Copula-based Modeling for IBNR Claim Loss Reserving

Samira Zaroudi[a], **Mohammad Reza Faridrohani**[b,*], **Mohammad Hassan Behzadi**[c],**and Hadi Safari-Katesari**[d]

[a]School of Mathematical and Statistical Sciences, Southern Illinois University Carbondale, IL 62901, USA, samira.zaroudi@siu.edu

[b,*]Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran, *m_faridrohani@sbu.ac.ir*. Phone: +98-21-2990-3010. Mobile: +98-912-570-3459 (corresponding author)

[c]Department of Statistics, Science and Research Islamic Azad University, Tehran, Iran, behzadi@srbiau.ac.ir

[d]Department of Mathematical Sciences, Stevens Institute of Technology, Hoboken, NJ 07030, USA, hsafarik@stevens.edu

Abstract

There are growing concerns for reserves estimation of incurred but not reported (IBNR) claims in actuarial sciences. In this paper, we propose a copula-based dependency model to capture the relationship between two main IBNR reserve variables, i.e., the “time between two successive occurrences” and “delay time”. A maximum likelihood estimation method is used to estimate the parameters of the model. A simulation study is conducted to evaluate the validity of the theoretical results. Moreover, the proposed method is applied to predict the number of claims for the next years of a portfolio from a major automobile insurer and is compared to the classical CL model forecasting.

Keywords: Copula; Event and report times; IBNR reserves; Run-off triangle; Third-party insurance.

1 Introduction

Reserves estimation for incurred but not reported (IBNR) losses in the insurance policy period is one of the concerns of the actuarial profession. IBNR claims can stay open for a long period of time due to the juristic regulation processes and the size of claims. The difference between the time of occurrence and the time of payment by insurer will change the insurer’s expected obligations and result to wrong amount of claim reserves. This is the place that the role of predicting and computing IBNR claim loss reserving is highlighted. In this paper, we aim to estimate the number of IBNR claims by considering the dependency between the event time and report time of the losses for insurance companies. To do so, we use copula function to build the joint distribution of event and report times of the claims. In order to compute the IBNR claim loss reserves, the classical methods apply data exploration to predict future expected losses, such as Bornhuetter-Ferguson (BF) method proposed by Bornhuetter and Ferguson [1], Benktander-Hovinen method proposed by Benktander [2] and Hovinen [3], Cape Cod method proposed by Bühlmann and Straub [4] and Stanard [5], and Chain-Ladder (CL) method proposed by Mack [6]. For more information about deficiency and properties of these methods, one can refer to [7-9]. Another method to estimate the IBNR claim reserves is copula approach which models the dependency between occurring time and reporting time of an event. Copula is a powerful tool for modeling the dependency between different random variables. In statistical literature, there are many applications of copula such as [10] in aborts data, [11] in travels data, and [12] in biological networks. For insurance data, Pettere and Kollo [13] modeled the size of claims and delay time (between occurrence and report of the claims) by using Archimedean copula family. Zhao et al. [14] and Zhao and Zhou [15] presented a model for individual claims development by using semiparametric techniques of survival analysis and copula methods. Moreover, Shi and Freez [16] used a copula regression model to predict the unpaid losses to obtain the dependency between different lines of a business. Badescu et al. [17] showed that reported and IBNR claim processes are marked Cox processes, while Avanzi et al. [18] used Cox process method to predict the number of IBNR claims by using a dataset of Australian general insurer to model the reporting delay and risk exposure. Landriault et al. [19] computed the moments of total discounted IBNR claims by using a compound renewal process at a given time greater than zero. Also, they considered joint moments of total discounted IBNR claims and incurred and reported claims by using reporting lags and arrival times. Crevecoeur et al. [20] considered the problem of incurred but not yet reported (IBNYR) by using a granular method to model the time between occurrence and observation of claims. For more information about modeling IBNR and IBNYR claims, see [21-26].

Another method to compute the insurance reserves for future obligations is multiplying the average of claims size to the average of claims number in each development time unit. The development time refers to the difference between the time that loss is occurred and the time that the loss is reported to the insurance company. In this paper, we estimate the number of IBNR claims through the following three steps:

Step 1: Copula is applied to model the joint distribution of two marginal variables i.e., the "event time" and "report time" or equivalently "time between two successive occurrences" and "delay time".

Step 2: The individual conditional probability of reporting a claim happened in the development years is estimated based on Step 1 modeling.

Step 3: The average claim size of a IBNR in the development years is estimated.

Similar to Weissner [27], we assume that the marginal distributions i.e., "difference between two occurrences" and "delay time", are exponential distributions with two different rates. We use copula to obtain the dependence between "difference between two successive occurrences" and "delay time". This is while that Zhao and Zhou [15] applied copula approach to obtain the dependency between event time and report time.

The rest of the paper is organized as follows. Section 2 reviews CL method and copula model. Section 3 specifies Clayton copula with event-report time variables as marginal distributions, and demonstrates estimation procedure of the IBNR claim numbers. Section 4 conducts simulation study and real data application by using an automobile insurance dataset. Finally, Section 5 concludes remarks.

