss reserving method is that the latter allows for the incorporation of external knowledge or prior information in the prediction. For this reason, and also because all subportfolios might not satisfy the same homogeneity assumption, Merz and Wüthrich (2009a) combined the two methods into one integrated framework, where the chain-ladder method could be applied to one subportfolio and the additive loss reserving method to the other. Thus, we have two combinations as shown in Table 9. From this perspective, the copula model offers similar flexibility by allowing different marginal specifications for different lines of business. This might explain the comparable results from the bootstrapped copula model and the combined multivariate chain-ladder and additive loss reserving method.

Table 9. Comparison of Different Approaches

Method	Estimated Reserves	Prediction Std Error
Copula	6,906,329	191,849
Bootstrap	6,921,032	328,991
Brehm(2002)	6,917,133	218,599
Taylor&McGuire(2007)	6,964,043	345,532
Merz&Wuthrich(2008)	6,927,224	339,649
Merz&Wuthrich(2009b)	7,584,956	344,603
Merz&Wuthrich(2009a)-I	7,634,670	368,917
Merz&Wuthrich(2009a)-II	6,877,302	320,497

5 Risk Capital Implication

The predictive distribution of unpaid losses helps actuaries to determine appropriate reserve ranges, it is also helpful to risk managers in determining the risk capital for an insurance portfolio. This section examines the implication of dependencies among loss triangles on the risk capital calculation. Risk capital is the amount of fund that property-casualty insurers set aside as a buffer against potential losses from extreme events. We consider two numerical measures that have been widely used by actuaries, the value-at-risk (VaR) and conditional tail expectation (CTE). The VaR (α) is simply the $100(1-\alpha)$th percentile of the loss distribution. The CTE (α) is the expected losses conditional on exceeding the VaR (α).

We calculate both risk measures for the insurance portfolio that consists of the personal auto and commercial auto lines. The risk capital estimates and corresponding confidence intervals are displayed in Table 10. One way to examine the role of dependencies is to calculate the risk measure for each subportfolio (i.e. the personal auto line and the commercial auto line), and then use the simple sum as the risk measure for the entire portfolio. This is the result reported under the silo method. We also report the risk measures calculated from the product copula and Gaussian copula models. The product copula treats the two lines of business as unrelated, and the Gaussian copula captures the association between the two triangles.

Table 10. Estimates of Risk Capital with 95% Confidence Interval (in thousand dollars)

	VaR (10%)	Confidence Interval		VaR (5%)	Confidence Interval		VaR (1%)	Confidence Interval	
Personal	6,721,641	6,711,531	6,731,751	6,799,452	6,786,617	6,812,287	6,949,072	6,925,425	6,972,719
Commercial	514,803	512,823	516,783	529,891	527,454	532,328	559,175	554,695	563,654
Silo	7,236,444	7,226,138	7,246,750	7,329,342	7,316,261	7,342,423	7,508,246	7,484,113	7,532,379
Product Copula	7,192,280	7,182,021	7,202,539	7,271,122	7,258,064	7,284,180	7,422,907	7,399,051	7,446,763
Gaussian Copula	7,155,400	7,145,523	7,165,277	7,231,093	7,218,782	7,243,404	7,377,212	7,354,518	7,399,906
	CTE (10%)	Confidence Interval		CTE (5%)	Confidence Interval		CTE (1%)	Confidence Interval	
Personal	6,824,360	6,812,432	6,836,288	6,891,787	6,876,286	6,907,288	7,026,968	6,997,131	7,056,805
Commercial	534,803	532,497	537,109	547,958	544,939	550,976	574,563	568,690	580,436
Silo	7,359,164	7,347,031	7,371,297	7,439,744	7,424,002	7,455,486	7,601,531	7,571,126	7,631,936
Product Copula	7,296,396	7,284,293	7,308,499	7,364,716	7,348,940	7,380,492	7,501,576	7,471,473	7,531,679
Gaussian Copula	7,255,551	7,244,076	7,267,026	7,321,340	7,306,343	7,336,337	7,453,552	7,424,767	7,482,337

