International Journal of Statistics and Probability; Vol. 11, No. 3; May 2022
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Bayesian Bivariate Cure Rate Models Using Copula Functions

Jie Huang[1] & Haiming Zhou[1] Nader Ebrahimi[1]

[1] Division of Statistics, Northern Illinois University, DeKalb, IL, USA

Correspondence: Jie Huang, Division of Statistics, Northern Illinois University, DeKalb, IL, USA

Received: January 10, 2022 Accepted: March 2, 2022 Online Published: March 9, 2022

doi:10.5539/ijsp.v11n3p9 URL: https://doi.org/10.5539/ijsp.v11n3p9

Abstract

Bivariate survival cure rate models extend the understanding of time-to-event data by allowing for a cured fraction of the population and dependence between paired units and make more accurate and informative conclusions. In this paper, we propose a Bayesian bivariate cure rate mode where a correlation coefficient is used for the association between bivariate cure rate fractions and a new generalized Farlie Gumbel Morgenstern (FGM) copula function is applied to model the dependence structure of bivariate survival times. For each marginal survival time, we apply a Weibull distribution, a log normal distribution, and a flexible three-parameter generalized extreme value (GEV) distribution to compare their performance. For the survival model fitting, DIC and LPML are used for model comparison. We perform a goodness-of-fit test for the new copula. Finally, we illustrate the performance of the proposed methods in simulated data and real data via Bayesian paradigm.

Keywords: bivariate cure rate, Copulas, Goodness-of-fit, Bayesian approach, survival analysis

1. Introduction

In survival analysis, it is of primary interest to measure the association between two time-to-event random variables associated with one individual. In many cases, the lifetime of paired unites from same individual would affect each other. For instance, visual loss for one eye could affect another eye for the same patient. Another case in cancer study, for example, a fraction of patents may response positively to the treatment. On the other word, this fraction of patients will not experience death within the follow-up period, and they have long term survival times. In the literature, frailty models with a single shared frailty were popular. They account for unobserved heterogeneity by including random effect. The main feature of the shared frailty models is all units share the same frailty. Because of this limitation, they have been extended to model data with more complex dependence relations. Yashin and Iachine (1999) involved correlated stochastic hazard in a given frailty of survival distribution. Peng and Taylor (2011) applied different random effects to model the correlations for cure patients and uncured patients, respectively. Gallardo, Gómez, and de Castro (2018) proposed a cure rate model and applied the competing risks approach to the latent causes of the event of interest.

An alternative is the use of copula functions. Unlike the frailty approaches, the copula approach models the joint distribution by connecting the two marginal distributions through a copula function. Modelling dependence is one of the primary interests in multivariate analysis. The advantages of the copula are as follows. First, the copula models the marginal distributions and the dependence parameters separately which allows flexibility in marginal models and straightforward construction of covariate effects. Secondly, the copula can handle the censoring through the marginal distributions. Thirdly, the conditional distributions can be obtained through the copula model. Romeo, Tanaka, and Pedroso-de Lima (2006) introduced the Archimedian copula family for modeling the dependence of bivariate lifetime components where a Weibull distribution is considered as the marginal distribution. Louzada, Suzuki, and Cancho (2013) proposed an FGM long-term bivariate survival copula model. They assumed a mixture cure rate model for the marginal distribution of each lifetime and assumed fixed cure fraction for the entire population. C.-M. Chen, Lu, and Hsu (2013) applied the pairwise odds ratio to the association of the insusceptibility of the individuals and adopted Clayton copula to measure the association of susceptible individuals with a semiparametric distribution as marginal regression model. Lakhal-Chaieb and Duchesne (2017) introduced a link function to relax subject-specific-effect assumption which improves the range of potential association and add flexibility to dependence structure.

The literature has introduced many other modelling approaches for bivariate long term data using copula functions, as for example the paper introduced by Louzada et al. (2013). But in that paper, the authors only present more simple cure fraction survival model situation assuming dependence with a FGM copula function structure. In this paper, a mixture model is applied to analyze a bivariate censored data with different susceptibilities. A correlation coefficient is applied for cure rates and a generalized bivariate Farlie Gumbel Morgenstern (FGM) copula, proposed by Bekrizadeh and Jamshidi (2017), is applied for the association of bivariate failure times. As studied in Louzada et al. (2013), it showed a weak

dependence between tow lifetimes in the diabetic retinopathy study data. The FGM commonly used to model very weak linear dependences and give more reliable estimates. One of the advantages of the generalized FGM copula has wider range of correlation compared to regular FGM copula. According to Bekrizadeh and Jamshidi (2017), the estimated correlations based on the generalized FGM copula is closer to the actual correlation of the observed data compared to FGM copula. This conclusion is also shown in diabetic retinopathy study data. Also, the Spearmans ρ of two lifetimes in the diabetic retinopathy study data is 0.376 which is greater than the upper bound of ρ for the FGM copula. Whereas the generalized FGM copula has one more parameter with upper bound of ρ up to 0.385. In Bayesian analysis of proposed model, we perform standard MCMC method in consideration of cure fractions and censoring for both lifetimes. We employ the flexible three-parameter generalized extreme value (GEV) distribution on the marginal survival time. We also apply Weibull distribution and log normal distribution for model comparison.

Misspecifying the copula model may have impact on the inference procedure. Therefore, it may be necessary for our proposed methodology to use a goodness-of-fit test for adequacy check. In the literature, specification tests have been extensively investigated such as rank based test as in Wang and Wells (2000), kernels as in Fermanian (2005), and blanket tests. Genest, Rémillard, and Beaudoin (2009) show that all of blank tests have no power in differentiating some copulas such as Gaussian copula and Student's t copula. Also, it's difficult for deriving analytically in the test statistics of Student's t copula and vine copula since blank tests require certain probability integral transformation. S. Zhang, Okhrin, Zhou, and Song (2016) propose an alternative specification test for semi-parametric copulas which does not require any probability integral transformation. The proposed test is referred to as pseudo-in-and-out-sample test (PIOS) which takes a form of ratio constructed via in sample pseudo-likelihood and out of sample pseudo-likelihood.

Compared to the POIS test proposed by S. Zhang et al. (2016), our work makes the following new contributions. First of all, the test is extended to be applicable to a parametric copula model of right-censored survival times. Secondly, the test is extended to the case of mixture cure rate model for individual survival function. Finally, we discuss how to identify the susceptible subjects in the mixture cure rate model in order to produce a PIOS test statistic.

2. The Model

Suppose two lifetimes T_1 and T_2 associated to the same subject. Let d_k be an indicator variable showing a subject is susceptible to the kth event and a corresponding cure probability is $P_k = Pr(d_k = 0)$, $k = 1, 2$. We assume mixture models for T_1 and T_2, given, respectively, by

$$S_1(t_1) = Pr(T_1 > t_1) = P_1 + (1 - P_1)S_{10}(t_1)$$
$$S_2(t_2) = Pr(T_2 > t_2) = P_2 + (1 - P_2)S_{01}(t_2),$$

where $S_{10}(t_1) = Pr(T_1 > t_1|d_1 = 1)$ and $S_{01}(t_2) = Pr(T_2 > t_2|d_2 = 1)$ are survival functions associated with T_1 and T_2 when the subject is susceptible for the underlying event.

The joint survival function for T_1 and T_2 is given by

$$S(t_1, t_2) = \sum_{d_1,d_2} S(t_1, t_2|d_1, d_2)Pr(d_1, d_2). \tag{1}$$

where $S(t_1, t_2|d_1 = 1, d_2 = 1)$, for example, is the joint survival function of T_1 and T_2 for the susceptible individuals.

