A FRAILTY-COPULA MODEL FOR DEPENDENT COMPETING RISKS IN RELIABILITY THEORY

YIN-CHEN WANG
Graduate Institute of Statistics, National Central University, Taiwan
email: peacebubu@gmail.com

TAKESHI EMURA*
Graduate Institute of Statistics, National Central University, Taiwan
email: takeshiemura@gmail.com

Dependent competing risks often arise in industrial life tests, where multiple types of failure determine the total lifespan of a unit. To make inference on multiple failure time distributions, two different models have been developed in reliability theory: copula models and frailty models. The objective of this paper is to propose a frailty-copula model for reliability theory, which is a hybrid model including both a frailty term (for heterogeneity) and a copula function (for dependence). We derive properties of the models that are useful to assess the reliability of units. We develop likelihood-based inference methods based on competing risks data. We also develop a model-diagnostic procedure based on the nonparametric estimation of the sub-distribution functions. We conduct simulations to examine the performance of the proposed methods.

Keywords: Competing risks, Copula, Frailty, Survival analysis, Weibull distribution.

1. Introduction

Dependent competing risks often arise in industrial life tests, biomedical follow-up studies, and animal experiments, where multiple types of failure determine the total lifespan of a unit. To make inference on multiple failure time distributions, copula-based dependent failure time models (Zheng and Klein 1995; Escarela and Carriere 2003; Lo and Wilke 2010; Shih and Emura 2018; Hsu et al. 2017; Emura and Chen 2016; Emura and Michimae 2017; Zhou et al. 2018; Zhang et al. 2018; Shih et al. 2018; Emura et al. 2019) have been considered in recent studies. The idea of copula models is to make inference on marginal failure time distributions by imposing a copula structure among different failure types.

---

* Corresponding Author

In the literature of reliability and competing risks theory, copula models are getting more popular than those traditional bivariate parametric models, such as a bivariate Weibull model (Moeschberger 1974; David and Moeschberger 1978; Fan et al. 2019), a bivariate normal model (Nádas 1971; Moeschberger 1974; Basu and Ghosh 1978) and an independent Lindley model (Mazucheli and Achcar 2011). An important advantage of copula models is that the copula parameters are separately interpretable from the marginal models. Recently, Hsu et al. (2017) and Zhou et al. (2018) proposed copula-based approaches for marginal regression models for reliability theory.

Another strategy to analysis of dependent competing risks is based on a shared frailty model (Liu 2012) or mixed model (Lo et al. 2017). The frailty model considers an unobserved frailty value, which specifies frail units via a large value (easy to fail) or robust units via a small value (difficult to fail). Liu (2012) proposed a gamma frailty model for reliability theory, where the conditional failure times follow a log-location-scale model. The mixed model is essentially the same as the frailty model. It should be emphasized that the frailty or mixed model considers a conditional model (given a frailty term), which is different from the copula models that consider marginal (unconditional) models. Thus, the interpretation of the copula and frailty models is essentially different.

A one-to-one relationship between a frailty distribution and an Archimedean copula does not imply the equivalence between the frailty model and copula model. For instance, a frailty model with the conditional piecewise constant hazard of Lo et al. (2017) no longer has piecewise constant hazards in their marginal. Hence, their model is not equivalent to a copula model with marginal piecewise constant hazards of Emura and Michimae (2017), even though the two models can have the same copula. Consequently, one cannot directly compare the frailty model and the copula model since their parametrizations are different (Section 3.3.4 of Duchateau and Janssen 2008).

A key in distinguishing between the frailty model and the copula model are the parameter interpretation. In the frailty model, the variance parameter (Duchateau and Janssen 2008) is interpreted as the degree of unobserved "heterogeneity" among units. Hence, the variance parameter in the frailty distribution is not purely representative of "dependence" among units, as it also influences the marginal distributions. In general, the copula model is suitable for purpose of studying dependence among units while the frailty model is suitable for purpose of studying heterogeneity among units.