Table 11. Bootstrapping Estimates of Risk Capital (in thousand dollars)

	VAR (10%)	Standard Error	VAR (5%)	Standard Error	VAR (1%)	Standard Error
Simulation	7,155,400	5,039	7,231,093	6,281	7,377,212	11,578
Bootstrap	7,504,340	6,174	7,623,322	7,893	7,829,356	14,335
	CTE (10%)	Standard Error	CTE (5%)	Standard Error	CTE (1%)	Standard Error
Simulation	7,255,551	5,855	7,321,340	7,652	7,453,552	14,686
Bootstrap	7,654,597	6,673	7,749,974	9,026	7,923,715	17,053

Table 10 shows that the risk measures from both copula models are smaller than the silo method for this particular insurer, although the subadditivity of VaR is not guaranteed in general. The corresponding confidence intervals suggest that the differences among the three methods are statistically significant. This observation is attributed to the diversification effect in the portfolio. The silo method implicitly assumes a perfect positive linear relationship among subportfolios, which does not allow any forms of diversification. The implication of this example is that by taking advantage of the diversification effect, an insurer could reduce the risk capital for risk management or regulatory purposes. A comparison of the copula models shows that both VaR and CTE are smaller for the Gaussian copula model than the product copula model. This is explained by the negative association between the two lines of business. One expects that the risk measures are larger for the Gaussian copula model than the product copula model, if the two triangles present positive association. The above results indicate that the silo method leads to more conservative risk measures, while the copula model leads to more aggressive risk measures. For this particular insurer, the risk measures from various assumptions do not substantively, though statistically, differ. Again, as discussed in Section 4.1, this is due to the disproportional size of the two subportfolios. The result implies that the insurer should increase the volume of the commercial auto line, if he is considering expanding, to take better advantage of the negative dependency. We verify these patterns by increasing the number of simulations and by analyzing the data for other large property-casualty insurers. Case studies can be found in Appendix A.3.

To examine the effect of uncertainty in parameter estimates, we re-calculate both VaR and CTE following the parametric bootstrap procedure in Appendix A.1. The calculations are based on the Gaussian copula model and the results are displayed in Table 11. For comparison purposes, the VaR and CTE from the simulation method are reproduced from the rows of the Gaussian copula in Table 10. Consistently, the bootstrap estimates are larger than the simulation estimates. Recall that the bootstrap produces a fatter predictive distribution, where one typically goes further in the tail to achieve a desired percentile. We find that the effect of incorporating the uncertainty in parameter estimates is even larger than that of diversification, which is explained by the dominating size and the small volatility of the personal auto line in the insurance portfolio.

6 Summary and Concluding Remarks

We considered using copulas to model the association among multiple run-off triangles, and showed that dependencies are critical in the determination of reserve ranges and risk capitals for property-casualty insurers.

Our parametric approach enjoys several advantages in the prediction of unpaid losses for an insurance portfolio. First, the copula model allows for different parametric regression for different lines of business. In fact, the data supported a lognormal regression for the personal auto line and a gamma regression for the commercial auto line, when we applied the method to the data of a major US insurer. Second, in addition to point estimates and their prediction errors, predictive distributions can be derived for unpaid losses, and thus loss reserves. We demonstrated the accident year and calendar year reserves for both individual lines and aggregated lines. Third, due to the parametric nature, a modern bootstrap can be easily performed to incorporate the uncertainty in parameter estimates and to examine the potential modeling overfitting.

We investigated a synthetic insurance portfolio that consists of the personal auto and commercial auto lines. Using the data of a major US insurer, our analysis suggested that the association among lines of business played a more important role in calculating reserve ranges than point estimates. In fact, we showed the point estimates from the copula model were close to the chain-ladder predictions, and the predictive distributions (or mean square error of prediction) were comparable to the results from alternative approaches. Finally, the calculation of risk capitals implied that the diversification effect relied on the magnitude of the dependency as well as the comparative size of each individual line.

References

Ajne, B. (1994). Additivity of chain-ladder projections. *ASTIN Bulletin 24*(2), 311–318.