Assuming that covariance between d_1 and d_2 is ρ, we have

$$Pr(d_1 = 1, d_2 = 1) = (1 - P_1)(1 - P_2) + \rho \triangleq \varphi_{11}, \tag{2}$$
$$Pr(d_1 = 1, d_2 = 0) = (1 - P_1)P_2 - \rho \triangleq \varphi_{10}, \tag{3}$$
$$Pr(d_1 = 0, d_2 = 1) = (1 - P_2)P_1 - \rho \triangleq \varphi_{01}, \tag{4}$$
$$Pr(d_1 = 0, d_2 = 0) = P_1P_2 + \rho \triangleq \varphi_{00}. \tag{5}$$

Now, using (1) - (5) we get

$$S(t_1, t_2) = \varphi_{11}S(t_1, t_2|d_1 = 1, d_2 = 1) + \varphi_{10}S_{10}(t_1) + \varphi_{01}S_{01}(t_2) + \varphi_{00},$$

where $S(t_1, t_2|d_1 = 1, d_2 = 1)$, S_{10} and S_{01} are defined as above. Now one possibility is to use different parametric distributions for $S(t_1, t_2|d_1 = 1, d_2 = 1)$. Another possibility is to use copula functions which link marginal distributions with a joint distribution. Throughout this paper, we use copula functions for joint survival of susceptible individuals. Thus, the joint survival function for T_1 and T_2 can be written as,

$$S(t_1, t_2) = \varphi_{00} + \varphi_{10}S_{10}(t_1) + \varphi_{01}S_{01}(t_2) + \varphi_{11}C(S_{10}(t_1), S_{01}(t_2)), \tag{6}$$

where $C(\cdot, \cdot)$ is a bivariate copula function.

2.1 Distributional Assumptions on $S_{01}(\cdot)$ and $S_{10}(\cdot)$

Let $\{S_{\boldsymbol{\gamma}}(\cdot)\}$ denote a parametric family of survival functions with support on $\mathbb{R}^+$, where $\boldsymbol{\gamma}$ is a vector of parameters. In this paper, we consider three families: Weibull, log normal and log generalized extreme value (GEV). The Weibull distribution has survival function

$$S_{\boldsymbol{\gamma}}(t) = \exp\Big[-\Big(\frac{t}{\mu}\Big)^{\lambda}\Big], \text{ where } \boldsymbol{\gamma} = (\mu, \lambda) \in \mathbb{R}^+ \times \mathbb{R}^+. \tag{7}$$

The log normal has survival function

$$S_{\boldsymbol{\gamma}}(t) = 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right), \text{ where } \boldsymbol{\gamma} = (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+. \tag{8}$$

The log GEV has survival function

$$S_{\boldsymbol{\gamma}}(t) = \begin{cases} 1 - \exp\Big\{-\Big(1 + \xi\frac{\log t - \mu}{\sigma}\Big)_+^{-\frac{1}{\xi}}\Big\} & \text{if } \xi \neq 0 \\ 1 - \exp\Big\{-\exp\Big(-\frac{\log t - \mu}{\sigma}\Big)\Big\} & \text{if } \xi = 0, \end{cases} \tag{9}$$

where $\boldsymbol{\gamma} = (\mu, \sigma, \xi) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}$ and $x_+ = \max(0, x)$. Now, given each of the above distribution families, we assume $S_{01}(t_1) = S_{\boldsymbol{\gamma}_1}(t_1)$ and $S_{10}(t_2) = S_{\boldsymbol{\gamma}_2}(t_2)$.

The Weibull distribution can produce only monotonic hazard rates. In contrast, the shape of the hazard function for a logGEV is various such as U-shaped, or bell shaped, or a combination of both.

2.2 A Generalized FGM Bivariate Copula

In this paper, we use the generalized class of Farlie-Gumbel-Morgenstern (FGM) copula proposed by Bekrizadeh and Jamshidi (2017) which is given by

$$C_\theta^p(s,t) = st[1 + \theta(1-s)(1-t)]^p, p \in [1, \infty], \ \theta \in [-p^{-1}, p^{-1}], \ \forall (s,t) \in [0,1]^2. \tag{10}$$

When $p = 1$, it reduces to the symmetric FGM copula. The Spearman's ρ can be written as $\rho = 12\sum_{k=1}^{p}\binom{p}{k}\theta^k\Big[\frac{1}{(k+1)(k+2)}\Big]^2$, where the upper bound of ρ is up to 0.3805 approximately, and the lower bound is equal to -0.3333 which is same as that of symmetric FGM. Thus, the range of ρ in this generalized FGM is $[-0.3333, 0.3805]$. A good example is when $p = 3$, $\theta = 0.33 < \frac{1}{3}$, the estimated Spearmans ρ, $\rho \approx 0.3583$, which is out of the range of Spearmans ρ for FGM copula, that is $[-1/3, 1/3]$. The Kendall's τ can be written as $\tau = 4\int_0^1\int_0^1 c(s,t)C(s,t)dsdt - 1$. The estimated Kendals $\tau \approx 0.2397$ which is out of the range of τ for FGM copula, that is $[-2/9, 2/9]$. As we can see that the generalized FGM improve the correlation range which can be applied for more data. Also, the Spearmans ρ increases as p increases and θ is fixed. And Spearmans ρ increases as θ increases and p is fixed. As discussed in Bekrizadeh and Jamshidi (2017), if true correlation is within the range of FGM, the estimated correlations based on the generalized FGM copula showed strong consistency and was closer to the correlations which come from the observed data compared to the regular FGM copula.

2.3 Likelihood and MCMC

Denote (T_{i1}, T_{i2}) and (C_{i1}, C_{i2}) as bivariate lifetimes and corresponding censored bivariate times, respectively, for $i = 1, \cdots, n$. For each individual i, observed time can be denoted as $t_{ij} = \min(T_{ij}, C_{ij})$ by assuming (T_{i1}, T_{i2}) and (C_{i1}, C_{i2}) are independent. Denote $\delta_{ij} = I(t_{ij} = T_{ij})$ as a censoring indicator, $j = 1, 2$.

Let $\boldsymbol{\theta} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\varphi}, \theta, p)$ denote the set of model parameters, where $\boldsymbol{\varphi} = (\varphi_{00}, \varphi_{10}, \varphi_{01}, \varphi_{11})$, and $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ are parameters for $S_{01}(\cdot)$ and $S_{10}(\cdot)$ respectively. Considering the joint survival function, $S(t_{i1}, t_{i2})$, given by the equation (6) with the copula function given by the equation (10), the log-likelihood of i-th individual is given by

$$\begin{aligned} l_i(\boldsymbol{\theta}) = {} & \delta_{i1}\delta_{i2}\log f(t_{i1}, t_{i2}) + \delta_{i1}(1-\delta_{i2})\log\left(-\frac{\partial S(t_{i1}, t_{i2})}{\partial t_{i1}}\right) \\ & + \delta_{i2}(1-\delta_{i1})\log\left(-\frac{\partial S(t_{i1}, t_{i2})}{\partial t_{i2}}\right) + (1-\delta_{i1})(1-\delta_{i2})\log S(t_{i1}, t_{i2}), \end{aligned} \tag{11}$$

where

$$\begin{aligned} f(t_{i1}, t_{i2}) =& \frac{\partial^2 S(t_{i1}, t_{i2})}{\partial t_{i1} \partial t_{i2}} \\ =& \varphi_{11} f_1 f_2 [1 + \theta F_1 F_2]^{p-2} \Big\{ (1 + \theta F_1 F_2)[1 + \theta F_1 (F_2 - pS_2)] \\ &+ \theta p S_1 [(1 - 2F_2)(1 + \theta F_1 F_2) + \theta(p-1) F_1 F_2 (1 - F_2)] \Big\}, \end{aligned}$$

$$-\frac{\partial S(t_{i1}, t_{i2})}{\partial t_{i1}} = f_1 \Big\{ \varphi_{10} + \varphi_{11} S_2 (1 + \theta F_1 F_2)^{p-1} [1 + \theta F_2 (F_1 - pS_1)] \Big\},$$

$$-\frac{\partial S(t_{i1}, t_{i2})}{\partial t_{i2}} = f_2 \Big\{ \varphi_{01} + \varphi_{11} S_1 (1 + \theta F_1 F_2)^{p-1} [1 + \theta F_1 (F_2 - pS_2)] \Big\}, \text{ and}$$

$$S(t_{i1}, t_{i2}) = \varphi_{00} + \varphi_{10} S_1 + \varphi_{01} S_2 + \varphi_{11} S_1 S_2 (1 + \theta F_1 F_2)^p.$$