The objective of this paper is to propose a frailty-copula model for reliability theory, which is a model including both a frailty term (for heterogeneity) and a copula function (for dependence). An important advantage of this hybrid model is that it allows to compare the frailty model and copula model in a systematic manner. In the literature, frailty and copula models have been separately studied. Only Rotolo et al. (2013) and Emura et al. (2017) considered survival models having both frailty and copula for “clustered” competing risks data. Our proposed model is motivated by the joint frailty-copula model of Emura et al. (2017). The purpose and setting of their paper are largely different from our objective. We suitably modify their model for purpose of reliability theory.

This paper is organized as follows. Section 2 presents our proposed method. Section 3 provides simulations to check the performance of the proposed method.

2. Proposed method

2.1. Competing risks data

Let T_{ij} be continuous failure time of a unit $i \in \{1, \ldots, n\}$ due to type $j \in \{1, 2\}$ failure mode. Also, let C_i random censoring time or the length of testing duration. Under competing risks, one can observe the first occurring event time $T_i = \min(T_{i1}, T_{i2}, C_i)$ and one of the three event types (Event 1, Event 2, and Censoring). We define two event indicators $\delta_i = \mathbf{I}(T_i = T_{i1})$ and $\delta_i^* = \mathbf{I}(T_i = T_{i2})$. What we observe are $\{ (T_i, \delta_i, \delta_i^*) : i = 1, \ldots, n \}$.

2.2. Proposed model

We assume that the individual units under an experiment are heterogeneous in the sense that some units are fragile (easy to fail) and some are robust (difficult to fail). Fragile items may have short failure times for Event 1 and Event 2, while robust items may have long failure times for them. Hence, this heterogeneity usually yields positive correlation between T_{i1} and T_{i2} (Duchateau and Janssen 2008).

To model the heterogeneity, we consider an unobserved frailty term Z_i that is a positive random variable with mean=1 and variance=η. We specifically consider a gamma frailty model

$$f_Z(z) = \frac{1}{\Gamma(1/\eta)\eta^{1/\eta}} z^{1/\eta - 1} \exp\left(-\frac{z}{\eta}\right), \qquad \eta > 0, \quad z > 0.$$

The term Z_i represents a factor that influences the two failure types: $Z_i > 1$ specifies fragile units, $Z_i < 1$ specifies robust units, and $Z_i \approx 1$ specifies normal units. The variance η represents the degree of heterogeneity.

We assume that the conditional survival function of T_{ij} given Z_i follows a

Weibull distribution

$$S_{T_{ij}|Z_i}(t \mid Z_i = z) = P(T_{ij} > t \mid Z_i = z) = P(T_{ij} > t \mid Z_i = 1)^z$$
$$= \left(\exp\left[-\left\{ \frac{t}{\exp(\mu_j)} \right\}^{\frac{1}{\sigma_j}} \right] \right)^z = \exp\left[-z \left\{ \frac{t}{\exp(\mu_j)} \right\}^{\frac{1}{\sigma_j}} \right],$$

where μ_j is a location parameter and σ_j is a scale parameter.

In the frailty model of Liu (2012), it is assumed that the pairs (T_{i1}, T_{i2}) are conditionally independent, i.e.

$$P(T_{i1} > t_1, T_{i2} > t_2 \mid z) = \exp\left[-z \left\{ \frac{t_1}{\exp(\mu_1)} \right\}^{\frac{1}{\sigma_1}} \right] \exp\left[-z \left\{ \frac{t_2}{\exp(\mu_j)} \right\}^{\frac{1}{\sigma_2}} \right]. \quad (1)$$

Emura et al. (2017) proposed a joint frailty-copula model by relaxing the conditional independence assumption (1). Emura et al. (2017) considered the Clayton copula for dependence and the semiparametric Cox models for the marginal regression under clustered competing risks data. To modify the idea of the frailty-copula model to be adaptive to the Weibull parametric competing risks model in reliability theory (also to a non-clustered model), we restrict our attention to the Gumbel copula instead of the Clayton copula.