Braun, C. (2004). The prediction error of the chain ladder method applied to correlated run-off triangles. *ASTIN Bulletin 34*(2), 399–424.

Brehm, P. (2002). Correlation and the aggregation of unpaid loss distributions. In *CAS Forum*, pp. 1–23.

de Alba, E. (2006). Claims reserving when there are negative values in the runoff triangle: Bayesian analysis using the three-parameter log-normal distribution. *North American Actuarial Journal 10*(3), 45–59.

de Alba, E. and L. Nieto-Barajas (2008). Claims reserving: a correlated Bayesian model. *Insurance: Mathematics and Economics 43*(3), 368–376.

de Jong, P. (2010). Modeling dependence between loss triangles using copula. *Working Paper*.

de Jong, P. and G. Heller (2008). *Generalized Linear Models for Insurance Data*. Cambridge University Press.

England, P. and R. Verrall (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal 8*(3), 443–518.

Frees, E., P. Shi, and E. Valdez (2009). Actuarial applications of a hierarchical insurance claims model. *ASTIN Bulletin 39*(1), 165–197.

Frees, E. and E. Valdez (2008). Hierarchical insurance claims modeling. *Journal of the American Statistical Association 103*(484), 1457–1469.

Frees, E. and P. Wang (2005). Credibility using copulas. *North American Actuarial Journal 9*(2), 31–48.

Frees, E. and P. Wang (2006). Copula credibility for aggregate loss models. *Insurance: Mathematics and Economics 38*(2), 360–373.

Genest, C., B. Rémillard, and D. Beaudoin (2009). Goodness-of-fit tests for copulas: a review and a power study. *Insurance: Mathematics and Economics 44*(2), 199–213.

Hess, K., K. Schmidt, and M. Zocher (2006). Multivariate loss prediction in the multivariate additive model. *Insurance: Mathematics and Economics 39*(2), 185–191.

Holmberg, Randall, D. (1994). Correlation and the measurement of loss reserve variability. In *CAS Forum*, pp. 1–247.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall.

Kirschner, G., C. Kerley, and B. Isaacs (2002). Two approaches to calculating correlated reserve indications across multiple lines of business. In *CAS Forum*, pp. 211–246.

Kirschner, G., C. Kerley, and B. Isaacs (2008). Two approaches to calculating correlated reserve indications across multiple lines of business. *Variance 2*(1), 15–38.

Kremer, E. (1982). IBNR claims and the two way model of ANOVA. *Scandinavian Actuarial Journal 1*, 47–55.

Landsman, Z. and E. Valdez (2003). Tail conditional expectations for elliptical distributions. *North American Actuarial Journal 7*(4), 55–71.

Mack, T. (1991). A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. *ASTIN Bulletin 21*(1), 93–109.

Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin 23*(2), 213–225.

Merz, M. and M. Wüthrich (2008). Prediction error of the multivariate chain ladder reserving method. *North American Actuarial Journal 12*(2), 175–197.

Merz, M. and M. Wüthrich (2009a). Combining chain-ladder and additive loss reserving methods for dependent lines of business. *Variance 3*(2), 270–291.

Merz, M. and M. Wüthrich (2009b). Prediction error of the multivariate additive loss reserving method for dependent lines of business. *Variance 3*(1), 131–151.

Merz, M. and M. Wüthrich (2010). Paid-incurred chain claims reserving method. *Insurance: Mathematics and Economics 46*(3), 568–579.

Meyers, G. (2009). Stochastic loss reserving with the collective risk model. *Variance 3*(2), 239–269.

Nelsen, R. (2006). *An Introduction to Copulas*. Springer.

Renshaw, A. and R. Verrall (1998). A stochastic model underlying the chain-ladder technique. *British Actuarial Journal 4*(4), 903–923.

Schmidt, K. (2006). Optimal and additive loss reserving for dependent lines of business. In *CAS Forum*, pp. 319–351.

Shi, P. and E. Frees (2010). Long-tail longitudinal modeling of insurance company expenses. *Insurance: Mathematics and Economics*, forthcoming.

