International Journal of Scientific Research in Mathematical and Statistical Sciences **Research Paper**
Vol.6, Issue.2, pp.91-98, April (2019) **E-ISSN:** 2348-4519
DOI: https://doi.org/10.26438/ijsrmss/v6i2.9198

Analysis of Australian Twin Data by using Shared Frailty Models based on Reversed Hazard

A. Pandey[1], Lalpawimawha[2*], P.K. Misra[3]

[1] Dept. of Statistics, Central University of Rajasthan, Ajmer, India
[2]Dept. of Statistics, PUC, Mizoram University, Aizawl, India
[3]Dept. of Mathematics and Statistics, Dr Shakuntala Misra National Rehabilitation University, Lucknow, India

[*]*Corresponding Author: raltelalpawimawha08@gmail.com, Tel.: +91-98623-07640*

Available online at: www.isroset.org

Received: 21/Mar/2019, Accepted: 12/Apr/2019, Online: 30/Apr/2019

***Abstract*—**In this paper, we propose shared frailty models based on reversed hazard rate with exponentiated moment exponential distribution as baseline distribution. The Bayesian approach of Markov Chain Monte Carlo (MCMC) technique was employed to estimate the parameters involved in the models. A simulation study also performed to compare the true value and the estimated value of the parameters. Comparison with the non-parametric model also done by using Bayesian comparison techniques . We apply to the Australian twin data set and better model is suggested.

Keywords*—*Bayesian comparison technique, Exponentiated moment exponential distribution, MCMC, Reversed hazard, Shared frailty model.

I. INTRODUCTION

Frailty models provide a very useful way to introduce random effects, dependence and unobserved heterogeneity in the model for bivariate survival data. For example, in clinical studies, the drug effects or treatments are differ substantially for each patient. In that case, the patients have different frailties and that those who are most frail will die earlier than others. The most common model is based on the assumption that the frailty is a common random effect, which is acting multiplicatively on the baseline hazard function. The model assumes that the hazard function for lifetime T given an unobserved random variable U = *u* is $h(t=U) = uh_0(t)$, where $h_0(t)$ is the baseline hazard function.

In many situations, hazard rate function is not suitable for the survival data especially when the survival data is left censored. In that case, reversed hazard rate functions are more suitable. In the study of time to event or recurrence event related with survival status, it is possible to get that there are many situations where lifetime data are left censored. For example, suppose T represents the lifetime of a bulb, the manufacturing company determine that the lifetime of bulb would not last more than m months, and then the lifetime of this bulb is left censored. In that case, a reversed hazard rate is more convenient tool to analyze lifetime data than a hazard rate due to the fact that there is instability of estimators of hazard rates when lifetime data are left censored. Barlow et al. (1963) [3]proposed reversed hazard rate (RHR) as a dual to the hazard rate as m(t) = f(t)/F(t) where F(t) and f(t) represents the distribution function and the probability density function. Block et al. (1998)[4] first given the explicit expression of reverse hazard rate for lifetime T as

$$m(t) = \lim_{\Delta t \to 0} \frac{P(t - \Delta t < T < t \mid T \le t)}{\Delta t} \qquad (1.1)$$

The reversed hazard rate (RHR) specifies the instantaneous rate of death or failure at time t, given it failed before time t > 0. Thus in a small interval, $m(t)\Delta t$ is the approximate probability of failure in the interval, given failure before the end of the interval $(t-\Delta t, t]$. Andersen et al. (1993)[1], Gurler (1996)[11] and Lawless (2003)[17] have discussed the use of reversed hazard rate for the analysis of right truncated or left censored data.

Sengupta and Nanda (2011)[22] also introduced the proportional reversed hazards rate model in a semi-parametric setup. Duffy et al. (1990)[8] considered Australian Twin data that incorporates with information regarding the age at appendectomy of monozygotic (MZ) and dizygotic (DZ) twins. There have been twenty one pairs with missing age at onset and those are the left censored observations. Duffy et al. (1990)[8] excluded these left

© 2019, IJSRMSS All Rights Reserved

censored observations in the analysis. it is therefore, appropriate to model common random effect by together with those left censored observations, which can done by developing frailty models using RHR. Consequently, Sankaran and Gleeja (2008)[20] introduced frailty as a common random effect that acts multiplicatively on reversed hazard rates, that is helpful for the analysis of left censored data. In this paper, we consider shared frailty model with gamma distribution as frailty distributions and exponentiated moment exponential distribution as baseline distribution because of the flexibility in modeling real-life data. Pamdey et al. (2018) [2] also studied correlated frailty model based on logistic exponential baseline distribution.

In order to compare the proposed Bayesian comparison techniques such as Akaike information criteria (AIC), Bayesian information criteria (BIC), Deviance information criteria (DIC) and Bayes factor were used. Here, parametric and semi-parametric methods of estimation is consider, though semi-parametric estimation offers more flexibility, the parametric estimation is more powerful if the form of baseline hazard is known in advance. Besides, it provides more simple technique of estimation.

The remaining sections are categories as follows-the introduction of shared frailty model is presented in section-2. In sections 3 and 4, the baseline distribution and proposed models are given. Estimation strategies are presented in section 5, section 6 for simulation study. Sections 7 and 8 for applications of the proposed models and conclusion.