Accordingly, we propose a frailty-copula model by specifying the conditional bivariate survival function with the Gumbel copula:

$$S_{T_{i1}, T_{i2}|Z_i}(t_1, t_2 \mid z) = P(T_{i1} > t_1, T_{i2} > t_2 \mid z)$$
$$= \exp\left\{ -\left(\left[-\log\left\{ S_{T_{i1}|Z_i}(t_1 \mid z) \right\} \right]^{\theta+1} + \left[-\log\left\{ S_{T_{i2}|Z_i}(t_2 \mid z) \right\} \right]^{\theta+1} \right)^{\frac{1}{\theta+1}} \right\}$$
$$= \exp\left(-z \left[\left\{ \frac{t_1}{\exp(\mu_1)} \right\}^{\frac{\theta+1}{\sigma_1}} + \left\{ \frac{t_2}{\exp(\mu_2)} \right\}^{\frac{\theta+1}{\sigma_2}} \right]^{\frac{1}{\theta+1}} \right), \quad (2)$$

where $\theta \geq 0$ is a parameter of the Gumbel copula. If $\theta = 0$, the proposed model reduces to the model of Liu (2012) in Equation (1). Hence, the proposed model is a generalization of the model of Liu (2012) with an additional parameter θ.

2.3 Properties of the proposed model

We derive several useful quantities in reliability theory, such as the quantiles under the proposed model (2). Although our model is a variant of the joint frailty-copula model of Emura et al. (2017), such derivations have not been considered.

We rewrite the proposed model (2) as $\exp(-zA_\theta)$ where

$$A_\theta \equiv A_\theta(t_1,t_2) \equiv \left[\sum_{j=1}^{2} \left\{ \frac{t_j}{\exp(\mu_j)} \right\}^{\frac{\theta+1}{\sigma_j}} \right]^{\frac{1}{\theta+1}} .$$

The unconditional survival function of (T_{i1},T_{i2}) can be explicitly written as

$$S_{T_{i1},T_{i2}}(t_1,t_2) = \int_0^\infty S_{T_{i1},T_{i2}|Z_i}(t_1,t_2 \mid z) f_Z(z)dz = \left(1+\eta A_\theta\right)^{-\frac{1}{\eta}} \qquad (3)$$

The marginal survival function and the p-th quantile function follow the three-parameter Burr XII model (Burr 1942):

$$S_{T_{ij}}(t) = \left[1+\eta \left\{ \frac{t}{\exp(\mu_j)} \right\}^{\frac{1}{\sigma_j}} \right]^{-\frac{1}{\eta}}, \quad t_{p,j} = \exp\left(\mu_j\right) \left\{ \frac{\left(1-p\right)^{-\eta}-1}{\eta} \right\}^{\sigma_j},$$

for $j=1$ and 2. The model is often used in reliability theory due to its flexibility and good mathematical properties (Watkins 1999; Belaghi and Asl 2016).

2.4 Likelihood-based inference

To estimate parameters of the proposed model (2), we propose a likelihood-based method. The sub-density function $f(t,j)$ for the j-th failure type is

$$f(t,j) = -\left.\frac{\partial S_{T_{i1},T_{i2}}(t_1,t_2)}{\partial t_1}\right|_{t_1=t_2=t} = \left(1+\eta \left[\left\{ \frac{t}{\exp(\mu_1)} \right\}^{\frac{\theta+1}{\sigma_1}} + \left\{ \frac{t}{\exp(\mu_2)} \right\}^{\frac{\theta+1}{\sigma_2}} \right]^{\frac{1}{\theta+1}} \right)^{-\frac{1+\eta}{\eta}}$$

$$\times \left[\left\{ \frac{t}{\exp(\mu_1)} \right\}^{\frac{\theta+1}{\sigma_1}} + \left\{ \frac{t}{\exp(\mu_2)} \right\}^{\frac{\theta+1}{\sigma_2}} \right]^{\frac{1}{\theta+1}-1} \times \frac{t^{\frac{\theta+1}{\sigma_j}-1}}{\sigma_j \exp\left\{ \mu_j \left(\frac{\theta+1}{\sigma_j} \right) \right\}} .$$