2 Model Specification

Copula is a tool to obtain the joint distribution of random variables, when the marginal distributions are available. It is also a strong technique to measure the size of both linear and nonlinear dependency between random variables. Similar to Zhao and Zhou [15], we use copula approach to model the dependence structure of IBNR claim loss reserving but with different marginal distributions. Zhao and Zhou [15] applied copula approach to model the event and delay time for individual claim loss modeling. But, we use copula to obtain the joint distribution and the dependence structure of the duration time between two successive events and the waiting time (reporting delay). In the Archimedean copula family, the Clayton copula [28] is the only absolutely continuous copula, which preserves the bivariate truncation. Oakes [29] applied Clayton model to obtain the joint distribution of the survival times, T_1 and T_2, which is interpreted as the ratio of the hazard rates of the conditional distribution of T_1 given $T_2 = t_2$ to T_1 given $T_2 > t_2$. In order to obtain the joint distribution and the dependence structure of the event and delay times to predict the number of IBNR claims, we propose a new dependence model via copula based on individual number of claims. In our approach, the joint distribution of the marginal distributions i.e., the "difference between two successive occurrences" and "delay time", are modeled by a parametric copula. Moreover, a Poisson process is fitted to the arrival process of claims. Similar to Jewell [30,31], the difference between the two successive occurrences and delays are fitted by using two exponential distributions. This model framework is more flexible than the competitive models for modeling IBNR claims. Moreover, we expect this framework generates more impressive and precise prediction for the number of IBNR claims. The evaluation of the accuracy of our framework is compared to the competitive models in the Section 4. Here, we define the specification of our model framework and a traditional method for modeling the number of IBNR claim called CL method. First, we introduce the CL approach.

2.1 Chain-Ladder Method

Consider a portfolio of an automobile insurance company which is consist of $N > 1$ run-off triangles of observations. Suppose that n $(1 \leq n \leq N)$ indicates the number of portfolios (triangles), i $(0 \leq i \leq I)$ shows the accident years (rows), and j $(0 \leq j \leq J)$ stands for the development years (columns). The number of claims in a portfolio with sample size n for the accident year i and development year j is given by $X^n_{i,j}$ and the cumulative claims of the accident year i up to the development year j are denoted by

$$C^n_{i,j} = \sum_{k=0}^{j} X^n_{i,j}, \tag{1}$$

where $X^n_{i,j} = 0$ for all $j > J$. The individual development factors for the accident year i and development year j are given as

$$f^n_{i,j} = \frac{\sum_{i=1}^{I-j} C^n_{i,j}}{\sum_{i=1}^{I-j} C^n_{i,j-1}}, \quad f^n_{i,j} = (f^1_{i,j}, \cdots, f^N_{i,j})^\top, \tag{2}$$

$$\widetilde{C}^n_{i,j} = C^n_{i,I-i} \prod_{j=I-i}^{J-1} f^n_j, \tag{3}$$

where $n \in \{1, \ldots, N\}, i \in \{1, \ldots, I\}$ and $j \in \{1, \ldots, J\}$, and $\widetilde{C}^n_{i,j}$ is the estimated number of IBNR reserve for the accident year i and the development year j [32]. Recently, the CL method faced high interest in insurance applications such as [33-36].

2.2 Copula Specification

The concept of copula was introduced by Sklar theorem [37]. Nowadays, copula is a main technique to build the dependence structure for insurance and finance datasets. A copula $\mathcal{C}_\theta : [0,1]^n \to [0,1]$ is a multivariate cumulative distribution function on $[0,1] \times [0,1]$ with marginal uniform distributions, where θ is an unknown dependence parameter of the copula. Sklar's Theorem states that any multivariate joint distribution can be written in terms of their univariate marginal distribution functions together with a copula. In the bivariate case, any joint distribution function $F_{T,S}$ corresponding to a bivariate random variable (T, S) with univariate marginal distribution functions F_T and F_S can be obtained by

$$F_{t;s}(x, y) = \mathcal{C}_\theta(F_T(t), F_S(s)),$$

where $\mathcal{C}_\theta(\cdot)$ is the copula function with the dependence parameter θ. One of the well-known class of copulas is Archimedean copulas. The advantage of Archimedean copula family is that the majority of copulas in this family have closed-form distribution functions. This is while that the copulas in the Gaussian copula family does not have closed-form distribution functions. Another characteristic of Archimedean copulas is that they allow to model the dependence structure of random variables in arbitrarily high dimensions with only one parameter. Here, we define Archimedean copulas. Let ϕ be a continuous and strictly decreasing function from $[0, 1]$ to $[0, \infty]$ such that $\varphi(1) = 0$. The pseudo-inverse of ϕ is the function $\phi^{[-1]}$ with domain $[0, \infty]$ and range $I = [0, 1]$ which is given by

$$\phi^{[-1]}(z) = \begin{cases} \phi^{-1}(z), & 0 \le z \le \phi(0) \\ 0, & \phi(0) \le z \le \infty \end{cases}. \tag{4}$$