Sun, J., E. Frees, and M. Rosenberg (2008). Heavy-tailed longitudinal data modeling using copulas. *Insurance Mathematics and Economics 42*(2), 817–830.

Taylor, G. (2000). *Loss Reserving: An Actuarial Perspective*. Kluwer Academic Publishers.

Taylor, G. and G. McGuire (2007). A synchronous bootstrap to account for dependencies between lines of business in the estimation of loss reserve prediction error. *North American Actuarial Journal 11*(3), 70.

Wright, T. (1990). A stochastic method for claims reserving in general insurance. *Journal of the Institute of Actuaries 117*, 677–731.

Wüthrich, M. and M. Merz (2008). *Stochastic Claims Reserving Methods in Insurance*. John Wiley & Sons.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association 57*(298), 348–368.

Zhang, Y. (2010). A general multivariate chain ladder model. *Insurance: Mathematics and Economics 46*(3), 588–599.

AUTHOR INFORMATION:

Peng Shi
Division of Statistics
Northern Illinois University
DeKalb, Illinois 60115 USA
e-mail: pshi@niu.edu

Edward W. Frees
School of Business
University of Wisconsin
Madison, Wisconsin 53706 USA
e-mail: jfrees@bus.wisc.edu

A. Appendix

A.1 Simulation and Bootstrap

One advantage of a parametric approach is that a predictive distribution of reserves can be obtained by simulating from the parameters. The simulation can be easily performed for the copula regression model. One could generate the pseudo unpaid losses according to the following procedure:

(1) For accident year i and development lag j that stratify $i+j > I$, generate a N_{ij}-dimensional realization $(u_{ij}^{(1)}, \cdots, u_{ij}^{(N_{ij})})$ from the copula function $c(\cdot; \hat{\boldsymbol{\phi}})$, where $\hat{\boldsymbol{\phi}}$ is the estimate of $\boldsymbol{\phi}$.

(2) Simulate the unpaid losses for accident year i and j by $y_{ij}^{(n)} = F^{(n)(-1)}(u_{ij}^{(n)}; \hat{\eta}_{ij}^{(n)}, \hat{\boldsymbol{\gamma}}^{(n)})$ for $i+j > I$ and $n = 1, \cdots, N_{ij}$. Here $\hat{\eta}_{ij}^{(n)}$ and $\hat{\boldsymbol{\gamma}}^{(n)}$ denote the estimate of $\eta_{ij}^{(n)}$ and $\boldsymbol{\gamma}^{(n)}$, respectively.

(3) For the insurance portfolio, we have:

- The unpaid losses for accident year i ($i = 1, \cdots, I$) is

$$\sum_{j=I+1-i}^{J} \sum_{n=1}^{N_{ij}} \omega_i^{(n)} y_{ij}^{(n)}$$

.

- The unpaid losses for calendar year k ($k = I+1, \cdots, I+J$) is

$$\sum_{i+j=k} \sum_{n=1}^{N_{ij}} \omega_i^{(n)} y_{ij}^{(n)}$$

.

- The total unpaid losses is

$$\sum_{i=1}^{I} \sum_{j=I-i+1}^{I} \sum_{n=1}^{N_{ij}} \omega_i^{(n)} y_{ij}^{(n)}$$

.

Thus, the predictive distribution for the accident year reserve, calendar year reserve, and total reserve can be derived by repeating the above procedures.

One merit of the parametric specification of the copula regression model is that the uncertainty in parameter estimates can be incorporated into the statistical inference by modern bootstrapping. The bootstrap technique serves the same purpose as the Bayesian analysis. Briefly, we use the pseudoresponses to compute the bootstrap distribution of parameters, based on which, we generate the predictive distribution of unpaid losses. The parametric bootstrap procedure is summarized as follows:

(1) Create a set of pseudoresponses of normalized incremental paid losses $y_{ij,r}^{*(n)}$, for i, j such that $i + j \leq I$ and $n = 1, \cdots, N_{ij}$, following the above simulation technique.