II. SHARED FRAILTY MODEL

The shared frailty models assume that the individuals within a group share the same frailty. Which creates dependency between the survival times of an individual's, it means that due to frailty or unobserved covariates the survival times are dependent. The two survival times are conditionally independent, given shared frailty. When there is no variability in the frailty variable distribution, the frailty variable is said to have degenerate distribution. Otherwise, there is positive dependence. A more detailed discussion of shared frailty models illustrated with some examples along with estimation methods can be seen in the books of Duchateau and Janssen (2008)[7], Kheri (2007)[16], Wienke (2010)[23], and Hanagal(2011)[12].

Suppose there are n individuals under study having first and second survival times (T_{1k}, T_{2k}) (k = 1, 2,…, n). Suppose X0 is common covariate, X1 and X2 are the observed covariates incorporates with T_{1k} and T_{2k}. Let a vector $X_{lk}=(X_{1lk},\ldots,X_{klk})$, (l = 0,1,2) for the k^{th} individual where X_{alk}(a = 1,2,3,…,m) represents the value of the a^{th} observed covariate for the k^{th} individual. Here we assume that the first and the second survival times for each individual share the same value of the covariates. Let U_k be shared frailty for the k^{th} individual. Assuming that the frailties are acting multiplicatively on the baseline RHR and both the survival times of individuals are conditionally independent for a given frailty, the conditional RHR for the k^{th} individual at the j^{th}(j = 1, 2) survival time t_{jk} for given frailty $U_k = u_k$ has the form:

$$m(t_{jk} \mid u_k, \underline{X}_k) = u_k m_0(t_{jk}) e^{\underline{X}_{0k}\underline{\beta}_0 + \underline{X}_{jk}\underline{\beta}_j}, \; j = 1,2 \quad (2.1)$$

where m0(t) is baseline reversed hazard at time $t_{jk} > 0$ and β is a vector of order k of regression coefficients. The conditional cumulative RHR for kth individual at j^{th} lifetime $t_{jk} > 0$ for a given frailty $U_k = u_k$ is

$$M(t_{jk} \mid u_k, \underline{X}_k) = u_k M_0(t_{jk}) \eta_{0k} \eta_{jk} \quad (2.2)$$

where $\eta_{0k} = e^{\underline{X}_{0k}\underline{\beta}_0}$, $\eta_{jk} = e^{\underline{X}_{jk}\underline{\beta}_j}$ and $M_0(t_{jk})$ is cumulative baseline RHR at time $t_{jk} > 0$. The conditional distribution function for j^{th} individual at k^{th} lifetime $t^{jk} > 0$ for given frailty $U_k = u_k$ is

$$F(t_{jk} \mid u_k, \underline{X}_k) = e^{M(t_{jk}|u_k, \underline{X}_k)}$$

$$= e^{-u_k M_{jk} \eta_{0k} \eta_{jk}} \quad (2.3)$$

Under the assumption of independence, the bivariate conditional distribution function for a given frailty $U_k = u_k$ at time t_{1k} and t_{2k} is

$$F(t_{1k}, t_{2k} \mid u_k, \underline{X}_k) = F(t_{1k} \mid u_k, \underline{X}_k) F(t_{2k} \mid u_k, \underline{X}_k)$$

$$= e^{-u_k (M_{01}(t_{1k})\eta_{1k} + M_{02}(t_{2k})\eta_{2k})\eta_{0k}} \quad (2.4)$$

The unconditional bivariate distribution function can be obtained by integrating over the frailty variable U_k having the probability function $f_U(u_j)$, for the k^{th} individual:

$$F(t_{1k}, t_{2k} \mid \underline{X}_k) = \int_{U_k} F(t_{1k}, t_{2k} \mid u_k) f_U(u_k) du_k$$

$$= \int_{U_k} e^{-u_k (M_{01}(t_{1k})\eta_{1k} + M_{02}(t_{2k})\eta_{2k})\eta_{0k}} f_U(u_k) du_k$$

$$= L_{U_k}\left[(M_{01}(t_{1k})\eta_{1k} + M_{02}(t_{2k})\eta_{2k})\eta_{0k}\right] \quad (2.5)$$

where L_{Uk} is the Laplace transform of the frailty variable U_k for the k^{th} individual. Here onwards we represent $F(t_{1k},t_{2k}|X_k)$ as $F(t_{1k}, t_{2k})$.

© 2019, IJSRMSS All Rights Reserved

III. SHARED GAMMA FRAILTY MODEL

In this paper, we consider gamma distribution as the frailty distribution because as the gamma variates are positive, it fits the non-negative criterion of frailties with no transformation.

The gamma distribution is one of the most commonly used distributions to model variables that are necessarily positive. Gamma distributions have been used for many years to generate mixtures in exponential and Poisson models. The popularity of the model is due to the fact that the model functions are very easy to derive because of the simplicity of the derivatives of the Laplace transform. The cross ratio function is constant for the gamma frailty model (see Clayton, 1978[5]). Sankaran and Gleeja (2006)[19] introduced a measure of association based on RHR in a similar manner as was introduced by Clayton (1978)[5] based on hazard rate. In case of gamma frailty, the measure of association given by Sankaran and Gleeja (2006)[19] is also constant. Assume that a common continuous random variable U follows a gamma distribution. For identifiability, we assume U has expected value equal to one. Under this restriction, the density function and the Laplace transformation of a gamma distribution reduces to

$$f(v)=\begin{cases}\dfrac{1}{\Gamma(\frac{1}{\zeta})}\dfrac{1}{\zeta}^{\frac{1}{\zeta}}v^{\frac{1}{\zeta}-1}e^{-\frac{v}{\zeta}};v>0,\xi>0\\ 0 \qquad\qquad ;otherwise\end{cases}$$

and $L_U(s)=(1+\zeta s)^{1/\zeta}$ with variance of U as ζ . The frailty variable U is degenerate at U = 1 when tends to zero. Replacing the Laplace transform in Eq. (2.5), we get the unconditional bivariate distribution function for the kth individual as

$$F(t_{1k},t_{2k})=[1+\zeta\eta_{0k}(M_{01}(t_{1k})\eta_{1k}+M_{01}(t_{2k})\eta_{2k})]^{1/\zeta} \quad (2.7)$$

where $M_{01}(t_{1k})$ and $M_{02}(t_{2k})$ are the cumulative baseline reversed hazard functions of the lifetime T_{1k} and T_{2k}, respectively.