Notice that $\phi^{[-1]}$ is continuous and non-increasing function on $[0, \infty]$, and strictly decreasing function on $[0, \phi(0)]$. Furthermore, we have $\phi^{[-1]}(\phi(u)) = u$ on I, and

$$\phi(\phi^{[-1]}(z)) = \begin{cases} z, & 0 \le z \le \phi(0) \\ \phi(0), & \phi(0) \le z \le \infty \end{cases} = min(z, \phi(0)). \tag{5}$$

Finally, if $\phi(0) = \infty$ then $\phi^{[-1]} = \phi^{-1}$ [28]. Let $\mathcal{C}$ be a copula function from I^2 to I given by

$$C_\phi(u; v) = \varphi^{(-1)}\left(\varphi(u) + \varphi(v)\right). \tag{6}$$

It is easy to see that the copulas are invariant under monotone transformations of the marginal distribution. Therefore, monotone association measures such as copula-based Kendall's tau with the expression

$$\tau = 4 \int_{[0,1]^2} C(u, v)\ dC(u, v) - 1 \in [-1, 1] \tag{7}$$

are used to obtain the size of dependency between marginal random variables [38]. This is while that the classical correlation measures such as Pearson's correlation coefficient only measures linear associations between marginal distributions. There are many studies to discuss how to select a copula for a given dataset, see [13,39]. The Clayton copula is an asymmetric Archimedean copula, which is able to measure positive dependency between random variables. It is also the most often applied and famous Archimedean copula in experimental applications [40]. The Clayton copula function with association parameter θ is defined as

$$\mathcal{C}_\theta(w; t) = (t^{-\theta} + w^{-\theta} - 1)^{-1/\theta}, \quad \theta \ge 0. \tag{8}$$

Therefore, the joint density function of the Clayton copula is obtained as

$$c(t; w) = (\theta + 1) \times (t\ w)^{-(\theta+1)} \times \left(t^{-\theta} + w^{-\theta} - 1\right)^{-\left(2+\frac{1}{\theta}\right)}, \quad \theta \ge 0. \tag{9}$$

For further illustration and additional properties of the Clayton copulas, see [38,41,42]. Since the association parameter in Clayton copula only accepts positive values, this copula is convenient merely for positively associated random variables. When the dependence parameter θ converges to zero, Clayton copula demonstrates the independent between marginal random variables. The relationship between copula-based Kendall's τ correlation measure and Clayton copula is given as $\tau = \dfrac{\theta}{\theta+2}$, which enables us to measure the size of the copula-based Kendall's τ with known θ. Moreover, the maximal value of τ is captured when θ goes to infinity. For more information, one can refer to [43,44].

3 Estimation of IBNR claim number

3.1 The Number of IBNR Claim with Event-Report Time Modeling

Let T_i and S_i denote the occurring time and the reporting time of an event, respectively. One can model the relationship between T_i and S_i, directly, to predict the number of IBNR claims. Alternatively, we model this relationship indirectly according to the duration time between two successive events and the waiting time (reporting delay). Following Jewell [30,31], we assume that the positive random waiting times, denoted by W_i's, are independent and identically distributed (i.i.d) according to a common exponential distribution. We show the corresponding density of this distribution by $f_{W_i}(.|\beta_2)$, where β_2 is an unknown parameter. In our indirect method, the period of time for occurring the next event plays a pivotal role. We denote the duration time between two successive events by T^* that has exponential distribution with parameter β_1. The joint density function of (T^*, W) based on copulas is given as

$$f_{(T^*,W)}\left(t, w|\beta_1, \beta_2\right) = f_{T^*}\left(t|\beta_1\right) f_W\left(w|\beta_2\right)\ c\left[F_{T^*}\left(t|\beta_1\right), F_W\left(w|\beta_2\right)\right], \tag{10}$$

where $c\left(t, w\right) = \frac{\partial^2 C(\mathrm{t,s})}{\partial t \partial w}$ is the density function of the copula C. Unfortunately, the recording of (T_i^*, W_i)'s are not possible, and so we cannot obtain the likelihood function of (β_1, β_2) based on the joint density function defined in Eq. (10). Instead, observations of the occurring event time T_i and the reporting time S_i are available. Therefore, we obtain the joint density function of $(T_i,\ S_i)$ by using the joint density function of (T_i^*, W_i) represented in Eq. (10).