(2) Use the pseudoresponses to form the rth bootstrap sample $\{(y_{ij,r}^{*(n)}, \mathbf{x}_{ij}^{(n)}) : i + j \leq I\}$, and from which to derive the bootstrap replication of the parameter vector $(\hat{\eta}_{ij,r}^{*(n)}, \hat{\boldsymbol{\gamma}}_r^{*(n)}, \hat{\boldsymbol{\phi}}_r^{*})$.

(3) Repeat the above two steps for $r = 1, \cdots, R$. Then based on the bootstrapping distribution of parameters, we are able to simulate the predictive distribution of unpaid losses or reserves for each individual line as well as the insurance portfolio.

A.2 Elliptical Copula

Elliptical copulas are extracted from elliptical distributions. Consider a N-dimensional random vector $\mathbf{Z}$ that follows multivariate elliptical distribution with location parameter $\mathbf{0}$ and correlation matrix $\mathbf{\Sigma}$. Let

$$h_{\mathbf{Z}}(\mathbf{z}) = \frac{c_N}{\sqrt{\det \mathbf{\Sigma}}} g_N \left(\frac{1}{2} \boldsymbol{z}' \mathbf{\Sigma}^{-1} \boldsymbol{z} \right),$$

be the density function of $\mathbf{Z}$, and $H_{\mathbf{Z}}(\mathbf{z})$ be the corresponding distribution function. Here, c_N is a normalizing constant and $g_N(\cdot)$ is known as density generator function. See Landsman and Valdez (2003) for discussions of commonly used elliptical distributions in actuarial science.

The N-dimensional elliptical copula, a function of $(u_1, \cdots, u_N) \in [0, 1]^N$, is defined by:

$$C(u_1, \cdots, u_N) = H_{\mathbf{Z}}(H^{-1}(u_1), \cdots, H^{-1}(u_N)),$$

with the corresponding probability density:

$$c(u_1, \cdots, u_N) = h_Z(H^{-1}(u_1), \cdots, H^{-1}(u_N)) \prod_{n=1}^{N} \frac{1}{h(H^{-1}(u_n))}.$$

Here, h and H are the density and distribution function for the marginal, respectively. One nice property of elliptical copulas is that any sub-copula belongs to the same family as the parent copula.

A.3 Case Studies

This section summarizes supplementary results on the dependent loss reserving studies. We provide evidences including: first, the association between lines of business varies across insurers; second, the implication on the loss reserve of dependencies among triangles relies on the construction of the insurance portfolio. In doing so, we apply the copula regression model to the insurance portfolio of the personal auto and commercial auto lines for two other major property-casualty insurers (denoted by Insurer A and Insurer B thereafter). One insurer shows negative association between the two claims triangles, while the other one shows positive association. The portfolios for both insurers are approximately equally distributed into the personal auto and commercial auto lines. The following displays the predictive distributions of unpaid losses and the risk measures of the insurance portfolio for both insurers.

Insurer A exhibits significant negative association between the personal auto and commercial auto lines. For a Gaussian copula regression with gamma models for both lines, the estimated association parameter is -0.44. Figure 12 presents the simulated unpaid losses for two lines under the product and Gaussian copulas. Unlike the insurer in Section 4, the two lines are of similar size.

This explains the stronger diversification effect on the predictive distribution, as shown in Figure 13. The predictive density for the portfolio from the Gaussian copula model is tighter than the independent case.

In contrast, Insurer B presents significant positive association between the two lines of business. We use a gamma model for the personal auto line and a log-normal model for the commercial auto line, the association parameter in the Gaussian copula is 0.34. The simulated unpaid losses and the predictive distribution are provided in Figure 14 and Figure 15, respectively. Note that in the presence of positive association, the predictive density from the Gaussian copula model is fatter than the independent case.

Furthermore, we provide the estimated risk measures of the insurance portfolio for both insurers in Table A.1 and A.2. Three assumptions are examined for each firm. The silo method assumes a perfect positive correlation between the losses from the personal auto and commercial auto lines. The product copula considers the independence case. The Gaussian copula uses the association inferred from the data. The small standard deviation indicates the significant difference between various approaches. The silo method gives the largest estimates of risk measures, because it does not account for any diversification effect in the portfolio. The lower the correlation between sub-portfolios is, the more significant the diversification effect we will observe. As a result, the risk measures from the Gaussian copula model are smaller than those from the product copula model for Insurer A, while the relationship is opposite for Insurer B.

