Copula-based Modeling for IBNR Claim Loss Reserving

Samira Zaroudi\textsuperscript{a}, Mohammad Reza Faridrohani\textsuperscript{b,∗}, Mohammad Hassan Behzadi\textsuperscript{c} and Hadi Safari-Katesari\textsuperscript{d}

\textsuperscript{a}School of Mathematical and Statistical Sciences, Southern Illinois University Carbondale, IL 62901, USA, samira.zaroudi@siu.edu
\textsuperscript{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)
\textsuperscript{c}Department of Statistics, Science and Research Islamic Azad University, Tehran, Iran, behzadi@srbiau.ac.ir
\textsuperscript{d}Department of Mathematical Sciences, Stevens Institute of Technology, Hoboken, NJ 07030, USA, hsa-farik@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 Freex [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 \(X_{i,j}^n\) and the cumulative claims of the accident year \(i\) up to the development year \(j\) are denoted by

\[
C_{i,j}^n = \sum_{k=0}^{j} X_{i,j}^n, \quad (1)
\]

where \(X_{i,j}^n = 0\) for all \(j > J\). The individual development factors for the accident year \(i\) and development year \(j\) are given as

\[
f_{i,j}^n = \frac{\sum_{l=1}^{I} C_{i,l}^n}{\sum_{l=1}^{I} C_{i,l}^{n-1}}, \quad f_{i,j}^n = (f_{i,j}^1, \ldots, f_{i,j}^N)^\top, \quad (2)
\]

\[
\tilde{C}_{i,j}^n = C_{i,I}^n \prod_{j=I}^{j} f_{j}^n, \quad (3)
\]

where \(n \in \{1, \ldots, N\}, i \in \{1, \ldots, I\}\) and \(j \in \{1, \ldots, J\}\), and \(\tilde{C}_{i,j}^n\) 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 \(C_\theta : [0, 1]^n \to [0, 1]\) is a multivariate cumulative distribution function on \([0, 1] \times [0, 1]\) with marginal uniform distributions, where \(\theta\) 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

$$F_{T,S}(x, y) = C_\theta(F_T(t), F_S(s)),$$

where $C_\theta(\cdot)$ is the copula function with the dependence parameter $\theta$. 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 $\phi$ be a continuous and strictly decreasing function from $[0, 1]$ to $[0, \infty]$ such that $\phi(1) = 0$. The pseudo-inverse of $\phi$ is the function $\phi^{-1}$ with domain $[0, \infty]$ and range $I = [0, 1]$ which is given by

$$\phi^{-1}(z) = \begin{cases} \phi^{-1}(z), & 0 \leq z \leq \phi(0) \\ 0, & \phi(0) \leq z \leq \infty \end{cases}.$$  \hspace{1cm} (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

$$\phi(\phi^{-1}(z)) = \begin{cases} z, & 0 \leq z \leq \phi(0) \\ \phi(0), & \phi(0) \leq z \leq \infty \end{cases} = \min(z, \phi(0)).$$ \hspace{1cm} (5)

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

$$C_\phi(u; v) = \varphi^{(-1)}(\varphi(u) + \varphi(v)).$$ \hspace{1cm} (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

$$\tau = 4 \int_{[0,1]^2} C(u, v) \ dC(u, v) - 1 \in [-1, 1]$$ \hspace{1cm} (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 $\theta$ is defined as

$$C_\theta(w; t) = (t^{-\theta} + w^{-\theta} - 1)^{-1/\theta}, \quad \theta \geq 0.$$ \hspace{1cm} (8)

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

$$c(t; w) = (\theta + 1) \times (t w)^{-(\theta+1)} \times (t^{-\theta} + w^{-\theta} - 1)^{-(2+\frac{1}{\theta})}, \quad \theta \geq 0.$$ \hspace{1cm} (9)
The joint density function of duration time between two successive events by In our indirect method, the period of time for occurring the next event plays a pivotal role. We denote the correlation measure and Clayton copula is given as the independent between marginal random variables. The relationship between copula-based Kendall’s correlation measure and Clayton copula is given as \( \tau = \frac{\theta}{\theta + 2} \), which enables us to measure the size of the copula-based Kendall’s \( \tau \) with known \( \theta \). Moreover, the maximal value of \( \tau \) is captured when \( \theta \) goes to infinity. For more information, one can refer to [38,41,42].

### 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 \( f_{W_i}(\cdot|\beta_2) \), where \( \beta_2 \) 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 \( \beta_1 \).