Notice that the occurrence time of the i^{th} event, T_i, is obtained by summing over all duration times between two successive events up to that time, i.e., $T_i = T_1^* + T_2^* + \cdots + T_i^* = T_{i-1} + T_i^*$. Then, T_i has the Gamma distribution $\Gamma(i, \beta_1)$, because T_i^*'s are iid and follow exponential distribution. On the other hand, it is easy to see that $S_i = T_i + W_i = T_{i-1} + T_i^* + W_i$. Therefore, the joint density function of (T_i, S_i) is obtained as

$$\begin{aligned} f_{(T_i,S_i)}(t,s) =& f_{(T_i,W_i)}(t, s-t) = f_{(T_i^*+T_{i-1},W_i)}(t, s-t) \\ =& \int_0^t f_{(T_i^*,T_{i-1},W_i)}(t-u, u, s-t)du \\ =& \int_0^t f_{(T_i^*,W_i|T_{i-1})}(t-u, s-t|u) f_{T_{i-1}}(u)du \\ =& \int_0^t f_{(T_i^*,W_i)}(t-u, s-t) f_{T_{i-1}}(u)du, \end{aligned} \tag{11}$$

where the $(i-1)^{th}$event time, T_{i-1}, is independent from (T_i^*, W_i) and has Gamma distribution $\Gamma(i-1, \beta_1)$. Moreover, the joint distribution between T_i^* and W_i is obtained by using the Clayton copula defined in Eq. (9) as follows

$$\begin{aligned} f_{(T_i,S_i)}(t,s) = &\int_0^t \frac{e^{-(t-u)\beta_1}(t-u)^{(i-2)}\beta_1^{(i-1)}}{\Gamma(i-1)})\beta_1\beta_2(\theta+1) \\ &\times e^{-(\beta_1 u)}e^{-\beta_2(s-u)}((1-e^{-(\beta_1 u)})(1-e^{-\beta_2(s-u)}))^{-(\theta+1)} \\ &\times ((1-e^{-(\beta_1 u)})^{-\theta} + (1-e^{-\beta_2(s-u)})^{-\theta} - 1)^{-(2+\frac{1}{\theta})}du. \end{aligned} \tag{12}$$

Then, the likelihood function of (β_1, β_2) based on (T_i, S_i) is as follows

$$\begin{aligned} L(\beta_1, \beta_2, \theta; (t_1, s_1), \cdots, (t_n, s_n)) = &\prod_{i=1}^n f_{(T_i,S_i)}(t_i, s_i) \\ = &\prod_{i=1}^n \int_0^{t_i} \frac{e^{-(t_i-u)\beta_1}(t_i-u)^{(i-2)}\beta_1^{(i-1)}}{\Gamma(i-1)})\beta_1\beta_2(\theta+1) \\ &\times e^{-(\beta_1 u)}e^{-\beta_2(s_i-u)}((1-e^{-(\beta_1 u)})(1-e^{-\beta_2(s_i-u)}))^{-(\theta+1)} \\ &\times ((1-e^{-(\beta_1 u)})^{-\theta} + (1-e^{-\beta_2(s_i-u)})^{-\theta} - 1)^{-(2+\frac{1}{\theta})}du. \end{aligned} \tag{13}$$

The maximum likelihood estimation (MLE) of β_1, β_2, and θ can be obtained by maximizing the likelihood function in Eq. (13).

3.2 Delay probability

After estimating the joint density function of $f_{(T_i,S_i)}(t,s)$ defined in Eq. (11), we are able to predict the number of claims reported in the next years. By using the information about i^{th} event occurrences in the j^{th} year, we can estimate the probability of reporting this event in the next $(i+j)^{th}$ years as follows

$$\hat{p}_{i,j}^{(l)} = \hat{P}(S_i \in I_{j+l}|T_i \in I_j) = \frac{\hat{P}(T_i \in I_j, S_i \in I_{j+l})}{\hat{P}(T_i \in I_j)}, \quad l = 1, \ldots, n_J - j, \tag{14}$$

where n_J is the upper bound of delay time.

3.3 IBNR claim number estimation

In order to estimate the number of IBNR claims, we need to obtain $\hat{N}_j^l$, which is the expected number of occurrences related to the reporting the event in the next $(j+l)^{th}$ years for $j = 1, \cdots, n_J$. Therefore, it can predict the number of claims incurred in the year $(j+l)$. Hence, one needs to estimate the expected number of IBNR claims by using following equation

$$\hat{N}_{i,j}^l = \sum_{k=1}^{n_i} \hat{p}_{k,j}^{(l)}, \quad i = 1, \cdots, n_I. \tag{15}$$

4 Data Analysis

In this section, we apply the proposed methods in Section 3 in simulation study and a real dataset. We conduct comparison study to compare the proposed methods with the competitor methods. Moreover, the performance of the maximum likelihood estimator of $(\beta_1, \beta_2, \theta)$ defined in Eq. (13) is considered. By using the estimator introduced in Eq. (15), we predict the claim number in the next years in a third-party insurance policy of an insurance company in Iran. The performance of the proposed model is compared with the CL model forecasting.

4.1 Simulation Study

As mentioned in section 3, T_i^*'s and W_i's are dependent random variables and have exponential distributions with different rate parameters. In order to generate a sequence of dependent observations t_i^* and w_i from random variables T_i^* and W_i, respectively, we apply accept-reject algorithm as follows. Let $Y_i = f_{W_i|T_i^*=t^*}(w|t^*)$ and $V = f_{W_i}(w) \sim exp(\beta_1)$, where f_{Y_i} and f_{W_i} have common support with $M = \sup f_{Y_i}/f_{W_i} < \infty$. consider $Y \sim f_{Y_i}$. Then,

a) generate $U \sim uniform(0,1)$ and $V = f_{W_i}(w)$ independently,

b) if $U < \dfrac{1}{M} f_Y(V)/f_V(V)$, set $Y = V$; otherwise, return to step a).