Here $S_1 = S_{10}(t_{i1})$, $S_2 = S_{01}(t_{i2})$, $F_1 = 1 - S_{10}(t_{i1})$, $F_2 = 1 - S_{01}(t_{i2})$, $f_1 = -\partial S_{10}(t_{i1})/\partial t_{i1}$ and $f_2 = -\partial S_{10}(t_{i2})/\partial t_{i2}$. Then the likelihood function of $\boldsymbol{\theta}$ for entire population is given by

$$L(\boldsymbol{\theta}) = \exp\Big(\sum_{i=1}^{n} l_i(\boldsymbol{\theta}) \Big). \tag{12}$$

We assume independent priors on the model parameters as

$$\pi(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\varphi}, \theta, p) = \pi(\boldsymbol{\gamma}_1)\pi(\boldsymbol{\gamma}_2)\pi(\boldsymbol{\varphi})\pi(\theta)\pi(p), \tag{13}$$

where a Dirichlet prior for $\boldsymbol{\varphi} \overset{set}{=} (\varphi_1, \varphi_2, \varphi_3, \varphi_4)$ with hyperparameter value $w_1 = w_2 = w_3 = w_4$ is

$$\pi(\boldsymbol{\varphi}) = \frac{\Gamma(\sum_{i=1}^{4} w_i)}{\prod_{i=1}^{4} \Gamma(w_i)} \prod_{i=1}^{4} \varphi_i^{w_i - 1}. \tag{14}$$

Also a $Beta(a_\theta, b_\theta)$ distribution is assigned to $\frac{1}{2}(1 - \theta)$ and an inverse gamma distribution $IG(a_p, b_p)$ is assigned to $p - 1$. Therefore, the joint posterior distribution can be written as

$$\pi(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\varphi}, \theta, p | \{t_{ij}\}) \propto L(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\varphi}, \theta, p) \times \pi(\boldsymbol{\gamma}_1)\pi(\boldsymbol{\gamma}_2)\pi(\boldsymbol{\varphi})\pi(\theta)\pi(p).$$

In order to guarantee proper posteriors, we adopt proper priors with known hyper-parameters. Thus, the following prior distributions are assigned to parameters of marginal distributions (1) for the log GEV distribution, we assume $\pi(\boldsymbol{\gamma}_1)\pi(\boldsymbol{\gamma}_2) = \pi(\mu_1)\pi(\sigma_1)\pi(\xi_1)\pi(\mu_2)\pi(\sigma_2)\pi(\xi_2)$, where $\mu_1, \mu_2 \sim N(0, \sigma^2_{\mu_j})$, $\sigma_1, \sigma_2 \sim IG(a_{\sigma_j}, b_{\sigma_j})$ and $\xi_1, \xi_2 \sim N(0, \sigma^2_{\xi_j})$; (2) for the Weibull distribution, we assume $\pi(\boldsymbol{\gamma}_1)\pi(\boldsymbol{\gamma}_2) = \pi(\mu_1)\pi(\mu_2)\pi(\lambda_1)pi(\lambda_2)$, where $\mu_j \sim Gamma(a_{\mu_j}, b_{\mu_j})$ and $\lambda_j \sim Gamma(a_{\lambda_j}, b_{\lambda_j})$, $j = 1, 2.$; (3) for the log normal distribution, we assume $\pi(\boldsymbol{\gamma}_1)\pi(\boldsymbol{\gamma}_2) = \pi(\mu_1)\pi(\sigma_1)\pi(\mu_2)\pi(\sigma_2)$, where $\mu_1, \mu_2 \sim N(0, \sigma^2_{\mu_j})$ and $\sigma_1, \sigma_2 \sim IG(a_{\sigma_j}, b_{\sigma_j})$.

Since its integration is not easy to perform, we use MCMC techniques to construct sample chains which are progressively more likely realizations of the distribution of the target distribution. Specifically, we simulate samples of parameters via Metropolis-Hastings (HM) steps within the Gibbs sampler. More details on the algorithm can be found in Web Appendix A

2.4 Model Comparison Criteria

To set notation, let $\mathcal{D}$, $\mathcal{D}_i$ and $\mathcal{D}_{-i}$ be the observed dataset, the ith data point, and the dataset with $\mathcal{D}_i$ removed, respectively, $i = 1, \ldots, n$. Let $L(\mathcal{D}|\boldsymbol{\theta})$ be the likelihood function based on observed data $\mathcal{D}$, and $L_i(\cdot|\boldsymbol{\theta})$ be the likelihood contribution based on $\mathcal{D}_i$ where $\boldsymbol{\theta}$ is the entire collection of model parameters under a particular model. Suppose $\{\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(\mathcal{L})}\}$ are random samples drawn from the full posterior $p_{post}(\boldsymbol{\theta}|\mathcal{D})$ and $\hat{\boldsymbol{\theta}} = \sum_{l=1}^{\mathcal{L}} \boldsymbol{\theta}^{(l)}/\mathcal{L}$ is the posterior mean estimate for $\boldsymbol{\theta}$.

Several model comparison methods are proposed in the literature. In the paper we will consider the following criteria.

(1) The deviance information criterion (DIC), a generalization of the Akaike information criterion (AIC), is commonly used for comparing complex hierarchical models for which the asymptotic justification of AIC is not appropriate

(Burnham & Anderson, 2004; Vaida & Blanchard, 2005). This criterion can be incorporated during the Monte Carlo simulation. Lower values of DIC indicate better adjustment. The expression of DIC can be written as

$$\text{DIC} = -2\log L(\mathcal{D}|\hat{\boldsymbol{\theta}}) + 2p_D, \tag{15}$$

where

$$p_D = 2\left(\log L(\mathcal{D}|\hat{\boldsymbol{\theta}}) - \frac{1}{\mathcal{L}}\sum_{l=1}^{\mathcal{L}}\log L(\mathcal{D}|\boldsymbol{\theta}^{(l)})\right).$$