The bivariate distribution in the presence of covariates, when the frailty variable is degenerate is given by

$$F(t_{1k},t_{2k})=e^{-\eta_{0k}(M_{01}(t_{1k})\eta_{1k}+M_{01}(t_{2k})\eta_{2k})} \quad (2.8)$$

IV. METHODOLOGY

Relevant details should be given including experimental design and the technique(s) used along with appropriate statistical methods used clearly along with the year of experimentation (field and laboratory).

V. BASELINE DISTRIBUTION

Exponentiated Moment Exponential Distribution

If a continuous random variable T follows the two-parameter exponentiated moment exponential distribution (EMED) proposed by Hasnain et al. (2015)[13], then the cumulative distribution function, hazard function and cumulative hazard function are, respectively,

$$F(t)=\left[1-\left(1+\frac{t}{\lambda}\right)e^{-\frac{t}{\lambda}}\right]^{\alpha},\ t>0,\alpha>0,\lambda>0 \quad (3.1)$$

$$m(t)=\frac{\frac{\alpha t}{\lambda^2}\left[1-\left(1+\frac{t}{\lambda}\right)e^{-\frac{t}{\lambda}}\right]^{\alpha-1}e^{-\frac{t}{\lambda}}}{1-\left[1-\left(1-\frac{t}{\lambda}\right)e^{-\frac{t}{\lambda}}\right]^{\alpha}},\ t>0 \quad (3.2)$$

$$M(t)=-\log\left[1-\left(1-\frac{t}{\lambda}\right)e^{-\frac{t}{\lambda}}\right]^{\alpha},\ t>0 \quad (3.3)$$

where α and λ are the scale and shape parameters of EMED. The hazard function m(t) is decreasing with time for α < 0:5 at first (Burn-in), then remains constant with respect to time(useful-life). Monotonically increasing at α = 0.5 and the hazard function m(t) is increasing with time for α > 0.5 at first then remains constant with respect to time. The hazard function m(t) is an increasing function of t for $\alpha \geq 1$. Furthermore, $m(t)\to 0$ when $t\to 0$, α > 1 and $m(t)\to 1/\lambda$ when $t\to\infty$.

VI. PROPOSED MODELS

Substituting cumulative reverse hazard function for the exponentiated moment exponential distribution in equations (2.7) and (2.8), we get the unconditional bivariate survival

functions at time t1k > 0 and t2k > 0 as,

$$S(t_{1k},t_{2k})=\left[1+\zeta(\eta_{1k}(-\log[1-(1+\frac{t_{1k}}{\lambda_1})e^{-\frac{t_{1k}}{\lambda_1}}]^{\alpha})+\eta_{2k}(-\log[1-(1+\frac{t_{2k}}{\lambda_2})e^{-\frac{t_{2k}}{\lambda_2}}]^{\alpha}))\eta_{0k}\right]^{-1/\zeta} \quad (4.1)$$

$$S(t_{1k},t_{2k})=\exp\left\{(\eta_{1k}(-\log[1-(1+\frac{t_{1k}}{\lambda_1})e^{-\frac{t_{1k}}{\lambda_1}}]^{\alpha})+\eta_{2k}(-\log[1-(1+\frac{t_{2k}}{\lambda_2})e^{-\frac{t_{2k}}{\lambda_2}}]^{\alpha}))\eta_{0k}\right\} \quad (4.2)$$

© 2019, IJSRMSS All Rights Reserved

Here onwards we call equation (4.1) and (4.2) as Model I and Model II respectively. Model I is model with frailty and Model II is model without frailty. The semi parametric model with frailty is given by

$$S(t_{1k},t_{2k})=\exp\left\{-u\eta_{0k}(\eta_{1k}(-\log[1-(1+\frac{t_{1k}}{\lambda_1})e^{-\frac{t_{1k}}{\lambda_1}}]^{\alpha})+\eta_{2k}(-\log[1-(1+\frac{t_{2k}}{\lambda_2})e^{-\frac{t_{2k}}{\lambda_2}}]^{\alpha}))\right\} \quad (4.3)$$

and the semi-parametric model without frailty is given by

$$S(t_{1k},t_{2k})=\exp\left\{-\eta_{0k}(\eta_{1k}(-\log[1-(1+\frac{t_{1k}}{\lambda_1})e^{-\frac{t_{1k}}{\lambda_1}}]^{\alpha})+\eta_{2k}(-\log[1-(1+\frac{t_{2k}}{\lambda_2})e^{-\frac{t_{2k}}{\lambda_2}}]^{\alpha}))\right\} \quad (4.4)$$

for more details see Hanagal (2011). Here onwards we call Eq. (4.3) as Model III, for U is gamma frailty and Eq. (4.4) as Model IV.