Here, our goal is to generate the data from W_i which are dependent of T_i^*. That is we have $Y_i = f_{W_i|T_i^*=t^*}(w|t^*)$. The simulated datasets are generated by using the accept-reject algorithm to be used to estimate different parameters of the model, i.e., β_1, β_2, and θ. The MLE of the parameters are conducted for different sample sizes, i.e., 50, 150 and 200, where the number of replication is 100,000. Moreover, the initial values of the scale parameters for the MLE algorithm are considered as the mean of random sample. For determining the initial values for θ, we computed the Kendall's tau ($\hat{\tau}$) for generated samples and obtained the initial value of θ by using $\theta = 2\tau/(1-\tau)$. The mean of the MLEs, mean square errors, and bias of the estimated parameters are reported in Table 1. Note that in this simulation, we selected the real parameters as $\beta_1 = 0.5, \beta_2 = 0.5, \theta = 1.5$. Table 1 demonstrates the average of MLE's, their mean squared error (MSE)'s and biases for parameters β_1, β_2, and θ with real values 0.5 , 0.5, and 1.5, respectively.
In Table 2, the ratios of the simulated number of claims reported in a typical year, i.e., 2016, but occurred over the past 7 years, i.e., during 2010-2016, are reported.

4.2 Real Data Application

In this section, we apply our proposed copula model and CL method to a real dataset from a major automobile insurer in Iran. In particular, we used the observations of a subsample of 140,228 policies recorded in the portfolio of the insurance company during 7 years from 2010 to 2016. We fitted the exponential distribution to marginal distributions, i.e., the "duration time between two successive events" and the "reporting delay time" in our dataset. We carried out Kolmogorov-Smirnov test in which p-values are 0.141 and 0.214, respectively. Therefor, we can assume that the marginal distributions of our copula model are following exponential distributions. As mentioned in Section 3, we estimate all parameters using the MLE method. Notice that we provided Tables 1-8 in the Appendix. First, we apply CL method to this dataset. The upper triangle of Table 3 provided the real number of cumulative claims and the lower triangle of this table demonstrated the estimated number of cumulative claims based on the CL method for the years between 2010 and 2015. In Table 3, first, we obtained the number of claims in each development year for different accident years by using Eq. (3). Then, we obtained the number of cumulative claims. The development year refers to the difference between the year that loss is occurred and the year that the loss is reported to the insurance company. For example, the development year equal zero means that the occurrence time and reporting time of the losses are in the same year and the development year equal 3 means that the losses are reported 3 years after occurrence of the loss. Also, $f^n_{i,j}$ is the individual development factors for the accident year i and development year j defined in Eq. (2). Similarly, we provided the predicted number of cumulative claims based on the CL method for the years 2010-2016 in Table 4. Now, we apply copula method to this dataset. We provided the estimated number of cumulative claims based on the copula method for the years between 2010 and 2015 in Table 6, and for the years 2010 to 2016 in Table 7. We obtained the number of claims in each development year for different accident years by using Eq. (15).

In order to compare the performance of our proposed copula model and CL method in predicting the number of reported claims during different development years, we provided the percentage of the proportional absolute value of errors based on CL method for the years 2010-2015 in Table 5 and based on copula model for the years 2010-2015 in Table 8. The percentage of the proportional absolute value of errors in Tables 5 is computed by subtracting the values of Tables 4 from corresponding values of Table 3, which result is divided to the corresponding values of Table 3. Similarly, The percentage of the proportional absolute value of errors in Tables 8 is computed by subtracting the values of Tables 7 from corresponding values of Table 6, which result is divided to the corresponding values of Table 6. Obviously, there is not any error value for the year 2016 in Tables 5 and 8. For more illustration, we provided an example, which shows how to compute the error values in Tables 5 and 8. The predicted number of claims based on CL method in Table 3 for accident year 2015 and development year 1 is 23719. This is while that the real value of the number of claims in Table 4 is 22769. The percentage of the proportional absolute value of error based on CL method in Table 7 is equal to $|22769 - 23719| \times 100/23719 = 4.0052$. The corresponding percentage of the proportional absolute value of error based on copula method in Table 8 for accident year 2015 and development year 1 is obtained as $|22769 - 22779| \times 100/22779 = 0.0439$. Therefor, the percentage of the proportional absolute value of error based on copula method (0.0439) is smaller than the error term based on CL method 4.0052. Similarly, we can obtain all percentage of the proportional absolute value of error in Table 5 and Table 8. By comparing the results of the percentage of the proportional absolute value of errors based on CL method in Table 5 and copula method in Table 8, we can conclude that our proposed copula method is performing better than CL method.