(2) The conditional predictive ordinate (CPO) method represents a posterior predictive approach that has proven useful in Bayesian model selection Box (1980); M.-H. Chen, Ibrahim, and Sinha (2002); Gelfand and Dey (1994). CPO provides a useful cross-validation approach that is computationally efficient, requiring only a sample from the posterior distribution. Larger values for the CPO_i imply better models and lower values indicate influential observations. The conditional predictive ordinate (CPO) statistic for data point $\mathcal{D}_i$ is given by

$$\text{CPO}_i = f(\mathcal{D}_i|\mathcal{D}_{-i}) = \int L_i(\mathcal{D}_i|\boldsymbol{\theta})p_{post}(\boldsymbol{\theta}|\mathcal{D}_{-i})d\boldsymbol{\theta},$$

where $p_{post}(\cdot|\mathcal{D}_{-i})$ is the posterior density of $\boldsymbol{\theta}$ give $\mathcal{D}_{-i}$. As noted by (Gelfand & Dey, 1994), one can use importance sampling to estimate CPO_i by

$$\widehat{\text{CPO}}_i = \left\{\frac{1}{\mathcal{L}}\sum_{l=1}^{\mathcal{L}}\frac{1}{L_i(\mathcal{D}_i|\boldsymbol{\theta}^{(l)})}\right\}^{-1}. \tag{16}$$

The LPML can be written in terms of CPO as

$$\text{LPML} = \sum_{i=1}^{n}\log\widehat{\text{CPO}}_i. \tag{17}$$

3. Goodness of Fit Test

In this section, we are going to extend the pseudo in-and-out-of-sample test so called PIOS test proposed by J. Zhang and Peng (2007), which perform a goodness-of-fit on the hypotheses given that the marginal distribution for susceptible subjects are fully specified. The hypotheses are stated as below.

$$\begin{aligned} H_0 &: C_0 \in C = \{C(\cdot;\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta | F_{01}(t_1); F_{10}(t_2)\} \\ &\quad vs \\ H_1 &: C_0 \notin C = \{C(\cdot;\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta | F_{01}(t_1); F_{10}(t_2)\}, \end{aligned}$$

where $C_0(\cdot)$ is the true copula function, $\Theta \subset \Re^2$ is a 2-dimensional parameter space, and $F_{01}(t_1), F_{10}(t_2)$ are the CDF for the susceptible individuals in the lifetimes T_1 and T_2, respectively.

To derive the goodness-of-fit test statistic, two-step estimation technique was applied. According to Shih and Louis (1995), the first step is to estimate parameters $\mathbf{P}_j$, $\boldsymbol{\gamma}_j$ and $\boldsymbol{\varphi}$ by maximizing the marginal likelihood function $\sum_{i=1}^{n}\Big[\delta_{ij}\log f_j(t_{ij}) + (1-\delta_{ij})\log S_j(t_{ij})\Big]$, where $f_j(t) = -\partial S_j(t)/\partial t$, and denote the estimates as $\hat{\mathbf{P}}_j$, $\hat{\boldsymbol{\gamma}}_j$ and $\hat{\boldsymbol{\varphi}}$. The second step is to obtain a pseudo maximum likelihood estimates (PMLE) of (θ, p) by maximizing $\sum_{i=1}^{n} l_i(\hat{\boldsymbol{\gamma}}_1, \hat{\boldsymbol{\gamma}}_2, \hat{\boldsymbol{\varphi}}, \theta, p)$, and denote the estimates by $\hat{\theta}$ and $\hat{p}$. The PMLE $(\hat{\theta}, \hat{p})$ in "in-sample" pseudo log likelihood is obtained using the full data. And the PMLE $(\hat{\theta}_{(-i)}, \hat{p}_{(-i)})$ in "out-sample" pseudo log likelihood is obtained using the subset of data with ith observation removed. Then test statistic can be written as

$$T_n = \hat{l}_{in} - \hat{l}_{out} = \sum_{i=1}^{n} l_i(\hat{\boldsymbol{\gamma}}_1, \hat{\boldsymbol{\gamma}}_2, \hat{\boldsymbol{\varphi}}, \hat{\theta}, \hat{p}) - \sum_{i=1}^{n} l_i(\hat{\boldsymbol{\gamma}}_1, \hat{\boldsymbol{\gamma}}_2, \hat{\boldsymbol{\varphi}}, \hat{\theta}_{(-i)}, \hat{p}_{(-i)}), \tag{18}$$

where the log-likelihood function of *ith* data can be written as

$$\begin{aligned} l(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\varphi}, \theta, p; D_i) &= \delta_{i1}\delta_{i2}\log f(t_{i1}, t_{i2}) + (1-\delta_{i1})\delta_{i2}\log\left\{-\frac{\partial S(t_{i1}, t_{i2})}{\partial t_{i2}}\right\} \\ &\quad + (1-\delta_{i2})\delta_{i1}\log\left\{-\frac{\partial S(t_{i1}, t_{i2})}{\partial t_{i1}}\right\}. \end{aligned} \tag{19}$$

There are several good properties regarding to this test statistic. First, T_n in Equation (18) converges in probability to the dimension of the parameter vector of the copula under the null hypothesis. Secondly, T_n is asymptotically subject to normal distribution under null hypothesis. The proofs can be found in J. Zhang and Peng (2007).

Instead of estimating the asymptotic variance analytically, we use the following bootstrap technique to approximates the asymptotic variance of the test statistic in finite samples.

1: Calculate test statistic T_n for the bivariate survival times using the original n pairs of observations.

2: Sample with replacement with size n.

3: Calculate the test statistic, denoted as $T_n^{(b)}$, using the sampled data from step 2.

4: Repeat 2 and 3 B times so have B test statistics, denoted as T_n, $T_n^B = \{T_n^{(b)}, b = 1, \cdots, B\}$. Calculate the standard deviation, $sd\{T_n^B\}$. Finally, calculate the p-value which is $2\left[1 - \Phi\left(\left|\frac{T_n - m}{sd\left(T_n^B\right)}\right|\right)\right]$, where Φ is the cdf of a standard normal distribution and m is the dimension of the parameter vector of copula function ($m = 2$ in our case).

There is a practical issue of choosing B, the number of bootstrap samples. Scholars have recommended more bootstrap samples as available computing power has increased. However, increasing the number of samples cannot increase the amount of information in the original data. It can only reduce the effects of random sampling errors which can arise from a bootstrap procedure itself (Kloke, McKean, & McKean, 2015). Racine and MacKinnon (2007) discuss this issue at length and proposed a method for choosing the number of bootstrap samples. Theoretical results derived by Olive (2017) suggest using $B \geq [nlogn]$. We choose the number of bootstrap samples same as sample size because of computing power.

One thing I need to mention here is before we obtain a PMLE of copula parameters (θ, p) in step 2, we need to estimate $\boldsymbol{\varphi}$ using equations in (1) which is same to estimate $\rho = cov(d_{i1}, d_{i2})$, where d_{ij} is an indicator variable showing the ith subject is susceptible for the jth event and $P(d_{ij} = 1) = 1 - P_j$. For that we have to identify the subjects that are susceptible for the events and we should know how many subjects are actually susceptible for the jth event according to our model. Since we have cure rate estimates $\hat{P}_j$, it indicates $n(1 - \hat{P}_j)$ subjects are susceptible for the jth event. Note that the subject with an uncensored observation is susceptible. Let D_j be the number of uncensored observations for the jth event. Then we need to choose $n(1 - \hat{P}_j) - D_j \triangleq n_j$ observations from censored observations. Actually, any subject having censored lifetime observation could be susceptible for the event. However, the smaller the value of observation, the smaller the value of the survival function, the more likely the subject having this survival time is susceptible. Thus, we can make $d_{ij} = 1$ for the subjects having the first n_j smallest survival times among those censoring times. Once we have d_{ij}, we use the sample covariance, $\hat{\rho}$, between d_{i1} and d_{i2} to estimate ρ. Plugging $\hat{\rho}$ and $\hat{P}_j$ into equations in (1), we can get an estimate $\hat{\boldsymbol{\varphi}}$ for $\boldsymbol{\varphi}$.