VII. ESTIMATION STRATEGIES AND MODEL COMPARISION

The likelihood function can be obtained by blending the failure times of the k^{th} individuals (k = 1, 2, 3. . . n) and censoring times by assuming independence between censoring scheme and individuals lifetimes and is given by

$$L(\underline{\Psi},\underline{\beta},\zeta)=\prod_{k=1}^{n_1}f_1(y_{1k},y_{2k})\prod_{k=1}^{n_2}f_2(y_{1k},d_{2k})\prod_{k=1}^{n_3}f_3(d_{1k},y_{2k})\prod_{k=1}^{n_4}f_4(d_{1k},d_{2k}) \quad (21)$$

where $\underline{\Psi}$, β and ζ are vectors of baseline parameters, regression coefficients and frailty distribution parameter. The likelihood function for without frailty is given as

$$L(\underline{\Psi},\underline{\beta})=\prod_{k=1}^{n_1}f_1(y_{1k},y_{2k})\prod_{k=1}^{n_2}f_2(y_{1k},d_{2k})\prod_{k=1}^{n_3}f_3(d_{1k},y_{2k})\prod_{k=1}^{n_4}f_4(d_{1k},d_{2k}) \quad (21)$$

where n_1, n_2, n_3 and n_4 are the random number of observations observed to lie in the range (y_{1k}, y_{2k}) lie in the ranges $y_{1k} < d_{1k}$, $y_{2k} < d_{2k}$; $y_{1k} < d_{1k}$, $y_{2k} > d_{2k}$; $y_{1k} > d_{1k}$, $y_{2k} < d_{2k}$ and $y_{1k} > d_{1k}$, $y_{2k} > d_{2k}$ respectively and the contribution of the k^{th} individual in the likelihood function as

$$\begin{aligned} f_1(y_{1k},y_{2k}) &= \frac{\partial^2 S(y_{1k},y_{2k})}{\partial y_{1k}\partial y_{2k}} \\ f_2(y_{1k},d_{2k}) &= -\frac{\partial^2 S(y_{1k},d_{2k})}{\partial y_{1k}} \\ f_3(d_{1k},y_{2k}) &= \frac{\partial^2 S(d_{1k},y_{2k})}{\partial y_{2k}} \\ f_4(d_{1k},d_{2k}) &= S(d_{1k},d_{2k}) \end{aligned} \quad (22)$$

Putting equation (22) in equations (20) and (21), we get the likelihood function for the mixture shared frailty model and shared frailty model.

The joint posterior density of the parameters given failure times is given as

$$\pi(\alpha_1,\lambda_1,\alpha_2,\lambda_2,\zeta,\underline{\beta})\propto L(\alpha_1,\lambda_1,\alpha_2,\lambda_2,\zeta,\underline{\beta})\times p_1(\alpha_1)p_2(\lambda_1)p_3(\alpha_2)p_4(\lambda_2)p_5(\zeta)\prod_{i=1}^{5}g_i(\underline{\beta_i})$$

,

where $p_i(.)(i = 1,\ldots,5)$ indicates the prior density function with known hyper parameters of corresponding arguments for baseline parameters and frailty variance; $g_i(.)$ is prior density function for regression coefficient β_i; $\underline{\beta_i}$ represents a vector of regression coefficients except β_i, i = 1, 2,…,a and likelihood function (.) is given by equation (19) or (20). Here it is assumed that all the parameters are independently distributed.

The expression of the likelihood function in equations (20) and (21) are not easy to solve by using the Newton-Raphson method. MLEs fail to converge as it involved a large number of parameters. Therefore, the Bayesian approach was utilized to estimate the parameters involved in the models, which does not endure any such kind of troubles.

Prior distributions are used as follows - gamma distribution with mean 1 and large variance $G(\Psi, \Psi)$ is used as a prior distribution for frailty parameter with a small value of Ψ. Normal distribution with mean zero and large variance is used as prior for the regression coefficient, say ϕ^2. The same type of prior distributions considered in Ibrahim et al. (2001)[14] and Sahu et al. (1997)[18] and non- informative prior assumed as the baseline parameters since we do not have any information about the baseline parameters. $G(a_1, b_1)$ and $U(a_2, b_2)$ are used as non-informative prior distributions. All the hyper-parameters Ψ, ϕ, a_1, a_2, b_1 and b_2 are assumed to be known. Here $G(a_1, b_1)$ represents gamma distribution with shape parameter a_1 and scale parameter b_1 and $U(a_2, b_2)$ is the uniform distribution over the interval a_2 to b_2. We set hyper-parameters as $\Psi = 0.0001$, $\phi^2 = 1000$, $a_1 = 1$, $b_1 = 0.0001$, $a_2 = 0$, and $b_2 = 100$.

To estimate the parameters in the models fitted with the above prior density function and likelihood equations (20) and (21), Metropolis-Hasting Algorithm and Gibbs Sampler was utilized. The convergence of the Markov chain to a stationary distribution is also observed by Geweke test and Gelman-Rubin Statistics as suggested by Geweke (1992)[10] and Gelman et al. (1992)[9]. To check the behavior of the chain, to decide burn-in period and autocorrelation lag, we used trace plots, coupling

© 2019, IJSRMSS All Rights Reserved

from the past plots and sample autocorrelation plots respectively.