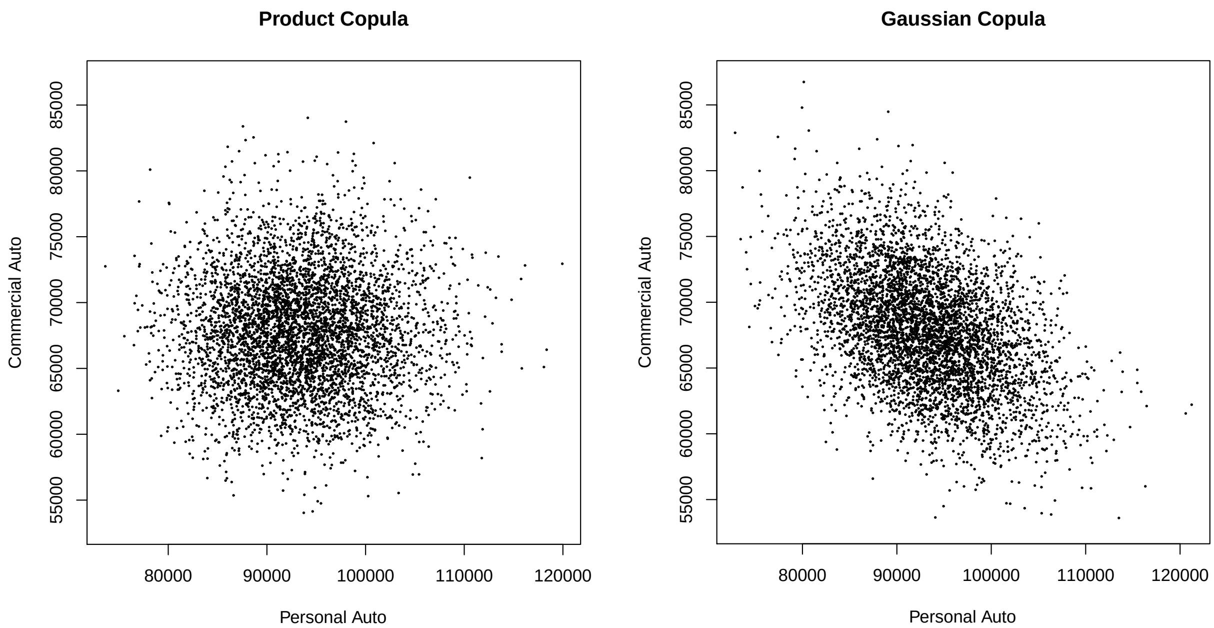

Figure 12: Plots of simulated total unpaid losses of the personal auto and commercial auto lines for Insurer A.