The joint density function of \((T^*, W)\) based on copulas is given as

\[
f_{(T^*, W)}(t, w|\beta_1, \beta_2) = f_{T^*}(t|\beta_1) f_W(w|\beta_2) \cdot c[F_{T^*}(t|\beta_1), F_W(w|\beta_2)],
\]

where \( c(t, w) = \frac{\partial^2 C(t,s)}{\partial t \partial w} \) 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 \((\beta_1, \beta_2)\) 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

\[
f_{(T_i, S_i)}(t, s) = f_{(T^*_i, W_i)}(t, s-t) = f_{(T^*_i + T_{i-1}, W_i)}(t, s-t) = \int_0^t f_{(T^*_i, T_{i-1}, W_i)}(t-u, u, s-t) du
\]

\[
= \int_0^t f_{(T^*_i, W_i, T_{i-1})}(t-u, s-t, u) f_{T_{i-1}}(u) du
\]

\[
= \int_0^t f_{(T^*_i, W_i)}(t-u, s-t) f_{T_{i-1}}(u) du,
\]

where

\[
f_{(T^*_i, W_i)}(t-u, s-t) = \frac{\beta_1 e^{-\beta_1 (t-u)}}{\Gamma(i, \beta_1)}\int_0^t f_W(w|\beta_2) \cdot c[F_{T^*_i}(t-u|\beta_1), F_W(w|\beta_2)] du.
\]
where the \((i - 1)\text{th}\) 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

\[
f_{(T_i, S_i)}(t, s) = \int_0^t \frac{e^{-(t-u)^{\beta_1}(t-u)(i-2)^{\beta_2(i-1)}}}{\Gamma(i-1)} \beta_1 \beta_2 (\theta + 1) \times e^{-(\beta_1 u)} e^{-(\beta_2(s-u))((1-e^{-\beta_1 u})(1-e^{-\beta_2(s-u)})^{-(\theta+1)})} \times ((1-e^{-\beta_1 u})^{-\theta} + (1-e^{-\beta_2(s-u)})^{-\theta} - 1)^{-(2+\frac{1}{\theta})} du.
\]

Then, the likelihood function of \((\beta_1, \beta_2)\) based on \((T_i, S_i)\) is as follows

\[
L(\beta_1, \beta_2, \theta; (t_1, s_1), \ldots, (t_n, s_n)) = \prod_{i=1}^n f_{(T_i, S_i)}(t_i, s_i) = \prod_{i=1}^n \int_0^{t_i} \frac{e^{-(t_i-u)^{\beta_1}(t_i-u)(i-2)^{\beta_2(i-1)}}}{\Gamma(i-1)} \beta_1 \beta_2 (\theta + 1) \times e^{-(\beta_1 u)} e^{-(\beta_2(s_i-u))((1-e^{-\beta_1 u})(1-e^{-\beta_2(s_i-u)})^{-(\theta+1)})} \times ((1-e^{-\beta_1 u})^{-\theta} + (1-e^{-\beta_2(s_i-u)})^{-\theta} - 1)^{-(2+\frac{1}{\theta})} du.
\]

The maximum likelihood estimation (MLE) of \(\beta_1, \beta_2, \) and \(\theta\) 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 \(i\text{th}\) event occurrences in the \(j\text{th}\) year, we can estimate the probability of reporting this event in the next \((i+j)\text{th}\) years as follows

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

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 \(\tilde{N}_i^l\), which is the expected number of occurrences related to the reporting the event in the next \((j+l)\text{th}\) 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

\[
\tilde{N}_{i,j} = \sum_{k=1}^{n_i} \hat{P}_{k,j}^{(l)}, \quad i = 1, \cdots, n_I.
\]

### 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 \(Y_i = f_{W_i|T^*_i=t^*_i}(w|t^*_i)\)
and \(V = f_{W_i}(w) \sim \exp(\beta_1)\), where \(f_{Y_i}\) and \(f_{W_i}\) have common support with \(M = \sup \frac{f_{Y_i}}{f_{W_i}} < \infty\). consider
\(Y \sim f_{Y_i}\). Then,

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

b) if \(U < \frac{1}{M} \frac{f_{Y_i}(V)}{f_{V}(V)}\), 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 \(Y_i = f_{W_i|T^*_i=t^*_i}(w|t^*_i)\). The simulated datasets are generated by using the accept-reject algorithm to be used
to estimate different parameters of the model, i.e., \(\beta_1, \beta_2,\) and \(\theta\). 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 sam-
ple. For determining the initial values for \(\theta\), we computed the Kendall’s tau (\(\hat{\tau}\)) for generated samples and
obtained the initial value of \(\theta\) by using \(\theta = 2\hat{\tau}/(1 - \tau)\). The mean of the MLE’s, 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 \(\beta_1 = 0.5, \beta_2 = 0.5,\) and \(\theta = 1.5\). Table 1 demonstrates the average of MLE’s, their mean squared
error (MSE)’s and biases for parameters \(\beta_1, \beta_2,\) and \(\theta\) 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. Therefore, 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, \( f_{n_{ij}} \) 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 \( |22769 - 23719| \times \frac{100}{23719} = 4.0052 \). 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 \( |22769 - 22779| \times \frac{100}{22779} = 0.0439 \). 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