It is important to decide which model provides the best fit to the dataset; the model comparison was done by using Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), Deviance Information Criteria (DIC) and Bayes factor.
Suppose there are P parameters in a model and n observations in a dataset. AIC, BIC and DIC are elucidated as

$$\mathrm{AIC} = -2log\mathrm{L}(y|\hat{\theta}) + 2P, \quad (23)$$

$$\mathrm{BIC} = -2log\mathrm{L}(y|\hat{\theta}) + log(n)P, \quad (24)$$

$$\mathrm{DIC} = -2log\mathrm{L}(y|\hat{\theta}) + 2P_{D,} \quad (25)$$

where $P_D = E[2\log L(y \mid \theta)] - 2\log L(y \mid \hat{\theta})$

Smaller values of AIC, BIC and DIC for the models are considered as better models than higher values.
Bayes factor also employed for selection of Model M_u against Model M_v and defined as

$$BF_{uv} = \frac{P(y \mid M_u)}{P(y \mid M_v)},$$

$$P(y \mid M_k) = \int_Q P(y \mid \Omega, M_k)\pi(\Omega \mid M_k)d\Omega, \quad (26)$$

where Ω represents the number of unknown parameters of the model M_k, $\pi(\Omega|M_k)$ is the density of prior distribution of the model M_k and Q is the bolster of the parameters Ω.

In spite of the fact that, $2\log BF_{uv}$ is roughly equal to the differences in the values of BIC for the given models, we utilized the strategy given by Kass and Raftery (1995)[15], to compute P (y M) from the MCMC sample gotten from each of the model parameters.

$$P(y \mid M_k) = \left(\frac{\sum_{k=1}^{N} L(y \mid \Omega^k)^{-1}}{N} \right)^{-1},$$

where Ω^k and N symbolize sample and sample size of the posterior distribution.

A value of $2log\mathrm{BF}_{uv}$ more than 10 shows that greatly strong positive to favor model M_u over model M_v, whereas a value between 0 and 2 is adequate to prove to favor not one or the other model. A value between 2 and 6 or 6 and 10 shows a mellow or modestly strong confirmation respectively, to favor the numerator model.

VIII. SIMULATION

To evaluate the performance of the Bayesian estimation procedure we carry out a simulation study. For the simulation purpose we have considered only one covariate X_0 which we assume to follow normal distribution for Model I. The frailty variable U is assumed to have gamma distribution for Model I with known variance. Lifetimes (T_{1k}, T_{2k}) for the k^{th} individual are conditionally independent for a given frailty $U_k = u_k$. We assume that T_{jk}(j = 1, 2; k = 1,2,...,n) follows one of the baseline distribution exponentiated moment exponential distribution.

A widely used prior for the frailty parameter ζ is the G(0.0001, 0.0001). In addition, we assume that the prior for regression coefficient is N(0, 1,000). Similar types of the prior distributions are used in Ibrahim et al. (2001), Sahu et al. (1997), and Santos and Achcar (2010). We also employ the same non-informative prior for the frailty parameter and the regression coefficients. Since we do not have any prior information about the baseline parameters, the prior distributions are assumed to be at. We consider two different non-informative prior distributions for the baseline parameters, one is $G(a_1, b_2)$ and another is $U(a_2, b_2)$. All the hyper-parameters a_1, a_2, b_1, and b_2 are known. Here G(a, b) is the gamma distribution with the shape parameter a and the scale parameter b and $U(a_2,b_2)$ represent the uniform distribution over the interval a_2 to b_2. For Model I, we set α_1 = 1.2, λ_1 = 8.12, α_2 = 1.2, λ_2 = 8.12, ζ = 0.0025, and β = 0.033 and X~Normal(12, 0.45), censoring distribution as the exponential distribution with the parameter 0.02 each. We assume the value of the hyper-parameters as a_1 = 1, b_1 = 0.0001, a_2= 0, and b_2 = 100.

Table 1: Simulation study for Model-I

Parameter (value)	Estimate	SE	LCL	UCL	Geweke values	P values	GR values
burn in period = 6800 ; autocorrelation lag = 320							
α_1 (2.0010)	2.0017	0.0567	1.9034	2.0934	0.0129	0.5051	1.0003
α_2 (0.1380)	0.1372	0.0423	0.0909	0.2492	0.0115	0.5046	1.0003
λ_1 (0.1810)	0.1816	0.0525	0.1115	0.2980	-0.0001	0.4999	1.0034
λ_2 (4.7500)	4.7592	0.5405	4.0073	5.7916	-0.0151	0.4939	1.0011
ζ(0.0018)	0.0018	0.0005	0.0010	0.0028	0.0100	0.5040	1.0054
β (0.0007)	0.0007	5.4e-05	0.0006	0.0008	-0.0055	0.4977	1.0003

We run two parallel chains for Model I using two sets of the prior distributions with the different starting points using the Metropolis-Hastings algorithm and the Gibbs sampler based on normal transition kernels. We iterate both the chains for 100,000 times. There is no effect of the prior distribution on the posterior summaries because the estimates of parameters are nearly the same and the convergence rates of the Gibbs sampler for both the prior sets are almost the same. Also for both the chains the results were somewhat similar. For all models, the trace plots, the coupling from the past plots, the

© 2019, IJSRMSS All Rights Reserved

running mean plots, and the sample autocorrelation plots for the simulation study are not provided due to lack of space. Table 1 present the estimates, the credible intervals, the Geweke test (Geweke 1992[10]), and the GelmanRubin statistics (Gelman and Rubin, 1992[9]) for all the parameters of the Model I based on the simulation study.

Estimated values of the parameters are close to the true values and the biases for the estimated values are small. Standard errors for the estimated values are quite small. The Gelman-Rubin convergence statistic values are nearly equal to one and also the Geweke test values are quite small and the corresponding p-values are large enough to say that the chain attains stationary distribution.