5 Conclusions

In this paper, we proposed a copula method to predict the IBNR claims. To do so, we applied a well-known family of copulas called Archimedean family. Particularly, we used Clayton copula to find the joint distribution between "difference between two occurrences" and "delay time". In order to assess the performance of the proposed method, we applied a well-known and competitive CL method and compared the results through simulation and real data application. The simulation study indicates that the proposed procedure can produce efficient estimates and improve predictions for the event delay numbers for the next year. Moreover, we used an empirical observation dataset from an insurance portfolio of a major automobile insurer in Iran. The results indicated that the performance of our proposed copula-based method has superior to CL method. As future directions, our method can be extended to the case that the actual event times are forgotten. Moreover, one can extend this method to the non-exponential marginal distributions.

References

[1] Bornhuetter, R.L. and Ferguson, R.E. "November. The actuary and IBNR", *In Proceedings of the casualty actuarial society*, **59**(112), pp. 181-195 (1972).

[2] Benktander, G. "An Approach to Credibility in Calculating IBNR for Casualty Excess Reinsurance", *Actuarial Review*, p7 (1976).

[3] Hovinen, E. *Additive and continuous IBNR*, ASTIN Colloquium, Loen, Norway, (1981).

[4] Bühlmann, H., and Straub, E. "Estimation of IBNR reserves by the methods chain ladder, Cape Cod and complementary loss ratio", *In International Summer School*, (1983).

[5] Stanard, J. N. *A simulation test of prediction errors of loss reserve estimation techniques* (Doctoral dissertation, New York University, Graduate School of Business Administration), (1985).

[6] Mack, T. "Distribution-free calculation of the standard error of chain ladder reserve estimates", *ASTIN Bulletin: The Journal of the IAA*, **23**(2), pp. 213-225 (1993).

[7] Bettonville, C., d'Oultremont, L., Denuit, M., et al. "Matrix calculation for ultimate and 1-year risk in the Semi-Markov individual loss reserving model", *Scandinavian Actuarial Journal*, **5**, pp. 380-407 (2021).

[8] Wang, Z., Wu, X. and Qiu, C. "The impacts of individual information on loss reserving", *ASTIN Bulletin: The Journal of the IAA*, **51**(1), pp. 303-347 (2021).

[9] Yanez, J.S. and Pigeon, M. "Micro-level parametric duration-frequency-severity modeling for outstanding claim payments", *Insurance: Mathematics and Economics*, **98**, pp. 106-119 (2021).

[10] Barbiero, A. "A proposal for modeling and simulating correlated discrete Weibull variables", *Scientia Iranica. Transaction E, Industrial Engineering* **25**(1), pp. 386-397 (2018).

[11] Rasaizadi, A. and Kermanshah, M. "Mode choice and number of non-work stops during the commute: Application of a copula-based joint model", *Scientia Iranica*, **25**(3), pp.1039-1047 (2018).

[12] Purutcuoglu, V. and Farnoudkia, H. "Copula Gaussian graphical modelling of biological networks and Bayesian inference of model parameters", *Scientia Iranica*, **26**(4), pp. 2495-2505 (2019).

[13] Pettere, G. and Kollo, T. "Modelling claim size in time via copulas", *In Transactions of 28th International Congress of Actuaries*, **206**, (2006).

[14] Zhao, X., and Zhou, X. "Applying copula models to individual claim loss reserving methods", *Insurance: Mathematics and Economics*, **46**(2), pp. 290-299 (2010).

[15] Zhao, X.B., Zhou, X. and Wang, J.L. “Semiparametric model for prediction of individual claim loss reserving”, *Insurance: Mathematics and Economics*, **45**(1), pp. 1-8 (2009).

[16] Shi, P. and Frees, E.W. “Dependent loss reserving using copulas”, *ASTIN Bulletin: The Journal of the IAA*, **41**(2), pp. 449-486 (2011).

[17] Badescu, A.L., Lin, X.S. and Tang, D. “A marked Cox model for the number of IBNR claims: Theory”, *Insurance: Mathematics and Economics*, **69**, pp. 29-37 (2016).

[18] Avanzi, B., Wong, B. and Yang, X. “A micro-level claim count model with overdispersion and reporting delays”, *Insurance: Mathematics and Economics*, **71**, pp. 1-14 (2016).

[19] Landriault, D., Willmot, G.E. and Xu, D. “Analysis of IBNR claims in renewal insurance models”, *Scandinavian Actuarial Journal*, **7**, pp. 628-650 (2017).

[20] Crevecoeur, J., Antonio, K. and Verbelen, R. “Modeling the number of hidden events subject to observation delay”, *European Journal of Operational Research*, **277**(3), pp. 930-944 (2019).

[21] Atatalab, F. and Payandeh Najafabadi, A.T. “Prediction of outstanding IBNR liabilities using delay probability”, *Journal of Mathematics and Modeling in Finance*, (2021).

[22] Cheung, E.C., Rabehasaina, L., Woo, J.K., et al. “Asymptotic correlation structure of discounted Incurred But Not Reported claims under fractional Poisson arrival process”, *European Journal of Operational Research*, **276**(2), pp. 582-601 (2019).