4. Simulation Studies

4.1 Estimation

In this section, we are going to use the results from simulation studies to illustrate the performance of the proposed methodology. The following four models are selected for model comparison by choosing different survival functions $S_{01}(\cdot) = S_{\gamma_1}(\cdot)$ and $S_{10}(\cdot) = S_{\gamma_2}(\cdot)$ (Weibull, log normal or log GEV) and the copula functions $C(\cdot, \cdot)$ (FGM or generalized FGM) in the model specification of (6):

Model 1: $S_{\gamma_1}(\cdot)$ and $S_{\gamma_2}(\cdot)$ are from Weibull; $C(s,t) = st[1 + \theta(1-s)(1-t)]$;

Model 2: $S_{\gamma_1}(\cdot)$ and $S_{\gamma_2}(\cdot)$ are from Weibull; $C(s,t) = st[1 + \theta(1-s)(1-t)]^p$;

Model 3: $S_{\gamma_1}(\cdot)$ and $S_{\gamma_2}(\cdot)$ are from log GEV; $C(s,t) = st[1 + \theta(1-s)(1-t)]^p$;

Model 4: $S_{\gamma_1}(\cdot)$ and $S_{\gamma_2}(\cdot)$ are from log normal; $C(s,t) = st[1 + \theta(1-s)(1-t)]^p$.

The simulation study includes a total of 48 simulated data sets based upon the four models, four sample sizes ($n = 50, 100, 200, 100$) and three censoring rates (L: 15% to 20%, M: 30% to 40%, and H: 45% to 60%). Once the setting is fixed, follow the steps below to get one simulated data set.

Step 1: Draw two independent uniform random variables (u_{i1}, v_{i2}).

Step 2: Set $u_{i2} = C_{2|1}^{-1}(u_{i1}, v_{i2})$, where $C_{2|1}^{-1}$ denotes the pseudo-inverse of $C_{2|1}$. More specifically, solve the following equation for u_{i2} when $C(s,t) = st[1+\theta(1-s)(1-t)]^p$: $u_{i2}[1+\theta(1-u_{i1})(1-u_{i2})]^{p-1}[1+\theta(1-u_{i1})(1-u_{i2}) - \theta p u_{i1}(1-u_{i2})] - v_{i2} = 0$.

Step 3: Generate a bivariate survival times (T_{i1}, T_{i2}) from (u_{i1}, u_{i2}) via $T_{i1} = F_{\gamma_1}^{-1}(u_{i1})$ and $T_{i2} = F_{\gamma_2}^{-1}(u_{i2})$, where $F_{\gamma_j}^{-1}(\cdot)$ is the quantile function of the distribution corresponding to $S_{\gamma_j}(\cdot)$.

Step 4: Generate latent indicator values (d_{i1}, d_{i2}) according to the distribution of $P(d_1 = i, d_2 = j) = \varphi_{ij}$, where i, j is 0 or 1 and d_{ij} is an indicator variable with 1 indicating that the ith subject is susceptible for the jth event. If $d_{ij} = 0$, we change T_{ij} to be a big number, say $10,000$, since the ith subject is cured for the jth event.

Step 5: Simulate the censoring time C_{ij} from Weibull distributions, which results in censoring rate for different levels (L, M, H) where $i = 1, \cdots, n$ and $j = 1, 2$.

Step 6: Obtain the observed data $D = \{(t_{i1}, t_{i2}, \delta_{i1}, \delta_{i2}), i = 1, \cdots, n\}$, where $t_{i1} = \min(T_{i1}, C_{i1})$ and $t_{i2} = \min(T_{i2}, C_{i2})$, $\delta_{i1} = I(T_{i1} \le C_{i1})$ and $\delta_{i2} = I(T_{i2} \le C_{i2})$.

For true values of parameters, we have $(\mu_1, \lambda_1) = (30, 5)$, $(\mu_2, \lambda_2) = (20, 4)$,
$\theta = 0.6$ for model 1 and 2, $p = 1.5$ for model 2, 3 and 4 and $(\mu_1, \sigma_1, \xi_1) = (3, 0.2, 0.1)$, $(\mu_2, \sigma_2, \xi_2) = (2, 0.3, 0.2)$ for model 3. We have $(\mu_1, \sigma_1) = (2.5, 1)$,
$(\mu_2, \sigma_2) = (2, 1.5)$ for model 4. We have $(\varphi_{11}, \varphi_{10}, \varphi_{01}, \varphi_{00}) = (0.70, 0.15, 0.10, 0.05)$, $(\varphi_{11}, \varphi_{10}, \varphi_{01}, \varphi_{00}) = (0.40, 0.30, 0.20, 0.10)$ and $(\varphi_{11}, \varphi_{10}, \varphi_{01}, \varphi_{00}) = (0.15, 0.40,$
$0.25, 0.20)$ for low censoring, medium censoring and high censoring, respectively.

MATLAB 2017b is our first choice for all the computation and R is used to generate the tables and graphs of the results. We plot running means of variables of interest vs iteration number such as sample trace plots. By visual inspection, the chains are began from over-dispersed starting points. Therefore, we decided to discard the first 5,000 iterations which contributes the burn-in phase and run another 50,000 iterations. To reduce the correlations of successive samples, we store every 10th values of the chain after burn-in phase which result in 5000 samples for the posterior analysis.

Table 1 provides summary results of model comparison for the cases with $n = 200$ and $n = 500$. For each sample size, a total 12 data sets are generated from each one of the four models and three censoring rates. The bold entries indicate the best fitted model according to DIC and LPML. The effective number of parameter p_D indicates the model complexity. According to DIC and LPML in Table 1, the best model and true model are consistent except for one case with $n = 200$ and medium censoring rate. However, Model 3 performs better than any other models even the true model is not Model 3 in small sample sizes such as $n = 50$ and $n = 100$. This information is available in Table 1 in Web Appendix B.

The posterior mean, standard deviation (SD) and %95 HPD interval under different scenarios are available in Tables 2 - 5 in Web Appendix B. As we can see that the true values of parameters are inside the 95% highest posterior density (HPD) interval. The posterior mean of association parameters are essentially unbiased under the true model. The estimated standard deviation is fairly small. Figure 1 shows the estimated survival function based upon Kaplan-Meier estimator and four different models when the Model 3 is true model. It shows that the estimated survival function based on Model 3 is closer to Kaplan-Meier estimates compared to other three models. However, the estimated survival function of Model 3 does not show big difference from Kaplan-Meier estimate, even Model 3 is not true model. This information can be seen in Figures 1 - 3 in Web Appendix B. The examination of these tables and figures shows misspecification of model can lead to significantly biased estimates which result in inaccurate inference and incorrect conclusions.

Table 1. Model comparison results for n=200 and 500.