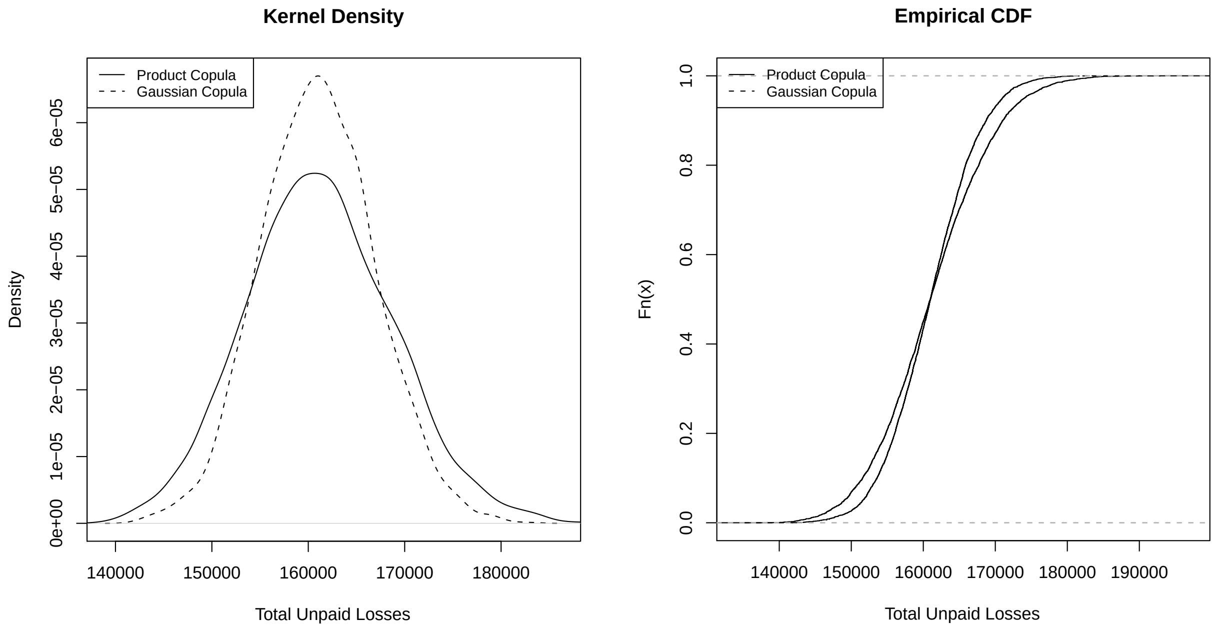

Figure 13: Predictive distributions of total unpaid losses from the product and Gaussian copula models for Insurer A.

Table A.1. Estimates of Risk Capital for Insurer A (in thousand dollars)

	VAR (10%)	Standard Error	VAR (5%)	Standard Error	VAR (1%)	Standard Error
Silo	175,019	200	179,241	252	187,396	471
Product Copula	171,110	195	174,093	243	179,807	449
Gaussian Copula	168,899	155	171,225	192	175,687	355
	CTE (10%)	Standard Error	CTE (5%)	Standard Error	CTE (1%)	Standard Error
Silo	180,604	232	184,267	302	191,628	586
Product Copula	175,045	226	177,618	293	182,752	560
Gaussian Copula	171,971	179	173,980	232	177,993	444

Table A.2. Estimates of Risk Capital for Insurer B (in thousand dollars)

	VAR (10%)	Standard Error	VAR (5%)	Standard Error	VAR (1%)	Standard Error
Silo	402,368	227	407,201	285	416,431	524
Product Copula	397,571	226	400,941	281	407,331	500
Gaussian Copula	399,337	265	403,261	326	410,737	589
	CTE (10%)	Standard Error	CTE (5%)	Standard Error	CTE (1%)	Standard Error
Silo	408,733	260	412,893	334	421,163	643
Product Copula	402,001	256	404,890	328	410,603	615
Gaussian Copula	404,493	298	407,861	383	414,532	734