IX. ANALYSIS OF AUSTRALIAN TWIN DATA

Now we apply the proposed models to the Australian twin data given in Duffy et a1. (1990)[8]. The data consists of six zygote categories. We consider the subset of the data with zygote category 2. The data consists of male's gender only and consist of 567 pair of twins with 23 and 17 censored in twin 1 and twin 2, respectively. An individual's having age at onset less than 11 are considered as left censored observations. The data has information on the age at appendectomy of twins. The genetic effect involved in the risk of appendectomy is the frailty variable. Here there is a common covariate age for both T_1 and T_2 and one covariate each for T_1, T_2, i.e., presence or absence of appendectomy. To check goodness of fit of Australian twin data set, we obtain Kolmogorov-Smirnov (KS) statistics and their p-values for T_1 and T_2 separately. For Model I observe that p-values for lifetimes T_1 and T_2 are 0.69714 and 0.47366, respectively. For Model II observe that p-values for lifetimes T_1 and T_2 are 0.71885 and 70278, respectively. Thus from p-values of KS test are quite high. We can say that there is no statistical evidence to the reject the hypothesis that data are from these two models, in the univariate case and we assume that they also fit for bivariate case. Figure 1 shows the plot of the KaplanMeier estimates, and hypothesized theoretical distribution for the marginal distributions of Models I and II, respectively, demonstrates consistency between the two, i.e. non parametric and parametric survival curves, and assumes the same case with bivariate case also.

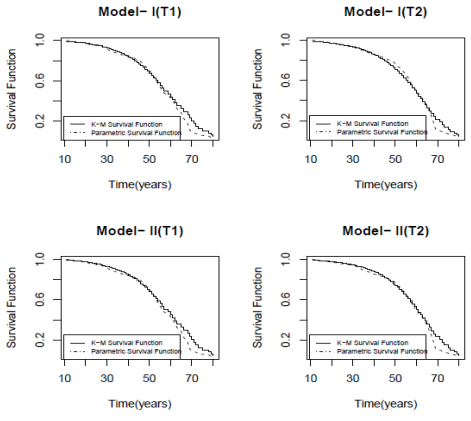

Figure 1: Non parametric versus parametric survival function for Australian twin data set.

As in the case of simulation, here also we assume the same set of prior distributions. We run two parallel chains for all models using two sets of prior distributions with the different starting points using the Metropolis-Hastings algorithm and the Gibbs sampler based on normal transition kernels. We iterate both the chains for 100,000 times. As seen in simulation study, here also we got nearly same estimates of parameters for both the set of priors, so estimates are not dependent on the different prior distributions. Convergence rates of Gibbs sampler for both the prior sets are almost the same. Also both the chains show somewhat similar results, so we present here the analysis for only one chain with G(1, 0.0001) as prior for the baseline parameters and G(0.0001, 0.0001) as the prior for the frailty parameter . Due to lack of space, we are presenting only for Model I (trace plots, coupling from the past plots, autocorrelation plots, autocorrelation plots after thinning and running mean plots) for the parameters in Figs. 2(a)-2(d). The Gelman-Rubin

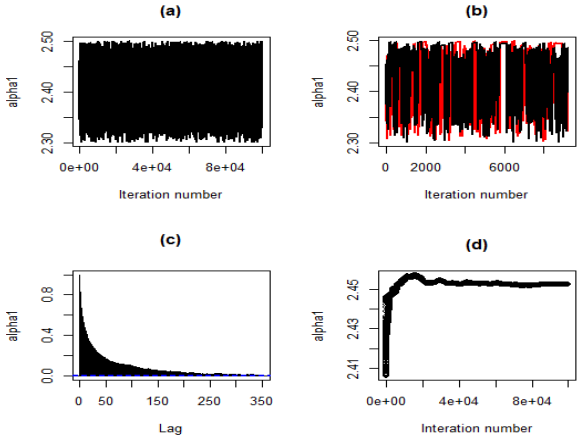

Figure 2: (a) Trace plot (b) Coupling from the past plot (c) ACF plot (d) Running mean plot for Model I.

Table 2: Posterior summary for Australian data set for Model-I

Parameter	Estimate	SE	LCL	UCL	Geweke values	P values	GR values
burn in period = 9200; autocorrelation lag = 350							
α_1	2.4484	0.0429	2.3412	2.4974	0.0034	0.5013	1.0001
λ_1	8.1946	0.1789	7.8445	8.5642	0.0023	0.5009	1.0001
α_2	0.8981	0.0482	0.8068	0.9845	0.0026	0.5010	1.0029
λ_2	9.9627	0.3249	9.3118	10.617	0.0004	0.5001	1.0035
ζ	0.0018	0.0004	0.0010	0.0027	-0.0016	0.4993	1.0003
β_0	0.0310	0.0015	0.0279	0.0337	-0.0033	0.4986	1.0000
β_1	-1.5930	0.0509	-1.6720	-1.4775	0.0014	0.4986	1.0003
β_2	0.0020	0.0004	0.0011	0.0030	-0.0014	0.4994	1.0001