**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.

<table>
<thead>
<tr>
<th>n</th>
<th>(\hat{\beta}_1)</th>
<th>(\hat{\beta}_2)</th>
<th>(\hat{\theta})</th>
<th>(\tau)</th>
<th>(\hat{\beta}_1)</th>
<th>(\hat{\beta}_2)</th>
<th>(\theta)</th>
<th>Bias</th>
<th>MLE</th>
<th>Kendall’s tau</th>
<th>MSE</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.570</td>
<td>0.587</td>
<td>1.746</td>
<td>0.466</td>
<td>0.051</td>
<td>0.063</td>
<td>0.097</td>
<td>0.07</td>
<td>0.087</td>
<td>0.246</td>
<td>0.063</td>
<td>0.097</td>
</tr>
<tr>
<td>150</td>
<td>0.545</td>
<td>0.561</td>
<td>1.675</td>
<td>0.456</td>
<td>0.009</td>
<td>0.006</td>
<td>0.013</td>
<td>0.045</td>
<td>0.061</td>
<td>0.175</td>
<td>0.013</td>
<td>0.017</td>
</tr>
<tr>
<td>200</td>
<td>0.507</td>
<td>0.523</td>
<td>1.537</td>
<td>0.434</td>
<td>0.004</td>
<td>0.008</td>
<td>0.009</td>
<td>0.007</td>
<td>0.023</td>
<td>0.037</td>
<td>0.009</td>
<td>0.012</td>
</tr>
</tbody>
</table>

**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.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.6400</td>
<td>0.2200</td>
<td>0.0600</td>
<td>0.0400</td>
<td>0.0200</td>
<td>0.0200</td>
<td>0.00</td>
</tr>
<tr>
<td>150</td>
<td>0.6933</td>
<td>0.1400</td>
<td>0.0733</td>
<td>0.0600</td>
<td>0.0200</td>
<td>0.0067</td>
<td>0.0067</td>
</tr>
<tr>
<td>200</td>
<td>0.7400</td>
<td>0.1600</td>
<td>0.0700</td>
<td>0.0150</td>
<td>0.0000</td>
<td>0.0050</td>
<td>0.0050</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Development year</th>
<th>Accident year</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>5,866</td>
<td>9,237</td>
<td>9,720</td>
<td>9,785</td>
<td>9,805</td>
<td>9,810</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>19,295</td>
<td>23,307</td>
<td>23,897</td>
<td>24,067</td>
<td>24,113</td>
<td>24,125</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>20,987</td>
<td>25,298</td>
<td>25,978</td>
<td>26,117</td>
<td>26,168</td>
<td>26,181</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>18,923</td>
<td>22,757</td>
<td>23,281</td>
<td>23,427</td>
<td>23,473</td>
<td>23,485</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>18,977</td>
<td>22,539</td>
<td>23,176</td>
<td>23,321</td>
<td>23,367</td>
<td>23,379</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>19,329</td>
<td>23,719</td>
<td>24,389</td>
<td>24,542</td>
<td>24,590</td>
<td>24,603</td>
<td></td>
</tr>
<tr>
<td>(f_{n,j})</td>
<td>1.227132</td>
<td>1.028251</td>
<td>1.006276</td>
<td>1.001950</td>
<td>1.000510</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 4: Estimated number of cumulative claims based on the CL method for the years 2010-2016.

<table>
<thead>
<tr>
<th>Development year</th>
<th>Accident year</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>5866</td>
<td>9237</td>
<td>9720</td>
<td>9785</td>
<td>9805</td>
<td>9810</td>
<td>9813</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>19295</td>
<td>23307</td>
<td>23897</td>
<td>24067</td>
<td>24113</td>
<td>24131</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>20987</td>
<td>25298</td>
<td>25978</td>
<td>26117</td>
<td>26174</td>
<td>26192</td>
<td>26206</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>18923</td>
<td>22757</td>
<td>23281</td>
<td>23397</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>18977</td>
<td>22539</td>
<td>22977</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>19329</td>
<td>22769</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>10946</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ f_{i,j}^{n} = \begin{array}{cccccc} 1.21794 & 1.02632 & 1.00591 & 1.00205 & 1.00068 & 1.00031 \end{array} \]

### 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.

<table>
<thead>
<tr>
<th>Development year</th>
<th>Accident year</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2011</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.0249</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.0229</td>
<td>0.0420</td>
</tr>
<tr>
<td>2013</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.8586</td>
<td>0.1193</td>
<td>0.0681</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>-</td>
<td>0.8586</td>
<td>0.8919</td>
<td>0.7532</td>
<td>0.7100</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

13
### Table 6: Estimated number of cumulative claims based on the copula method for the years 2010-2015

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

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

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

### 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.

<table>
<thead>
<tr>
<th>Development year</th>
<th>Accident year 0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2011</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.0373</td>
</tr>
<tr>
<td>2012</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.0115</td>
<td>0.0344</td>
</tr>
<tr>
<td>2013</td>
<td>-</td>
<td>-</td>
<td>0.0470</td>
<td>0.0511</td>
<td>0.0298</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>-</td>
<td>-</td>
<td>0.0174</td>
<td>0.0043</td>
<td>0.0560</td>
<td>0.0775</td>
</tr>
<tr>
<td>2015</td>
<td>-</td>
<td>0.0439</td>
<td>0.0428</td>
<td>0.0297</td>
<td>0.0085</td>
<td>0.0127</td>
</tr>
</tbody>
</table>

### 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.