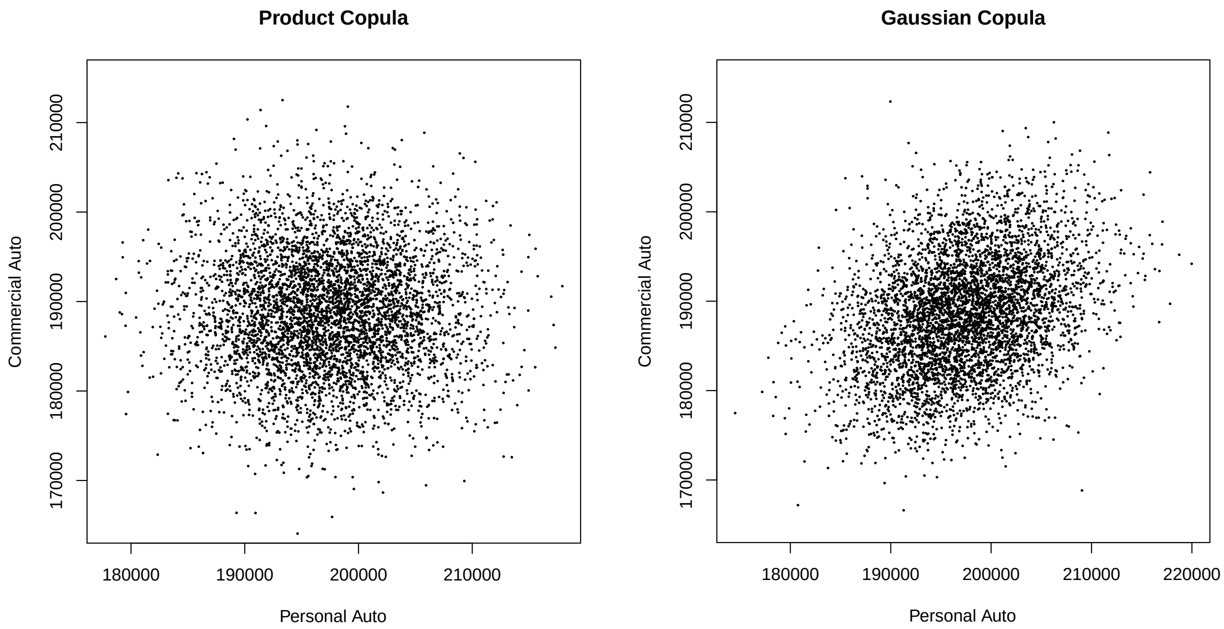

Figure 14: Plots of simulated total unpaid losses for the personal auto and commercial auto lines for Insurer B.

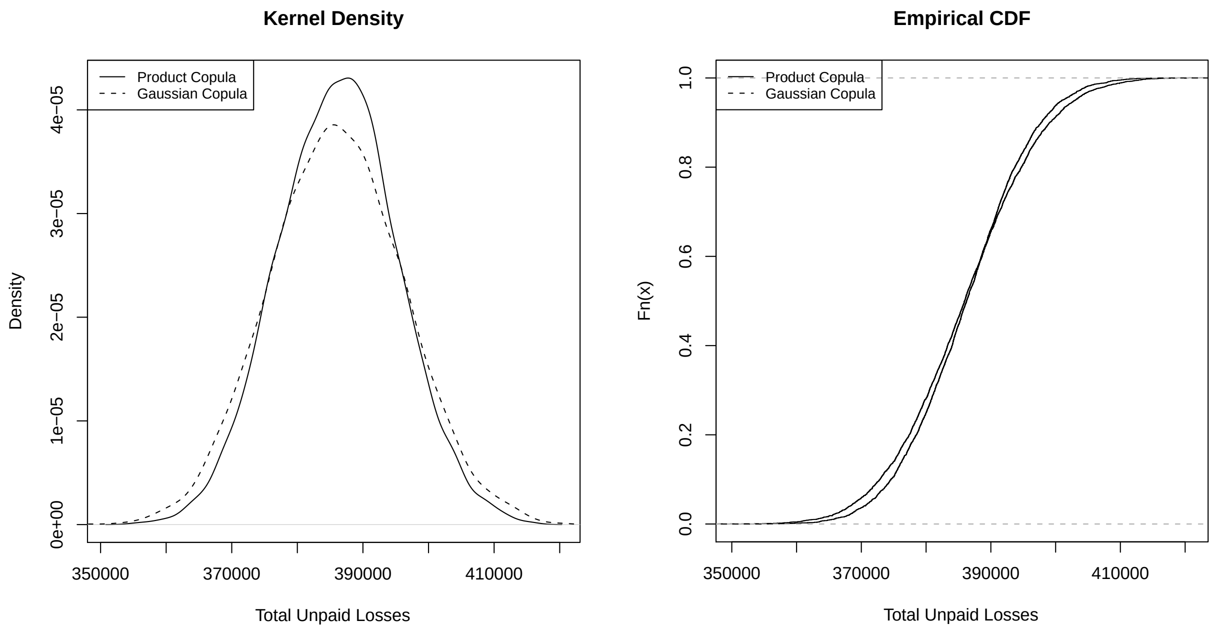

Figure 15: Predictive distributions of total unpaid losses from the product and Gaussian copula models for Insurer B.