© 2019, IJSRMSS All Rights Reserved

Table 3: Posterior summary for Australian data set for Model-II

Parameter	Estimate	SE	LCL	UCL	Geweke values	P values	GR values
burn in period = 9800; autocorrelation lag = 360							
α_1	2.3903	0.0368	2.2897	2.4289	0.0063	0.5025	1.0002
λ_1	8.2737	0.0555	8.1783	8.3663	0.0055	0.5022	1.0010
α_2	1.2320	0.0495	1.1559	1.3359	0.0103	0.5041	1.0002
λ_2	8.6888	0.0488	8.58.9	8.7561	-0.0041	0.4983	1.0001
β_0	0.0300	0.0010	0.0279	0.00321	-0.0026	0.4989	1.0004
β_1	-1.5705	0.0538	-1.6513	-1.4553	-0.0150	0.4989	1.0001
β_2	0.0020	0.0005	0.0011	0.0030	-0.0040	0.4983	1.0034

convergence statistic values are nearly equal to one and the Geweke test statistic values are quite small and the corresponding p-values are large enough to say the chains attain stationary distribution. The posterior mean and standard error with 95% credible intervals for the baseline parameters, the frailty parameter and the regression coefficients are presented in Tables 2-5. The posterior summary of all the four

Table 4: Posterior summary for Australian data set for Model-III

Parameter	Estimate	SE	LCL	UCL	Geweke values	P values	GR values
burn in period = 6500; autocorrelation lag = 300							
ζ	0.5202	0.0423	0.4357	0.6064	0.0120	0.5047	1.0025
β_0	0.0244	0.0019	0.0205	0.0285	-0.0033	0.4986	1.0004
β_1	-1.6707	0.0584	-	-	-0.0018	0.4986	1.0005
β_2	0.0025	0.0005	1.7573	1.5328	0.0075	0.5030	1.0010
			0.0016	0.0034			

Table 5: Posterior summary for Australian data set for Model-IV

Parameter	Estimate	SE	LCL	UCL	Geweke values	P values	GR values
burn in period = 7900; autocorrelation lag = 290							
β_0	0.0294	0.0010	0.0275	0.0314	-0.0064	0.4974	1.0001
β_1	-1.4908	0.0455	-	-	0.0022	0.4974	1.0003
β_2	0.0031	0.0005	1.5583	1.3901	0.0078	0.5031	1.0000
			0.0022	0.0040			

models are given in Tables 2-5. Tables 2-5 present the estimates, the credible intervals, the Geweke test, and the GelmanRubin statistics for all the parameters of the all the four models based on data. For Model I and Model II, the estimates of the shared frailty parameter (ζ) are, respectively, 0.0018, 0.5202. This shows that there is heterogeneity between the pairs of twins. Bayes factor for Model I against Model II is 13.8990, Model I against Model III is 83.6493 and Model I against Model IV is 11.4289. This is also a Bayesian test based on Bayes factor for testing $\zeta = 0$ against $\zeta > 0$ and which supports the alternative hypothesis, i.e., models with frailty fit better. The credible interval of the regression coefficient, β_0 , β_1 and β_2 does not contain zero for all models. Hence age of the patient and presence and absence of appendectomy are the significant covariates for all the models. The age at onset affects positively as the age of patient increases as the presence and absence of appendectomy on the age at onset. To compare six models, we first use AIC, BIC, and DIC values which are given in Table 6 and Bayes factor in Table 7. The AIC, BIC, and DIC values for Model I is least among all four models. On the basis of AIC, BIC, and DIC values, Model I is the best among all four models. Similarly the Bayes factors show that models with frailty (Model I and Model III) are better than the models without frailty. The parametric models are also better than semi-parametric models.

X. CONCLUSION

Our main aim of the study is to examine the role of the shared frailty model based on the RHR in the survival studies. For this, we used the shared gamma frailty model with the exponentiated moment exponential distribution as a baseline distribution and these models are compared with their baseline model based on the RHR. We also compare the parametric models with the semi-parametric models. We found that the parametric models are better than the semi-parametric models. We also found that the shared frailty models are better models as compared to their baseline model on the basis of AIC, BIC, and DIC values for Australian twin data set. The Bayes factors also support the shared frailty models. The simulation results indicate that the performance of the Bayesian estimation method is quite satisfactory. Bayes factor is used to test the frailty parameter $\zeta = 0$ and it is observed that the frailty parameter is significant in all frailty models.

The choice of the best model for Australian twin data is based on AIC, BIC, DIC, and the Bayes factor values. We found that Model I (shared frailty model based on reversed hazard rate) is a best model on the basis of AIC, BIC, DIC, and the Bayes factor values. The age of the patient and presence and absence of appendectomy are the significant

Table 6: AIC, BIC and DIC values for all the models fitted to Australian data set.

Model No.	AIC	BIC	DIC
Model I	8611.658	8646.381	8603.539
Model II	8626.130	8656.512	8615.897
Model III	8621.069	8655.792	8613.178
Model IV	8625.269	8655.651	8614.949

Table 7: Bayes factor values and decision for test of significance fitted to Australian data set

numerator model against denominator model	$2\log_e(B_{uv})$	range	Evidence against model in denominator
M_I against M_{II}	13.8990	> 10	Very Strong Positive
M_I against M_{III}	83.6493	> 10	Very Strong Positive
M_I against M_{IV}	11.4289	> 10	Very Strong Positive
M_{II} against M_{III}	69.7503	> 10	Very Strong Positive
M_{IV} against M_{II}	2.4701	≥ 2 and < 6	Positive
M_{III} against M_{IV}	72.2204	> 10	Very Strong Positive

© 2019, IJSRMSS All Rights Reserved

covariates for all models. The age at onset affects positively as the age of patient increases and also effected by presence and absence of appendectomy. By referring all the above analysis now we are in a position to say that, the shared gamma frailty model based on the RHR with the exponentiated moment exponential baseline is more suitable model for Australian twin data set, with left censored observations. It is also worth to be mentioned that shared gamma frailty model with exponentiated moment exponential is fit better to Australian twin data than Hanagal and Pandey (2016)[6]. The methods discussed in this paper may be extended into other frailty models and correlated frailty models with different baseline distributions, using the Bayesian approach, provided the models fit to the data.