Therefore, the likelihood function can be derived as

$$L\left(\mu_1,\mu_2,\sigma_1,\sigma_2,\eta,\theta\right) = \prod_{i=1}^{n} f(T_i,1)^{\delta_i} f(T_i,2)^{\delta_i^*} S_{T_{i1},T_{i2}}(T_i,T_i)^{1-\delta_i-\delta_i^*} .$$

The log-likelihood function can be re-expressed as

$$
\begin{aligned}
l(\mu_1,\mu_2,\sigma_1,\sigma_2,\eta,\theta) = -\Bigg\{ & m\log\sigma_1 + m^*\log\sigma_2 + m(\theta+1)\frac{\mu_1}{\sigma_1} + m^*(\theta+1)\frac{\mu_2}{\sigma_2}\Bigg\} \\
& + \sum_{i=1}^{n}\left\{\delta_i\left(\frac{\theta+1}{\sigma_1}-1\right)+\delta_i^*\left(\frac{\theta+1}{\sigma_2}-1\right)\right\}\log T_i \\
& - \frac{\theta}{\theta+1}\sum_{i=1}^{n}\left(\delta_i+\delta_i^*\right)\log\left[\sum_{k=1}^{2}\left\{\frac{T_i}{\exp(\mu_k)}\right\}^{\frac{\theta+1}{\sigma_k}}\right] \\
& - \sum_{i=1}^{n}\left(\delta_i+\delta_i^*+\frac{1}{\eta}\right)\log\left(1+\eta\left[\sum_{k=1}^{2}\left\{\frac{T_i}{\exp(\mu_k)}\right\}^{\frac{\theta+1}{\sigma_k}}\right]^{\frac{1}{\theta+1}}\right),
\end{aligned}
$$

where $m=\sum_{i=1}^{n}\delta_i$ and $m^*=\sum_{i=1}^{n}\delta_i^*$. Some restrictions on θ and η will be imposed to avoid the non-identifiability (Tsiatis 1975). What we use is the most common one, the known value of θ (Zheng and Klein 1995). Then, the maximum likelihood estimator (MLE) is obtained numerically, for instance by using “*nlm( )*” or “*optim( )*” in R.

2.5 Goodness-of-fit

We propose a method to examine the goodness-of-fit of the proposed model (2) to data. Our method consists of three steps: **(i)** parametric estimation of the sub-distribution function, **(ii)** nonparametric estimation of the sub-distribution function, **(iii)** comparison of the parametric and nonparametric estimators. If the two estimators are close, we conclude that there is no evidence against the model.

(i) Parametric estimation of sub-distribution functions

Under the proposed model (2) and by the formula of $f(t,j)$, the parametric estimator of the sub-distribution function (for Event *j*) is derived as

$$
\begin{aligned}
F_{\hat{\Phi}}(t,j) = \int_0^t & \left(1+\hat{\eta}\left[\left\{\frac{s}{\exp(\hat{\mu}_1)}\right\}^{\frac{\theta+1}{\hat{\sigma}_1}}+\left\{\frac{s}{\exp(\hat{\mu}_2)}\right\}^{\frac{\theta+1}{\hat{\sigma}_2}}\right]^{\frac{1}{\theta+1}}\right)^{-\frac{1+\hat{\eta}}{\hat{\eta}}} \\
& \times\left[\left\{\frac{s}{\exp(\hat{\mu}_1)}\right\}^{\frac{\theta+1}{\hat{\sigma}_1}}+\left\{\frac{s}{\exp(\hat{\mu}_2)}\right\}^{\frac{\theta+1}{\hat{\sigma}_2}}\right]^{\frac{1}{\theta+1}-1}\frac{s^{\frac{\theta+1}{\hat{\sigma}_j}-1}}{\hat{\sigma}_j\exp\left\{\hat{\mu}_j\left(\dfrac{\theta+1}{\hat{\sigma}_j}\right)\right\}}ds, \quad j=1,2.
\end{aligned}
$$