[23] Costa, L. and Pizzinga, A. “State-space models for predicting IBNR reserve in row-wise ordered runoff triangles: Calendar year IBNR reserves & tail effects”, *Journal of Forecasting*, **39**(3), pp. 438-448 (2020).

[24] Fung, T.C., Badescu, A.L. and Lin, X.S. “A new class of severity regression models with an application to IBNR prediction”, *North American Actuarial Journal*, pp. 1-26 (2020).

[25] Hendrych, R. and Cipra, T. “Applying State Space Models to Stochastic Claims Reserving”, *ASTIN Bulletin: The Journal of the IAA*, **51**(1), pp. 267-301 (2021).

[26] Manski, S., Yang, K., Lee, G.Y., et al. “Extracting information from textual descriptions for actuarial applications”, *Annals of Actuarial Science*, pp. 1-18 (2021).

[27] Weissner, E.W. “Estimation of the distribution of report lags by the method of maximum likelihood”, *In Proceedings of the Casualty Actuarial Society*, **65**, pp. 1-9 (1978).

[28] Clayton, D.G. “A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence”, *Biometrika*, **65**(1), pp. 141-151 (1978).

[29] Oakes, D. “On the preservation of copula structure under truncation”, *The Canadian Journal of Statistics*, **33**, pp. 465-468 (2005).

[30] Jewell, W. “Predicting IBNYR events and delays, part I continuous time”, *Astin Bulletin*, **19**, , pp. 25-56 (1989).

[31] Jewell, W. "Predicting IBNYR events and delays, part II discrete time", *Astin Bulleti*, **20**, pp. 93-111 (1990).

[32] Wüthrich, M.V. and Merz, M. *Stochastic claims reserving methods in insurance* 435, John Wiley and Sons, (2008).

[33] Badounas, I., Bozikas, A. and Pitselis, G. "A robust random coefficient regression representation of the chain-ladder method", *Annals of Actuarial Science*, pp. 1-32 (2021).

[34] Delong, Ł., Lindholm, M. and Wüthrich, M.V. "Collective reserving using individual claims data" *Scandinavian Actuarial Journal*, pp. 1-28 (2021).

[35] Fischinger, D. and Gach, F. "The 1-year premium risk", *European Actuarial Journal*, pp. 1-21 (2021).

[36] Portugal, L., Pantelous, A.A. and Verrall, R. "Univariate and multivariate claims reserving with Generalized Link Ratios", *Insurance: Mathematics and Economics*, **97**, pp. 57-67 (2021).

[37] Sklar, M. "Fonctions de répartition à n dimensions et leurs marges", *Université Paris* **8**, (1959).

[38] Nelsen, R. B. *An Introduction to Copulas*, New York: Springer Science Business Media, (2006).

[39] Liu, G., Long, W., Yang, B., et al. "Semiparametric estimation and model selection for conditional mixture copula models", *Scandinavian Journal of Statistics*, (2021).

[40] Cherubini, U., Luciano, E., and Vecchiato, W. *Copula Methods in Finance*, John Wiley and Sons, Chichester, (2004).

[41] Katesari, H. S., and Vajargah, B. F. "Testing Adverse Selection Using Frank Copula Approach in Iran Insurance Markets", *Mathematics and Computer Science*, **15**(2), pp. 154-158 (2015).

[42] Katesari, H.S. and Zarodi, S. "Effects of Coverage Choice by Predictive Modeling on Frequency of Accidents", *Caspian Journal of Applied Sciences Research*, **5**(3), pp. 28-33 (2016).

[43] Safari-Katesari, H. and Zaroudi, S. "Count copula regression model using generalized beta distribution of the second kind", *Statistics in Transition New Series*, **21**(2), pp. 1-12 (2020).

[44] Safari-Katesari, H. and Zaroudi, S. "Analysing the impact of dependency on conditional survival functions using copulas", *Statistics in Transition New Series*, **22**(1), pp. 217-226 (2021).

Appendix

Table 1: The average of MLE (Maximum Likelihood Estimation), MSE (Mean Squared Error), and biases for parameters $(\beta_1, \beta_2, \theta)$ with real values $(\beta_1 = 0.5, \beta_2 = 0.5, \theta = 1.5)$ for sample size $n = 50, 150$, and 250.

	MLE			Kendall's tau	MSE			Bias		
n	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\theta}$	τ	β_1	β_2	θ	β_1	β_2	θ
50	0.570	0.587	1.746	0.466	0.051	0.063	0.097	0.07	0.087	0.246
150	0.545	0.561	1.675	0.456	0.009	0.006	0.013	0.045	0.061	0.175
200	0.507	0.523	1.537	0.434	0.004	0.008	0.009	0.007	0.023	0.037

Table 2: Simulation results for the ratio of the number of claims reported in the year 2016 to the number of claims occurred over the years 2010-2016 with different sample sizes ($n = 50, 150$, and 200).