True Model	Censoring rate	Fitted Model	n = 200			n = 500		
			DIC	LPML	p_D	DIC	LPML	p_D
1	L	1	**2433**	**-1216**	7.78	**6040**	**-3020**	7.87
		2	2434	-1217	7.82	6041	-30201	7.95
		3	2453	-1226	8.70	6050	-3025	9.32
		4	2478	-1239	7.91	6119	-3060	7.94
	M	1	**2104**	**-1052**	7.77	**5343**	**-2672**	7.72
		2	2105	-1053	7.85	5344	-2673	7.92
		3	2108	-1055	8.39	5394	-2697	9.25
		4	2132	-1066	7.72	5415	-2708	7.67
	H	1	**1749**	**-875**	7.45	**4243**	**-2121**	7.27
		2	1750	-876	7.62	4244	-2122	7.58
		3	1754	-877	8.48	4258	-2129	10.98
		4	1790	-895	7.43	4332	-2166	7.44
2	L	1	2411	-1206	7.51	6017	-3009	7.57
		2	**2410**	**-1205**	7.86	**6016**	**-3008**	7.70
		3	2414	-1207	8.31	6030	-3018	9.15
		4	2450	-1225	7.67	6118	-3059	7.83
	M	1	2114	-1057	7.13	5231	-2615	7.44
		2	2113	-1056	7.45	**5230**	**-2614**	7.69
		3	**2109**	**-1054**	8.69	5262	-2630	9.36
		4	2150	-1075	7.87	5323	-2661	7.65
	H	1	1702	-852	6.96	4287	-2144	7.56
		2	**1700**	**-850**	8.54	**4286**	**-2143**	7.76
		3	1709	-855	8.57	4298	-2149	9.79
		4	1731	-866	7.66	4365	-2182	7.76
3	L	1	2465	-1235	6.97	6038	-3022	7.47
		2	2463	-1234	8.37	6037	-3021	7.81
		3	**2187**	**-1093**	9.91	**5473**	**-2736**	9.82
		4	2258	-1129	7.78	5651	-2826	7.90
	M	1	2169	-1089	7.12	5314	-2662	7.06
		2	2168	-1088	7.48	5313	-21661	7.76
		3	**1974**	**-987**	9.42	**4861**	**-2430**	9.71
		4	2019	-1010	7.65	4973	-2487	7.67
	H	1	1752	-883	7.00	4215	-2110	7.29
		2	1751	-882	7.22	4214	-2009	7.47
		3	**1596**	**-798**	8.87	**3944**	**-1972**	9.41
		4	1069	-535	7.68	3978	-1990	7.87
4	L	1	2594	-1297	6.80	6486	-3243	6.97
		2	2595	-1298	7.30	6487	-3244	7.48
		3	2581	-1290	8.89	6452	-3226	9.11
		4	**2575**	**-1288**	7.25	**6437**	**-3219**	7.43
	M	1	2283	-1142	7.50	5708	-2854	7.68
		2	2283	-1142	8.05	5709	-2855	8.24
		3	2271	-1136	9.80	5678	-2839	9.76
		4	**2266**	**-1133**	7.99	?**5665**	?**-2833**	8.19
	H	1	1826	-925	7.29	4566	-2312	7.47
		2	1827	-925	7.82	4567	-2312	8.01
		3	1817	-920	9.52	4542	-2299	9.49
		4	?**1813**	?**-918**	7.77	?**4532**	?**-2295**	7.96

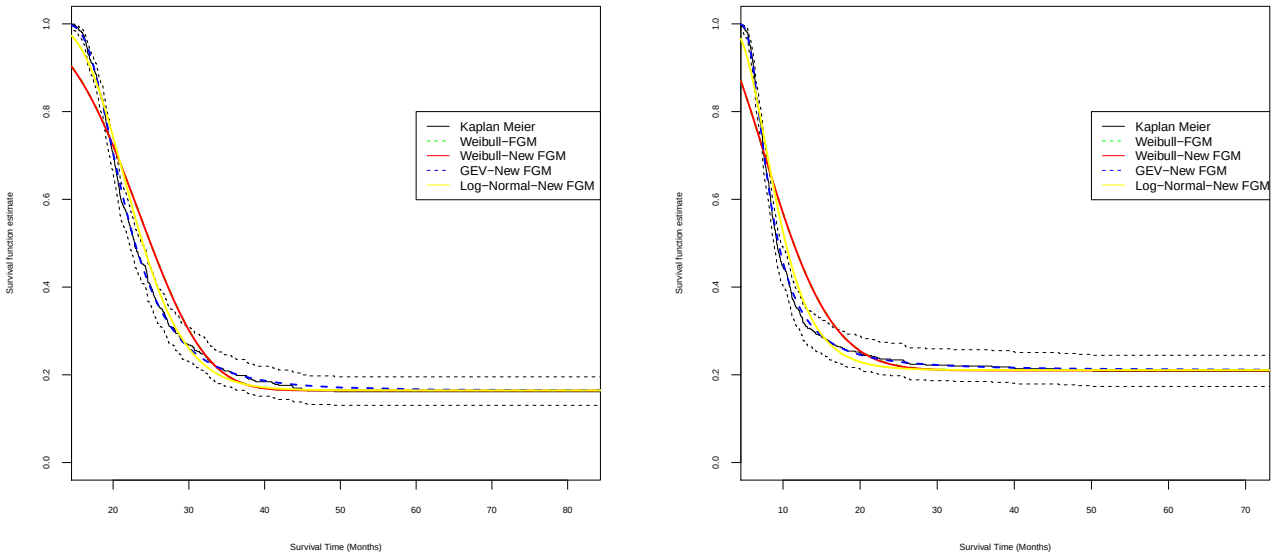

Figure 1. The plots of estimated survival functions of different models based upon simulated data sets from model 3 with n = 500 and low censoring rate. Left: Survival function estimate for treated eye; Right: Survival function estimate for untreated eye

4.2 Goodness of Fit Performance

In this section, we perform goodness of fit test for our proposed model rely on empirical type I error and test power. The following copulas are considered: (1) the generalized FGM, (2) Clayton, and (3) Gaussian. We generate a total of 36 simulated datasets for each different marginal distribution (Weibull, log GEV and log normal) based on three copulas, with four different sample sizes (100, 200, 500, 1000), under three levels of censoring rate (20%, 40%, 60%), and three levels of Kendalls τ, ($\tau = 0.16, 0.20, 0.24$). For each dataset, the empirical p-value is calculated using bootstrap samples.

First, we need to generate the data based on three copulas, four different sample sizes, three levels of censoring rate and three levels of Kendall's τ.

Secondly, we can follow the steps in Section 3 to calculate the test statistics and produce the p-value through bootstrap samples. Recall in our bootstrap steps, we state that the subject associated with smaller censored lifetime is more likely to be the susceptible subject. On the other hand, the subject associated with bigger censored lifetime is more likely to be the cured subject. Note that the likelihood function in Equation (18) is based upon the entire population which implies this test is to perform the goodness-of-fit for the model, not only for the copula. We can limit the population to the susceptible subjects which is the subset of the entire population. In this way, we can test the goodness-of-fit test only for copula. We can still follow the step in Section 3 to identify the susceptible subjects that have censored lifetime, and change the joint survival function and joint density function involved in Equation (18) to the ones of susceptible subjects. The type I errors and test powers are available in Tables 9 - 11 and Tables 15 - 17, respectively, in Web Appendix C.

Table 6 - 8 in Web Appendix C report the empirical type I error under different sample sizes and with three different marginal distributions for susceptible individuals. The type I error is empirical proportions to reject the null hypothesis when the null hypothesis is assumed to be true at significance level equal to 5%. As we can see from these tables, the overall performance of our test is good, especially for sample size equal to 1000. Regardless of marginal distributions and censoring rates, the type I error decreases as sample size increases. Type I error does not show certain trend when Kendall's τ increases.

Table 12 - 14 in Web Appendix C report the empirical power under different sample sizes and with three different marginal distributions for susceptible individuals. As we can see from these tables, the overall performance of our test in differentiating among new FGM, Clayton and Gaussian copula is good. However, n its hard to differentiate the Gaussian copula and Clayton copula as the power of rejecting the Clayton copula is low, even with big sample data simulated from Gaussian. This might be due to the right censoring of simulated data, which leads to insufficient information of the upper tail dependence from the data. It is noted that the test power increases as the sample sizes increases, or the censoring rate

decreases, but does show certain trend when Kendall's τ increases.