(ii) Nonparametric estimation of sub-distribution functions

For nonparametric estimation, we let $T_{(1)} < T_{(2)} < \cdots < T_{(k)}$ be distinct uncensored times. Then, the nonparametric estimators of sub-distribution functions (Escarela and Carriere 2003) are

$$\hat{F}(t, j) = \sum_{i:T_{(i)} \le t} \hat{S}(t) \frac{d_{ij}}{n_i}, \qquad j = 1, 2$$

where $n_i = \sum_{j=1}^{n} \mathbf{I}(T_j \ge T_{(i)})$, $d_i = \sum_{j=1}^{n} \mathbf{I}(T_j = T_{(i)})$, $d_{i1} = \sum_{j=1}^{n} \delta_i \mathbf{I}(T_j = T_{(i)})$, $d_{i2} = \sum_{j=1}^{n} \delta_i^* \mathbf{I}(T_j = T_{(i)})$, and $\hat{S}(t) = \prod_{i:T_{(i)} \le t} (1 - d_i / n_i)$.

(iii) The Cramér-von Mises type statistic

$$CvM = \sum_{i=1}^{n} \delta_i \left\{ F_{\hat{\Phi}}(T_i, 1) - \hat{F}(T, 1) \right\}^2 + \sum_{i=1}^{n} \delta_i^* \left\{ F_{\hat{\Phi}}(T_i, 2) - \hat{F}(T, 2) \right\}^2.$$

A large value of CvM means the lack of fit for the model (2). In practice, a graphical diagnostic for the value of CvM is helpful. We suggest plotting $F_{\hat{\Phi}}(t, j)$ and $\hat{F}(t, j)$ on the same graph for each j. We shall demonstrate this graphical model diagnostic through the simulations below.

3. Simulations

We examine the performance of the proposed likelihood method by simulations. We generated data of size n=100 or 200 from the model (2) under some specific parameters (Table 1). We also generated independent censoring times from a uniform distribution that gives 20% censoring percentage. Under competing risks, we only use the first occurring event time $T_i = \min(T_{i1}, T_{i2}, C_i)$ and two event indicators $\delta_i = \mathbf{I}(T_i = T_{i1})$ and $\delta_i^* = \mathbf{I}(T_i = T_{i2})$. Based on the observed data $\{ (T_i, \delta_i, \delta_i^*) : i = 1, \ldots, n \}$, the MLE is obtained by "*optim()*". Our simulations are based on 500 repetitions.

Table 1 shows the simulation results. It shows that all the parameters are almost unbiasedly estimated except for η. The standard deviation (SD) of the estimates decrease as the sample size increases from n=100 to 200. The SE is close the SD, except for η. This implies some difficulty of estimating η in small samples, but the bias and SD decrease as the sample size increases. The coverage probability (CP) of the 95% CI are mostly close to 0.95.

Figure 1 shows one simulation run for estimating the sub-distribution function by the parametric and nonparametric methods. Since the data were simulated from the correct model, the two estimators of the sub-distribution functions are close to each other. This implies that the two estimators consistently estimate the true sub-distribution function. However, if the model assumptions were not correct, the parametric estimator would be inconsistent.