Ratios							
	Years						
n	2016	2015	2014	2013	2012	2011	2010
50	0.6400	0.2200	0.0600	0.0400	0.0200	0.0200	0.00
150	0.6933	0.1400	0.0733	0.0600	0.0200	0.0067	0.0067
200	0.7400	0.1600	0.0700	0.0150	0.000	0.0050	0.0050

Table 3: Estimated numbers of cumulative claims based on CL method for the years 2010-2015

	Development year					
Accident year	0	1	2	3	4	5
2010	5,866	9,237	9,720	9,785	9,805	9,810
2011	19,295	23,307	23,897	24,067	24,113	**24,125**
2012	20,987	25,298	25,978	26,117	**26,168**	**26,181**
2013	18,923	22,757	23,281	**23,427**	**23,473**	**23,485**
2014	18,977	22,539	**23,176**	**23,321**	**23,367**	**23,379**
2015	19,329	**23,719**	**24,389**	**24,542**	**24,590**	**24,603**
$f^n_{i,j}$		1.227132	1.028251	1.006276	1.001950	1.000510

Table 4: Estimated number of cumulative claims based on the CL method for the years 2010-2016.

	Development year						
Accident year	0	1	2	3	4	5	6
2010	5866	9237	9720	9785	9805	9810	9813
2011	19295	23307	23897	24067	24113	24131	**24138**
2012	20987	25298	25978	26117	26174	**26192**	**26206**
2013	18923	22757	23281	23397	**23445**	**23469**	**23484**
2014	18977	22539	22977	**23113**	**23191**	**23213**	**23223**
2015	19329	22769	**23368**	**23517**	**23605**	**23634**	**23655**
2016	10946	**13332**	**13683**	**13763**	**13792**	**13801**	**13805**
$f^n_{i,j}$		1.21794	1.02632	1.00591	1.00205	1.00068	1.00031

Table 5: The percentage of the proportional absolute value errors of number of claims based on CL method in compared with the values presented in Table 4

	Development year					
Accident year	0	1	2	3	4	5
2010	-	-	-	-	-	-
2011	-	-	-	-	-	0.0249
2012	-	-	-	-	0.0229	0.0420
2013	-	-	-	0.8586	0.1193	0.0681
2014	-	-	0.8586	0.8919	0.7532	0.7100
2015	-	4.0052	4.1863	4.1765	4.0057	3.9385

Table 6: Estimated number of cumulative claims based on the copula method for the years 2010-2015

	Development year					
Accident year	0	1	2	3	4	5
2010	5866	9237	9720	9785	9805	9810
2011	19295	23307	23897	24067	24113	**24140**
2012	20987	25298	25978	26117	**26177**	**26189**
2013	18923	22757	23281	**23386**	**23465**	**23484**
2014	18977	22539	**22981**	**23143**	**23210**	**23234**
2015	19329	**22779**	**23389**	**23550**	**23620**	**23651**

Table 7: Estimated number of cumulative claims based on the copula method for the years 2010-2016.

	Development year						
Accident year	0	1	2	3	4	5	6
2010	5866	9237	9720	9785	9805	9810	9813
2011	19295	23307	23897	24067	24113	24131	**24137**
2012	20987	25298	25978	26117	26174	**26198**	**26206**
2013	18923	22757	23281	23397	**23453**	**23477**	**23480**
2014	18977	22539	22977	**23144**	**23197**	**23216**	**23223**
2015	19329	22769	**23379**	**23557**	**23618**	**23648**	**23653**
2016	10946	**15096**	**15677**	**15833**	**15887**	**15914**	**15918**

Table 8: The percentage of the proportional absolute value errors of number of claims based on copula method in compared with the values presented in Table 7.

	Development year					
Accident year	0	1	2	3	4	5
2010	-	-	-	-	-	-
2011	-	-	-	-	-	0.0373
2012	-	-	-	-	0.0115	0.0344
2013	-	-	-	0.0470	0.0511	0.0298
2014	-	-	0.0174	0.0043	0.0560	0.0775
2015	-	0.0439	0.0428	0.0297	0.0085	0.0127

Biographies

Samira Zaroudi is a research assistant at Southern Illinois University, USA. She received her PhD degree in statistics from Science and Research Azad University. Her current research focuses on copula models and

dependencies.

Mohammad Reza Faridrohani is an associate professor in the department of statistics, Shahid Beheshti University, Tehran, Iran. He received his PhD degree in applied statistics from Shahid Beheshti University. His current research focuses on nonparametric inference on multivariate and functional data.

Mohammad Hassan Behzadi is an associate professor in the department of statistics, science and Research Azad University, Tehran, Iran. He received his PhD degree in applied statistics from Science and Research Azad University. His current research focuses on statistical inference and probability applications.

Hadi Safari Katesari is a teaching assistant professor in the department of mathematical sciences, Stevens Institute of Technology, USA. He received his PhD degree in Mathematics, major in Statistics, from Southern Illinois University, USA. His current research focuses on Bayesian statistics, multivariate time series, copula models and dependencies.