REFERENCES

[1] P.K. Andersen, O. Borgan, R.D. Gill, N. Keiding, "*Statistical Methods Based on Counting Processes*", NewYork, Springer Verlag, **1993**.

[2] A. Pandey, Lalpawimawha , S. Bhushan, P.K. Misra, "Correlated Gamma Frailty Model Based on Logistic Exponential Baseline Distribution", *International Journal of Scientific Research in Mathematical and Statistical Sciences,* Vol. **5**, pp. **170-176, 2018.**

[3] R.E. Barlow, A.W. Marshal, F. Proschan, F," *Properties of the probability distribution with monotone hazard rate*", Ann. Math. Stat. . Vol. **34**, pp. **375-389**, **1963**.

[4] H.W. Block, T.H. Savits, H. Singh, H," *On the reversed hazard rate function*", Probab. Eng. Inf. Sci., Vol. **12**, pp. **69-90**, **1998**.

[5] D.G. Clayton, "*A model for association in bivariate life tables and its applications to epidemiological studies of familial tendency in chronic disease incidence*", Biometrica. Vol. **65**,pp. **141-151**, **1978**.

[6] D.D. Hanagal, A. Pandey,"*Gamma shared frailty model based on reversed hazard rate*", Communications in Statistics - Theory and Methods. Vol. **45(7)**, pp. **2071-2088**, **2016**.

[7] L. Duchateau, P. Janssen, "*The Frailty Model*", Springer, New York, **2008**.

[8] D.L. Duffy, N.G. Martin, J.D. Mathews,"*Appendectomy in Australian twins*", Aust. J. Hum. Genet., Vol. **47(3),**pp. **590-592**, **1990**.

[9] A. Gelman, D.B. Rubin,"*A single series from the Gibbs sampler provides a false sense of security*",In Bayesian Statistics 4 (J. M. Bernardo, J. 0.Berger, A. P. Dawid and A. F. M. Smith, eds.), Oxford: Oxford Univ. Press, pp. **625-632**, **1992**.

[10] J. Geweke,"*Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments*", In Bayesian Statistics 4 (eds. J.M. Bernardo, J. Berger, A.P. Dawid and A.F.M. Smith), Oxford: Oxford University Press, pp. **169-193**, **1992**.

[11] U. Gurler,"*Bivariate Estimation With Right-Truncated Data*",Journal of the American Statistical Association, Vol. **91(435)**, pp. **1152-1165**, **1996**.

[12] D.D. Hanagal,"*Modeling Survival Data Using Frailty Models*", Boca Raton,Chapman& Hall/CRC, **2011**.

[13] S.A. Hasnain, Z. Iqbal, M. Ahmad,"*On exponentiated moment exponential distribution*",Pakistan Journal of Statistics, Vol. **31(2)**, pp. **267-280**, **2015**.

[14] J.G. Ibrahim, M.H. Chen, D. Sinha, " *Bayesian Survival Analysis*",Springer, Verlag, **2001**.

[15] R.A. Kass, A.E. Raftery,"*Bayes' factor*",Journal of the American Statistical Association, Vol. **90**,pp. **773-795**, **1995**.

[16] S. Kheiri, A. Kimber, M.R. Meshkani,"*Bayesian analysis of an inverse Gaussian correlated frailty model*", Computational Statistics and Data Analysis, Vol. **51**,pp. **5317-5326**, **2007**.

[17] J.F. Lawless,"*Statistical Models and Methods for Lifetime Data*", (2nd ed.),New Jersey: John Wiley and Sons, **2003**.

[18] S.K. Sahu, D.K. Dey, H. Aslanidou, D. Sinha,"*A Weibull regression model with gamma frailties for multivariate survival data*", Life time data analysis, Vol. **3**,pp. **123- 137**, **1997**.

[19] P.G. Sankaran, V.L. Gleeja,"*On bivariate reversed hazard rates*", J. Jpn. Stat. Soc., Vol. **10**,pp. **181-198**, **2006**.

[20] P.G. Sankaran, V.L. Gleeja,"*Proportional reversed hazard and frailty models*", Metrika , Vol. **68**,pp. **333-342**, **2008**.

[21] C.A. Santos, J.A. Achcar,"*A Bayesian analysis for multivariate survival data in the presence of covariates*", Journal of Statistical Theory and Applications, Vol. **9**,pp. **233-253**, **2010**.

[22] D. Sengupta, A.K. Nanda,"*The proportional reversed hazards regression model*",Statistical Theory and Applications, Vol. **18**,pp. **461-476**, **2011**.

[23] A. Wienke,"*Frailty Models in Survival Analysis*" New York, CRC Press, **2010**.

AUTHORS PROFILE

Dr Arvind Pandey is currently working as assistant Professor , Department of Statistics at Central University of Rajasthan. He was working as assistant professor and head in the department of statistics, Pachhunga University College since 2005. He has more than 20 publications

Mr.Lalpawimawha is currently working as assistant professor, Department of Statistics, Pachhunga University College, Mizoram University, Aizawl, Mizoram. He was working in Centre for applied Mathematics, Central University of Jharkhand during 2011-2012. He has more than 15 publications.

Dr Praveen Kumar Misra is assistant professor, Department of Mathematics and Statistics, Dr Shakuntala Misra National Rehabilitation University, Lucknow , Uttar Pradesh. He has more than 10 years teaching experience and has many publications

© 2019, IJSRMSS All Rights Reserved