Table 1. Simulation results under $\theta = 6$ based on 500 repetitions.

	n	Parameter	True	Mean	SD	SE	CP%
CEN = 20%	100	μ_1	1	0.9991	0.0266	0.0247	0.9239
T1 = 60%		μ_2	1.5	1.5235	0.1981	0.1902	0.9506
T2 = 20%		σ_1	0.1	0.0974	0.0208	0.0189	0.9259
		σ_2	0.5	0.5093	0.1050	0.1059	0.9506
		η	0.5	0.5914	0.5677	0.4090	0.9362
	200	μ_1	1	0.9993	0.0182	0.0173	0.9346
		μ_2	1.5	1.5046	0.1336	0.1265	0.9366
		σ_1	0.1	0.0981	0.0128	0.0132	0.9366
		σ_2	0.5	0.4987	0.0756	0.0723	0.9448
		η	0.5	0.5443	0.2510	0.2490	0.9489

T1=the percentage that failure with Event 1. T2=the percentage that failure with Event 2. CEN=the percentage that lifetime is censored. SD=the sample standard deviation of the estimates. SE=the average of the standard error. CP%=the coverage ratio for the 95% confidence intervals.

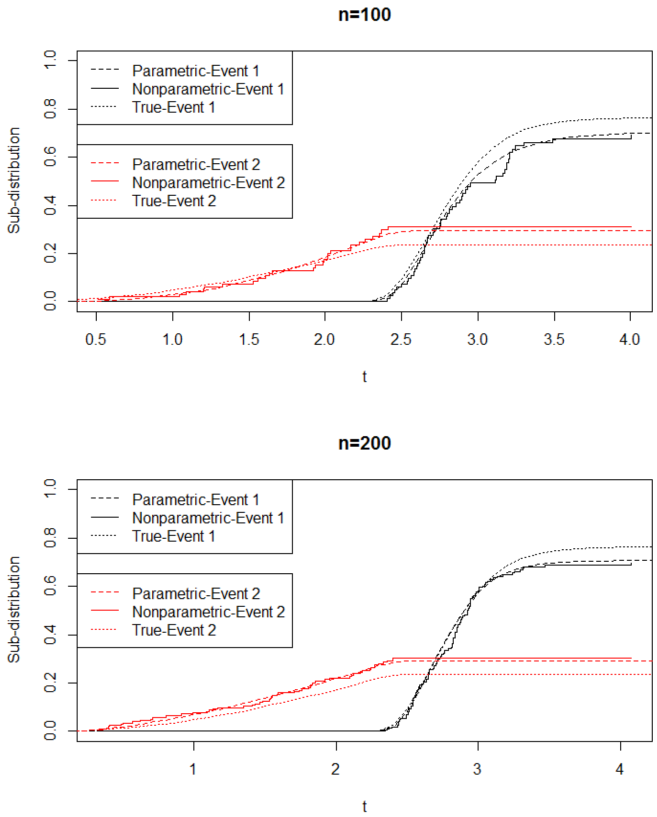

Figure 1 Estimates for the sub-distribution function under $\Phi \equiv (\mu_1, \mu_2, \sigma_1, \sigma_2, \eta, \theta)$=(1, 1.5, 0,1, 0.5, 0.5, 6) with 20% censoring.

References

[1] A.P. Basu, J.K. Ghosh. Identifiability of the multinormal and other distributions under competing risks model. Journal of Multivariate Analysis 8, 413-429, 1978

[2] R.A. Belaghi, M.N. Asl. Estimation based on progressively type-I hybrid censored data from the Burr XII distribution. Stat Pap, DOI: 10.1007/s00362-016-0849-5, 2016

[3] I.W. Burr. Cumulative frequency functions. Ann Math Stat 13(2), 215-232, 1942.

[4] H.A. David, M.L. Moeschberger. The Theory of Competing Risks. London: Griffin, 1978

[5] L. Duchateau, P. Janssen. The Frailty Model. Berlin: Springer, 2008.

[6] T. Emura, Y.H. Chen. Gene selection for survival data under dependent censoring, a copula-based approach, Stat Methods Med Res 25(6): 2840-57, 2016.