It is also noted that type I error and test power do not change significantly when marginal distributions are changed. It implies that our test works regardless of marginal distribution.

5. Real Data Analysis

In this section, we demonstrate our proposed model by using the Diabetic Retinopathy Study (DRS) data which was first considered by Huster, Brookmeyer, and Self (1989). There are 162 patients and each patient received laser treatment for one eye and no treatment for the other eye. In the analysis considered here, the time to blindness for the eye randomized to laser treatment and not received the treatment are denoted as T_1 and T_2, respectively. The blindness is defined as the time from initiation of treatment to the time when visual acuity dropped below 5/200 two visits in row. To check the data, there are 75 censored observations in T_1 and 84 censored observations in T_2. For those censored observations, some patients will not have the occurrence of blindness in the period of study because of drop-out and end of study.

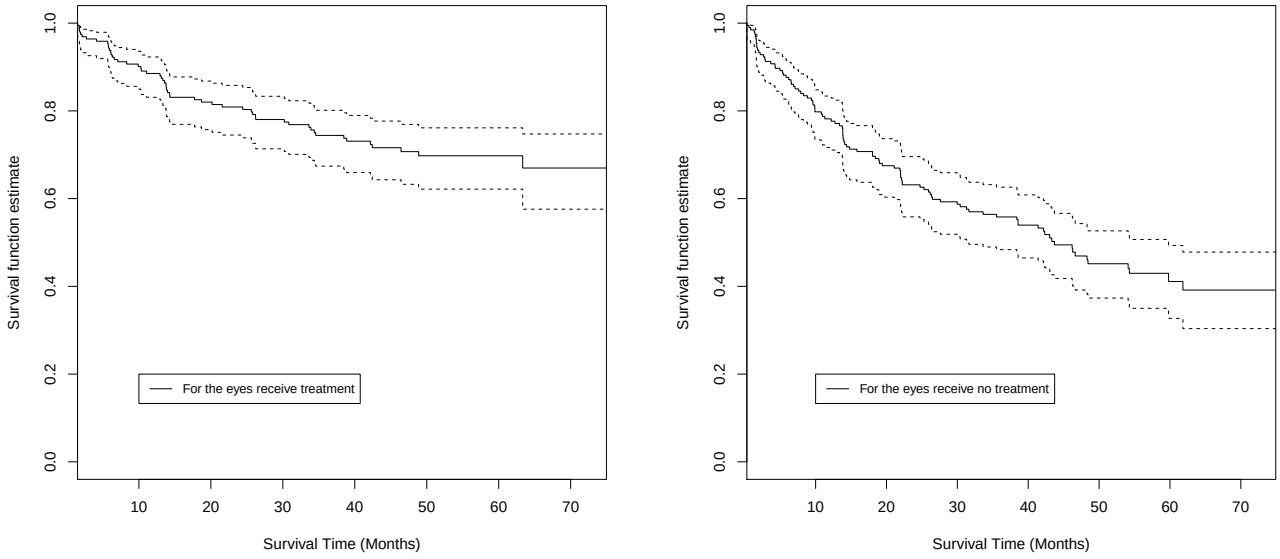

Figure 2. Kaplan-Meier nonparametric estimates for the survival function

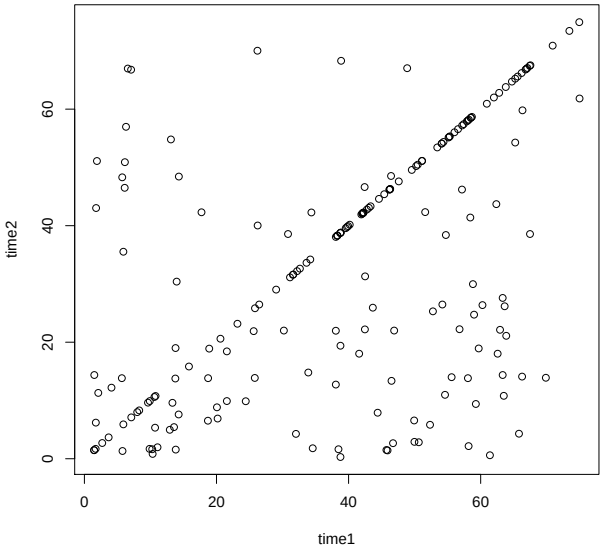

Figure 3. Scatter Plot for DRS data

As a preliminary analysis of the data, Figure 2 gives estimated survival function of T_1 and T_2 based upon the Kaplan Meier estimates. From this plot, we have some indication of cure fraction. Also, it seems to have better results for the eye that received treatment which has larger time to blindness in comparison with the eye that received no treatment. Figure 3 gives the scatter plot for T_1 and T_2. From this plot, we observe the two lifetimes spread out all over the place which means there is no certain trend for the relationship between T_1 and T_2. In other word, the correlation between T_1 and T_2 is weak. From Figure 2 and Figure 3, cure rate model should be considered when a significant proportion of patients are "cured" and copulas with week dependence should be considered.

We fit the following six models such as log GEV, log normal and Weibull with symmetric FGM and generalized FGM copulas. The model of log GEV with generalized FGM copula gives lowest DIC (1367.31) and highest LPML (-683.61) so that our proposed model with log GEV with generalized FGM copula is the best for fitting the data.

As we discussed in the simulation study, the first 5,000 iterations will be ignored, and another 50,000 iterations will be used to consider the simulation of each parameters. To get approximated uncorrelated values, we store every 10th values of the chain after burn-in phase which gives a final chain of size 5,000 for the posterior analysis. We also apply our test procedure to this data. The empirical estimate of Kendalls rank correlation is 0.376. The corresponding p-value of our test for the generalized FGM is 0.296. At the significant level 0.05, we failed to reject these this copula.

Table 2. Posterior summaries: Bivariate log GEV distribution based upon a generalized FGM copula

Parameter	Mean	SD	95% Credible Interval	DIC	LPML
μ_1	3.15	0.40	(2.37 , 3.91)		
μ_2	2.55	0.21	(2.15 , 2.96)		
σ_1	1.72	0.42	(1.01 , 2.59)		
σ_2	1.41	0.15	(1.15 , 1.71)		
ξ_1	-0.21	0.30	(-0.82 , 0.35)		
ξ_2	-0.60	0.17	(-0.92 , -0.26)	1367.31	-683.61
φ_{00}	0.23	0.08	(0.06 , 0.36)		
φ_{01}	0.15	0.10	(0.00 , 0.33)		
φ_{10}	0.06	0.05	(0.00 , 0.15)		
φ_{11}	0.57	0.12	(0.32 , 0.78)		
p	1.68	0.76	(1.09 , 2.79)		
θ	0.35	0.23	(-0.08 , 0.78)		
θ	0.42	0.20	(0.05 , 0.8)		

Table 3. The DRS data: Bayesian criteria for models proposed in Louzada et al. (2013)

	DIC	LPML
FGM Weibull	1522	-761.81
FGM Exponential	1525	-763.46
PSF Weibull	1527	-764.89
PSF Exponential	1524	-762.87
Frank Weibull	1522	-762.04
Frank Exponential	1525	-763.69
Clayton Weibull	1523	-762.47
Clayton Exponential	1525	-763.93
Independence Weibull	1528	-764.55
Independence Exponential	1530	-765.93

Table 2 shows the posterior summaries of interest assuming log GEV as marginal distribution for the lifetime T_1 and T_2 for susceptible subjects using a generalized FGM. The posterior mean of all the parameters are in the %95 HPD interval. The standard deviation is relatively small. The time to blindness of untreated eye has a lower cure rate compared to that of treated eye. The posterior estimates for other models are shown in tables 18 - 22 in Web Appendix D. Table 3 presents the model comparison criteria for the same DRS data which is discussed in Louzada et al. (2013). The results show that our method performs better than all the models proposed in Louzada et al. (2013).