[7] T. Emura, S. Matsui, V. Rondeau, Survival Analysis with Correlated Endpoints, Joint Frailty-Copula Models, JSS Research Series in Statistics, Springer, Singapore, 2019.

[8] T. Emura, H. Michimae. A copula-based inference to piecewise exponential models under dependent censoring, with application to time to metamorphosis of salamander larvae, Environ Ecol Stat 24(1) 151–73, 2017

[9] T. Emura, M. Nakatochi, K. Murotani, V. Rondeau. A joint frailty-copula model between tumour progression and death for meta-analysis, Stat Methods Med Res 26: 2649-66, 2017.

[10] T. Emura, J.H. Shih, I.D. Ha, R. Wilke. Comparison between the marginal hazard models and sub-distribution hazard models for competing risks data, Comp Stat Data Anal, in revision, 2019.

[11] G. Escarela, J.F. Carriere. Fitting competing risks with an assumed copula. Stat Methods Med Res 12: 333-349, 2003.

[12] T.H. Fan, Y.F. Wang, S.K. Ju. A competing risks model with multiply censored reliability data under multivariate Weibull distributions, IEEE T Reliab, DOI:10.1109/TR.2019.2907518.

[13] T. M. Hsu, T. Emura, T.H. Fan. Reliability inference for a copula-based series system life test under multiple type-I censoring, IEEE T Reliab 65 (2): 1069-80, 2016.

[14] X. Liu. Planning of accelerated life tests with dependent failure modes based on a gamma frailty model. Technometrics, 54(4), 398-409, 2012.

[15] S.M. Lo, R.A. Wilke. A copula model for dependent competing risks. JRSS Series C (Applied Statistics) 59(2), 359-376, 2010.

[16] S.M. Lo, G. Stephan, R.A. Wilke. Competing risks copula models for unemployment duration: An application to a German Hartz reform. Journal of Econometric Methods, 6(1), 2017

[17] J. Mazucheli, J.A. Achcar. The Lindley distribution applied to competing risks lifetime data. Comput Methods Programs Biomed 104(2), 188-192, 2011.

[18] M. L. Moeschberger. Life tests under dependent competing causes of failure. Technometrics, 16(1), 39-47, 1974.

[19] A. Nádas. The distribution of the identified minimum of a normal pair determines' the distribution of the pair. Technometrics, 13(1), 201-202, 1971.

[20] F. Rotolo, C. Legrand, I. Van Keilegom I. A simulation procedure based on copulas to generate clustered multi-state survival data. Comput Methods Programs Biomed, 109(3), 305-312, 2013.

[21] J.H. Shih, T. Emura, Likelihood-based inference for bivariate latent failure time models with competing risks under the generalized FGM copula, Comput Stat 33(3): 1293-23, 2018.

[22] J.H. Shih, W. Lee, L.H. Sun, T. Emura, Fitting competing risks data to bivariate Pareto models, Commun Stat-Theor, doi:10.1080/03610926.2018.1425450, 2018

[23] A. Tsiatis. A nonidentifiability aspect of the problem of competing risks. Proc. Natn. Acad. Sci. USA, 72: 20-22, 1975.

[24] A. J. Watkins. An algorithm for maximum likelihood estimation in the three parameter Burr XII distribution. Comp Stat Data Anal 32(1), 19-27, 1999.

[25] C. Zhang, L. Pan, S. Wang, X. et al. An accelerated life test model for solid lubricated bearings used in space based on time-varying dependence analysis of different failure modes. Acta Astronautica 152:352-359, 2018

[26] M. Zheng, J.P. Klein. Estimates of marginal survival for dependent competing risks based on an assumed copula. Biometrika, 82(1), 127-138, 1995.

[27] Y. Zhou, Z. Lu, Y. Shi, K. Cheng. The copula-based method for statistical analysis of step-stress accelerated life test with dependent competing failure modes. Proc. Inst. Mech. Eng, Part O: Journal of Risk and Reliability, 1748006X18793251, 2018