The 3D plot of the joint survival function and corresponding contour plot for model of log GEV with generalized FGM copula presented in Figure 4. It shows that the joint survival functions are decreasing as time t_1 goes up or time t_2 goes

up. It also shows that the joint survival function decreases slowly when t_1 goes up compared to when t_2 goes up. That implies that the treatment actually has positive effect on the eye.

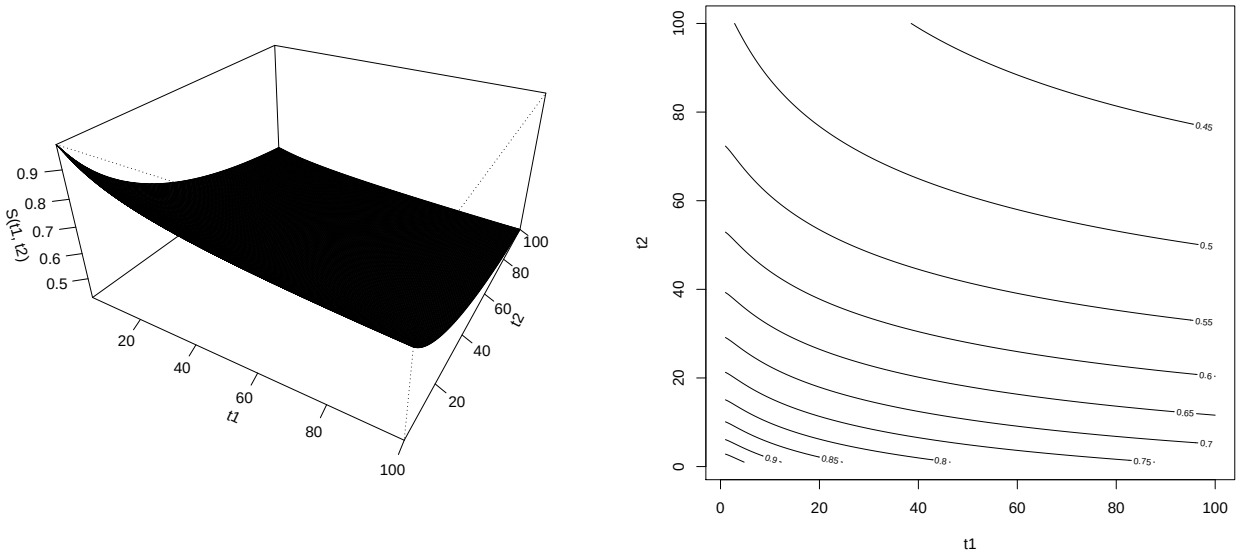

Figure 4. The 3D plot of joint survival function and corresponding contour plot under Model 3. Left: Joint survival function; Right: Contour plot

6. Discussion

The use of copula functions could be a good alternative way to analyze bivariate lifetime data. Observe that in many applications of lifetime modeling we could have individuals that are "long term survivors" or "cure individuals". An analytical structure of the statistical methodology was developed to model the dependence between cure rate fractions, and an extremely flexible generalized extreme value distribution was employed to model the logarithm of the survival time. It is very useful to use copulas to avoid the problem of the marginal distributions depending on the dependence structure, especially the use of the generalized FGM copula. This allows a broader range of correlations than the typical FGM copula indicating that the methods can be applied to more data contexts.

To check for adequacy of the generalized copula for our situation, we have extended the PIOS test to the new proposed test for right-censored bivariate survival times where the possibility of cure must be incorporated. The fact that the performance of our test does not depend on the choice of marginal distributions provides a lot of flexibility and avoids model misspecification issues. Also, the test is computationally straightforward and easily constructed.

References

Bekrizadeh, H., & Jamshidi, B. (2017). A new class of bivariate copulas: dependence measures and properties. *METRON*, *75*(1), 31–50.

Box, G. E. (1980). Sampling and bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society: Series A (General)*, *143*(4), 383–404.

Burnham, K., & Anderson, D. (2004). Model selection and multi-model inference. *Second. NY: Springer-Verlag*, *63*(2020), 10.

Chen, C.-M., Lu, T.-F. C., & Hsu, C.-M. (2013). Association estimation for clustered failure time data with a cure fraction. *Computational Statistics & Data Analysis*, *57*(1), 210–222.

Chen, M.-H., Ibrahim, J. G., & Sinha, D. (2002). Bayesian inference for multivariate survival data with a cure fraction. *Journal of Multivariate Analysis*, *80*(1), 101–126.

Fermanian, J.-D. (2005). Goodness-of-fit tests for copulas. *Journal of multivariate analysis*, *95*(1), 119–152.

Gallardo, D. I., Gómez, Y. M., & de Castro, M. (2018). A flexible cure rate model based on the polylogarithm distribution. *Journal of Statistical Computation and Simulation*, *88*(11), 2137–2149.

Gelfand, A. E., & Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society: Series B*, *56*(3), 501–514.

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

Huster, W. J., Brookmeyer, R., & Self, S. G. (1989). Modelling paired survival data with covariates. *Biometrics*, 145–156.

Kloke, J., McKean, J. W., & McKean, J. W. (2015). *Nonparametric statistical methods using r*. CRC Press Boca Raton.

Lakhal-Chaieb, L., & Duchesne, T. (2017). Association measures for bivariate failure times in the presence of a cure fraction. *Lifetime Data Analysis*, *23*(4), 517–532.

Louzada, F., Suzuki, A. K., & Cancho, V. G. (2013). The fgm long-term bivariate survival copula model: modeling, bayesian estimation, and case influence diagnostics. *Communications in Statistics-Theory and Methods*, *42*(4), 673–691.

Olive, D. J. (2017). Theory for linear models. In *Linear regression* (pp. 313–342). Springer.

Peng, Y., & Taylor, J. M. (2011). Mixture cure model with random effects for the analysis of a multi-center tonsil cancer study. *Statistics in medicine*, *30*(3), 211–223.

Racine, J. S., & MacKinnon, J. G. (2007). Simulation-based tests that can use any number of simulations. *Communications in StatisticsSimulation and Computation®*, *36*(2), 357–365.

Romeo, J. S., Tanaka, N. I., & Pedroso-de Lima, A. C. (2006). Bivariate survival modeling: a bayesian approach based on copulas. *Lifetime Data Analysis*, *12*(2), 205–222.

Shih, J. H., & Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 1384–1399.

Vaida, F., & Blanchard, S. (2005). Conditional akaike information for mixed-effects models. *Biometrika*, *92*(2), 351–370.

Wang, W., & Wells, M. T. (2000). Model selection and semiparametric inference for bivariate failure-time data. *Journal of the American Statistical Association*, *95*(449), 62–72.

Yashin, A. I., & Iachine, I. A. (1999). Dependent hazards in multivariate survival problems. *Journal of Multivariate Analysis*, *71*(2), 241–261.

Zhang, J., & Peng, Y. (2007). A new estimation method for the semiparametric accelerated failure time mixture cure model. *Statistics in medicine*, *26*(16), 3157–3171.

Zhang, S., Okhrin, O., Zhou, Q. M., & Song, P. X.-K. (2016). Goodness-of-fit test for specification of semiparametric copula dependence models. *Journal of Econometrics*, *193*(1), 215–233.